Title
A Bathtub Model of Traffic Congestion

Author
Arnott, Richard

Publication Date
2011-02-01

University of California Transportation Center
UCTC-FR-2011-18

A Bathtub Model of Traffic Congestion

Richard Arnott
University of California, Riverside
September 2011

A BATHTUB MODEL OF TRAFFIC CONGESTION

Richard Arnott*

February 2011

Abstract

In the standard economic models of traffic congestion, traffic flow does not fall under heavily congested conditions. But this is counter to experience, especially in the downtown areas of most major cities during rush hour. This paper presents a bathtub model of traffic congestion. The height of water in the bathtub corresponds to traffic density, velocity is negatively related to density, and outflow is the product of density and velocity. Above a critical density, outflow falls as density increases. The model indicates that, when demand is high relative to capacity, applying an optimal time-varying congestion tolls generates benefits that are considerably larger than those obtained from the standard models and exceed the toll revenue collected.

Keywords: traffic congestion, tolls, flow, density, velocity, hypercongestion, trip-timing equilibrium

JEL codes:

Department of Economics, University of California Riverside, Riverside, CA 92521
richard.arnott@ucr.edu.

*I would like to thank the University of California Transportation Center (grant number xxxx) for financial support, and Tola Kokoza and Huiling Zhang for excellent research assistance. In writing this paper, I have drawn on the knowledge gained from earlier collaborations with André de Palma, Elijah de Palma, Eren Inci, Marvin Kraus, Robin Lindsey, and John Rowse, as well as conversations with Kenneth Small. Finally, dedicating this paper to the memory of William Vickrey is appropriate since a conversation with him was its inspiration.

1 Introduction

There are two standard models of traffic congestion. In the first, which is familiar from undergraduate textbooks, trip cost increases in traffic flow, approaching infinity as capacity is reached. In the second, the bottleneck model (William Vickrey, 1969; Arnott, André de Palma, and Robin Lindsey, 1993), which treats the evolution of congestion over the rush hour, congestion is modeled as a deterministic queue behind a bottleneck of fixed flow capacity. In neither model does traffic flow as congestion increases. But casual experience and common sense[1] suggest that, especially in the downtown areas of major cities during rush hour, traffic flow falls as congestion increases (with zero flow in the limit as traffic becomes completely jammed). Surprisingly, documentation of this phenomenon has started only recently (Carlos Daganzo and Nikolas Geroliminis, 2008).

This paper develops a bathtub model of traffic congestion that captures this phenomenon[2]. Think of the bathtub as being Manhattan. In the morning rush hour, cars join the traffic on Manhattan streets, entering either across the bridges into Manhattan or from parking garages in Manhattan. These cars correspond to the flow of water into the bathtub. The density of traffic corresponds to the height of water in the bathtub, and the outflow corresponds to cars exiting Manhattan traffic, either across the bridges out of Manhattan or into parking garages in Manhattan. Traffic velocity is inversely related to traffic density, flow equals density times velocity, and, above a critical density, flow is negatively related to density[3]. In terms of the bathtub analogy, as the water level in the bathtub rises, a critical water level is reached above which the rate at which water drains out of the bathtub is

[1] There is abundant anecdotal evidence of downtown rush-hour traffic speeds of 3 to 5 mph in cities such as Moscow, Shanghai, Bangkok, Djakarta, Istanbul, and London before the congestion toll, but no reliable documentation (with average speeds below this level it would be faster to walk). And it seems implausible that traffic flow can be close to capacity at such speeds.

[2] The model of this paper was inspired by a conversation with Vickrey a few years before his death. He coined the term "bathtub model of traffic congestion", thinking of a bathtub and of Manhattan being shaped like a bathtub. Vickrey never published his model, and the unpublished notes he left were incompletely developed.

[3] In Vickrey's terminology, traffic is "hypercongested" when flow and density are negatively related, which corresponds to traffic jam situations.

inversely related to the height of water in the bathtub. In the early morning rush hour, the flow of water into the bathtub exceeds the rate at which the bath can drain, and the water level rises. If the water rises much above the critical level, the water in the bathtub takes a long time to drain. A planner regulating the flow of water into the bathtub would ensure that the water never rises above the critical level.

The bathtub analogy does have the ring of truth to it. But only recently has it been empirically investigated. Traffic engineers started to collect detailed data on freeway traffic flow using loop detectors in the early 1980's (Fred Hall, Brian Allen, and Margot Gunter, 1986). Analysis of these data (for example, Michael Cassidy and Robert Bertini, 1999) suggests that freeways contain bottlenecks whose discharge rates fall only slightly as the length of the queue behind them increases. Based partly on these analyses, the prevailing wisdom in urban transportation economics (Kenneth Small and Erik Verhoef, 2007) is that the aggregative or macroscopic[4] behavior of rush-hour traffic in metropolitan areas is broadly consistent with the bottleneck model. Only in the last five years have comparable data been collected for downtown city streets in major cities. Analysis of these data (Geroliminis and Daganzo, 2008; Daganzo, Vikash Gayah, and Eric Gonzales, 2011) provides strong support for what these authors refer to as the existence of a "stable urban-scale macroscopic fundamental diagram" – a stable graph relating traffic flow to traffic density at the level of the downtown area for a particular city, which includes a portion corresponding to heavily congested conditions over which traffic flow falls as density increases.

Section 2 briefly introduces traffic flow theory and the economic theory of traffic congestion. Section 3 presents the basic bathtub model, and applies it to determining traffic flow equilibrium in the morning rush hour in the absence of congestion tolls. Section 4 solves for the corresponding social optimum and the time-varying toll that decentralizes it. Section 5 compares the no-toll equilibrium and the social optimum, policy insights and directions for

[4]The prevailing wisdom has recently been disputed by Arnott and Eren Inci (2010). There is no dispute that hypercongestion can occur as a localized phenomenon. The dispute concerns whether hypercongestion can or does occur at the level of the downtown or metropolitan area.

future research, and section 6 concludes.

2 Background

The earliest well-articulated theory of traffic flow is known variously as kinematic wave theory, LWR traffic flow congestion theory, and classical flow congestion (James Lighthill and Gerald Whitham, 1955, and Paul Richards, 1956). The theory combines three elements: i) the fundamental identity of traffic flow theory, that flow equals density times velocity: $q = vk$; ii) the equation of continuity, which is the conservation of mass for a fluid; and iii) an assumed technological relationship between velocity and density; in this paper, Greenshields' Relation is assumed (Bruce Greenshields, 1935), which posits a negative linear relationship: $v = v_0\left(1 - \frac{k}{k_j}\right)$, where v_0 is free-flow velocity and k_j jam density. Elements i) and iii) imply that flow is related to density according to the formula:

$$q = kv_0\left(1 - \frac{k}{k_j}\right) \tag{1}$$

which describes a parabolic relationship, as shown in Figure **??**. For each level of flow, there are two levels of density, the lower one corresponding to congested travel (in Vickrey's terminology), which corresponds to free-flowing traffic, the higher one to hypercongested travel, which corresponds to traffic jam situations. The maximum level of flow, $q_c = \frac{v_0 k_j}{4}$, is referred to as capacity flow[5].

[5] In the transportation science literature, relationships such as (**??**) are termed macroscopic since they describe the behavior of traffic at an aggregate level, without reference to the behavior of individual cars. Microscopic models, in contrast, posit the behavior of the individual car, and then aggregate. The earliest microscopic models are simple car-following models, which relate a car's velocity or acceleration to the position and velocity of the car immediately ahead. When traffic is in steady state, car-following models yield simple relationships between velocity and density but not generally otherwise. As a result of the avalanche of traffic flow data obtained from sensors, as well as the development of traffic microsimulation models, there has been an explosion of microscopic models. Broadly speaking, there are now two schools in traffic flow theory (Sven Maerivoet and Bart De Moor, 2008). The European School insists on aggregation from microscopic models. The Berkeley School, in contrast, argues that traffic flow is so complex that, at an aggregate level at least, a more fruitful approach is to assume a technological relationship between flow and density even out of steady state. This paper conforms to the Berkeley School.

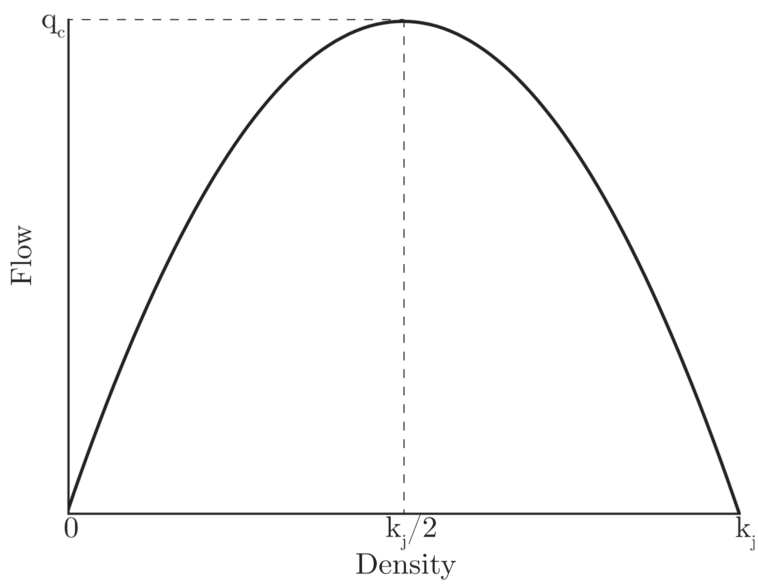

Figure 1: Flow-density diagram with Greenshields' Relation

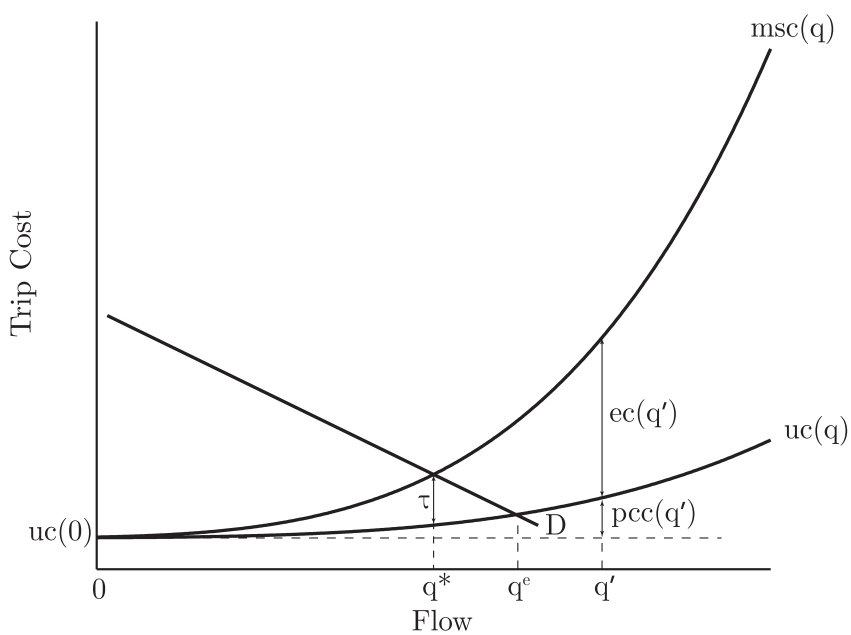

Figure 2: Textbook congestion diagram

In the transportation economics literature, there are two broad approaches to the study of rush-hour traffic congestion. The first, deriving from Arthur Pigou (xxxx) and Frank Knight (xxxx), applies the static theory of externalities to traffic congestion. Figure **??** presents the standard textbook analysis, which assumes identical cars and drivers. The user cost (or marginal private cost) of an individual driver[6] is an increasing and convex function of traffic flow: $uc(q)$. Since all drivers face the same user cost, user cost corresponds to average cost in cost theory, from which a graph relating marginal social cost to traffic flow can be derived: $msc(q)$. Since a driver slows down other drivers without compensating them, he generates an externality cost, $ec(q)$, equal to the difference between marginal social cost and user cost. The optimal level of flow, q^*, occurs where the marginal social cost curve intersects the demand curve, and the untolled equilibrium level of flow, $\hat{q}$, where the user cost curve intersects the demand curve. The externality can be internalized by imposing a congestion toll, τ, equal to the difference between the marginal social cost and the user cost of a trip, evaluated at the socially optimal level of traffic flow. Since it assumes that velocity is negatively related to flow, this approach ignores the possibility of hypercongestion. Most studies measuring the benefits of congestion pricing (e.g. David Anderson and Herbert Mohring, 1997) adopt this approach, and many assume the Bureau of Public Roads (BPR) link congestion function:

$$T = T_0 + T_1 \left(\frac{q}{k}\right)^4, \tag{2}$$

where T is travel time, q flow, $pcc(q) = uc(q) - uc(0)$, and T_0 and T_1 are link-specific constants. Defining private congestion cost to be the increase in a driver's trip cost due to congestion, (**??**) has the property that the ratio of the externality cost to the private congestion cost equals[7] 4.0.

[6]Different vehicle types are accommodated by treating them as passenger-car equivalents (PCE's).

[7]Consider a driver whose journey to work takes 15 minutes in uncongested traffic and 30 minutes in congested traffic. Her private congestion cost is the value of 15 minutes of time and, with the BPR function, she imposes an externality cost equal to the value of 60 minutes of time.

The second broad approach to the study of rush-hour traffic congestion entails application of Vickrey's bottleneck model. There are N identical commuters who must travel from home to work in the morning rush hour along a single road, and who have a common work start time, t^*. In the absence of a toll, there are two components of trip cost, travel time cost and schedule delay cost, which is the cost associated with arriving at work either early or late. Along the road there is a single bottleneck with fixed flow capacity, s. Waiting in a queue to pass through the bottleneck is the only cost of traveling on the road. Where α is the unit cost of travel time, β the unit cost of time early, γ the unit cost of time late, t departure time from home and $T(t)$ the travel time as a function of t, trip cost as a function of time is

$$\begin{aligned} c(t) &= \text{travel time cost}(t) + \text{time early cost}(t) + \text{time late cost}(t) \\ &= \alpha T(t) + max(0, \beta(t^* - t - T(t)) + max(0, \gamma(t + T(t) - t^*))). \end{aligned} \tag{3}$$

A trip-timing equilibrium is achieved when no commuter can reduce her trip price (which equals trip cost in the absence of a toll) by altering her departure time. If $\gamma < \alpha < \beta$ (which is supported by the empirical evidence: Small, 1982), a trip-timing equilibrium exists in which trip price is equalized over the departure interval. This is achieved via adjustment in the length of the queue behind the bottleneck, which is in turn achieved via adjustment in the departure rate from home. The commuter who arrives exactly on time faces no schedule delay cost and in the absence of a time-varying toll must therefore experience the highest travel time cost. The commuters who depart at either the beginning or end of the rush hour experience the highest schedule delay cost and no travel time cost. Thus, in the absence of a time-varying toll, the queue increases (linearly) in the early morning rush hour and then dissipates (linearly) in the late morning rush hour.

Figure **??** displays the no-toll equilibrium in terms of cumulative departures from home, $D(t)$, and cumulative arrivals at work, $A(t)$. The vertical distance between $D(t)$ and $A(t)$ is the queue length as a function of t, and the horizontal distance between the curves for the

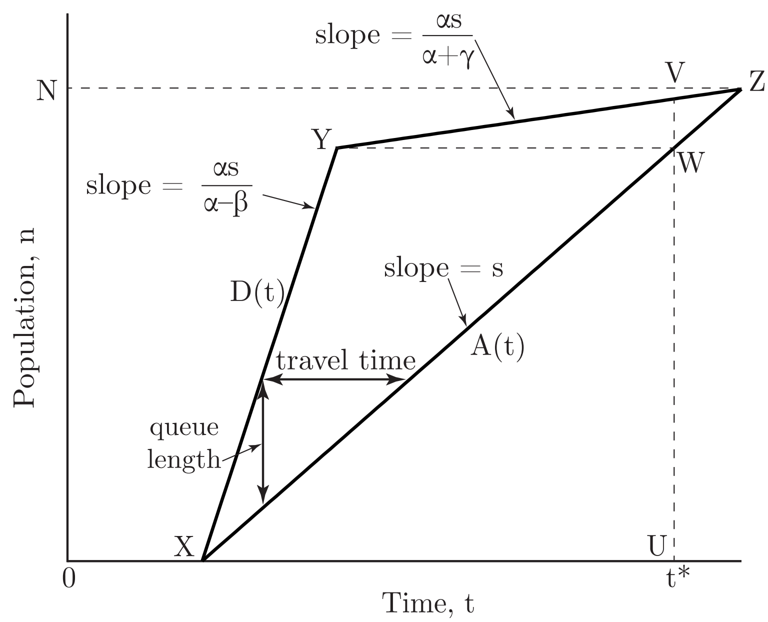

Figure 3: Cumulative inflow and outflow in the bottleneck model: No-toll equilibrium

nth driver indicates the time she spends in the queue. The slope of the cumulative departure function is the departure rate from home, and the slope of the cumulative arrival function the arrival rate at work, which equals the bottleneck's flow capacity, s. Total travel time is given by the area of the triangle XYZ, total time early by the area XWU, and total time late by the area WVZ.

Four properties of the model are particularly noteworthy:

1. Consider the effect of applying a time-varying toll equal to travel time cost in the no-toll equilibrium. The toll simply replaces travel time cost, completely eliminating the queue, with all commuters experiencing the same common trip price as in the no-toll equilibrium. Thus, the efficiency gain from applying this (optimal) time-varying toll equals the revenue it collects, so that commuters are better off from tolling if any fraction of the toll revenue collected is spent to their benefit.

2. Figure **??** plots reduced-form cost functions, in a diagram similar to Figure **??**, except

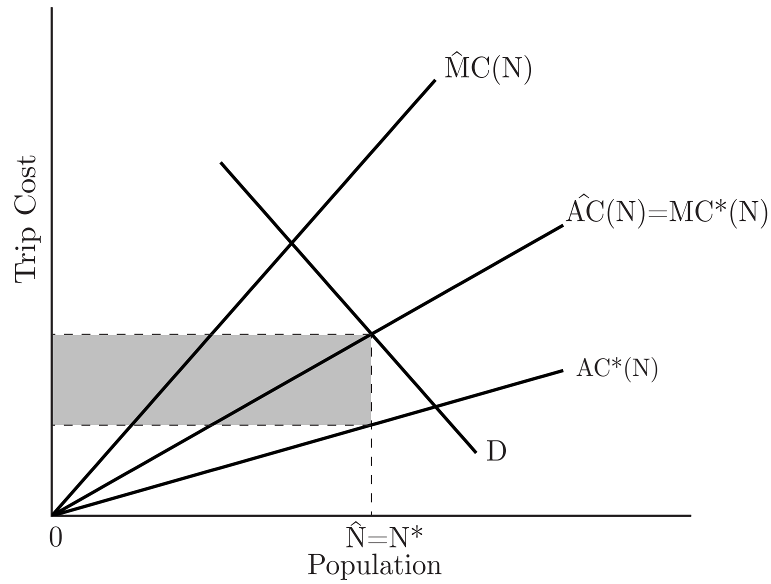

Figure 4: Reduced-form cost functions for the bottleneck model. The grey region shows the efficiency gain from applying this optimal congestion toll.

that the population of individuals over the rush hour, N, is on the x-axis rather than flow[8]. See Figure **??**. Since a doubling of population doubles the base and height of the triangles in Figure **??**, for the no-toll equilibrium, user cost, $\hat{UC}(N)$, and marginal social cost, $\hat{MSC}(N)$, are linear functions of population, with the marginal social cost function having double the slope of the average cost function. Accordingly, the textbook model and the bottleneck model can be qualitatively reconciled by interpreting the standard model as a reduced form of the bottleneck model, with flow replaced by population. Even with this reconciliation, however, there are important differences between the models.

3. In the textbook model so reinterpreted, with the BPR link congestion function, the elasticity of private congestion cost with respect to population is 4.0, while in the bottleneck model it equals 1.0. The value of this elasticity is important since it equals the ratio of the optimal toll (optimal average toll in the bottleneck model) to private congestion cost. The

[8] When demand is sensitive to price, the no-toll equilibrium occurs where the no-toll user cost curve intersects the demand curve and the optimal-toll equilibrium where the optimal-toll marginal social cost function interests the demand curve. Since the no-toll user cost curve coincides with optimal-toll marginal social cost function, the number of trips in the no-toll equilibrium equals that in the optimal-toll equilibrium.

model to be presented in this paper points to a way in which these very different elasticity values can be reconciled.

4. In the textbook model, the benefit from congestion tolling derives from the reduction in the number of trips it induces. In the bottleneck model, there is the added benefit that tolling reallocates trips over the rush hour, causing the user cost and marginal social cost curves to shift down.

3 The Bathtub Model and Its No-toll Equilibrium

The bottleneck model is now well entrenched in both transportation economics and transportation engineering. There is no doubt that the choice of trip time is central to rush-hour traffic congestion, but one wonders how accurately the bottleneck model describes rush-hour traffic congestion at the aggregate level. Its property that a doubling of population, holding fixed the traffic network, results in only a doubling of each driver's cost does not square with casual observation[9]. One reason the bottleneck model has not been extended to provide a richer treatment of traffic congestion was suggested earlier freeway data tend to support the simple bottleneck models treatment of congestion while only recently have data on downtown congestion been collected systematically. Another is that this extension has been attempted but has proved to be analytically difficult.

Vernon Henderson (1981) replaced the bottleneck with a single highway link subject to flow congestion, but in order to obtain analytical tractability resorted to the unappealing assumption that a car's speed along the link is inversely related to the entry flow rate at

[9]Casual empiricism suggests that time last cost at least is a convex function of time late. Suppose, for the sake of argument, that time late cost as a function of arrival time t' equals $c_1(t^*-t')+c_2(t^*-t')^2+c_3(t^*-t')^3$, where c_1, c_2, and c_3 are positive constants, and that time early costs are $c_4[c_1(t^*-t')+c_2(t^*-t')^2+c_3(t^*-t')^3]$ with c_4 a positive constant less than one. In the bottleneck model with these time late and time early cost functions, it is straightforward to show that, as population increases, holding bottleneck capacity fixed, the elasticity of the no-toll equilibrium trip cost rises from 1.0 to 3.0. Thus, sufficient convexity of the schedule delay cost functions has the potential to reconcile the different values of the elasticity of private congestion cost with respect to population between the textbook model and the bottleneck model. The empirical evidence, however, suggests that the schedule delay cost functions show only mild convexity.

the time it enters the road, which is inconsistent with the physics of fluid flow. Xuehao Chu (1995) considered the same model, except that a car's speed along the link is a function of the exit flow at the time it exits the road. Elijah DePalma and Arnott (2010) considered the same model but with LWR flow congestion (in particular, assuming Greenshields' Relation). By solving a non-linear partial differential equation, they were able to obtain a closed-form solution for the social optimum, though not for the no-toll equilibrium. In both the social optimum and the no-toll equilibrium, the elasticity of private congestion cost with respect to population is between 0.5 and 1.0. In none of these models does hypercongestion occur, in Henderson and Chu by assumption and in DePalma and Arnott due to the physics of traffic flow along a road of uniform width.

In a paper focusing on hypercongestion, Small and Xuehao Chu (2003) replaced the bottleneck with a dense, isotropic street network. Working with an isotropic space is an appealing simplification since it eliminates spatial differentiation. A natural starting point is to assume that commuters are identical in all respects (except that they are uniformly distributed over space) including having the same trip length, L. With the entry rate as the control variable, the temporal evolution of travel time and density is given by the following three equations:

$$\int_{t}^{t+T(t)} v(k(u))\,\mathrm{d}u = L$$

$$D(t) = A(t + T(t))$$

$$k(t) = D(t) - A(t)$$

The first equation states that the integral of velocity over an individual's travel time interval equals trip length, and implicitly defines $T(t)$; the second states that cumulative departures up to time t equal cumulative arrivals up to time $t + T(t)$; and the third states that density equals cumulative inflow minus cumulative outflow. Substituting out $k(t)$ and $A(t)$ gives an equation relating travel time to the cumulative entry function:

$$\int_t^{t+T(t)} v(D(u) - D(u - T(u)))\, \mathrm{d}u = L.$$

This is a delay integral equation with an endogenous delay, a class of problems that is at the research frontier in applied mathematics. Small and Chu finessed the inability to solve this equation by assuming that trip time equals trip length divided by exit velocity: $v(t + T(t))T(t + T(t)) = L$. This assumption is unappealing for the same reason that Henderson's and Chu's simplifying assumptions are unappealing – it is inconsistent with the physics of fluid flow. Small and Chu are however to be credited with having solved for the no-toll equilibrium under their assumption. The qualitative features of the no-toll equilibrium under their simplifying assumption are broadly similar to the qualitative features of the no-toll equilibrium obtained in this paper. Small and Chu seem not to have realized, however, that the reduced-form equilibrium may exhibit hypercongestion, defined as mean density-weighted density above capacity density[10].

3.1 The Model

Consider an isotropic space in which trip distance and traffic density per unit area are both well defined[11]. There are N identical commuters per unit area, each of whom must travel from home to work by car in the morning rush hour and has start time t^*. Congestion is described by a function relating velocity to density[12]

$$v(t) = v(k(t)), \tag{4}$$

[10]In flow-price space, a steady-state hypercongested equilibrium lies on the backward-bending portion of the user cost curve. In their model and in ours, in population-price space, the reduced-form user cost curve is upward sloping. This does not however imply that an equilibrium is not hypercongested according to this definition.

[11]For concreteness, one may think of a Manhattan network of streets covering the entire globe.

[12]At some points in the paper, v is treated as a function of k and at others k is treated as a function of v. Also, $v(t)$ denotes the velocity at time t, whether or not v is treated as a function of k.

Letting $d(t)$ be the departure rate at time t, and $a(t)$ be the arrival rate at time t, the evolution of density is given by

$$\dot{k}(t) = d(t) - a(t). \tag{5}$$

Letting $\underline{t}$ denote the time of the first departure from home, and $\bar{t}$ the time of the last departure from home:

$$\begin{aligned} D(\underline{t}) = A((\underline{t}) = 0 \\ D(\bar{t}) = N \end{aligned} \tag{6}$$

To make the model tractable, the following simplifying assumption is made:

Assumption A-1 implies that[13] *there is a negative exponential distribution of trip lengths, with mean L. Furthermore, an individual does not know his trip length until immediately after his departure, and bases his departure time choice on expected trip price.*

This assumption is made solely for analytical convenience. Trip lengths are not so distributed; individuals know their trip destinations when they depart; and knowing their trips lengths, they sort themselves across departure times based on trip length (those with shorter trips traveling closer to the rush hour peak). The assumption is defensible only to the extent that it results in a model whose aggregate properties conform to observation[14]. This assumption implies that

$$a(t) = \frac{k(t)v(k(t))}{L}; \tag{7}$$

$\frac{v(t)}{L}$ is the probability that an individual terminates his trip at time t. Substituting (**??**) into

[13] Assumption A-1 results in arrivals being generated by a time-varying Poisson process. Within any infinitesimal interval of time, dt, the probability that an individual in traffic will reach his destination is $\frac{v(k(t))dt}{L}$. Thus, the number of arrivals in this interval is $k(t)\frac{v(k(t)dt}{L}$.

[14] Small and Chu consider and reject this simplifying assumption. I prefer it, however, to their assumption, which is inconsistent with the physics of fluid flow.

(**??**) yields

$$\dot{k}(t) = d(t) - \frac{v(k(t))k(t)}{L}. \tag{8}$$

The social optimum is solved as an optimal control problem, with (**??**) as the differential equation constraint. Equilibrium is solved for by adding an equal trip-price constraint, whose form will differ depending on the form of tolling applied.

3.2 The No-toll Equilibrium – Analysis

Because a commuter does not know his trip length at the time he departs from home, when deciding when to depart he faces uncertainty. Since there is no toll, the trip-price condition is that no commuter can reduce his expected trip cost by departing at a different time. The cost function is given by (**??**).

- early departures

Expected trip cost at time t is

$$\begin{aligned} Ec(t) = \int_t^{\infty} \alpha(u-t)\left(\frac{v(u)}{L}\right) P(u;t)\,du + \int_t^{t^*} \beta(t^*-u)\left(\frac{v(u)}{L}\right) P(u;t)\,du \\ + \int_{t^*}^{\infty} \gamma(u-t^*)\left(\frac{v(u)}{L}\right) P(u;t)\,du, \end{aligned} \tag{9}$$

where $P(u;t) = exp\left[-\int_t^u \left(\frac{(v(x)}{L}\right) dx\right]$ is the probability that an individual departing from home at time t has not arrived at his destination by time u. The equal trip-cost condition may be written as $\dot{Ec}(t) = 0$. Setting the derivative of (**??**) with respect to t equal to zero ($P_t(t;t) = \frac{v(t)}{L}$), and simplifying, yields the early-morning trip-timing equilibrium condition:

$$\underline{c} = Ec(t) = \frac{\alpha L}{v(t)} + \beta(t^* - t). \tag{10}$$

The intuition for (??) is that an individual should be indifferent between departing now and departing an increment of time dt later. If he departs now, there is a probability $\frac{v(t)}{L}\,dt$ that he will exit in the interval of time between t and $t\,dt$, incurring a trip cost of $\beta(t^* - t)$, and a probability $1 - \frac{v(t)}{L}\,dt$ that he will not exit in the interval, incurring a trip cost of $\alpha\,dt + \underline{c}$. Rearranging (??) gives

$$v(t) = \frac{\alpha L}{Ec(t) - \beta(t^* - t)} = \frac{c_0}{c_1 + \beta t}\ \text{, where } c_1 = \underline{c} - \beta t^*\text{, and } c_0 = \alpha L. \tag{11}$$

Since there is a one-to-one relationship between velocity and density, (??) may be rewritten as $k'(v(t))\dot{v}(t) = d(t) - \frac{k(v(t))v(t)}{L}$, and hence

$$d(t) = k'(v(t))\dot{v}(t) + \frac{k(v(t))v(t)}{L}. \tag{12}$$

$\underline{t}$ is then determined by the condition that[15] $v(\underline{t}) = v_0$.

- late departures

Under moderately congested conditions, late departures do not occur since, even with no inflow, the bathtub does not drain sufficiently rapidly to satisfy the trip-timing condition. When late departures do occur, repeating the same procedure for late departures gives

$$Ec(t) = \frac{\alpha L}{v(t)} + \gamma(t - t^*) \tag{13}$$

$$v(t) = \frac{c_0}{c_2 - \gamma t}\ \text{, where } c_2 = \underline{c} + \gamma t^*, \tag{14}$$

[15] It needs to be demonstrated that there is not a mass of departures at the beginning of the rush hour. Suppose there were so that $v(\underline{t}^+) < v_0$ by a finite amount. Consider an individual who departs an interval of time dt prior to $\underline{t}$, where a departure mass occurs. There is the probability $\frac{v_0 dt}{L}$ that he will exit in the interval of time between $\underline{t} - dt$ and $\underline{t}$, incurring a trip cost of $\beta(t^* - \underline{t})$, and a probability $1 - \frac{v_0 dt}{L}$ that he will not exit in the interval, incurring a trip cost of $\alpha dt + \underline{c}$. Thus, $c(\underline{t} - dt) = \left(\frac{v_0 dt}{L}\right)\beta(t^* - \underline{t}) + \left(1 - \frac{v_0 dt}{L}\right)(\alpha dt + \frac{\alpha L}{v(t)} + \beta(t^* - \underline{t})) = \underline{c} + dt\left[\left(\frac{v_0}{L}\right)\beta(t^* - \underline{t}) + \alpha - \frac{\alpha v_0}{v(t)} - \frac{v_0}{L}\beta(t^* - \underline{t})\right] < \underline{c}$. A mass of departures at $\underline{t}$ is therefore inconsistent with the trip-timing equilibrium condition. Hence, $v_0 = \frac{c_0}{c_1 + \beta\underline{t}}$, so that $\underline{t} = \frac{\frac{\alpha L}{v_0} - \underline{c}}{\beta}$.

and (??) continues to apply, with $\bar{t}$ being determined by the condition that $d(\bar{t}) = 0$[16]. The constant $\underline{c}$ is obtained from the condition that the integral of the entry rate over the departure interval equals the population.

Over the early morning rush hour, density rises and velocity falls, and over the late morning rush hour, density falls and velocity rises. Since the density rises throughout the early morning rush hour, inflow must continue over this period and exceed the outflow. When traffic is only congested, outflow increases over this period, but when traffic is hypercongested, it falls. In the late morning rush hour, if inflow continues it must be less than outflow.

3.3 No-toll Equilibrium – Numerical Example

It will prove instructive to work out a numerical example. Greenshields' Relation is employed, which specifies a negative linear relationship between velocity and density[17], with a free-flow velocity of 20 mph, and jam density normalized so that capacity flow equals 1.0, implying that $k_j = 0.2$. Time is measured such that $t^* = 0$. $\alpha = 20$ (\$/hr), $\beta = 10$, and $\gamma = 40$, which is broadly consistent with their empirically estimated values for large US metropolitan areas, and mean trip length is 5.0 miles. To make the example striking, it is assumed that velocity at the peak of the rush hour is 2.5 mph. This is unrealistically high for US metropolitan areas, except perhaps Manhattan, but not for cities such as Delhi, Bangkok, Moscow, and Shanghai. These parameter values imply that $c_0 = 100$, $c_1 = \underline{c} = 40$, $c_2 = 40$, and $c_3 = \underline{c} = 40$, $\underline{t} = -3.5$, $\bar{t} = 0.625$, and $N = 0.6922$. The no-toll equilibrium is displayed in Figure ??, which plots cumulative departures from home, $D(t)$, and arrivals at work, $A(t)$, as function of time. Recall that the horizontal distance between the two curves gives travel time, and the vertical distance gives density, which is negatively related to velocity. The departure interval is 4.125 hours, the expected cost of a trip is \$40, and at the peak

[16]Thus, $\bar{t}$ is given implicitly by $k'(v(\bar{t}))\dot{v}(\bar{t}) + \frac{k(v(\bar{t}))v(\bar{t})}{L} = 0$. With Greenshields' Relation, $\bar{t} - t^* = \frac{\underline{c} - L\frac{\alpha+\gamma}{v_0}}{\gamma}$. Thus, late departures do not occur when $\underline{c} < \frac{L(\alpha+\gamma)}{v_0}$, i.e. when trip cost is less than $\frac{\alpha+\gamma}{\alpha}$ times uncongested trip cost, $\frac{\alpha L}{v_0}$.

[17]Relative to this negative linear specification, empirically estimated velocity-density relations tend to bulge out at low densities and bulge in at high densities.

of the rush hour traffic is so jammed that flow is only 44% of capacity flow. How inflow, outflow, density, and velocity evolve over the rush hour shall be discussed in section 5, which compares the no-toll equilibrium and the social optimum.

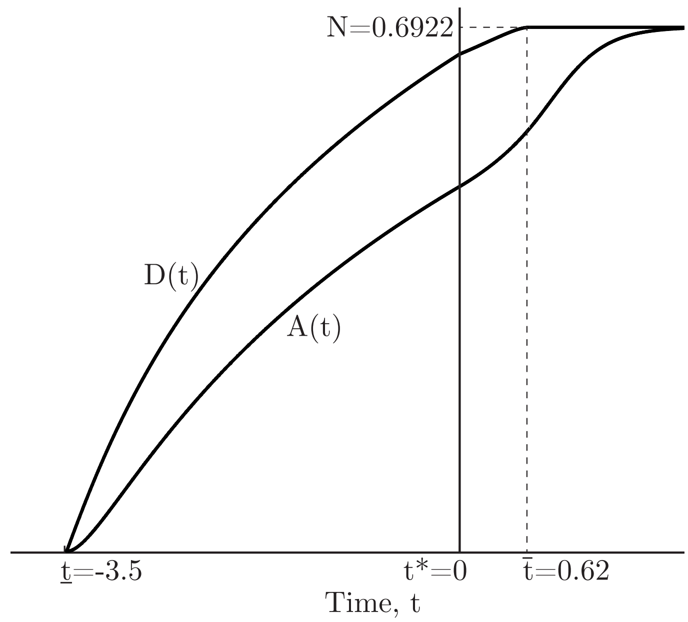

Figure 5: Cumulative departures and arrivals in the bottleneck model: No-toll equilibrium

The social optimum will be solved for in the next section. But for the moment the inefficiency of the no-toll equilibrium can be illustrated by comparing it to the situation where capacity flow is maintained throughout the departure interval — call it the benchmark allocation. This requires a departure mass of 0.1 at $\underline{t}$, followed by the inflow rate needed to maintain capacity flow for 2.961 hours. Thereafter, traffic density declines, with a corresponding increase in velocity. An upper bound on the expected cost of a trip can be established by assuming that velocity is 10 mph throughout the rush hour, and that the departure interval ends at t^*. Under these assumptions, expected travel time is 0.5 hours, expected travel time cost \$10, expected schedule delay cost[18] \$15.55 and expected trip cost

[18]The arrival rate would be 0.2 up to t^*. For the 0.5922 early arrivers, therefore, mean schedule delay cost would be \$14.80. For the 0.1 late arrivers, mean schedule delay would be 0.5 for a mean schedule delay cost of \$20.00. Overall mean schedule delay cost would therefore be $\frac{(0.5922)(14.80)+(0.1)(20.00)}{0.6922} = \$15.55$

\$25.55. Thus, expected trip cost and expected private congestion cost are consistently higher in the no-toll equilibrium that in the benchmark allocation.

The example illustrates the inefficiency of traffic flow over the rush hour in the absence of tolling. In the no-toll equilibrium, in order to satisfy the trip-timing condition, traffic congestion must increase steadily over the early morning rush hour, so that at the peak of the rush hour velocity is only 2.5 mph. But such a low velocity is associated with jammed conditions and an outflow rate only 44% of capacity flow. The combination of a low arrival rate at the start of the rush hour due to density considerably lower than capacity density and a low outflow rate around the peak of the morning hour due to density considerably higher than capacity density, lengthens the departure interval compared to the benchmark allocation. In terms of the bathtub analogy, in the benchmark allocation the planner adjusts the inflow rate so that the water flows out of the bathtub at capacity flow throughout the early morning rush hour, and then turns off the tap, allowing the bathtub to drain in the late morning rush hour. In the no-toll equilibrium, the inflow rate is such that the water level gradually rises until it reaches a height at which the outflow is less than one-half capacity flow, so that the bathtub takes a long time to drain.

3.4 The Reduced-Form Cost Curves

The economic importance of the elasticity of private congestion cost (PCC) with respect to population was explained earlier, as was the substantial difference in this elasticity between the textbook model and the bottleneck model. The value of this elasticity for this model is now explored.

A condition for equilibrium is that the integral of the departure rate over the departure interval equal population:

$$\int_{\underline{t}}^{\bar{t}} d(t)\,\mathrm{dt} = N. \tag{15}$$

From (**??**), $d(t) = k'(v(t))\dot{v}(t) + \frac{k(v(t))v(t)}{L}$. Expressions for $v(t)$ for the early and late morning rush hours are given by (**??**) and (**??**). Setting $t^* = 0$ and combining these results gives

$$d(t) = -k'\left(\frac{\alpha L}{\underline{c} + \beta t}\right)\frac{\alpha L \beta}{(\underline{c} + \beta t)^2} + k\left(\frac{\alpha L}{\underline{c} + \beta t}\right)\frac{\alpha}{\underline{c} + \beta t}$$

for the early departures. For late departures there are two cases to consider, the first without late arrivals, the second with late arrivals. For the rest of the subsection, Greenshields' Relation is assumed, for which

$$\begin{aligned} d(t) &= \frac{k_j \alpha}{\underline{c} + \beta t}\left(1 - \frac{L(\alpha - \beta)}{v_0(\underline{c} + \beta t)}\right) && \text{early}^{19} \\ &= \frac{k_j \alpha}{\underline{c} - \gamma t}\max\left\{0, \left(1 - \frac{L(\alpha + \gamma)}{v_0(\underline{c} - \gamma t)}\right)\right\} && \text{late} \end{aligned} \tag{16}$$

When population is zero and when therefore there is no congestion, expected user cost is $\frac{\alpha L}{v_0}$. Define $\theta \equiv \underline{c} \div \frac{\alpha L}{v_0}$, which has the interpretation as the ratio of user cost to uncongested user cost. In the first case $\theta < \frac{\alpha+\gamma}{\alpha}$, congestion is moderate and there are no late arrivals; in the second, $\theta > \frac{\alpha+\gamma}{\alpha}$, congestion is heavy and there are late arrivals.

- Case 1: Moderate Congestion $\left(\theta < \frac{\alpha+\gamma}{\alpha}\right)$

  Integrating and then substituting in $\underline{c} + \beta \underline{t} = \frac{\alpha L}{v_0}$ yields

[19] This equation may be rewritten as $d(t) = \frac{k_j v(t)}{L}\left(1 - (\frac{\alpha-\beta}{\alpha})\frac{v(t)}{v_0}\right)$ and $\dot{d}(t) - (\frac{k_j \dot{v}}{L})\left(1 - 2(\frac{\alpha-\beta}{\alpha})(\frac{v(t)}{v_0})\right)$. Since $\dot{v} < 0$, the inflow rate is increasing if $v(t) > \frac{\alpha v_0}{2(\alpha-\beta)}$, and decreasing otherwise. If β is only slightly greater than zero, the inflow rate is increasing if traffic is congested and decreasing if it is hypercongested. If β is greater than $\frac{\alpha}{2}$, the inflow rate is decreasing throughout the rush hour.

$$N = k_j \left[\left(\frac{\alpha}{\beta} \right) ln\theta - \left(\frac{\alpha - \beta}{\beta} \right) \left(1 - \frac{1}{\theta} \right) \right] \tag{17}$$

$$\frac{dN}{d\theta} = k_j \left(\frac{\alpha/\beta}{\theta} - \frac{(\alpha - \beta)/\beta}{\theta^2} \right). \tag{18}$$

- Case 2: Heavy Congestion $\left(\theta > \frac{\alpha+\gamma}{\alpha}\right)$

  Integrating and then substituting in $\underline{c} + \beta \underline{t} = \frac{\alpha L}{v_0}$ and $\underline{c} - \gamma \bar{t} = \frac{(\alpha+\gamma)L}{v_0}$, yields

$$N = k_j \left[\left(\frac{\alpha}{\beta} + \frac{\alpha}{\gamma} \right) \left(ln\theta - 1 + \frac{1}{\theta} \right) + 1 + \left(\frac{\alpha}{\gamma} \right) ln \left(\frac{\gamma}{\alpha + \alpha} \right) \right] \tag{19}$$

$$\frac{dN}{d\theta} = k_j \left(\frac{\alpha}{\beta} + \frac{\alpha}{\gamma} \right) \left(\frac{1}{\theta} - \frac{1}{\theta^2} \right). \tag{20}$$

Total cost is $\frac{\alpha L N \theta(N)}{v_0}$, user cost is $\frac{\alpha L \theta(N)}{v_0}$, and marginal cost is $\alpha L \left(\frac{\theta(N) + N\theta'(N)}{v_0} \right)$. Now, define $\epsilon = \theta - 1$, which is the ratio of private congestion cost to uncongested trip cost. For light congestion and heavy congestion, respectively:

$$E_{\epsilon:N} = \frac{\beta(\epsilon+1)^2}{(\alpha\epsilon + \beta)\epsilon} \left(\frac{\alpha}{\beta} \right) ln(\epsilon + 1) - \left(\frac{\alpha - \beta}{\beta} \right) \left(\frac{\epsilon}{\epsilon + 1} \right)$$

$$E_{\epsilon:N} = \left(\frac{\epsilon + 1}{\epsilon} \right)^2 \left[\left(ln(1+\epsilon) - \frac{\epsilon}{\epsilon + 1} \right) + \frac{\beta}{\beta + \gamma} ln \left(\frac{\alpha}{\alpha + \gamma} \right) + \frac{\beta\gamma}{\alpha(\beta + \gamma)} \right]. \tag{21}$$

With the BPR form of congestion function, in the textbook model the elasticity of private congestion cost with respect to flow is constant, with a value of 4.0 typically being assumed. When flow in that model is interpreted as the number of trips over the day or rush hour, it corresponds to population (the number of travelers). In the bottleneck model, the elasticity of private congestion with respect to population equals 1.0. In the bathtub model, the elasticity of private congestion cost with respect to population equals $E_{\epsilon:N}$. Its value approaches zero as N approaches zero, increases monotonically with N, and asymptotically increases with the order[20] N as N approaches infinity. Thus, when flow in the textbook model is interpreted as population, the elasticity of this paper's model spans those of the textbook and bottleneck models, with the bottleneck model's elasticity corresponding to moderate congestion and the textbook model's elasticity corresponding to heavy congestion. In this sense at least, the textbook model and the bottleneck model can be regarded as special cases of the bathtub model. Figure **??** displays the reduced-form private congestion cost and marginal variable social cost (marginal social cost minus the fixed cost of a trip, $\frac{\alpha L}{v_0}$) functions for the example.

4 The Social Optimum

The planner's problem is to choose the departure rate over the rush hour so as to minimize total trip costs, which equal total travel time costs plus total schedule delay costs. Total travel time costs alone would be minimized by having an infinitely long departure interval, so that there would never be any congestion. Total schedule delay costs alone would be minimized by having a mass of departures at the beginning of the rush hour of sufficient size to bring density up to capacity density, followed by a period with entry at capacity flow, during which outflow would be at capacity, and concluded by a period with exit but no entry in which density dissipates[21]. The solution to the planner's problem is an "average" of these two solutions, with relatively more weight being put on reducing total schedule costs, the higher is the ratio of $\frac{\beta}{\alpha}$.

[20] In the numerical example of the previous section, $N = 0.6922$, $\epsilon = 7$, and travel is heavily congested (case 2), so that (**??**) applies. $\frac{dN}{d\epsilon} = 0.05468$ and $E_{\epsilon:N} = 1.8084$.

[21] This is the same departure pattern as in the benchmark example in the previous section.

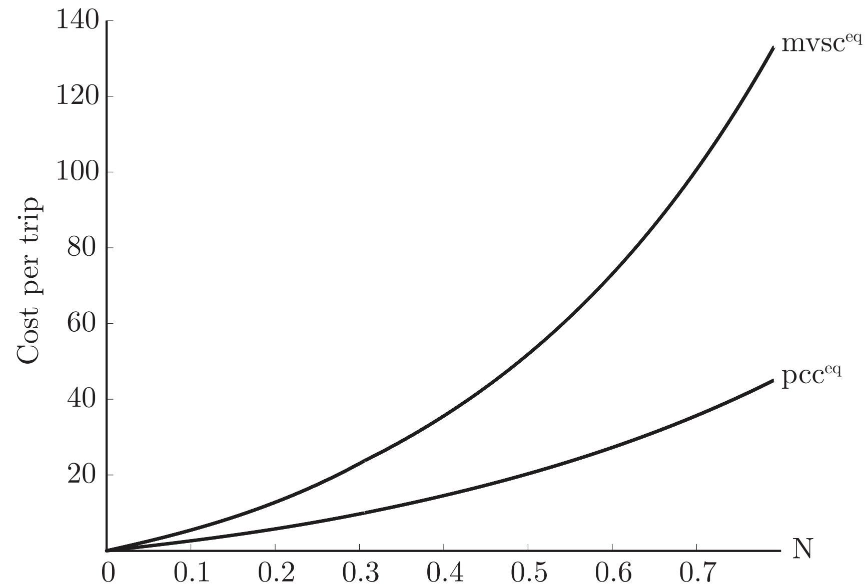

Figure 6: Private congestion cost and marginal variable social cost as functions of population: No-toll equilibrium

4.1 Optimal Control Problem

The social optimum problem is solved for using optimal control theory. There is one control variable, $d(t)$, and two state variables, $A(t)$, and $k(t)$. The objective is to minimize the sum of total travel time costs, total time early costs, and total time late costs (all per unit area). $t = 0$ is taken to be the time of the first departure[22], so that t^* and $\bar{t}$ are taken as choice variables. Total travel time costs can be written as $\int_0^\infty \alpha k(t)\, dt$, with the travel time on the road between t and $t+dt$ being simply the density of cars on the road at that time; total time early costs as $\int_0^{t^*} \beta A(t)\, dt$, with the time early between t and $t + dt$ equaling the number of individuals who have already arrived at work; and total time late costs as $\int_{t^*}^\infty \gamma(N - A(t))\, dt$, with the time late between t and $t + dt$ equaling the number of individuals who have not arrived at work. The two differential equation constraints are $\dot{k}(t) = d(t) - \frac{k(t)v(k(t))}{L}$ (eq. (**??**)) and $\dot{A}(t) = \frac{k(t)v(t)}{L}$. There are also an isoperimetric constraint, that the integral of

[22]Note that the time origin is different from that of the previous section, where t^* is the origin.

the departure rate over the departure interval equal the population, initial and terminal conditions on the state variables, and a non-negativity constraint on $d(t)$.

The optimal control problem is

$$min \left[\int_0^{t^*} \alpha k(t)\, dt + \int_0^{t^*} \beta A(t)\, dt \right] + \left[\int_{t^*}^{\infty} \alpha k(t)\, dt + \int_{t^*}^{\infty} \gamma (N - A(t))\, dt \right]$$

such that

$$\begin{aligned}
&\text{i) } \dot{k}(t) = d(t) - \frac{k(t)v(k(t))}{L} && \lambda \\
&\text{ii) } \dot{A}(t) = \frac{k(t)v(t)}{L} && \mu \\
&\text{iii) } N \leq \int_0^{\bar{t}} d(t)\, dt && \rho \\
&\text{iv) } k(0) = A(0) = 0 \\
&\text{v) } k(\infty) = 0 \quad A(\infty) = N \\
&\text{vi) } d(t) \geq 0
\end{aligned} \tag{22}$$

Ignoring constraint vi) since it turns out not to bind, for the early morning rush hour the Hamiltonian is

$$H = \alpha k + \beta A + \lambda \left(\frac{d - kv(k)}{L} \right) + \mu \left(\frac{kv(k)}{L} \right) - \rho d \tag{23}$$

The optimality conditions are:

$$d : \lambda - \rho = 0 \tag{24}$$

$$k : \dot{\lambda} = - \left[\alpha - \frac{\lambda (v + kv')}{L} + \frac{\mu (v + kv')}{L} \right] \tag{25}$$

$$A : \dot{\mu} = -\beta \tag{26}$$

Since ρ is the shadow price associated with an isoperimetric constraint (equaling the marginal social cost of a trip), it is independent of t. So therefore from (**??**) is $\lambda(t)$, which is the shadow price of density at time t. Solving (**??**) gives that $\mu(t)$, the shadow price of an arrival at time t, equals $\mu(0) - \beta t$. Since the cost of time early is $\beta(t^* - t)$, $\mu(0) = \beta t^*$. Combining these results yields

$$\alpha L - [\lambda - \beta(t^* - t)]\,(v + kv') = 0. \tag{27}$$

This equation can be explained through a perturbation argument. Increase traffic density at time t by one unit and then decrease it back to its original level a period dt later. This increases traffic density by one unit for a period of time dt. The travel time cost associated with this is αdt. The arrivals rate at time t is $\frac{k(t)v(k(t))}{L}$, which increases by $\frac{v(k)+kv'(k)}{L}$ for the increment of time dt, and thereafter returns to its original level. Since the cost of an extra arrival at time t is $\beta(t^* - t)$, time early costs increase by $\beta\frac{(t^*-t)(v(k)+kv'(k))}{L}$. To restore density to its original level, requires subtracting $\frac{(v(k)+kv'(k))}{L}$ individuals at time $t + dt$ at a saving of $\frac{\lambda(v(k)+kv'(k))}{L}$. Since the initial allocation was optimal, this perturbation has no effect on total trip costs. Thus, $[\alpha L + \beta(t^* - t)(v + kv') - \lambda(v + kv')]dt = 0$.

Since there is no congestion at $t = 0$, $v(0) = (v + kv')_0 = v_0$, so that from (**??**)

$$\lambda = \frac{\alpha L}{v_0} + \beta t^*. \tag{28}$$

Inserting (**??**) into (**??**) yields

$$\alpha L - \left(\frac{\alpha L}{v_0} + \beta t\right)(v + kv') = 0. \tag{29}$$

Since $v + kv' > 0$, traffic is never hypercongested during the early morning rush hour. Also, density increases continuously over the early morning rush hour. (**??**) may be solved for $k^e(t)$, the optimal time path of density over the early morning rush hour. From this, $v^e(t) = v(k^e(t)), a^e(t) = \frac{k^e(t)v^e(t)}{L}, d^e(t) = \dot{k}^e(t) + a^e(t)$ may be calculated. The first-order condition with respect to t^* is

$$\beta A(t^*) - \gamma[N - A(t^*)] = 0, \tag{30}$$

so that

$$t^* = A^{-1}\left(\frac{\gamma N}{\beta + \gamma}\right). \tag{31}$$

If departures continue into the late morning rush hour, the condition analogous to (**??**) is

$$\alpha L - \left(\frac{\alpha L}{v_0} + \beta t^* - \gamma(t - t^*)\right)(v + kv') = 0, \tag{32}$$

and $\bar{t}$ is determined by the condition $d(\bar{t}) = 0$. Suppose, for the sake of argument, that departures do continue into the late morning rush hour. Then (**??**) can be solved for $k^l(t)$, the optimal time path of density over the portion of the late morning rush hour during which there are departures, and so too can $d^l(t) = \dot{k}^l(t) + \frac{k^l(t)v^l(t)}{L}$. If the t that solves $d^l(t) = 0$ exceeds t^*, then departures do continue into the late morning rush hour, and otherwise not. After the last departure, whether this occurs at t^* or later, traffic density falls according to the equation $\dot{k}(t) = -\frac{k(t)v(t)}{L}$. Traffic is never hypercongested in the late morning rush hour. If there are late departures, (**??**) implies that traffic is not hypercongested in the late morning departure interval, and after departures cease traffic density decreases and therefore cannot become hypercongested. If there are no late departures, traffic density falls continuously after t^*, and since traffic is not hypercongested at t^*, it is not hypercongested after t^*.

Finally, it bears note that, *measured from the start of the rush hour*, the time path of the inflow rate during the early morning rush hour is independent of population, with population determining the length of the morning rush hour. The reason is that density must increase from zero at such a rate that the marginal social cost of a trip remain constant over time. An analogous result holds for the no-toll equilibrium, in which density must increase from zero at such a rate that trip cost remain constant over time.

4.2 Social Optimum – Numerical Example

The social optimum for the example considered in the previous section is now solved for. Assuming Greenshields' Relation, (??) and (??) become

$$k(t) = \frac{k_j}{2} - \left(\frac{\alpha L k_j}{v_0}\right) \div 2\left(\frac{\alpha L}{v_0} + \beta t\right) \qquad \text{for } t < t^*$$

$$= \frac{k_j}{2} - \left(\frac{\alpha L k_j}{v_0}\right) \div 2\left(\frac{\alpha L}{v_0} + \beta t^* - \gamma(t - t^*)\right) \qquad \text{for } t \in (t^*, \bar{t}) \tag{33}$$

Also,

$$\dot{k}(t) = -\frac{k(t)v(k(t))}{L} \qquad \text{for } t > \bar{t}$$

For early arrivals

$$\dot{A}(t) = \frac{k(t)v(k(t))}{L} = \frac{k_j v_0 - \left(\frac{\alpha^2 L^2 k_j}{v_0}\right) \div \left(\frac{\alpha L}{v_0} + \beta t\right)^2}{4L}.$$

Since A(0) = 0,

$$A(t) = \frac{k_j v_0 t + \left(\frac{\alpha^2 L^2 k_j}{\beta v_0}\right) \div \left(\frac{\alpha L}{v_0} + \beta t\right)}{4L} - \frac{\alpha k_j}{4\beta}. \tag{34}$$

Also, from (??), t^* is determined by the equation

$$A(t^*) = \left(\frac{\gamma}{\beta + \gamma}\right) N = \frac{k_j v_0 t^* + \left(\frac{\alpha^2 L^2 k_j}{\beta v_0}\right) \div \left(\frac{\alpha L}{v_0} + \beta t^*\right)}{4L} - \frac{\alpha k_j}{4\beta}. \tag{35}$$

The departure rate is given by

$$d(t) = \dot{k}(t) + \dot{A}(t) = \frac{\frac{k_j v_0}{L} - \frac{(\alpha)(\alpha-2\beta)Lk_j}{v_0} \div \left(\frac{\alpha L}{v_0} + \beta t\right)^2}{4}. \tag{36}$$

With the parameters of the numerical example, $k(t) = \frac{0.2t}{1+2t}$, $\dot{k}(t) = \frac{0.2}{(1+2t)^2}$, $v(t) = \frac{20(1+t)}{1+2t}$, $a(t) = \frac{0.8(1+t)t}{(1+2t)^2}$, $d(t) = 0.2$, $A(t) = \frac{0.4t^2}{1+2t}$, $t^* = 3.201$, and $\lambda = 37.01$. For $t \in (t^*, \bar{t})$, $k(t) = 0.1 - \frac{0.5}{\lambda+40(t^*-t)}$, from which $\dot{k}(t)$, $a(t) = \frac{k(t)v(t)}{L}$, and $d(t) = \dot{k}(t) + a(t)$ can be calculated. $\bar{t}$ is calculated from the condition that $d(\bar{t}) = 0$, and in the numerical example equals 3.847. Thus, the departure interval is slightly shorter in the social optimum than in the no-toll equilibrium. For $t > \bar{t}$, there is no inflow, and so density falls at an increasing proportional rate as cars exit. $k(t)$ is calculated from $k(\bar{t})$ and $\dot{k}(t) = -\frac{k(t)v(k(t))}{L}$.

Cumulative departures and arrivals are shown for the example in Figure **??**. Differences between the no-toll equilibrium and the optimum will be commented on in the next section.

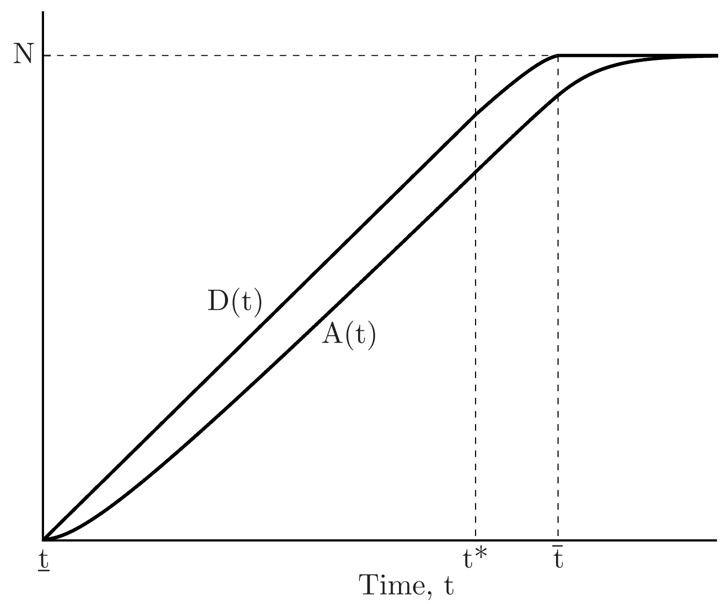

Figure 7: Cumulative departures and arrivals in the social optimum

The marginal social cost function may be calculated using the relationships from (**??**) and (**??**):

$$\lambda(N) = \frac{\alpha L}{v_0} + \beta t^*(N) = \frac{\alpha L}{v_0} + \beta A^{e-1}\left(\frac{\gamma N}{\beta+\gamma}\right). \tag{37}$$

Rewrite (??) as

$$\beta k_j v_0 t^* \left(\frac{\alpha L}{v_0} + \beta t^*\right) - \left(\frac{\alpha^2 L^2 k_j}{v_0}\right) + \left(\alpha L k_j - 4\left[\frac{\beta\gamma}{\beta+\gamma}\right] NL\right)\left(\frac{\alpha L}{v_0} + \beta t^*\right) = 0$$

or

$$t^{*2}(\beta^2 k_j v_0) + t^*\left[2\alpha L k_j - 4\left(\frac{\beta\gamma}{\beta+\gamma}\right) NL\right] + \left[-4\alpha\left(\frac{\gamma}{\beta+\gamma}\right)\frac{NL^2}{v_0}\right] = 0. \tag{38}$$

This is a quadratic equation in t^*, which has the form $at^{*2} + [b_0 - b_1 N]t^* - eN = 0$, for which the solution is $t^* = \frac{-(b_0 - b_1 N) + \left((b_0 - b_1 N)^2 + 4aeN\right)^{1/2}}{2a}$. t^*, and hence marginal variable social cost and private congestion cost, rise with N in the order between $\frac{1}{2}$ and 1. Total private congestion costs are

$$\int_0^N \beta t^*(n)\, dn = \int_0^N \beta\left[\frac{-(b_0 - b_1 n) + ((b_0 - b_1 n)^2 + 4aen)^{1/2}}{2a}\right] dn.$$

For the numerical example, numerical integration gives that private congestion cost is xxxx and hence average trip costs is xxxx. COMMENT.

Figure **??** displays marginal variable social cost and private congestion cost as functions of N. The difference between them is the average congestion externality cost or equivalently the average value of the optimal time-varying toll.

4.3 The Optimal Time-Varying Toll

The optimal time-varying toll, $\tau(t)$, which is applied at the beginning of the trip, may be calculated as the difference between the marginal social cost, λ, and the user cost, evaluated at the social optimum. In the decentralized optimum, expected trip price must be the same

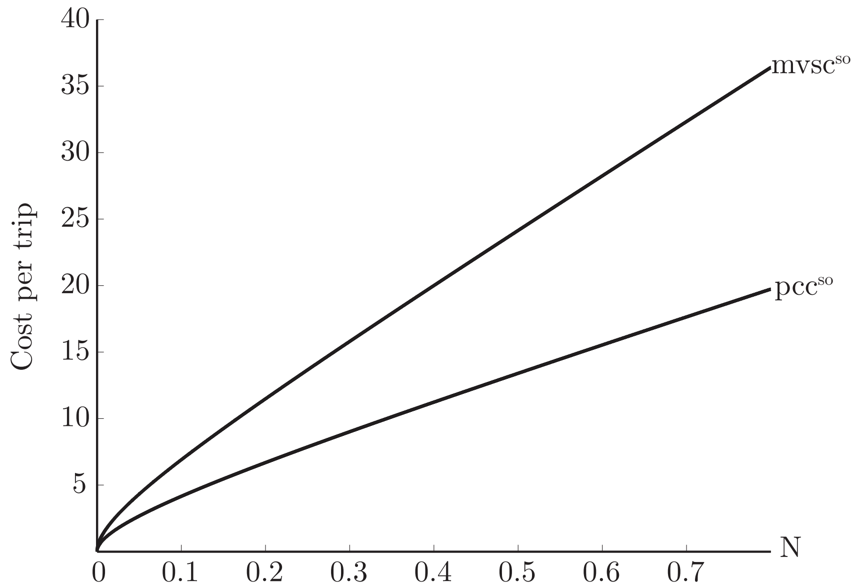

Figure 8: Private congestion cost and marginal variable social cost as functions of population: Social optimum

for all t in the departure interval. Expected trip price for a departure at time $t < t^*$ is

$$
\begin{aligned}
Ep(t) = \lambda = \int_t^{\infty} \alpha(u-t)\left(\frac{v(u)}{L}\right) P(u;t)\,du + \int_t^{t^*} \beta(t^*-u)\left(\frac{v(u)}{L}\right) P(u;t)\,du \\
+ \int_{t^*}^{\infty} \gamma(u-t^*)\left(\frac{v(u)}{L}\right) P(u;t)\,du + \tau(t).
\end{aligned}
$$

Integrating the first term on the right-hand side by parts yields

$$
\begin{aligned}
Ep(t) = \lambda = \int_t^{\infty} \frac{\alpha L}{v(u)}\left(\frac{v(u)}{L}\right) P(u;t)\,du + \int_t^{t^*} \beta(t^*-u)\left(\frac{v(u)}{L}\right) P(u;t)\,dt \\
+ \int_{t^*}^{\infty} \gamma(u-t^*)\left(\frac{v(u)}{L}\right) P(u;t)\,du + \tau(t).
\end{aligned}
\tag{39}
$$

Using (**??**), (**??**), and (**??**), one may write

$$\lambda = \int_t^{\infty} \left(\frac{\alpha L}{v(u) + k(u)v'(k(u))} \right) \frac{v(u)}{L} P(u;t)\, du + \int_t^{t^*} \beta(t^* - u) \frac{v(u)}{L} P(u;t)\, du + \int_{t*}^{\infty} \gamma(u - t^*) \frac{v(u)}{L} P(u;t)\, du. \tag{40}$$

Subtracting (**??**) from (**??**) yields

$$\tau(t) = \int_t^{\infty} \alpha \left(- \frac{k(u)v'(k(u))}{v(k(u)) + k(u)v'(k(u))} \right) P(u;t)\, du. \tag{41}$$

$-\frac{k(u)v'(k(u))}{v(u)+kv'(u)}$ is the ratio of the congestion externality cost to the user cost of travel time at time u. Since a commuter's user cost of travel between u and $u + du$ is αdu, the congestion externality cost he imposes during this interval is $[-\frac{kv'}{v+kv'}]_u du$. Multiplying this expression by the probability that the commuter is on the road at this time, and integrating over u, gives the expected congestion cost of a trip starting at time t. The externality operates entirely via traffic congestion and not at all via schedule delay[23].

5 Comparison of the No-toll Equilibrium and Social Optimum

5.1 Their Comparison in the Example

This section starts with a detailed discussion of the numerical example. In the numerical example, free-flow velocity equals 20 mph and velocity at capacity flow 10 mph, so that travel is hypercongested for speeds below 10 mph. The units of density are chosen so that capacity flow is 1.0; jam density is 0.2 per ml^2-hr, and capacity density 0.1 per ml^2. The units of flow are cars per ml-hr or car-miles per ml^2 per hr. Since mean trip distance is 5

[23]This result is due to the Envelope Theorem. Since the allocation is socially optimal, the adjustment in the departure time distribution induced by the addition of an individual at a particular point in time does not have a first-order effect on aggregate trip costs.

mls, the steady-state capacity inflow rate (that inflow rate that would sustain capacity flow on the street system) is 0.2. Time is measured in clock hours with the common work start time being 9:00 am.

Figure **??** contains three panels. The top panel displays velocity as a function of time, the middle panel density, and the bottom panel the departure rate. In each panel, the curve for the no-toll equilibrium is represented by a solid line and that for the social optimum by a starred line. The discussion will focus on the early morning rush hour. Consider first the top panel. In the no-toll equilibrium, $v(t) = \frac{10}{4-(9-t)}$. The early morning rush hour starts at 5:30 am, at which time traffic flows at free-flow speed. Traffic becomes increasingly congested over the early morning rush hour, with hypercongestion setting in at 6:00 am, and with velocity at the peak being 2.5 mph. In the social optimum, $v(t) = \frac{20[1+(t-5.799)]}{\left(1+\frac{2}{t-5.799}\right)}$. The early morning rush hour starts at 5:48 am. Traffic becomes increasingly congested over the early morning rush, but never becomes hypercongested, with velocity at the peak being 11.35 mph.

The middle panel displays density as a function of time. Since density is a negative linear function of velocity (Greenshields' Relation), this panel displays the same information as that in the top panel but from a perspective that will facilitate understanding why traffic becomes so much more congested in the no-toll equilibrium than in the social optimum even though the lengths of their rush hours are similar.

The bottom panel displays the departure rates as a function of time. In the no-toll equilibrium early morning rush hour, the departure rate is given by $d(t) = \left(\frac{k_j v_0}{L}\right)\left(\frac{v(t)}{v_0}\right)\left(1 - \frac{v(t)}{2v_0}\right) = 0.8\left(\frac{0.5}{(4-(9-t))}\right)\left(1 - \frac{0.25}{4-(9-t)}\right)$. The departure rate falls over the early morning rush hour, from 0.4 (twice the steady-state capacity departure rate) at the start of the rush hour, 5:30 am, to 0.3 at 6:00 am, to 0.175 at 7:00 am, to 0.09375 at 9:00 am. In line with the paper's central simplifying assumption, the arrival rate is proportional to the flow rate on the streets: $\frac{v(t)k(t)}{L} = 0.8\left(\frac{0.5}{4-(9-t)}\right)\left(1 - \frac{0.5}{4-(9-t)}\right)$. The arrival rate starts at zero at 5:30 am, rises to a maximum of 0.2 at 6:00 am, at which time hypercongestion sets in, and then falls to 0.15 at 7:00 am, and then 0.0875 at 9:00 am. To satisfy the trip-timing condition, density must rise

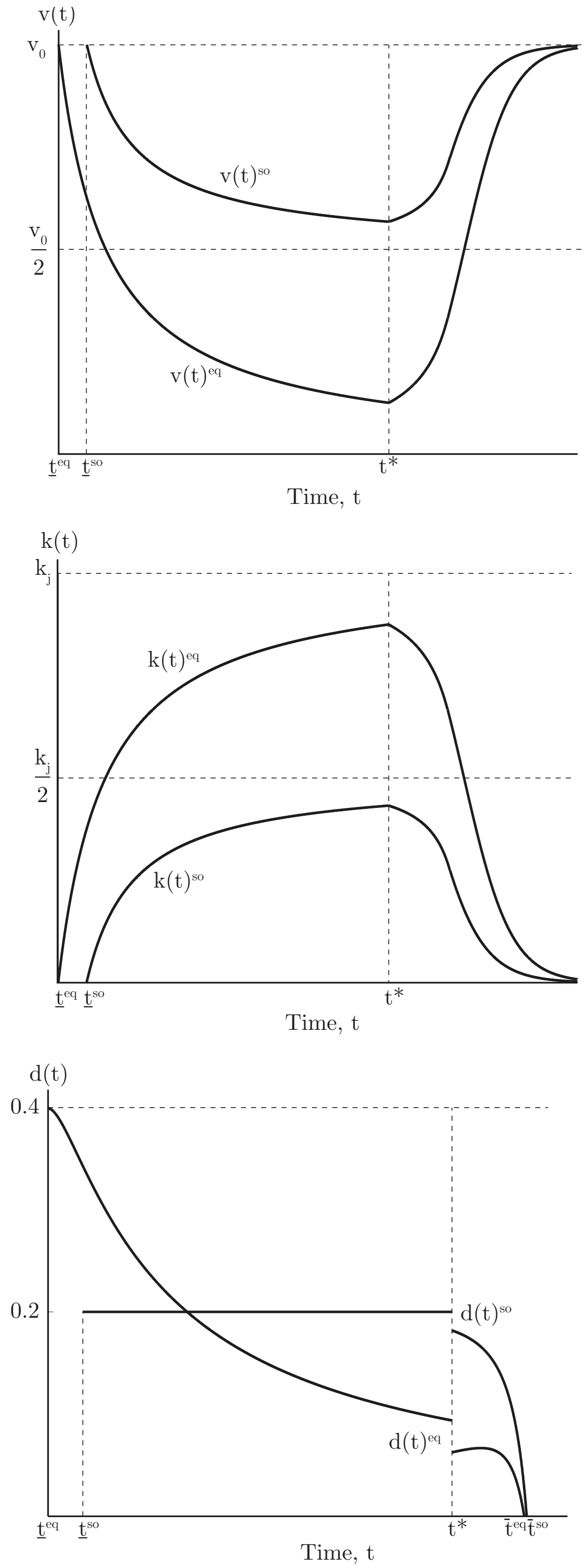

Figure 9: Velocity, density, and inflow rate over the morning rush hour: No-toll equilibrium and social optimum.

throughout the early morning rush hour, which requires that the departure rate exceed the arrival rate, even after hypercongestion has set in. In the social optimum in contrast, the departure rate equals the steady-state capacity departure rate throughout the early morning rush hour. As a result, density rises steadily but at a decreasing rate, asymptotically approaching capacity density. With this departure rate, user cost falls over the early morning rush hour, so that over this period decentralization of the optimum requires a toll that it is increasing in time.

In the no-toll equilibrium, the common trip price, which equals the common user cost since there is no toll, is \$40.00. In the decentralized social welfare optimum, the common trip price, which equals the marginal social cost of a trip, is \$37.01. Thus, in the example, the imposition of the optimal time-varying congestion toll would make commuters better off even if the toll revenue collected were completely squandered! Put alternatively, the revenue from the congestion toll is raised with negative burden.

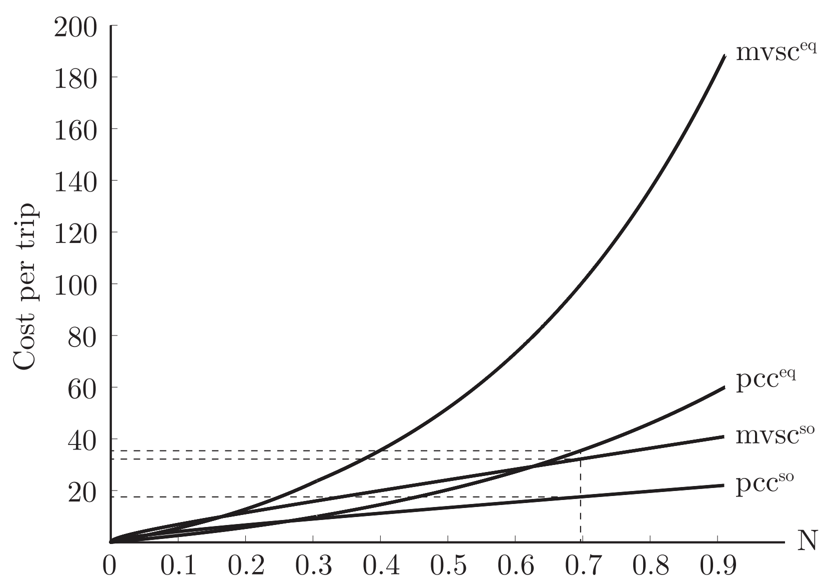

Figure 10: Private congestion cost and marginal variable social cost as functions of population: No-toll equilibrium and social optimum

Figure ?? combines two previous figures, Figures ?? and ??, displaying the marginal variable social cost and private congestion cost as functions of population (the fixed number of commuters over the morning rush hour, road capacity fixed). Again the curves for the no-toll equilibrium are drawn as solid lines, those for the social optimum as starred lines. The vertical distance between the private congestion cost in the no-toll equilibrium and that in the social optimum equals the per capita deadweight loss due to the no-toll equilibrium's inefficient pattern of departures over the rush hour. Particularly noteworthy is that with very heavy congestion – when road capacity is very small relative to population, as in most major cities outside Western Europe and North America – private congestion cost increases close to exponentially with population in the no-toll equilibrium but only linearly with population in the social optimum. In the no-toll equilibrium, as population gets very large, mean travel time cost increases without limit, as traffic is almost completely jammed over almost the entire rush hour. In the social optimum, in contrast, as population gets very large, mean travel time cost levels off since traffic is close to capacity flow for almost the entire rush hour, so that, first, the bulk of marginal social cost is schedule delay cost, and, second, a doubling of population results in a doubling of the length of the morning rush hour and hence a doubling of average schedule delay cost.

5.2 Policy insights

Most applied studies of congestion tolling have examined the policy using the textbook model, where the benefit from congestion tolling derives from a reduction in the number of trips taken (REFS). These studies typically find that the toll revenue raised is several times the efficiency gain achieved. Unless therefore the toll revenue is spent wisely, tolling can be harmful in aggregate, and even when it is beneficial in aggregate, some groups are hurt and others helped. There has therefore been considerable policy and academic discussion about how toll revenues can be redistributed to benefit all major user groups, or at least the super-majority of users needed for the policy to be politically attractive (REFS). In the basic bottleneck model, in which trip demand is inelastic, the benefit from congestion tolling derives from the reallocation of the fixed number of trips over the rush hour. The toll revenue exactly equals the efficiency gain, and so is raised with no burden. The bathtub model of this paper enriches the bottleneck model to allow for classical flow congestion, in which velocity is negatively related to density. With moderate congestion, the efficiency gain from congestion tolling falls short of the revenue raised, but when congestion is very heavy may exceed it many times. The optimal time-varying toll prevents hypercongestion – traffic jam situations – which can add enormously to trip cost when road capacity is low relative to population.

To simplify the analysis, this paper has assumed that trip demand is completely inelastic. Extend the model to treat price-sensitive demand, and consider the effect in the extended model of applying a flat (time-invariant) congestion toll. The toll would be successful in reducing the number of commuters, but would not prevent crippling traffic jams at the peak of the rush hour. An optimal time-varying toll, however, would. Thus, this paper strengthens the argument for employing congestion tolls that vary over time.

In several papers, Daganzo and his co-authors have pointed to the critical importance, under heavily congested conditions, of not allowing traffic to become hypercongested. They have argued for traffic restraint policies to achieve this. Where time-varying congestion tolling is infeasible, as is the case in most developing countries where the technology for its implementation would be impractically expensive and where the government's administrative and enforcement capacities are limited, traffic restraint policies might indeed be a desirable second best. At a conceptual level, a traffic restraint policy would restrain the entry rate into areas that have the potential to become badly jammed, similar to the way ramp meters do on freeways. In this way, traffic jams would be converted into queues, and hypercongestion avoided. Whether practical traffic restraint policies along these lines can be designed and implemented remains to be seen.

5.3 Directions for Future Research

This paper has presented just about the simplest possible model of bathtub traffic congestion over the morning rush hour[24], and even for this model devoted considerable space to a particular example and derived few general results. In light of its analytical tractability, it should not be unduly difficult either to derive results with general functional forms for the trip cost and congestion functions or to generalize the model, while still achieving analytical tractability, in many of the ways that the bottleneck model has been generalized. For example, the model could be extended relatively straightforwardly to allow for commuters who differ in their values of time and work start times, and for alternative modes (when traffic becomes heavily congested, bicycling and walking[25] become attractive options). It would also be very straightforward to apply the model to solve for optimal capacity,[26] and to extend it to treat downtown parking[27]. The model also seems well suited to endogenize activity scheduling within an urban area, most notably the distribution of work start times. The research program initiated by Daganzo and his co-authors in estimating macroscopic fundamental diagrams for heavily congested areas of particular cities should prove highly complementary to development of the bathtub model, leading to practical policy simulation based on bathtub models within a short period of time.

The model was based on three central assumptions, one more general, and the other two more specific. The more general assumption is that traffic flow at the level of the urban

[24] An alternative approach would extend the bottleneck model to, first, allow the capacity of the bottleneck to depend on the length of the queue behind it and, second, to replace FIFO (first-in, first-out) queuing with random access queuing so as to avoid having to solve a delay differential equation.

[25] A Punch cartoon from the early 1960's (before the construction of motorways in the UK, when holiday traffic jams were measured in miles) shows two hitchhikers getting out of a car, with the caption along the lines: "We've really enjoyed visiting with you, but we must get along now." When taking taxis in central London prior to implementation of the congestion-pricing program, several times the author had to get out of his taxi and walk in order to make his appointment.

[26] With perfectly inelastic demand, (second-best) capacity in the no-toll equilibrium would be higher than (first-best) capacity in the social optimum. As the demand elasticity is increased, induced demand in the no-toll equilibrium would increasingly undermine the benefits of capacity expansion, until a demand elasticity is reached beyond which optimal capacity in the no-toll equilibrium would be lower than that in the social optimum.

[27] Arnott, Rave and Schöb (2005), Arnott and Inci (2006), and Arnott and Rowse (2009) examine parking policy in steady-state bathtub models.

neighborhood can be described, with reasonable accuracy, by classical flow congestion, which assumes a negative relationship between velocity and traffic density. This assumption is a working hypothesis, and is based on the argument that traffic congestion on city streets is so complicated that the most fruitful approach is to model it at the aggregate level, in terms of a macroscopic relationship between flow, velocity, and density. The early tests of the hypothesis are promising, but its validity is open to question. The classical flow congestion model was originally derived for flow congestion on highways, and it is a leap to apply it to congested urban areas where nodal congestion – congestion at intersections – is arguably more important than flow congestion. The first more specific assumption is that individuals do not know the length of their trips when making their trip-timing decisions. This is obviously unrealistic, but it is hard to see how the basic insights derived from the paper would not carry through if the assumption were relaxed. Furthermore, relaxing the assumption by itself results in a model that is severely analytically intractable (even numerical analysis of it is difficult), while the alternative assumptions that have been made in the literature to achieve tractability all violate the physics of fluid flow, which raises doubts about the soundness of the results obtained. The second more specific assumption is that trip lengths are negative exponentially distributed. This assumption too is not realistic, but again does it compromise the macroscopic relationships the model is designed to elucidate?

6 Conclusion

In developed countries, traffic congestion is getting worse slowly but steadily; in developing countries, it is getting worse quickly. Even though most of us experience traffic congestion on a daily basis, it remains poorly understood, which at least partially explains why congestion mitigation seems so intractable. Previous economic models of traffic assume that congestion increases travel time but does not reduce flow. This assumption is counter to experience and recent empirical evidence. This paper developed a model of rush-hour traffic congestion in which, under heavily congested conditions, increased congestion causes traffic flow to fall. The broad intuition is easy to understand through analogy with a special type of bathtub for which the rate at which the bath drains depends on the height of water in the tub. Up to a critical height, the outflow is positively related to the height of water, but as the height of water increases above this critical level the drain becomes increasingly clogged by the weight of water above it. The bathtub represents Manhattan or any other heavily built-up urban area; the height of water represents traffic density (and the critical height to capacity density); and water entering from the tap and exiting via the drain represent the flow of traffic entering and exiting the city streets.

This paper developed a simple model along these lines and applied it to traffic flow in the morning rush hour. A fixed number of identical commuters with a common work start time travel from home to work over city streets, experiencing two types of cost, travel delay due to congestion and the inconvenience of arriving at work early or late. In equilibrium no commuter can reduce her trip cost by altering her departure time from home. The paper solved for the no-toll equilibrium and the social optimum, and compared their properties. In the no-toll equilibrium, travel delay increases over the early morning rush hour at a rate that exactly offsets the benefits from arriving less early, and over the late morning rush hour decreases. With heavy congestion (the capacity of the street system is low relative to the number of commuters), equilibrium traffic density rises above capacity density causing a traffic jam (hypercongestion) that is slow to dissipate. In the social optimum, in contrast, the planner regulates the departure of commuters ensuring that a traffic jam does not occur. The social optimum can be decentralized via a time-varying toll equal to the congestion externality imposed by drivers, evaluated at the social optimum. The paper developed an extended numerical example of a heavily congested area in which trip price in the no-toll equilibrium is higher than that in the social optimum, implying that the efficiency gain from tolling exceeds the revenue raised from the toll.

Casual empiricism, supported by recent evidence collected by transportation scientists, suggests that the central areas of many of the world's largest cities (though few in the US) are sufficiently heavily congested that efficiency gains of this order could be achieved from time-varying congestion pricing. However, most of these cities are in countries that lack the capacity to implement sophisticated tolling schemes. Whether there are traffic restraint policies akin to ramp metering on freeways that these cities could practically implement to reduce hypercongestion is an open question.

7 References

References

Anderson, D. and H. Mohring. 1997. Congestion costs and congestion pricing, in D. Greene and D. Jones, eds. The Full Costs and Benefits of Transportation: Contributions to Theory, Method, and Measurement. Berlin: Springer-Verlag.

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

Arnott, R., T. Rave, and R. Schb. 2005. Alleviating Urban Traffic Congestion. Cambridge, MA: MIT Press.

Arnott, R. and E. Inci. 2006. An integrated model of downtown parking and traffic congestion. *Journal of Urban Economics* 60: 418-442.

Arnott, R. and E. Inci. 2011. The stability of downtown parking and traffic congestion. *Journal of Urban Economics* 68: 260-276.

Cassidy, M. and R. Bertini. 1999. Some traffic features at freeway bottlenecks. *Transportation Research B* 33: 25-42.

Chu, X. 1995. Endogenous trip scheduling: The Henderson approach reformulated and compared with the Vickrey approach. *Journal of Urban Economics* 37: 324-343.

Daganzo, C., V. Gayah, and E. Gonzales. 2011. Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness, and instability. *Transportation Research B* 45: 278-288.

DePalma, E. and R. Arnott. 2010. Morning commute in a single-entry traffic corridor with no late arrivals. mimeo.

Geroliminis, N. and C. Daganzo. 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research B* 42, 759-770.

Greenshields, B. 1935. A study of traffic capacity. *Highway Research Board Proceedings* 14, 448-477.

Hall, F., B. Allen, and M. Gunter. 1986. Empirical analysis of freeway flow-density relationships. *Transportation Research A* 20: 197-210.

Henderson, J.V. 1981. The economics of staggered work hours. *Journal of Urban Economics* 9: 349-364.

Lighthill, M. and G. Whitman. 1955. On kinematic waves II: A theory of traffic flow on long, crowded roads. *Proceedings of the Royal Society (London)* A229: 317-345.

Lipman, T. 1996. Using road pricing revenue: Economic efficiency and equity considerations. *Transportation Research Record* 1558: 24-28.

Maerivoet, S. and B. De Moor. 2008. Traffic flow theory. arXiv:physics/0507126v1 [physics.soc-ph].

Parry, I. and K. Small. 2005. Does Britain or the United States have the right gasoline tax? *American Economic Review* 95: 1276-1289.

Richards, P. 1956. Shock waves on highways. *Operations Research* 4: 42-51.

Small, K. 1982. The scheduling of consumer activities: Work Trips. *American Economic Review* 72: 467-479.

Small, K. 1992. Using the revenues from congestion pricing. *Transportation*19: 359-381.

Small, K. and X. Chu. 2003. Hypercongestion. *Journal of Transport Economics and Policy* 37: 319-352.

Small, K. and E. Verhoef. 2007. The Economics of Urban Transporation. Abingdon, England: Routledge.

Vickrey, W. 1969. Congestion theory and transport investment. *American Economic Review* 59: 251-260.

Vickrey, W. 1994. Types of congestion pricing models. mimeo. Columbia University.

Velocity[1]	$v(t) = \begin{cases} \frac{c_0}{c_1+\beta t} = \frac{\alpha L}{\underline{c}-\beta t^*+\beta t}\,, & t \in [\underline{t}, t^*] \\ \frac{c_0}{c_2+\beta t} = \frac{\alpha L}{\underline{c}+\gamma t^*-\gamma t}\,, & t \in [t^*, \bar{t}] \end{cases}$
Density[1]	$k(t) = \left(1 - \frac{v(t)}{v_0}\right) k_j = \begin{cases} k_j - \frac{k_j \alpha L}{v_0(\underline{c}-\beta t^*+\beta t)}\,, & t \in [\underline{t}, t^*] \\ k_j - \frac{k_j \alpha L}{v_0(\underline{c}+\gamma t^*-\gamma t)}\,, & t \in [t^*, \bar{t}] \end{cases}$
Arrival Rate[1]	$a(t) = \frac{vk}{L} = \begin{cases} \frac{k_j}{L}\frac{c_0}{c_1+\beta t} - \frac{k_j c_0^2}{v_0 L (c_1+\beta t)^2}\,, & t \in [\underline{t}, t^*) \\ \frac{k_j}{L}\frac{c_0}{c_2-\gamma t} - \frac{k_j c_0^2}{v_0 L (c_2-\gamma t)^2}\,, & t \in (t^*, \bar{t}) \end{cases}$
Cumulative Arrival Rate	$A(t) = \int_0^t a(u)\, du$
Departure Rate Case I[2] (No Late Departures) Case II (Late Departures)	$d(t) = \dot{k}(t) + a(t) = \frac{k_j c_0}{L(c_1+\beta t)} + \frac{k_j c_0\left(\beta - \frac{c_0}{L}\right)}{v_0(c_1+\beta t)^2}\,, t \in [\underline{t}, t^*)$ $d(t) = \dot{k}(t) + a(t) = \begin{cases} \frac{k_j c_0}{L(c_1+\beta t)} + \frac{k_j c_0\left(\beta - \frac{c_0}{L}\right)}{v_0(c_1+\beta t)^2}\,, & t \in [\underline{t}, t^*) \\ \frac{k_j c_0}{L(c_2-\gamma t)} + \frac{k_j c_0\left(-\gamma - \frac{c_0}{L}\right)}{v_0(c_2-\gamma t)^2}\,, & t \in (t^*, \bar{t}] \end{cases}$
Cumulative Departure Rate	$D(t) = \int_0^t d(u)\, du$
Time of First Departure	$v(\underline{t}) = v_0 \rightarrow \underline{t} = \frac{\frac{c_0}{v_0} - c_1}{\beta}$
Time of Last Departure Case I (No Late Departures) Case II (Late Departures)	$\bar{t} = t^*$ $\bar{t} = \frac{c_2 - \frac{c_0+\gamma L}{v_0}}{\gamma}$
Trip Cost	$\underline{c} = \frac{\alpha L}{v(t^*)}$ or by solving $\int_{\underline{t}}^{\bar{t}} d(t) = N$

Table 1: No-toll Equilibrium

Notes: 1. After $\bar{t}$, there are no departures. Over this interval, density evolves according to the differential equation $\dot{k}(t) = -\frac{k(t)v(k(t))}{L}$, with $k(\bar{t})$ determined as indicated above. Over this interval, $v(t)$ is determined as $v(k(t))$ and $a(t)$ as $-\dot{k}(t)$.

2. Case I applies if $d(t^{*+}) < 0$.

3. Greenshields Relation is assumed: $v(k) = v_0(1 - \frac{k}{k_j})$.

4. t^* is exogenous.

Velocity[1]	$v(t) = v_0\left(1 - \frac{k}{k_j}\right) = \begin{cases} \frac{1}{2}v_0\left[1 + \frac{\frac{\alpha L}{v_0}}{\frac{\alpha L}{v_0}+\beta t}\right], & t \in [0, t^*] \\ \frac{1}{2}v_0\left[1 + \frac{\frac{\alpha L}{v_0}}{\frac{\alpha L}{v_0}+\beta t^* - \gamma(t-t^*)}\right], & t \in [t^*, \bar{t}] \end{cases}$
Density[1]	$k(t) = \begin{cases} \frac{k_j}{2}\left[1 - \frac{\frac{\alpha L}{v_0}}{\frac{\alpha L}{v_0}+\beta t}\right], & t \in [0, t^*] \\ \frac{k_j}{2}\left[1 - \frac{\frac{\alpha L}{v_0}}{\frac{\alpha L}{v_0}+\beta t^* - \gamma(t-t^*)}\right], & t \in [t^*, \bar{t}] \end{cases}$
Arrival Rate[1]	$a(t) = \frac{vk}{L} = \begin{cases} \frac{v_0 k_j}{4L}\left[1 - \frac{\left(\frac{\alpha L}{v_0}\right)^2}{\left(\frac{\alpha L}{v_0}+\beta t\right)^2}\right], & t \in [0, t^*] \\ \frac{v_0 k_j}{4L}\left[1 - \left(\frac{\frac{\alpha L}{v_0}}{\frac{\alpha L}{v_0}+\beta t^* - \gamma(t-t^*)}\right)^2\right], & t \in [t^*, \bar{t}] \end{cases}$
Cumulative Arrival Rate	$A(t) = \int_{\underline{t}}^{t} a(u)\, du$
Departure Rate Case I[2] (No Late Departures)	$d(t) = \dot{k}(t) + a(t) = \frac{v_0 k_j}{4L} + \frac{k_j}{4}\frac{\frac{\alpha L}{v_0}(2\beta - \alpha)}{\left(\frac{\alpha L}{v_0}+\beta t\right)^2}, \quad t \in [\underline{t}, t^*)$
Case II (Late Departures)	$d(t) = \dot{k}(t) + a(t) = \begin{cases} \frac{v_0 k_j}{4L} + \frac{k_j}{4}\frac{\frac{\alpha L}{v_0}(2\beta - \alpha)}{\left(\frac{\alpha L}{v_0}+\beta t\right)^2}, & t \in [\underline{t}, t^*) \\ \frac{v_0 k_j}{4L} - \frac{k_j}{4}\frac{\frac{\alpha L}{v_0}(2\gamma + \alpha)}{\left(\frac{\alpha L}{v_0}+\beta t\right)^2}, & t \in (t^*, \bar{t}] \end{cases}$
Cumulative Departure Rate	$D(t) = \int_{\underline{t}}^{t} d(u)\, du$
Time of Last Departure Case I (No Late Departures) Case II (Late Departures)	 $\bar{t} = t^*$ $\mathrm{d}(\bar{t}) = 0$

Table 2: Social Optimum

Notes: 1. After $\bar{t}$, there are no departures. Over this interval, density evolves according to the differential equation $\dot{k}(t) = -\frac{k(t)v(k(t))}{L}$, with $k(\bar{t})$ determined as indicated above. Over this interval, $v(t)$ is determined as $v(k(t))$ and $a(t)$ as $-\dot{k}(t)$.

2. Case I applies if $d(t^{*+}) < 0$.

3. Greenshields Relation is assumed: $v(k) = v_0(1 - \frac{k}{k_j})$.

4. Time is measured from the start of the rush hour, i.e. $\underline{t} = 0$. t^* is determined by the optimality condition $A(t^*) = \frac{\gamma}{\beta+\gamma}$