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

Richard Arnott[*], Anatolii Kokoza[†] and Mehdi Naji[‡]

July 26, 2016

Abstract

For a quarter century, a top priority in transportation economic theory has been to develop models of rush-hour traffic dynamics that incorporate hypercongestion – situations of heavy congestion where throughput decreases as traffic density increases. Unfortunately, even the simplest models along these lines appear to be analytically intractable, and none of the models that have made approximations in order to achieve tractability has gained widespread acceptance. This paper takes a different tack focusing on a special case – the isotropic model with identical commuters and the $\alpha - \beta - \gamma$ cost function – for which an analytical solution is possible. A complete, closed-form solution is presented for the no-toll equilibrium in which departures and arrivals occur in masses, and the solution for the social optimum is fully characterized.

JEL Code: L91, R41

Keywords: equilibrium, rush hour, traffic congestion

Acknowledgements:

We would like to thank Joshua Buli for helpful comments related to the existence and uniqueness of equilibrium in the model, Matthew Fitzgerald for research assistance, seminar participants at the University of Lille for helpful comments and criticisms, and Mogens Fosgerau, Nikolas Geroliminis, Raphael Lamotte, Lewis Lehe, Robin Lindsey, and Kenneth Small for very useful comments on an earlier version of the paper.

[*]Department of Economics, University of California, Riverside, 4106 Sproul Hall, Riverside, CA 92521-0427; richard.arnott@ucr.edu

[†]Department of Economics, University of Arizona, 1130 E Helen St, 401 McClelland Hall, Tucson, AZ 85721-0108; kokoza@email.arizona.edu

[‡]Department of Money and Foreign Currencies, Money and Banking Research Institute, Tehran, Iran; mehdi.naji@email.ucr.edu

1 Introduction

Until a decade ago, there were no aggregate data on traffic congestion (flow, density, and velocity) at the level of a downtown neighborhood or of the entire downtown area. Then, in a landmark paper, Geroliminis and Daganzo (2008), using a combination of stationary and mobile (taxis) sensors, measured traffic flow and density over a neighborhood of Yokohama, Japan essentially continuously over a period of weeks (see Figure 1). At this spatial scale, they found an inverse U-shaped relationship between traffic flow and density that was stable over the course of the day, and across days, which they termed the neighborhood's macroscopic fundamental diagram (MFD). Subsequent research has documented the same qualitative result for downtown areas in other cities, though the MFDs vary across neighborhoods within a downtown area and across the downtown areas of different cities.

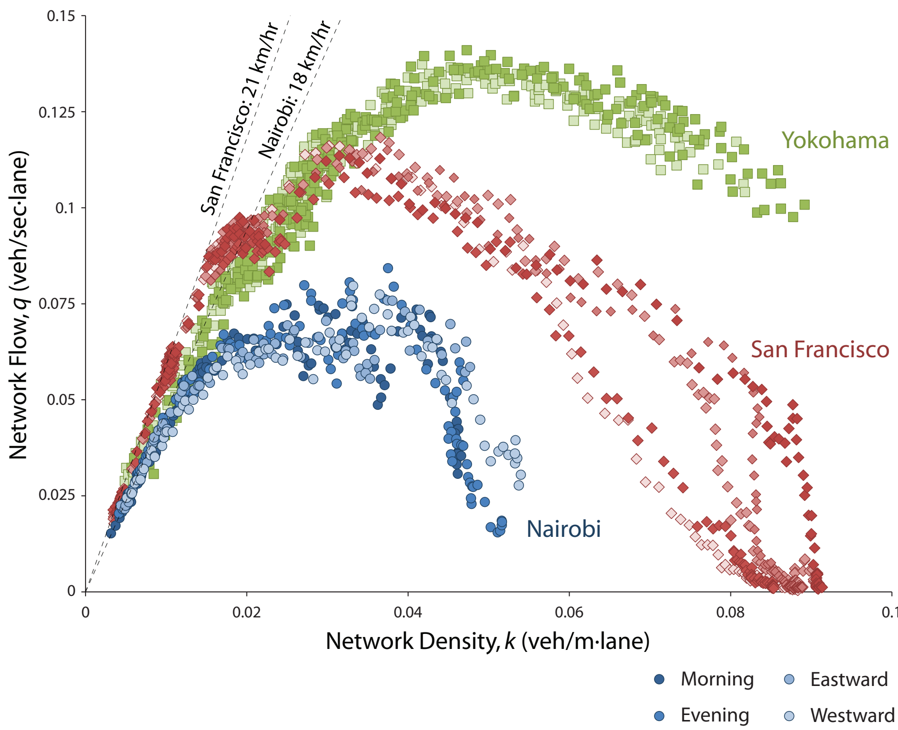


Figure 1: Macroscopic fundamental diagram for three cities
Note: Each dot corresponds to an observation. Observations were made at regular intervals throughout the business day and over days of the workweek

This research confirmed what many urban transportation economists and transportation scientists long suspected, that hypercongestion – situations where traffic flow is *negatively* related to traffic density – is a pervasive and quantitatively important feature of equilibrium rush-hour traffic dynamics at the scale of a downtown area.

William Vickrey's bottleneck model (1969) has been the workhorse model of metropolitan rush-hour traffic dynamics for a quarter century.[1] While it has proved very adaptable and has generated a host of useful insights, as a model of downtown rush-hour traffic dynamics it is flawed since it rules out hypercongestion, assuming instead that under congested conditions aggregate traffic flow is constant. Urban transportation economists have been searching for a model of rush-hour traffic dynamics that admits hypercongestion without sacrificing the elegant simplicity of the bottleneck model. There is an obvious alternative "proper" model. Unfortunately, it is in general analytically intractable. Some modelers have addressed this intractability by making approximating assumptions, but none of these approximating models has been widely accepted, partly because, without solution of the proper model, the accuracy of the approximations is unknown. This paper takes a different tack, solving for the equilibrium and the optimum for a special case that is analytically tractable.

To place this paper in context, we present the essentials of the alternative proper model, and explain why the equilibrium is in general analytically intractable but tractable for the special case we consider. The simplest variant of the alternative proper model describes an isotropic (spatially uniform) downtown area in which a fixed number of identical commuters per unit area, N, have to travel a given distance, L, from home to work over the morning rush hour. We term the class of models that considers an isotropic downtown area as "the isotropic model".[2] The congestion technology conforms to the macroscopic fundamental diagram, combining an assumed technological relationship in which traffic velocity, v, is inversely related to traffic density per unit area, k ($v = v(k)$, $v'(k) < 0$), with the fundamental identity of traffic flow, that flow per unit area, q, equals velocity times density per unit area: $q(k) = kv(k)$. A defining feature of no-toll rush-hour traffic equilibrium is the Vickrey trip-timing equilibrium condition: that no commuter can reduce her trip cost by altering her departure time. It is assumed that trip cost can be written as a function of departure time and travel time:

$$c(t, T(t)) = \underline{c} \quad \text{for all } t \in D \qquad \text{and} \qquad c(t, T(t)) \geq \underline{c} \quad \text{for all } t \notin D, \tag{1}$$

where $\underline{c}$ is the equilibrium trip cost, t is departure time, $T(t)$ is equilibrium travel time as a function of departure time, $c(\cdot)$ is the trip cost function, and D is the set of times at which departures occur. We assume for the moment that $c(\cdot)$ has continuous partial derivatives. The physical relationship that the integral of travel speed over the duration of a trip equal trip distance is

$$\int_t^{t+T(t)} v(k(u))du = L. \tag{2}$$

[1] Small (2015) offers a review of the bottleneck literature.

[2] This term is tentative since there is yet no consensus on terminology for this class of models.

Note that this relationship holds at all times, not just times in the departure set. Differentiating (2) with respect to t yields[3]

$$v\Big(k\big(t+T(t)\big)\Big)\big(1+\dot{T}(t)\big) = v\big(k(t)\big). \tag{3}$$

Differentiating (1) with respect to t gives

$$0 = c_t(t,T(t)) + c_T(t,T(t))\dot{T}(t) \qquad \text{for } t \in D. \tag{4}$$

Substituting the expression for $\dot{T}(t)$ from (4) into (3) yields

$$v\Big(k\big(t+T(t)\big)\Big) = \frac{v\big(k(t)\big)}{1-\frac{c_t(t,T(t))}{c_T(t,T(t))}} \qquad \text{for } t \in D. \tag{5}$$

which indicates what a commuter's velocity needs to be at the end of her trip, as a function of her velocity at the beginning of her trip, for the trip-timing condition to be satisfied. There is also the equation of motion

$$\dot{k}(t) = e(t) - x(t),$$

where $e(t)$ is the entry (or departure) rate per unit area and $x(t)$ is the exit (or arrival) rate per unit area at time t. There are also the constraints that $e(t) > 0$ for $t \in D$ and $e(t) = 0$ for $t \notin D$. Since the number of commuters who have exited by time $t + T(t)$, $X(\cdot)$, equals the number of commuters who have entered by time t, $E(\cdot)$, we have

$$X(t+T(t)) = E(t),$$

which implies that

$$x(t+T(t))(1+\dot{T}(t)) = e(t).$$

Since there is a period at the beginning of the rush hour during which there are entries but no exits, and a period at the end of the rush hour during which there are exits but no entries, the equation of motion may

[3]Consider two commuters, with commuter 2 departing an interval of time dt after commuter 1. For most of the trip they travel on the road together at the same (time-varying) velocity. However, when commuter 2 enters the road, commuter 1 has already traveled a distance $v(k(t))dt$. When commuter 1 exits the road, commuter 2 has this distance further to travel, which takes her an amount of time $dt' = v(k(t))dt/v(k(t+T(t))$. Since, commuter 2's trip takes $\dot{T}(t)dt$ more time than commuter 1's, $dt' = (1+\dot{T}(t))dt$.

be rewritten as

$$\dot{k}(t) = \begin{cases} e(t) & \text{for } t \in [\underline{t}, \underline{t}') \\ e(t) - x(t) = e(t) - e\big(t - S(t)\big)(1 - \dot{S}(t)) & \text{for } t \in [\underline{t}', \bar{t}] \\ -x(t) = -e\big(t - S(t)\big)(1 - \dot{S}(t)) & \text{for } t \in (\bar{t}, \bar{t}'] \end{cases} \tag{6}$$

where $\underline{t}$, $\underline{t}'$, $\bar{t}$, and $\bar{t}'$ are the times of the first entry, the first exit, the last entry, and the last exit, respectively, and $S(t)$ is the travel time[4] for a commuter who *arrives* at time t.

The difficulty lies in determining an entry rate function over the rush hour that is consistent with (5) and (6), the boundary conditions that $k(t)$ is zero at the beginning and end of the rush hour, the conditions that the entry rate is positive over the departure interval and zero outside it, and the condition that the integral of the entry rate over the departure interval equals the exogenous population, N. Not a single example has yet been found that has a closed-form or analytical solution.[5] This is very disappointing since the isotropic model is such a natural starting point for the study of downtown rush-hour traffic dynamics.

This paper investigates the special case of the isotropic model having the $\alpha - \beta - \gamma$ cost function. With this cost function,[6] for $t \in D$

$$\dot{T}(t) \begin{array}{ll} = \dfrac{\beta}{\alpha - \beta} & \text{with early arrival} \\ = -\dfrac{\gamma}{\alpha + \gamma} & \text{with late arrival,} \end{array} \tag{7}$$

with $\alpha > \beta$. Eqs. (3) and (7) together imply that, with the $\alpha - \beta - \gamma$ cost function, there is a discontinuous increase in $\dfrac{v(k(t + T(t)))}{v(k(t))}$ at $t' \equiv t^* - S(t^*)$. This requires that there be a departure mass at t', an arrival mass at t^* or both. Since a departure mass at time t' implies an arrival mass at t^*, and *vice versa*, both occur. This observation gives rise to the conjecture that, with the $\alpha - \beta - \gamma$ cost function, there is an equilibrium in which all departures occur in masses, which implies that all arrivals also occur in masses. The analysis that follows confirms this conjecture. Furthermore, with departure masses, we obtain a closed-form solution for equilibrium.

Thus, the motivation of the paper is as follows. In what seems to be the simplest and most natural model

[4] In terms of the arrival time, the conservation of cars is $X(t) = E(t - S(t))$, so that $x(t) = e(t - S(t))(1 - \dot{S}(t))$.

[5] To us at least, it is counterintuitive that the model is so difficult to solve. Imagine a road from A to B on which velocity is constant along the road at a point in time but changes over time. One would think that knowing how trip duration varies over time would provide enough information to solve analytically for how velocity varies over time. But such is not the case.

[6] With early arrival, $c(t, T(t)) = \alpha T(t) + \beta(t^* - (t + T(t)))$, where α is the unit value of travel time, β is the unit value of time early, and t^* is the desired arrival time. Over any interval of the early morning rush hour over which departures are continuous, the trip-timing condition implies that $dc(t, T(t))/dt = 0$, so that $\dot{T}(t) = \dfrac{\beta}{\alpha - \beta}$. Inserting this result into (5) gives that $v(k(t+T(t))) = [(\alpha-\beta)/\alpha]v(k(t))$. Thus, over any interval of the morning rush hour over which departures are continuous, a commuter's velocity at the end of her trip is $(\alpha - \beta)/\alpha$ times her velocity at the beginning of her trip.

of equilibrium rush-hour traffic dynamics in a downtown area, an analytical solution of equilibrium is not in general possible. Without an analytical solution, it is difficult to gain insight into the economic properties of equilibrium. There is, however, a special case with an analytical solution. Studying this special case, which entails departure masses, will hopefully provide economic insights that generalize.

Section 2 provides a brief review of the relevant literature. Section 3 presents the model and discusses issues related to the existence and uniqueness of equilibrium in the model. Section 4 provides a detailed analysis of equilibrium in the model. Section 5 presents a preliminary and incomplete analysis of the social optimum in the model, and a preliminary and incomplete comparison of equilibrium and optimum. Section 6 discusses extensions, and section 7 concludes.

2 Literature Review

The importance of having a strong theoretical basis for designing policies to address rush-hour traffic congestion has long been recognized. The literature on the subject derives from three classic articles. The first is Beckmann et al. (1956), which derives static (steady-state) equilibrium on a traffic network in which travel time on a link is an increasing function of the ratio of the volume (flow) to capacity on the link. The second is Walters (1961), which both develops the economic theory of steady-state traffic congestion that we employ today and introduces hypercongestion. The third is Vickrey (1969), which presents the bottleneck model.

Agnew (1976) presents a dynamic model in which the throughput of a congestion-prone system is a strictly concave function of its load. In the context of rush-hour traffic dynamics, throughput is the exit rate and load is traffic density. In that context, the model makes the simplifying assumption that the exit rate depends only on traffic density and not on the time pattern of entries.[7] It is surprising that the Agnew model has not been widely used by transportation economists, since it may provide a good approximation to actual rush-hour traffic dynamics when the average duration of a trip is short relative to the length of the rush hour. In unpublished notes, Vickrey (1991) sketches a model, which Vickrey termed the "bathtub model" of traffic congestion, that essentially adapts Agnew's model to traffic congestion in Manhattan. Arnott (2013) provides a formalization of Vickrey's bathtub model that explicitly incorporates schedule delay costs and, in the equilibrium variant, applies Vickrey's equilibrium trip-timing condition. The paper shows that Agnew's assumption that the exit rate depends only on traffic density holds if trip distances have the same negative exponential distribution over the rush hour, but also argues that this will generally not hold since trip-cost-minimizing commuters will order themselves by trip distance, those with shorter trips traveling closer to the peak.

[7]This assumption is supported by the empirical work by Geroliminis and Daganzo (2008) mentioned earlier.

Several paper have modeled rush-hour traffic dynamics by extending the Vickrey bottleneck model so that the capacity of the bottleneck depends on the length of the queue behind it. Most recently Fosgerau and Small (2013) sidesteps the analytical intractability of a proper model by treating a discrete number of capacity levels. Yet other papers avoid the intractability of the proper isotropic model by making the approximation that a commuter's travel time on a trip depends on the density of cars when the commuter begins (Geroliminis and Levinson 2009) or ends (Small and Chu 2003) her trip.

None of the models that avoid the intractablity of the proper model by making approximating assumptions has gained widespread acceptance. There are three grounds for criticizing an approximating assumption. First, it may be "inconsistent with rational economic behavior"; this is a valid criticism since rational behavior is necessary for sound welfare analysis. Second, it may be "inconsistent with the laws of physics"; this is a valid criticism, though it needs to be made precise what laws of physics are violated and how. And third, it may result in an inaccurate approximation; this is a potentially valid criticism, but to gauge the inaccuracy of an approximation it is necessary to know the exact solution, and, remarkably, no one has yet solved a proper model numerically.

The model of this paper is consistent with rational economic behavior. It also entails no approximations. However, its congestion technology can be viewed as inconsistent with laws of physics. For one thing, when a car enters the street system it is assumed to instantaneously travel at the velocity of the prevailing traffic, which entails infinite acceleration, and when a car exits the road it is assumed to reduce its velocity from that of the prevailing traffic to zero instantaneously, which entails infinite deceleration. This inconsistency is more severe in our model since a departure mass entails a mass of cars simultaneously entering and immediately traveling at the velocity consistent with the density of cars in the departure mass. To what extent these physical inconsistencies cause the aggregate behavior of the model to be unrealistic is a matter of judgment.

Fosgerau (2015) is the only paper in the literature to date that works with the proper isotropic model (which Fosgerau refers to as the bathtub model).[8] In contrast to this paper, it specifies a trip cost function that is sufficiently smooth that departure masses do not occur. Also, in contrast to this paper, the two variants of the model that it explores treat heterogeneous commuters. In the first, drivers are heterogeneous with respect to trip distance; in the second, they are heterogeneous with respect to both trip distance and the analog in his model to desired arrival time. For both variants, "regular sorting" is assumed, under

[8] Fosgerau (2015) and Vickrey (1991) (with Arnott (2013) following Vickrey) use the term "bathtub" model in different senses. In the hydrological literature, a bathtub model of a water table is so called since water is assumed to distribute itself over space such that the height of the water table is the same everywhere, just as the surface of the water in a bathtub is flat. Fosgerau (2015) refers to a bathtub model of traffic flow in this sense since traffic is assumed to distribute itself over space such that traffic density is the same everywhere. Vickrey (1991) uses the term bathtub in a double sense to model rush-hour traffic dynamics in Manhattan. First, Manhattan is shaped like a bathtub, and one may view Manhattan streets as filling up with traffic in the early morning rush hour and then draining in the late morning rush hour. Second, the exit rate from traffic depends on traffic density, just as the rate at which water flows out of a bathtub depends on the height of water in the bathtub.

which drivers with longer trip distances both depart earlier and arrive later than those with shorter trips. Under this assumption, the model can be formulated in terms of ordinary differential equations. In the first variant, the paper obtains the strong result that the equilibrium and the optimum coincide. In the second variant, among drivers with the same trip distance, a driver with an earlier desired arrival time departs and arrives earlier than a driver with a later desired arrival time, and tolling is effective. The paper presents numerical examples with hypercongestion in the equilibrium but none with hypercongestion in the social optimum, which leads to the conjecture that hypercongestion does not arise with socially optimal rush-hour traffic dynamics. How limiting the assumption of regular sorting is remains to be seen.[9] The paper's analysis suggests, paradoxically, that the mathematics with a continuum of commuters who differ in terms of a naturally ordered characteristic, such as trip distance or desired arrival time, may be easier than the mathematics with identical individuals, by smoothing the problem.

The bottleneck model is remarkable in having gained almost universal currency. In contrast, none of the models reviewed above has been widely adopted.

Trained in the spirit of microeconomics, transportation economic theorists have resisted solving models exclusively using numerical analysis. Numerical examples are welcome to supplement and quantify theory but not as a substitute for theoretical analysis. But to break the current logjam in the study of rush-hour traffic dynamics with hypercongestion, it may be necessary to rely more heavily on numerical analysis. Numerical analysis will provide exact solutions that can be used to determine the accuracy of alternative approximating models. As well, numerical analysis can be used to generate empirical regularities that will guide theory. It can be used, for example, to investigate whether the Fosgerau (2015) regularity condition holds for most realistic rush-hour traffic conditions or only for a highly restricted subset. It can also be used to resolve an important question that this paper does not answer. We show that, under our assumptions, an equilibrium exists in which all departures occur in masses with contiguous travel time intervals, but do not demonstrate uniqueness. Numerical analysis might generate an equilibrium that is different from the one we identify.

3 The Basic Model and the Properties of its Equilibrium

3.1 The Basic Model

Throughout most of the paper we work with normalized units, while unnormalized variables, as well as functions of unnormalized variables, are indicated with a $\hat{}$. The model is spatial and its space is an

[9] Regular sorting implies that all drivers be on the road at the same time, which seems unrealistic for large cities.

isotropic downtown area: home locations, job locations, and road capacity are uniformly distributed over space.[10] Per unit area, $\hat{N}$ commuters must travel from home to work over the morning rush hour. All commuters have the same commuting distance, L, and the same desired arrival time, $\hat{t}^*$.

The form of congestion is flow congestion in which commuters' velocity is decreasing in the density of commuters. To simplify the algebra, Greenshields' Relation is assumed, in which traffic velocity, $\hat{v}$, is negative linearly related to traffic density, $\hat{k}$:

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

where v_f is free-flow velocity and Ω is jam density. Thus, travel time per unit distance is

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

Greenshields' Relation has the properties that maximum or capacity flow occurs when traffic density equals one-half jam density. Traffic is said to be congested at densities below this level and hypercongested at densities above this level.

The physical relationship that trip distance equals the integral of velocity over trip duration is

$$\int_{\hat{t}}^{\hat{t}+\hat{T}(\hat{t})} \hat{v}(\hat{k}(u))du = L, \tag{8}$$

where $\hat{t}$ is departure time, and $\hat{T}(\hat{t})$ is travel time with departure time $\hat{t}$.

The familiar $\alpha - \beta - \gamma$ trip cost function is employed:

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

Where $\hat{c}(\hat{t})$ is trip cost as a function of departure time:

$$\hat{c}(\hat{t}) = \alpha\hat{T}(\hat{t}) + \beta\max(0, \hat{t}^* - \hat{t} - \hat{T}(\hat{t})) + \gamma\max(0, \hat{t} + \hat{T}(\hat{t}) - \hat{t}^*). \tag{9}$$

We impose the standard condition that the unit cost of travel time exceeds the unit cost of time early, $\alpha > \beta$. In the morning rush hour, each commuter chooses when to depart from home so as to minimize her trip price. A commuter's trip price equals her trip cost plus the toll she pays.

Four other conditions complete the model. The first is the accumulation equation, that density at time $\hat{t}$ equals cumulative entries at time $\hat{t}$ minus cumulative exits at time $\hat{t}$. Two other conditions are the boundary

[10] One may think of the downtown area as having a dense, grid network of streets extending infinitely in both the north-south and east-west directions. Alternatively, one may think of the downtown area as covering the outside of a large torus. In either case, space is isotropic so that there are no edge effects.

conditions that density equal zero right before the start of the rush hour and right after the end of the rush hour. Together these equations imply the conservation of vehicles. The fourth condition is that cumulative exits at time $\hat{t} + \hat{T}(\hat{t})$ equal cumulative entries at time $\hat{t}$. Along with (8), this condition ensures that all commuters have trip length L.

3.2 Equilibrium

A *morning rush hour equilibrium* is a time path of departures over the morning rush hour, and the induced time paths of traffic density and arrivals over the morning rush hour, such that no commuter can reduce her trip price by altering her departure time. When no toll is applied, we refer to this as a no-toll morning rush hour equilibrium. "Trip price" in the definition of equilibrium may then be replaced by trip cost. Except section 6, the paper focuses on the basic model in which commuters are identical – having the same α, β, γ, L and $\hat{t}^*$ – with no late arrivals admitted. Analytically, "admitting no late arrivals" is the limiting case as the unit value of time late, γ, approaches infinity.

In the no-toll morning rush hour equilibrium with identical commuters, the trip-timing condition reduces to the condition that trip cost is the same at all times at which there are departures and is at least as high at all times at which there are no departures. Where D denotes the set of departure times at which there are departures, this condition is

$$\hat{c}(\hat{t}) = \underline{\hat{c}} \quad \text{for all } \hat{t} \in D \qquad \text{and} \qquad \hat{c}(\hat{t}) \geq \underline{\hat{c}} \quad \text{for all } \hat{t} \notin D,$$

where $\hat{c}(\hat{t})$ is given by (9).

3.3 Existence and Uniqueness of a No-Toll Morning Rush Hour Equilibrium

Even though the above model is simple, its mathematical structure is not. There are no off-the-shelf results from the theory of integral/differential equations that can be applied to establish the existence and uniqueness of equilibrium in the isotropic model. One reason is that the model gives rise to non-standard differential equations[11] (i.e., not ordinary differential equations) for which the literature on existence and uniqueness of equilibrium is not as well developed as the corresponding literature on ordinary differential equations. Another is that all theorems on existence and uniqueness of solutions to differential equations apply to particular classes of functions (such as analytical functions) not to any function, and there is no good

[11] Differentiation of (3) gives $v'(k(t + T(t)))\dot{k}(t + T(t))(1 + \dot{T}(t))^2 + v(k(t + T(t)))\dddot{T}(t) = v'(k(t))\dot{k}(t)$. With Greenshields' Relation, this simplifies to $\dot{k}(t+T(t))(1+\dot{T}(t))^2 - v(k(t+T(t))\dddot{T}(t) - \dot{k}(t) = 0$. Over the departure interval, trip cost is constant. Taking this trip cost as a parameter, the functions $T(t)$, $\dot{T}(t)$, and $\dddot{T}(t)$ are exogenous. Therefore the differential equation is a delay differential equation with an endogenous delay.

justification for restricting our model's equilibrium departure rate function to be a member of particular class of functions. Thus, investigation of the existence and uniqueness of equilibrium in the basic model requires *ad hoc* reasoning. In this subsection, we assume that no late arrivals are permitted.

Lemma 1. *The last arrival must be at $\hat{t}^*$.*

Proof. Suppose not and that the last arrival is at $\hat{t}' < \hat{t}^*$. A deviating commuter who departs a period of time $\widehat{dt}$ after the commuter who arrives at $\hat{t}'$ will travel a distance $\hat{v}\Big(\hat{k}\big(\hat{t}' - S(\hat{t}')\big)\Big)\widehat{dt}$ after the commuter who arrives at $\hat{t}'$ completes her journey. Since she travels at free-flow speed over this distance, she arrives a period of time $\frac{\hat{v}\left(\hat{k}\left(\hat{t}'-S(\hat{t}')\right)\right)\widehat{dt}}{v_f}$ less early than the commuter who arrives at $\hat{t}'$. Since $\hat{v}\Big(\hat{k}\big(\hat{t}' - S(\hat{t}')\big)\Big)$ is less than or equal to v_f, the travel time of the deviating commuter is no greater than that of the commuter who arrives at $\hat{t}'$ while her schedule delay is less, which is inconsistent with the definition of equilibrium. ■

Let $\hat{\hat{t}}$ denote the departure time corresponding to arrival at $\hat{t}^*$.

Lemma 2. *There is an arrival mass at $\hat{t}^*$ and a corresponding departure mass at $\hat{\hat{t}}$.*

Proof. Suppose not, so that there is neither an arrival mass at $\hat{t}^*$ nor a departure mass at $\hat{\hat{t}}$. Then, since late arrival is not admitted, $\hat{k}(\hat{t}^{*-}) = 0$. Either there are departures in the interval of time $\widehat{dt}$ prior to $\hat{\hat{t}}$ or there are not. Consider first the case where there are departures in the interval of time prior to $\hat{\hat{t}}$. The condition that trip cost be the same at $\hat{\hat{t}} - \widehat{dt}$ as at $\hat{\hat{t}}$ implies that

$$\dot{\hat{T}}(\hat{\hat{t}}^-) = \beta/(\alpha - \beta).$$

Differentiating (8) with respect to $\hat{t}$ and evaluating at $\hat{\hat{t}}^-$ gives

$$\hat{v}\Big(\hat{k}\big(\hat{\hat{t}}^- + \hat{T}(\hat{\hat{t}}^-)\big)\Big)\big(1 + \dot{\hat{T}}(\hat{\hat{t}}^-)\big) = \hat{v}\big(\hat{k}(\hat{\hat{t}}^-)\big).$$

Since $\dot{\hat{T}}(\hat{\hat{t}}^-) = \beta/(\alpha - \beta)$, and since $\hat{\hat{t}}^- + T(\hat{\hat{t}}^-) = \hat{t}^{*-}$,

$$\hat{v}\Big(\hat{k}(\hat{t}^{*-})\Big)\left[\frac{\alpha}{\alpha - \beta}\right] = \hat{v}(\hat{k}(\hat{\hat{t}}^-)),$$

which implies that $\hat{v}\big(\hat{k}(\hat{t}^{*-})\big) < \hat{v}(\hat{k}(\hat{\hat{t}}^-))$. But since $\hat{v}\big(\hat{k}(\hat{t}^{*-})\big) = v_f$ is the maximum possible velocity, this leads to a contradiction. This leaves the other case, in which there are no departures in the time interval immediately prior to $\hat{\hat{t}}$. Let $\hat{t}''$ be the latest time prior to $\hat{\hat{t}}$ at which there are departures. Since cars may have exited the downtown street system between $\hat{t}''$ and $\hat{\hat{t}}$, and since no cars have entered the downtown street system in this time interval, the travel time at $\hat{t}''$ must be at least as high as that at $\hat{\hat{t}}$. Since the schedule delay cost for departure at $\hat{t}''$ is strictly higher than that at $\hat{\hat{t}}$, the trip cost for departure at $\hat{t}''$ is strictly higher than that at $\hat{\hat{t}}$, which is inconsistent with equilibrium. ■

A broad intuition is as follows. Since the last commuter to arrive faces no schedule delay cost, she must face the highest travel time cost. If there were no arrival mass at t^*, the last commuter to arrive would be traveling at free-flow velocity when at the end of her journey, which is inconsistent with her having the highest travel time cost.

Lemmas 1 and 2 lead to Conjecture 1, that there is an equilibrium departure rate function with all departures occurring in masses. Actually, we came to this conjecture by a different route. We noticed that when the number of commuters is small relative to the capacity of the street system, a single departure mass arriving at $\hat{t}^*$ is an equilibrium. Any commuter departing after the mass arrives after $\hat{t}^*$, which is not is permitted. If the number of commuters is sufficiently small, any commuter who departs before the mass experiences a schedule delay cost that is greater than the travel cost saving.

Lemmas 1 and 2 also lead to additional conjectures: Conjecture 2, that, among departure rate functions in which all departures occur in masses, there is a unique equilibrium, and Conjecture 3, that this equilibrium is in fact the unique equilibrium when no restrictions are put on the form of the departure rate function.

3.3.1 Conjecture 1

Conjecture 1 is that there is an equilibrium departure rate function with all departures occurring in masses. Since we provide a proof by construction in the next section, here we just provide a heuristic argument.

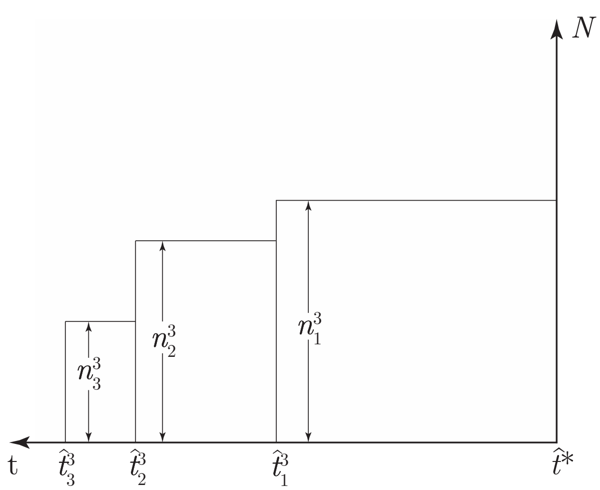


Figure 2: Three contiguous departure masses

Figure 2 displays an equilibrium with three departure masses. The departure masses are indexed such that the departure mass arriving at $\hat{t}^*$ is mass 1, the next latest departure mass is mass 2, etc. The travel time intervals of the three masses are contiguous – they do not overlap and there is no travel time interval between the travel time intervals of the masses. Thus, for example departure mass 1 departs immediately after departure mass 2 arrives.

Letting $\hat{n}_i^m$ denote the size of departure mass i when there are m departure masses, the velocity of cars in mass i is $\hat{v}_i^m = v_f\left(1 - \frac{\hat{n}_i^m}{\Omega}\right)$. The travel time of a commuter in mass i is trip distance divided by velocity, $\hat{T}_i^m = \left(\frac{L}{v_f}\right)\left[\frac{\Omega}{\Omega - \hat{n}_i^m}\right]$. The trip cost of a commuter in mass i is the unit cost of travel time multiplied by travel time in mass i plus the unit cost of time early multiplied by the time early, which equals the sum of the travel times of all later departure masses, $\hat{c}_i^m = \alpha\hat{T}_i^m + \beta\sum_{j=1}^{i-1}\hat{T}_i^m$. In equilibrium the trip cost of all the masses is the same and the trip cost of a commuter who departs before the earliest departure mass is greater than the common trip cost in the masses.

The departure function displayed in Figure 2 is an equilibrium. If a commuter deviates by departing after the latest departure mass, she arrives late, which is not permitted. If she deviates by departing between two departure masses, her trip cost is a weighted average of the trip cost associated with traveling in each departure mass, and since those trip costs are the same so is their weighted average. If she deviates by departing before the earliest departure mass, her trip cost is strictly higher than the equilibrium trip cost. Extending the argument to an arbitrary number of departure masses confirms Conjecture 1. Thus, we have

Theorem 1. *In the isotropic model with identical individuals, no late arrivals allowed, and a cost function that is linear in travel time and time early, for any set of parameter values there exists a "restricted" equilibrium, where "restricted" entails all departures occurring in departure masses with contiguous travel time intervals.*

3.3.2 Conjectures 2 and 3

Conjecture 2 is that, among departure rate functions in which all departures occur in masses, there is a unique equilibrium, which is the unique restricted equilibrium.

Lemma 3. *In the no-toll equilibrium, there can be no time interval in the interior of the rush hour with no cars on the road.*

Proof. Suppose not, and that a departure configuration such as that displayed in Figure 3 is an equilibrium. Consider a deviating commuter who departs an interval of time $\widehat{dt}$ after the departure mass at $\hat{t}_2^2$. The deviating commuter travels most of her journey with the departure mass. When the departure mass exits she has a distance $\hat{v}\big(\hat{k}(\hat{t}_2^2)\big)\widehat{dt}$ left to travel, which she does at free-flow speed. Not only is her trip duration lower than the trip duration of those traveling in the departure mass but also her schedule delay is lower, which is inconsistent with the equal trip-cost condition. ■

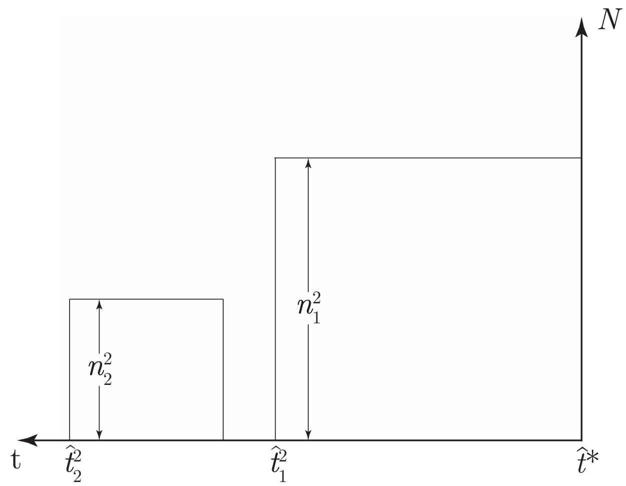


Figure 3: Two departure masses with an interval with no cars on the road during the rush hour

We now investigate whether, in the no-toll equilibrium, the travel time intervals of different departure masses may overlap. Unfortunately, we have not been able to resolve the issue one way or the other, though we do have a negative result for two departure masses.

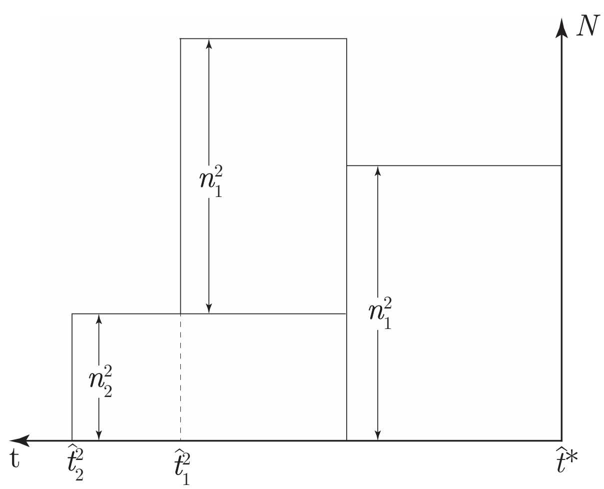


Figure 4: Two departure masses with both masses on the road for part of the rush hour

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

Proof. With two departure masses with overlapping travel time intervals, the rush hour is described by three intervals, as shown in Figure 4. To avoid confusing intervals and departure masses, we shall refer to the three intervals in terms of time – earliest, middle, and latest. Only commuters who depart in departure mass 2 travel in the earliest interval, all commuters travel in the middle interval, and only commuters in mass 1 travel in the latest interval. To ease burden of notation, we normalize units so that $\alpha = 1$, $L = 1$, $v_f = 1$, and $\Omega = 1$. We also set $\theta = \frac{\beta}{\alpha}$, which by assumption is less than 1, and $N = \frac{\hat{N}}{\Omega}$. Let δ be the proportion of the commuting distance traveled by all commuters in the middle interval. Commuters in departure mass 1 incur a normalized cost of $\frac{\delta}{1-N}$ for travel in the middle interval and $\frac{1-\delta}{1-n_1^2}$ for travel in the latest interval. Commuters in departure mass 2 incur a normalized cost of $\frac{\delta}{1-N}$ for travel in the middle interval and $\frac{1-\delta}{1-n_2^2}$ for travel in the earliest interval, and incur schedule delay equal to the duration of the latest interval, at a cost of $\frac{\theta(1-\delta)}{1-n_1^2}$. The condition that commuters in the two departure masses incur the same cost is

$$c = \frac{1-\delta}{1-n_2^2} + \frac{\delta}{1-N} + \frac{\theta(1-\delta)}{1-n_1^2} = \frac{\delta}{1-N} + \frac{1-\delta}{1-n_1^2}. \tag{10}$$

This requires that $N \in (\theta, 1)$. If $N < \theta$ then the n_2^2 that solves (10) is negative, which is inconsistent with equilibrium. And if $N > 1$, density in the middle interval would exceed jam density, which is inconsistent with equilibrium. Now consider a deviating commuter who travels a distance δ solo before joining the commuters in departure mass 2, with whom she travels a distance $1-\delta$. She travels the distance δ solo at velocity 1 and the remaining distance $(1-\delta)$ at velocity $\dfrac{1}{1-n_2^2}$, and her schedule delay is $\dfrac{\delta}{1-N} + \dfrac{1-\delta}{1-n_1^2}$ for a trip cost of

$$c' = \delta + \frac{\theta\delta}{1-N} + \frac{1-\delta}{1-n_2^2} + \frac{\theta(1-\delta)}{1-n_1^2}.$$

Thus,

$$c' - c = \frac{\delta(N-\theta)}{1-N}, \tag{11}$$

so that her trip cost is lower than that of the other commuters if $N \in (\theta, 1)$. Thus for all values of N the configuration shown in Figure 4 is inconsistent with equilibrium. ■

Consider a downtown area where N is growing over time, and assume that all departures occur in masses. Initially, there is a single departure mass. As N grows, a critical N, which is derived in the next section, is reached at which a second departure mass forms. Per Lemma 4, in equilibrium these departure masses have contiguous travel time intervals. As N grows further, another critical N is reached at which a third departure mass forms. Assuming that the travel time intervals associated with the first two departure masses remain contiguous, the line of reasoning employed above can be applied to establish that the travel time interval of the third departure mass is contiguous to that of the second departure mass, and so on.

This observation is consistent with Conjecture 2 but does not establish it, since the definition of equilib-

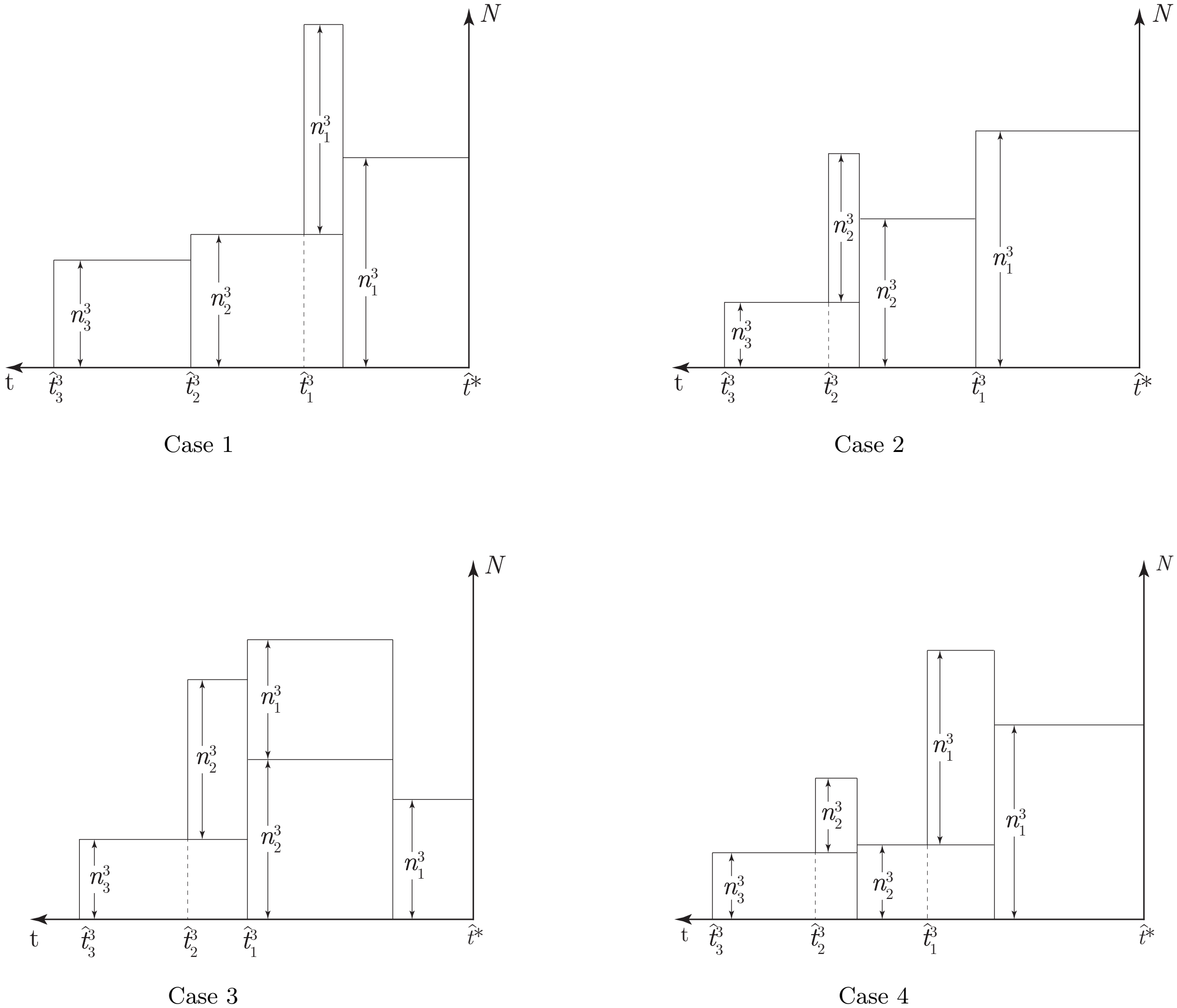


Figure 5: Four possible departure patterns for three departure masses

rium is a static one and makes no reference to adjustment dynamics or stability. A logical next step is to investigate whether an equilibrium exists with three departure masses with overlapping travel time intervals. Figure 5 displays the mutually exclusive and collectively exhaustive qualitative departure patterns for three masses with at least one overlapping travel time interval and no interval with no cars on the road.

Case 1	$\hat{t}_3^3 < \hat{t}_3^{3'} = \hat{t}_2^3 < \hat{t}_1^3 < \hat{t}_2^{3'} < \hat{t}_1^{3'}$
Case 2	$\hat{t}_3^3 < \hat{t}_2^3 < \hat{t}_3^{3'} < \hat{t}_1^3 = \hat{t}_2^{3'} < \hat{t}_1^{3'}$
Case 3	$\hat{t}_3^3 < \hat{t}_2^3 < \hat{t}_3^{3'} = \hat{t}_1^3 < \hat{t}_2^{3'} < \hat{t}_1^{3'}$
Case 4	$\hat{t}_3^3 < \hat{t}_2^3 < \hat{t}_3^{3'} < \hat{t}_1^3 < \hat{t}_2^{3'} < \hat{t}_1^{3'}$

This is as far as we have proceeded. One possibility is that all these cases can be eliminated as possible equilibria and that the arguments used to establish this can be applied to eliminate as possible equilibria patterns with more than three departure masses and at least one overlapping travel time interval. If this can be established, then applying the arguments in the limit as the number of departure masses approaches infinity would confirm Conjecture 2. Another possibility is that there exists an equilibrium with three departure masses and at least one overlapping travel time interval, which would disprove the conjecture. The final possibility is that examination of these cases is inconclusive – all the above cases can be eliminated as possible equilibria, but the arguments used to establish this cannot be applied to patterns with more than three departure masses and at least one overlapping travel time interval. Conjecture 3 will be even more difficult to prove or disprove.

In the remainder of the paper, we restrict analysis to the situation where the travel time intervals associated with each departure mass are contiguous. In the next section, we prove by construction that for all (positive) parameter values for which $\alpha > \beta$ a restricted equilibrium exists and is unique.

4 (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. In this section, we assume that individuals are identical and that no late arrivals are admitted; these assumptions are relaxed in section 6. 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 over which each corresponding departure mass travels from home to work, with the latest departure mass arriving exactly on time.*

We define a restricted equilibrium to be an equilibrium, conditional on departures satisfying assumption 1. In what follows, we shall demonstrate by construction that a restricted equilibrium exists.

4.1 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 above 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.

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 latest 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 trip 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.

4.1.1 One departure mass

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,

$$c_1^1(N) = \frac{\alpha L}{\hat{v}} = \frac{\alpha L}{v_f \left(1 - \frac{\hat{N}}{\Omega}\right)} \tag{12}$$

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 employ several normalizations, but 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 distance, the normalized time unit is the length of time it takes to travel the trip distance at free-flow velocity, the normalized money unit is the cost of travel per normalized time unit, and normalized population density is jam density. With these normalizations, (12) reduces to

$$c_1^1(N) = \frac{1}{1 - N}, \tag{13}$$

the velocity of the mass is $1 - N$, and its travel time is $\dfrac{1}{1 - N}$. 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, 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, and 1 normalized population density unit equals Ω unnormalized population density units.[12]

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.

We now proceed with the analysis in normalized units. Consider an infinitesimal commuter who departs a period $\Delta t \leq 1$ 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 $N_{1,2}^e = \theta$. 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.

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

[12]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 shall employ, we assume that $v_f = 15$ mph and $L = 5$ miles, 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.

N. From (13), when there is a single departure mass in equilibrium, thus when $c^e(N) = c_1^1(N)$, total cost is

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

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

$$MSC^e(N) = \frac{dTC^e_{(1)}}{dN} = \frac{1}{(1-N)^2}$$
$$MCE^e(N) = MSC^e(N) - c^e(N) = \frac{N}{(1-N)^2}.$$

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 the cost imposed on other commuters 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

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

We bring together the above results in

Proposition 1. *A restricted 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 $s^e(N) = \frac{N}{1-N}$.*

4.1.2 Two departure masses

Now we consider two departure masses, that is $m = 2$. 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 travel time of $\frac{1}{1-n_2^2}$ and schedule delay of $\frac{1}{1-n_1^2}$. 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 $N \le \theta(3-\theta)$. Thus, equilibrium entails two departure masses for $N \in (\theta, \theta(3-\theta))$. For N in this interval

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

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

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

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

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

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

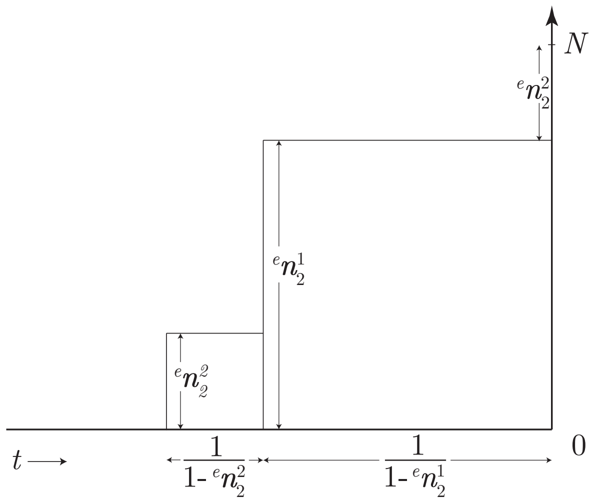
$N_{2,3}^e$ is that N for which a commuter is indifferent between departing in departure mass 2 and departing in departure mass 3 by herself. If she departs in departure mass 3 by herself, her travel time cost decreases by $\frac{1}{1-n_2^2} - 1$ and her schedule delay cost increases by $\frac{\theta}{1-n_2^2}$.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

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

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

We bring together the results for two departure masses in

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

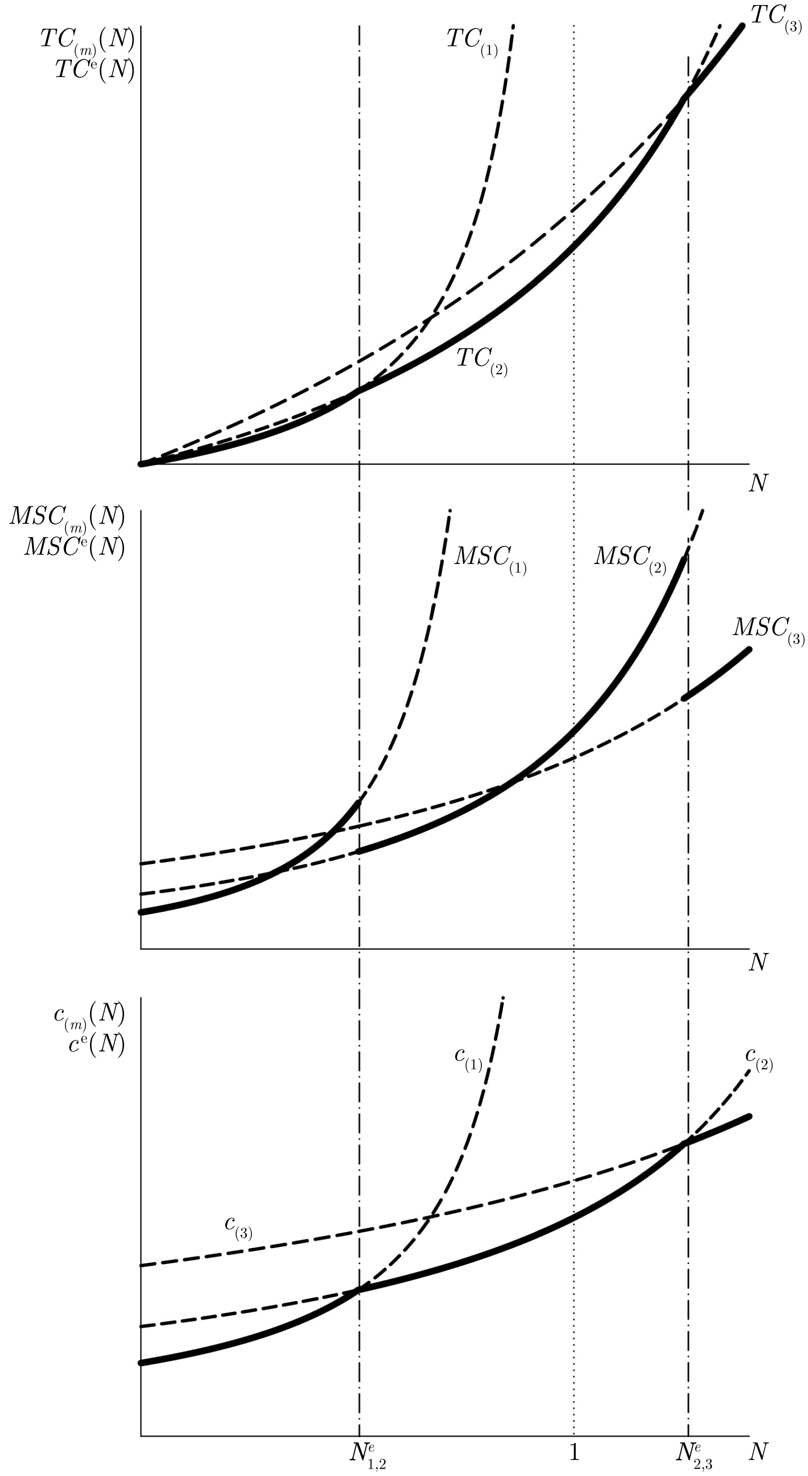
Figure 6: No-toll equilibrium with two departure masses

Figure 6 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: $\alpha = \$20/\text{hr}$, $L = 5$ miles, and $v_f = 15$ mph. $L/v_f = 0.333$ hrs is the assumed trip duration at free-flow speed. The numerical results are recorded in Table 1. Complementing Table 1 is Figure 7, 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 7. 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.

Starting with the bottom panel, $c_{(1)}(N) = \dfrac{1}{1-N}$ gives trip cost as a function of N when the entire population travels in a single departure mass. Since the mass arrives on time, the entire trip cost is travel time



$TC_{(m)}(N)$
$TC^e(N)$
$TC_{(1)}$
$TC_{(2)}$
$TC_{(3)}$
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 7

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 is 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 $c_{(2)}(\theta) = c_{(1)}(\theta) = \dfrac{1}{1-\theta}$, $c_{(2)}(1) = \dfrac{2-\theta}{1-\theta}$, 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 two, and as a dashed line outside this interval. $c_{(3)}(N) = \dfrac{3\theta - 3 - \theta^2}{(3-N)(1-\theta)^2}$. Since a switch occurs from m to $m+1$ 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, $c^e(N)$, is the lower envelope of the trip cost functions for specific numbers of departure masses. As a result it has an escalloped shape, shown as a solid bold line.

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 the 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 m to $m+1$ departure masses in equilibrium. 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 in section 4.4. 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 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

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.

decreases discontinuously at each critical population density where a departure mass is added; and iii) the length of the rush hour increases continuously 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 $\theta > N > 1/2$, and does not occur when $\theta \leq 1/2$; with the assumed parameter value of $\theta = 1/2$, hypercongestion does not occur. 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 departure mass 1 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.

4.2 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,

$$c_j^m = \frac{1}{1-n_j^m} + \theta \sum_{i=1}^{j-1} \frac{1}{1-n_i^m}. \tag{23}$$

The trip-timing equilibrium condition implies that

$$1 - n_{j+1}^m = \frac{1-n_j^m}{1-\theta}. \tag{24}$$

Combining (24) 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

$$n_1^m = 1 - \frac{m-N}{\frac{1-\theta}{\theta}\frac{1-(1-\theta)^m}{(1-\theta)^m}}. \tag{25}$$

Combining (25) and (23) for $j = 1$, and noting that in equilibrium the trip cost is the same for all departure masses, yields the equilibrium trip cost

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

The total trip cost can then be calculated as $TC_{(m)}^e(N) = Nc_{(m)}^e(N)$:

$$TC_{(m)}^e = \frac{N}{m-N}\frac{1-\theta}{\theta}\left[\frac{1-(1-\theta)^m}{(1-\theta)^m}\right]. \tag{27}$$

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

$$MSC_{(m)}^e = \frac{m}{N(m-N)} TC_{(m)}^e.$$

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

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

The equilibrium number of departure masses is now calculated as a function of N. By the equal trip cost condition, 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 (27), this reduces to

$$N^e_{m,m+1} = m - \frac{1-\theta}{\theta}\Big(1-(1-\theta)^m\Big). \tag{28}$$

This can be rewritten as a recursive relationship:

$$N^e_{m+1,m+2} = \theta(m+1) + (1-\theta)N^e_{m,m+1}. \tag{29}$$

Using (24) and (25), the duration of the rush hour with m departure masses and population density N is

$$D^e_{(m)} = \sum_{i=1}^{m} \frac{1}{1-{}^en_i^m} = \left[\frac{1-(1-\theta)^m}{\theta}\right]^2 \frac{(1-\theta)^{1-m}}{m-N}$$

We also have that

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

and

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

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

4.3 Comparative Static and Dynamic 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 2: Algebraic results in normalized form: equilibrium with no late arrivals

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

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

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 present the comparative static results, holding constant number of departure masses, and in the last row of Table 4 indicate whether an increase in the exogenous parameter can cause an increase or a decrease in the equilibrium number of departure masses.

The signs of the comparative static derivatives are the same with three or more departure masses as they are with two. To more easily convey the intuition, we focus on the case when there are two departure masses in equilibrium, which was presented in Section 4.1.2.

The only comparative static derivative with respect to population density 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 departure mass 2, and as $\hat{N}$ increases the proportion of the population in departure mass 2 increases (see (16)). The only comparative static derivative with respect to jam density worthy of remark is that the population in departure mass 1 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 departure mass 1 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.

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

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

Notes: Let $A(m,\theta) = \dfrac{1-\theta}{\theta}\left[\dfrac{1-(1-\theta)^m}{(1-\theta)^m}\right]$. 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 distance at free-flow velocity; thus, to convert to unnormalized units, multiply by $\dfrac{L}{v_f}$, trip distance 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.

The comparative static properties with respect to free-flow velocity derive from Greenshields' 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 distance 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, i.e. 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

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	+	$?^9$	+
$\dfrac{M\hat{S}C^e_{(m)}}{\hat{c}^e_{(m)}}$	+	-	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	$?^5$	$?^{11}$	+
$\dfrac{T\hat{T}C^e_{(m)}}{S\hat{D}C^e_{(m)}}$	$-^1$	+	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 a 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 (29). 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 $N\in(1,5/4)$.

comparative static results. One striking result is that an increase in β can lead to a decrease in total schedule delay costs, which implies that schedule delay falls more than in proportion to the rise in β. Recall that in the example presented in Section 4.1.1, a second departure mass starts to form when 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 departure mass 2 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 departure mass 2 to shrink from $\dfrac{\Delta}{2 - \theta_0}$ to $\dfrac{\Delta - (\theta_1 - \theta_0)}{2 - \theta_1}$.

As θ increases, a point is reached where the number of commuters in departure mass 2, 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 production process. More specifically, by deconcentrating departures across departure masses, the rise in α causes travel time in departure mass 1 to fall. If travel is severely hypercongested in departure mass 1, 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 departure mass 1 experience no schedule delay, their trip cost, which is the equilibrium trip cost, equals their travel time cost.

4.4 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. For instance, a typical driver in Bangkok spent forty four hours per year sitting in gridlock (Gibbs, 1997). Earlier in the paper, in Table 1 (section 4.1.2), we presented a numerical example, but discussed only those cases where equilibrium entailed two departure masses. Now that we have generalized our analysis, we discuss the remaining cases, with larger N, which provides insight into the behavior of our model under heavily congested conditions. In the numerical example, we assumed that $\theta = 1/2$. From (25), 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 departure mass 1 is entirely travel time cost, and since the cost of travel time is normalized at 1, travel time in departure mass 1 too is related to N according to $2^{N+1} - 1$, while travel speed as function of N is $\frac{1}{2^{N+1} - 1}$. In particular, speed in departure mass 1, 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, the number of departure masses in the restricted equilibrium is $m = N + 1$. The severity of congestion,

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

then reduces to 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, $PCC(N)$. Private congestion cost is defined as trip cost minus trip cost with no congestion, which is 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

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

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 $\theta = \frac{1}{2}$,

$$\hat{D}(N) = \frac{(2^{N+1}-1)(2-(\frac{1}{2})^N)}{3}.$$

5 Social Optimum with Identical Individuals

Our focus in this paper is on the no-toll equilibrium. A later paper is planned that focuses on the social optimum. Here we restrict analysis of the social optimum to departure patterns in which all departures occur in departure masses with contiguous travel time intervals, referring to the resulting allocation of commuters to departure times as the restricted social optimum. This allows comparison with the restricted no-toll equilibrium treated in the previous section. While we have not proved that the social optimum with no late arrivals and a cost function that is linear in travel time and schedule delay entails departure masses, the restricted social optimum that we consider is at least a local social optimum in the sense that the marginal social cost of no commuter can be reduced by altering her departure time. Furthermore, we restrict our analysis to two departure masses. We do this since, in contrast the restricted no-toll equilibrium, we have been unsuccessful in deriving closed-form solution for more than two departure masses.[13] Thus, the results of this section are suggestive rather than exhaustive.

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 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

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

[13] We were able to obtain closed-form solution for the restricted no-toll equilibrium by exploiting the recursive relationship (24), in particular by finding the infinite series solution and then calculating the finite series solution as the difference between two infinite series solution. Unfortunately, we have been unsuccessful in applying the same "trick" to obtain closed-form solution for the social optimum.

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

$$MSC_1^2 = \frac{1}{(1-n_1^2)^2} + \frac{\theta n_2^2}{(1-n_1^2)^2}, \tag{30}$$

$$MSC_2^2 = \frac{1}{(1-n_2^2)^2} + \frac{\theta}{1-n_1^2}. \tag{31}$$

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

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

Thus

$${}^*c_1^2 = \frac{1}{1 - {}^*n_1^2} = \frac{1+A}{A(2-N)} \tag{32}$$

$${}^*n_2^2 = N - {}^*n_1^2 = \frac{A-(1-N)}{1+A} \tag{33}$$

$${}^*c_2^2 = \frac{1}{1 - {}^*n_2^2} + \frac{\theta}{1 - {}^*n_1^2} = \frac{(A+\theta)(1+A)}{A(2-N)} \tag{34}$$

$$\begin{aligned} {}^*MSC_1^2 &= \frac{1+\theta n_2^2}{(1-n_1^2)^2} = \frac{(A+1+\theta)(1+A)}{A(2-N)^2} \\ {}^*MSC_2^2 &= \frac{1}{(1-n_2^2)^2} + \frac{\theta}{1-n_1^2} = \frac{(A+1+\theta)(1+A)}{A(2-N)^2}. \end{aligned} \tag{35}$$

By setting ${}^*n_1^2 = N$, we obtain the critical population density at which there is a switch from one to two departure masses at the social optimum, $N^*_{1,2}$:[14]

$$N^*_{1,2} = \frac{(2+\theta) - (\theta^2+4)^{1/2}}{2}.$$

We next solve for $N^*_{2,3}$. To calculate this, we solve for the N for which ${}^*MSC^3(N^*_{2,3}) ={}^* MSC^2(N^*_{2,3})$. Now, when $N = N^*_{2,3}$, ${}^*n^3_3 = 0$, and

$${}^*MSC^3(N^*_{2,3}) = 1 + \frac{\theta}{1 \ - \ {}^*n^2_2(N^*_{2,3})} + \frac{\theta}{1 \ - \ {}^*n^2_1(N^*_{2,3})}. \tag{36}$$

The first term is the marginal social travel time of a commuter in departure mass 3 when ${}^*n^3_3 = 0$ and the last two terms are the marginal schedule delay cost of this commuter. Now, from (31),

$${}^*MSC^3(N^*_{2,3}) = \ {}^*MSC^2_2(N^*_{2,3}) = \frac{1}{(1- \ {}^*n^2_2(N^*_{2,3}))^2} + \frac{\theta}{1- \ {}^*n^2_1(N^*_{2,3})}. \tag{37}$$

Comparing (36) and (37) permits a closed-form solution for ${}^*n^2_2(N^*_{2,3})$. Substituting this into $MSC^2_2(N^*_{2,3}) = MSC^2_1(N^*_{2,3})$ from (30) and (31) gives a closed-form solution, albeit an ugly one, for ${}^*n^2_1(N^*_{2,3})$.

The results we have obtained for one or two departure masses are recorded below.

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 restricted social optimum than in the restricted no-toll equilibrium, which is consistent with the downward shift in $MSC^e(N)$ at $N^e_{1,2}$ as population increases that was noted in the previous section. In the no-toll equilibrium with $N = N^e_{1,2}$, 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. The travel time externality cost is higher by adding a commuter to departure mass 1 than to departure mass 2. Furthermore, adding a commuter to departure mass 1 generates a schedule delay externality while adding a commuter to departure mass 2 does not.

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

For all sets of parameters,[15] in the restricted social optimum as population density increases departure

[14] There is a negative root and a positive root. The positive root has ${}^*n^2_1 > 1$, which does not make economic sense.

[15] We have imposed the restriction that $\theta \in (0,1)$. As in the bottleneck model, this is a necessary condition for equilibrium to exist. This can be seen by comparing (14) and (15) in the case of two departure masses. Since the schedule delay of departure mass 2 equals the travel time of departure mass 1, with $\theta > 1$ the schedule delay cost alone of traveling in departure mass 2 would exceed the trip cost of traveling in departure mass 1, which comprises only travel time cost. Thus, all commuters would choose to travel in departure mass 1, but if $N > 1$ this would not be possible. In contrast, the restricted social optimum in the isotropic model is well defined with $\theta > 1$ (as it is in the bottleneck model). In this case, hypercongestion can arise in the social optimum. To see this, consider the limiting case in which θ approaches infinity. With $N < 1$, it would be optimal for all commuters to travel in a single mass since all would be on time.

mass 2 is created before hypercongestion arises in departure mass 1. In contrast, there is a single hypercongested departure mass in the no-toll equilibrium when N is greater than $\frac{1}{2}$ but less than θ.

Result 3. $N^e_{2,3} = 3\theta - \theta^2 > N^*_{2,3}$ *for* $\theta \in (0,1)$

On the basis of the intuition provided for Result 1, it in natural to conjecture that $N^e_{m,m+1} > N^*_{m,m+1}$.

Result 4. $\frac{1}{2} \; > \; {}^*n^2_1 \; > \; {}^*n^2_2 \quad$ *for* $\quad \theta \in (0,1)$

It is natural to conjecture that ${}^*n^m_i \; > \; {}^*n^m_{i+1}$ for all m and for all i from 1 to $m-1$ as in the equilibrium, that traffic density increases monotonically during the morning rush hour. It is also natural to conjecture that hypercongestion does not occur in the restricted social optimum for $\theta \in (0,1)$.

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,

$$
{}^*\tau^2_i \, = \, {}^*MSC^2_i \, - \, {}^*c^2_i, \; i = 1, 2,
$$

which can be calculated from (32), (34), and (35):

$$
\begin{aligned}
{}^*\tau^2_1 &= \frac{(1+A)(A+N-1+\theta)}{A(2-N)^2} \\
{}^*\tau^2_2 &= \frac{(1+A)(1+(N-1)(A+\theta))}{A(2-N)^2}
\end{aligned}
$$

We now construct a numerical example with two departure masses in the social optimum, and compare the social optimum and the no-toll equilibrium. To obtain particular qualitative results, we assume that $\theta = 0.9$, in contrast to $\theta = 0.5$, which was assumed in the earlier example. Otherwise, the parameters are the same as in the earlier example. Per the procedure outlined above, we obtain $N^*_{2,3} = 0.776$, and assume $N = 0.75$ in order to obtain almost as congested as possible a social optimum with two departure masses. 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 numerically the social optimum and the no-toll equilibrium for 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 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

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	4	16	3	3	0	4
social opt.	2	0.4148	0.3352	1.709	3.042	3.801	1.729	1.213	0.5155	3.213

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.

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 social optimum, in contrast, commuters distribute themselves between two departure masses. Travel speed in the more congested departure mass is 8.778 mph compared to a free-flow travel speed of 15 mph, while that in departure mass 2 is 9.972 mph. Commuters in departure mass 1 experience a travel time of 34.17 minutes, for a normalized cost of 1.709, while commuters in mass 2 have a 30.09 minute commute, which, along with the 34.17 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 of that in the no-toll equilibrium. Even though the optimum has two departure masses, its rush hour is 64.26 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 higher 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 example they all depart at the same time), but the concentration of departure times creates such severe hypercongestion that the length of the rush hour increases. Congestion tolling, by causing commuters to face the external cost, results in them deconcentrating their departures, eliminating hypercongestion and shortening the rush hour.

Table 5 also illustrates another 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 2.092 (\$13.93) for commuters traveling in departure mass 1 and 0.7586 (\$5.05) for those traveling in departure mass 2. In normalized units, the total revenue is 1.122, while the efficiency gain from congestion tolling is 1.271 (\$11.30 per commuter). Thus, the example illustrates the very considerable efficiency gains achievable through congestion tolling 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.

6 Extensions

In this section we present only the most straightforward extensions and treat only the no-toll equilibrium, not the social optimum.

6.1 Price-sensitive Demand

As in the bottleneck model, the function relating trip cost to the number of commuters in the restricted no-toll equilibrium can be regarded as a reduced-form supply curve. The reduced-form supply function for the restricted no-toll equilibrium is given by (26) and (28). Eq. (28) identifies the population intervals over which there are various numbers of departure masses. For each of these population intervals, (26) relates trip cost to population. The reduced-form supply curve is upward sloping over the entire range of population. Adding to this a demand curve relating the number of commuters to trip cost permits solution of the restricted no-toll equilibrium with price-sensitive demand, as shown in Figure 8. An increase in demand results in movement up the upward-sloping reduced form supply curve, and hence to an increase in both the equilibrium trip cost and the equilibrium number of commuters. This stands in contrast to the stationary-state reduced-form supply curve derived in Arnott and Inci (2010), which is backward bending. There high *stationary-state* (flow) demand relative to capacity could be accommodated only through the trip price rising to a level entailing hypercongestion. Here there is the extra margin of the length of the rush hour that adjusts to achieve equilibrium when (stock) demand for trips over the rush hour is high relative to capacity. Thus, one may say of the no-toll equilibrium in this paper's isotropic model of rush-hour traffic dynamics that, while hypercongestion may exist at the peak of the rush hour (in the sense that traffic density exceeds capacity density) it does not occur at the aggregate level (in the sense that the aggregate supply curve for trips is upward sloping).

The social optimum with price-sensitive demand and identical commuters occurs where the marginal social benefit of trips over the morning rush hour equals the marginal social cost. Diagrammatically, this corresponds to the point of intersection of the demand curve for trips and the marginal social cost of trips

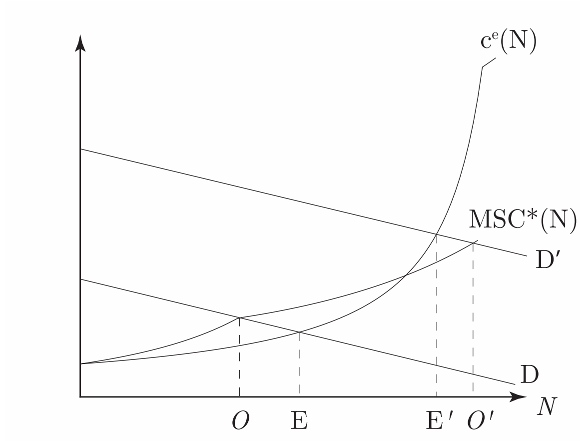


Figure 8: No-toll equilibrium trip cost and marginal social cost with price-sensitive demand

(O and O'), shown in Figure 8. It is important to note that the marginal social cost curve for the social optimum is different from that in the no-toll equilibrium. As in the bottleneck model, efficient pricing in the isotropic model has two effects. The first is to alter the timing of departures over the rush hour, holding the number of commuters fixed, and the second is to ensure the efficient number of trips.

A final point to note is that the reduced-form supply curve for trips in the no-toll equilibrium intersects the marginal social cost for trips in the social optimum, as shown in Figure 8.[16] For population levels below the population level where the two curves intersect, trip price is lower and the number of trips higher in the no-toll equilibrium than in the social optimum ($E > O$), and the revenue raised from the optimal time-varying toll is greater than the welfare gains it achieves. For population levels above the population level where the two curves intersect, trip price (trip cost plus the optimal time-varying toll) is lower and the number of trips higher in the social optimum than in the competitive equilibrium ($O' > E'$), and the revenue raised from the optimal time-varying toll is less than the welfare gain it achieves. Thus, as in Arnott and Inci (2010), in cities in which demand is high relative to capacity, optimal congestion tolling would be beneficial even if the toll revenue were completely squandered.

6.2 Late Arrivals

Earlier we simply asserted that the restricted no-toll equilibrium with no late arrivals permitted is the limiting case of the corresponding no-toll equilibrium with late arrivals permitted but an infinite value of time late. The proof proceeds in two steps. The first step is to prove that, with the $\alpha - \beta - \gamma$ cost function, there is an arrival mass at t^*. This can be proved using the same line of argument used to prove Lemma 2 in section 3.3. The discontinuity in $\dot{T}(t)$ for t corresponding to arrival at time t^* requires that there be an

[16] All the points made in this paragraph were made for Arnott's bathtub model in Arnott (2013).

arrival mass at t^*. The second step is to show that the no-toll equilibrium derived in section 4 is achieved as the limiting case of the equilibrium with the $\alpha - \beta - \gamma$ cost function as the unit time late cost approaches infinity. We now demonstrate this somewhat informally.

To keep the indexation consistent with that earlier in the paper, we shall denote by $i = -1$ the first late departure mass, $i = -2$ the second late departure mass, and so on, with the departure masses arriving early or on time are indexed as before.[17]

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. 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. 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. Earlier we proved that, when late arrivals are not permitted, a second departure mass starts to form when $N = \theta$. When late arrivals are permitted, the same result holds when the second departure mass to form is a second early departure mass. 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. In the limit as γ approaches ∞, as long as population density is finite, a late departure mass never forms, and the equilibrium is the same as when late arrivals are not admitted.

6.3 Commuters Who Differ in θ

In the no-toll equilibrium of the bottleneck model with no late arrivals, commuters order themselves over the rush hour according to the value of β relative to α, that is according to θ, independent of the absolute size of the units, those with a higher θ departing later. The same is true of the isotropic model. The logic that was applied earlier to the case of identical commuters to prove that there is a mass of arrivals at t^* can be adapted to the case of commuters who differ in θ. Thus, there exists an equilibrium in which all departures occur in masses with contiguous travel time intervals. The analysis is easier if we assume that there is a continuous

[17] It would be more intuitive to index early departure masses with a $-$ and late departure masses with a $+$. We have not done so only to achieve notational consistency in the paper.

cumulative distribution of commuters, $F(\theta)$. 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 departure mass 1, 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.

$$\frac{1}{1 - N(1 - F(\theta_{1,2}))} = {}^{e}c_1^2(\theta_{1,2}) = {}^{e}c_2^2(\theta_{1,2}) = \frac{1}{1 - NF(\theta_{1,2})} + \frac{\theta_{1,2}}{1 - N(1 - F(\theta_{1,2}))} \tag{38}$$

$$1 + \frac{\underline{\theta}}{1 - NF(\theta_{1,2})} + \frac{\underline{\theta}}{1 - N(1 - F(\theta_{1,2}))} = {}^{e}c_3^3(\underline{\theta}) \geq {}^{e}c_2^2(\underline{\theta}) = \frac{1}{1 - NF(\theta_{1,2})} + \frac{\underline{\theta}}{1 - N(1 - F(\theta_{1,2}))} \tag{39}$$

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 into (38) 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 (39) 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 value of θ 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.

7 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, the model has a serious deficiency. It does a bad job of modeling downtown traffic congestion when it is at its worst. In particular, it assumes that throughput is the same whether downtown traffic congestion is moderate or severe. Experts have long believed that throughput falls sharply when congestion becomes severe, but only recently was this confirmed empirically.

Since downtown traffic congestion is a critical problem when it is severe, it is time to move beyond the bottleneck model.

Urban transportation economists have long understood how the bottleneck can be extended or replaced to treat hypercongestion – severe congestion in which throughput falls as traffic density increases. The problem is that, even though easy to write down, all "proper" models – those that respect rational economic behavior and the physics of traffic congestion – have proven to be analytically intractable. All result in delay differential equations with an endogenous delay with both an initial and a terminal condition on the state variable, whose study is at the research frontier in applied mathematics.

Several papers have attempted to break this impasse by introducing approximations that restore tractability. None has gained wide acceptance since their approximations have been challenged on a priori grounds. This paper took a different tack, working out a special case that generates a closed-form solution for a no-toll equilibrium without making any approximations. The special case adapts the simplest bottleneck model, which has identical commuters, a trip cost function that is linear in travel time and schedule delay, and no late arrivals, to an isotropic downtown area with flow congestion. As in the bottleneck model, the central equilibrium condition is that no commuter can reduce her trip price by altering her departure time. This special case has a no-toll equilibrium in which all departures are in contiguous masses, which we referred to as a restricted no-toll equilibrium.

The big advantages of the special case are that it entails no approximations, its properties can be explored analytically, and its economics and physics are intuitive. Thus, it provides a promising starting point. Whether or not this promise is realized will depend on the robustness of the special case, which remains to be explored. Robustness has several aspects: whether the restricted no-toll equilibrium is the unique equilibrium for the special case, whether its qualitative properties are robust, and how far the special case can be extended in the direction of realism without sacrificing closed-form solution or at least analytical tractability. Since departure masses are not observed and would generate turbulence, which the model does not incorporate, it remains to be seen whether the departure masses can be smoothed out – by, for example, introducing a distribution of desired arrival times – without sacrificing tractability.

Developing a theory of rush-hour traffic dynamics that is economically and physically sound, mathematically tractable, and admits hypercongestion has proved to be difficult. This paper falls far short of fully developing such a theory. Rather, it introduces a promising line of attack that merits further development and exploration.

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

Beckmann, M., McGuire, C., and Winsten, C. (1956). *Studies in the Economics of Transportation*. Yale University Press.

Fosgerau, M. (2015). Congestion in the bathtub. *Economics of Transportation*, 4(4):241–255.

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.

Gibbs, W. W. (1997). Transportation's perennial problems. *Scientific American*, 277:54–57.

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

Small, K. A. (2015). The bottleneck model: An assessment and interpretation. *Economics of Transportation*, 4(12):110 – 117.

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.

Walters, A. A. (1961). The Theory and Measurement of Private and Social Cost of Highway Congestion. *Econometrica*, 29(4):676–699.

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 cost in mass i conditional on m masses
$c_{(m)}$	common trip cost in no-toll equilibrium with m departure masses
e	equilibrium
$e(t)$	entry rate as a function of departure time
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^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)
$x(t)$	exit rate as a function of departure time
A	intermediate variable ($\equiv (1-\theta(1-N))^{1/2}$)
$A(m,\theta)$	constant term ($\equiv \frac{1-\theta}{\theta}\left[\frac{1-(1-\theta)^m}{(1-\theta)^m}\right]$)
D	set of departure times
$D_{(m)}$, $\hat{D}_{(m)}$	normalized, unnormalized duration of rush hour with m masses
E	elasticity
$E(t)$	cumulative number of commuters who have entered by time t
$F(\theta)$	cumulative distribution function of θ
L	unnormalized trip distance (normalized trip distance equals 1)
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
SDC, $S\hat{D}C$	normalized, unnormalized total schedule delay cost
$S(N)$	alternative measure of the severity of congestion ($\equiv \frac{MCE(N)}{c(N)-1}$)
$S(t)$	trip time for a driver who arrives at time t
$T(t)$	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, $TT\hat{C}$	normalized, unnormalized total travel time cost
$X(t)$	cumulative number of commuters who have exited by time t

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_i^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