Transportation Research Part B 55 (2013) 142–153

Contents lists available at SciVerse ScienceDirect

Transportation Research Part B

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

Development and comparison of choice models and tolling schemes for high-occupancy/toll (HOT) facilities

Lauren M. Gardner [a,b,*], Hillel Bar-Gera [c], Stephen D. Boyles [d]

[a] School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia
[b] NICTA, Sydney, NSW 2052, Australia
[c] Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
[d] Civil, Architectural and Environmental Engineering Department, University of Texas at Austin, Austin, TX 78712, USA

ARTICLE INFO

Article history:
Received 12 November 2012
Received in revised form 5 June 2013
Accepted 6 June 2013

Keywords:
High-occupancy/toll lanes
Tolling
Managed lanes
Value of time

ABSTRACT

A high-occupancy/toll (HOT) lane is an increasingly popular form of traffic management strategy which reserves a set of freeway lanes for HOVs and transit users, while allowing low-occupancy vehicles (LOVs) to enter for a fee. In turn, HOT lanes maintain a minimal level of service by regulating the volume of entering LOVs. The focus of this paper is how to model the choice process of individual drivers, which dictates the volume of LOVs that choose to pay and take the HOT lane. Such models and the insights they provide can be very helpful for the toll setting process. Two simple formulations (an all-or-nothing assignment and an additive logit model) are compared with a proposed formulation based on the population value of time (VOT) distribution. Both static and dynamic toll setting algorithms are studied based on the proposed lane choice model, and their performance is compared under deterministic traffic behavior.

1. Introduction

High-occupancy/toll (HOT) lanes refer to high-occupancy vehicle (HOV) facilities that allow lower-occupancy vehicles (LOVs) to pay a toll to enter. HOT lanes represent one way to utilize the remaining capacity of HOV lanes, and are becoming an increasingly prevalent form of congestion pricing in the US and internationally. The first HOT lane was implemented in 1995 on State Route 91 in Orange County, California. Since 1995 more than a dozen states have successfully implemented HOT lanes, with multiple HOT projects in development. In addition, many authorities of current HOT lanes are using dynamic pricing.

HOT lane operation policies have multiple objectives – to maximize passenger throughput of the entire freeway (both general purpose (GP) and toll lanes), and to provide free-flow traffic service on the managed lane. Because HOV lanes are designed "first and foremost to provide less congested conditions for carpoolers and transit users" (Munnich, 2006), priority is often given to the second objective.

In order to guarantee a minimal level of service on the HOT lane the volume of entering LOVs must be regulated. The volume of LOVs that will choose to pay and take the HOT lane is dependent on the toll value, the user's value of time (VOT), value of reliability and the travel time savings gained from taking the HOT lane relative to the GP lane (Devarasetty et al., 2012).

* Corresponding author at: School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia. Tel.: +61 (2) 9385 5562.
E-mail addresses: l.gardner@unsw.edu.au (L.M. Gardner), bargera@bgu.ac.il (H. Bar-Gera), sboyles@austin.utexas.edu (S.D. Boyles).

0191-2615/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.trb.2013.06.006

Modeling the choice process of HOT lanes by individual drivers is crucial to understanding how tolls should be set. The core contribution of this paper, presented in Section 3.2, concerns the choice model. Two simple formulations (an all-or-nothing assignment and an additive logit model) are first discussed, followed by a formulation we believe to be superior, based on the population VOT distribution.

Using the VOT lane choice formulation, we examine possible tolling strategies based on inflow rates by vehicle occupancy class and the VOT distribution. We show that under deterministic (or perfect forecast) conditions, the optimal toll-setting algorithm in terms of minimizing average passenger delay is to maximize inflow to the HOT lane without exceeding capacity. We compare the results of this optimal dynamic tolling scheme with static tolling options, relative to base scenarios where either all lanes are GP, or only HOVs can use the managed lane. The simplification of deterministic conditions enables to clarify basic insights. The discussion also offers a foundation for future research on the implication of various types of stochastic aspects, as well as operational implications such as traffic monitoring locations, and traffic classification abilities.

A specific facility configuration is used throughout the paper, as a basis for numerical evaluation, as well as for concrete demonstration of the discussed concepts. The configuration consists of a freeway corridor with a single managed lane and three GP lanes that merge to two lanes at the downstream bottleneck.

The contributions of this paper are (1) a novel VOT choice model and comparison to conventional choice models; (2) development of an optimal toll-setting algorithm under deterministic conditions (or perfect forecast); and (3) numerical comparison of several tolling alternatives using the proposed choice model.

The remainder of this paper is organized as follows: Section 2 reviews related literature, Section 3 describes the proposed methodology, Section 4 provides numerical results for a demonstrative facility, and Section 5 concludes the paper.

2. Background

Previous literature offers background on several aspects relevant for this paper, including: the prevalence of HOT facilities and their classification; methods used for determining toll values; models used for HOT corridor scenario evaluation, particularly the component in the evaluation model used for the representation of HOT lane choice by toll-paying drivers; and empirical research regarding VOT and VOT distributions.

Multiple region-specific case studies have been conducted to investigate the impact of implementing HOT lanes and other types of managed lanes (Burris et al., 2009, 2011; Bar-Gera, 2012; Cao et al., 2011; Parkany, 1998; Chung and Choi, 2010; Fosgerau, 2011; Gordon et al., 2004; Kuhn et al., 2005). In addition, current HOT facilities have been evaluated in attempts to identify user preferences and improve future performance (Borrell-Rovira and Supernak, 2011; Goel and Burris, 2011; SCAG, 2004).

HOT facilities can be classified by the tolling strategy used: pre-scheduled by time of day, or determined in real time (Chung and Recker, 2011). As of June 2013 examples of HOT facilities operating with a pre-scheduled time of day toll include SR-91 in Orange County, California, I-25 in Denver, Colorado and I-10Wand US 290 in Houston, Texas. Examples of HOT facilities implementing tolls determined in real time include I-15 in Salt Lake City, Utah, I-15 in San Diego, California, I-110 in Los Angeles, California, I-95 in Miami, FL, SR-167 in Seattle, WA, I-394 and I-35W in Minneapolis, MN and I-85 in Atlanta, Georgia.

In a relatively early study Dahlgren (2002) evaluated HOT lanes by assuming "...that as long as there is congestion on the main lanes, the toll can be set so that the HOT lane is fully utilized but not congested." We shall refer to this option as *full-utilization* tolls. For evaluation purposes, assuming that full-utilization tolls can be obtained is not necessarily realistic. For example Lou et al. (2011) claim that "since the moving-bottleneck effect always exists, HOT lanes may never be fully utilized." However, full-utilization may be an appropriate idealized reference point for comparison with other alternatives, especially if it can be proven as optimal for certain scenarios – as we show in this paper.

Setting toll values in practice is a challenge in all HOT facilities, particularly if they are determined in real time. For example, the toll rates for I-394 HOT lanes in Minnesota can be adjusted as often as every 3 min. In this operational system, HOT lane density is monitored continuously, and the rate is adjusted according to a predefined lookup table (Halvorson et al., 2006). In contrast to pre-defined lookup tables, various optimization based approaches have been proposed in the literature. Michalaka et al. (2011) developed a scenario-based robust toll optimization model, and used simulation to compare their model results with network performance under a density lookup table. The results showed substantially less toll fluctuations, and better performance for the proposed model. Morgul and Ozbay (2011) proposed a methodology to extend the application of dynamic tolling to two neighboring tolled facilities. A tolling algorithm was developed to calculate toll rates and define route decisions of users of two parallel routes. Microscopic simulation of the Holland and Lincoln tunnels from NJ to NYC was conducted to compare static and dynamic density-based tolls. The simulation results showed that average occupancies decreased 36% for Holland Tunnel and 11% for Lincoln Tunnel, and average speeds increased by 24% and 4%, respectively, as a result of the dynamic pricing case compared with the static pricing case. Ozbay et al. (2011) conducted mesoscopic simulation of bridges in Manhattan, comparing prevailing static tolls with density-based dynamic tolls set by proportional–integral–derivative (PID) controller. Simulation results again support the use of dynamic pricing for reducing congestion, especially at peak periods, when the occupancies at some of the tolled crossings are high. In addition, 16% higher toll revenues would be collected in the hypothetical dynamic pricing scenario compared to the simulated static tolled scenario.

Lou et al. (2011) extended the work by Yin and Lou (2009) and further developed a reactive self-learning approach for determining time-varying tolls in response to the detected traffic arrivals with a more realistic representation of traffic dynamics and an explicit formulation for toll optimization. The approach learns motorists' willingness to pay and determines toll rates to maximize the freeway's throughput while ensuring superior travel service to the users of the toll lanes. Studies of other network configurations have examined time-varying tolls for bottlenecks under conditions in which optimal solutions can be analytically derived (Arnott et al., 1998; Chu, 1995; Liu and McDonald,1999; Yang and Huang,1997; Kuwahara, 2001).

Models for HOT corridor scenario evaluation have two main components: traffic flow and HOT lane choice. As mentioned above, several different approaches have been used for traffic flow representation, including microscopic simulation (e.g. Morgul and Ozbay, 2011) and mesoscopic simulation (e.g. Ozbay et al., 2011). A key dynamic aspect relevant for HOT corridor evaluation is queue accumulation, which can also be represented by analytic models, such as a point queue, single-bottleneck model (e.g. Vickrey, 1963).

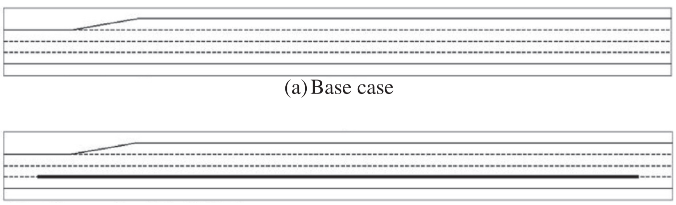
HOT lane choice by individual drivers is represented in most of the above mentioned studies by a logit model (e.g. Lou et al., 2011). The ratio of cost and time parameters in the logit model can be interpreted as a "representative" VOT value. However, VOT heterogeneity in the population is not explicitly addressed by the logit model, an issue that plays a major role in other toll-related research. Yang and Huang (2005, p. 4) state that "The concept of value of time (VOT) plays a pivotal role in road pricing analysis as it describes how users make tradeoffs between cost and time." They describe two main approaches to model VOT heterogeneity: by a discrete set of VOTs for several distinct user classes; or by a continuously distributed VOT. Often VOTs are assumed known, or to follow a known distribution. Leurent (1996) proposed and evaluated a dual criteria assignment model which accounted for a lognormally distributed VOT. The assignment model provides a means to incorporate information about trip makers trade-offs between time and money. Mayet and Hansen (2000) developed a model to determine the optimal toll for a congestible facility under different objective functions based on the assumption of a continuously distributed VOT. The results highlight the role of the objective function in toll setting, specifically the units in which welfare is quantified (monetary or time); however no specific VOT distributions were specified or evaluated.

To implement choice models that consider VOT distribution, an appropriate distribution (or distribution family) should be chosen, and empirical research is needed for the calibration of distribution parameters. Various studies have attempted to infer actual VOTs through the use of stated preference surveys (Lowery et al., 2011) or examination of travel patterns which provide the revealed preferences of commuters (Prince, 2011; Goodall and Smith, 2010; Liu et al., 2007; Liu et al., 2011; Lou et al., 2011). Results are highly variable, with calculated VOT's ranging from \$7–12/h (Lowery et al., 2011) to \$40–60/h, higher than the average \$36/h wage (Prince, 2011). Goodall and Smith (2010) found that the actual proportion of HOT users was determined mainly by time of day, and the impact of toll was rather modest. Cho et al. (2011) discovered limited correlation between time saved and proportion of travelers using HOT. It seems that additional research is needed on VOT distributions among drivers in the context of HOT lane choices.

As a temporary alternative we conjecture that the distribution of VOT values may share similar features with income distributions. Kleiber and Kotz (2003) review several families of distributions used to model income, including the four-parameter general beta distribution of the second kind (GB2), as well as its special case three-parameter Singh-Maddala distribution, a.k.a. Burr XII or simply the Burr distribution. The CDF of the latter is $F(x) = 1 - [1 + (x/b)^a]^{-q}$ (Kleiber and Kotz, 2003; p. 198, Eq. (6.43)). In this paper we use a further simplification two-parameter variant of this distribution, presented by Eq. (13).

3. Model

This section describes the model used for the experiments in this paper, first discussing the traffic flow model, lane choice model, and toll algorithms and then describing the simulation process. The facility is depicted in Fig. 1. Both the GP and HOT lane groups are spatially homogeneous, with bottlenecks at the downstream end of capacity $\bar{q}_{GP}$ and $\bar{q}_{HOT}$. The free-flow time on the two lane groups are denoted τ^0_{GP} and τ^0_{HOT}, both assumed to be integers in the unit system chosen. We assume there are no onramps or offramps in the region of interest to focus our investigation on the issue of the lane choice model. We discretize time into $T+1$ intervals of equal length, and index these intervals with $t \in \{0, 1, \ldots, T\}$, where the index t refers to the time at the start of the t th interval.



(a) Base case

(b) Bottleneck with Managed Lane

Fig. 1. Case study facility. (a) Base case – all lanes are for general purpose (GP); and (b) one lane is managed.

Let V be the set of vehicle classes (for instance, single-occupant vehicles, HOVs, and transit) and $\overline{V} \subseteq V$ the set of vehicle classes which must pay the toll. The median value of time for vehicles of class v is ζ_v. Let d_v^t be the number of vehicles of class v arriving at the GP/HOT diverge point during time interval t. The mean vehicle occupancy for class v is o_v. The toll on the HOT lane in time interval t is c^t. Note that in our model as stated here, all vehicle classes consume the same amount of lane capacity. It is straightforward to generalize this model so that certain classes (such as heavy vehicles or transit) require additional roadway space, but this complicates the notation and, in our opinion, obscures the contributions of this paper.

The following quantities are endogenous: the travel times by lane group over time (τ_{GP}^t and τ_{HOT}^t are the travel times of a vehicle entering the system at the start of time interval t), the proportion of vehicles of class v choosing the HOT lane over time (p_v^t), the average vehicle travel time (*AVTT*), and the average passenger travel time (*APTT*). The latter two are calculated by:

$$AVTT = \frac{1}{\bar{d}} \cdot \sum_t \sum_v d_v^t (p_v^t \tau_{HOT}^t + (1 - p_v^t) \tau_{GP}^t) \tag{1}$$

$$APTT = \frac{1}{\bar{o}} \sum_t \sum_v o_v d_v^t (p_v^t \tau_{HOT}^t + (1 - p_v^t) \tau_{GP}^t) \tag{2}$$

where $\bar{d} = \sum_t \sum_v d_v^t$ is the total number of traveling vehicles and $\bar{o} = \sum_t \sum_v o_v d_v^t$ is the total number of traveling persons (occupants). The toll revenue collected is given by

$$R = \sum_t \sum_{v \in \overline{V}} d_v^t p_v^t c^t \tag{3}$$

3.1. Traffic flow model

Our traffic flow model is simple, consisting of a facility with a single bottleneck at the downstream end, and a higher capacity upstream. We implicitly assume that the total demand never exceeds this upstream capacity – this is reasonable if the freeway immediately upstream of the modeling area has the same characteristics as the section in the model, since the flow across the boundary can never exceed the capacity. We further assume that the bottleneck capacity and study area are large enough that the entering flows are never restricted by a queue at the downstream bottleneck. This model resembles the single-bottleneck pricing models of Vickrey (1963) and others, with the distinction that our model will incorporate variation in value of time, rather than departure time choice.

To describe the evolution of congestion and vehicle flows, we use upstream and downstream cumulative counts: $N_l^{\uparrow}(t)$ and $N_l^{\downarrow}(t)$ are the arrival and departure curves, respectively representing the total number of vehicles that have passed the upstream and downstream end of lane group $l \in \{\mathrm{GP}, \mathrm{HOT}\}$ at the end of time interval t. These values completely define the state of our traffic model. Our model is discrete, so we calculate these cumulative counts only at times $t \in \{0, 1, \ldots, T\}$ and linearly interpolate to obtain intermediate counts. The upstream count equations are:

$$N_{GP}^{\uparrow}(t) = N_{GP}^{\uparrow}(t-1) + \sum_v d_v^t (1 - p_v^t) \tag{4}$$

$$N_{HOT}^{\uparrow}(t) = N_{HOT}^{\uparrow}(t-1) + \sum_v d_v^t p_v^t \tag{5}$$

for $t > 0$ with $N(0) = 0$ for all lane groups. If there were no bottlenecks downstream, we would have $N_l^{\downarrow}(t) = N_l^{\uparrow}(t - \tau_l^0)$; however, the bottleneck constrains the exiting flow in any time interval to be no greater than $\bar{q}_l$. Thus

$$N_l^{\downarrow}(t) = \min\{N_l^{\downarrow}(t-1) + \bar{q}_l, N_l^{\uparrow}(t - \tau_l^0)\} \tag{6}$$

for $l \in \{\mathrm{GP}, \mathrm{HOT}\}$. To reflect system conditions at times not corresponding to one of the discretization points t, linear interpolation is used:

$$N(t) = (t - \lfloor t \rfloor) N(\lfloor t \rfloor) + (\lceil t \rceil - t) N(\lceil t \rceil) \quad \text{if } t \in (0, T) \text{ is not an integer} \tag{7}$$

The same formula applies to both upstream and downstream counts of all lane groups. Using this continuous formula, an inverse function can be meaningfully defined:

$$T(n) = \min\{t : N(t) \geqslant n\} \tag{8}$$

representing the time at which the nth vehicle passes the point where N is measured (omitting subscripts and superscripts for brevity). Since this traffic flow model assumes first-in, first-out ordering, the travel time for a vehicle entering lane group l at time t is

$$\tau_l(t) = T_l^{\downarrow}(N_l^{\uparrow}(t)) - T_l^{\uparrow}(N_l^{\uparrow}(t)) \tag{9}$$

3.2. Lane choice model

The calculation of p_v^t is the major object of study in this paper, and three rules are presented here. The first two (all-or-nothing and additive logit) are discussed briefly, along with reasons we believe them both to be lacking. This leads to discussion of the third rule based on VOT distribution, which we believe to be superior.

All of these rules are based on two quantities: the travel time savings from the HOT lane $\Delta\tau = \tau_{GP} - \tau_{HOT}$ and the toll c for paying vehicles. We seek a mapping $P_v : \Re^2 \to [0,1]$ which gives the proportion of vehicles choosing the HOT lane as a function of the travel time differential and cost: $p_v^t = P_v(\Delta\tau, c)$. Time superscripts are omitted in this section for brevity. When there is a "tie" between the two lane groups (that is, whenever $\zeta_v \Delta\tau = c$ for a demand group v), demand from such vehicle groups splits between the free lanes and HOT lanes. This split is proportional to capacity, after accounting for the vehicles which can use the HOT lane for free (classes in the set $V \setminus \overline{V}$). More precisely, the tiebreaking proportion p^* is

$$P^* = \left[\bar{q}^{HOT} - \sum_{v \in V \setminus \overline{V}} d_v\right] \bigg/ \left(\left[\bar{q}^{HOT} - \sum_{v \in V \setminus \overline{V}} d_v\right] + \bar{q}_{GP}\right) \tag{10}$$

The simplest possible rule is an all-or-nothing assignment

$$P_v^{AON}(\Delta\tau, c) = \begin{cases} 1 & \zeta_v \Delta\tau > c \\ p^* & \zeta_v \Delta\tau = c \\ 0 & \zeta_v \Delta\tau < c \end{cases} \tag{11}$$

assigning all vehicles to the lane group with lower generalized cost. This mapping is far too simplistic: it assumes there is literally no variation in VOT across vehicles of class v, and it is not continuous in $\Delta\tau$ or c, making it highly unlikely that a stable solution will emerge if there is any congestion at all. The one situation where this rule is plausible is with vehicle classes which do not pay the toll – in this case, the model reduces to simply choosing the lane group with lower travel time (which should be the HOT lane).

A simple improvement to the all-or-nothing assignment is a logit model, with the utility of each lane being given by $U_l^v = -\zeta_v \tau_l - c + \varepsilon$, adding independent and identically distributed Gumbel disturbance terms to the generalized cost, and calculating $p_v = \Pr(U_{HOT}^v > U_{GP}^v)$. The resulting mapping is

$$P_v^{LOGIT}(\Delta\tau, c) = \frac{1}{1 + \exp(\theta(c - \zeta_v \Delta\tau))} \tag{12}$$

where θ reflects the inverse of the standard deviation in the disturbance terms. (More details on this and other elementary discrete choice models can be found in standard overviews such as Koppelman and Bhat, 2006). Unlike P_v^{AON}, this mapping is continuous in its parameters. However, it suffers from several deficiencies as well: adding the disturbance term to the generalized cost suggests that the variation in population preferences is due to factors independent of cost and travel time. As shown below, this specification leads to counterintuitive results (if both lane groups are at free flow, single-occupant vehicles will use both lanes even if the HOT lane is tolled).

Instead, we feel that the primary variation in lane choice preferences is due to the distribution of VOT across the population of individual drivers: even within a single vehicle class, different drivers will have different values of time due to demographic factors, trip purposes, and individual heterogeneity. In this case, the proportion of travelers choosing the HOT lane is exactly the proportion of travelers whose VOT exceeds $c/\Delta\tau$. If F_v denotes the cumulative distribution function (CDF) of VOT for vehicle class v, then we have

$$P_v^{VOT}(\Delta\tau, c) = 1 - F_v(c/\Delta\tau) \tag{13}$$

with $F_v(c/\Delta\tau) = 1$ if $\Delta\tau$ is zero and c is nonzero, and $F_v(c/\Delta\tau) = p^*$ if both $\Delta\tau$ and c are zero. Assuming that F_v is continuous and strictly increasing, this mapping is continuous except when $\Delta\tau = c = 0$ (when effectively there is only a single lane group), and has the advantage of directly reflecting variation in VOT across the population. Notice that if the toll is not zero and the travel times of both lanes are equal, no toll-paying travelers will choose the HOT lane, alleviating the problem seen with the simple logit model.

The remaining question is which VOT distribution to use. For the purposes of this paper, this is not a critical decision. Following the discussion in Section 2, we use the following simplified variant of the Burr distribution:

$$F\left(\frac{c}{\Delta\tau}; \zeta, \gamma\right) = 1 - \frac{1}{\left[1 + \left(\frac{c}{\zeta \cdot \Delta\tau}\right)^{\gamma}\right]} \tag{14}$$

where γ is a shape parameter affecting the relative width of the VOT distribution. Eq. (14) can be inverted to obtain the cost to travel-time-savings ratio corresponding to a certain proportion of travelers using the HOT lane:

$$\frac{c}{\Delta\tau} = F^{-1}(1 - p; \zeta, \gamma) = \zeta\left(\frac{1}{p} - 1\right)^{1/\gamma} \tag{15}$$

Using the United States income distribution in 2008, the 25th percentile to median ratio suggests $\gamma \approx 1.5$, and the 75th percentile to median ratio suggests $\gamma \approx 2$. The experiments reported here show results for both values. Fig. 2a illustrates the *probability* of a traveler choosing the HOT lane and Fig. 2b illustrates the *proportion* of travelers choosing the HOT lane, relative to the cost to travel-time-savings ratio. The curves are representative of the PDF and CDF of the simplified version of the Burr distribution presented in Eq. (14) for $\gamma = 1.5$ and $\gamma = 2$.

Fig. 3 illustrates the fundamental difference between P_v^{LOGIT} and P_v^{VOT} graphically, showing contour plots of p in $(c, \Delta\tau)$ space. The lower-right panel (p_4) representsP_v^{VOT}, while the other panels represent P_v^{LOGIT} with different specifications of the logit parameters. The diagonal proportion contour line in p_1 is reasonable, suggesting that half of drivers will use the HOT lane if the travel time difference is equivalent to the toll in terms of median VOT. However, the diagram also says that even with zero travel time difference, p declines very gradually with c. It also implies that with zero toll, p increases only gradually as $\Delta\tau$ increases. Calibrating the parameters of the logit function does not change anything fundamental. Adding a constant (mode bias) in p_2 shifts the diagram to the right, but we still get the same problematic parallel lines, and now observed VOT does not meet the median VOT even for the proportion of 0.5. Changing the cost sensitivity in p_3 changes the density of the lines, but otherwise still nothing much has changed. By contrast, if the most critical stochastic feature is VOT variability among travelers, the cost and time observations at a given facility for a constant toll with a given capacity ratio (and therefore a given HOT proportion) should correspond to the VOT of the same proportion in the population. This immediately suggests that the contour lines of the choice function should form a fan of rays, as in p_4 (the P_v^{VOT} model).

3.3. Toll algorithms

In this paper, we compare two toll-setting methods: a constant toll, and a time-varying toll adjusted to maximize inflow to the HOT lane without exceeding its bottleneck capacity (the "full utilization" algorithm), taking as input the travel time

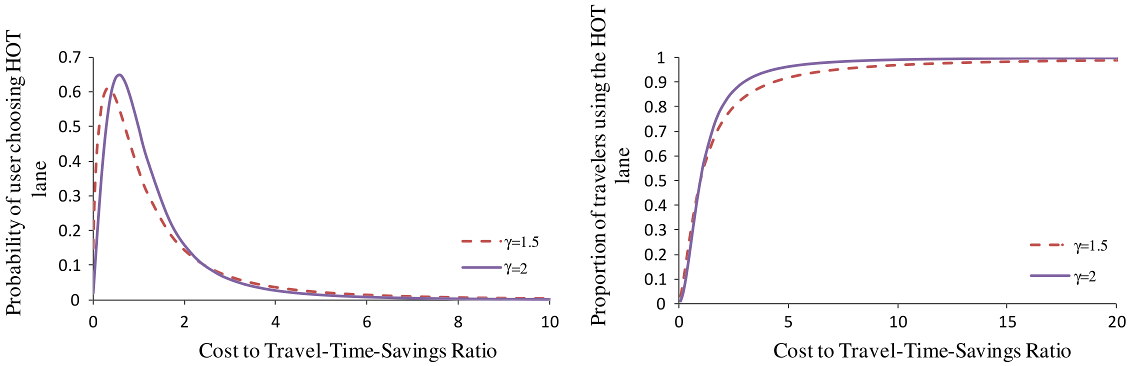


Fig. 2. Burr distributed value of time. (a) Probability density function (PDF); (b) cumulative distribution function (CDF).

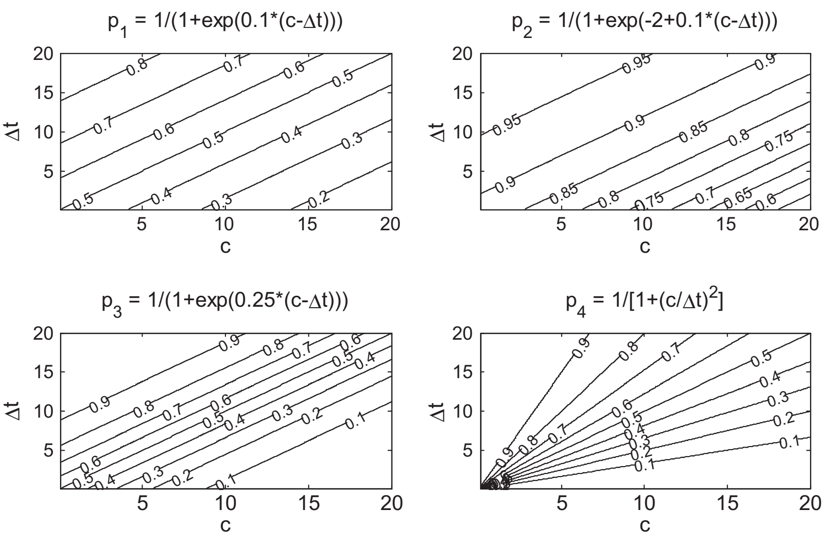


Fig. 3. Choice probability contour plots for additive logit (p_1, p_2 and p_3) and for Burr VOT distribution (p_4).

difference between the lanes, the inflow rates by vehicle class, and the VOT distributions by user class. In a deterministic setting, the full utilization algorithm minimizes the total occupancy-weighted travel time: any solution that does not fully utilize the capacity of the HOT lane cannot be optimal, because it is possible to move some LOVs from the GP to the HOT lane, while keeping the travel time of the HOT lane at free flow. This means that vehicles that previously used the HOT lane are not affected, and all others gain.

In all the solutions that fully utilize the capacity of all lanes, the total duration of congestion will be the same, and the non-weighted AVTT will be the same. It is just a queue, and if some get ahead in the queue their gain is identical to the loss of others. In terms of occupancy-weighted APTT, it is clear that any switch allowing an HOV to get a head in the queue at the expense of an LOV improves the outcome. So a solution where HOVs travel at free flow is optimal. To find the toll corresponding to full utilization (hereby referred to as FU toll), we solve the following equation for c:

$$\sum_{v\in\overline{V}} d_v P_v^{VOT}(\Delta\tau, c) + \sum_{v\in V\setminus\overline{V}} d_v = \min\left\{\bar{q}_{HOT}, \sum_v d_v\right\} \tag{16}$$

In this way, the inflow to the HOT lane, represented by the left-hand side of (16), is the highest possible value for the given demand level without exceeding bottleneck capacity, represented by the right-hand side. As the left-hand side is continuous and strictly decreasing in c, a unique solution is guaranteed, assuming $\sum_{v\in V\setminus\overline{V}} d_v < \bar{q}_{HOT}$ (which is logical, since otherwise the lane could be fully used as an HOV lane without adding a toll) and $\Delta\tau > 0$ (otherwise, there is no congestion). Time superscripts are omitted for clarity, and this equation is solved at the start of each time interval to determine the dynamic toll for that period. For the special case when only a single class of vehicles $\bar{v}$ must pay the toll, and when this class has a VOT distribution described by the Burr distribution, the solution to this equation can be expressed analytically by:

$$c = \Delta\tau \cdot F^{-1}\left(1 - \frac{\min\{\bar{q}_{HOT}, \sum_v d_v\} - \sum_{v\neq\bar{v}} d_v}{d_{\bar{v}}}\right) = \Delta\tau \cdot \zeta \cdot \left(\frac{d_{\bar{v}}}{\min\{\bar{q}_{HOT}, \sum_v d_v\} - \sum_{v\neq\bar{v}} d_v} - 1\right)^{1/\gamma} \tag{17}$$

where $P_v^{VOT}(\Delta\tau, c) = 1 - F_v(c/\Delta\tau)$, and the right most expression is obtained by substituting Eq. (15). This is the formula used in the experiments reported in this paper.

3.4. Simulation algorithm

The system described above is implemented in a simulation program written in C. This simulation performs the following steps:

1. Initialize: set $t = 0$, travel times on GP and HOT lanes to free flow, and all N values to zero.
2. Calculate toll c (either constant or based on the full-utilization formula (17)).
3. Calculate lane choice probability p_v^t.
4. Using one of the lane choice mappings described by (11), (12) or (13).
5. Propagate flow, updating upstream counts using (4) and (5), and updating downstream counts using (6).
6. Update statistics (lane group travel time, revenue, and other metrics)
7. If $t < T$, increment t and return to step 2. Otherwise, terminate.

Note that our model recalculates the toll at every time step. Thus, the time step should be adjusted based on the temporal resolution of input data, the desired response rate, and any regulatory constraints on how frequently the toll can be changed. In our numerical analyses the time step is set to 1 min.

4. Numerical analysis

The results presented in this section are based on the facility depicted in Fig. 1. The HOT lane has a capacity of 1800 vph and each GP lane has a capacity of 2100 vph. The length is 10 km and the free flow speed is 100 km/h. The demand profile (provided in Table 1) is chosen such that the LOV demand exceeds the GP lane capacity and a queue is formed at the bottleneck.

Table 1
Peak period travel demand profile.

	Average occ.	7:00–8:00	8:00–9:00	9:00–10:00
LOV	1.2	6300	5100	3900
HOV	4	600	600	600
Transit	40	300	300	300
Total		7200	6000	4800

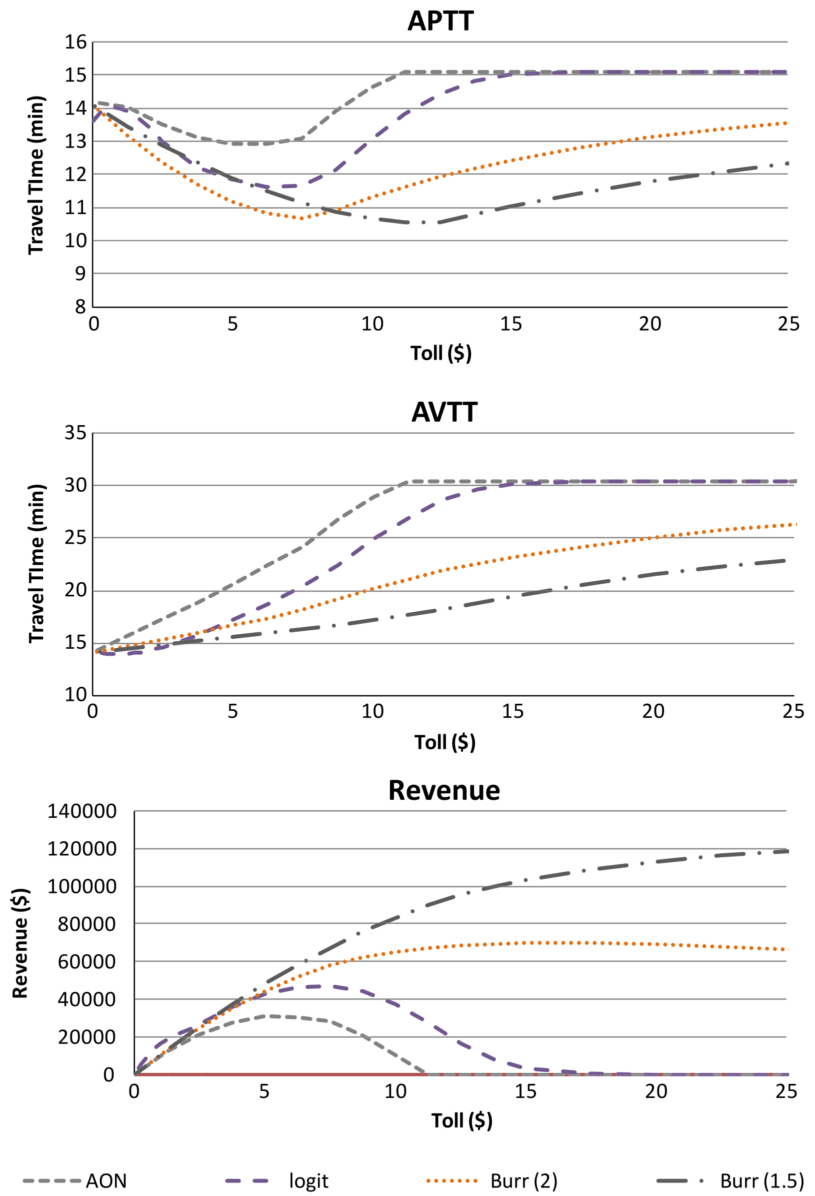


Fig. 4. APTT, AVTT, and revenue as a function of fixed toll value.

The facility performance is evaluated based on three measures: APTT, AVTT, and revenue, using Eqs. (1)–(3). Results are provided for the base case where all lanes are GP, the HOV case where only HOV and transit users use the managed lane, and each of the lane choice models considered under the HOT case. In addition, a more detailed level of analysis is provided to depict lane choice behavior and lane properties temporally over the entire peak period, including entering flow volumes, percent of LOVs choosing to take the HOT lane, and travel time differences between the HOT and GP lane as a function of time.

The parameters for the analysis presented are: median VOT of 15$/h, equivalent to 0.5$ per 2 min interval used by Lou et al. (2011); the logit parameter is set at $\theta = 0.25$, to obtain comparable results with the Burr model.[1] The Burr distribution was evaluated for both $\gamma = 1.5$ and $\gamma = 2$.

4.1. Evaluation of fixed tolls by different lane choice models

Fig. 4 illustrates the three performance metrics as a function of the fixed toll value for the AON, logit and VOT lane choice models. The APTT and AVTT for the base case are both 14 min, and the APTT and AVTT for the HOV case are 15.1 and

[1] At 20 min travel time saving (typical for our scenarios), 15% of travelers choose the HOT lane at a toll of 11.9$ and 15.9$ using the Burr model with $\gamma = 1.5$ and $\gamma = 2$ respectively; and at a toll of 11.9$ using the Logit model with $\theta = 0.25$.

30.4 min, respectively. The managed lane in underused in the HOV case, resulting in an 8% increase in APTT and a 117% increase in AVTT, relative to the base case.

In Fig. 4 similar APTT and AVTT behavior under the VOT lane choice model is illustrated for both Burr parameters, with the main difference being a higher optimal fixed toll value for $\gamma = 1.5$. The AVTT increases as the fixed toll value increases for all models, though at a slower rate for the VOT model. For a high enough toll value the AVTT converges with the HOV case under all models, at which point no LOV users will opt to pay. The maximum toll value some users are willing to pay is highest for the VOT model and lowest for the AON model.

Under the AON and VOT models, a toll of zero results in the base case AVTT. It should be mentioned that with zero toll the logit model predicts that more LOVs will use the HOT than required to equalize travel time, as a result GP travel time is shorter than the HOT travel time, and all HOVs use the GP lane. This unrealistic behavior is a direct result of the fundamental structure of the additive logit choice model. A related fundamental counter-intuitive anomaly is the upward trend of APTT at low toll values with the logit choice model, as shown in Fig. 4.

4.2. Evaluation of fixed and full-utilization tolls by a VOT choice model

For the full-utilization (FU) case the tolls vary dynamically over the course of the peak travel period, and are calculated as described in Section 3.4. For both $\gamma = 1.5$ and $\gamma = 2$, FU toll values follow a similar pattern over the peak period (Fig. 5). The wider distribution of VOT when $\gamma = 1.5$ results in an inflated toll value. The impact of these inflated tolls is evident in the increased facility revenue for the case when $\gamma = 1.5$; the revenue for the FU case with $\gamma = 1.5$ is \$68,030, while only \$43,053 when $\gamma = 2$. The APTT and AVTT for the FU cases are 9 min and 14 min, respectively. The AVTT for both Burr

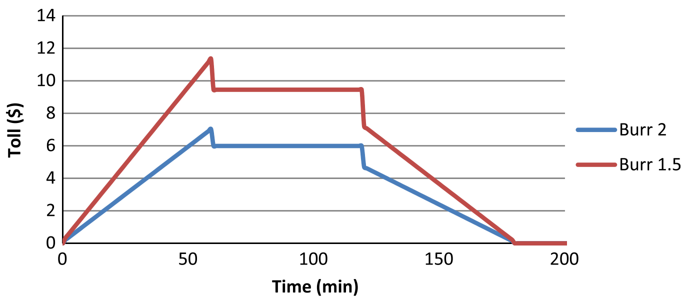


Fig. 5. FU tolls over the peak period.

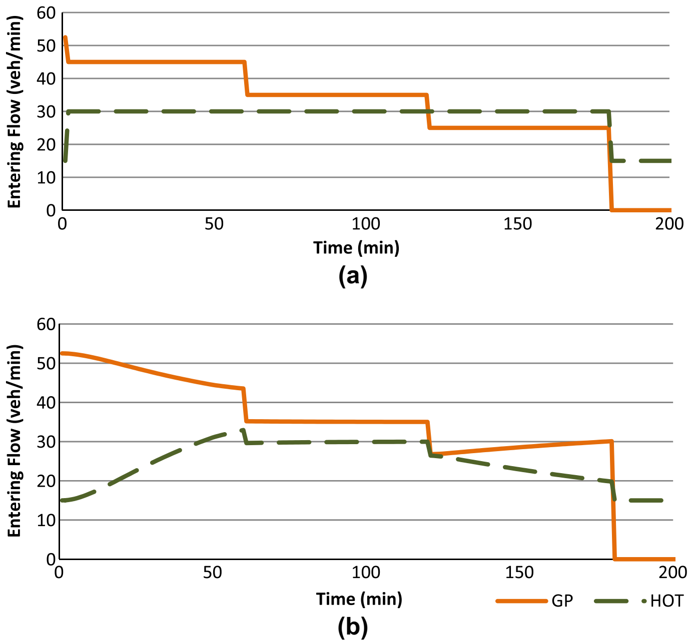


Fig. 6. Entering Flow per Lane for (a) FU tolls and (b) fixed toll = \$7.50 during the peak period, assuming VOT distribution is Burr with $\gamma = 2$.

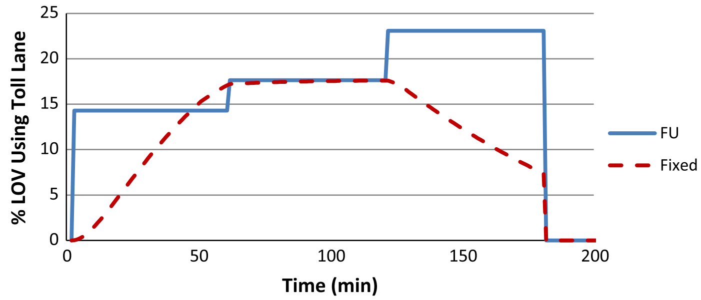


Fig. 7. Percent of LOVs choosing HOT lane during peak period.

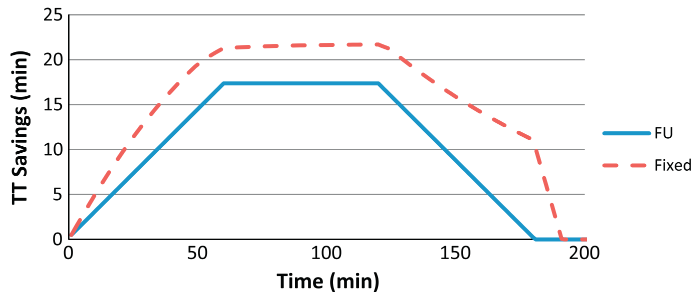


Fig. 8. Travel time savings by taking HOT lane during the peak period.

parameters under FU tolls is equal to the base case AVTT, which makes the most optimal use of the facility in terms of vehicle travel. The FU tolls also result in the same APTTs for both Burr parameters, and a 33% savings relative to the base case, because the FU tolls maximize the entering HOT lane volume subject to its capacity, as shown in Fig. 6a. Additionally, based on Fig. 4, a static toll has the potential to save a significant portion of the average person travel time savings achieved by a dynamic FU toll (i.e. the difference between the base case and the full utilization toll). For example, a static toll of \$7.5 saves 3.3 min out of a potential APTT saving of 5 min, i.e. two thirds of the benefit.

For the remainder of the analysis presented the Burr parameter is set at $\gamma = 2$ and the fixed toll is set to \$7.50, which is the fixed toll resulting in the lowest APTT for this lane choice map. Fig. 6 depicts total volume entering each lane throughout the peak period for the VOT model with FU and fixed tolls. For both cases, the maximum volume entering the HOT lane is equal to the lane capacity of the HOT lane, 1800 vph, i.e. 30 vpm. Based on the demand profile, the HOV inflow rate is 900 vph, therefore the remaining 900 vph is made up of toll-payers.

The increase in percent of LOVs using the HOT lane (see Fig. 7) is due to a combined impact of decreasing demand over the peak period, and the FU toll objective of keeping the HOT lane at capacity. The increase in LOV usage occurs despite a decrease in travel time savings, because of the falling toll. For fixed toll this is not the case, and the percent of LOV users follows a similar profile to the travel time savings. The travel time savings are greater for the fixed toll, meaning a larger discrepancy in lane performance between the HOT and GP lanes, therefore a less equitable tolling policy.

Fig. 7 also indicates that 17% of LOV users pay to use the HOT lane during the middle of the peak period under both tolling schemes. Fig. 8 illustrates travel time savings achieved by taking the HOT lane. Under a fixed toll of \$7.50 users save an average of 21.6 min whereas the FU toll value is \$6 over this same time interval and users save an average of 17 min. Both tolling schemes equate to a willingness to pay \$0.35/min of travel time savings, or \$21/h which is 40% higher than the median VOT of \$15/h assumed in this work. The apparent discrepancy between "observed" VOT and the median VOT in the population is due to the fact that the proportion of LOV users choosing the HOT lane is 17% rather than 50%.

5. Conclusions

This paper studied the impact of various lane choice models and toll algorithms in a deterministic context. Based on this investigation, we believe that the lane choice model should be directly linked to the VOT distribution in the population: it is a continuous mapping, capturing the essential stochastic feature in the model. Numerical experiments comparing the proposed choice model to other alternatives demonstrate its superior plausibility. We also showed that the modeling framework enables the evaluation of "full utilization" tolls and fixed tolls in comparison to the reference options of regular HOV lane, and base case (no managed lanes).

The main limitations of this study should be recognized: we considered a simple facility with only a single downstream bottleneck, deterministic demand, and a known VOT distribution which is perfectly reflected in the demand at all times. This study can serve as a foundation for subsequent research using similar modeling methodology and evaluation approach, while relaxing the above mentioned assumptions in order to generalize the applicability of the model and to identify the relative significance of these factors. In particular, it can be expected that under stochastic conditions full utilization of the HOT lane may not necessarily be the optimal strategy, and a tolerance for disruptions would be beneficial. Adding operational considerations, and particularly traffic monitoring possibilities, is another important topic for future research. Other useful expansions are to include travel time reliability, to address more sophisticated networks involving on-ramps and off-ramps, and to study the sensitivity of the model to errors in the input data. Finally, empirical estimation of VOT distributions among drivers in the context of HOT-lane usage is critical for the improvement of modeling and operations of HOT facilities.

References

Arnott, R., de Palma, A., Lindsey, R., 1998. Recent developments in the bottleneck model. In: Button, Kenneth J., Verhoef, Erik T. (Eds.), Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility, pp. 79–110.

Bar-Gera, H., 2012. The fast lane to Tel-Aviv: high-occupancy toll project with a Pareto package. In: Proceedings of the 91st Annual Meeting of the Transportation Research Board, Washington, DC, No. 0712.

Borrell-Rovira, A., Supernak, J., 2011. Peak-period utilization of I-15 corridor in San Diego: long term impact of the Fastrak Program. In: Proceedings of the 90th Annual Meeting of the Transportation Research Board, Washington, DC, No. 3751.

Burris, M.W., Ungemah, D.H., Mahlawat, M., Pannu, M.S., 2009. Investigating the Impact of Tolls on High-Occupancy-Vehicle Lanes Using Managed Lanes. Transportation Research Record: Journal of the Transportation Research Board, No. 2099, 113–122.

Burris, M.W., Geiselbrecht, T.C., Goodin, G.D., MacGregor, M.E., 2011. Predetermining performance-based measures for managed lanes. In: Proceedings of the 90th Annual Meeting of the Transportation Research Board, Washington, DC, No. 1482.

Cao, X., Xu, Z., Huang, A., 2011. Safety benefits of converting HOV lanes to HOT lanes: case study of I-394 MNPass. In: Proceedings of the 90th Annual Meeting of the Transportation Research Board, Washington, DC, No. 0361.

Cho, Y., Goel, R., Gupta, P., Bogonko, G., Burris, M.W., 2011. What are I-394 HOT lane drivers paying for? In: Proceedings of the 90th Annual Meeting of the Transportation Research Board, Washington, DC.

Chu, X., 1995. Endogenous trip scheduling: the Henderson approach reformulated and compared with the Vickrey approach. Journal of Urban Economics 37 (3), 324–343.

Chung, Y., Choi, K., 2010. Dynamic toll concept to assess feasibility of high-occupancy vehicle lane on Kyungbu Freeway, South Korea. Transportation Research Record: Journal of the Transportation Research Board, No. 2178, 162–169.

Chung, C., Recker, W., 2011. State-of-the-art assessment of toll rates for high-occupancy and toll lanes. In: Proceedings of the 90th Annual Meeting of the Transportation Research Board, Washington, DC, No. 0496.

Dahlgren, J., 2002. High-occupancy/toll lanes: where should they be implemented? Transportation Research Part A 36 (2002), 239–255.

Devarasetty, P., Burris, M., Shaw, W., 2012. The value of travel time and reliability – evidence from a stated preference survey and actual usage. Transportation Research Part A 46 (8), 1227–1240.

Fosgerau, M., 2011. How a fast lane may replace a congestion toll. Transportation Research Part B 45 (6), 845–851.

Goel, R., Burris, M.W., 2011. HOT-lane policies and their implications. In: Proceedings of the 90th Annual Meeting of the Transportation Research Board, Washington, DC, No. 0270.

Goodall, L., Smith, B.L., 2010. What drives decisions of single-occupant travelers in high-occupancy vehicle lanes? Investigation using archived traffic and tolling data from MnPASS express lanes. Transportation Research Record: Journal of the Transportation Research Board, No. 2178, 156–161.

Gordon, P., Moore II, J.E., Richardson, H.W., Williamson, C., 2004. The Economic Effects of Highway Widening: Tolled Lanes vs. General-Purpose Lanes – Using an Integrated Impact Model. Report FHWA/CA/OR-2005/12. <http://ntl.bts.gov/lib/24000/24900/24922/01-14_P_Gordon_Final.pdf>.

Halvorson, R., Nookala, M., Buckeye, K.R., 2006. High occupancy toll lane innovations: I-394 MnPASS. Public Works Management & Policy 10 (1), 242–255.

Kleiber, C., Kotz, S., 2003. Statistical Size Distributions in Economics and Actuarial Sciences. Wiley. ISBN 0-471-15064-9.

Koppelman, F., Bhat, C., 2006. A Self-Instructing Course in Mode Choice Models: Multinomial and Nested Logit Models. Prepared for U.S. Department of Transportation, Federal Transit Administration. <http://www.ce.utexas.edu/prof/bhat/COURSES/LM_Draft_060131Final-060630.pdf> (accessed 10.03.13).

Kuhn, B., Goodin, G., Ballard, A., Brewer, M., Brydia, R., Carson, J., Chrysler, S., Collier, T., Eisele, W., Fenno, D., Fitzpatrick, K., Ullman, B., Ullman, G., Venglar, S., 2005. Findings from Texas: Five Years of Research on Managed Lanes. Report Number: FHWA/TX-06/0-4160-25. <http://tti.tamu.edu/documents/0-4160-25.pdf> (accessed 21.07.11).

Kuwahara, M., 2001. A theoretical analysis on dynamic marginal cost pricing. In: Proceedings of the Sixth Conference of Hong Kong Society for Transportation Studies, pp. 28–39.

Leurent, F., 1996. The theory and practice of a dual criteria assignment model with a continuously distributed value-of-time. In: Lesort, J.B. (Ed.), Transportation and Traffic Theory. Elsevier Science, pp. 455–477.

Liu, L.N., McDonald, J.F., 1999. Economic efficiency of second-best congestion pricing schemes in urban highway systems. Transportation Research Part B 33 (3), 157–188.

Liu, H.X., He, X., Recker, W., 2007. Estimation of the time-dependency of values of travel time and its reliability from loop detector data. Transportation Research Part B 41 (4), 448–461.

Liu, X., Zhang, G., Lao, Y., Wang, Y., 2011. Quantifying the attractiveness of high occupancy toll (HOT) lane under various traffic conditions using traffic sensor data. In: Proceedings of the 90thAnnual Meeting of the Transportation Research Board, Washington, DC, No. 2882.

Lou, Y., Yin, Y., Laval, J., 2011. Optimal dynamic pricing strategies for high-occupancy/toll lanes. Transportation Research Part C 19 (1), 64–74.

Lowery, J., Burris, M.W., Geiselbrecht, T., Goodin, G., 2011. Reaction to value pricing by different suburban populations. In: Proceedings of the 90th Annual Meeting of the Transportation Research Board, Washington, DC, No. 0122.

Mayet, J., Hansen, M., 2000. Congestion pricing with continuously distributed values of time. Journal of Transport Economics and Policy 34 (3), 359–370.

Michalaka, D., Lou, Y., Yin, Y., 2011. Proactive and robust dynamic pricing strategies for high-occupancy–toll (HOT) lanes. In: Proceedings of the 90th Annual Meeting of the Transportation Research Board, Washington, DC, No. 2617.

Morgul, E.F., Ozbay, K., 2011. Simulation-based evaluation of a feedback based dynamic congestion pricing strategy on alternate facilities. In: Proceedings of the 90th Annual Meeting of the Transportation Research Board, Washington, DC, No. 3535.

Munnich, L.W., 2006. Easing the commute: analysis shows MnPASS system curbing congestion on Interstate 394. Downtown Journal/Skyway News. <http://www.downtownjournal.com/articles/2006/07/24/news/news04.txt> (24.07.06).

Ozbay, K., Morgul, E.F., Ukkusuri, S., Iyer, S., Yushimito, W.F., 2011. Mesoscopic simulation evaluation of dynamic congestion pricing strategies for New York City crossings. In: Proceedings of the 90th Annual Meeting of the Transportation Research Board, Washington, DC, No. 3612.

Parkany, E., 1998. Can HOT Lanes Encourage Carpooling? A Case Study of Carpooling Behavior on the 91 Express Lanes. UCI-ITS-WP-98-5, <http://www.its.uci.edu/its/publications/papers/ITS/UCI-ITS-WP-98-5.pdf> (accessed 24.07.11).

Prince, R.J., 2011. SR-91 Express lanes: understanding user travel patterns from a behavioral perspective. In: Proceedings of the 90th Annual Meeting of the Transportation Research Board, Washington, DC, No. 1883.
South California Association of Governors (SCAG), 2004. Regional South-California High-Occupancy Vehicle Lane System Performance Study. <http://www.scag.ca.gov/publications/pdf/2005/HOVStudy_Final1104.pdf> (accessed 24.07.11).
Vickrey, W.S., 1963. Congestion theory and transport investment. American Economic Review 59 (2), 251–260.
Yang, H., Huang, H.J., 1997. Analysis of the time-varying pricing of a bottleneck with elastic demand using optimal control theory. Transportation Research Part B 31 (6), 425–440.
Yang, H., Huang, H.J., 2005. Mathematical and Economic Theory of Road Pricing. Elsevier, Amsterdam.
Yin, Y., Lou, Y., 2009. Dynamic tolling strategies for managed lanes. ASCE Journal of Transportation Engineering 135 (2), 45–52.