Journal of Urban Economics 76 (2013) 110–121

Contents lists available at SciVerse ScienceDirect

Journal of Urban Economics

www.elsevier.com/locate/jue

A bathtub model of downtown traffic congestion

Richard Arnott

Department of Economics, University of California Riverside, Riverside, CA 92521, United States

ARTICLE INFO

Article history:
Available online 17 January 2013

JEL classification:
R41

Keywords:
Rush hour
Traffic congestion
Equilibrium
Optimum
Toll

ABSTRACT

In 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 major cities during rush hour. This paper analyzes a bathtub model of downtown rush-hour traffic congestion that builds on ideas put forward by William Vickrey. Water flowing into the bathtub corresponds to cars entering the traffic stream, water flowing out of the bathtub to cars exiting from it, and the height of water in the bathtub to traffic density. Velocity is negatively related to density, and outflow is proportional to the product of density and velocity. Above a critical density, outflow falls as density increases (traffic jam situations). When demand is high relative to capacity, applying an optimal time-varying toll generates benefits that may be considerably larger than those obtained from standard models and that exceed the toll revenue collected.

1. Introduction

There are two standard models of traffic congestion employed by economists. In the first, which is familiar from undergraduate economics textbooks, as demand increases, equilibrium traffic flow increases. In the second, the bottleneck model (Vickrey, 1969; Arnott et al., 1993), which treats morning rush-hour traffic dynamics, congestion is modeled as a deterministic queue behind a bottleneck of fixed flow capacity. In neither model does traffic flow fall as demand increases. But casual experience and common sense[1] suggest that, in the downtown areas of heavily congested cities, due to traffic jams traffic flow is lower at the peak of the rush hour than during less congested periods of the day. Only very recently have traffic engineers started to measure traffic flows at the level of downtown neighborhoods, and early results provide strong support for this phenomenon (Geroliminis and Daganzo, 2008; Daganzo et al., 2011). This phenomenon has important implications for the management of downtown traffic congestion in very congested cities. First, the time loss due to rush-hour congestion could be sharply reduced if traffic restraint policies restricting entry flow into the downtown area were implemented. Second, the benefits from applying optimal time-varying congestion tolls are substantially higher that those estimated from the standard models, and exceed the toll revenue raised.

This paper develops a bathtub model of downtown traffic congestion that captures traffic-jam situations.[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 and tunnels into Manhattan or from parking spaces in Manhattan. These cars correspond to the inflow of water into the bathtub. Similarly, cars leaving the traffic stream, either exiting Manhattan or entering parking spaces in Manhattan, correspond to the outflow of water from the bathtub. Traffic density corresponds to the height of water in the bathtub. Traffic velocity is assumed to be inversely proportional to traffic density. Via the fundamental identity of traffic flow, traffic flow equals traffic density times traffic velocity. And outflow from the bathtub is assumed to be proportional to traffic flow. The relationship between traffic velocity and traffic density is such that outflow is increasing in the height of water in the bathtub up to some critical height, and decreasing in the height of the water above

E-mail address: richard.arnott@ucr.edu

[1] There is abundant anecdotal evidence of downtown rush-hour traffic speeds of 2–5 mph in cities such as Central London before the congestion toll, and central Beijing, Cairo, Djakarta, Istanbul, Moscow, and Mexico City, but no reliable documentation. It seems implausible that traffic flow can be close to its maximum (capacity flow) at such low 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 both of Manhattan and of an (imperfect) hydraulic analogy. A dated outline of a model and incomplete notes on it were found in his files after his death (Vickrey, 1991). I am grateful to Keith Knapp for pointing out that the term "bathtub model" is used in biology and hydrology to refer to a dynamic model (of an aquifer, for example) in which a disturbance at one location is instantaneously propagated to all other locations. That characteristic is a property of this paper's bathtub model too.

0094-1190/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jue.2013.01.001

that height.[3] Vickrey used the term "hypercongestion" to refer to traffic jam situations, where traffic flow is inversely related to density. In the early morning rush hour, the inflow into the bathtub exceeds the rate at which the bath drains, and the water level rises. If the water level rises much about the critical level, the bathtub takes a long time to drain. A planner regulating the inflow into the bathtub would ensure that the water level never rises above the critical height.[4]

Vickrey's bathtub analogy for downtown traffic congestion has the ring of truth about it. But only recently has the relationship between average traffic density and average traffic velocity at the level of downtown neighborhoods received strong empirical confirmation. Traffic engineers started to collect detailed data on freeway traffic flow in the late 1970s (e.g., Hall et al., 1986). Analysis of such data (e.g., Cassidy and Bertini, 1999) suggests that freeways contain bottlenecks whose discharge rates fall only modestly as the length of the queue behind them increases. Based partly on these analyses, the prevailing wisdom in urban transportation economics is that the aggregative or macroscopic behavior of rush-hour traffic in metropolitan areas is broadly consistent with the bottleneck model, with flow being approximately constant and congestion delay taking the form of quasi-queues.[5] Only in the last 5 years have comparable data been collected for downtown neighborhoods in larger cities[6], using a network of stationary sensors, supplemented by mobile sensors (GPS devices in taxis). Analysis of these data (e.g., Geroliminis and Daganzo, 2008) provides strong support for what the authors refer to as the existence of a stable, urban-scale macroscopic fundamental diagram (MFD)—a stable graph relating traffic flow to density at the level of a downtown neighborhood, which includes a hypercongested portion.[7] The results of these empirical studies are broadly supported by the current generation of downtown traffic microsimulation models, which incorporate elements permitting flow to fall as density increases.[8] Fig. 1 reproduces Fig. 3 from Gonzales et al. (2011). It plots flow versus density for a large neighborhood in Yokohama (measured), San Francisco (simulated), and Nairobi (simulated). The graphs for Nairobi and San Fransisco are particularly striking, showing a sharp reduction in flow with high levels of congestion.

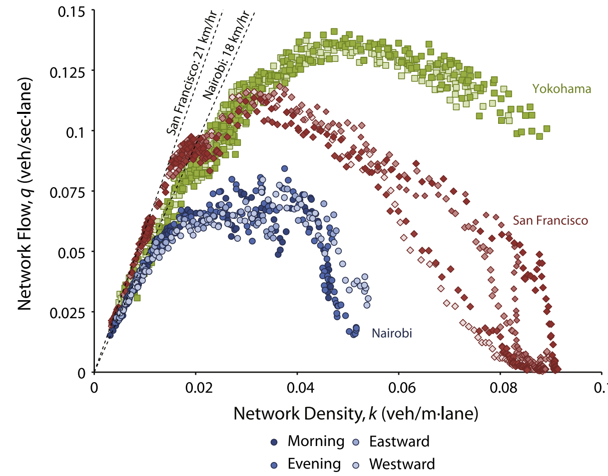

Fig. 1. Comparison of macroscopic flow-density phase paths for Yokohama (measured; Geroliminis and Daganzo, 2008), Nairobi (simulated; Gonzales et al., 2011), and San Francisco (simulated; Geroliminis and Daganzo, 2007).

While the paper restricts attention to traffic congestion, the bathtub model can be adapted to other congestible facilities for which heavy loading results in decreased output, such as brown-outs and black-outs in electrical systems and jammed switches in telephone circuits.

Section 2 presents a particularly simple bathtub model of the morning rush hour, in which velocity is linearly inversely related to density, commuters are identical, and the user cost function corresponds to that in the bottleneck model. Section 3 solves for equilibrium in the morning rush hour in the absence of tolls. Section 4 solves for the corresponding social optimum and for the time-varying congestion toll that supports it. Section 5 compares the no-toll equilibrium and the social optimum for a particular numerical example calibrated to correspond to a downtown area where demand is high relative to capacity. Section 6 discusses policy insights and directions for future research, and Section 7 concludes.

2. The bathtub model

Consider an isotropic[9] downtown area. There are N identical commuters per unit area, each of whom must travel from home to work in the morning rush hour and has work start time t^*, and experiences travel time cost and schedule delay cost (the cost of arriving at work inconveniently early or late). Classical flow congestion is

[3] In the traffic engineering literature, the "capacity" of a section of road is defined to be its maximum sustainable flow. In this paper, the congestion technology has the property that flow cannot exceed the maximum sustainable flow, so that "capacity" or "capacity flow" will refer simply to maximum flow. "Capacity density" and "capacity velocity" will refer to the density and velocity corresponding to capacity flow. Here, capacity density corresponds to the critical height of water in the bathtub.

[4] I thank Kenneth Small for making me aware of two papers antedating Vickrey (1991) that share at least part of his conception, in particular his "macroscopic" conception of a well-defined relationship between traffic aggregates at the scale of a downtown area. Olszewski and Suchorzewski (1987) presented a complex macroscopic model of traffic congestion in the city center of Warsaw that relates aggregate traffic flow to average speed. Ardekani and Herman (1987) estimated the parameters of Herman and Prigogine's (1979) macroscopic two-fluid model of "town" traffic for Austin and Dallas, assuming a stable relationship between mean density, mean flow, mean velocity, and the fraction of vehicles stopped. Neither paper, however, considers the evolution of congestion over the rush hour that is an essential feature of bathtub model.

[5] The prevailing wisdom is also based partly on the assumption, often implicit, that downtown traffic congestion has the same qualitative properties as freeway traffic congestion.

[6] Traffic flow data on individual streets have been collected on a periodic basis for many years, and provide the basis for calibrating the traffic microsimulation models that all major cities in developed countries now employ to simulate the effects of improvements to the transportation infrastructure. But Daganzo and Geroliminis were the first to measure traffic flow simultaneously over an entire downtown neighborhood for complete rush hours, and to show that the traffic flows so measured aggregate into a stable diagram relating flow and density. Olszewski and Suchorzewski (1987) assumed specific functional forms drawn from the traffic engineering literature, estimated their parameters for Warsaw, and then employed the fitted functional forms for policy simulation. Ardekani and Herman (1987) assumed specific functional forms, based on the two-fluid model of Herman and Prigogine (1979), and estimated their parameters for Austin and Dallas. May et al. (2000) and Liu et al. (2011) investigated the aggregative properties of calibrated traffic microsimulation models for Cambridge and York, England.

[7] A stable, urban-scale macroscopic fundamental diagram should not be universally expected. Field experiments and traffic microsimulations (Mazloumian et al., 2010; Geroliminis and Sun, 2011) indicate that the diagram is more stable (in the sense of the observations being less scattered) the more similar are links in the network and the more "redundant" the network (the more paths there are on average between an origin and a destination). The latter result derives from drivers having the opportunity to divert to avoid jammed links—a smoothing phenomenon.

[8] Think of a downtown network of one-way streets with signalized intersections. Assume that each intersection operates at capacity if the queues in both directions of traffic are sufficiently long. As demand increases, an increasingly high proportion of intersections satisfy this condition and operate at capacity, and in the limit all intersections operate at capacity. In this simple model of downtown traffic congestion, flow increases as demand increases. If flow is to fall as demand increases, the model must be modified. One realistic modification is spillbacks—the queue at an intersection becomes sufficiently long that it spills back into the upstream intersection, blocking it. Other realistic modifications include pedestrians who cross against the light, and queued drivers who, being distracted, are slow to advance when the light turns green.

[9] "Isotropic" means spatially symmetric. In this context, one implication of isotropy is that the density of economic activity is uniform over the space. An infinite homogeneous plane and the homogeneous surface of a sphere are isotropic spaces in two dimensions. For concreteness, one may imagine a dense Manhattan network of uniform streets on an infinite homogeneous plain.