High Occupancy Toll Lanes: Do They Reduce Congestion?

David Wang

Professor Charles Becker, Faculty Advisor
Professor Michelle Connolly, Faculty Advisor

Honors thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University

Duke University
Durham, North Carolina
2014

Acknowledgements

I would like to thank Professor Charles Becker and Professor Michelle Connolly for their copious and invaluable comments in guiding me through the research process. I would also like to thank my peers in two semesters of thesis seminar class for their helpful suggestions and challenging questions. Lastly, I would like to extend many thanks to my friends and family for supporting me through this process.

JEL classification: R41, R48

Keywords: HOV lanes, HOT lanes, Transportation economics, Tolls, Congestion Pricing

Table of Contents

Table of Figures

I. Introduction

In 2009, according to data from the American Community Survey, ninety percent of workers in the U.S. used a privately owned vehicle when commuting. For an average commuter, the annual traffic delay in urban areas has increased from below fifteen hours in 1982 to more than thirty-five hours in 2007 (Winston, 2013). Furthermore, the annual cost of congestion, including travel delays and fuel expenditures, exceeds $100 billion a year (Winston, 2013). From a welfare standpoint, these travel delays cause a total welfare cost of $45 billion a year (Langer, Winston, & Baum-Snow, 2008). Governments have considered a variety of solutions to combat this congestion, the most prevalent being high occupancy vehicle (HOV) lanes and congestion pricing, including high occupancy toll (HOT) lanes.

The federal government heavily encouraged the construction of HOV (high occupancy vehicle) lanes, passing the Intermodal Surface Transportation Efficiency Act of 1991. It was thought that the speed differential between HOV lanes and general-purpose (GP) lanes would lead drivers to switch to carpools, thereby reducing the number of vehicles on the roads and the amount of congestion. However, in practice, HOV lanes have not been very successful and single occupancy vehicle (SOV) users often complain about underutilized HOV lanes. These observations mirror transportation researchers' criticisms of the ineffectiveness of HOV lanes in reducing congestion. Dahlgren (1998) argues that adding a GP lane to existing highways is more effective at lowering delay costs than adding an HOV lane.

Another strategy, congestion pricing, involves placing a price on using roadways to offset the social cost arising from such use. Under Vickrey's theory, the charges should match the marginal social cost of each trip as closely as possible (Vickrey, 1963). In a standard highway with only GP lanes, the personal cost of traveling, in the form of the value of time, often does not equal the social cost imposed on other commuters on the highway. This scenario gives rise to a mismatch in incentives and to a tragedy of the commons. In actual application, few tolling schemes for congestion pricing exist. Since technology has made collecting tolls cheaper, the main challenge now is public resistance. Current implementations of congestion pricing include Singapore's Electronic Road Pricing system and U.S. HOT lanes.

High occupancy toll (HOT) lanes have potential as a politically feasible policy to improve utilization of HOV lanes and generate revenue. HOT lanes give solo drivers the option to pay a toll for use of HOV lanes. The HOT lanes help address several issues, for example balancing the load and reducing congestion by shifting some solo drivers from GP lanes to HOV lanes, giving drivers the option of traveling on less congested lanes, and generating revenue for highway operators (Poole & Orski, 2000). Many studies of HOT lanes use California State Route 91 as an example of a HOT highway and seek to model any welfare gains from its implementation (Liu & McDonald, 1998; Small & Yan, 2001). Opened in 1995, the SR-91 serves as a good case study because it was one of the first HOT operations in the U.S. The literature also explores the distributional effects of congestion pricing over the Washington, D.C., metropolitan area, looking at a network of roads rather than a single one (Safirova et al., 2004). They emphasize that much of the benefit from HOT lanes comes from undoing the inefficiency created by existing HOV lanes. Safirova et al.'s study is one of few empirical ones in the literature. Nevertheless, theoretical models, such as that by Konishi and Mun (2010), could be used as a basis for future empirical studies. The economics literature covers welfare gains, but does not address empirically how the congestion reduction from HOV and HOT lanes has changed over time. Although theory may predict welfare gains in the short term, it is unclear what the long run effects may be. In this paper, we seek to examine the short-run effects of the conversion of HOV to HOT lanes on highway congestion.

This paper has six main sections. Section II reviews the relevant literature on HOV and HOT lane efficiency. Section III describes highway usage data from the California Department of Transportation and Section IV presents a theoretical model. Section V gives an empirical specification. Section VI discusses the results and Section VII concludes with policy implications and thoughts for future research.

II. Literature Review

The literature on carpooling and congestion pricing comes from engineering, policy, and economics backgrounds. These varied perspectives provide different methods for analyzing congestion. It is well established that the socially efficient solution for optimizing congestion results from first-best marginal cost pricing (Yang & Huang, 1999). In a first-best pricing policy, all lanes and alternative roads are charged tolls. However, the first-best solution is not politically feasible, as the public is unlikely to consent to free highways becoming toll roads. Consequently, most of the literature on congestion pricing focuses on second-best solutions, in which tolls or travel restrictions apply to only part of the routes, not all. In other words, travelers must have a way to move from point A to B without tolls. To measure how much maximum efficiency is gained, the literature compares various second-best pricing schemes to first-best and no-toll schemes.

Early literature is mostly theoretical and focuses on whether high occupancy lanes (HOV) lanes are more effective than general-purpose (GP) lanes. Dahlgren (1998) finds that HOV lanes are more effective than GP lanes only if there is a significant travel time differential (35 min or more) between the two lanes and if the HOV is well used (at least 20% of cars). For most situations, due to opposing factors, adding a GP lane is more effective at reducing congestion. The incentive to shift from GP to HOV depends on the difference in travel times between the two lanes. As people shift to HOVs, the difference decreases and fewer will be motivated to switch. Dahlgren uses queuing theory with a logit model to explain her findings and does not make any statements regarding the effects of HOV choice on social welfare. Later literature incorporates congestion pricing and the travel cost associated with congestion. Liu and McDonald (1998) conclude that socially optimal price of second-best tolls are quite low and that welfare gains are small compared to those in a first-best pricing regime, where all lanes are subject to tolls. In their model, travelers have homogeneous preferences and demand as a function of the travel cost, which depends on the traffic volume across two periods (peak or non-peak), and tolls. These travelers have the choice of two parallel roads connecting the same two points, except one road requires tolls and the other is free. Finally, for simplicity, they ignore carpool vehicles. They also conclude that a profit-maximizing toll scheme causes the largest welfare loss, due to the high tolls charged.

Extending the previous study, Small and Yan (2001) claim that accounting for heterogeneous preferences in the value of time improves the performance of differential price policies. The highway model and demand function are the same as in Liu and McDonald (1998), except it adds a parameter for value of time and uses only a single period. As expected, a higher value of time results in more people choosing to pay the toll and improves the performance of second-best pricing policies. However, the results still indicate that these second-best policies are poor substitutes for more thorough first-best policies.

More recently, Konishi and Mun (2010) explore in detail the theoretical effect of converting HOV lanes to HOT lanes. They also use a model of commuters using a highway with multiple lanes. However, they differ by making commuters heterogeneous in carpool organization costs. The costs of travel are then a sum of the carpool costs, travel cost as a function of traffic volume, and toll prices. They conclude that converting HOV lanes to HOT lanes reduces traffic in the GP lanes, but the additional traffic in the HOV lane may cause the conversion to not be Pareto improving. Although they make many simplifying assumptions, such as ignoring the heterogeneity in value of time, the reinvestment of revenues from tolls, or the dependence of carpool costs on the size of the carpool, their model provides a solid basis to build upon. Further work can relax these assumptions.

Although most of the literature is theoretical, a couple of empirical studies on the welfare effects of congestion pricing do exist. Small et al. (2006) once again study SR-91, but use survey data of travelers that use the highway to build an estimated demand function through discrete choice modeling. They use many parameters and interaction terms to account for perceived heterogeneity in preferences, including socioeconomic factors and characteristics of the journey, such as trip distances, tolls, travel times, and the reliability of the travel times, to estimate dollar values of time and reliability. Using 1,124 observations across two revealed preference samples and one state preference sample, the model estimates the marginal utilities of the toll costs, median travel times, and travel reliability, where the toll costs and travel information are observed for time periods corresponding to those covered by the survey data. The value of time and value of reliability are calculated as ratios of marginal utilities of travel time and reliability, respectively, to the marginal utility of toll cost. They find that the median value of time is $19.63 per hour and the median value of reliability is $20.76 per hour, only slightly below the average wage rate of $23 per hour in the sample area. Using the estimated demand function and a standard U.S. Bureau of Public Roads supply side congestion model, they simulate changes in consumer surplus for various pricing and operational policies. These policies include no toll, HOV, HOT, one-route toll (one tolled lane with no discount for carpools), two-route toll (two tolled lanes at different levels), and two-route HOT (same as two-route toll and free for carpools). To categorize the policies, HOT and one-route tolls are second-best policies, while the two-route tolls are first-best policies. Their simulation results are in line with previous theoretical papers that find that the first-best policy produces the largest welfare gain over HOV and HOT policies. However, they suggest that a limited two-route HOT, in which tolls are charged on all lanes unless driving a carpool, is the best policy to compromise between full first-best pricing and less efficient HOV or HOT policies.

Small et al.'s (2006) methods advanced the literature by adding empirical results to theoretical models. However, their empirical methods are difficult to replicate without similar survey data on driver choices. Demographic factors likely affect the degree to which people change their behaviors in response to policy shocks, so an empirical model of demand may be a better fit than a theoretical model with many simplifying assumptions. The massive California Department of Transportation's Performance Management System database provides opportunities for study of traffic patterns resulting from changing HOT policies. This paper fills the gap in the empirical economic literature on the short-term effects of HOV to HOT conversion on congestion, using raw traffic data.

III. Data

This study uses traffic data from the California Department of Transportation (Caltrans) Performance Measurement System (PeMS). This database provides real-time vehicle volume counts and travel speeds from detector stations placed along nearly all major highways in California. Depending on the state transportation district and highway, the data range from 2001 to January 2014. Due to the large size of the dataset (millions of data points), this paper uses only data from one to three years leading up to and immediately following the introduction of HOT lanes. The PeMS traffic data provides a real-world measure of traffic spanning enough time to analyze the highways before the construction of HOT lanes, a base case scenario.

In this section, I first describe the physical characteristics of the HOT lanes and then explain the construction of the dataset. Although the first HOT lane opened in 1995 in California, only a few California HOT lanes are in operation today, including SR-237 and I-680 in the San Francisco Bay Area, SR-91 in Orange County, I-15 in San Diego, and I-10 and I-110 in Los Angeles County. However, the SF Bay Area's Metropolitan Transportation Committee plans to build to 550 miles of express lanes by 2035, most of which will be converted from HOV lanes (BAIFA, 2014). The large government interest in the performance of HOT lanes as a congestion reduction tool partially motivates this study. I consider a wide cross sectional study of multiple highways, but each HOT system has unique characteristics that make such a study difficult with limited data on non-California highways. Therefore, in this study, I focus on one highway in the SF Bay Area, I-680 Southbound.

Table 1. Highways used in study

Highway	HOT Date	HOT Times	Peak Traffic	HOT Length	HOT Lanes	GP Lanes	Previous System	Barrier
I-680S	September, 20 2010	5 a.m.-8 p.m	5 a.m.-9 a.m.	14 mi	1	3-4	HOV	Double line

The I-680 Southbound HOT lane spans 14 miles from SR-84 to the SR-237 freeway junction, with a total of three entry points and three exit points.[1] The HOT lane is in operation from 5 a.m. to 8 p.m. Monday through Friday, but peak traffic occurs in the morning from 5 a.m. to 9 a.m. The HOT lane employs two solid white lines to separate the lane from general-purpose (GP) lanes. Under state law, drivers cannot cross these white lines, but no physical barrier exists to prevent drivers from crossing into the lane. California Highway Patrol enforces the lane visually through physical patrols. Prior to conversion to HOT, the lane operated as a HOV2+ lane, requiring two passengers per vehicle for use. Under HOT policy, cars with two or more passengers keep free access to the lane, but single occupancy vehicles gain the option to pay for access. Drivers pay tolls upon entry at specified entry points, at which radio tag readers detect the radio transponders adhered to the vehicle windshields. Tolls vary by traffic conditions and distance traveled.[2] Both before and after HOT opening, the stretch of highway upstream and downstream from the 14-mile stretch of HOV/HOT lane contains only GP lanes.

Construction on I-680 began in October 2008 and continued with light landscaping after the HOT lane officially opened in September 2010 (AlamedaCTC, 2012). To account for the traffic prior to construction, this paper uses traffic data compiled in one-hour increments from October 2007 to December 2013, but focuses on the periods before and after construction to avoid any traffic disruption due to construction. A finer measure of construction delay was not available through Caltrans archives or news sources. To include the effects of the HOT lane on downstream

---

[1] See Appendix 1.
[2] On I-680S, tolls range from $1.00 to $7.50 during the morning peak hours and $0.30 to $7.50 during off peak hours. The toll is set to maintain a 45 mph speed in the HOT lane. Detectors at each entry point allow for computation of distance traveled, which varies due to the three entries and exits.

(after the lane) traffic, the data spans an extra eight miles south (downstream) beyond the express lane.[3]

I code the dataset to reflect information related to the highways, but not directly included in the PeMS database. To control for the number of cars that travel on the two highways independently of the HOT policy, I include yearly population counts for the Alameda and Santa Clara counties in California, where I-680S is located. This data comes from the American Community Survey. In addition, I denote HOT section or HOT portion as the locations of the sections of highway that will have (prior to HOT existence) or have HOT lanes. On these two highways, all HOT lanes are converted from HOV lanes.

Detector stations located at points along the highway record the traffic data across lanes. The stretch of I-680 under study has 52 detector stations, but only five were in operation in 2007, before construction on the HOT began. Since the majority of new detectors in the HOT section of the highway were installed after the start of HOT construction, panel analysis has selection bias if all stations are used. Therefore, I only keep observations from five stations, the ones that exist through the whole period of interest, 2007-2014. Furthermore, the detectors are not perfectly balanced and do not have perfect data quality, so I drop observations that are unbalanced or recorded incorrectly.[4]

---

[3] Upstream detectors are located prior to an interchange with another major highway, I-580, and are left out to avoid any unobservable effects due to the interchange.

[4] New detectors were installed along the highway with the majority installed in the HOT section after 2010, the opening of the HOT lane. Therefore, the dataset is unbalanced and may have selection bias. Visual inspection reveals that at between certain dates, traffic patterns repeated exactly with same flow every Monday, every Tuesday, etc. These values are likely recorded incorrectly.

Table 2. PeMS Variable Definition

PeMS Data Variable	Description	Unit
Station Length	Highway length covered by the station as measured from midpoints between detector stations.	Miles
Total Flow	Total number of vehicles, *across all lanes*, passing through the detector station in the hour.	Vehicles/ Hour
Average Speed	Flow-weighted average of speeds *across all lanes*.	Mph
Delay (V_t=60)	Sum of extra time spent by all vehicles to travel across detector station length above the time it takes at 60 mph.	Vehicle-Hour (In Hours)
Lane N Flow	Total number of vehicles, *for lane N*, passing through the detector station in the hour.	Vehicles/ Hour
Lane N Average Speed	Flow-weighted average of speeds *for lane N*.	Mph

The dataset for I-680S has approximately 86,000 observations across four detector stations. The dataset contains enough observations that dropping some is not a problem. The detector stations provide key variables of interest for this study: total flow, lane flow, average speed, lane speed, and time delay. Table 2 summarizes these measurements. The flow variable measures the number of vehicles passing the detector station in the hour, in total across all lanes or on a per lane basis. The average speed measure gives the flow-weighted speed across all lanes at a detector station, giving a weighted average speed by each car that passes through. The delay measure at a given detector station shows the amount of extra time required to travel through the detector station's highway section when traveling at speeds below a threshold. In this study, I focus on delay at a threshold of 60 mph, the expected free flow speed (Kwon & Varaiya, 2008).

Tables 3 through 6 present sample summary statistics for the above measure. I check the dataset in detail to identify potential outliers using graphical analysis, a sample provided in the appendix. In addition, the summary statistics give insight into whether certain values may be outliers. Tables 3 and 4 present statistics for the flows, average speed, and delay for all detectors along the

Table 3. Summary of Traffic Before HOT Exists on I-680S at 8 AM, All Observations[5]

N = 914	Mean	Median	Std. Dev.	Min	Max
Total Flow (Vehicles/Hr)	4682.2	4675.5	913.6	581.0	6230.0
HOV/T Flow	1208.4	1247.0	276.6	95.0	1627.0
GP Lane Flow	1400.7	1455.0	231.3	162.0	1690.0
Average Speed	61.9	63.1	5.6	33.2	75.8
HOV/T Speed	67.3	68.3	4.5	42.4	77.7
GP Lane Speed	59.9	61.2	6.0	30.1	74.8
Delay at 60 mph Threshold	11.5	2.9	20.4	0.0	263.0

Table 4. Summary of Traffic After HOT Exists on I-680S at 8 AM, All Observations[6]

N = 1246	Mean	Median	Std. Dev.	Min	Max
Total Flow (Vehicles/Hr)	5269.4	5936.0	1676.7	12.0	7821.0
HOV/T Flow	1116.1	1219.5	388.4	0.0	2160.0
GP Lane Flow	1246.2	1334.0	279.1	0.0	1629.5
Average Speed	63.3	65.5	7.0	27.2	75.2
HOV/T Speed	68.8	70.5	6.7	21.1	81.9
GP Lane Speed	61.0	63.4	7.1	29.6	72.6
Delay at 60 mph Threshold	3.3	0.2	8.9	0.0	173.1

HOT portion of I-680S at 8 AM on weekdays. The 8 AM hour is during peak morning traffic and should have positive traffic flows, even if heavily congested. However, the minimum value of total flow is zero in Table 4. These points may reflect traffic incidents, construction periods, or incorrectly recorded data. Therefore, I define an acceptable range using two standard deviations from the mean value. I calculate separately the means by detector station, the existence of the HOT lane, hour of day, and weekday. This calculation runs within groups.

[5] Summary statistics restricted to weekdays at 8 AM on the section of highway that will have or has HOT lane.
[6] Ibid.

Table 5. Summary of Traffic Before HOT Exists on I-680S at 8 AM, Outliers Removed

N = 849	Mean	Median	Std. Dev.	Min	Max
Total Flow (Vehicles/Hr)	4806.9	4696.0	723.1	2580.0	6230.0
HOV/T Flow	1238.1	1255.0	239.0	504.0	1627.0
GP Lane Flow	1433.3	1466.0	166.5	803.0	1690.0
Average Speed	62.4	63.1	4.6	49.4	70.6
HOV/T Speed	67.7	68.4	3.5	56.3	74.0
GP Lane Speed	60.4	61.3	4.8	45.2	70.5
Delay at 60 mph Threshold	9.3	2.8	13.3	0.0	79.0

Table 6. Summary of Traffic After HOT Exists on I-680S at 8 AM, Outliers Removed

N = 1133	Mean	Median	Std. Dev.	Min	Max
Total Flow (Vehicles/Hr)	5493.2	6102.0	1480.4	1642.0	7821.0
HOV/T Flow	1166.0	1236.0	317.5	147.0	2020.0
GP Lane Flow	1294.6	1345.0	196.1	533.8	1629.5
Average Speed	63.5	65.7	6.7	38.6	75.2
HOV/T Speed	69.0	70.6	6.2	46.4	79.4
GP Lane Speed	61.2	63.5	6.9	35.0	71.0
Delay at 60 mph Threshold	2.9	0.2	6.9	0.0	32.0

IV. Theoretical Framework

The objective of this study is to examine the impacts on congestion of the conversion of HOV lanes to HOT lanes. Many metropolitan areas in the U.S. have adopted HOT lanes, including the Silicon Valley, Los Angeles, Minneapolis, Miami, Seattle, Denver, Salt Lake City, San Diego, Houston, and Orange County (Metro, 2012). Additional cities are now evaluating new HOT lane polices, including Raleigh for I-40.

The most prevalent studies in transportation research employ two methods to analyze the effects of HOV and HOT lanes. One method examines how traffic patterns change following the introduction of HOV/HOT lanes, using performance measures such as speed, travel time, vehicle and person throughput, bottlenecks, vehicle occupancy, and lane violations (Kittelson Associates, 2013). This method is primarily empirical, using data collected by traffic sensors and visual observation. For HOV analysis, these studies appear on both an aggregate level across highways and an individual level for a single highway. However, due to the lack of widespread HOT adoption, HOT studies typically focus on single highways. Another common model analyzes driver decisions given choices of carpooling, paying tolls, driving in the GP lanes, or not driving at all. These studies are usually theoretical and use a representative highway. Outcomes of various highway policies and designs are simulated based on variations in travel times, reliability, carpool costs, toll costs, and driver preferences.

While any researcher aims to find a solution that maximizes efficiency, the actual definition of this efficiency can vary greatly. Traditionally, transportation planners aim to maximize traffic flows, by either reducing the total number of drivers on the roads or making sure the drivers on the roads move through more quickly. However, maximizing traffic flows does not necessarily maximize social welfare. If the marginal driver’s benefit does not exceed the travel time cost he imposes on other drivers, his addition to the traffic flow would not be welfare improving. Thus, planners have begun to account for driver and social welfare when evaluating new projects. Examining drivers' demand functions for highway use is particularly relevant in the analysis of HOT lanes, since the tolls set directly affect total travel costs.

In this paper, I seek to examine whether HOT lanes improve congestion and therefore social welfare. Thus, I define increased efficiency as a reduction in aggregate net travel costs across all drivers, assuming homogeneous preferences. For a single potential driver, the net travel costs include the opportunity cost of time, toll costs, and lost benefits if the trip does not occur. The standard demand-side framework of congestion pricing considers a two-route highway. In this two-route model, one, none, or both routes can be subject to tolls. Users differ in their demand for travel and their losses of utility from congestion. However, for simplicity I impose homogenous losses of utility from congestion. Equilibrium occurs under two scenarios, when the user costs of using either route are equal (the user is indifferent to the two options) or when only users with positive net benefits make the trip (value from the trip exceeds the user cost). The user cost is the sum of the travel time and toll costs.

Later models relax the assumption that users have the same loss of utility from travel delay. These models account for drivers' heterogeneous preferences through a value of time (VOT) measure. The VOT captures as a monetary value the lower bound for a traveler's loss of utility from the travel delay. This metric is important for calculating user welfare and comparing with tolls and taxes. Empirically, Small et al. (2006) apply a mixed logit model to both state and revealed preference data to estimate this value of time and also a value of reliability (VOR). Many studies use such estimates when calculating the total benefit due to travel time savings from HOV or HOT lanes.

HOT lanes allow for more flexible levels of price discrimination than HOV lanes do because the tolls directly vary the cost of using the lanes. With HOV lanes, the cost of carpooling often exceeds any benefit derived from the savings in travel time, so the lanes are underutilized. HOT lanes can increase utilization of the HOV lanes by creating travel costs that are less than or equal to drivers' willingness to pay. A driver's decision between a GP and HOV lane involves two factors, the carpool cost and his savings from the reduced travel time. The decision between a GP and HOT lane includes three factors: the toll cost, potential carpool cost, and travel time savings. A highway operator cannot control an individual driver's carpool cost, but can set the toll cost. Thus, we expect a highway with an HOT system to allow for more optimal pricing than an HOV system. If the HOV system exhibits underutilization due to improper pricing, then the HOT system should increase utilization and traffic flow.

V. Empirical Specification

Before estimating the effect of introducing an HOT lane on highway traffic, I first define a proxy for the level of congestion. Much of the literature uses vehicle flow and average speed as proxies for congestion. These measures are used to calculate travel times and traffic delay, which also form good proxies for congestion. The vehicle flow measures the total number of cars passing a detector in a set time interval. The average speed is the mean speed at which vehicles are traveling, across all lanes at a detector. Finally, traffic delay is the increase in travel time across a length of highway that results when the average speed drops below 60 mph, the standard free flow speed (Kwon & Varaiya, 2008). Assuming excess demand exists for a highway, higher vehicle flow, higher average speed, and lower delay signify less congestion. These metrics form a relatively complete picture of the movement of automobiles on the road, but cannot directly capture drivers' decisions. For example, severe congestion may deter drivers from a certain route or minimal traffic may attract drivers. Due to this circularity, any analysis of changes in traffic should control for changes in demand for the route.

Therefore, I use the vehicle flow measure as a proxy for drivers' choices to drive on a certain highway. This vehicle flow is endogenous to the model, so estimation follows a two-stage least squares regression with fixed effects. The first stage regresses total flow against instruments of county population, HOT opening, HOT location, and controls for time, including the year, month, day of the week, and hour of the day. These measures are uncorrelated with the error terms of the model, as drivers do not set the HOT policy or the time. There is a possibility that the location of the HOT lane may be endogenous for having high traffic. However, I do not examine a cross section of highways so this should not be an issue. In addition, the flow is a nonlinear measure of traffic. A value of zero flow could result from two scenarios: no traffic, when no cars are on the roads or heavy traffic, when cars are on the roads but cannot get through. I control for this bimodal distribution through hour dummies, which help capture the two modes that result from peak and off-peak traffic. The fixed effects estimation allows me to control for some time invariant unobservable highway characteristics, such as the quality of the roads, physical characteristics, and destinations of the vehicles for each station length along the highway. For characteristics that do change over time, I code the changes using dummy variables. The following regressions are the culmination of iterations of tests and theory. To ensure that the model only captures the effect of the HOT conversion on traffic and no other traffic factors away from the HOT lane, I further restrict the dataset to only the HOT section of the highway. The HOT section is the portion of highway that will have or has HOT lanes after conversion.

V.I Congestion Equations

I include two congestion equations to compare the effect of having an instrument for flow to a nonlinear effect of the HOT lane existing. Equations (1) and (2) specify fixed effects regressions that show the effect of HOT policy on congestion. Average speed and average delay serve as proxies for congestion. The average delay measure incorporates both vehicle speeds over the detector and the distances traveled to calculate the time delayed.[7] Formally, the average delay is the increase in travel time required compared to travel at speeds of 60 mph.

Equation (1):

$$\begin{aligned} Congestion = \alpha &+ \beta_1 HOT\ exist_{it} + \beta_2 HOV/T\ FlowIV_{it} + \beta_3 HOV/T\ FlowIV_{it}^2 \\ &+ \beta_4 GP\ FlowIV_{it} + \beta_5 GP\ FlowIV_{it}^2 + \gamma_1 \#\ of\ Lanes_{it} \\ &+ \gamma_2 County\ Population_{it} + \sum_{j=1}^{n} \delta_{ij} Time\ Dummies_{ij} + u_{it} \end{aligned}$$

[7] Distance defined by the length of highway observed by the detector, measured from the midpoints between detectors.

Equation (2):

$$Congestion = \alpha + \beta_1 HOT\ exist_{it} + \gamma_1 \#\ of\ Lanes_{it} + \gamma_2 County\ Population_{it} + \sum_{j=1}^{n} \phi_{ij} HOTexist_{it} * Hour\ Interactions_{ij} + \sum_{k=1}^{m} \delta_{ij} Time\ Dummies_{ik} + u_{it}$$

HOT exist is a dummy variable that equals one after the HOT is constructed and operating. Eq. (1) controls for the number of vehicles on the highway through the *FlowIV* instrumental variables. Flow measures the number of vehicles passing detector station *i* at time *t*. *HOV/T FlowIV* is the flow on the HOV/T lane and *GP FlowIV* is the flow on the GP lanes at the detector station. Since the number of cars, flow, on the road is endogenous to the level of congestion, I use a two stage least squares regression to estimate Eq. (1). I include a $FlowIV^2$ term to capture the effect of variable pricing on congestion. As congestion rises, not only does the travel time cost increase, but also the tolls on the HOT lane increase in order to maintain speeds of at least 45 mph in the lane. Therefore, the price of driving on the highway is nonlinear. In Eq. (2), the interaction of *HOT exist* and the hour dummies reflect what the $FlowIV^2$ terms capture, that the effect on congestion after the HOT exists should vary by the hour, since tolls will vary with the given level of congestion and thus changes driver cost. *# of Lanes* is the number lanes at detector station *i* at time *t*. *County Population* is the sum of yearly ACS estimates of the populations of the local counties.[8] I treat this value as a sum to avoid the multicollinearity issue between the populations of the local counties. *Time Dummies* are dummy variables for the hour of day, day of the week, month, and year.

[8] American Community Survey, see Table A1 for estimates.

V.II Estimating the Vehicle Flow Instrumental Variable

Since the number of cars on the road is endogenous to the model, Equation (3) instruments for flow using various exogenous variables. The exogenous instruments used in this equation are also present in the main regression (1). These variables are valid instruments because they are correlated with *Flow*, but not correlated with the error terms in Eq. (1).

Equation (3): Stage 1 of 2SLS Fixed Effects Regression Equation

$$Flow_{it} = \alpha + \pi_1 HOT\ exist_{it} + \pi_2 \#\ of\ Lanes_{it} + \pi_3 County\ Population_{it} + \sum_{j=1}^{n} \phi_{ij} HOT\ exist_{it} * Hour\ Interactions_{ij} + \sum_{k=1}^{m} \delta_{ij} Time\ Dummies_{ij} + v_{it}$$

The dependent variable, *Total Flow*, measures the number of vehicles passing detector station *i* at time *t*. *HOT exist* is a dummy variable that equals one after the HOT is completed and operating. *# of Lanes* is the number of lanes at detector station *i* at time *t*. *County Population* is the sum of yearly ACS estimates of the populations of the local counties. I control for time of day, day of the week, and seasonality by including dummy variables for the hour, day of week, month, and year. The interaction terms control for different marginal effects of the HOT lane on total flow at each hour.

VI. Results

I first estimate Eq. (3) as the first stage of the 2SLS Fixed Effects Regression. Table 7 presents results from the estimation of Eq. (3). The table shows that for I-680S, vehicle flows in the HOV/T and GP lanes decrease following the opening of the HOT lane. This result is not consistent with the positive increase predicted by the theoretical model and within the literature. Since an HOT lane opens capacity to the HOV lane, overall capacity on the highway should increase and allow higher throughput of vehicles. Instead, after the HOT starts its initial operation, predicted flow in the HOT lane during the 6 AM hour decreases by 253.4 (-356.3+102.9) vehicles per hour. Both components of the net decrease are significant, so I expect the net amount to be significant to the 1% level as well. For 7 AM, the drop in flow in the HOT lane is the largest, but the value is not significant. Considering the expanded capacity of the HOT lane, it is unusual that flow drops at all in the lane. For the GP lane, predicted values reflect the average un-weighted flow across the general-purpose lanes. Column (2) predicts a similar drop in flow after HOT policy. As in Column (1), the predicted flow is lowest for 7 AM and highest for 9 AM, but the net effect is a decrease at all hours 6, 7, 8, and 9 AM.[9] For example, for 7 AM, the predicted drop in flow is 157.6 (-232.8-75.2) vehicles per hour. It is unclear what drives the decrease in flow despite a supply improving policy change. Furthermore, the County Population measure is not significant and has a very small effect on the flow of vehicles. In addition, the *# of lanes* estimates are positive and significant at the 1%. This measure implies that for every additional lane constructed, total flow increases by 148.5 vehicles per hour. Ceteris paribus, an additional free-flow lane should increase flow. This increase in flows from the lane is also relatively large compared to the decrease of flows in the HOT and GP lane. It may even be better to build a new GP lane instead of a HOT lane.

---

[9] Hours of 6, 7, 8, 9 AM

Table 7. Impact on Flow around Detector Stations, HOV/T Section Only

Outcome: Flow (Vehicles/Hour)	(1) HOV/T Flow	(2) GP Flow
HOT in Existence	-356.3***	-232.8***
	(8.696)	(6.667)
County's Populations	0.000571	-0.00113
	(0.00128)	(0.000983)
5 AM[10]	323.3***	353.5***
	(8.937)	(6.852)
6 AM	772.7***	821.8***
	(8.902)	(6.825)
7 AM	1,043***	1,120***
	(8.986)	(6.890)
8 AM	1,075***	1,172***
	(8.989)	(6.892)
9 AM	879.2***	950.1***
	(9.013)	(6.910)
HOT Exist * 5 AM	145.8***	150.2***
	(11.73)	(8.991)
HOT Exist * 6 AM	102.9***	48.40***
	(11.70)	(8.971)
HOT Exist * 7 AM	-7.557	-75.18***
	(11.77)	(9.027)
HOT Exist * 8 AM	47.94***	-47.81***
	(11.78)	(9.032)
HOT Exist * 9 AM	148.5***	87.99***
	(11.80)	(9.049)
Lanes	208.0***	285.8***
	(3.654)	(2.801)
Constant	-651.3***	-876.9***
	(15.33)	(11.76)
N	48,297	48,297
R^2	0.823	0.891
Detector Station FE	YES	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Time dummy variables not all shown (10 AM - 4 AM)

[10] 5 AM corresponds to the hour from 5 AM –6 AM.

Table 8. Impact on Congestion around Detector Stations for I-680S

VARIABLES	(3) Avg Speed	(4) Avg Speed	(5) Avg Speed
HOT in Existence	-4.163***	-3.582***	-3.250***
	(0.193)	(0.179)	(0.185)
HOV/T Flow		-0.00414***	-0.0374***
		(0.00104)	(0.00379)
HOV/T Flow^2			2.12e-05***
			(2.11e-06)
GP Flow		0.00268**	0.0529***
		(0.00128)	(0.00342)
GP Flow^2			-2.40e-05***
			(1.35e-06)
# of Lanes	9.102***	9.200***	5.144***
	(0.0811)	(0.213)	(0.410)
County's Population	0.000889***	0.000901***	0.000962***
	(2.84e-05)	(2.88e-05)	(3.44e-05)
HOT Exist * 5 AM	2.232***		
	(0.260)		
HOT Exist * 6 AM	1.575***		
	(0.260)		
HOT Exist * 7 AM	2.014***		
	(0.261)		
HOT Exist * 8 AM	1.472***		
	(0.261)		
Constant	30.36***	28.81***	38.82***
	(0.340)	(0.678)	(1.331)
N	48,297	48,297	48,297
R^2	0.373		
R^2 within	.	0.343	0.145
Detector Station FE	YES	YES	YES
Vehicle Flow IV	NO	YES	YES
# of Detector Stations		2	2

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Average speed in miles per hour
Time dummy variables not all shown (9 AM - 4 AM)
Time interactions relative to 12 AM (omitted)

Table 8 displays estimates of Eq. (1), the 2SLS regression, and Eq. (2), the time interaction regression, for average speed on the HOT portion of I-680S. The first column gives estimates with no instrumental variable for the number of vehicles on the road, instead using the interaction term between the HOT-exist dummy and the time of day to capture nonlinearity in the predicted effects. Columns (4) and (5) include the flow instrumental variable and drop the time interaction term. In column (3), average speed declines with statistical significance following introduction of HOT policy. Columns (4) and (5) give estimates consistent with those in column (3) and predict drops in average speed of 3.58 mph in (4) and 3.25 mph in (5). The time interaction term is unneeded in this regression because the flow instrumental variables account for the time variation in travel demand. The coefficients on the flows of the HOV/T lane and the GP lane are significant to at least 5%, but are opposite in direction. As flow in the HOV/T lane increases, the overall speed around the detector station decreases. Conversely, as flow in the GP lane increases, the overall speed on the highway increases. These effects seem to be reversed and inconsistent with theory. Perhaps HOT flow increases if the highway is more congested as drivers shift to the HOT lane. It appears that the HOT policy worsens congestion, as measured by average speed. Due to the weighting used to calculate the average speed, it may give extra weight to the speeds of cars in the GP lanes because there are simply more cars. Therefore, I examine the effects on the HOT and GP lanes separately in Table 9. I present only estimates for the 2SLS regression and leave the time interaction regression in the Appendix.

Table 9 predicts a decrease in speed in both the HOT and GP lanes following HOT policy. Under columns (6) and (7), the HOT lane suffers a 4.48 and 2.95 mph drop in speed, respectively, compared to a smaller decrease in speed of 3.27 mph and 2.64 mph in (8) and (9), respectively. After the start of HOT policy, holding vehicle flow as given, the ratio of vehicle using the HOT lane against the GP lane should increase. Therefore, with a larger relative increase in cars, the HOT lane

Table 9. Impact on Congestion around Detector Stations for I-680S

VARIABLES	(6) HOT Speed	(7) HOT Speed	(8) GP Speed	(9) GP Speed
HOT in Existence	-4.483***	-2.948***	-3.266***	-2.639***
	(0.188)	(0.163)	(0.194)	(0.196)
HOV/T Flow	-0.0165***	-0.0309***	-0.00689***	-0.0358***
	(0.00109)	(0.00335)	(0.00112)	(0.00401)
HOV/T Flow^2		1.84e-05***		2.09e-05***
		(1.86e-06)		(2.23e-06)
GP Flow	0.0197***	0.0379***	0.00867***	0.0519***
	(0.00135)	(0.00302)	(0.00139)	(0.00362)
GP Flow^2		-1.65e-05***		-2.29e-05***
		(1.19e-06)		(1.43e-06)
# of Lanes	8.158***	6.693***	8.293***	4.879***
	(0.225)	(0.362)	(0.231)	(0.434)
County's Population	0.000472***	0.000514***	0.000903***	0.000958***
	(3.03e-05)	(3.03e-05)	(3.13e-05)	(3.64e-05)
Constant	37.88***	41.65***	30.06***	38.40***
	(0.714)	(1.175)	(0.736)	(1.410)
N	48,297	48,297	48,297	48,297
R^2 within	0.199	0.267	0.321	0.158
Detector Station FE	YES	YES	YES	YES
Vehicle Flow IV	YES	YES	YES	YES
# of Detector Stations	2	2	2	2

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Speed in miles per hour

should experience a larger decline in speed than that experienced by the GP lane. Even though the speed in the HOT lane decreases more, the lane operates at a higher speed due to its higher pre-HOT speed. If we consider a smaller loss in speed in the GP lane as a gain compared to the loss in the HOT lane, this result appears to support the findings of Konishi and Mun (2010), that gains from a decrease in cars in the GP lanes do not offset the cost of the increase in cars in the HOT lane. The coefficients on the flow variables have the same sign as in Table 4.

VII. Conclusion

This study examines the effect of HOV to HOT conversion on highway congestion within the context of determining whether HOT policy is an optimal solution to reducing congestion than HOV policy. In this paper, an optimal solution is a reduction in net travel costs aggregated across all drivers on the highway. One way to reduce travel costs is to increase the travel speed and thus decrease the travel time. Though inconsistent with prediction from theory, the HOT project on I-680S appears to reduce travel speeds, after controlling for the flow on the highway. With the highway already operating near capacity, converting a HOV lane into a HOT lane will yield little benefits as the HOV lane is already highly utilized and any additional SOV entering the lane may only have a short time savings.

As many metropolitan areas explore the possibility of HOT lanes as an improvement over HOV lanes, they must consider whether the costs of implementing a program are worth the little, if any improvement in traffic conditions. The paper attempts to correct for the endogeneity present in the flow measure by using instruments, a common technique with economic literature, but uncommon in feasibility studies run by policymakers. The analysis presented in this paper explains basic traffic patterns, but cannot properly capture demand shift, as the traffic data does not include information on individual drivers. The PeMS database is updated yearly and is expected to gain access to California Highway Patrol Accident data and FasTrak data this year. The FasTrak data would provide information on tolls collected alongside the traffic conditions, leading to perfect extension of this project. With such tolling data, true measure of social costs could be measured directly by calculating value of time. As HOT lanes gain popularity and data availability increases, perhaps cross-sectional studies could incorporate route substitution with alternate routes.

Works Cited

AlamedaCTC. (2012). I-680 Sunol Express Lanes - Southbound Fact Sheet: Alameda County Transportation Commission.

BAIFA. (2014, 2014). Introducing a Network of Express Lanes in the Bay Area. *BAIFA Express Lanes.* Retrieved April 15, 2014, from http://www.baifaexpresslanes.org/projects/express_lanes/

Dahlgren, J. (1998). High occupancy vehicle lanes: Not always more effective than general purpose lanes. *Transportation Research Part A: Policy and Practice, 32*(2), 99-114. doi: http://dx.doi.org/10.1016/S0965-8564(97)00021-9

Kittelson Associates, I. (2013). Southbound I-680 Express Lane Performance Evaluation - An After Study: Alameda County Transportation Commission.

Konishi, H., & Mun, S.-i. (2010). Carpooling and Congestion Pricing: HOV and HOT Lanes. *Regional Science and Urban Economics, 40*(4), 173-186. doi: http://www.elsevier.com/wps/find/journaldescription.cws_home/505570/description#description

Kwon, J., & Varaiya, P. (2008). Effectiveness of California's High Occupancy Vehicle (HOV) system. *Transportation Research Part C: Emerging Technologies, 16*(1), 98-115. doi: http://dx.doi.org/10.1016/j.trc.2007.06.008

Langer, A., Winston, C., & Baum-Snow, N. (2008). Toward a Comprehensive Assessment of Road Pricing Accounting for Land Use. *Brookings-Wharton Papers on Urban Affairs*, 127-175. doi: 10.2307/25609550

Liu, L. N., & McDonald, J. F. (1998). Efficient Congestion Tolls in the Presence of Unpriced Congestion: A Peak and Off-Peak Simulation Model. *Journal of Urban Economics, 44*(3), 352-366. doi: http://dx.doi.org/10.1006/juec.1997.2073

Metro. (2012, April 19, 2012). HOT Lanes in the U.S., from http://www.metro.net/projects/expresslanes/expresslanes_us/

Poole, R. W., Jr., & Orski, C. K. (2000). HOT lanes: A better way to attack urban highway congestion. *Regulation, 23*(1), 15-20.

Safirova, E., Gillingham, K., Parry, I., Nelson, P., Harrington, W., & Mason, D. (2004). WELFARE AND DISTRIBUTIONAL EFFECTS OF ROAD PRICING SCHEMES FOR METROPOLITAN WASHINGTON DC. *Research in Transportation Economics, 9*(0), 179-206. doi: http://dx.doi.org/10.1016/S0739-8859(04)09008-0

Small, K. A., Winston, C., Yan, J., Baum-Snow, N., & Gómez-Ibáñez, J. A. (2006). Differentiated Road Pricing, Express Lanes, and Carpools: Exploiting Heterogeneous Preferences in Policy Design. *Brookings-Wharton Papers on Urban Affairs*, 53-96.

Small, K. A., & Yan, J. (2001). The Value of "Value Pricing" of Roads: Second-Best Pricing and Product Differentiation. *Journal of Urban Economics, 49*(2), 310-336. doi: http://dx.doi.org/10.1006/juec.2000.2195

Vickrey, W. S. (1963). Pricing in Urban and Suburban Transport. *The American Economic Review, 53*(2), 452-465. doi: 10.2307/1823886

Winston, C. (2013). On the Performance of the U.S. Transportation System: Caution Ahead. *Journal of Economic Literature, 51*(3), 773-824. doi: doi: 10.1257/jel.51.3.773

Yang, H., & Huang, H.-J. (1999). Carpooling and congestion pricing in a multilane highway with high-occupancy-vehicle lanes. *Transportation Research Part A: Policy and Practice, 33*(2), 139-155. doi: http://dx.doi.org/10.1016/S0965-8564(98)00035-4

Appendix

Figure A1. I-680 Southbound Express Lane

Source: http://www.680expresslane.org/I-680_Map.asp

Table A1. Population by County (Estimated July 1)

County	2007	2008	2009	2010[11]	2011	2012	2013
Alameda	1,455,715	1,477,208	1,498,539	1,513,527	1,531,324	1,553,960	1,578,891
Santa Clara	1,712,026	1,740,964	1,765,137	1,786,539	1,811,297	1,836,025	1,862,041
San Mateo	693,849	703,830	713,617	719,756	728,288	738,681	747,373

[11] Year 2010 estimated population on April 1, to coincide with 2010 Census Data

Table A2. Impact on Congestion around Detector Stations for I-680S

VARIABLES	(4) HOT Speed	(5) HOT Speed	(6) HOT Speed	(7) GP Speed	(8) GP Speed	(9) GP Speed
HOT in Existence	-3.654***	-4.483***	-2.948***	-4.085***	-3.266***	-2.639***
	(0.173)	(0.188)	(0.163)	(0.205)	(0.194)	(0.196)
HOV/T Flow		-0.0165***	-0.0309***		-0.00689***	-0.0358***
		(0.00109)	(0.00335)		(0.00112)	(0.00401)
HOV/T Flow^2			1.84e-05***			2.09e-05***
			(1.86e-06)			(2.23e-06)
GP Flow		0.0197***	0.0379***		0.00867***	0.0519***
		(0.00135)	(0.00302)		(0.00139)	(0.00362)
GP Flow^2			-1.65e-05***			-2.29e-05***
			(1.19e-06)			(1.43e-06)
# of Lanes	10.37***	8.158***	6.693***	9.340***	8.293***	4.879***
	(0.0727)	(0.225)	(0.362)	(0.0861)	(0.231)	(0.434)
County's Population	0.000435***	0.000472***	0.000514***	0.000882***	0.000903***	0.000958***
	(2.55e-05)	(3.03e-05)	(3.03e-05)	(3.02e-05)	(3.13e-05)	(3.64e-05)
HOT Exist * 6 AM	-0.594**			0.943***		
	(0.233)			(0.276)		
HOT Exist * 7 AM	-0.319			1.474***		
	(0.234)			(0.277)		
HOT Exist * 8 AM	-0.955***			1.046***		
	(0.235)			(0.278)		
HOT Exist * 9 AM	-0.309			1.592***		
	(0.235)			(0.278)		
Constant	31.83***	37.88***	41.65***	27.86***	30.06***	38.40***
	(0.305)	(0.714)	(1.175)	(0.361)	(0.736)	(1.410)
N	48,297	48,297	48,297	48,297	48,297	48,297
R^2	0.485			0.376		
R^2 within	.	0.199	0.267	.	0.321	0.158
R^2 overall	.	0.281	0.331	.	0.154	0.219
R^2 between	.	1	1	.	1	1
Detector Station FE	YES	YES	YES	YES	YES	YES
Vehicle Flow IV	NO	YES	YES	NO	YES	YES
# of Detector Stations		2	2		2	2

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Average speed in miles per hour; Delay in vehicle-hours
Time dummy variables not all shown (10 AM - 5 AM)
Time interactions relative to 12 AM (omitted)