High Occupancy Toll lane Performance Under Alternative Pricing Policies

by Thomas Light

This paper explores how alternative pricing and operating policies influence revenue generation, level of service, and travel time costs for high occupancy toll (HOT) lane facilities. A framework for modeling HOT lanes is applied to a hypothetical facility. The analysis suggests that the way in which tolls are set can have a non-trivial influence on competing measures of HOT lane performance. Other operating characteristics, such as the number of lanes designated as free and priced and whether carpools are allowed to ride free or must pay a toll to access the HOT lanes, are shown to significantly influence performance as well.

INTRODUCTION

High occupancy toll (HOT) lanes are receiving increased attention in the United States, in part because of their potential to better utilize scarce road capacity, but also because they raise revenue that can support investments in transportation systems. Versions of the HOT lane concept have been implemented on a number of highways, including Interstate 15 (I-15) in San Diego, California; Interstate 394 (I-394) near Minneapolis, Minnesota; State Route 167 (SR-167) near Seattle, Washington; and State Route 91 (SR-91) in Orange County, California. These facilities allow high occupancy vehicles (HOVs) to utilize the HOT lanes for free (or at a discount) and single occupant vehicles (SOVs) to access the HOT lanes by paying tolls that potentially vary by time of day and day of the week or with real-time traffic conditions. SOVs which opt not to pay the tolls can use free general purpose (GP) lanes that run parallel to the HOT lanes.

A review of existing HOT lane facilities suggests that their performance can vary widely, particularly in terms of their abilities to generate revenue. Factors that can influence the revenue generated by these facilities include the level of congestion and users' willingness to pay, as well as design features such as the share of capacity that is HOT, the way carpools are treated for the purpose of pricing, how tolls are determined, and whether or not the facility is priced dynamically or according to a static toll schedule. This paper seeks to understand how some of these design features influence the performance of HOT lane facilities with emphasis on the implications of setting tolls to achieve different goals (i.e., maximize revenue or maintain a target level of service). To do this, a simple but relatively flexible HOT lane model is presented and examined numerically.

The model considers a highway during a peak period in one direction with a fixed number of HOVs and SOVs using it. Users of the vehicles in the HOV and SOV classes vary in their value of travel time. In the model, as long as the HOT lanes offer travel time savings, carpools will use them since there is no access cost. Drivers of SOVs, however, choose between the GP and HOT lanes, with those having the highest value of time opting to pay the toll to access the HOT lanes.

This paper builds on Light (2009) where motorists vary in their value of time and choose between a priced and free route but only one vehicle class is considered. Toll and capacity policies, as well as the fiscal implications of pricing the facility to maximize welfare are explored and characterized. The distributional implications of adopting pricing on one route are also explored. Two clear results arise for the case when road capacity is the same before and after pricing is implemented and the transportation authority's budget is balanced via a uniform tax paid by all travelers: (1) value pricing will make some users worse off and (2) the user who is left worst off is the one who is indifferent between using the priced and free alternative. Unlike Light's (2009) previous research, this analysis incorporates carpools into the model and explores the implications of a larger set of pricing objectives. The findings and methodology of this paper are likely to interest practitioners involved in evaluating HOT lanes. In particular, the model is relatively simple to set up and solve and relies on a limited set of assumptions. Because it is capable of predicting conditions under a broad set of HOT lane policies, it is appropriate for sketch planning revenue estimation and operational analysis.

The results of the numerical example suggest some generalizations about the efficiency and revenue potentials of HOT lanes that do not depend on facility design features, including the ways the tolls are set. Consistent with other research (Small and Yan 2001, and Verhoef and Small 2004), it is found that tolls that maximize revenue are higher than those that minimize aggregate travel time cost and produce large speed differences between GP and HOT lanes, all else equal. Additionally, it is found that tolls that are set to maintain a target level of service in the HOT lanes may be set above or below the toll that meets other objectives, depending on corridor demand and other factors.

The paper proceeds as follows. The next section discusses the different motivations for adopting HOT lanes in the United States, and describes four existing HOT lane facilities. Next, the theoretical model is presented followed by the results of the numerical analysis.

HOT LANES IN THEORY AND PRACTICE

Economists have focused on the potential ability for road pricing to improve the efficiency of road networks by causing motorists to internalize the externalities they impose on other motorists and the public at large. The delay costs imposed on others from a vehicle's contribution to congestion is the primary external cost discussed in the congestion pricing literature. However, other external costs such as those associated with accidents, emissions, and noise can be incorporated into congestion pricing. HOT lanes represent a relatively dull instrument for inducing the internalization of vehicle externalities since, in most applications, relatively few travelers in HOT lane corridors actually pay a toll. Small et al. (2006) have noted that HOT lanes have the potential to improve the efficiency of urban highways by making better use of underutilized carpool lanes and providing faster and more reliable travel to those who value it most.

In practice, however, the recent attention given to HOT lanes is more likely due to their ability to generate revenue. As the costs of transportation improvements have escalated over time, funding for transportation has not kept up the pace. Many metropolitan areas are now looking for new ways to fund transportation improvements, including the implementation of HOT lanes. As an example, the San Francisco Bay Area is looking into converting the existing regional carpool lanes network into a network of HOT lanes with the revenues generated to be used to expand the HOT lane network in the future (Metropolitan Transportation Commission 2008).

In addition to these motivating reasons for adopting HOT lanes, the HOT lane concept is attractive because it continues to encourage carpooling by granting carpools preferential access to a faster travel option. This is particularly important given the recent evidence that HOV lanes may not be an efficient means of reducing traffic congestions (Dahlgren 1998, Poole and Balaker 2005), making HOT lanes a more appealing alternative. In order to encourage carpool formation, it has been suggested that tolls should be set high enough to maintain a high minimum level of service in the HOT lanes. Furthermore, in California and other states, there can be legal restrictions which place level of service requirements on HOT lanes. For instance, the legislation enabling the development of the I-15 HOT lane in San Diego requires that a level of service B or Caltrans HOV lane standard be maintained in the I-15 HOT lanes (Turnbull 1997). This requires speeds in the I-15 HOT lanes to meet or exceed 45 miles per hour.

Small and Yan (2001) extend a model first developed by Liu and McDonald (1998) to account for user heterogeneity on a facility resembling SR-91. They highlight the importance of accounting for user heterogeneity in preferences for travel time savings when determining the benefits of HOT lane pricing. Small and Yan (2001) simulate both welfare maximizing and revenue maximizing tolls for facilities that have priced lanes running parallel to free GP lanes. Their findings suggest that revenue maximization is likely to produce significantly higher tolls than welfare maximization and may result in poorer outcomes than if the priced lanes were converted to free GP lanes.

Sullivan and Burris (2006) conduct a benefit cost analysis for the SR-91 Express Lane facility. Their analysis suggests that the facility produces benefits that are greater than the amortized investment and operating costs. Their analysis also indicates that the majority of benefits from the SR-91 Express Lane facility come about through travel time savings as opposed to reductions in fuel use or emissions. Safirova et al. (2007) studied the impacts of converting existing HOV lanes to HOT lanes in Washington, D.C.[1] Their analysis assumes that tolls are set to maximize aggregate social welfare. Their findings suggest that while all income groups benefit from the conversion to HOT lanes, higher-income households tend to see the greatest welfare gains. Small et al. (2006) studied a variety of HOT lane pricing policies using a model estimated from data collected for SR-91 in Orange County, California. Similar to Safirova et al. (2007), they find that the highest-income users of the corridor tend to benefit more than lower-income users when HOT lane pricing is adopted.

Design and Performance of Four HOT Lane Facilities

Because of the many appealing features of HOT lanes, there are numerous examples of their use, four of which are briefly discussed below.[2]

Interstate 15 (San Diego, California). The I-15 HOT lanes consist of two reversible lanes that are approximately eight miles in length. Construction is underway to extend the facility and add two additional HOT lanes. Currently, the toll for SOVs varies dynamically with congestion in the HOT lanes and is updated every six minutes. Message signs near the access point to the HOT lanes indicate the charge, which typically varies between $0.50 and $4.00, although it may be as high as $8.00 in periods of high demand. Carpools with two or more passengers can use the I-15 HOT lanes for free.

Interstate 394 (Minneapolis, Minnesota). The I-394 MnPASS facility allows vehicles with two or more persons to travel free while SOVs may access the HOT lanes if they pay a toll. The facility is dynamically priced. A three-mile stretch of it consists of two reversible lanes that accommodate peak direction traffic, while the remainder consists of one lane in each direction.

State Route 167 (Seattle, Washington). The SR-167 HOT lane facility opened in 2008 and is located near Seattle, Washington. The SR-167 HOT lanes allow vehicles with two or more occupants (HOV 2+) to use it free while SOV must pay a toll to access the lanes. The facility is dynamically priced every five minutes using real-time traffic data on speed and volumes, with tolls varying between $0.50 and $9.00. Motorists who wish not to use the HOT lane can use two parallel GP lanes that operate in each direction.

State Route 91 (Orange County, California). The SR-91 Express Lane facility was constructed as part of a private, for profit toll road venture authorized by the California Legislature in 1989. However, the Orange County Transportation Authority subsequently took control of the facility after its first three years of operation. While all vehicles are charged a toll during peak periods on the 10-mile State Route 91 Express Lane facility, vehicles with three or more occupants receive a 50% discount on the toll charge. The toll varies by time of day and day of the week according to a fixed schedule.

The I-15, I-394, and SR-167 facilities all operate under an HOV 2+ policy and utilize dynamic pricing algorithms to adjust toll levels based on real-time traffic conditions. Because the tolling algorithms are proprietary, it is unclear how they work and what they are designed to achieve. However, it is known that their algorithms have minimum and maximum tolls as well as a mechanism to adjust the toll so that specified HOT lane level-of-service requirements are not breached. SR-91 is unique in that it charges all vehicles during peak hours and uses a static toll schedule, which is updated every few months.

Perhaps the most striking performance feature of these examples is the range of revenue generation. Despite all four facilities being of fairly similar length, Table 1 suggests SR-91 is much more profitable than the others, with over $43 million in gross toll revenues generated during 2009.[3] I-15 and I-394 generate revenues of approximately $1 million annually while SR-167 produced only $0.3 million in its initial year of operations. The differences in revenue are at least partly due to the treatment of carpool vehicles for the purposes of tolling, as well as the number of priced lanes, the level of congestion in the corridor, and differences in willingness-to-pay of motorists. The model presented in the next section can potentially accommodate variation in these facility features as well as different approaches in setting toll levels.

Table 1: Description and Performance of Four Facilities

	I-15 (San Diego, CA)	**I-394 (Minneapolis, MN)**	**SR-167 (Seattle, WA)**	**SR-91 (Orange County, CA)**
Year Opened	1997	2005	2008	1995
Distance	8 miles	11 miles	11 miles northbound/ 9 miles southbound	10 miles
Number of GP Lanes	4 lanes in each direction	2 lanes in each direction	2 lanes in each direction	4/5 lanes in each direction
Number of HOT Lanes	2 reversible lanes	1 lane in each direction for 8 miles, 2 reversible lanes for 3 miles	1 lane in each direction	2 lanes in each direction
Carpool Policy	HOV 2+ ride free	HOV 2+ ride free	HOV 2+ ride free	HOV 3+ ride free during off-peak/ HOV 3+ pay discounted toll during peak
Pricing Approach	Dynamically Priced	Dynamically Priced	Dynamically Priced	Static Toll Schedule
Hours/Days of Operation	12.5 hours per day/ 5 days per week	9 hours per day/ 5 days per week	14 hours per day/ 7 days per week	24 hours per day/ 7 days per week
Annual Gross Revenue	Approximately $1.2 million (2007)	Approximately $1.0 million (2007)	$0.3 million (2008)	$43.7 million (2009)

Source: Washington State Department of Transportation (2009), Orange County Transportation Authority (2009), Cambridge Systematics (2002), Federal Highway Administration (2007), Federal Highway Administration (undated).

THE MODEL

The model considers one direction of a highway facility that is divided into GP and HOT lanes. It applies to a single period (say a typical peak hour or perhaps a five-minute period if dynamic pricing is implemented) in which N_S SOVs and N_C HOVs utilize the facility.[4] Each vehicle is associated with a value of time, $v \in [\underline{v}, \overline{v}]$, which varies from one vehicle to the next. The fraction of SOVs and HOVs with a value of time less than or equal to v is given by $F_S(v)$ and $F_C(v)$, respectively.

Both the GP and HOT lanes are congestible. The amount of time it takes to travel in the GP and HOT lanes is given by $t_G(x_G)$ and $t_H(x_H)$, respectively, where x_G and x_H denote the number of vehicles using the GP and HOT lanes and $x_G + x_H = N_S + N_C$. The travel time functions are assumed to be increasing and strictly convex. Furthermore, it is reasonable to assume that $t_G(0) = t_H(0)$ since a free flowing HOT lane should operate at about the same speed as a parallel, free flowing GP lane, although this assumption can be relaxed without impacting the modeling.

Under the base case HOT lane policy, the tolling authority allows HOVs to use the HOT lanes at no cost while SOVs must pay a toll $p > 0$ to access the HOT lanes. The equilibrium condition requires that no driver can change his or her lane choice and be made better off given the choices of other drivers. That is, drivers choose the travel option that minimizes their generalized cost of travel in equilibrium. This implies that a SOV with value of time v will select the HOT lane if $vt_H^* + p < vt_G^*$ and the GP lanes if $vt_H^* + p > vt_G^*$ where t_G^* and t_H^* represent the equilibrium travel times in the GP and HOT lane respectively. An SOV is indifferent between the two alternatives if $vt_H^* + p = vt_G^*$. If it is optimal for SOV users with value of time v to choose the HOT lanes, it must be optimal for all SOV users with a value of time greater than v to choose the HOT lane as well. Similarly, if SOV users with value of time v do not choose the HOT alternative, all SOVs with a value of time less than v must also choose the GP lanes. Carpoolers are exempt from paying the toll and therefore are only concerned with the relative travel times in the GP and HOT lanes. They will use the HOT lane as long as $t_H^* < t_G^*$ and will be indifferent between these two options if $t_H^* = t_G^*$.

Three types of equilibrium can occur depending on the SOV and HOV volumes and the toll level that is selected by the tolling authority. In Case 1, if travel times in the HOT lanes are greater than in the GP lanes when all SOVs use the GP lanes and all HOVs use the HOT lanes (which implies that, $t_G(N_S) \geq t_H(N_C)$), then regardless of how the toll is set, a pooling equilibrium occurs in which all SOVs stay in the GP lanes and carpool vehicles allocate themselves to the GP and HOT lane until travel times equalize (i.e., $t_H^* = t_G^*$). In Case 2, if travel times in the HOT lane are less than in the GP lanes when all SOVs use the GP lanes and all HOVs use the HOT lanes (which implies that $t_G(N_S) > t_H(N_C)$) and the toll is set such that the SOV driver with the highest value of time would find it advantageous to buy his or her way into the HOT lanes (i.e., $p \in (0, \overline{v}(t_G(N_S) - t_H(N_C))]$), a separating equilibrium occurs in which all HOVs use the HOT lane and the SOV users with the highest value of time pay the toll to use the HOT lane. Specifically, all SOVs with a value of time greater than (less than) $\hat{v}$ use the HOT (GP) lanes, where $\hat{v}$ solves,

$$(1) \quad \hat{v}t_G(N_S F_S(\hat{v})) = \hat{v}t_H(N_C + N_S(1 - F_S(\hat{v}))) + p .$$

Here $\hat{v}$ represents the value of time of the SOV users who are indifferent between using the GP lane and paying the toll to use HOT lanes in equilibrium. In Case 3, if travel times in the HOT lanes are less than in the GP lanes when all SOVs use the GP lanes and all HOVs use the HOT lanes (which implies that $t_G(N_S) > t_H(N_C)$ but the toll level is set so high that the SOV users with the highest value of time would not find it advantageous to buy their way into the HOT lanes (i.e., $p > \overline{v}(t_G(N_S) - t_H(N_C))$), then despite the fact that traffic in the HOT lanes are moving faster than in the GP lanes the toll is set so high that no SOV users will opt to use the HOT lane in equilibrium. Therefore, the HOT lane effectively operates like a carpool lane.

In Case 1, there are so many HOVs that speed in the HOT lanes degrades to the same level as in the GP lanes and no SOVs are willing to buy their way into the HOT lane. In practice, if a facility is operating in this state often, the tolling authority should consider converting the HOT lane to a GP lane, raising the carpool occupancy requirement, or converting a GP lane to a HOT lane so that there is additional HOT capacity. Case 2 is more interesting. In this case, the HOT lane is used by all carpools and, by the highest value of time, SOV users. The total number of vehicles using the GP and HOT lanes is $x_G(\hat{v}) = N_S F_S(\hat{v})$ and $x_H(\hat{v}) = N_C + N_S(1 - F_S(\hat{v}))$, respectively. In Case 3, the HOT lane is made so expensive that no SOVs use it, so it effectively operates as a carpool lane. Throughout the rest of this paper, it is assumed that the facility is operating under the conditions of Case 2.

HOT Lane Pricing Policies

Having described equilibrium under HOT lane pricing, optimal toll policies are characterized next under three pricing objectives: minimizing the aggregate cost of travel, maximizing the revenue generated by the HOT lanes, and maintaining a target level of service in the HOT lanes. These toll setting problems can be formulated as mathematical programming problems.

When solving for optimal policies, it is helpful to note that there exists a unique toll $p > 0$ that will induce any feasible $\hat{v}$ in the equilibrium associated with Case 2. In particular, this toll ($p^e(\hat{v})$) can be found by rearranging Eq. (1) to get,

(2) $$p^e(\hat{v}) = \hat{v}\left[t_G(N_S F_S(\hat{v})) - t_H(N_C + N_S(1 - F_S(\hat{v})))\right].$$

Notice that the range of marginal values of time that are feasible are defined by $\hat{v} \in (v^o, \overline{v}]$ where,

(3) $$v^o = [v \in [\underline{v}, \overline{v}] : t_G(x_G(v)) = t_H(x_H(v))]\cdot$$

Here v^o represents the lowest feasible marginal value of time that can be induced by a strictly positive toll level. A marginal value of time will approach v^o as the toll level approaches zero. As this happens, travel times on the GP and HOT lanes will also converge.

Minimizing Aggregate Travel Time Cost. Under this pricing policy, the tolling authority seeks to minimize the aggregate travel time cost in the corridor. Since total demand for travel on the highway is assumed perfectly inelastic, the welfare of motorists in the model will vary with the time and monetary cost of completing trips. Under the additional assumption that toll revenues are returned to motorists as lump-sum transfers, minimizing aggregating travel time costs will be consistent with welfare maximization.[5]

Aggregate travel time cost in the corridor can be broken down into the travel time costs of GP ($C_G(\hat{v})$) and HOT ($C_H(\hat{v})$) lane users, which are respectively calculated as,

(4) $$C_G(\hat{v}) = t_G(x_G(\hat{v})) \times x_G(\hat{v}) \times \frac{\int_{\underline{v}}^{\hat{v}} v\,dF_S(v)}{F_S(\hat{v})}$$

(5) $$C_H(\hat{v}) = t_H(x_H(\hat{v})) \times N_S(1 - F_S(\hat{v})) \times \frac{\int_{\hat{v}}^{\overline{v}} v\,dF_S(v)}{1 - F_S(\hat{v})} + t_H(x_H(\hat{v})) \times N_C \times \int_{\underline{v}}^{\overline{v}} v\,dF_C(v)$$

The above expressions for aggregate travel time cost reflect the fact that SOVs sort between the GP and HOT lanes based on their value of time, with all SOVs with a value of time less than $\hat{v}$ opting to use the GP lanes and all those SOVs with a value of time greater than $\hat{v}$ opting to purchase their way into the HOT lanes. The sorting of SOVs affects the calculation of both the number of SOVs using the GP and HOT lanes, and how their average value of time is calculated. Specifically, the aggregate travel time cost experienced in the GP lanes ($C_G(\hat{v})$) represents the product of the travel time in the GP lanes ($t_G(x_G(\hat{v}))$), the number of SOVs using the GP lanes ($x_G(\hat{v}) = N_S F_S(\hat{v})$), and the average value of time of SOVs who use the GP lanes ($\int_{\underline{v}}^{\hat{v}} vdF_S(v) \big/ F_S(\hat{v})$).[6] The aggregate travel time cost associated with the HOT lanes ($C_H(\hat{v})$) can be decomposed into the aggregate travel time cost for SOV and HOV users of the HOT lanes. The aggregate travel time cost for the SOV users of the HOT lane is calculated as the product of the HOT lane travel time ($t_H(x_H(\hat{v}))$), the number of SOVs using the HOT lane ($N_S(1-F_S(\hat{v}))$), and the average value of time of SOVs using the HOT lane ($\int_{\hat{v}}^{\overline{v}} vdF_S(v) \big/ (1-F_S(\hat{v}))$). The aggregate travel cost of the HOV users in the HOT lane is calculated as the product of the HOT lane travel time ($t_H(x_H(\hat{v}))$), the number of HOVs in HOT lane (N_C), and the average value of time of HOVs ($\int_{\underline{v}}^{\overline{v}} vdF_C(v)$). The aggregate travel time cost in both the GP and HOT lanes is $C(\hat{v})$ and it is equal to $C(\hat{v}) = C_G(\hat{v}) + C_H(\hat{v})$.

A tolling authority wishing to minimize aggregate travel time cost ($C(\hat{v})$) will want to induce through tolling a marginal value of time $\hat{v}*$ such that, $\partial C(\hat{v}*)/\partial \hat{v}* = 0$. The Appendix shows this derivative. Using this expression and the equilibrium implied by Eq. (2), the Appendix further shows that the toll that minimizes aggregate travel time cost equals the difference between the marginal external congestion cost (*MEC*) in the HOT and GP lanes. Specifically, this toll is,

$$(6) \quad p^* = MEC_H(\hat{v}*) - MEC_G(\hat{v}*)$$

Here,

$$(7) \quad MEC_G(\hat{v}) = \frac{\partial t_G(x_G(\hat{v}))}{\partial x_G(\hat{v})} \times x_G(\hat{v}) \times \frac{\int_{\underline{v}}^{\hat{v}} vdF_S(v)}{F_S(\hat{v})}$$

$$(8) \quad MEC_H(\hat{v}) = \frac{\partial t_H(x_H(\hat{v}))}{\partial x_H(\hat{v})} \times \left(N_S(1-F_S(\hat{v})) \times \frac{\int_{\hat{v}}^{\overline{v}} vdF_S(v)}{1-F_S(\hat{v})} + N_C \times \int_{\underline{v}}^{\overline{v}} vdF_C(v) \right)$$

are the marginal external congestion cost calculated for the GP and HOT lanes, respectively.[7] The *MEC* represents the monetized value of the congestion externality generated by a marginal increase in use in the GP or HOT lanes. The toll described by Eq. (6) cause motorists to internalize their contribution to congestion in the HOT and GP lanes when making a decision over which set of lanes to utilize. In particular, if a motorist opts to use the HOT lanes, he or she will increase the travel time cost experienced by other HOT lane users by $MEC_H(\hat{v}*)$, but reduce travel time cost in the GP lanes by $MEC_G(\hat{v}*)$. The toll that minimizes aggregate travel time cost causes SOVs to take both of these congestion related effects into account when deciding whether or not to use the HOT lanes.

Revenue Maximization: The second tolling approach seeks to maximize the revenue that will be generated by the HOT lane. This approach coincides with that which would likely be pursued by a private HOT lane operator. One might also observe a publicly owned HOT lane facility priced in this way if the operator is financially constrained.

Having defined the equilibrium pricing function in Eq. (2), it is possible to quantify the revenue raised by the HOT lane as a function of any feasible $\hat{v}$. In particular, the HOT lane revenue function is given by $R(\hat{v}) = N_S(1-F_S(\hat{v})) \times p^e(\hat{v})$. It represents the product of the number of users paying to use the HOT lanes and the toll charge. If $\hat{v}^{**}$ is the solution to the revenue maximization problem, then the first order condition requires $\partial R(\hat{v}^{**}) / \partial\hat{v}^{**} = 0$. As shown in the Appendix, the first order condition implies that the revenue maximizing toll level will equal,

$$(9) \quad p^{**} = \frac{1 - F_S(\hat{v}^{**})}{f_S(\hat{v}^{**})} \frac{\partial p^e(\hat{v}^{**})}{\partial \hat{v}^{**}}$$

Here, $f_S(\hat{v}) = \partial F_S(\hat{v}) / \partial \hat{v}$.

Maintain a Minimum Level of Service (LOS): Under the third objective, the HOT lane is priced so as to maintain a minimum level of service, which can be defined as the minimum speed in the HOT lane. Let $\tilde{s}$ be the desired minimum speed measured in miles per hours (mph). If travel time is measured in hours and the length of the facility in miles and equals *d*, then a tolling authority wishing to maximize use of the HOT lanes without exceeding the desired minimum speed will solve,

$$(10) \quad \max_{\hat{v} \in (v^o, \bar{v}]} x_H(\hat{v}) \text{ such that } \frac{d}{t_H(x_H(\hat{v}))} \geq \tilde{s} .$$

Let $\hat{v}^{***}$ represent a solution to Eq. (10). The price required to induce a HOT lane speed of $\tilde{s}$ is given by $p^{***} = p^e(\hat{v}^{***})$. In fact, any toll greater than or equal to p^{***} will induce a speed of at least $\tilde{s}$. For the purposes of the numeric example presented in the next section, it is assumed that the tolling authority sets a minimum level of service in the HOT lane that corresponds to a minimum speed of 50 mph (i.e., $\tilde{s}$ = 50 mph).

Notice that the goal of maintaining a minimum level of service can be combined with other objectives by simply incorporating the constraint in Eq. (10) into other toll optimization problems. For instance, one could solve for the toll that minimizes aggregate social cost or maximizes revenue subject to a minimum level of service constraint simply by introducing the constraints in Eq. (10) into the aggregate cost and revenue maximization equation, respectively.

A NUMERIC COMPARISON OF ALTERNATIVE PRICING OBJECTIVES

In this section, the relationships between the HOT lane toll level, aggregate travel time cost, revenue, and GP and HOT lane speed under different pricing objectives and facility design features are examined.

Specification of the Numeric Example

In order to examine numerically the implications of different tolling policies, vehicle and facility characteristics must be specified. It is assumed that the facility is 10 miles long (i.e. d = 10 miles) and originally operates with three GP lanes and one carpool lane in each direction. Also, it is assumed that the travel time functions are of the form suggested by the Bureau of Public Roads (1964). Specifically,

$$(11) \quad t(x) = \frac{10}{60}\left(1 + 0.20\left(\frac{x}{2,000k}\right)^4\right)$$

where *k* represents the number of GP or HOT lanes provided and *x* is the volume of vehicles using those lanes per hour. The implied free flow speed for this volume delay function is 60 miles per hour. The empirical evidence supporting the use of the BPR function is discussed in Small (1992). Next, it is assumed that there are 8,972 SOVs and 1,028 HOVs per hour. These traffic volumes were selected to give speeds in the GP lanes and HOT lane of approximately 30 and 59 mph, respectively, when the HOT lane is operated as an HOV lane. Under these conditions, there could be advantages to converting the HOV lane to a HOT lane to relieve some congestion in the GP lanes. In sensitivity analysis, both the total corridor volume and the share of vehicles that are HOVs are varied. Finally, it is assumed that the value of time of SOVs and HOVs follow a log-normal distribution. The mean and standard deviation of SOVs' values of time are assumed to be $20/hour and $10/hour, respectively, while HOVs are assumed to have a mean value of time of $40/hour with a standard deviation of $20/hour. These value of time assumptions imply that the mean and standard deviations of the logarithms of values of time are 2.88 and 0.47 for SOVs and 3.58 and 0.47 for HOVs, respectively.

The mean SOV value of time assumption is in line with estimates from Small et al. (2005), who studied the State Route 91 Express Lane facility. As is the case in other similar studies, such as Sullivan and Burris (2006), HOVs are assumed to have a higher average value of time, due primarily to their higher vehicle occupancy. Hensher (2008) provides empirical evidence for this assumption and discusses some implications for toll road evaluations. Sensitivity analysis is conducted to illustrate how the results change when the mean HOV value of time is reduced to $30/hour.

In addition to the three alternative tolling objectives, three other changes to the facility are considered. The first deals with carpools. The implications of moving from a policy where carpools can use the HOT lanes for free to one where every vehicle must pay to use the HOT lanes are examined. This case is similar to the express lane concept on State Route 91 and may be appropriate if there is limited capacity to sell or if revenue generation becomes a priority. The second policy variation is with respect to the share of capacity provided as GP and HOT. In particular, the implications of converting an additional lane to HOT so that there are a total of two GP lanes and two HOT lanes operating in total in each direction in considered. The third and final policy considered involves converting the HOV lane to a GP lane, so that all four lanes on the highway operate as GP lanes. No revenue is generated in this case, but it serves as a useful benchmark for comparison.

Hot Lane Performance Under Alternative Scenarios

Table 2 summarizes the conditions under the carpool lane scenario, an all-GP lanes scenario, and nine different tolling conditions, carpool policy, and capacity scenarios. In all of the HOT lane scenarios except the eighth, which sets the toll to maximize revenue and converts one GP lane to a HOT lane, the HOT alternatives operate at a lower aggregate travel time cost than the carpool lane base case. In the three scenarios where aggregate travel time cost is minimized (Scenarios one, four, and seven), there is a 22% to 23% reduction in aggregate travel time cost relative to the HOV scenario. When there is only a small amount of capacity to sell, as is the case in the first through the third scenario, the different HOT lane pricing objectives perform similarly in terms of aggregate travel time cost. However, in the cases where there is more capacity to sell, there is greater variation in aggregate travel time cost between the scenarios. It is also interesting to note that the conversion of the HOT lane to a GP lane would produce considerable benefits as measured by the reductions in aggregate travel time costs.

In terms of revenue generation, the seventh through ninth scenarios show large differences between the pricing objectives. Revenue under the level of service objective are more than double those found under aggregate travel time cost minimization, while revenue maximization produces revenue which are four times larger than those associated with aggregate travel time cost minimization. These large differences in revenue generation across pricing objectives do not prevail when there is limited capacity to sell as is the situation in scenario one through three.

Under revenue maximization, toll levels tend to be much higher, which causes fewer motorists to pay to use the HOT lane. As a result, there is more congestion in the GP lanes while speed in the HOT lane remains near the free flow level. When aggregate travel time cost is minimized,

Table 2: Summary of Peak Hour Performance Under Alternative HOT Lane Policies

	HOV Lane	All GP Lanes	HOT Lane(s): Cost Min	HOT Lane(s): Rev Max	HOT Lane(s): Maintain LOS
			3 GP Lanes, 1 HOT Lane; HOVs Ride Free		
Scenario	*Base Case*		*(1)*	*(2)*	*(3)*
GP Volume (x_G)	8,972.0	10,000.0	7,787.1	8,248.7	8,000.0
HOV/HOT Volume (x_H)	1,028.0	NA	2,212.9	1,751.3	2,000.0
GP Speed (MPH)	30.0	40.3	38.3	35.0	36.8
HOV/HOT Speed (MPH)	59.2	NA	46.2	53.7	50.0
Marginal VoT ($\hat{v}$)	NA	NA	\$30.32	\$34.67	\$32.06
Toll (p)	NA	NA	\$1.35	\$3.45	\$2.31
Aggregate Travel Cost ($C(\hat{v})$)	\$69,472	\$54,725	\$53,739	\$55,842	\$54,202
Toll Revenue ($R(\hat{v})$)	NA	NA	\$1,603	\$2,495	\$2,245
			3 GP Lanes, 1 HOT Lane; HOVs Must Pay		
Scenario			*(4)*	*(5)*	*(6)*
GP Volume (x_G)			7,831.3	8,812.3	8,000.0
HOT Volume (x_H)			2,168.7	1,187.7	2,000.0
GP Speed (MPH)			38.0	31.1	36.8
HOT Speed (MPH)			47.0	58.5	50.0
Marginal VoT ($\hat{v}$)			\$28.45	\$35.42	\$29.36
Toll (p)			\$1.44	\$5.35	\$2.11
Aggregate Travel Cost ($C(\hat{v})$)			\$53,593	\$62,284	\$53,954
Toll Revenue ($R(\hat{v})$)			\$3,125	\$6,350	\$4,229
			2 GP Lanes, 2 HOT Lanes; HOVs Ride Free		
Scenario			*(7)*	*(8)*	*(9)*
GP Volume (x_G)			5,423.5	7,676.2	6,000.0
HOT Volume (x_H)			4,576.5	2,323.8	4,000.0
GP Speed (MPH)			35.8	16.2	29.8
HOT Speed (MPH)			44.7	58.7	50.0
Marginal VoT ($\hat{v}$)			\$20.27	\$29.52	\$21.98
Toll (p)			\$1.13	\$13.23	\$2.98
Aggregate Travel Cost ($C(\hat{v})$)			\$53,484	\$95,765	\$55,902
Toll Revenue ($R(\hat{v})$)			\$3,995	\$17,150	\$8,848

Notes: GP = general purpose; HOT = high occupancy toll; HOV = High occupancy vehicle; MPH = miles per hour; VoT = value of time; Cost Min = cost minimization; Rev Max = revenue maximization; Maintain LOS = maintain a minimum level of service.

there is less service differentiation between the GP and HOT lanes than is observed in revenue maximization. When the toll is set to maintain a specific level of service, the speed in the GP lanes varies inversely with the target HOT lane speed.

The finding that revenue maximization tends to result in higher tolls than would be efficient is consistent with Liu and McDonald (1998), Small and Yan (2001), and Verhoef and Small (2004). Public officials involved in the management of HOT lanes are likely to face a tradeoff between revenue generation and achieving travel time cost efficiency. Because aggregate travel time cost cannot be as easily measured as revenue, there is likely to be a natural tendency to put revenue generation above it.

Sensitivity of Results to Alternative Toll Levels

Figure 1 shows how aggregate travel time cost, revenue, and speed in the GP and HOT lanes vary as the toll is adjusted for the case where there are three GP lanes and one HOT lane, and carpools ride free in the HOT lane. This is the same capacity and carpool policy explored in scenario one through three. In the graphs, when the toll equals zero, the HOT lane effectively operates as a GP lane and speeds across the two lane types are the same. As the figure shows, there are modest benefits in terms of aggregate travel time cost reductions from HOT lane tolling. But, if the toll is set high, HOT lane capacity can be underutilized, causing the facility to operate worse in terms of aggregate travel time cost than if the HOT lane is simply converted to a GP lane. Figure 1 highlights the fact that the way in which HOT lanes are priced can have a non-trivial impact on the potential travel time benefits and revenue generated. As a consequence, considerable thought should go into selecting tolling algorithms and pricing parameters when implementing HOT lanes.

Sensitivity of Results to Alternative Demand Level

The estimates above are most likely representative of a facility operating under peak demand. In this section, demand is varied to illustrate how a HOT lane priced under different objectives might operate during other times of the day. To vary total corridor demand, it is assumed that the share of carpool vehicles remains constant at 10.2% of total vehicle demand. Let $N = N_S + N_C$ and rewrite $N_S = (1 - 0.102)N$ and $N_C = 0.102N$.

Figure 2 shows how tolls, aggregate travel time cost, and revenue vary with traffic volume under each pricing objective and the assumptions implicit in scenario one through three. Figure 3 extends Figure 2 to show GP and HOT lane speed. Under revenue maximization, the toll level tends to be about twice as high as the toll level that minimizes aggregate travel time cost. The toll that maintains a minimum HOT lane level of service behaves much differently. As shown in Figure 3, below 8,000 vehicles per hour (i.e., $N = 8{,}000$) the HOT lane effectively becomes a GP lane (i.e., $p = 0$) under the minimum HOT lane level of service objective and the speed in the GP lanes and HOT lane are equalized above the minimum required HOT lane speed of 50 miles per hour. When traffic volume exceeds 8,000 vehicles per hour, however, pricing keeps the HOT lane moving at 50 miles per hour.

The second pane of Figure 2 shows how aggregate travel time cost varies under each objective and the case where the HOT lane is operated as an HOV lane. Interestingly, it is difficult to differentiate between the alternative HOT lane objectives in this graph. However, the HOV lane case lies above the HOT lane cases for all levels of traffic but most noticeably when corridor demand is highest. The bottom pane of Figure 2 illustrates revenue generation under each HOT lane pricing objective. Revenues under aggregate travel time cost minimization tend to be about 60% of that obtained under revenue maximization. The minimum level of service objective produces a discontinuous level of revenue for the reasons discussed above.

Figure 1: Relationship Between Toll Level and Aggregate Travel Time Cost, Revenue, and Speed in Scenarios 1–3 (3 GP Lanes, 1 HOT Lane; HOVs Ride Free)

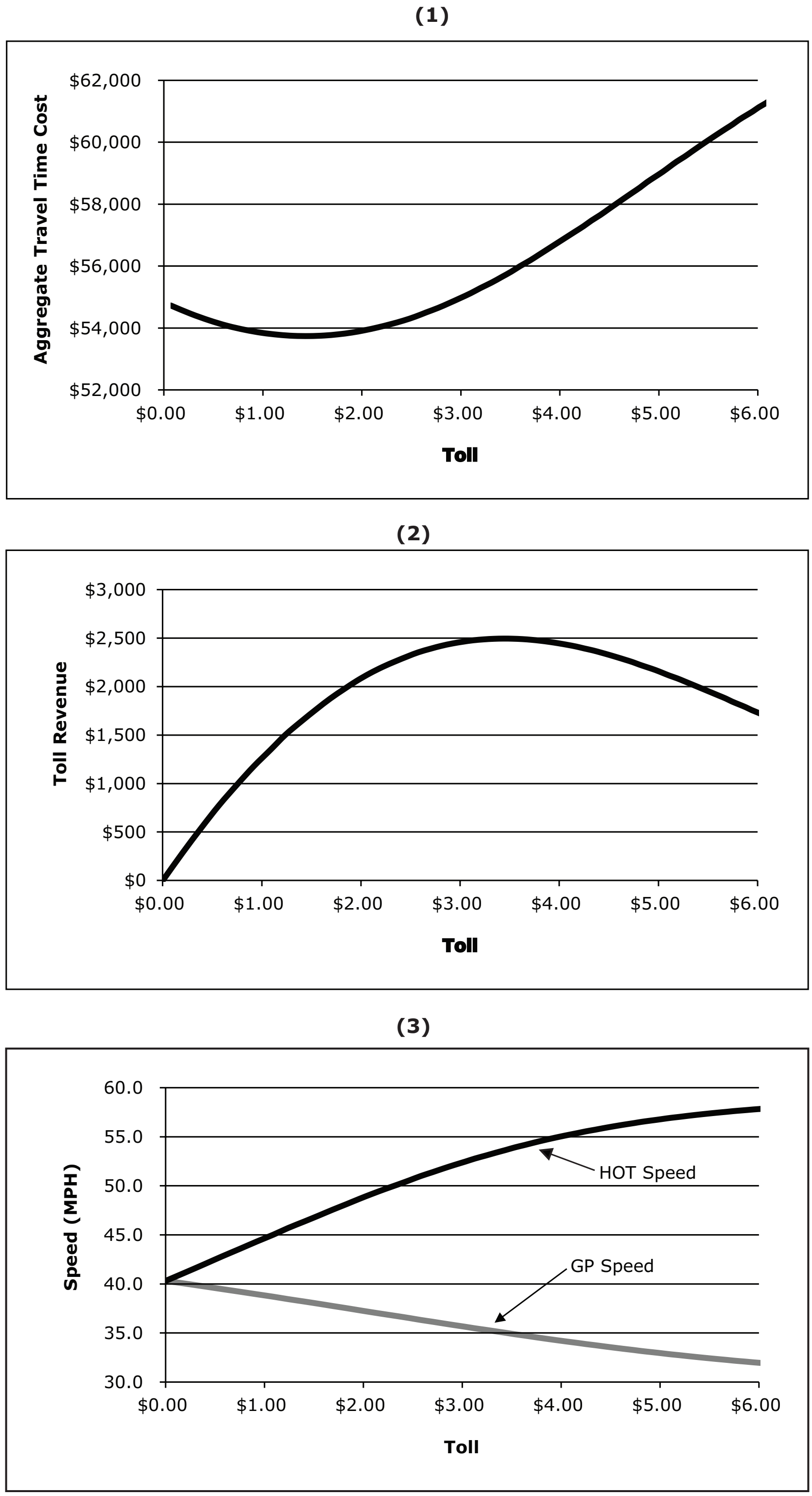
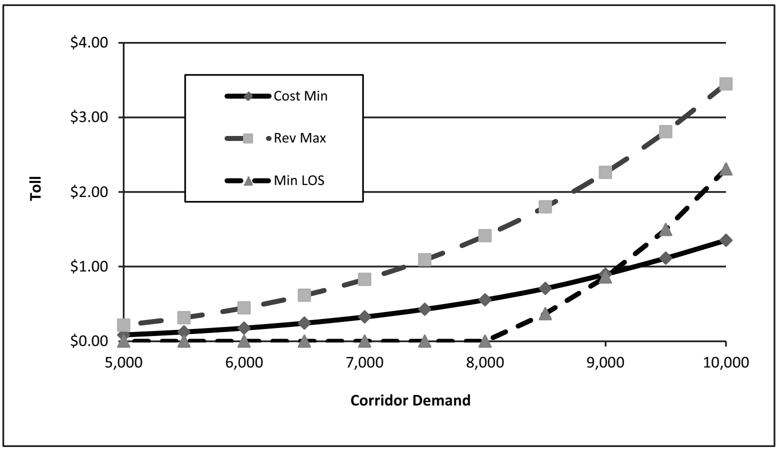
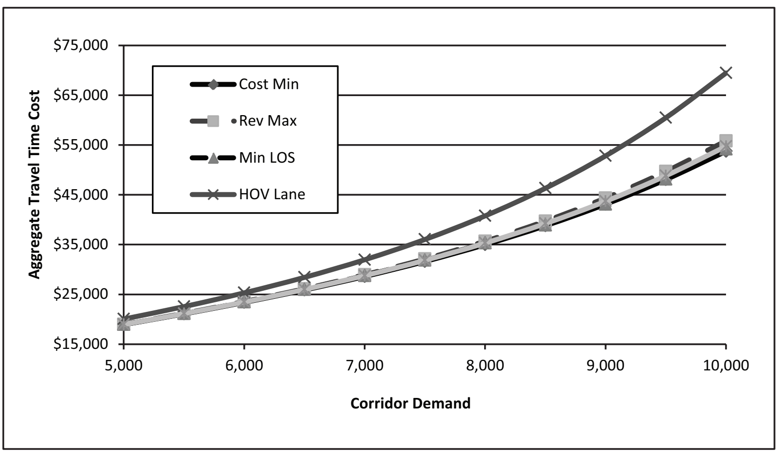


Figure 2: Relationship Between Toll Level and Aggregate Travel Time Cost, and Revenue Under Varying Corridor Demand and Tolling Objectives in Scenarios 1–3 (3 GP Lanes, 1 HOT Lane; HOVs Ride Free)

(1)



(2)



(3)

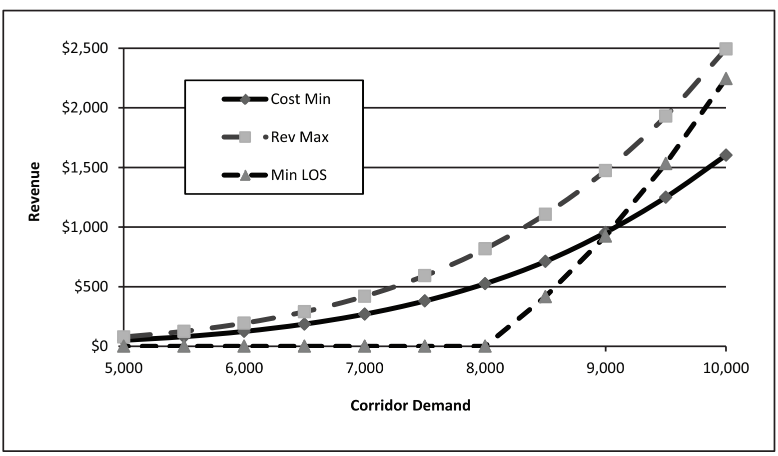
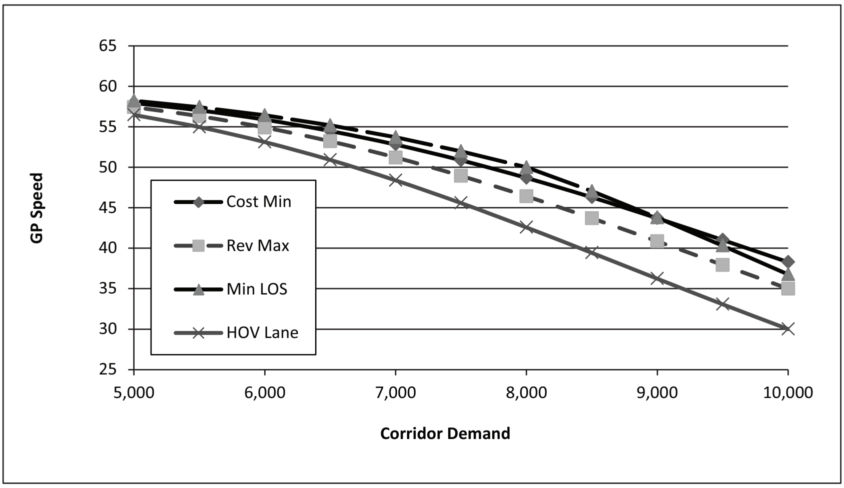
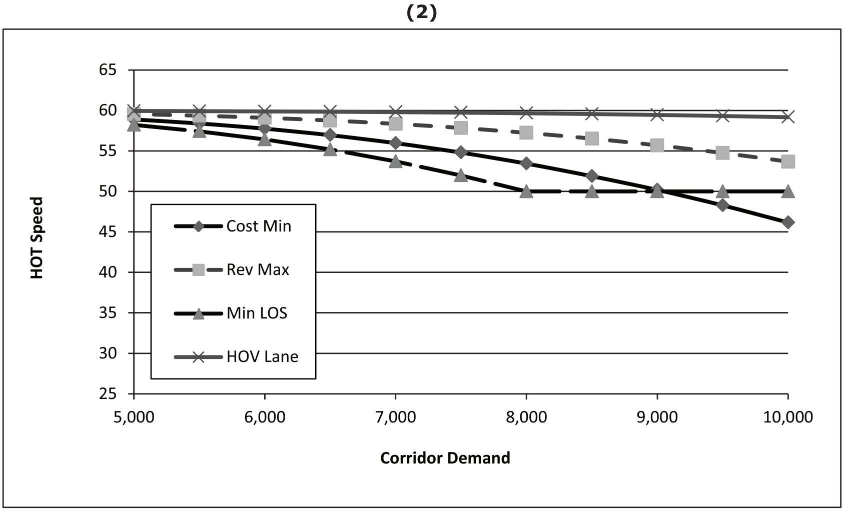


Figure 3: Relationship Between GP and HOT Lane Speed Under Varying Corridor Demand Tolling Objectives in Scenarios 1–3 (3 GP Lanes, 1 HOT Lane; HOVs Ride Free)

(1)



(2)



As shown in Figure 3, the HOV lane policy produces a slower speed in the GP lanes than the HOT lane policies. This is indicative of the fact that some SOV users purchase their way into the HOT lane under the HOT lane policy, but otherwise would have been forced to remain in the GP lanes under an HOV policy. In general, as the toll level in the HOT lane is lowered, more SOVs will purchase their way into the HOT lanes. This will improve the GP lane speed, but at the expense of potentially reducing the HOT lane speed.

Sensitivity of Results to a Higher Carpool Rate

The results presented above assume that approximately 10% of vehicles using the facility are carpool vehicles. In HOT lane evaluations, the share of carpool vehicles in the traffic stream will vary by location of the facility, direction, time of day, and day of week. In other analyses performed by the author using traffic count data for freeway facilities in the San Francisco Bay Area, the observed share of vehicles using carpool lanes during the morning peak averaged around 10%, with higher levels observed during the afternoon peak period.

How do the results change as the share of carpool vehicles using the facility increase? Table 3 presents the results of a sensitivity analysis to identify the effect of increasing the share of carpool vehicles in the corridor to 15%, holding corridor traffic volume constant at 10,000 vehicles per hour (i.e. $N = 10,000$, $N_S = 8,500$, and $N_C = 1,500$). In this case, when carpools are allowed to use the HOT lane for free, the tolling authority effectively has less capacity to sell to SOVs. While there continues to be the potential for aggregate travel time cost reductions from converting the HOV lane to a HOT lane, these benefits are less than in the previous estimates where HOVs made up a small share of the traffic stream. For example, when the share of vehicles that are HOVs is approximately 10%, the change in aggregate travel time cost from going from an HOV policy to the HOT lane policy described by scenario one is $15,733 per hour (= $69,472 - $53,739, see Table 2). When the share of HOVs in the traffic stream is increased to 15%, the change in aggregate travel time cost generated by converting from an HOV policy to the HOT lane policy described by scenario one falls to $6,493 per hour (= $62,569 - $56,076, see Table 3). Revenue generating potential is also impaired as the share of vehicles that qualify as carpools (and can therefore ride free in the HOT lane) increases, making it less likely that the HOT lane will be financially self-supporting. For example, the revenue generated in scenario one falls by 32%, from $1,603 (in Table 2) to $1,087 (in Table 3) per hour, when the carpool share is increased from 10% to 15%. If, however, HOVs must pay to use the HOT lane, a higher share of HOV motorists using the facility can increase both aggregate travel-time cost reductions and the revenue potential of converting an HOV lane to a tolled lane. This can be seen by comparing the results for scenarios four through six in Tables 2 and 3 and stems from the fact that HOVs tend to have a higher value of time than SOVs.

Sensitivity of Results to Lower Carpool Values of Time

The final sensitivity analysis in Table 4 shows results when HOVs are assumed to have a lower mean value of time. By reducing the mean value of time associated with HOV users, the travel time benefits generated from providing faster travel to HOV users are given less weight in the aggregate travel time cost calculation. To illustrate the implications of this,, the mean HOV value of time is reduced from $40/hour to $30/hour in the results presented in Table 4, but all other assumptions about the vehicle population are held constant at their original values. When carpools are allowed to use the HOT lane for free, this change only impacts traffic volume, speed, the toll level, and revenue when tolls are set to minimize aggregate travel time costs. To illustrate this, consider the results for scenario one shown in Tables 2 and 4. Reducing the mean HOV value of time from $40/hour to $30/hour causes the toll level to fall from $1.35 (Table 2) to $1.17 (Table 4) for scenario one. As a result, more SOVs purchase their way into the HOT lane, and speed in the GP lanes improve slightly from 38.3 miles per hour (Table 2) to 38.6 miles per hour (Table 4). When HOVs must pay to use the HOT lane (scenarios four through six), the revenue generating potential of the facility is reduced as the mean HOV value of time is reduced. For example, the revenue generated under scenario 5 falls from $6,350 per hour (Table 2) to $5,946 per hour (Table 4).

CONCLUSION

HOT lanes have emerged as the most politically feasible form of congestion pricing in the United States. In the coming decade, it seems likely that more examples of the HOT lane concept will be put into practice. This paper has presented a simple HOT lane model that can be used for "sketch-planning" analysis to understand the likely performance of these facilities. Using a realistic numeric example, the model was used to gain insights into how HOT lane performance varies with the way in which facilities are designed and priced. The analysis suggests that different pricing objectives can

Table 3: Sensitivity Analysis with HOV Volume Increased to 1,500 Vehicles per Hour and SOV Volume Decreased to 8,500 Vehicles per Hour

	HOV Lane	All GP Lanes	HOT Lane(s) Cost Min	Rev Max	Maintain LOS
	Base		*3 GP Lanes, 1 HOT Lane; HOVs Ride Free*		
Scenario	*Case*		*(1')*	*(2')*	*(3')*
GP Volume (x_G)	8,500.0	10,000	7,785.1	8,018.4	8,000.0
HOV/HOT Volume (x_H)	1,500.0	NA	2,214.9	1,981.6	2,000.0
GP Speed (MPH)	33.2	40.3	38.3	36.6	36.8
HOV/HOT Speed (MPH)	56.4	NA	46.1	50.3	50.0
Marginal VoT ($\hat{v}$)	NA	NA	$34.30	$37.80	$37.46
Toll (p)	NA	NA	$1.52	$2.80	$2.70
Aggregate Travel Cost ($C(\hat{v})$)	$62,569	$57,081	$56,076	$56,721	$56,626
Toll Revenue ($R(\hat{v})$)	NA	NA	$1,087	$1,351	$1,349
			3 GP Lanes, 1 HOT Lane; HOVs Must Pay		
Scenario			*(4')*	*(5')*	*(6')*
GP Volume (x_G)			7,820.8	8,828.1	8,000.0
HOT Volume (x_H)			2,179.2	1,171.9	2,000.0
GP Speed (MPH)			38.0	31.0	36.8
HOT Speed (MPH)			46.8	58.6	50.0
Marginal VoT ($\hat{v}$)			$29.72	$37.68	$30.80
Toll (p)			$1.46	$5.74	$2.22
Aggregate Travel Cost ($C(\hat{v})$)			$55,792	$64,989	$56,133
Toll Revenue ($R(\hat{v})$)			$3,189	$6,726	$4,436
			2 GP Lanes, 2 HOT Lanes; HOVs Ride Free		
Scenario			*(7')*	*(8')*	*(9')*
GP Volume (x_G)			5,422.6	7,302.3	6,000.0
HOT Volume (x_H)			4,577.4	2,697.7	4,000.0
GP Speed (MPH)			35.8	18.6	29.8
HOT Speed (MPH)			44.7	57.6	50.0
Marginal VoT ($\hat{v}$)			$21.13	$29.74	$23.10
Toll (p)			$1.17	$10.81	$3.13
Aggregate Travel Cost ($C(\hat{v})$)			$55,715	$84,911	$58,143
Toll Revenue ($R(\hat{v})$)			$3,605	$12,943	$7,821

Notes: GP = general purpose; HOT = high occupancy toll; HOV = High occupancy vehicle; MPH = miles per hour; VoT = value of time; Cost Min = cost minimization; Rev Max = revenue maximization; Maintain LOS = maintain a minimum level of service.

Table 4: Sensitivity Analysis with Mean Carpool Value of Time Decreased to $30/hour

	HOV Lane	All GP Lanes	HOT Lane(s) Cost Min	HOT Lane(s) Rev Max	HOT Lane(s) Maintain LOS
			3 GP Lanes, 1 HOT Lane; HOVs Ride Free		
Scenario	*Base Case*		*(1'')*	*(2'')*	*(3'')*
GP Volume (x_G)	8,972.0	10,000	7,746.3	8,248.7	8,000.0
HOV/HOT Volume (x_H)	1,028.0	NA	2,253.7	1,751.3	2,000.0
GP Speed (MPH)	30.0	40.3	38.6	35.0	36.8
HOV/HOT Speed (MPH)	59.2	NA	45.4	53.7	50.0
Marginal VoT ($\hat{v}$)	NA	NA	$30.02	$34.67	$32.06
Toll (p)	NA	NA	$1.17	$3.45	$2.31
Aggregate Travel Cost ($C(\hat{v})$)	$65,940	$52,177	$51,493	$53,928	$52,146
Toll Revenue ($R(\hat{v})$)	NA	NA	$1,430	$2,495	$2,245
			3 GP Lanes, 1 HOT Lane; HOVs Must Pay		
Scenario			*(4'')*	*(5'')*	*(6'')*
GP Volume (x_G)			7,795.1	8,788.3	8,000.0
HOT Volume (x_H)			2,204.9	1,211.7	2,000.0
GP Speed (MPH)			38.2	31.2	36.8
HOT Speed (MPH)			46.3	58.4	50.0
Marginal VoT ($\hat{v}$)			$26.89	$32.95	$27.87
Toll (p)			$1.23	$4.91	$2.01
Aggregate Travel Cost ($C(\hat{v})$)			$51,190	$59,297	$51,593
Toll Revenue ($R(\hat{v})$)			$2,712	$5,946	$4,015
			2 GP Lanes, 2 HOT Lanes; HOVs Ride Free		
Scenario			*(7'')*	*(8'')*	*(9'')*
GP Volume (x_G)			5,375.8	7,676.2	6,000.0
HOT Volume (x_H)			4,624.2	2,323.8	4,000.0
GP Speed (MPH)			36.3	16.2	29.8
HOT Speed (MPH)			44.2	58.7	50.0
Marginal VoT ($\hat{v}$)			$20.14	$29.52	$21.98
Toll (p)			$0.99	$13.23	$2.98
Aggregate Travel Cost ($C(\hat{v})$)			$51,167	$94,013	$53,846
Toll Revenue ($R(\hat{v})$)			$3,564	$17,150	$8,848

Notes: GP = general purpose; HOT = high occupancy toll; HOV = High occupancy vehicle; MPH = miles per hour; VoT = value of time; Cost Min = cost minimization; Rev Max = revenue maximization; Maintain LOS = maintain a minimum level of service.

produce different results and that these differences in performance grow as there is more capacity available to sell. When comparing outcomes under different tolling objectives, the analysis suggests that the revenue maximizing toll level will be above the toll that minimizes aggregate travel time cost. When the toll is set to maintain a target level of service, it may be set above or below the toll, which will achieve other goals depending on the level of service target. General statements about the ability of HOT lanes to generate revenue or reduce aggregate travel time cost will depend critically on the characteristics of the facility, including how it is priced.

While the model presented here is capable of predicting outcomes under a broad range of policies with limited information, it has some shortcomings that should be highlighted and which represent areas for future extensions. First, the approach assumes a fixed number of motorists use the facility during the period under study. In reality, transportation policies such as HOT lanes can influence a variety of driving behaviors, which make this assumption questionable. In particular, motorists may vary when they make trips, where they are going, and whether trips are made alone, with others, via transit, or not at all.

One modification to the model that can potentially allow total demand to vary involves enabling motorists to choose between using the GP lanes, HOT lanes, or a third option, which represents not traveling on the facility during the period being modeled. In this case, the model must be calibrated to reproduce actual travel counts and a reasonable demand response to changes in the monetary and travel time cost of driving. A mode choice decision could also be integrated into the modeling framework, as is done in Yang and Huang (1999). This adds complexity to the model and the requirement for information on the mode choice decisions of corridor users, which may not be available. Finally, it may be possible to integrate a departure time choice into the modeling and spread traffic between different modeling periods, as in Arnott et al. (1993).

Acknowledgements

This paper has benefited from feedback from Martin Wachs, the editor Kofi Obeng, a reviewer, and seminar participants at the 52nd Annual Transportation Research Forum and RAND's Logistics Seminar Series.

Appendix

Derivation of Eq. (6)

The value of time $\hat{v}^*$ of the indifferent user which minimizes aggregate travel time cost will solve the first order condition $\partial C(\hat{v}^*)/\partial\hat{v}^* = 0$. This first order condition is,

$$\text{(A.1)}\quad \begin{aligned}\partial C(\hat{v}^*)/\partial\hat{v}^* = {} & \frac{\partial t_G(x_G(\hat{v}^*))}{\partial x_G(\hat{v}^*)}\times N_S f(\hat{v}^*)\times x_G(\hat{v}^*)\times\frac{\int_{\underline{v}}^{\hat{v}^*} v\,dF_S(v)}{F_S(\hat{v}^*)} + t_G(N_S F_S(\hat{v}^*))\times\hat{v}^* N_S f(\hat{v}^*)\\ & -\frac{\partial t_H(x_H(\hat{v}^*))}{\partial x_H(\hat{v}^*)} N_S f(\hat{v}^*)\times\left(N_S(1-F_S(\hat{v}^*))\times\frac{\int_{\hat{v}^*}^{\overline{v}} v\,dF_S(v)}{1-F_S(\hat{v}^*)} + N_C\times\int_{\underline{v}}^{\overline{v}} v\,dF_C(v)\right)\\ & -t_H(x_H(\hat{v}^*))\times\hat{v}^* N_S f(\hat{v}^*) = 0.\end{aligned}$$

Dividing through by $N_S f(\hat{v}^*)$ and rearranging the terms, Eq. (A.1) can be rewritten as,

$$\text{(A.2)}\quad \begin{aligned}&\hat{v}^*\left[t_G(N_S F_S(\hat{v}^*)) - t_H(N_C + N_S(1-F_S(\hat{v}^*)))\right] =\\ &\frac{\partial t_H(x_H(\hat{v}^*))}{\partial x_H(\hat{v}^*)}\times\left(N_S(1-F_S(\hat{v}^*))\times\frac{\int_{\hat{v}^*}^{\overline{v}} v\,dF_S(v)}{1-F_S(\hat{v}^*)} + N_C\times\int_{\underline{v}}^{\overline{v}} v\,dF_C(v)\right)\\ &-\frac{\partial t_G(x_G(\hat{v}^*))}{\partial x_G(\hat{v}^*)}\times x_G(\hat{v}^*)\times\frac{\int_{\underline{v}}^{\hat{v}^*} v\,dF_S(v)}{F_S(\hat{v}^*)}.\end{aligned}$$

From the equilibrium requirement in Eq. (2), the toll that can induce $\hat{v}^*$ is given by

$$\text{(A.3)}\quad p^* = p^e(\hat{v}^*) = \hat{v}^*\left[t_G(N_S F_S(\hat{v}^*)) - t_H(N_C + N_S(1-F_S(\hat{v}^*)))\right].$$

Notice that both Eq. (A.2) and Eq. (A.3) equal $\hat{v}^*\left[t_G(N_S F_S(\hat{v}^*)) - t_H(N_C + N_S(1-F_S(\hat{v}^*)))\right]$. Thus, Eq. (A.2) and Eq. (A.3) are equivalent, implying that

$$\text{(A.4)}\quad \begin{aligned}&p^* = p^e(\hat{v}^*) = \hat{v}^*\left[t_G(N_S F_S(\hat{v}^*)) - t_H(N_C + N_S(1-F_S(\hat{v}^*)))\right] =\\ &\frac{\partial t_H(x_H(\hat{v}^*))}{\partial x_H(\hat{v}^*)}\times\left(N_S(1-F_S(\hat{v}^*))\times\frac{\int_{\hat{v}^*}^{\overline{v}} v\,dF_S(v^*)}{1-F_S(\hat{v}^*)} + N_C\times\int_{\underline{v}}^{\overline{v}} v\,dF_C(v)\right)\\ &-\frac{\partial t_G(x_G(\hat{v}^*))}{\partial x_G(\hat{v}^*)}\times x_G(\hat{v}^*)\times\frac{\int_{\underline{v}}^{\hat{v}^*} v\,dF_S(v)}{F_S(\hat{v}^*)}.\end{aligned}$$

Using the definition of $MEC_G(\hat{v})$ and $MEC_H(\hat{v})$ from Eq. (7) and (8) respectively, one arrives at Eq. (6), $p^* = MEC_H(\hat{v}^*) - MEC_G(\hat{v}^*)$. $MEC_G(\hat{v})$ and $MEC_H(\hat{v})$ represent the changes in aggregate travel time cost from a marginal increase in use in the GP and HOT lanes, respectively. For each vehicle class using either the GP or HOT lanes, it can be decomposed into the product of the change in travel time due to a marginal increase in use, the number of vehicles using the route, and the average value of time associated with those vehicles.

Derivation of Eq. (9)

The first order condition association with revenue maximization problem implies that the $\hat{v}^{**}$ that maximizes revenue satisfies the following first order condition

$$\text{(A.5)}\quad \partial R(\hat{v}^{**})/\partial\hat{v}^{**} = -N_S f_S(\hat{v}^{**})\times p^e(\hat{v}^{**}) + N_S(1-F_S(\hat{v}^{**}))\frac{\partial p^e(\hat{v}^{**})}{\partial\hat{v}^{**}} = 0.$$

From the equilibrium condition, Eq. (2), the toll that induces the revenue maximizing value of time of the indifferent user is given by $p^{**} = p^e(\hat{v}^{**})$. Thus,

$$\text{(A.6)}\quad -N_S f_S(\hat{v}^{**})\times p^{**} + N_S(1-F_S(\hat{v}^{**}))\frac{\partial p^e(\hat{v}^{**})}{\partial\hat{v}^{**}} = 0.$$

Next, adding $N_S f_S(\hat{v}^{**})\times p^{**}$ to both sides of Eq. (A.6) and dividing both sides by $N_S f_S(\hat{v}^{**})$ reveals Eq.

$$\text{(9)}\quad p^{**} = \frac{1-F_S(\hat{v}^{**})}{f_S(\hat{v}^{**})}\frac{\partial p^e(\hat{v}^{**})}{\partial\hat{v}^{**}}.$$

Endnotes

1. Because aggregate social welfare is not something that can easily be measured, other HOT lane pricing objectives tied to observable metrics such as revenue generation or speed are likely to be given greater weight in practice.

2. Other HOT lane facilities are currently operating on I-10 and US-290 in Houston, Texas; I-25 and US-36 in Denver, Colorado; I-95 in Miami, Florida; and I-680 southbound in Alameda and Santa Clara County, California.

3. Revenues from the SR-91 Express Lane facility were even higher in previous years, reaching $50 million in 2007 (Orange County Transportation Authority 2007).

4. Carpool formation is not explicitly modeled here because it adds considerable complexity to the analysis. For an example of how one might model carpool formation in the context of tolled facilities that provide free access to carpools, see Yang and Huang (1999).

5. It is generally assumed in the congestion pricing literature that the toll revenue generated from congestion pricing will be returned to citizens through tax reductions or other means. As a result, the toll cost experienced by users is exactly offset by the redistribution of toll revenue back to citizens in welfare calculations. That is, the collection and redistribution of revenues effectively cancel each other out and are ignored in welfare calculations.

6. The distribution of the values of time for the SOV users that use the GP lanes reflects the left-hand side of the overall SOV value of time distribution, while the distribution of values of time for the SOV users that opt to buy into the HOT lanes represents the right-hand side of the overall SOV value of time distribution. For both groups of SOVs (i.e., those that travel in the GP lanes and those that travel in the HOT lanes), the respective value of time distributions represent a truncated distribution, with the truncation occurring on either the right or left of $\hat{v}$. The calculation of the mean value of time for each group reflects this truncation.

7. The derivation of Eq. (6) is in Light (2009) for the case where all vehicles, regardless of the number of occupants, must pay the toll to access a set of priced lanes that run parallel to a set of free GP lanes. Verhoef et al. (1996) derive a related expression for optimal toll on a priced route when there is a free alternative. In their analysis, demand is elastic but users are assumed not to vary in their value of time, and there is no distinction made between carpool and non-carpool vehicles.

References

Arnott, Richard, Andre de Palma, and Robin Lindsey. "A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand." *American Economic Review* 83 (1), (1993): 161-179.

Bureau of Public Roads. Traffic Assignment Manual. U.S. Dept. of Commerce. Washington, D.C., June 1964.

Cambridge Systematics. Twin Cities HOV Study: Volume 1. http://www.dot.state.mn.us/information/hov/pdfs/full_study.pdf, March 14, 2010.

Dahlgren, Joy. "High Occupancy Vehicle Lanes: Not Always More Effective Than General Purpose Lanes." *Transportation Research Part A: Policy and Practice* 32 (2), (1998): 99-114.

Federal Highway Administration. A Guide for HOT Lane Development. http://ntl.bts.gov/lib/jpodocs/repts_te/13668_files/images/13668.pdf, March 15, 2010

Federal Highway Administration. HOV/HOT Lanes. http://tmcpfs.ops.fhwa.dot.gov/meetings/uploaded_files/HOV%20HOT%20Lanes%20Brochure.pdf, March 15, 2010.

Hensher, David. "Influence of Vehicle Occupancy on the Valuation of Car Driver's Travel Time Savings: Identifying Important Behavior Segments." *Transportation Research Part A: Policy and Practice* 42 (1), (2008): 67-76.

Light, Thomas. "Optimal Highway Design and User Welfare Under Value Pricing." *Journal of Urban Economics* 66 (2), (2009): 116-24.

Liu, Louie and John F. McDonald. "Efficient Congestion Tolls in the Presence of Unpriced Congestion: A Peak and Off-Peak Simulation Model." *Journal of Urban Economics 44* (3), (1998): 352-66.

Metropolitan Transportation Commission. Bay Area High Occupancy Toll (HOT) Network Study December 2008 Update. http://www.mtc.ca.gov/planning/hov/12-08/HOT_Network_Report_12-08.pdf, March 18, 2010.

Orange County Transportation Authority. "91 Express Lane 2009 Annual Report." http://www.91expresslanes.com/generalinfo/91annualreport.pdf, April 5, 2010.

Orange County Transportation Commission. "91 Express Lane 2007 Annual Report." http://www.octa.net/91annual07b/test.pdf, April 7, 2010.

Poole Jr., Robert W. and Ted Balaker. "Virtual Exclusive Busways: Improving Urban Transit While Relieving Congestion." Policy Study No. 337 (Los Angeles: Reason Foundation, September 2005).

Safirova, Elena, Kenneth Gillingham, and Sebastien Houde. "Measuring Marginal Congestion Costs of Urban Transportation: Do Networks Matter?" *Transportation Research Part A: Policy and Practice* 41 (8), (2007): 734-749.

Small, Kenneth. *Urban Transportation Economics. Vol. 51 of Fundamentals of Pure and Applied Economics series*, Harwood Academic Publishers, 1992.

Small, Kenneth, Clifford Winston, and Jia Yan. "Differential Road Pricing, Express Lanes and Carpools: Exploiting Heterogeneous Preferences in Policy Design." *Brookings-Wharton Papers on Urban Affairs*, (2006): 53-96.

Small, Kenneth, Clifford Winston and Jia Yan. "Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability: Implications for Road Pricing." *Econometrica* 73 (4), (2005): 1367-1382.

Small, Kenneth and Jia Yan. "The Value of Value Pricing' of Roads: Second-Best Pricing and Product Differentiation." *Journal of Urban Economics* 49 (2), (2001): 310-336.

Sullivan, Edward and Mark Burris. "Benefit-Cost Analysis of Variable Pricing Projects: SR-91 Express Lanes." *Journal of Transportation Engineering 132 (*3), (2006): 191-198.

Turnball, Katherine. "HOV, HOT Congestion Pricing in the United States." *European Transport Conference Proceedings*, (1997): 97-104.

Verhoef, Erik, Peter Nijkamp, and Piet Rietveld. "Second-Best Congestion Pricing: The Case of an Untolled Alternative." *Journal of Urban Economics* 40 (3), (1996): 279-302.

Verhoef, Erik and Kenneth Small. "Product Differentiation on Roads: Constrained Congestion Pricing with Heterogeneous Users." *Journal of Transport Economics and Policy* 38 (1), (2004): 127-156.

Washington State Department of Transportation. SR 167 HOT Lanes Pilot Project First Annual Performance Summary: May 2008 – April 2009. http://www.wsdot.wa.gov/NR/rdonlyres/31FB3D24-79CC-4332-82F7-EBECEBE1CA71/0/HOTLanesAnnualReport2009.pdf, March 11, 2010.

Yang, Hai and Hai-Jun Huang. "Carpooling and Congestion Pricing in a Multilane Highway with High-Occupancy-Vehicles." *Transportation Research Part A: Policy and Practice* 33 (2), (1999): 139-155.

***Thomas Light** is an economist at RAND and a professor of economics at the Pardee RAND Graduate School. His research focuses on public finance issues in the transportation and energy sectors. Prior to joining RAND, Light worked at ECONorthwest, where he was involved in over a dozen high occupancy toll lane feasibility studies. He holds a Ph.D. in economics from Cornell University.*