Regional Science and Urban Economics 43 (2013) 764–782

Contents lists available at ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regec

Trip-timing decisions with traffic incidents☆

Mogens Fosgerau [a,b,*], Robin Lindsey [c]

[a] Technical University of Denmark, Denmark
[b] Centre for Transport Studies, Sweden
[c] Sauder School of Business, University of British Columbia, Canada

ARTICLE INFO

Article history:
Received 28 June 2012
Received in revised form 30 June 2013
Accepted 8 July 2013
Available online 19 July 2013

JEL classification:
C61
D62
R41

Keywords:
Departure-time decisions
Bottleneck model
Traffic incidents
Congestion
Scheduling utility
Morning commute
Evening commute

ABSTRACT

This paper analyzes traffic bottleneck congestion when drivers randomly cause incidents that temporarily block the bottleneck. Drivers have general scheduling preferences for time spent at home and at work. They independently choose morning departure times from home to maximize expected utility without knowing whether an incident has occurred. The resulting departure time pattern may be compressed or dispersed according to whether or not the bottleneck is fully utilized throughout the departure period on days without incidents. For both the user equilibrium (UE) and the social optimum (SO) the departure pattern changes from compressed to dispersed when the probability of an incident becomes sufficiently high. The SO can be decentralized with a time-varying toll, but drivers are likely to be strictly worse off than in the UE unless they benefit from the toll revenues in some way. A numerical example is presented for illustration. Finally, the model is extended to encompass minor incidents in which the bottleneck retains some capacity during an incident.

1. Introduction

Traffic congestion imposes a heavy burden in urban areas. The Texas Transportation Institute conducts an annual survey of traffic congestion in the US. According to its 2012 report, in 2011 congestion caused an estimated 5.5 billion hours of travel delay and 2.9 billion gallons of extra fuel consumption with a total cost of $121 billion (Schrank et al., 2012). The average cost per automobile commuter in the urban areas studied was $818. Nonrecurring traffic congestion due to accidents, bad weather, special events, and other shocks accounts for a large fraction of the total delays. According to Schrank et al. (2011, Appendix B, p. B-27) incident-related delays alone contribute 52–58% of total delay in US urban areas.[1]

Unanticipated travel delays upset peoples' travel plans, and may cause them to arrive late with serious consequences for commuting, business, and other types of trips. Travelers can sometimes adjust to the threat of delays by changing their transport mode or destination, or even canceling trips, but a more common response is to adjust departure times. Researchers have long been interested in studying the adjustment process, and they have adopted various modeling approaches. In an early and insightful study, Gaver (1968) derived the optimal departure time for a driver faced with stochastic travel time who incurs costs from both travel time and schedule delay. The optimal policy, which Gaver called a headstart strategy, entails a probabilistic trade-off between arriving early and arriving late. Gaver assumed that travel time has a constant and exogenous variance, and he did not attempt to derive an endogenous travel time distribution as a dynamic equilibrium. His approach was adopted and extended by Knight (1974), Hall (1983), Noland and Small (1995), and Noland (1997).

All these studies use models with flow congestion. An alternative approach is to use the Vickrey (1969) bottleneck model in which congestion delay takes the form of queuing. A series of studies by Arnott et al. (1991, 1999) and Lindsey (1994, 1999) introduced stochasticity into the bottleneck model by assuming that capacity and/or demand fluctuate randomly from day to day, but are constant during the period of use

☆ We are grateful to the editor and two referees for very helpful comments. Mogens Fosgerau would also like to thank Stefan Mabit and Lars Peter Østerdal for comments on an earlier draft. Both authors thank participants at the Kuhmo-Nectar Conference, Stockholm, June 30, 2011. Mogens Fosgerau has been supported by the Danish Strategic Research Council.
* Corresponding author. Tel.: +45 54 25 65 21.
E-mail addresses: mf@transport.dtu.dk (M. Fosgerau), robin.lindsey@sauder.ubc.ca (R. Lindsey).

[1] The 2012 Report does not repeat this estimate. As Hall (1993) observes, the contribution of nonrecurrent congestion is difficult to determine because it depends on the magnitude and timing of recurrent congestion, and vice versa. Drivers may underestimate the prevalence of nonrecurrent congestion because incident-induced queues can persist long after the incidents are cleared away.

0166-0462/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.regsciurbeco.2013.07.002

on a given day. For want of a better term, we will call this the "daily-shocks" model.[2]

Our paper differs from these earlier bottleneck-model studies in three ways. First, they adopted the traditional specification of trip-timing preferences used by Vickrey (1969) in which individuals have a preferred time to arrive at their destination and incur a schedule delay cost proportional to the amount of time they arrive earlier or later. Following Börjesson et al. (2012) we will call this the "step" model. Here we adopt a more general scheduling utility function approach that incorporates preferences for time spent at different activities. We apply the model to commuting trips by specifying preferences for time spent at home and at work.[3]

Second, and more fundamentally, we assume that capacity can fluctuate while trips are being made rather than being determined before travel begins. Third, we assume that capacity reductions are due to incidents caused by drivers during their trip. The timing of shocks is therefore endogenous to the model rather than exogenous as in earlier studies. Since drivers are responsible for most incidents, this within-day, endogenous specification of capacity fluctuations accounts for a significant portion of nonrecurring congestion that occurs. It also provides the basis for assessing tolling and other policies to reduce the costs of congestion by altering peoples' travel decisions. For most of the paper we assume that capacity is reduced to zero by an incident although in a final section we examine a variant of the model in which loss of capacity is partial.

Two unpublished studies cover part of the same ground as we do. Schrage (2006) derives the unregulated and socially optimal departure rates for a single road link when the accident rate is a function of the inflow rate and therefore endogenous. Her model differs from ours in three main respects. First, she uses the Henderson (1974) flow congestion model in which a driver's travel time is determined by the aggregate departure rate when he starts his trip. This model has no state variable analogous to queue length in the bottleneck model. Second, capacity is reduced only partially in an incident and it subsequently recovers slowly, and deterministically, rather than all at once. Third, drivers are assumed to know whether and when an accident has occurred before they depart. Schrage derives the optimal time-varying and state-dependent toll that decentralizes the social optimum, but she does not solve for the timing of departures in either the unregulated user equilibrium or the social optimum. In independent work, Peer et al. (2010) use the bottleneck model to analyze incidents in which, like Schrage (2006), capacity loss is partial. They treat incident timing as exogenous and assume that an incident persists until all drivers have completed their trips. They also adopt the "step" model of trip-timing preferences. Finally, they limit attention to the unregulated user equilibrium and do not examine the social optimum or tolling.

In our paper we undertake a systematic analysis of both user (i.e., Nash) equilibrium and socially optimal trip-timing decisions when drivers do not know whether an incident has occurred before they decide when to depart. We solve for the optimal time-varying (but state-independent) toll that decentralizes the social optimum. One of the questions we address is whether the bottleneck operates at capacity throughout the travel period on days when no incident occurs, or whether some capacity goes "unused". We show that for both the user equilibrium and social optimum, spare capacity does exist for part, or all, of the travel period if incidents are sufficiently probable.[4] In contrast to the daily-shocks model, departures can be more spread out in the user equilibrium than in the social optimum. Another difference is that the socially-optimal departure rate can decrease, rather than increase, over time.

The paper is organized as follows. Section 2 describes the model. Section 3 summarizes the main features of user equilibrium and social optimum for the deterministic variant of the model with no incidents. Section 4 derives properties of the user equilibrium with incidents. Section 5 conducts a parallel analysis of the social optimum. Section 6 presents a numerical example calibrated for morning commutes, and then considers a variant for evening commutes. Section 7 undertakes a partial analysis of an extension of the model in which the bottleneck retains some capacity during an incident. Finally, Section 8 concludes with a summary and ideas for extension.

2. The model

A continuum of N identical individuals drive alone from a common origin through a bottleneck to a common destination.[5] To be concrete, in most of the paper the trip is assumed to be a morning commute from home (H) to work (W). (However, an evening commute is also examined in the example section.) Departure time from home is denoted by t. Drivers[6] depart at a rate $\rho(t)$ during a set of times T; cumulative departures are thus $R(t) = \int_{\{v \in T | v \le t\}} \rho(v) dv$.[7] Free-flow travel time before and after reaching the bottleneck is normalized to zero. A driver departing at t encounters a queuing delay of $q(t)$ at the bottleneck and reaches work at time $a = t + q(t)$. Drivers have scheduling preferences[8] described by the utility function

$$u(t,a) = \int_{t_H}^{t} \beta(v)dv + \int_{a}^{t_W} \gamma(v)dv. \tag{1}$$

The limits of integration, t_H and t_W, are chosen such that all travel takes place within the interval $[t_H, t_W]$. Function $\beta(\cdot) > 0$ denotes the flow of utility from being at home, and function $\gamma(\cdot) > 0$ denotes utility from being at work. Functions $\beta(\cdot)$ and $\gamma(\cdot)$ are assumed to be continuously differentiable with derivatives $\beta' < 0$ and $\gamma' > 0$ and to intersect at time t^*.[9] Utility from time spent driving is normalized to zero. These assumptions ensure that, for any fixed trip duration, there is a unique departure time t, $t < t^*$, that maximizes scheduling utility. They also assure that $u(t, a)$ is strictly increasing in t, strictly decreasing in a, and globally strictly concave. Two final assumptions, $\underset{v \to t_H}{Lim}\, \beta(v) = \infty$ and $\underset{v \to t_W}{Lim}\, \gamma(v) = \infty$, will ensure existence of a Nash equilibrium in departure times.[10]

2 Arnott et al. (1991, 1999) and Li et al. (2008) analyze user equilibrium in the daily-shocks model, whereas Lindsey (1994, 1999) focuses on the social optimum. Other recent papers have also studied random travel times using the bottleneck model. Xin and Levinson (2007) assume that travel times are exogenous and independently distributed over time, and their model does not feature incidents per se. Fosgerau (2010) shows how the dynamics of random congestion induce characteristic loops in the relationship between the mean and the variance of travel time over different times of day. de Palma and Fosgerau (2011) analyze random queue sorting whereby travel time is random from the perspective of individual travelers, but capacity and demand are fixed.

3 Jenelius et al. (2011) use a similar scheduling utility function approach to study the effects of unpredictable travel time shocks on trip-timing decisions. They apply the model to a full day of activity including morning and evening commutes. Their model differs in featuring shocks that are exogenous and independent of time of day. There is also no traffic congestion in their model.

4 Holding spare capacity is broadly consistent with policies of reserving shoulder lanes for use during accidents and other disruptions.

5 A notational glossary is provided at the end of the paper.

6 Throughout the paper we will refer to "drivers" even though individuals are treated as a continuum in the model so that there are no discrete or atomic agents. Reference to "drivers", "users", "commuters" and so on is common in the bottleneck model literature, and it facilitates exposition.

7 All statements about ρ in the paper will be "almost surely", since ρ can take arbitrary values on sets of Lebesgue measure zero without affecting aggregate behavior or welfare. To ease exposition this detail will be ignored.

8 This formulation of scheduling preferences originates from Vickrey (1969, 1973) and has been used by Tseng and Verhoef (2008), Fosgerau and Engelson (2011), Fosgerau and de Palma (2012), Jenelius et al. (2011), and Börjesson et al. (2012).

9 The notation differs from that in the step model where β denotes the cost per minute of arriving before t^*, and γ denotes the cost per minute of arriving after t^*. The assumptions $\beta' < 0$ and $\gamma' > 0$ rule out the step model because the (implicit) $\beta(\cdot)$ and $\gamma(\cdot)$ functions in that model are constants except at t^* where $\gamma(\cdot)$ steps up. This is not particularly restrictive since the step-model preferences can be approximated arbitrarily closely by differentiable functions. Nevertheless, the assumptions could be generalized as in Fosgerau and Engelson (2011).

10 These assumptions are relaxed in the example of Section 6 where $\beta(\cdot)$ and $\gamma(\cdot)$ are linear functions.