E. J. Ourso College of Business
Department of Economics

DEPARTMENT OF ECONOMICS WORKING PAPER SERIES

Dynamic Road Pricing and the Value of Time and Reliability

(Revised June 2017)

Daniel A. Brent
Louisiana State University

Austin Gross
University of Washington

Working Paper 2016-07
http://faculty.bus.lsu.edu/papers/pap16_07.pdf

Department of Economics
Louisiana State University
Baton Rouge, LA 70803-6306
http://www.bus.lsu.edu/economics/

Dynamic Road Pricing and the Value of Time and Reliability

Daniel A. Brent*[a] and Austin Gross[b]

[a]Department of Economics, Louisiana State University
[b]Department of Economics, University of Washington

June 2017

Abstract

High Occupancy Toll (HOT) lanes that use dynamic pricing to manage congestion and generate revenue are increasingly popular. In this paper we estimate the behavioral response of drivers to dynamic pricing in a HOT lane. The challenge in estimation lies in the simultaneity of price and demand: the structure of dynamic tolling ensures that prices increase as more drivers enter the HOT lane. Prior research has found that higher prices in HOT lanes increase usage. We find that after controlling for simultaneity HOT drivers instead respond to tolls in a manner consistent with economic theory. The average response to a 10% increase in the toll is a 1.6% reduction in usage. Drivers primarily value travel reliability over time savings, although there is heterogeneity in the relative values of time and reliability based on time of day and destination to or from work. The results highlight the importance of both controlling for simultaneity when estimating demand for dynamically priced toll roads and treating HOT lanes with dynamic prices as a differentiated product with bundled attributes.

Keywords: transportation economics; tolling; value of time; value of reliability; dynamic pricing; congestion management; high occupancy toll lanes

*Corresponding author: LSU Economics, 2300 Business Education Complex, Louisiana State University, Baton Rouge, LA. 70803; *dbrent@lsu.edu*. The authors are grateful to Mark Hallenbeck for many insightful discussions and endless patience with data requests. Sara Myers and Tyler Patterson from WSDOT also were incredibly helpful and generous with data requests. Robert Halvorsen, Neil Bruce, David Layton, Hendrik Wolff, Robert Plotnik and Clark Williams-Derry, as well as seminar participants at AERE and WEAI conferences also provided useful feedback. Any errors or omissions are the responsibility of the authors.

1 Introduction

Road transportation comprises a substantial proportion of the United States economy. The vast majority of infrastructure is owned, maintained, and operated by local, state, or federal agencies. According to Winston (2010) in 2007 American consumers spent over $1 trillion dollars on gasoline and vehicles. In metropolitan areas road congestion led consumers to purchase 2.9 billion *additional* gallons of fuel and spend 5.5 billion hours sitting in traffic (Schrank et al., 2012). These costs are likely lower bounds due to unpriced congestion externalities for local air pollution (Currie and Walker, 2011; Gibson and Carnovale, 2015), carbon emissions (Weitzman, 2009), and increased sprawl (Anas and Rhee, 2006).[1] Despite the enormous annual cost of traffic congestion most roads do not have even the simplest form of congestion pricing. Combating congestion in the short run by increasing capacity is challenging due to strained transportation budgets such as the perennial projected insolvency of the Highway Trust Fund (Kirk and Mallett, 2013). Furthermore, according to the fundamental law of highway capacity (Downs, 1962, 2004; Duranton and Turner, 2011), increasing road capacity is met with proportional increases in demand - meaning augmenting supply is not likely to solve the problem of traffic congestion. Based on these facts, implementing appropriate congestion pricing has the potential to produce large welfare gains. Recent estimates from Couture et al. (Forthcoming) place the deadweight loss of congestion in the United States at 30 billion dollars a year.

One particular example of congestion pricing that is gaining traction is the High Occupancy Toll (HOT) lane, where High Occupancy Vehicles (HOV) travel do not pay to access the road and Single Occupancy Vehicles (SOV) are charged an access toll.[2] From an engineering perspective redistributing cars from congested general-purpose (GP) lanes to free-flowing HOV lanes can reduce congestion delays and associated externalities (Dahlgren, 2002). The tolls on HOT lanes often vary by time of day or traffic conditions moving trans-

[1] Tolling may not affect the urban structure as shown in Arnott (1998).

[2] The definition of an HOV varies by location and roadway, but all HOV require at least two occupants and some require three or more.

portation infrastructure closer towards pricing congestion externalities that the economics field has espoused for several decades (Agnew, 1977).

HOT lanes generate revenue for local and state agencies that is particularly valuable in the context of aging infrastructure and diminished revenue from gas taxes which has declined in real terms since 1993. Furthermore, HOT lanes are politically palatable compared to unpopular uniform tolls or vehicle miles traveled taxes because GP lanes are left untolled, maintaining a free alternative for low income or cost-sensitive drivers Lindsey (2010). These characteristics of HOT lanes in theory can lead to broad welfare gains (Safirova et al., 2004). However, there remain concerns over equity of access: wealthy drivers can avoid congestion while others must sit in traffic, leading to HOT lanes being disparagingly termed 'Lexus Lanes'. While HOT lanes are a popular form of managing valuable public roadways,[3] there is relatively little empirical economic research on the behavioral response of drivers to HOT implementation. The presence of free GP lanes as a veritable substitute makes HOT lanes with dynamic prices an opportunity to uncover how drivers respond to congestion management.

Our primary contributions are to generate empirical estimates for the price elasticity of demand and calculate the value of time and reliability on dynamically priced HOT lanes using micro-level data on SR167, in the Seattle-Tacoma metropolitan area of Washington State. While most economists are not surprised by a downward sloping demand curve, the existing empirical literature on dynamically priced HOT lanes estimates a positive price response (Liu et al., 2011b; Janson and Levinson, 2014). In addition to identifying a more plausible demand elasticity we show that failure to properly identify the behavioral response to price produces invalid estimates of the value of time and reliability.

The most common explanation put forth for the positive effect of price on HOT demand is that price acts as a signal of future congestion (Liu et al., 2011b; Janson and Levinson,

[3]According to the Bureau of Economic Analysis the value of the stock of highways is equal to $3,264.5 billion. Data were accessed from Table 7.1B. Current-Cost Net Stock of Government Fixed Assets on 1/13/2014 - http://www.bea.gov/iTable/index.cfm

2014), and therefore higher prices are associated with greater time savings. This explanation confounds expectations about time savings with the pure behavioral response to price. *Ceteris paribus*, consumers prefer to purchase the same amount of time savings at a lower price. Not accounting for expectations of time savings introduces omitted variable bias into the coefficient on the pure price effect. A more plausible explanation for a positive price elasticity is that previous work did not adequately control for the simultaneity of price and congestion in dynamic tolling algorithms. As more drivers enter the managed lane conditions deteriorate (speed decreases) and the tolls increase, leading to a positive correlation between price and usage. Traffic conditions are persistent and exhibit a high degree of autocorrelation, leading to biased estimates of price, which is a function of the lagged dependent variable.

Without formally addressing the problems of simultaneity and omitted variable bias in the setting of dynamic tolling algorithms it is premature to conclude that HOT lanes cause a positive response to prices. Our identification relies on an instrumental variable and first-differences approach that overcomes the simultaneity of price and quantity that generates the positive demand response for Liu et al. (2011b). We also control for travel reliability and expectations of time savings using micro-level data, unlike Janson and Levinson (2014) who examine aggregate differences in usage after experimental changes in the toll rates.

Contrary to the previous literature that finds a positive demand response we estimate price elasticities ranging between -0.16 and -0.21, with a preferred estimate of -0.16. This is the first estimate of a negative price elasticity (to our knowledge) for dynamically priced HOT lanes, which are a critical part of many cities' future transportation management plans. Our second contribution is to jointly estimate value of time (VOT) and value of reliability (VOR) for a dynamically priced HOT lane. Prior studies (Brownstone et al., 2003; Liu et al., 2011b; Burris et al., 2012; Janson and Levinson, 2014) of dynamically priced HOT lanes construct a simple estimate of the VOT by dividing the toll by the realized time savings. This method produces unrealistically high estimates of VOT that can exceed $100/hr (Burris et al., 2012; Janson and Levinson, 2014), whereas the U.S. Department of Transportation uses 50% of median household hourly income for the VOT for personal travel, which equates to roughly $14 for Seattle Metro.[4] Simply dividing the toll by time savings is problematic because the toll contains a bundle of attributes including improved reliability.[5] Though some of the authors (Burris et al., 2012; Janson and Levinson, 2014) mention these limitations there are no large scale revealed preference studies that jointly estimate VOT and VOR for HOT lanes.[6]

The joint estimation of VOT and VOR is axiomatically linked to the challenge of properly identifying the demand response for HOT lanes with dynamic pricing. Without identifying the demand response, which represents the (negative) marginal utility of income, it is impossible to estimate the marginal rates of substitution between time savings and money and reliability and money. This leads to problems in the simple methods for estimating VOT on HOT lanes that assigns all of the benefits associated with the HOT lane to time savings. Devarasetty et al. (2012) show in a stated preference study that VOR can be larger than VOT on HOT lanes, indicating that most of the simple revealed preference VOT estimates are too large. Our results are even more stark; drivers vastly value reliability over time savings on the HOT lane. We estimate that VOT is only $7/hour for the preferred specification while VOR is over $22/hour. In aggregate 68% of the benefits to HOT users are from increased reliability, though time savings is relatively more important for some subsets of the roadway. In related work Bento et al. (2014) find that drivers primarily value urgency, or on-time arrival, when using HOT lanes, which is consistent with our findings of reliability being the more important factor. Our estimation also allows us to calculate the back of the envelope benefits to toll users associated with decreased travel time and increased travel reliability. In the base specification the benefits are approximately $3.4 million; the average driver paying

[4]Sources are from (U.S. Department of Transportation, 2014; Bureau of Labor Statistics, 2014).

[5]There are other attributes other than VOT contained in the purchase of HOT access such as the mental stress from being in traffic, particularly when watching cars pass by in the free-flowing HOT lane.

[6]Carrion and Levinson (2013) estimates value of reliability for three different lanes using GPS data, but the sample only contains 18 observations.

the toll receives benefits that are roughly twice the cost of the toll.

The remainder of the paper is organized as follows. The next section discusses how this research fits into the existing literature. Section 3 describes the project setting. Section 4 describes the econometric methodology and Section 5 reports the data used in the analysis. The results are presented in Section 6, and Section 7 offers concluding remarks and discussion. Details on the econometric specification tests, as well as additional tables and figures are available in the Appendix.

2 Literature Review

There is a long literature on congestion pricing, and in particular dynamically priced toll roads that explicitly target the congestion externality Vickrey (1963); Agnew (1977); Arnott (1998); Verhoef and Rouwendal (2004); Lindsey (2010). While the literature shows that the optimal congestion charge usually requires all lanes to be tolled this is politically unpopular. An increasingly common compromise is implementing an HOT lane where SOVs can pay to access a HOV lane. Theory suggests wide benefits from HOT lanes in the presence of driver heterogeneity (Dahlgren, 2002), in practice welfare impacts depend on the distribution of VOT and VOR (Small and Yan, 2001; Small et al., 2005), trends in commuting demand, tolling structure (Chung and Recker, 2011) and distortions in connected markets (Parry and Bento, 2002). Konishi and Mun (2010) show that converting from an HOV to an HOT lane has ambiguous welfare effects, in part due to discouraging carpooling. Many of the models used in these studies include price elasticity as a parameter without providing support for its magnitude and direction.

While we focus on estimating driver demand for a HOT lane in a specific location, the overall context is critical in an attempt to generalize the welfare results. Li (2001) attempts to explain the determining characteristics of HOT use on SR91 in California by analyzing survey data. His findings indicate that income, occupancy, trip purpose and age are important factors. From a theoretical perspective Parry (2002) conducts an analysis of congestion tax alternatives using simple models with three assumptions: (1) equate the marginal social cost of trips both between peak and off-peak travel (2) equate the marginal social cost across travel modes at a given point in time, and (3) sort high and low time-cost drivers by lane. He finds that, given driver time cost heterogeneity, a two lane road with tolls to separate high and low cost users achieves the maximum efficiency, while a uniform toll across both lanes achieves 90% of the optimum by 'spreading out the commute of lower time cost commuters to before and after the toll'. He notes that single-lane tolls are more politically feasible given the 'hostility' from motorists to congestion taxes. Parry and Bento (2002) partially extend this analysis by incorporating distortions into the welfare calculations of a congestion tax. They find that the distortions can cause 'substantial' changes to welfare that must be considered as part of a policy change.

According to these findings HOT lanes probably do not achieve the social optimum since there is still a lane with unpriced congestion externalities. However, dynamic pricing of a HOT lane shifts policy towards internalizing congestion externalities by raising private costs as congestion in the tolled lane increases, as well as allowing drivers with different levels of VOT and VOR to sort appropriately. This is consistent with Dahlgren (2002) who models the addition of different lane types to an existing transportation environment: 'mixed' lanes perform better than HOV lanes when 'initial maximum delay is very high but the proportion of HOVs is not sufficient to fully utilize an HOV lane'. Thus, HOT lanes are an attractive choice in locations where HOV lanes have excess capacity.

From the perspective of individual drivers, two of the primary benefits for paying to access a toll lane are travel time savings and improved reliability. Therefore, related to the welfare implications of HOT lanes is the estimation of VOT and VOR. Li et al. (2010) and Carrion and Levinson (2012) survey the literature and describe three main models for modeling travel time reliability: (1) the mean-variance model, (2) the scheduling model, and (3) the mean lateness model. We use a variety of the mean-variance model based on Small et al. (2005) that uses expected travel time savings and the difference between the median and 80th percentile of travel times at different starting times as a metric for reliability. Small et al. (2005) estimates a VOT of $21 and a VOR of $27 in a HOT lane implementing time-of-day pricing using revealed preference data, with lower estimates using stated preference data. Janson and Levinson (2014) estimate that VOT for a dynamically priced HOT lane in Minnesota ranges from $60 - $124, although the authors acknowledges that the estimates are higher than typical estimates of VOT for a variety of reasons including not accounting for VOR. Burris et al. (2012) also estimates relatively high VOT of $49 for a dynamically priced HOT lane; similar to Janson and Levinson (2014) this estimate does not account for VOR.

A related concept is the sensitivity of drivers to toll rates. Matas and Raymond (2003) shows in a review of the literature that travel demand with respect to tolls is relatively inelastic, with elasticity estimates ranging from -.03 to -.5. In their setting in Spain, Matas and Raymond (2003) finds elasticities ranging from -.21 to -.83. Finkelstein (2009) finds travel demand is quite inelastic with respect to tolls, with an elasticity of -.05; drivers are even less responsive at facilities that use electronic toll collection systems. In the HOT context, Liu et al. (2011b) recovered coefficient values on price in a logit framework ranging from 0.214 to 0.600. More recent research by Janson and Levinson (2014) also found a positive effect of price on HOT usage with elasticities ranging from 0.03 to 0.85. However, as described above, there are some issues with estimation strategies of the aforementioned studies. Our results fit within the standard literature on how drivers respond to tolls.

3 Background

Our project setting is State Route 167 (SR167) in the greater metropolitan area of Seattle, Washington. It is a connector road between the communities of Renton and Auburn south of Seattle and the I405 freeway, which then feeds either into Seattle via I5 or Bellevue via I405.

Instituted in May 2008 by the Washington State Department of Transportation (WSDOT), the HOT lanes pilot project converted a ten mile stretch, in both directions, of SR167's HOV lanes into HOT lanes and continues to operate as of June 2017 (Figure 1). This location was selected due to severe congestion in the GP lanes and excess capacity in the existing HOV lane. At the onset of the project the objective was to fill the excess capacity in the HOV lane by allowing some SOVs to purchase access. Congestion not only causes delays but reduces the total carrying capacity of a road, so shifting cars from the GP to the HOV lane can conceivably increase the total throughput in *both* lanes (Dahlgren, 2002). From a national perspective 10 HOT Lanes are operating in eight states according to the U.S. Department of Transportation.[7] SR167 is likely a smaller road handling fewer cars than other HOT lanes.

Figure 1: SR167 HOT Lanes Map, (WSDOT, 2013)

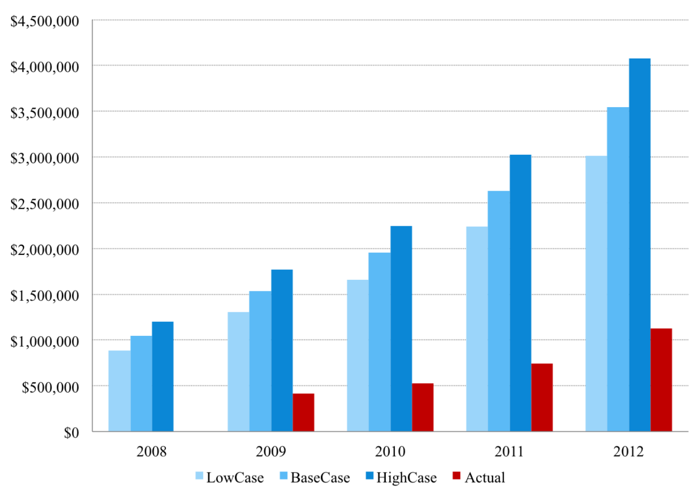
The primary role of the toll is to regulate access to the HOT lane and maintain a minimum level of service, thereby not discouraging use of the lane by HOV and transit. Prior to implementation the GP lanes averaged 30-35 miles per hour during congested periods, with a speed limit of 60 mph, resulting in delays of roughly 50% relative to free flow. Prior to converting to HOT lanes the HOV lanes experienced little to no congestion (Wilbur Smith Associates, 2006). As of 2012 WSDOT attributed the HOT lanes with a host of desirable

[7]Information available at https://ops.fhwa.dot.gov/publications/fhwahop12031/fhwahop12025/index.htm.

outcomes including: decreased congestion in the GP lanes, decreased peak congestion, maintained free flow in the HOT lanes, increased capacity of the corridor, increased safety and revenue neutrality (WSDOT, 2012). The pilot program, originally set to expire in 2012, has been extended and there are plans to convert and additional six miles from HOV to HOT lanes.

Assessing the willingness to pay for toll lanes is a requirement to determining toll levels that meet traffic volume priorities. Problematic assumptions by WSDOT in terms of the demand for HOT usage manifested in poor revenue forecasts as seen in Figure 2. Although revenue generation was not the primary objective for WSDOT, it is clear that there was a fundamental misunderstanding of the trajectory of usage and the driver response to the introduction of a tolled alternative lane.[8]

Figure 2: SR167 Revenue: Forecasts vs Actual, (Wilbur Smith Associates, 2006)



Along the ten mile stretch that comprises our project setting, SR167 has three northbound lanes and three southbound. In each direction there are two GP lanes with the third lane reserved for HOT use. HOVs require no additional equipment to use the lane, but SOVs that use the HOT lane must have purchased and installed a WSDOT 'Good to Go' (GTG) pass that registers a vehicle's passage and collects the posted toll. Transponder detectors

[8]Other factors including a general decrease in travel demand due to the global financial crisis also contributed to the erroneous forecast.

are installed at ‘gates’ that are the only legal entry and exit points for the HOT lane. There are six gates in the northbound direction and four gates southbound, with a double white line separating the GP and HOT lanes between gates.

The GTG passes can be used for all tolling facilities operated by WSDOT, including the SR520 Bridge and the Tacoma Narrows Bridge in addition to the SR167 HOT lanes. Both individuals and businesses can purchase GTG passes. While WSDOT does track both commercial and individual accounts, individuals with many users (a large family) could also purchase a commercial pass. WSDOT could not provide us with separate or tagged samples of usage broken out by account type so we were unable to uncover what effect this has on our estimates. However, purchasers of HOT access who do not bear the burden of the price would be expected to decrease the magnitude of our estimates of price sensitivity, and so we consider our elasticity estimates to be lower bounds in absolute value.

While we do not observe each unique driver in our HOT usage data we do have a summary of those who use the HOT lane by zip code and frequency of use. One factor that impacts HOT use is the penetration of GTG passes. The unique number of HOT users on SR167 has been rising steadily, from 21,623 unique drivers in 2008 to 38,025 in 2011.[9] We also obtained summary statistics for the frequency of use for individual drivers. We find that most drivers use the HOT lane only sparingly. The mean annual number of tips is just under seventeen, but the median is only two with a large number of drivers only paying once or twice in a year. While we may expect there to be different behavior between frequent and infrequent users, Liu et al. (2011a) find that the behavior is very similar between these two user classes. We therefore are not concerned about the skewness of the frequency data.

[9] It should be noted that the HOT lanes were implemented in May 2008 so the 2008 figure is incomplete. Likewise through the end of September 2012 the HOT had 29,623 unique paying users. For reference, there are approximately 44,000 zip codes in the U.S., and 170 zip codes in the Seattle metropolitan area

4 Methodology

Our objective is to identify the pure behavioral response of toll prices on the demand for purchasing access to the HOT lane. The structure of the tolling algorithm dictates that prices increase as HOT speeds decrease and HOT volume increases. The changes in speed and volume are computed every five minutes taking the difference between data at the five minute mark and the average of the previous four minutes. For instance, conditions from 7:59-8:00 are compared to the average conditions from 7:55 through 7:59 to calculate the price at 8:00-8:05. Therefore, drivers do not impact their own toll rate, but rather for drivers traveling behind them. However, traffic conditions exhibit a high degree of persistence leading to autocorrelation in the variables of interest (see Figure 3). Not accounting for such a high degree of autocorrelation results in biased coefficients. The econometric challenge to identification in our setting can be outlined based on the following equation:

$$y_{it} = \beta p(y_{It-1})_{it} + \theta X_{it} + \epsilon_{it} \tag{1}$$

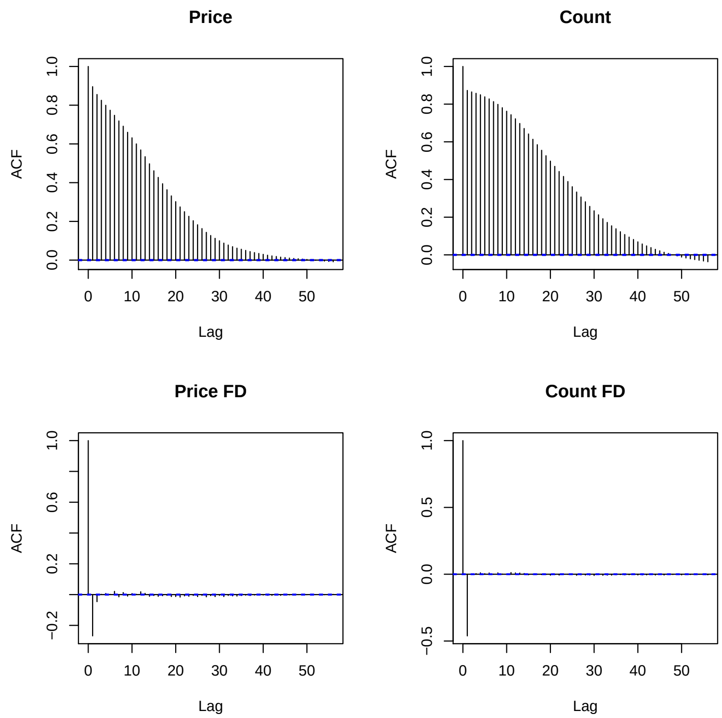
In this specification y_{it} is the count of SOV drivers in the HOT lane at gate i and time t, and $p(y_{It-1})_{it}$ is the price at gate i and time t, which is a function of traffic at the current and downstream gates $I \in [i, i+1, ...i+n]$, in the previous $(t-1)$ period where n is the terminal gate.. Since current prices depend on lagged counts of HOT users, the OLS estimates of β will be biased in the presence of autocorrelation. This can be seen in equation 2 by substituting in the value of y_{it-1} in the for $p(y_{It-1})_{it}$, which contains ϵ_{it-1}. Formally, the OLS estimates are biased if $E[p(\epsilon_{it-1})\epsilon_{it}] \neq 0$.

$$y_{it} = \beta p\left(\beta p(y_{It-1})_{it} + \theta X_{it-1} + \epsilon_{it-1}\right)_{it} + \theta X_{it} + \epsilon_{it} \tag{2}$$

Since, traffic is highly persistent and unobserved factors in the previous five minutes are correlated with current HOT usage, which likely biases the OLS estimates. We therefore first difference (FD) the data to reduce the serial correlation (see Figure 3). This requires the less stringent assumption that the first differenced error terms exhibit zero autocorrelation ($E[p(\Delta\epsilon_{it-1})\Delta\epsilon_{it}] \neq 0$). The first differenced equation is seen in equation 3.[10] The interpretation of differenced data is also more attractive in the context of dynamic tolling. We know that both the toll and usage will be high during periods of congestion, but what we hope to recover is how drivers respond to *changes* in the toll rate. Increasing the toll as congestion increases is the central tenant of dynamic congestion pricing and is critical to managing HOT lanes.

$$\Delta y_{it} = \beta \Delta p(y_{It-1})_{it} + \theta \Delta X_{it} + \Delta \epsilon_{it} \tag{3}$$

Figure 3: Price and Count ACF



Notes: The autocorrelation function (ACF) of price and count and their respective first differences.

Since it is possible that $E[f(\Delta\epsilon_{it-1})\Delta\epsilon_{it}] \neq 0$, in addition to differencing the data we also instrument price at the gate using downstream traffic and price. Recall that prices at gate i at time t are determined not just by traffic at gate i at time $t-1$, but also

[10]Prior to estimation we perform several test for unit roots adapted for time series data do not find unit roots in the data. Details can be found in Section A.3 in the Appendix.

gates $i+1, ...i+n$. Therefore, we expect the FD-IV specification requires the less stringent assumption that $E[f(\Delta\epsilon_{i+1,t-1})\Delta\epsilon_{it}] = 0$. The intuition is that when traffic conditions are more severe downstream, drivers upstream will drivers will face higher tolls than would be dictated by the traffic at their gate. A demand shock at gate 2 that does not impact traffic at gate 1 will cause variation in the price at gate 1 that is uncorrelated with demand shocks at gate 1.

Empirical Specification

Our specific estimating equation takes the form:

$$y_{it} = \beta p_{it} + \gamma GP_{it} + \phi E[TTsave|t] + \lambda R_t + c_i + h_t + d_t + u_{it} \tag{4}$$

where the dependent variable, y_{it}, is the count of SOVs in the HOT lane at gate i and time t, and takes values $y = \{0, 1, 2, 3, \dots\}$. Equation 4 is similar to our equation 1 where we expand the variables and parameters contained in θX_{it}. Our parameter of interest is β, the coefficient on the displayed HOT price, p_{it}. We also want to recover the VOT and VOR, the marginal rates of substitution between time and money and reliability and money. VOT and VOR are represented by the ratios $-\frac{\phi}{\beta}$ and $-\frac{\lambda}{\beta}$ respectively. Estimating all the preference parameters jointly presents a more reliable methodology for estimating VOT and VOR compared to simply dividing the toll by time savings, which is confounded with unobservables (Janson and Levinson, 2014). We include the expected time savings ($E[TTSave|t]$) based on the information available to drivers at time t as described in Section 5.3. Unlike Liu et al. (2011b), who include realized travel time, we model this as an expectation based on the expected difference in travel time since the actual time savings are unknown to the driver when she makes the HOT purchase decision. Reliability (R_t) is the difference in reliability between the GP and HOT lanes based on the reliability metric advocated in Small et al. (2005). Additional controls include the speed in the GP lanes (GP_{it}). We also include fixed effects for the gate of entrance (c_i), hour of day (h_t), and day-of-week (d_t). The idiosyncratic error term is represented by u_{it}.

First-differencing equation 4 gives

$$\Delta y_{it} = \beta \Delta p_{it} + \gamma \Delta GP_{it} + \phi \Delta E[TTsave|t] + \lambda \Delta R_t + \Delta u_{it} \tag{5}$$

where the coefficient of interest is β. We perform the first-differencing (FD) between time periods on the same day and at the same gate. The fixed effects drop out as a result of the first difference. Our FD instrumental variables estimation (FD-IV) replaces Δp_{it} with $\Delta \hat{p}_{it}$, where the excluded instruments are downstream traffic and price ($\Delta GP_{i+1,t}$, $\Delta y_{i+1,t}$, and $\Delta p_{i+1,t}$).

5 Data

Our primary dataset consists of information collected by highway loop detectors and automated tolling systems.[11] Loop detectors yield volume (number of vehicles) and occupancy (percentage of time a vehicles are on top of the detector) for specific lengths of individual lanes at five minute intervals. The data are publicly available through the Washington State Transportation Center (TRAC) based at the University of Washington. Tolling data, obtained from the WSDOT through a public disclosure request, includes date and time of toll collection, entry-exit gate combination and the price paid. There are several challenges in generating a viable dataset from the loop detector data. First we remove all observations that have a data quality flag indicating infrastructure malfunction.[12] Next we drop all observations on weekends and holidays as these are not representative of normal commuting behavior. This leaves us with a time series of volume and occupancy for all loop detectors on

[11] Prior analysis also accounted for weather and gas price. Weather variables are reported hourly from SeaTac airport, at a distance of 4.1 miles from SR167. Gas prices are the weekly average for the area of study. In pursuing a first-difference estimation at the five-minute level these other controls vary too infrequently to contribute.

[12] Of the 16,870,248 loop detector observations 4.8% were removed due to quality flags.

the route for every valid five minute interval during our sample period. Speed is computed from volume data based on Athol (1965)'s formula.

$$v = \frac{q}{o \times g}$$

$$\begin{aligned} \text{where } v &= \text{mean speed in mph} \\ q &= \text{volume of vehicles} \\ o &= \text{percent of lane occupancy} \\ g &= \text{speed parameter, given by WSDOT as 2.4} \end{aligned}$$

Using imputed speed, TRAC also provided estimated whole-route travel times for the northbound and southbound directions, divided into HOT and GP, at five minute intervals. Our final sample includes 1,071,743 observations of drivers entering the HOT lane between 2008 and 2011.

5.1 HOT Tolls

The tolling algorithm is designed to determine the price at five minute intervals using data from the HOT lane and ensuring a minimum speed of 45 mph in the HOT lane. The algorithm compares the current speed and flow with an average of the previous four minutes. *Ceteris paribus* when either speed or flow is increasing (decreasing) the toll rate will decrease (increase). We obtained the tolling algorithm without exact parameter values from WSDOT under the condition that we not reproduce it, since it is proprietary to the consulting company that designed it. Importantly for our study, the algorithm only incorporates HOT data and does not consider traffic in the GP lanes. Additionally, while the toll at a given gate is based upon data from loops around the gate, the toll may be overridden if downstream gates are computing a higher toll. We exploit this feature of the tolling algorithm by using downstream (further along the route) traffic and prices to generate instruments for the price that drivers face. WSDOT provided anonymous transponder recordings from SR167 including time of day, amount charged, as well as entry and exit gate. The prevalent usage pattern, in both directions, is to enter at the first gate and stay in the HOT lane until it ends (NB1 to NB6 and SB1 to SB4; see Figure A.1). The next greatest usage is characterized by entering at the second gate and staying through the end, etc. A small minority of drivers pay the toll and exit before the HOT terminates. These customers are not used in the estimation since our instrument requires dropping data from terminal gates. Figure A.2 in the Appendix displays the toll rates by time of day and gate location.

5.2 Survey

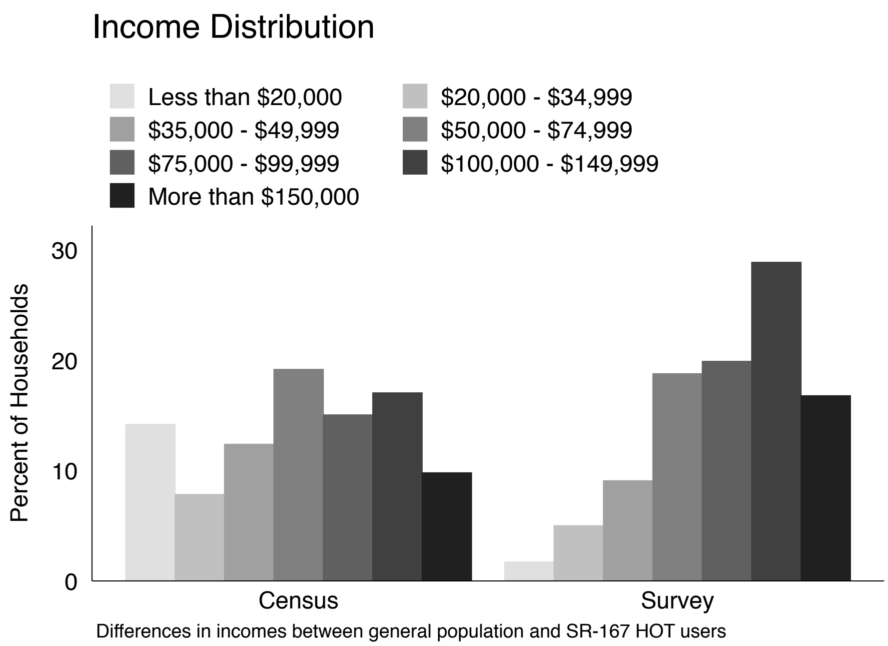
To capture demographic characteristics we obtained yearly surveys of SR167 HOT users from WSDOT. The survey is sent to all email addresses attached to a GTG account that have used the SR167 HOT lane at least once. It covers a broad range of topics including questions about demographics and attitudes towards the HOT lanes. We focus on the income of HOT users since previous research (Li, 2001) has identified this as an important characteristic of use. Income is also a driver of VOT and will help put the VOT and VOR results in context for generalizing the results. There are likely selection issues for estimating the income distribution of SR167 from the survey data, but a priori the effects are ambiguous.[13]

To approximate the SR167 income distribution we use a weighted average of zip code level data from the 2010 US Census, where the weights are the proportion of all HOT users that came from a specific ZIP code. This method places more weight on zip codes that use the SR167 HOT lane. We assume that the spatial distributions of HOT users and GP users are similar. We cannot account for differences within a zip code between the average household and the average GP user. Since HOT users may be more affluent than GP users our method of constructing the SR167 income distribution is likely to be represent an upper bound. Figure 4 presents the income from the survey of HOT users compared to our estimate

[13]For example, the survey sample may have higher income if they are more likely to have internet access, or have lower income if they have a lower opportunity cost for completing the survey.

of the income of all SR167 users from the weighted census data. It is clear that even the upper bound of SR167 users' income is substantially lower than the HOT users.

Figure 4: Differences in Income



Notes: Census data are weighted by ZIP code frequency of 167 GTG users, and may be considered an upper bound of income. Survey data are from annual WSDOT surveys of HOT users.

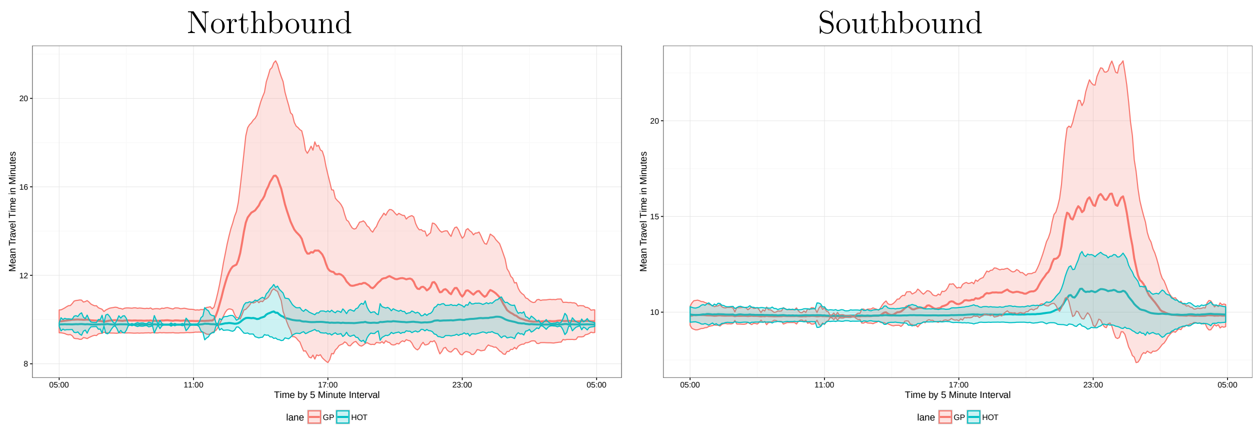
5.3 Travel Time and Reliability

To illustrate the difference in average travel time and reliability between the GP and HOT lanes we plot the distribution of travel time over the course of the day in Figure 5 for both the northbound and southbound routes. The travel times were computed by TRAC for every five minute interval for both the HOT and GP lanes. The thick line represents the average travel time at a given time of day and the shaded region is one standard deviation in travel times at that time over all days in the sample. There are several noteworthy features in Figure 5. First, the peak congestion periods are dramatic: there is a steep spike in traffic for the GP lanes during the morning in the northbound direction and during the evening in the southbound direction. The free flow rate, as evidenced by travel times in the middle of the night, is approximately 10 minutes in either direction. On average, the HOT lane maintains close to free flow conditions throughout the day in the northbound direction and experiences very minimal congestion during the evening peak in the southbound route. Comparing mean HOT travel times to the average GP travel times during the peak commute shows that drivers are saving roughly 3-6 minutes by paying for HOT access. The summary statistics for the sample are presented in Table 1. Roughly 2.5 drivers purchase access to the HOT in the average 5-minute period, and pay an average price of $0.68. The average time savings is 1.85 minutes, and the average difference in reliability is 1.17 minutes.

Previous research by Small et al. (2005) (among others) show that reliability is also an important determinant of the HOT use decision. The shaded region shows that one standard deviation in travel times in the GP lane can often exceed 20 minutes during the commuting period.[14] There is little variation in travel times in the HOT lane, indicating that reliability is also a key attribute of the good. We follow Small et al. (2005) in constructing a reliability measure, estimating the median and 80th percentiles in travel-time savings for each 5-minute interval throughout the day for both the GP and HOT lanes using quantile regressions. Coefficients for northbound and southbound directions are estimated separately. The difference between the fitted values for the 80th and 50th quantiles is a measure of dispersion that approximates reliability for each lane. The reliability variable in the regression is the difference in reliability between the GP and HOT lanes. Expected travel-time savings are estimated from the fitted values of a linear regression of time savings on 5-minute fixed effects, HOV counts, and GP speeds. Our specification for expected travel time savings is: $TTSave_{it} = \beta_1 HOV_{it} + \beta_2 GP_{it} + D_m + e_{it}$, where $TTSave_{it}$ is realized travel time savings, GP_{it} is GP speed, HOV_{it} is the count of HOV drivers in the HOT lanes, and D_m are fixed effects for each 5-minute period in the day. This produces a forward thinking prediction of drivers' expectation of travel time savings when they make the decision to enter the HOT

[14] Although the graph shows a rough measure of travel time reliability, Peer et al. (2012) show that commuters form expectations over time that generate perceptions of reliability. Thus, the raw data are likely not the exact measure of travel time variability, and in the regression we control for as many factors as possible such as removing weekends and holidays and including day-of-week fixed effects.

Figure 5: Travel Time and Reliability



Notes: The thick line the mean travel time and the shaded region bounded by the thin lines represents one standard deviation in travel time. The colors distinguish between the GP and HOT lanes. Free flow travel time is approximately 10 minutes in either direction. The graphs consist of 1,566 days of 5-minute traffic data.

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Count	1,071,743	2.41	3.96	0	46
Price	1,071,743	0.68	0.44	0.50	6.50
SOV	1,071,743	102.78	38.85	0.00	209.00
GP Speed	1,071,743	53.52	11.18	5.00	70.00
Reliability	1,071,743	1.17	0.93	−0.00	5.31
Expected Time Savings	1,071,743	1.85	1.56	−2.16	13.55

lane. Our specification assumes drivers form their expectation of travel time savings from the HOT lane based on the time of day and traffic conditions.

6 Results

Coefficient estimates for both the OLS, FD, and FD-IV HOT counts are presented in Table 2. Standard errors in OLS models (columns (1) and (2)) are robust to heteroskedasticity and clustered by date and gate of entrance and the FD (columns (3) and (4)) are adjusted using a first-difference robust variance matrix clustered by date and gate of entrance (Wooldridge, 2010). The OLS estimates are presented to show the impact of both simultaneity due to autocorrelation and the omitted variable bias from not including expected travel time and reliability. Since the data are count we also estimate the regression in columns (1) and (2) using a Poisson model. The results qualitatively similar and are available upon request. Column (1) shows that in a simple OLS framework the price response is positive, significant, and reasonably large in magnitude.[15] Simply controlling for travel time and reliability, as the seen in column (2), vastly decreases the magnitude of the estimate of the demand response but it is still positive and significant. Columns (3) and (4) show the results of the FD model and the FD-IV both produce a demand response that is negative and statistically significant at the 1% level. While there are differences when using instruments for price the results are relatively similar, indicating that autocorrelation is the primary driver of endogeneity.[16]

Our preferred specification, the FD-IV model shown in column (4), estimates that a \$1 increase in the price decreases HOT users by 0.579 within a 5 minute interval. This finding stands in contrast to earlier research that identified a positive price response (Liu et al., 2011b; Janson and Levinson, 2014). Transforming this to an elasticity at the average quantity and price yields an estimated elasticity of -0.16. As expected higher GP speeds decrease SOV purchases of HOT access since faster GP speeds decrease the benefits of the HOT lane relative to the GP lane. The point estimates of VOT and VOR are \$6.7 and \$22.3 respectively indicating that drivers care more about reliability than time savings when using the HOT lanes. VOT and VOR are investigated in more detail in Section 6.2. While we do not have causal identification framework for estimating the parameters on $E[TTsave|t]$ and $Reliability$ and we believe that these variables do not suffer from the same endogeneity concerns as the price variables after first differencing the data. These variables are constructed by distributional statistics from the entire sample, so an individual driver at a given point in time has little ability to influence these variables. Additionally, we model these as expectations that are not impacted by idiosyncratic demand shocks at time t.

[15] Since the data in columns (1) and (2) are counts we also estimate count models including Poisson and negative binomial that produce similar results and are available upon request.

[16] Another explanation is that the FD model in column (3) still utilizes the quasi-random variation of current prices being overridden by downstream prices.

Table 2: First Difference and OLS Regressions

	(1) OLS	(2) OLS	(3) FD	(4) IV-FD
Price	3.537***	1.245***	-0.709***	-0.579***
	(0.0697)	(0.0588)	(0.0225)	(0.0262)
GP Speed	-0.0341***	0.00734***	-0.0103***	-0.0129***
	(0.00186)	(0.00203)	(0.000711)	(0.000744)
Reliability		0.549***	0.344***	0.215***
		(0.0372)	(0.0203)	(0.0195)
E[TT Save]		1.035***	0.0714***	0.0650***
		(0.0231)	(0.00648)	(0.00646)
Elasticity	1.00	0.35	-0.20	-0.16
VOT		-49.9	6.0	6.7
VOR		-26.5	29.1	22.3
Observations	1,071,743	1,071,743	1,056,997	839,394
F-Statistic				267.9

Note: The dependent variable is the count of HOT users in a five minute interval. The dependent and all independent variables are in levels in columns (1) and (2) and first-differenced in columns (3) and (4). The decrease in observations from in the FD model is due to dropping the first period. The decrease in observations in the IV model is due to dropping the last gate, which has no spatial lead. Standard errors are robust to heteroskedasticity and clustered at the entry gate. *p<0.1; **p< 0.05; ***p<0.01

6.1 Heterogeneity in Demand Elasticity

The price elasticity, as well as the VOT and VOR, depend on the features of the trip so we investigate two important sources of heterogeneity: the time of day and trip direction. Time of day is associated with congestion and also captures drivers on the traditional daily commute to and from work. The peak period is defined as 6:00am-9:00am in the northbound (NB) direction and 3:00pm-6:00pm in the southbound (SB) direction corresponding to the morning and evening commutes.[17] The heterogeneity with respect to route direction is motivated by the argument that drivers face different incentive structures for utilizing the HOT lanes driving to and from work.

Table 3 presents results of regressions that subset the sample by direction and peak congestion period. Columns (1) and (2) subset the sample by direction, columns (3) and (4) subset the sample by peak period, and columns (5) and (6) subset the sample by direction in the peak period. The response to the toll is relatively consistent for most of the subsamples with the exception of the off-peak sample. This may be due to lower range of prices that

[17] The peak period can be seen on the travel time graphs in Figure 5.

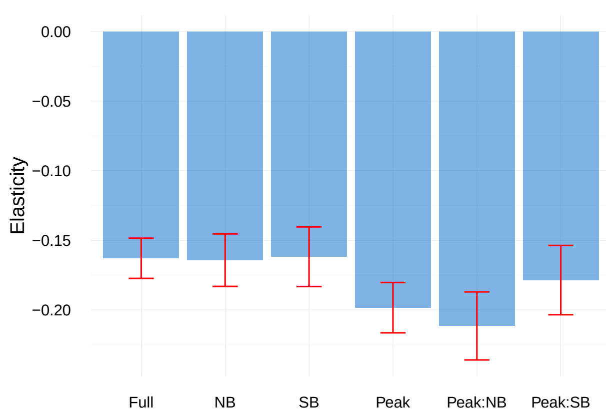
off-peak users face and/or the fact that off-peak HOT users may reflect commercial drivers that do not pay the toll on their own.[18] Figure 6 shows the different elasticity estimates for the subsamples. Excluding the off-peak period shows a consistent demand elasticity ranging form -.16 to -.21. The next section discusses both base effects of the value of time and reliability and issues of heterogeneity.

Table 3: Heterogeneity by Gates, Congestion, and Direction

	(1) NB	(2) SB	(3) Peak	(4) Off-Peak	(5) Peak:NB	(6) Peak:SB
Price	-0.584***	-0.575***	-0.705***	-0.178***	-0.751***	-0.634***
	(0.0342)	(0.0389)	(0.0327)	(0.0272)	(0.0443)	(0.0451)
GP Speed	-0.0131***	-0.0119***	-0.0169***	-0.0104***	-0.0175***	-0.0138***
	(0.000838)	(0.00147)	(0.00160)	(0.000581)	(0.00182)	(0.00286)
Reliability	0.373***	0.0785**	0.216***	0.208***	0.625***	0.0476
	(0.0245)	(0.0292)	(0.0303)	(0.0168)	(0.0615)	(0.0344)
E[TT Save]	0.0318***	0.118***	0.120***	0.0394***	0.0547**	0.258***
	(0.00861)	(0.00954)	(0.0158)	(0.00499)	(0.0203)	(0.0248)
Elasticity	-0.16	-0.16	-0.20	-0.050	-0.21	-0.18
VOT	3.3	12.3	10.2	13.3	4.4	24.4
VOR	38.4	8.2	18.4	70.4	49.9	4.5
Observations	622,468	216,926	179,970	659,424	133,231	46,739

Note: All models are based on the same specification in column (4) of Table 2. Columns (1) and (2) subset the sample by direction, columns (3) and (4) subset the sample by peak period, and columns (5) and (6) subset the sample by direction in the peak period. Standard errors are robust to heteroskedasticity and clustered at the entry gate. *p<0.1; **p< 0.05; ***p<0.01

Figure 6: Price Elasticities by Model



Notes: The thick bars represent the mean estimates for average elasticity and the thin bars are 95% confidence intervals calculated by the delta method. The estimates are based on the price parameter as well as average quantity and price in the relevant subsample for regressions in column (1) - (3) of Table 3.

[18] The average peak price is \$1.15 compared to the the average off-peak price of \$0.55.

6.2 Value of Time and Reliability

A simple estimate of the Value of Time (VOT) is just the toll divided by the realized time savings. Though as stated above this measure produces unrealistically high VOT estimates on dynamically priced HOT lanes.

$$VOT = \frac{\text{Toll}}{TT_{GOP} - TT_{HOT}}$$

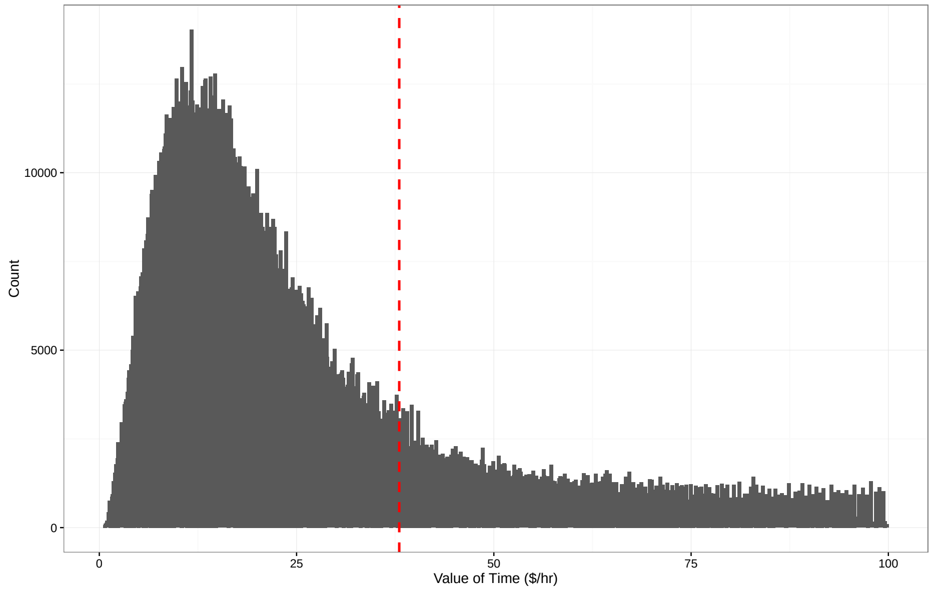
Before presenting the jointly estimated VOT and VOR from the regressions we show the distribution of the simple VOT from SR167 in Figure 7. VOT is constructed by simply dividing the toll by the difference in travel time between the GP and HOT lane for each 5-minute interval in the sample.[19] It should be noted that the simple VOT is the minimum that a driver is willing to pay for the realized time savings, assuming all benefits are due to time savings. The average VOT using the simple method is $38 dollar per hour and is designated by the red dashed line in Figure 7. Approximately 0.7% of drivers experience a negative simple VOT where the travel time in the GP lane was faster than the HOT lane.

Columns (1) and (2) (as well as (5) and (6)) of Table 3 show that drivers have a relatively similar price responsiveness in both directions, but NB driver primarily value reliability and SB drivers primarily value time savings. Figure 8 presents the estimates of VOT and VOR defined as the ratio of the preference parameters. It is important to note that this requires a negative coefficient on price in order to obtain a valid estimate of the marginal utility of income defined as negative one times the dis-utility of the toll. The estimates in Figure 8 are based on the regression models presented in Table 3, as well as the base model from column (4) of Table 2. Since VOT and VOR are nonlinear combinations of parameters Figure 8 reports the mean and 95% confidence interval using the delta method.

When interpreting the relative magnitudes of VOT and VOR it is important to consider that although they are both measured in hours these values are based on different vari-

[19] All observations with negative time savings and time savings above $100 are not shown in the figure but are used to construct the average.

Figure 7: Simple Value of time



Notes: Simple VOT is based on toll and loop detector data. The red dashed line is the average. The figure is does not show negative values or values above $100/hr to assist in the viability.

ables in the regression. However, based on the summary statistics provided in Table 1 the means and variance of expected times savings are relatively similar. Additionally, when we aggregate the values over the observed time savings and reliability we get similar relative magnitudes as our base estimates of VOT and VOR.[20] The main result is that VOR is more important than VOT. In the base specification the reliability ratio (VOR/VOT), is 3.3 and statistically different than 1, indicating that reliability is more important in using the HOT lane than time savings. These results suggest that the simple estimates of VOT on HOT lanes overestimate the true VOT, and that much of the purchase decision is actually based on improved reliability.

There is substantial heterogeneity in VOT and VOR. Northbound travelers greatly prefer reliability to time savings, which may indicate the need to arrive at work at a specified time. Conversely, the difference between VOT and VOR for southbound drivers is not statistically

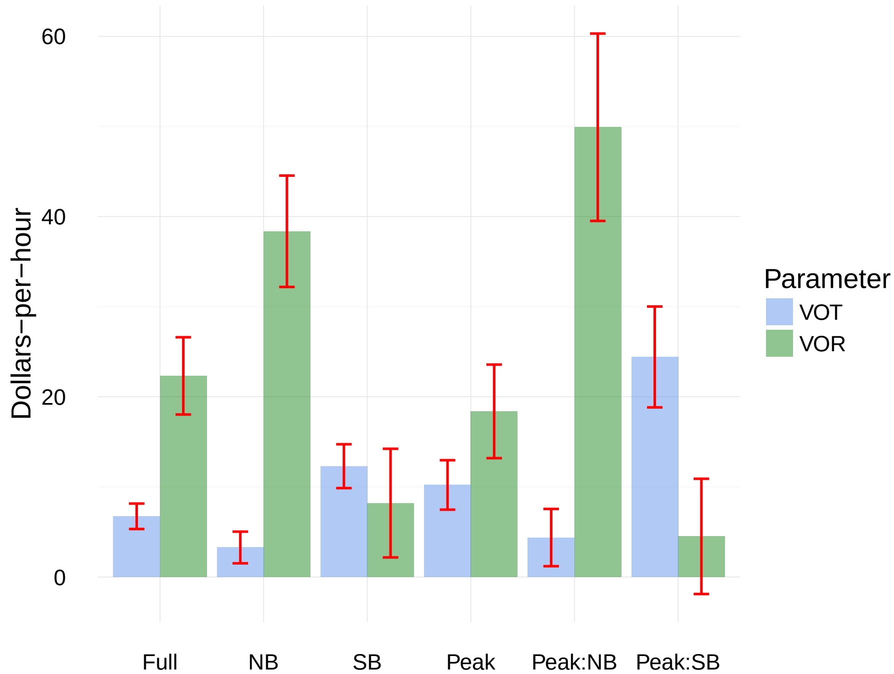
[20] We find that for our parameters VOR represents 77% of the total value (VOR/(VOT+VOR) = 77%) and aggregating over the observed time savings and reliability improved reliability generates 68% of aggregate benefits.

significant. The peak VOT is statistically significantly larger than the base specification and the reliability ratio decreases to 1.8 during the primary commuting period.

Breaking down the heterogeneity further by focusing on the peak period shows that drivers in the morning commute to work (NB) value reliability over time savings while drivers returning home (SB) prefer time savings. This is intuitive given that drivers need to get to work on time and they just want to return home quickly.

The heterogeneity also suggests that transportation managers can optimize the toll algorithms for HOT lanes based on simple observable differences in usage behavior. Drivers that value reliability may not be as sensitive to the toll rate and will purchase HOT access at a wide range of prices. Conversely those who value time savings may be more sensitive to the toll rate and traffic conditions when deciding to use the HOT. However, since time savings and reliability are correlated a more detailed analysis is required to investigate the relationship between VOT, VOR, and price elasticity.

Figure 8: Value of Time and Reliability



Notes: The thick bars are the mean value of time and reliability and the thin error bars are 95% confidence intervals calculated by the delta method. The means are based on dividing the time savings and reliability parameters by the price coefficient from the regressions in Table 3. The confidence intervals are created using the delta method.

6.3 Robustness

We also perform several robustness checks presented in Table 4. All regressions presented in Table 4 have the same basic from as the our preferred regression (Table 2 column (4)), which is shown in column (1) of Table 4 for reference. The first three robustness checks add additional control variables: column (2) adds gate-specific time trends, column (3) adds HOT speed, and column (4) adds HOV Volume. All the parameter estimates are relatively stable across these specifications. Next we test the robustness of the instrument by using two $(i + 2)$ and three $(i + 3)$ spatial leads of traffic conditions and prices. This addresses the concern that since we are using 5-minute bins some of the drivers in gate i may also be registered at gate $i + 1$ at time t. The main parameters of interests are quite similar with the exception of an elasticity value that is larger in absolute value when using three spatial leads. The higher elasticity may reflect that we need to drop almost half of the observations when using three spatial leads as opposed to one spatial lead. Lastly, we also estimate a log-log specification where the dependent variable is the natural log of counts and the price variable also undergoes the logarithmic transformation. This is an alternate estimate of the elasticity, but we don't estimate VOT and VOR because the interpretation of the marginal rates of substitution changes when using the log of price. The elasticity is slightly lower in absolute value, but relatively similar in magnitude.

Table 4: Robustness

	(1) Base	(2) Time Trend	(3) HOT Speed	(4) HOV Volume	(5) 2-Leads	(6) 3-Leads	(7) Logs
Price	-0.579***	-0.579***	-0.631***	-0.559***	-0.559***	-0.979***	
	(0.0262)	(0.0262)	(0.0379)	(0.0297)	(0.0297)	(0.0636)	
ln(Price)							-0.116***
							(0.00609)
GP Speed	-0.0129***	-0.0129***	-0.0143***	-0.0123***	-0.0123***	-0.0141***	-0.00373***
	(0.000744)	(0.000744)	(0.000853)	(0.000802)	(0.000802)	(0.00120)	(0.000165)
Reliability	0.215***	0.215***	0.271***	0.195***	0.195***	0.338***	0.0511***
	(0.0195)	(0.0195)	(0.0246)	(0.0197)	(0.0197)	(0.0331)	(0.00408)
E[TT Save]	0.0650***	0.0649***	0.0652***	0.0415***	0.0412***	0.0672***	0.0152***
	(0.00646)	(0.00646)	(0.00826)	(0.00704)	(0.00705)	(0.0119)	(0.00158)
Elasticity	-0.16	-0.16	-0.18	-0.16	-0.16	-0.28	-0.12
VOT	6.7	6.7	6.2	4.5	4.4	4.1	
VOR	22.3	22.3	25.8	21.0	20.9	20.7	
Observations	839,394	839,394	621,916	664,674	664,674	420,266	839,394

Note: All models are based on the same specification in column (4) of Table 2, which is reproduced in column (1) for reference. Column (2) adds gate-specific time trends, columns (3) and (4) add HOT speed volume respectively as control variables. Columns (5) and (6) use two and three spatial leads for the instrument respectively. Column (7) estimates the equation using the using the natural logarithm of counts and prices. Standard errors are robust to heteroskedasticity and clustered at the entry gate. *p<0.1; **p< 0.05; ***p<0.01

6.4 Aggregate Time Savings and Reliability Benefits to HOT Users

Combining the VOT and VOR estimates with the realized time savings and improvements in reliability generates monetary benefits to drivers paying the toll on SR167.[21] These benefits focus on the dollar value of time savings and reliability for those that purchase access to the HOT lane and omits other attributes of the toll, such as reducing the dis-utility of being stuck in traffic or improved safety. Thus, the estimates presented should be considered lower bounds of the benefits to HOT drivers. Our base specification produces aggregate benefits

[21] Aggregate benefits from time savings are equal to $VOT \times \sum_{it} TTsave_{it}$ and aggregate benefits from reliability are equal to $VOR \times \sum_{it} Reliability_{it}$.

of \$3.4 million, and 68% of the benefits come from improved reliability. This corresponds to roughly \$1.30 in consumer benefits per trip, which is roughly twice the value of the average toll. If we generate benefits using the separate estimates for NB and SB drivers the total benefits increase slightly to \$3.9 million. In the NB direction 87% of the benefits are from improved reliability, whereas SB the majority of the benefits stem from time savings - only 35% of the benefits are from better reliability.

The benefits exclusively focus on the benefits to SOV drivers purchasing access to the HOT lane. WSDOT claims (WSDOT, 2012) that during the sample period HOT lane usage increased while maintaining 45 MPH over 99% of the time so there are unlikely to be negative impacts on HOV users and public transit riders. Additionally, GP speeds increased during peak periods according to WSDOT. WSDOT's estimates should not be interpreted as causal impacts of HOT implementation on traffic on SR167, but if the drivers in the GP lane experience improvements in reliability and decreases in travel times then substantial benefits will accrue to drivers on the GP lanes. However, since HOT users are wealthier on average we caution the extrapolation of VOT and VOR parameters from HOT users to the GP lane. Overall, it appears that drivers paying the toll achieve significant gains from conversion of SR167 to a HOT lane.

7 Conclusion

The scale of congestion costs in the United States warrants new approaches to managing our roads. A burgeoning approach to managing congestion is to introduce a HOT lane, either by converting existing HOV lanes or when adding new road capacity. The use of new technology, such as real time congestion pricing, gives road operators a powerful tool for managing congestion while at the same time collecting much-needed revenue. While there has been applied work in the transportation literature, as well as theoretical economic research on HOT lanes, there has been little empirical economics research using revealed preference data on how consumers respond to HOT lanes. The few studies using revealed preference data on HOT lanes have estimated a positive price response that is inconsistent with economic theory. If higher prices actually cause drivers to move into the HOT lane the basic premise of dynamically priced HOT lanes is flawed. We provide evidence that prior results are due to a failure to address issues of serial correlation and simultaneity inherent in dynamic pricing, as well as not accounting for the bundle of attributes that HOT lanes deliver.

We employ a first difference estimation strategy to recover a negative price elasticity by overcoming simultaneity in the dynamic pricing structure due to autocorrelation in travel demand. The negative demand response enables us to jointly estimate the value of time and reliability. We find a negative and substantial elasticity of approximately -0.16, indicating that causal behavioral response of drivers to higher tolls is to reduce the quantity demanded. This negative elasticity has important policy implications: if the demand for HOT lanes is not downward sloping then the entire premise of dynamically priced HOT lanes as a congestion management mechanism is fatally flawed. Given a positive price response higher prices will induce higher usage, and the cycle will continue until the lanes reaches its performance constraint.[22]

The elasticity estimates are relatively low in absolute value and may reflect that many drivers on SR167 may have set patterns - they either always use the HOT lane or always use the GP lane. Thus, we are identifying our elasticity estimate only based on the drivers who are sensitive to the toll. It should be noted that the elasticity estimate depends on the features on the SR167 HOT lane, including the pricing algorithm. The elasticity estimate can improve revenue forecasts for HOT lanes, and provide insight when developing dynamic pricing algorithms. Consumers may be relatively insensitive to price at the price intervals on SR167; more than 95% of observed prices in our sample ranged between \$0.50 and \$2.50 and the maximum price was \$6.50. Transportation planners may need to charge higher prices

[22] In our setting it was necessary to restrict HOT Lanes to HOV traffic between 0.33% and 0.14% of the time during toll hours. The variance arises from measuring closure rates at the road segment level.

is to be able to deter SOV drivers from entering HOT lanes the prices. Inelastic demand also means that setting higher toll rates will likely increase the revenue generated from HOT lanes. Further research that examines multiple HOT lanes with different pricing structures can determine the extent that the tolling algorithm impacts demand parameters.

The set of drivers sensitive to the toll is also likely a function of the dynamic pricing structure, and HOT lanes that have high prices that rapidly respond to traffic conditions may increase the set of drivers who are sensitive to the toll, and consequently the magnitude of the price elasticity. The analysis of VOT and VOR show that drivers primarily value reliability rather than time savings. There is heterogeneity in VOT and VOR; drivers value reliability during the morning commute and time savings during the evening commute. The aggregate benefits to HOT users on STR167 is estimated to be $3.4 million. The monetary benefits show that there are large benefits to drivers from using HOT lanes. If the the GP lanes also the benefit from decreased congestion due to traffic diverted to the HOT lanes the benefits may be significantly larger.

References

Agnew, Carson E, "The theory of congestion tolls," *Journal of Regional Science*, 1977, *17* (3), 381–393.

Anas, Alex and Hyok-Joo Rhee, "Curbing excess sprawl with congestion tolls and urban boundaries," *Regional Science and Urban Economics*, 2006, *36* (4), 510–541.

Arnott, Richard, "Congestion tolling and urban spatial structure," *Journal of Regional Science*, 1998, *38* (3), 495–504.

Athol, Patrick, "Interdependence of certain operational characteristics within a moving traffic stream," Technical Report 1965.

Bento, Antonio M., Kevin Roth, and Andrew Waxman, "The Value of Urgency: Evidence from Congestion Pricing Experiments," 2014.

Brownstone, David, Arindam Ghosh, Thomas F Golob, Camilla Kazimi, and Dirk Van Amelsfort, "Drivers' willingness-to-pay to reduce travel time: evidence from the San Diego I-15 congestion pricing project," *Transportation Research Part A: Policy and Practice*, 2003, *37* (4), 373–387.

Bureau of Labor Statistics, "Occupational Employment and Wages,," Technical Report, U.S. Department of Labor 2014.

Burris, Mark, Scott Nelson, Pete Kelly, Partha Gupta, and Youngjae Cho, "Willingness to Pay for High-Occupancy Toll Lanes: Empirical Analysis from I-15 and I-394," *Transportation Research Record: Journal of the Transportation Research Board*, 2012, (2297), 47–55.

Carrion, Carlos and David Levinson, "Value of travel time reliability: A review of current evidence," *Transportation Research Part A: Policy and Practice*, 2012, *46* (4), 720–741.

_ and _ , "Valuation of travel time reliability from a GPS-based experimental design," *Transportation Research Part C: Emerging Technologies*, 2013, *35*, 305–323.

Chung, Chih-Lin and Will Recker, "State-of-the-art assessment of toll rates for high-occupancy and toll lanes," in "Transportation Research Board 90$^{\text{th}}$ Annual Meeting" number 11-1331 2011.

Couture, Victor, Gilles Duranton, Matthew A Turner et al., "Speed," *Review of Economics and Statistics*, Forthcoming.

Currie, Janet and Reed Walker, "Traffic Congestion and Infant Health: Evidence from E-ZPass," *American Economic Journal. Applied Economics*, 2011, *3* (1), 65.

Dahlgren, Joy, "High-occupancy/toll lanes: where should they be implemented?," *Transportation Research Part A: Policy and Practice*, 2002, *36* (3), 239–255.

Devarasetty, Prem Chand, Mark Burris, and W Douglass Shaw, "The value of travel time and reliability-evidence from a stated preference survey and actual usage," *Transportation Research Part A: Policy and Practice*, 2012, *46* (8), 1227–1240.

Downs, Anthony, "The law of peak-hour expressway congestion," *Traffic Quarterly*, 1962, *16* (3).

_ , *Still Stuck in Traffic: Coping with Peak-Hour Traffic Congestion*, Brookings Institution Press, 2004.

Duranton, Gilles and Matthew A Turner, "The Fundamental Law of Road Congestion: Evidence from US Cities," *American Economic Review*, 2011, *101* (6), 2616–2652.

Finkelstein, Amy, "E-ZTAX: TAX SALIENCE AND TAX RATES.," *Quarterly Journal of Economics*, 2009, *124* (3).

Gibson, Matthew and Maria Carnovale, "The effects of road pricing on driver behavior and air pollution," *Journal of Urban Economics*, 2015, *89*, 62 – 73.

Holtz-Eakin, Douglas, "Testing for individual effects in autoregressive models," *Journal of Econometrics*, 1988, *39* (3), 297–307.

_ **, Whitney Newey, and Harvey S Rosen**, "Estimating vector autoregressions with panel data," *Econometrica*, 1988, pp. 1371–1395.

Im, Kyung So, M.Hashem Pesaran, and Yongcheol Shin, "Testing for unit roots in heterogeneous panels," *Journal of Econometrics*, 2003, *115* (1), 53–74.

Janson, Michael and David Levinson, "HOT or not," *Research in Transportation Economics*, 2014, *44*, 21–32.

Kirk, Robert S and William J Mallett, "Funding and Financing Highways and Public Transportation," *Congressional Research Service*, 2013, p. 23.

Konishi, Hideo and Se il Mun, "Carpooling and congestion pricing: HOV and HOT lanes," *Regional Science and Urban Economics*, 2010, *40* (4), 173–186.

Li, Jianling, "Explaining high-occupancy-toll lane use," *Transportation Research Part D: Transport and Environment*, 2001, *6* (1), 61–74.

Li, Zheng, David A Hensher, and John M Rose, "Willingness to pay for travel time reliability in passenger transport: A review and some new empirical evidence," *Transportation research part E: logistics and transportation review*, 2010, *46* (3), 384–403.

Lindsey, Robin, "Reforming road user charges: A research challenge for regional science," *Journal of Regional Science*, 2010, *50* (1), 471–492.

Liu, Xiaoyue, Guohui Zhang, Yao-Jan Wu, and Yinhai Wang, "Measuring the Quality of Service for High Occupancy Toll Lanes Operations," *Procedia - Social and Behavioral Sciences*, 2011, *16*, 15–25.

_ , _ , Yunteng Lao, and Yinhai Wang, "Quantifying the Attractiveness of High-Occupancy Toll Lanes with Traffic Sensor Data Under Various Traffic Conditions," *Transportation Research Record: Journal of the Transportation Research Board*, 2011, *2229* (-1), 102–109.

Matas, Anna and José-Luis Raymond, "Demand elasticity on tolled motorways," *Journal of Transportation and Statistics*, 2003, *6* (2), 91.

Parry, Ian W.H. and Antonio Bento, "Estimating the Welfare Effect of Congestion Taxes: The Critical Importance of Other Distortions within the Transport System," *Journal of Urban Economics*, 2002, *51* (2), 339–365.

Parry, I.W.H., "Comparing the efficiency of alternative policies for reducing traffic congestion," *Journal of Public Economics*, 2002, *85* (3), 333–362.

Peer, Stefanie, Carl C Koopmans, and Erik T Verhoef, "Prediction of travel time variability for cost-benefit analysis," *Transportation Research Part A: Policy and Practice*, 2012, *46* (1), 79–90.

Pesaran, M Hashem, "Estimation and inference in large heterogeneous panels with a multifactor error structure," *Econometrica*, 2006, *74* (4), 967–1012.

Pesaran, M. Hashem, "A simple panel unit root test in the presence of cross-section dependence," *Journal of Applied Econometrics*, 2007, *22* (2), 265–312.

Safirova, Elena, Kenneth Gillingham, Ian Parry, Peter Nelson, Winston Harrington, and David Mason, "Welfare and Distributional Effects of Road Pricing Schemes for Metropolitan Washington, DC," *Research in Transportation Economics*, 2004, *9*, 179–206.

Schrank, David, Bill Eisele, and Tim Lomax, "TTI's 2012 urban mobility report," *Proceedings of the 2012 annual urban mobility report. Texas A&M Transportation Institute, Texas, USA*, 2012.

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

_ , Clifford Winston, and Jia Yan, "Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability," *Econometrica*, 2005, *73* (4), 1367–1382.

U.S. Department of Transportation, "The Value of Travel Time Savings: Departmental Guidance for Conducting Economic Evaluations Revision 2 (2014 Update)," Technical Report, U.S. Department of Transportation 2014.

Verhoef, Erik T and Jan Rouwendal, "Pricing, capacity choice, and financing in transportation networks," *Journal of Regional Science*, 2004, *44* (3), 405–435.

Vickrey, William S, "Pricing in urban and suburban transport," *The American Economic Review*, 1963, pp. 452–465.

Weitzman, Martin L, "On modeling and interpreting the economics of catastrophic climate change," *The Review of Economics and Statistics*, 2009, *91* (1), 1–19.

Wilbur Smith Associates, "SR167 HOT Lanes Pilot Project Traffic and Revenue Analysis," Technical Report, WSDOT 2006.

Winston, Clifford, *Last exit: Privatization and deregulation of the US transportation system*, Brookings Institution Press, 2010.

Wooldridge, JM, *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT press, 2010.

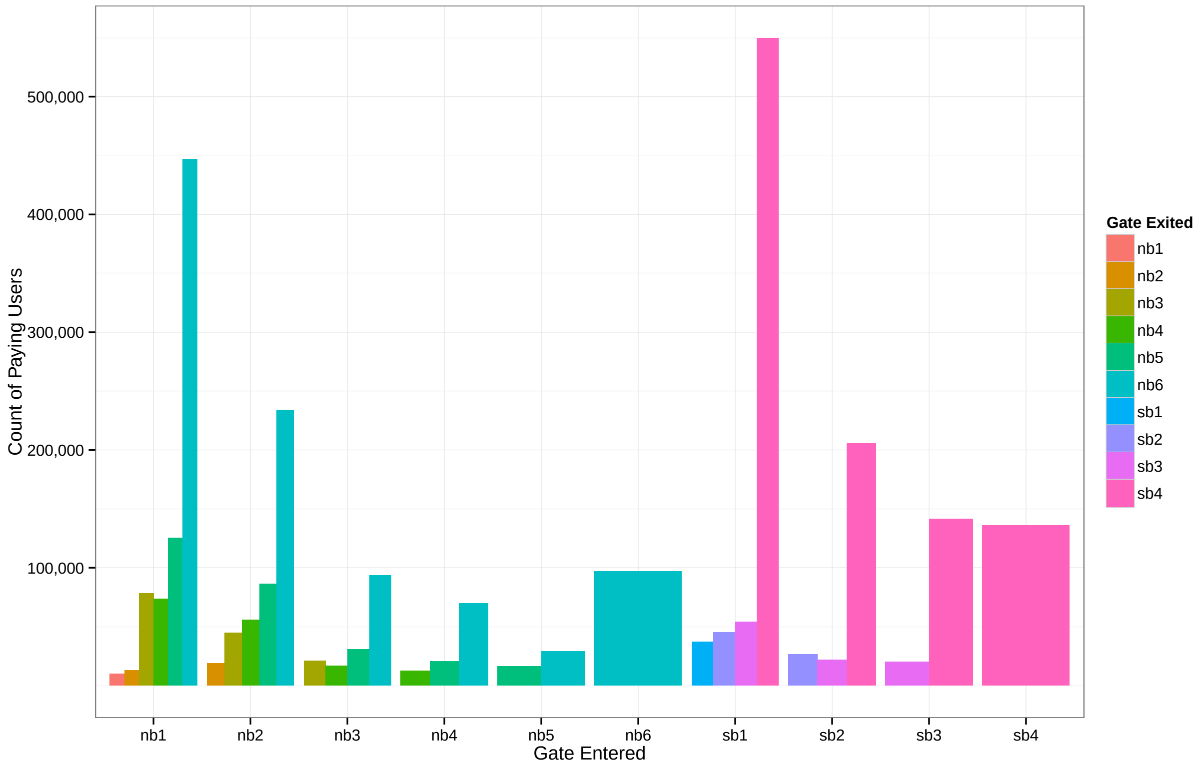
WSDOT, "SR 167 HOT Lanes Pilot Project Fourth Annual Performance Summary," Technical Report May 2008, WSDOT 2012.

_ , "SR 167 HOT Lanes - Project Library," Technical Report 2013.

A Online Appendix

A.1 Trip Combinations

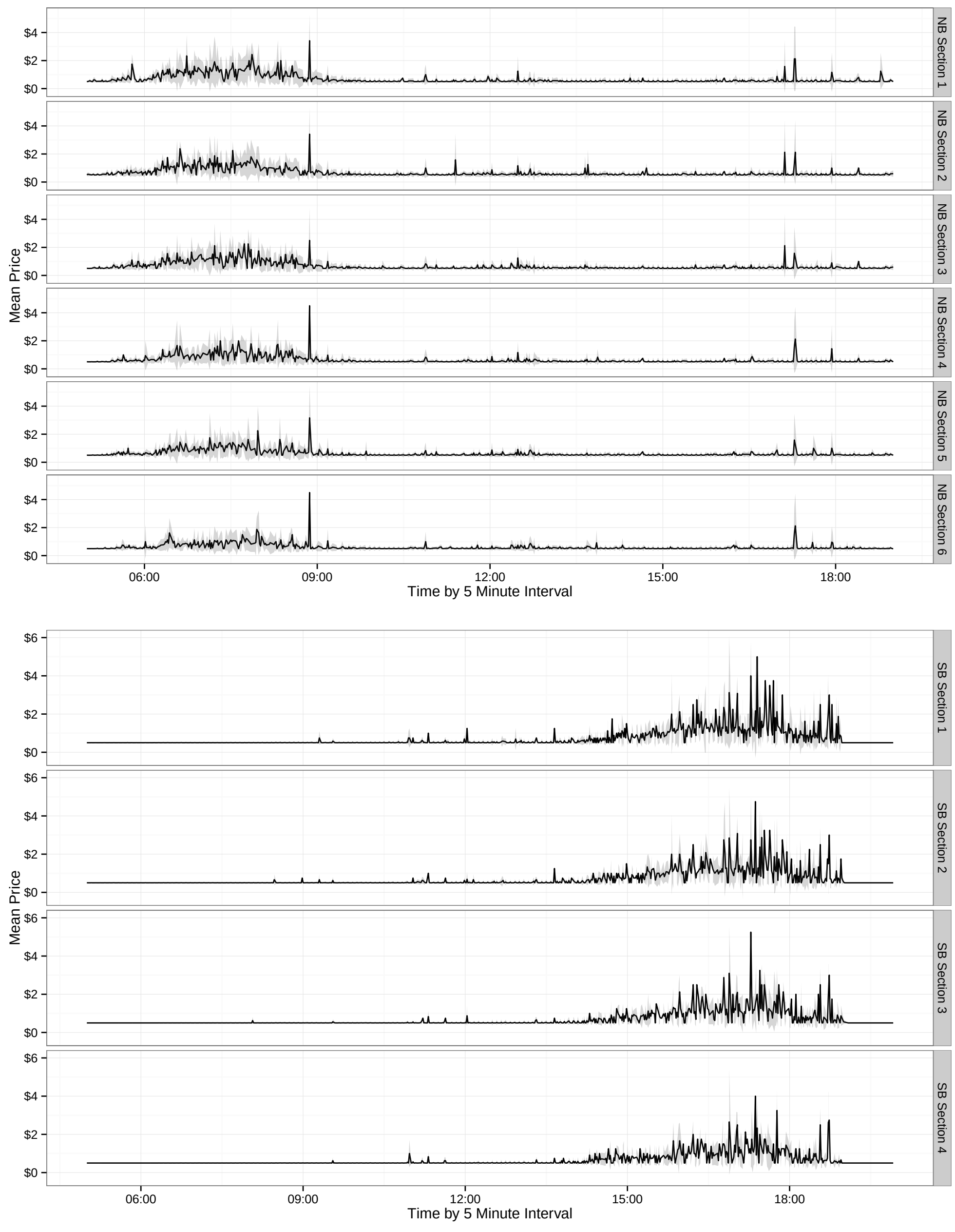
Figure A.1: SR167 trip combinations



Notes: The horizontal axis indicates the entry gate and the colors designate exit gates. The number of purchased tolls for each combination is on the vertical axis. The thickness does not convey variation in the data, and is meant to properly space the figure.

A.2 Prices by gate

Figure A.2: Price by Section



Gate locations were determined using this tool.

North Bound:

- NB1 13.83
- NB2 15.10
- NB3 16.60
- NB4 18.61
- NB5 20.28
- NB6 22.90
- ends 25.09

South Bound:

- SB1 25.64
- SB2 23.70
- SB3 20.53
- SB4 18.99
- ends 17.03

A.3 Panel Data

Our data can be interpreted as a 'pseudo' panel in the sense that we repeatedly observe usage at individual entry gates. Structured this way, $N = 10$ (six northbound plus four southbound gates) and $T = 1,215,000$ (although T will vary according to different missing data at different gates). 'Pseudo' panels such as we are proposing here are likely to violate the dynamic homogeneity assumption underlying true panel data models (Im et al., 2003). Understanding that the data might display dynamic heterogeneity informs our choice of unit-root tests presented later in this section.

We employ three tests to explore the heterogeneity of our 'pseudo' panel based on Holtz-Eakin et al. (1988); Holtz-Eakin (1988):

1. an F-test for parameter equality from an Augmented Dickey-Fuller estimation with third order lag (ADF(3)),

2. another F-test from a third order autoregressive (AR(3)) regression across all variables and

3. finally White's test for groupwise heteroskedasticity.

A rejection of the null in the F-tests indicates heterogeneity across parameters while a rejection of the null in White's test indicates an inequality of variance. We performed White's test using a regression of the residuals from the ADF(3) regression on the original regressors and their squares. The test statistic is $(NT) * R^2 \sim \chi^2$, where the degrees of freedom are the number of regressors from the second stage. Table A.1 presents the statistics, all of which reject the null of parameter and variance homogeneity at the 1% significance level.

Table A.1: Dynamic Heterogeneity Tests

Test	White	ADF	AR
F Statistic	261017.62	2037.24	1018.45

Notes: White's test tests the equality of variance, the ADF(3) and AR(3) test for parameter equaltiy across an ADF and AR equation respectively. The number of obseations is $1,282,349$

Given parameter and variance heterogeneity we employ two unit root tests for panel data: that developed by Im et al. (2003) and a cross-sectionally augmented version of Im et al. (2003) proposed by Pesaran (2007). We begin with the following formula:

$$y_{it} = \alpha_i + \beta_i y_{i,t-1} + \epsilon_{it} \tag{6}$$

where $i = 1, \ldots, N$ for each gate, $t = 1, \ldots, T$ represents the time periods and y_{it} represents each series in the panel. The traditional panel unit root test developed by Im et al. (2003) fits the following ADF equation

$$\Delta y_{it} = \alpha_i + \beta_i y_{i,t-1} + \sum_{j}^{p_i} \rho_{ij} \Delta y_{i,t-j} + \epsilon_{it} \tag{7}$$

where p_i is the number of lags (here three). This is estimated for each variable in the panel across each cross section. A 't-bar' statistic is formed as the average of the individual cross sectional ADF(3) statistics according to the following equation

$$t\text{-bar} = \frac{1}{N}\sum_{i=1}^{N} t_{\rho_i} \tag{8}$$

The null is that each series in the 'pseudo' panel contains a unit root. Rejection of the null indicates that there is no unit root in any cross sectional series. Im et al. (2003) show that the statistic is normally distributed under the null and provide critical values for given N and T. The t-bar for each series is presented in Table A.2 and reject the null of unit roots at about the 1% significance level.

This traditional Im et al. (2003) unit root test can be modified to account for cross-sectional dependence. Pesaran (2006) shows that the effects of unobserved common factors in panel data can be eliminated by filtering the cross-sectional mean. Extending this work to unit roots in Pesaran (2007) leads to the following cross-sectionally augmented DF (CADF) regression

$$\Delta y_{it} = \alpha_i + \beta_i y_{it-1} + \gamma_i \bar{y}_{t-1} + \delta_i \Delta \bar{y}_t + \epsilon_{it} \tag{9}$$

where $\bar{y}_t$ is the cross-sectional mean. The t-statistics on the β coefficient are estimated from each unit i of the panel, with the average forming the CIPS statistic

$$\text{CIPS} = t\text{-bar}\frac{1}{N}\sum_{i=1}^{N} t_{\beta_i} \tag{10}$$

Results from the tests are presented in Table A.2 and indicate there are no unit roots in the series.

Table A.2: IPS Panel Unit Root Tests

Variables	ADF	ADFtrend	CADF	CADFtrend
Count	-56.67	-56.79	-57.55	-57.69
Price	-45.67	-45.80	-42.48	-42.51
Volume After Gate	-56.25	-56.97	-60.10	-64.28
Speed Before Gate	-68.20	-68.25	-68.55	-68.92
Speed at Gate	-56.74	-57.07	-61.90	-62.54
Speed After Gate	-63.15	-64.50	-66.17	-69.22
Value of Reliability	-40.68	-40.68	-28.83	-28.83
Expected Time Savings	-29.38	-29.38	-15.53	-15.51
Proportion	-59.75	-61.07	-61.84	-64.23

Notes: All tests are conducted using three lags. Columns with 'trend' include an individual time trend in the regression. All tests reject the null of an unit root at greater than 1% significance. Total number of observations is $1,282,349$.