Behavioral Responses to Congestion Pricing

Austin Gross

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Robert Halvorsen, Chair

Jacques Lawarree

David Layton

Program Authorized to Offer Degree:
Economics

University of Washington

Abstract

Behavioral Responses to Congestion Pricing

Austin Gross

Chair of the Supervisory Committee:
Professor Robert Halvorsen
Economics

Traffic congestion is costly both in terms of time lost as well as resources squandered. Road infrastructure is also expensive to create and maintain, with current funding mechanisms falling short. While there is strong theoretical support for road pricing solutions they have been sparsely deployed in the United States. We begin with a consideration of the current theory and politics underpinning road pricing. We then turn to empirical examples of two different road pricing mechanisms: one that adapts to traffic conditions in a High Occupancy Toll (HOT) lane and another that imposes a toll varying only by time of day on a bridge. In both cases we will estimate the behavioral response of drivers to toll pricing.

In the dynamic pricing context 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 and that their responses to value of time and value of reliability are very large. 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.

In the time-of-day pricing context we estimate commuters' responses by measuring the shift from the tolled bridge to a parallel free bridge and also how commuters adjust the timing of trips to less expensive periods. We estimate own and cross-price elasticities for the two bridges finding results in line with previous research. We also estimate the time adjustment that commuters make in response to changing toll rates. These estimates allow us to compute the benefits to toll paying commuters from decreased congestion resulting from the imposition of the tolls.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to Robert Halvorsen, David Layton, Jacques Lawarree, Mark C Long and Robert Plotnik, all of whom provided excellent suggestions and helpful criticism. Special thanks are owed to Daniel Brent for originally planting the seed of studying congestion policies and also for being a wonderful collaborator and friend.

The author is also grateful to Mark Hallenbeck for many insightful discussions and endless patience with data requests. Sara Myers and Tyler Patterson from WSDOT were also incredibly helpful and generous. Seminar participants at AERE and WEAI conferences also provided useful feedback.

DEDICATION

to my dear wife Liane and our children Ariana, Oliver and Nikolas

Chapter 1

CONGESTION PRICING POLICY

1.1 Introduction

With transportation infrastructure spending again under consideration in Washington DC it is timely to review the state of congestion pricing as both a revenue mechanism and demand management tool. The American Society of Civil Engineers estimate a $836 billion backlog of operations and maintenance projects for bridges and roads.[1] The current iteration of the presidential plan for infrastructure involves raising $1 trillion dollars through 'public-private' partnership.[2] Yet transportation projects themselves will be unattractive to private capital without associated revenue streams. With revenue most likely to come from some variety of direct user-fees, issues of equity, that have plagued transport pricing for years [Levinson, 2010], will again come to the fore.

Revenue is not the only problem facing American roadways. The news is filled every day with references to congestion and attempts to solve it.[3] Ranging from an outright ban of vehicles from city centers[4] to dynamic tolling systems that respond to real-time traffic levels [Brent and Gross, 2017], solutions to the single problem of congestion are legion.

---

[1] http://www.infrastructurereportcard.org/roads/funding-future-need/ accessed on 2017-05-13.

[2] https://www.usnews.com/news/the-report/articles/2017-03-01/donald-trumps-changed-tune-encouraging-for-infrastructure-reform accessed on 2017-03-04.

[3] A Google news search for 'traffic congestion' returned 664K 'recent' results. Searching for 'traffic congestion solutions' returned 29.5K results.

[4] http://www.theguardian.com/environment/2015/oct/19/oslo-moves-to-ban-cars-from-city-centre-within-four-years accessed on 2017-03-04.

Nor is it merely a case of journalistic sensationalism. Huge investments on behalf of the public have been poured into road transportation. The final cost of the Interstate System alone, estimated in 1991, was \$128.9 billion and only included mileage from the Interstate Construction Program [Federal Highway Administration, 2015]. The variable costs associated with consumer use of the public road network is also enormous: Winston [2010] estimates that in 2007 American consumers spent over \$1 trillion on gasoline and vehicles. Road congestion is a very large portion of this cost: Schrank et al. [2012] estimate that in 2011 consumers purchased an *additional* 2.9 billion gallons of fuel and spent 5.5 billion hours stuck in traffic at a total cost of \$121 billion. Despite the enormous annual cost of congestion most roads do not have even the simplest form of congestion pricing.

Economists have long argued for road pricing and analyzed the myriad ways in which consumers would benefit [Vickrey, 1963, Winston, 1985, Small, 1983, Winston, 2010]. Yet despite strong positivist arguments for increased consumer welfare from optimal location decisions in the presence of road pricing [Parry and Bento, 2002, Small, 1997, De Lara et al., 2013] only recently has pricing of public roads gathered momentum. Nonetheless, there are large political obstacles to the imposition of tolls on currently free-to-use roads. Politicians charged with care of the public infrastructure have been exceedingly slow in implementing congestion pricing and weak in defending it once implemented [Altshuler, 2010].

It is surprising to note that, given the public's perception of roads as a 'free' resource, for a majority of European history roads were constructed and maintained through tolling. Even Adam Smith saw fit to discuss direct user fees in his section on public works [Smith, 1817]. Lindsey [2006] provides an excellent overview of the historical development of road pricing in economics. Indeed, Lindsey notes that equity is one of Adam Smith's first points in the consideration of road pricing in that it directly ties maintenance of the roads to those who degrade them through use (tolls in proportion to weight). But this form of 'equity', in that those who benefit bear the cost, is found to be perverted as reported by the Seattle Times regarding the I405 High Occupancy Toll (HOT) lanes, "The benefits would be more fairly and efficiently distributed if that capacity were in the form of general-purpose lanes" an opponent to the lanes argues. [5] This flies in the face of years of research into demand management of transportation resources where consensus has developed again and again around pricing road ways to combat congestion. Even in the time of Jules Dupuit [Lindsey, 2006] (he developed models to calculate the toll that would recover costs) we had the mathematical tools to manage usage through pricing. Indeed, Dupuit used the idea of diminishing marginal utility to develop the downward sloping demand curve, an idea that can sometimes be overlooked in modern work (see Brent and Gross [2017]). He also incorporated ideas of travel time and understood tolls from the standpoint of revenue generation rather than of congestion management per se.

Ding and Song [2012] present three classic paradoxes of supply side congestion management with accompanying solutions.

1. Pigou-Knights-Down: adding capacity to congested routes may not decrease congestion since traffic may shift to the upgraded route. If the increase in capacity is twice the travel flow then there will be a decrease in congestion.

2. Downs-Thompson: increased capacity of road segments decreases usage of competing transit link, worsening overall performance. Charging a toll that minimizes the total social cost corrects the externality.

3. Braess: adding a new link to a transit system can cause total travel time to increase. Again, appropriate tolls correct the externality.

All of these are well established in the literature. Indeed, the major contribution of Ding and Song [2012] was to present these paradoxes and their solutions in a 'more general form'.

---

[5] http://www.seattletimes.com/seattle-news/transportation/the-i-405-toll-lane-experiment-hows-it-working-for-drivers-state/ highlights many strong opinions on both sides.

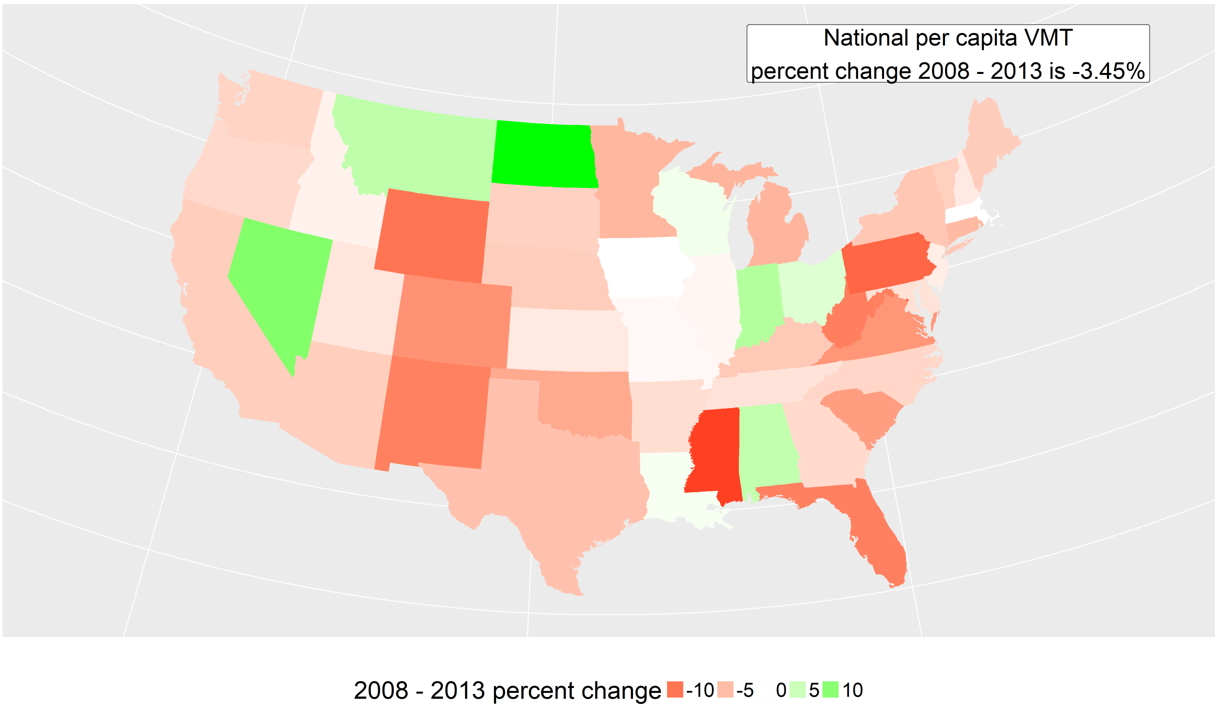
Nonetheless, it seems that among some portions of the population the lesson has yet to be learned. There are mixed signs of progress: policy makers are turning more readily to tolling as a means of alleviating years of budget woes, but they may be politically incapable of charging the appropriate tolls. I will discuss this at further length in Section 1.4.

Returning to the idea of 'free' access quoted above, the public seems unwilling to grapple with the perennially imminent failure of historical funding mechanisms. Road infrastructure budgets rely on gasoline taxes at the federal and state levels. And while we have been discussing the worsening problem of traffic congestion, it is important to note that per capita Vehicle Miles Traveled (VMT) is falling in the US (Figure 1.1) suggesting both that the Highway Trust Fund will continue to shrink and that *overall* commuting is not the problem to be solved as much as specific time of day trips. Indeed, according to the United States Department of Transportation [2015] only ten states have increased per capita VMT between 2008 and 2013[6] with North Dakota and Nevada increasing by 14% and 10% respectively. While this decrease is likely due to the recession during that period, if these trends continue then budget woes will only worsen at a time when peak congestion is itself intensifying.

---

[6] Alabama, Indiana, Iowa, Louisiana, Massachusetts, Montana, Nevada, North Dakota, Ohio, Wisconsin

Figure 1.1: US per capita VMT percent change 2008 to 2013, [United States Department of Transportation, 2015]



1.2 History

Currently, the Federal Highway Administration considers the four categories of road pricing as [Federal Highway Administration, a] demand management:

1. priced lanes,

2. priced highways,

3. priced zones,

4. and priced road networks.

If we consider that a majority of the travel on roads in the US is by gasoline vehicle, then in a very real sense we are currently pricing the entire road network through the use of gasoline taxes. The Highway Trust Fund (HTF), which is maintained by the federal government, is funded by a $0.184 per gallon tax on gasoline. Insofar as any driving is then subject to this tax it is likely effective in reducing overall Vehicle Miles Traveled (VMT) [Knittel and Sandler, 2010]. Even here, however, the efficiency of the current level of gasoline tax is an active area of research. Parry and Small [2005] investigates the optimal tax rates for the US and Britain, finding that US rates should double and British rates should be halved to reach an optimum. Importantly, they note that their research suggests little effect on congestion since there are adjustment mechanisms other than simply driving less. They also note that taxing VMT would be more effective, but this again ignores the high degree of variability in demand for specific roads at specific times and also that per capita VMT is already decreasing (see Figure‘1.1).

The gasoline tax, however, ignores the heterogeneity of demand for specific roadways at specific times. It is thus a very blunt instrument for reducing congestion, which arises from a combination of local factors. Furthermore, since fuel consumption varies widely across vehicles, increases in the gas tax might simply be met with the purchase of more efficient vehicles. Clearly, then, some combination of zonal, segment and lane pricing could be more effective depending on various factors.

The Federal Highway Administration [b] provides a short introduction to the history of toll facilities in the US. Noting that the first turnpike in the US was chartered in 1792, there are examples in every stage of this nation’s history of using tolls to produce the revenue required for construction and maintenance of bridges, tunnels and roads. Yet, despite this long-running direct financial relationship between the resource and those who use it, the modern era of US transportation policy has largely been defined by the dissemination of federal funds for construction and maintenance. This shift was already under way long before the construction of the modern interstate highway. In 1921 Congress passed the Federal Highway Act to provide financial assistance to the states to build roads and bridges. In 1956 the Federal-Aid Highway Act was signed into law, using tax-payer funds for the construction of the coast-to-coast Interstate System. This led to the predictable dying-out of new toll projects, the last of which was completed in 1963 having been planned before the Federally funded Interstate System. It fundamentally changed the way we consider transport costs and benefits: national interest prompted the construction of a national network that suffers from local congestion and local shortfalls in funding.

Contemporaneous to the passage of the Federal Highway Act were developments by Pigou [1920] in using tolls to combat congestion. He notes that 'taxing' the most popular route would redistribute some 'carts', increasing their 'trouble' while leaving other carts on the original route better off. This argument led to a long strand of articles arguing back and forth about the necessity of taxation when private-ownership would lead to an optimal result based on quality differentiation. This in turn was criticized as only being true under strict assumptions [Lindsey, 2006].

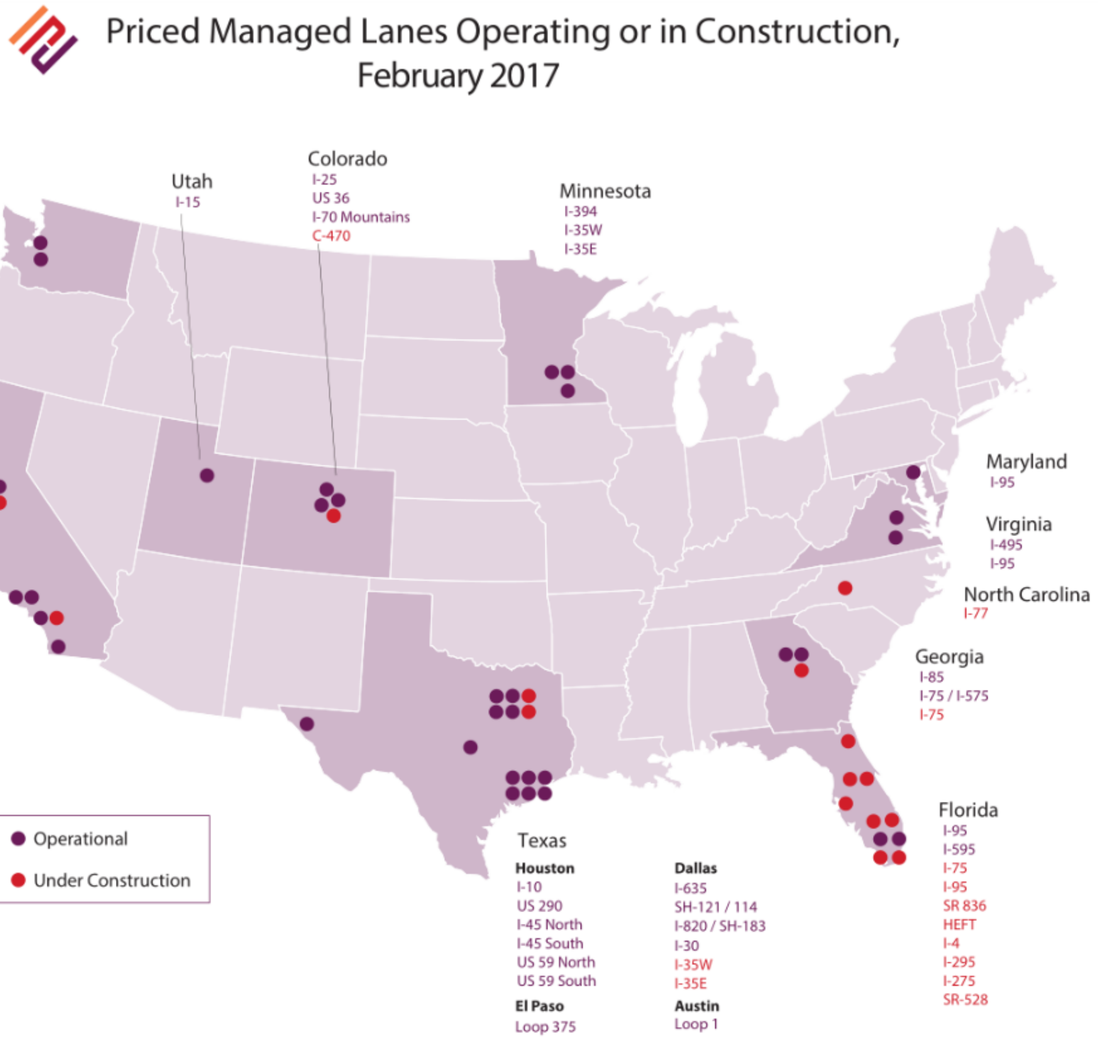
Priced lanes are being constructed in diverse locations throughout the US. Figure 1.2 shows road pricing projects as of 2017. It is clear both from the number of existing facilities and from those 'under construction' that priced lanes are very popular, at least with state transportation agencies. Table 1.1 gives a flavor of the diversity of years, mileage and pricing methods employed in these priced lanes. Perhaps most important is the concentration of facilities in a handful of states. It seems from Figure 1.2 that a majority of states have not been willing to experiment with HOT lanes.

Taking Tennessee as an example, three reasons provided by a traffic consultant as to why HOT lanes are not being considered in the state[7] are illuminating of the general perception

---

[7] http://www.bizjournals.com/nashville/news/2016/12/16/why-an-easy-way-to-reduce-traffic-and-fund-

Figure 1.2: Source: FHWA



working against them:

1. the technology to collect fees from drivers is costly,

---

transit.html

Table 1.1: Selection of Priced Lanes in US

Location	Name	Length	Pricing	Pricing Began
Alameda/S. Clara	I-680	14 miles	Dynamic	2010
Atlanta	I-85	15.5 miles	Dynamic	2011
Dallas	TEXpress Lanes	70 miles	Dynamic	2014
Denver	I-25	7 miles	Variable	2006
Houston	Metro Lanes	13 miles	Dynamic	1998
Houston	Katy Managed Lanes	13 miles	Variable	2009
Miami-Ft. Lauderdale	I-95	7 miles	Dynamic	2008
Minneapolis-St. Paul	I-394	11 miles	Dynamic	2005
Minneapolis-St. Paul	I-35W	16 miles	Dynamic	2009
Orange County	SR 91	10 miles	Variable	1995
Salt Lake City	I-15	40 miles	Dynamic	2006
San Diego County	I-15	16 miles	Variable	1996
Seattle	SR 167	11 miles	Dynamic	2008
Seattle	I-405	15 miles	Dynamic	2015

2. congestion pricing is often seen as a regressive tax,
3. they are illegal in Tennessee, where state law does not allow tolling of existing roads.

Compare these to the findings from Altshuler [2010]:

1. they are low-cost alternatives to adding lanes,
2. they are Pareto-improving in that a subset benefits while leaving no one worse off by alleviating conditions in the general purpose lane,

3. they recapture "wasted" capacity from under-utilized High Occupancy Vehicle (HOV) lanes.

It seems the debate has been poorly framed. The most interesting point from the consultant's list is the illegality of HOT lanes in Tennessee. That legislation exists at the state level to block these facilities is a testament to the resentment against the idea of pricing some public resources. Even though Florida has many new HOT facilities, it remains unpopular enough that legislation to end tolling after bonds have been repaid was introduced into the stat legislature.[8]

1.3 Why Should it Work?

Vickery argued for short-run marginal cost pricing of road ways, and indeed that variations in the short-run marginal cost should be immediately addressed [Lindsey, 2006]. This is the essence of a dynamic pricing mechanism such as the one employed by WSDOT on SR167. He further advocated network-wide tolling of roads in Washington DC. Ideas such as these eventually led to the first area licensing scheme to be introduced in Singapore in 1976. This created cordon pricing to enter the six km area of the central business district. This led to a 77% reduction in traffic [Spencer and Sien, 1985]. HOT lane pricing finally came to the DC area in 2008 with the installation of the 495 Express lanes.

Recent work by Liu et al. [2014] consider various road-pricing, rationing and cordoning policies and their effects on social-welfare. While their work employs a linear mono-centric city their simplifying assumptions produce interesting results in that the first-best price is not Pareto-improving, while various combinations of less efficient policies do result in Pareto improvement. This is an attractive result in that it reflects the complexity of interacting influences on commute decisions by consumers. Indeed their results cover a wide array of

---

[8] http://www.miaminewtimes.com/news/i-95-express-lanes-could-be-banned-under-proposed-law-finally-9044048 article regarding Sen. Frank Artiles introducing bill into Florida legislature.

possibilities such as rationing commute ability, road pricing, a cordon around the CBD and park and ride installations. That varying combinations are not always Pareto-improving speaks to the delicate balance required to construct resilient systems.

Contrasting work by Ekström et al. [2014] attempts to solve a similar problem, but instead constructs a Wardrop [1952] user equilibrium[9] with variables corresponding to toll location, level, travel flow and demand. The authors use a sensitivity analysis heuristic to solve for the optimal toll levels and locations. They present a case study of Stockholm, finding that optimizing both the cordon toll rate and adding tolls to linked roads would create improvements in social welfare. Indeed all the scenarios they consider result in improvements to social welfare, and while Liu et al. [2014] find more nuance in which policies are Pareto improving, it demonstrates the degree to which it is easier to benefit society as a whole rather than ensure that the benefit improves all groups within society.

Work by Chung and Recker [2011] into the toll mechanisms for HOT lanes finds that most lanes are priced according to some combination of flow, density and speed. They end with a suggestion that travel time savings and reliability be included in the toll calculation. But if we zoom out for a moment, consider that this is essentially the problem that private sector variable service providers are also trying to solve. Banerjee et al. [2016] construct a model to describe two-sided market of drivers and passengers. The interesting result in a HOT context is that dynamic pricing can help to discover the ‘correct’ price by adjusting between high and low prices in various conditions. From this perspective the HOT platforms are ripe for price experimentation to continuously adjust prices according to a host of unobservable factors, including the heterogeneity of Value of Time (VOT) and Value of Reliability (VOR).

More recent work by Lupton [2017] proposes a simple (practical) rule for setting tolls that optimizes the flow through any given segment of the road network. Citing the near-

---

[9] A Wardrop user equilibrium states that users have chosen their trip paths such that travel times are minimized across all available routes. A Wardrop social equilibrium states that the average journey time across the entire system is at a minimum.

impossibility of discovering the demand curve for all cohorts of drivers, Lupton argues that the optimal Pigouvian toll might be higher than the flow maximizing toll. He is essentially arguing for a VOT derived toll, in that the toll is proportional to the difference between actual and free flow travel-times. This, however, ignores the VOR and thus potentially the most important variable in the decision to use the toll facility. The main results of Brent and Gross [2017] show that, depending on specification, VOR can be as much as 5x more important in the HOT use decision than VOT.

There seems, however, to be little consideration for these factors when implementing new tolling regimes. For instance, in the 2007 construction of the tolling algorithm for SR 167 in Seattle, the only consideration is speed and flow in the HOT lane. As discussed above, prices constructed in this fashion could only hope to address the marginal social cost of an additional driver in the HOT lane. Chung and Recker [2011] construct optimal tolls using VOT and VOR, which incorporates performance of the General Purpose (GP) and HOT lanes, finding in an empirical application to SR 91 in California that by ignoring the benefit to drivers tolls were too low when speeds in the GP lane were low and too high when speeds in the GP were high. This mismatch could lead to distortions of over use and degraded performance and will be discussed in Section 1.4.

1.4 Does it Actually Work?

Perhaps some of the best evidence for a variety of congestion management strategies can be recovered from Zimmerman et al. [2015] where the USDOT provided approximately \$800 million in grants to six metropolitan areas to reduce congestion. Table 1.2 is taken from the report and shows the breakdown of the Urban Partnership Alliance (UPA) projects by category and metro. Four of the six metros included conversion of HOV to HOT lanes, while two experimented with either variable pricing on parking or variable tolling on a bridge. Five out of the six also included changes to transit, while all six invested in remote work programs and also new technology to support the related infrastructure changes.

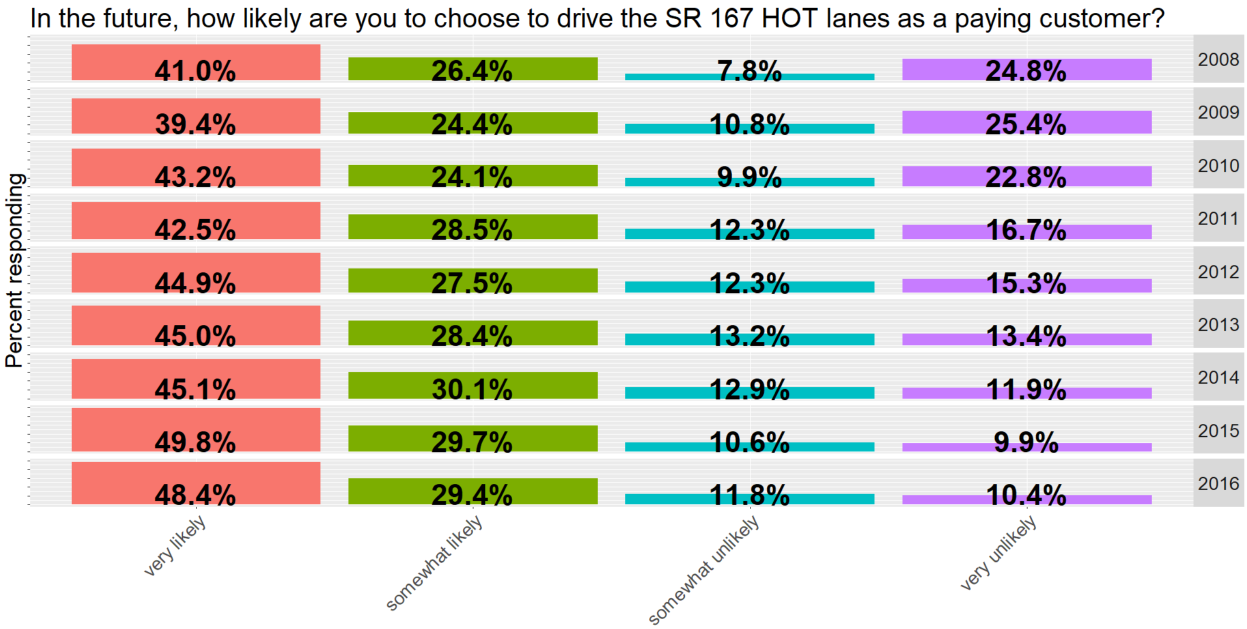
Given the degree to which many variables changed simultaneously it would be difficult to disentangle the effects of any one change towards the outcomes observed. Seemingly with that in the mind the report offers some high-level reporting using pre and post values for some joint measures (not all projects collected data on all measures). Table 1.3 presents selected values for average travel time pre and post, for general purpose and HOV/HOT lanes. The results are mixed in terms of sign (travel time increases in GP lanes in Atlanta Southbound) and magnitude (many of the gains are fractional minutes). These changes at the mean might mask more substantial shifts elsewhere in the distribution, but are nevertheless suggestive of what might be expected for the next $800 million in spending. They of course do not represent a true policy evaluation arrived at by counterfactual methods. The report notes that public perception of many of the changes was positive. These mixed results are in line with research regarding London's congestion tax, where initial benefits decreased over time and the authors found other policies predating the cordon to also have been effective [Givoni, 2012].

1.4.1 Seattle

The Seattle metropolitan area has moved decisively over the last decade in the direction of either segment or lane pricing (see Table 1.1). Between 2008 and 2018 Seattle metro has added priced lanes on SR-167 and I-405, and also reintroduced whole-segment tolling on SR-520.

In the yearly survey of SR167 users there has been a marked decrease in the percent of users responding that they were 'very unlikely' to use the HOT lanes again as a paying customer (see Figure 1.3). Concurrent with this was a degradation in the performance of the HOT lanes, where the frequency of lane closure to new paying users increased, indicating that the demand for the lane was too high for the toll rate charged. The lowering of reported resistance to lane usage coupled with the observed over-use of the lane indicates that initial resistance to paying for access has eroded over time and that current pricing schedules need to be updated to account for this increased demand.

Figure 1.3: SR 167 survey responses



Similar findings apply to I-405, where analysis of traffic data indicates that the facility is not meeting the mandated minimum speeds of 45mph at least 90% of the time [Washington Joint Transportation Committee, 2018]. Furthermore there was no significant change in average speed in the general purpose lanes after the HOT lanes were constructed. As in SR167 the culprit appears to be the degree to which toll rates are reaching their maximum, indicating that the willingness to pay for the lanes exceeds the current price schedule. For example, in March 2017 the maximum toll was charged on approximately 25% of trips in the north and south bound directions. The authors note that prices are also out of sync with traffic momentum: drivers paid as much as $4 less at the start of their trip than the price required by conditions at the end of their trip. Recommendations by the authors point to changing the price algorithm to be more sensitive to changing conditions as well as raising the maximum price.

The findings from the dynamic tolling facilities can be compared with the time of day tolls on SR520, where dramatic reductions in traffic volume (36% decrease) after tolling was introduced have been followed by slow recovery of volume (between 1 - 5.9% increases) afterwards [CDM Smith, 2017]. These increases on SR520 have been matched by increases on I-90, an alternative route, and although the relative increases were modest compared to SR520 they are on a base roughly twice as large as SR520. And while these results are not directly comparable to a HOT lane, since the entire bridge was tolled, they indicate changing attitudes towards paying for public transportation resources.

Table 1.2: UPA projects by category and metro [Zimmerman et al., 2015]

Atlanta	**Los Angeles**	**Miami**	**Minneapolis**	**San Francisco**	**Seattle**
Oct. 2011	I-110 in Nov. 2012; I-10 in Feb. 2013; Express Park in May 2012.	Dec. 2008 northbound; Jan. 2010 southbound.	Phase 1 in Sept. 2009; Phase 2 in Nov. 2010.	Sept. 2011	Dec. 2011
Tolling					
HOV to HOT lane conversion on I-85 (Express Lanes).	HOV to HOT lane conversion on I-10 and I-110 (ExpressLanes). A second HOT lane was added on I-10. Also, demand-based parking pricing (LA Express Park™).	HOV to HOT lane conversion on I-95 (95 Express). Also added a second HOT lane in each direction.	HOV to HOT lane conversion and addition of new HOT lanes on I-35 W (MnPass). Also added priced dynamic shoulder lane.	Demand-based pricing of city-owned on-street and off-street parking spaces in 7 pilot areas (SF*park*).	Variable tolling on SR 520 Bridge.
Increased HOV requirement from 2+ to 3+.	Maintained existing HOV requirements, 2+ on I-110 and 3+ during peak periods on I-10.	Increased vehicle occupancy requirement from 2+ to 3+.	Maintained 2+ carpool occupancy requirement.		Registered vanpools and buses ride for free.
Carpools required to register and use toll tag.	Carpools required to register and use switchable transponder.	3+ carpools required to register and display special decal but no transponder.	Carpools do not need to register or use transponder.		Registered vanpools and buses required to use transponder.
Transit					
12 new commuter buses in the corridor.	59 new clean-fuel buses to 4 service providers.	23 new buses	27 new buses.		44 new buses.
3 new Xpress bus routes in the corridor.	More frequent bus rapid transit service and municipal feeder service.	3 new transit routes.			Addition of 90 one-way peak period bus trips.
4 new or expanded park-and-ride lots.	Expanded parking capacity at transit stations.	500 park-and-ride spaces added.	6 new or expanded park-and-ride lots.		Enhancements to two park-and-ride lots.
	Transit signal priority on two streets in downtown LA.	Bus rapid transit in HOT lanes.	Bus bypass lane at Highway 77/Highway 62 interchange.		
	Other enhancements, e.g., new transit operating and maintenance facility.		Double contraflow lanes in downtown Minneapolis (MARQ2).		
Telecommuting /TDM					
Outreach to encourage formation of 3-person carpools.	Promotion to increase registered vanpools and employer-based ridesharing.	Outreach to encourage 3-person carpool formation as well as other existing employer-based programs for ridesharing, telecommuting and flex-time.	eWorkPlace telework program.	Use of existing alternate commute outreach to distribute brochures on SF*park* and 511 parking information.	Continued programs already in use by agencies and other employers that aim to reduce trips in the region.
Technology					
Automated toll enforcement systems.	Dissemination of parking information on-line and on variable message signs.	Introduction of ramp metering at 22 locations.	Active traffic management.	Real-time information on parking price and availability disseminated by websites, telephone, and mobile apps.	Active traffic management system on SR 520 and I-90 corridors.
	Transit signal priority on selected streets in downtown Los Angeles.	Transit signal priority on selected routes leading to I-95.	Real-time transit and traffic signs.		New travel time signs near key interchanges for SR 520.
			Driver assist system for shoulder running buses.		

Table 1.3: UPA project effectiveness [Zimmerman et al., 2015]

Measure	Pre-Deployment	Post-Deployment	Change
Atlanta I-85, A.M. Peak Period, Southbound			
General Purpose Freeway Lanes			
Mean Travel Time	16.1 min	16.9 min	+0.80 min
HOV/HOT Lanes			
Mean Travel Time	14.1 min	13.8 min	-0.30 min
Los Angeles I-110, A.M. Peak Period, Northbound			
General Purpose Freeway Lanes			
Mean Travel Time	27.09 min	27.07 min	-0.02 min
HOV/HOT Lanes			
Mean Travel Time	12.40 min	14.29 min	+1.90 min
Los Angeles I-10, A.M. Peak Period, Westbound			
General Purpose Freeway Lanes			
Mean Travel Time	30.88 min	33.90 min	+3.02 min
HOV/HOT Lanes			
Mean Travel Time	15.96 min	15.08 min	-0.85 min
Miami I-95, A.M. Peak Period, Southbound			
General Purpose Freeway Lanes			
Mean Travel Speed	15 mph	51 mph	+36 mph
HOV/HOT Lanes			
Mean Travel Speed	20 mph	64 mph	+44 mph
Minneapolis I-35W, A.M. Peak Period, Northbound			
General Purpose Freeway Lanes			
Mean Travel Time	18.9 min	16.8 min	-2.10 min
HOV/HOT Lanes (Section south of I-494 only)			
Mean Travel Time	6.4 min	6.0 min	-0.4 min

1.5 Bringing the Public Along for the Ride

Given the degree to which the effectiveness of these projects can be brought into question (see Section 1.4) it is then important to analyze how they ever become implemented. It seems that the consensus in the political science literature is that they are implemented against a consensus of local citizens.

Taking a specific example, Hysing [2015] investigates the Gothenburg congestion tax. Hysing argues it as an implementation of representative government, in that politicians determined the best policy with minimal citizen involvement. A key concern with this style of democracy, however, is the continual erosion of trust in government, either the legislators or the institutions. From the standpoint of enacting unpopular policies that are later deemed beneficial, this policy proved remarkably inexpensive politically, in that politicians associated with it were subsequently re-elected. As in the case of SR 167 above (Figure 1.3) post-implementation resistance decreased over time. Yet there is theoretically a tipping point where citizens will resist 'beneficial' policies being foisted upon them by well-meaning administrators. Use of this recipe is likely a one-shot deal, at least on sufficiently short time horizons.

A key ingredient to the Gothenberg implementation was tying a new infrastructure investment package to the tax, making the cost-benefit analysis explicit for the local citizenry. This effect was potentially magnified by the perception that the city had been left out of previous rounds of infrastructure updates. This mirrors the situation around the toll introduction on SR520 in Seattle, where bonds necessary for financing the bridge reconstruction are secured through the tolls collected. Furthermore, there was political consensus among the majority parties that road pricing was the appropriate solution. The agreement was evidently strong enough amongst the main parties that a new single-issue party opposing the congestion tax was formed, but did not fare well in elections.

Noordegraaf et al. [2014] perform a meta-analysis of six road-pricing projects, systematically categorizing the factors involved in implementation. A striking result is that of the 106 papers analyzed and the 61 implementation factors identified, only six factors were present in all six cases. This is a surprisingly small common coverage considering all cases revolved around similar transportation policies. Furthermore, these six common factors (political support, public support, information campaign, actor perceptions, transport system characteristics and marketing) only account for an average of 27% of all implementation factors, suggesting there is a large amount of heterogeneity between similar policies in different regions (all politics are local after all). This sentiment agrees well with the expansion of road pricing schemes in the US (see Figure 1.2) where many new facilities are undertaken by areas with existing road-pricing projects. These results stand in opposition to other results in the literature that seek to build heuristics around the implementation of road-pricing policies [Sørensen et al., 2014, Isaksson et al., 2017, Ramjerdi and Fearnley, 2014]. I find it likely that Noordegraaf et al. [2014] has the correct formulation, in that the variability of constituencies, experience, political moment and infrastructure requirements will prevent truly universal rules of implementation from arising.

1.6 Conclusion

A recent special report in the Economist magazine[10] discusses the future disruption to transportation coming from the introduction of self-driving vehicles. Ease of access to point-to-point transportation may dramatically increase the use of our existing infrastructure as commuters switch away from public transit to private car sharing/autonomous vehicles. The US, as a society, might finally be running out of roadway to address existing infrastructure shortfalls with existing 'dumb' transport architecture. At the same time, however, there will be more data available to optimize transport systems than ever before. And cars whose presence on specific roadways at specific hours is more easily tracked open themselves to a

---

[10] https://www.economist.com/na/printedition/2018-03-03 accessed on 2018-03-04.

highly fine-grained pricing policy.

The question is then more pertinent than ever before regarding the efficacy and implementation of road-pricing schemes. The city of NY was recently considering bypassing the state legislature in the introduction of a cordon toll.[11] This was based on a claim that the city of NY itself has the authority to create the toll zone, rather than relying on the state legislature where previous efforts have failed. And as discussed above, the proposal is directly tying revenues raised from the congestion tax to local infrastructure improvements.

---

[11] https://www.wsj.com/articles/proposal-for-congestion-charge-on-new-york-city-motorists-1496581200 accessed on 2018-03-04.

Chapter 2

DYNAMIC CONGESTION PRICING

2.1 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.

---

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

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 transportation infrastructure closer towards pricing congestion externalities, a perennial favorite of the economics field [Agnew, 1977].

HOT lanes generate revenue for local and state agencies, a particularly valuable feature in the context of aging infrastructure and diminished revenue from declining (in real terms since 1993) gas taxes. 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

---

[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.

[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 2014-01-13 - http://www.bea.gov/iTable/index.cfm

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., 2011a, Janson and Levinson, 2014]. In addition to identifying a more plausible demand elasticity we show that failing 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., 2011a, Janson and Levinson, 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 itself 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. Ignoring this relationship generates the positive demand response for Liu et al. [2011a]. We exploit a feature of the tolling algorithm where downstream traffic affects prices upstream, generating variation in prices that is uncorrelated with traffic conditions at the location where the driver pays the toll. 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., 2011a, 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

---

[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.

reliability and money. This leads to problems in simple methods for estimating VOT on HOT lanes, assigning 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 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 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 2.3 describes the project setting. Section 2.4 describes the econometric methodology and Section 2.5 reports the data used in the analysis. The results are presented in Section 2.6, and Section 2.7 offers concluding remarks and discussion. Details on the econometric specification tests, as well as additional tables and figures are available in Appendix A.

2.2 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], and 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.

The theoretical properties of HOT lanes are similar to second best pricing with a tolled route and untolled alternative. Verhoef et al. [1996] show that the efficiency of tolling one route depends on relative cost parameters, demand elasticity, and the cost of implementing the toll. Arnott et al. [1992] and Arnott et al. [1994] examine route choice and tolling with heterogeneous users. This research highlights how efficiency of a toll depends on the spatial and temporal integration of the routes, and that tolls without rebates benefit commuters with high schedule delay costs and hurt those with low schedule delay costs. Extending to a dynamic model van den Berg and Verhoef [2011] shows the distributional costs of tolling are not monotonic with respect to schedule delay costs, and that first and second best tolling generate similar distributional effects. 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.[7] 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

---

[7] While the overall framework is similar we cannot capture all the same features of alternative models that incorporate schedule delay costs. Therefore, one limitation in our study is not being able to precisely capture schedule delay costs.

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. [2011a] 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. 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.

2.3 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 2.1). 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.[8] SR167 is likely a smaller road handling fewer cars than other HOT lanes.

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 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 2.2. 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

---

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

Figure 2.1: SR167 HOT Lanes Map, [WSDOT, 2013]

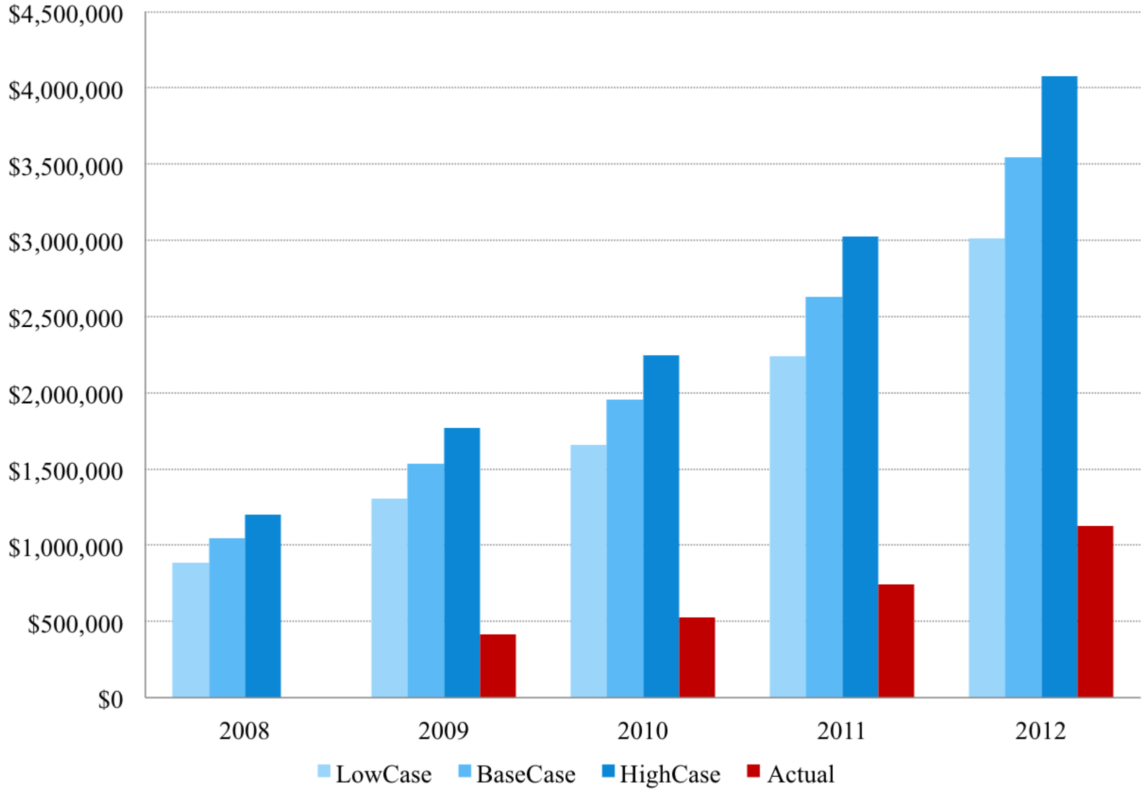
introduction of a tolled alternative lane.[9]

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

---

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

Figure 2.2: SR167 Revenue: Forecasts vs Actual, [Wilbur Smith Associates, 2006]



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.[10] 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. [2011b] find that the behavior is very similar between these two user classes. We therefore are not concerned about the skewness of the frequency data.

2.4 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

---

[10] 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

to autocorrelation in the variables of interest (see Figure 2.3). 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:

$$y_{it} = \beta p(y_{It-1})_{it} + \theta X_{it} + \epsilon_{it}. \tag{2.1}$$

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 2.2 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 $E[p(\epsilon_{it-1})\epsilon_{it}] \neq 0$.

$$y_{it} = \beta p\left(\beta p(y_{It-1})_{it} + \theta X_{it-1} + \epsilon_{it-1}\right)_{it} + \theta X_{it} + \epsilon_{it} \tag{2.2}$$

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 2.3). 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 2.3.[11] 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.

---

[11] 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.4 in the Appendix.

$$\Delta y_{it} = \beta \Delta p(y_{It-1})_{it} + \theta \Delta X_{it} + \Delta \epsilon_{it} \tag{2.3}$$

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 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 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.

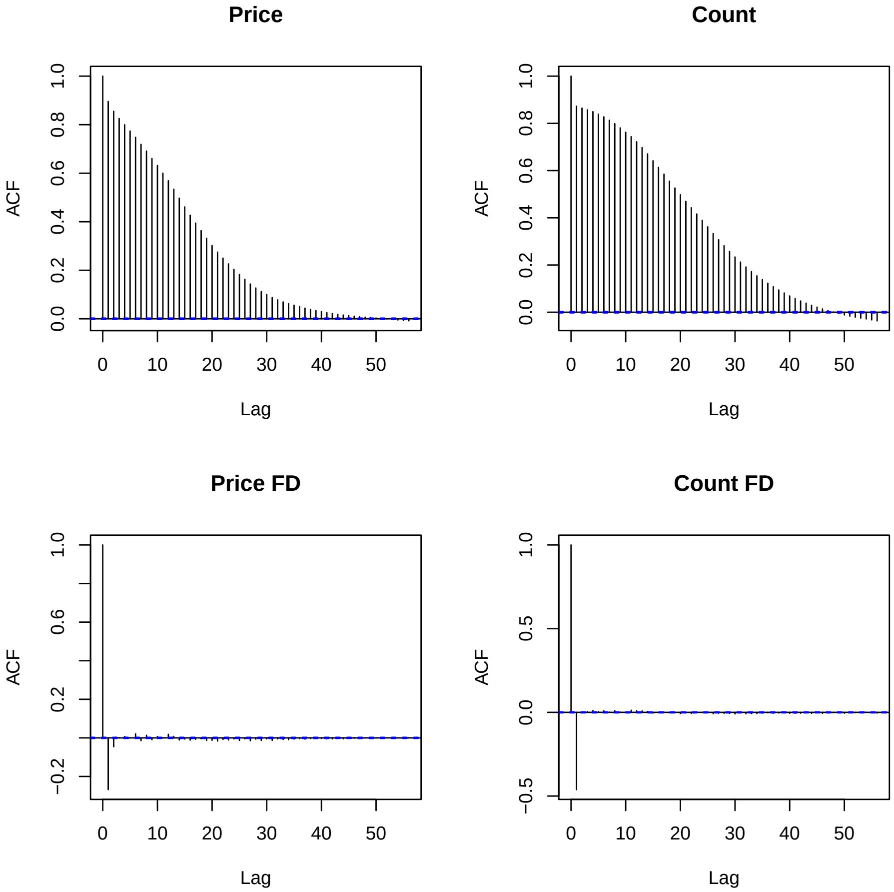
2.4.1 Empirical Specification

Our specific estimating equation takes the form:

$$y_{it} = \beta p_{it} + \gamma GP_{it} + \phi E[TTsave|t] + \lambda R_t + c_i + h_t + d_t + u_{it}, \tag{2.4}$$

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 2.4 is similar to our equation 2.1 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 2.5.3. Unlike Liu et al. [2011a], who include realized travel time, we model this as an

Figure 2.3: Price and Count ACF



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 2.4 gives

$$\Delta y_{it} = \beta \Delta p_{it} + \gamma \Delta GP_{it} + \phi \Delta E[TTsave|t] + \lambda \Delta R_t + \Delta u_{it}, \tag{2.5}$$

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 $\Delta p_{i+1,t}$).

2.5 Data

Our primary dataset consists of information collected by highway loop detectors and automated tolling systems.[12] 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.[13] 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 the route for every valid five minute interval during our sample period. Speed is computed from volume data based on Athol [1965]'s formula.

---

[12] 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.

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

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

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.

2.5.1 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 in the Appendix). 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.

2.5.2 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.[14]

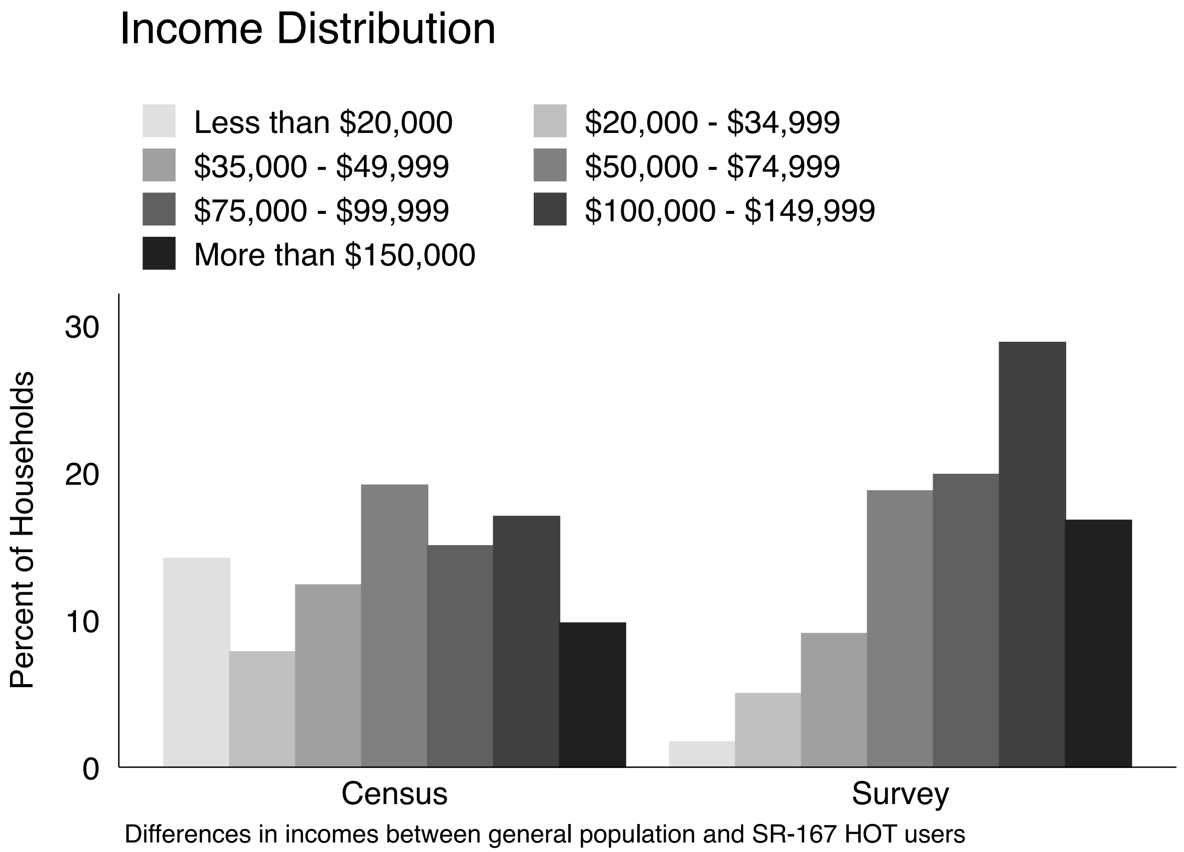
To approximate the SR167 income distribution we use a weighted average of zip code level data from the 2010 US Census, where the weights are the proportion of all HOT users that came from a specific ZIP code. This method places more weight on zip codes that use the SR167 HOT lane. We assume that the 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 2.4 presents the income from the survey of HOT users compared to our

---

[14] 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.

estimate 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 2.4: Differences in Income



Note: 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.

2.5.3 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 for both the northbound (Figure 2.5) and southbound (Figure 2.5) 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. 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 2.1. 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.[15] 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

---

[15] 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.

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 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.

Table 2.1: 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

Figure 2.5: Travel Time and Reliability

Northbound

Southbound

Note: The thick line is 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.

2.6 Results

Coefficient estimates for both the OLS, FD, and FD-IV HOT counts are presented in Table 2.2. 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.[16] 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.[17]

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., 2011a, 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

---

[16] 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.

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

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 2.6.2. While we do not have a 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.[18]

2.6.1 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.[19] 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 2.3 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

---

[18] Despite these arguments, we acknowledge the limitation that we do not have fully random variation in travel times and reliability.

[19] The peak period can be seen on the travel time graphs in Figure 2.5.

Table 2.2: 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

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.[20] Figure 2.6 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 2.3: Heterogeneity by Gates, Congestion, and Direction

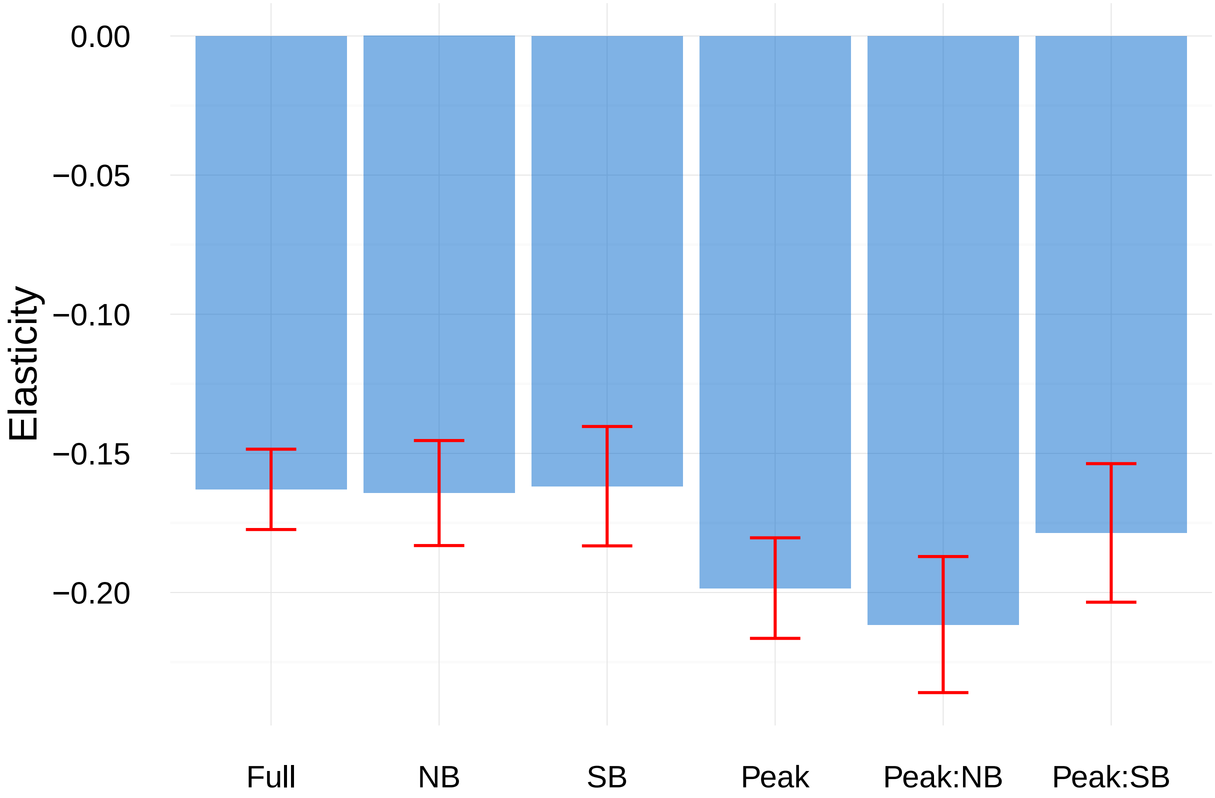
	(1) NB	(2) SB	(3) Peak	(4) Off-Peak	(5) Peak:NB	(6) Peak:SB
Price	-0.584^{***}	-0.575^{***}	-0.705^{***}	-0.178^{***}	-0.751^{***}	-0.634^{***}
	(0.0342)	(0.0389)	(0.0327)	(0.0272)	(0.0443)	(0.0451)
GP Speed	-0.0131^{***}	-0.0119^{***}	-0.0169^{***}	-0.0104^{***}	-0.0175^{***}	-0.0138^{***}
	(0.000838)	(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 2.2. 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

---

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

Figure 2.6: Price Elasticities by Model



Note: 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 2.3.

2.6.2 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 2.8. 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.[21] 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 2.8. Approximately 0.7% of drivers experience a negative simple VOT where the travel time in the GP lane was faster than the HOT lane.

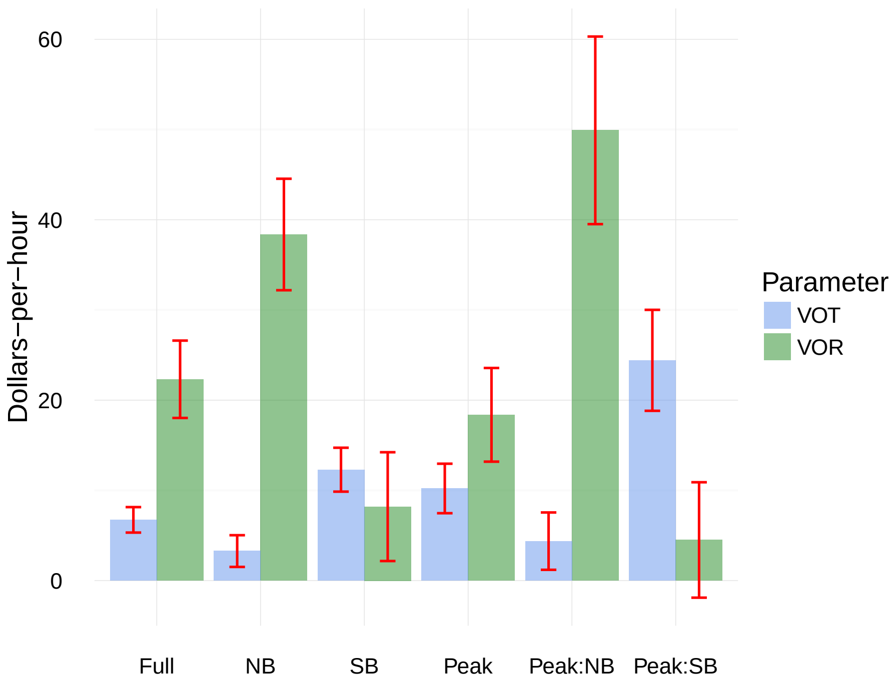
Columns (1) and (2) (as well as (5) and (6)) of Table 2.3 show that drivers have a relatively similar price responsiveness in both directions, but NB driver primarily value reliability and SB drivers primarily value time savings. Figure 2.7 presents the estimates of VOT and VOR defined as the ratio of 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 2.7 are based on the regression models presented in Table 2.3, as well as the base model from column (4) of Table 2.2. Since VOT and VOR are nonlinear combinations of parameters Figure 2.7 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

---

[21] 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 2.7: Value of Time and Reliability



Note: 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 2.3. The confidence intervals are created using the delta method.

that although they are both measured in hours these values are based on different variables in the regression. However, based on the summary statistics provided in Table 2.1 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.[22] 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 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.

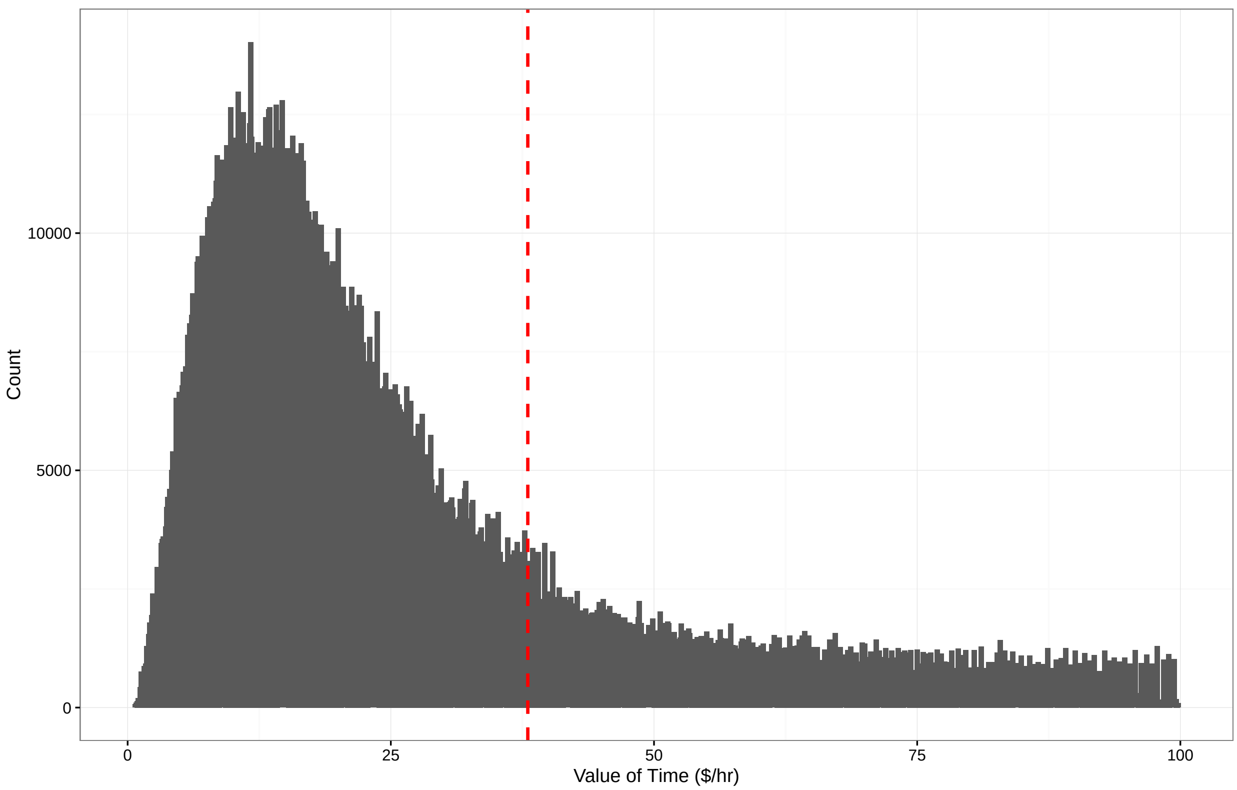
The heterogeneity also suggests that transportation managers can optimize the toll algorithms for HOT lanes based on simple observable differences in usage behavior. Drivers that value reliability may not be as sensitive to the toll rate and will purchase HOT access at a wide range of prices. Conversely 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

---

[22] 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.

savings and reliability are correlated a more detailed analysis is required to investigate the relationship between VOT, VOR, and price elasticity.

Figure 2.8: Simple Value of time



Note: 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.

2.6.3 Robustness

We also perform several robustness checks presented in Table 2.4. All regressions presented in Table 2.4 have the same basic from as the our preferred regression (Table 2.2 column (4)), which is shown in column (1) of Table 2.4 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.

2.6.4 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.[23] 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 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

---

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

Table 2.4: 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^{***}
							(0.0061)
GP Speed	-0.0129^{***}	-0.0129^{***}	-0.0143^{***}	-0.0123^{***}	-0.0123^{***}	-0.0141^{***}	-0.00373^{***}
	(0.0007)	(0.0007)	(0.0009)	(0.0008)	(0.0008)	(0.0012)	(0.0002)
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^{***}
	(0.0065)	(0.0065)	(0.0083)	(0.0070)	(0.0071)	(0.0119)	(0.0016)
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 2.2, 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

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.

2.7 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.[24]

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 is to be able to deter SOV drivers from entering HOT lanes the prices. Inelastic demand

---

[24] 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.

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.

Chapter 3

TIME OF DAY CONGESTION PRICING

3.1 Introduction

Congestion is an expensive problem [Schrank et al., 2012], resulting from the free access to publicly constructed and maintained roadways [Vickrey, 1963]. There is a long history of interest in the 'optimal toll' to offset the externality of the marginal driver [Baumol and Oates, 1988, Boardman and Lave, 1977, Anas and Rhee, 2006], but only recently have traffic engineers and planners moved to implement congestion tolling. There have been few empirical studies [Brownstone et al., 2003, Liu et al., 2011a, Janson and Levinson, 2014, Brent and Gross, 2017] testing the behavioral response of drivers to congestion pricing in an economic context, despite the existence of prediction models [McFadden, 1974].

In this paper we estimate the impact on traffic volume of imposing a time-of-day toll on one of two bridges crossing Lake Washington in the Seattle, WA metropolitan area. Since these bridges share a catchment of potential commuters we expect three general responses:

1. shift from the tolled route to the free route,
2. change of time on the tolled route to less expensive periods, or
3. change of mode from single-occupancy to public transport or carpool.

We assemble a dataset consisting of traffic volumes and travel times on the two bridges that cross Lake Washington. We leave modal shifts to future work, acknowledging that it will affect the estimates of price-sensitivity that we construct. Changes in commute time in response to time-of-day tolling is referred to as 'spreading', which is an attempt by drivers to avoid the most expensive tolls. Using a Difference-in-Differences (DD) approach we estimate the impact of introducing a toll, both on the magnitude of spreading and also on route choice. We also also include robustness checks on our DD estimates using Regression Discontinuity (RD) techniques.

Our results indicate that, at the mean, the introduction of time-of-day pricing has decreased volume on SR520 by approximately 360 cars per hour during the tolled period. We find a corresponding increase during the tolled hours on I90 of 168 cars per hour. We also find evidence of spreading around toll changes.

3.1.1 Literature

Vickrey [1963] puts forth compelling arguments for road pricing as a congestion management tool, noting copious examples where peak-demand pricing was already in place. He further argues that un-priced access creates an oversupply of roads and that sensible pricing will decrease demand and free valuable land for more productive purposes. The provision of public transport can alleviate this overburdening of the network but itself requires understanding of price sensitivity to achieve optimal flows.

McFadden [1974] did further work on estimating aggregate travel demand model from a sample of commuters in the Bay Area. The investigation focuses on mode choice and constructs estimates from a stated preference survey. He estimates elasticities of demand that are then used in forecasting urban travel demand resulting from a policy intervention. This remains a popular method [Wilbur Smith Associates, 2011] despite the growing amount of revealed preference data.

Though congestion pricing is only recently becoming more prevalent, empirical and theoretical work by as early as the 1970s advocates for dynamic tolling. Boardman and Lave [1977] calculate the optimal toll using observations of the speed-flow relationship from a two lane highway. Their optimal toll attempts to account for the marginal damage of an additional driver to overall welfare by adding themselves to the highway. Considering that the speed-flow relationship is dependent on road conditions, they argue that the optimal toll should be dynamically adjusted.

More recently, Parry [2002] conducts a theoretical analysis of congestion tax alternatives using simple models finding that 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'. This research highlights the connection between time-of-day pricing and fully dynamic pricing. While dynamic pricing may be optimal from a theoretical perspective, in practice time-of-day pricing may approach the welfare gains of dynamic pricing, with the added benefit of easier implementation. Additionally, time-of-day pricing reduces driver uncertainty with respect to the price of travel, which may affect their investments in alternative transport modes.

While tolls raise questions of social equity due to their regressive nature [Plotnik et al., 2010], there could be a net social gain from decreased congestion, increased revenue [Baumol and Oates, 1988], and reduced pollution from idling cars [Small, 1983]. Despite the immense benefits from using tolling to shift drivers away from the peak congestion period there is relatively scant empirical research on traffic spreading. An exception is recent work by Gibson and Carnovale [2015] that shows the removal of a constant daytime toll to enter the city center in Milan altered the timing and spatial distribution of driver behavior through attempts to avoid the toll. Additionally, they find large impacts on pollution reduction attributed to the toll. Our setting differs since the magnitude of the Milan toll did not vary throughout the day: the toll is constant from 7:30AM-7:30PM and zero all other times. And there were no free alternatives to enter the city.

Previous unpublished work by Foreman [2012] found evidence that time-of-day pricing induces traffic spreading manifested by decreases in trips during the peak demand period.

Identification was based upon toll increases of different magnitude on multiple tolled bridges in the Bay Area. This research setting differs in that we observe a change from free access to tolled, while also observing an alternate route that remains free to use. Thus we expect the magnitude of the responses in our study to be larger than Foreman's, since the introduction of the toll alone should force drivers with low value-of-time to seek other routes. Similar to other goods the elasticity of demand depends on the availability of substitutes in the market. The countervailing effect of increased travel time on the free alternative, as more drivers switch routes, would be expected to dampen the welfare loss of toll payers since drivers with low value of time can select the free route. However, the increased congestion on the free road reduces welfare gains from reducing congestion on the tolled road.

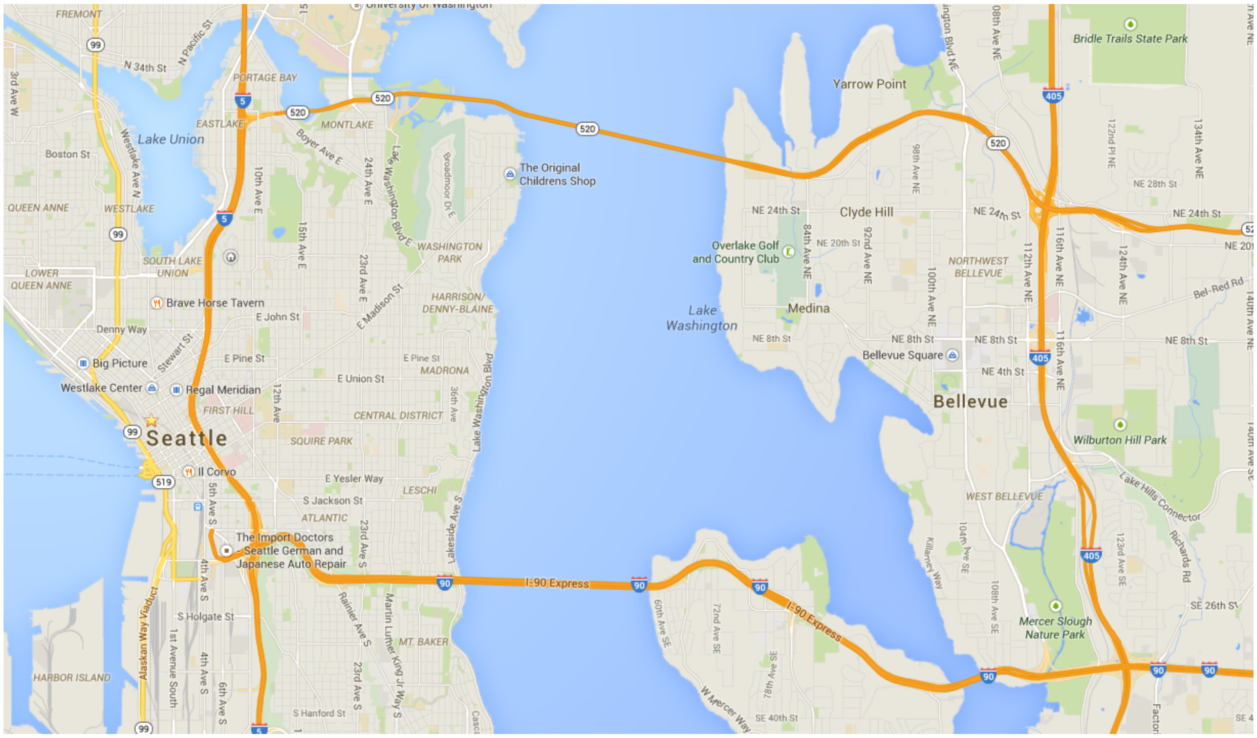
The remainder of the paper is organized as follows. Section 3.2 introduces the project setting and relevant data collected. Section 3.3 discusses the econometric methods and identification strategy. Section 3.4 presents and discusses the results. Section 3.5 offers concluding remarks. Details on econometric specification tests, as well as additional tables and figures are available in Appendix B.

3.2 Data

We examine driving behavior on two bridges that serve as major commuting routes in the Seattle, WA metropolitan area. Both bridges cross Lake Washington and connect Seattle to Bellevue and the eastern suburbs (see Figure 3.1). Most of the traffic is composed of commuters who live in the eastern suburbs and work in Seattle, though there is a substantial population that live in Seattle and work on the east side at companies such as Microsoft in Redmond. The toll was introduced on SR520 in December 2011 while I90 was left free to use. These two routes are the primary connector routes from the eastern suburbs to the central business district, which is designated by the star on Figure 3.1.

We obtained volume, occupancy and speed data for SR520 and I90 from the UW Traffic

Figure 3.1: 520 and 90 Crossing Lake Washington, map data:Google

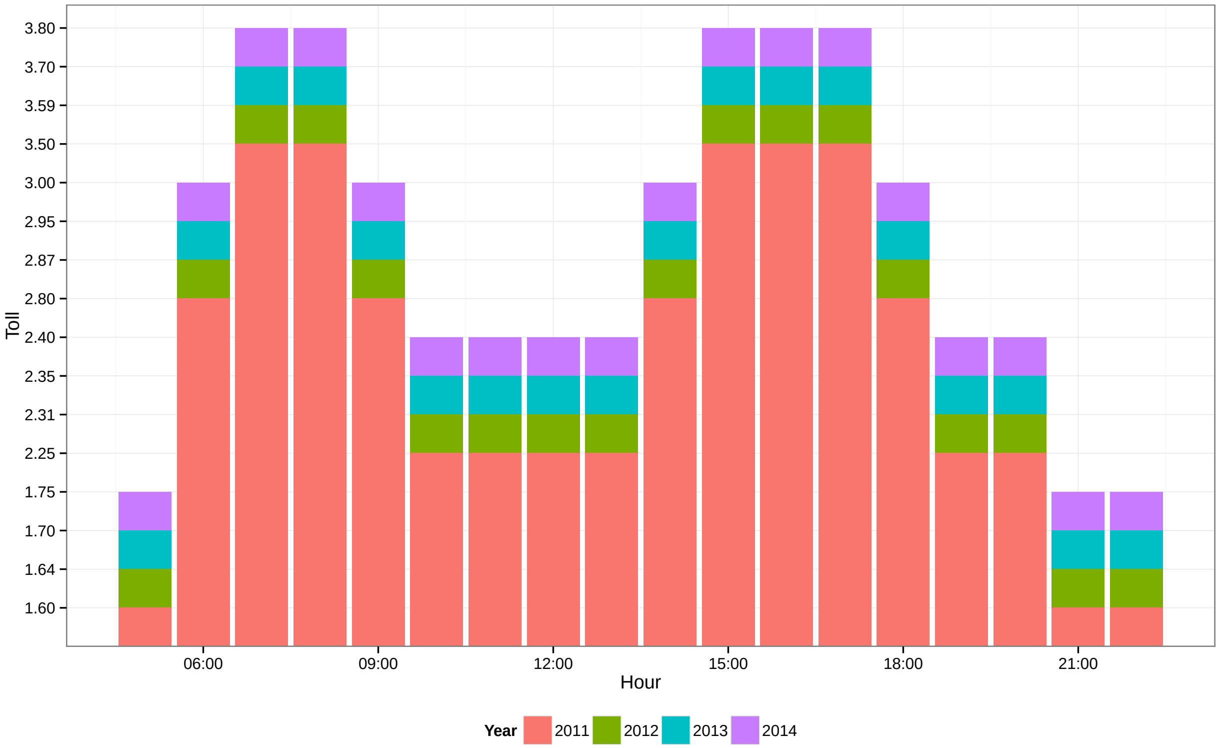


Analysis Center (TRAC) covering 2010 through July 2014. The data is collected by traffic loop detectors operated by the Washington State Department of Transportation (WSDOT) and reports information in 5 minute intervals. The specific loops used in the analysis are at milepost 1.58 on SR520 and 4.30 on I90. Each set of loops is located at the western extent of its respective bridge, before westbound traffic can exit and after eastbound traffic has passed its last exit. The presence of these loop detectors enables us to exclusively capture all the traffic crossing the bridge.

Tolls were introduced on SR520 on 29 December 2011 and have changed on 1 July of each subsequent year. A toll of varying price is required to cross the bridge in either direction from 5am until 11pm. Figure 3.2 shows the rate changes over time. The rates increase 2.5%

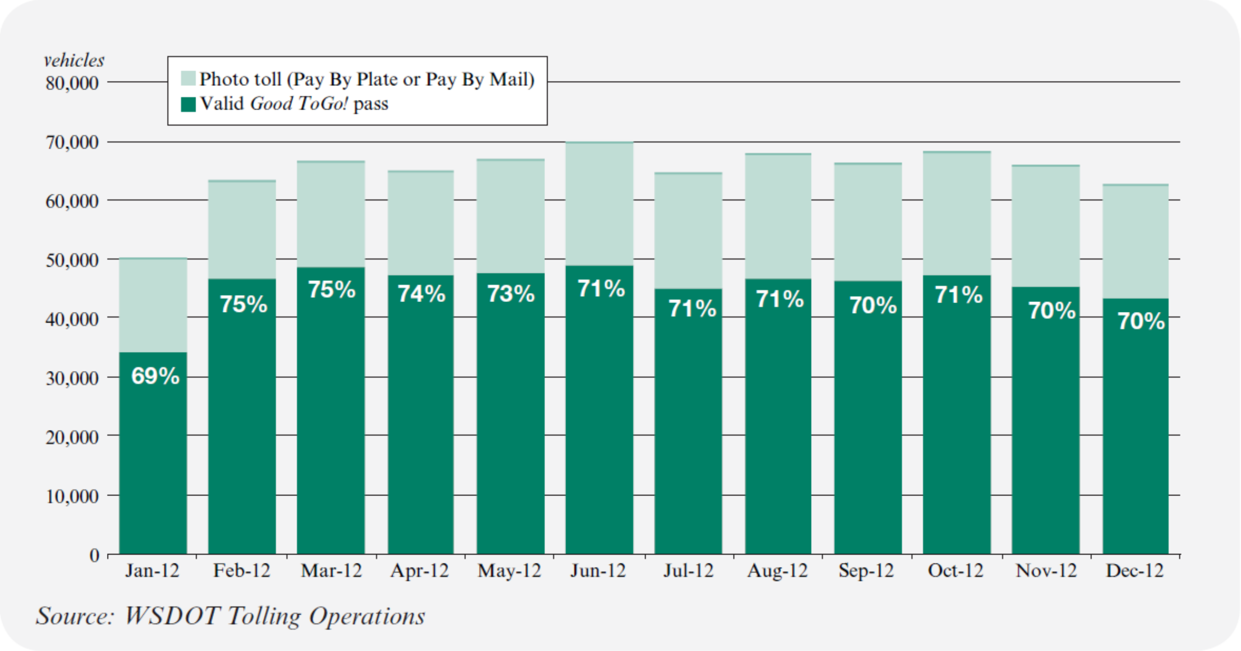
every year until 2015 when they will rise a final 15% [Stone, 2012], subject to legislature approval.

Figure 3.2: Toll Rates



Everyone who crosses SR520 pays a toll, but the rate depends on the technology used to collect the toll. Drivers who have purchased a transponder, named a Good-to-Go (GTG) pass by WSDOT, and attached it to their vehicle pay a reduced fare compared to those who cross the bridge without the transponder. Drivers without transponders have their license plates photographed and bills sent by mail. They pay $1.60 more than drivers with a transponder. This creates issues of endogeneity since drivers can choose which rate they pay. Since drivers without the GTG pass pay a higher toll, and may not be equally distributed across commute times, our results may be biased. However, as shown in Figure 3.3, a majority of users use a GTG pass when crossing SR520.[1]

Figure 3.3: Share of Transponder vs Non-transponder SR520 Users



Source: WSDOT Tolling Operations

Figure 3.4 shows the monthly mean of hourly traffic volume in each direction for both bridges. Please notice the lack of data on SR520 for the month of January 2012. The preferred loop detector (having the longest continuous time series) was not functioning during this period. We are not overly concerned, however, since it is the month immediately following the introduction of the toll and we expect a disequilibrium to prevail. The missing data likely biases a coefficient on treatment as drivers explore alternate routes/modes.[2]

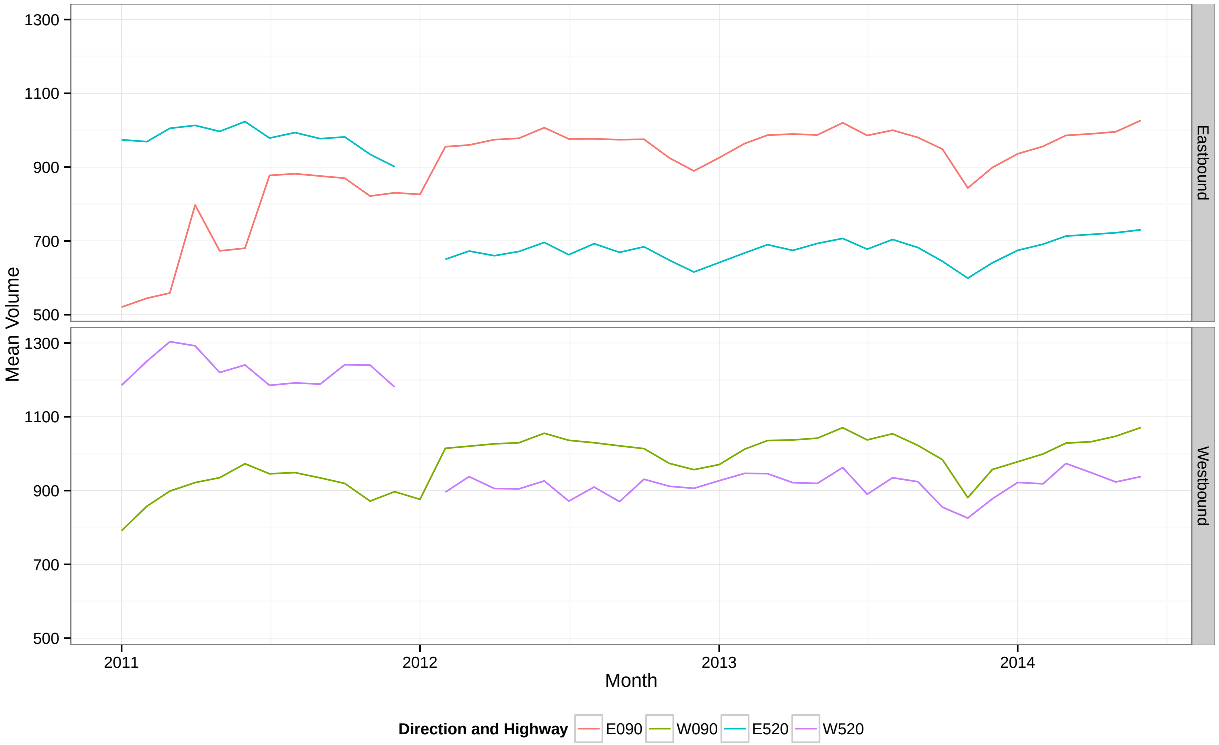
---

[1] To test and correct for this bias we have requested tagged tolling data from WSDOT and are awaiting delivery of the data.

[2] There are other loop detectors with data available from that time and we are currently investigating the feasibility of 'filling' in the missing data.

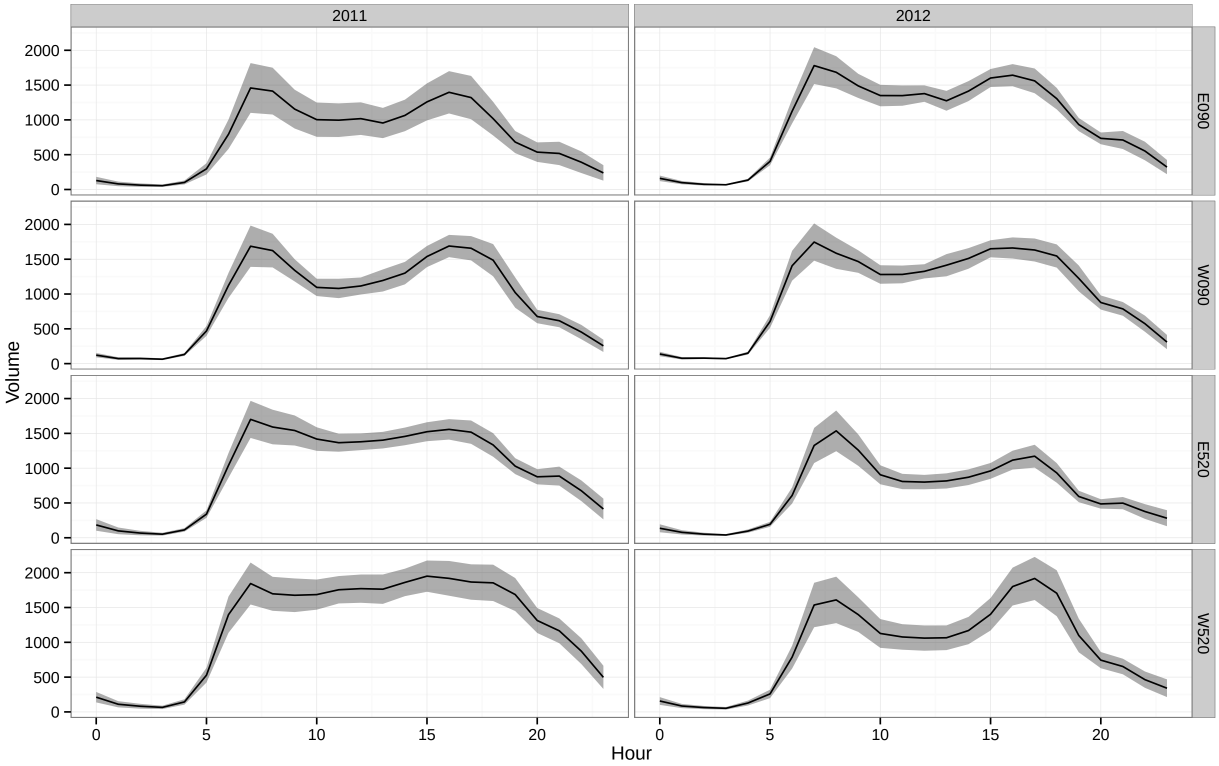
The average monthly volume time trends for both bridges in each direction appear to have similar trajectories pre- and post-toll. The eastbound means, however, do not share a common time trend; there is a sharp increase in traffic on I90 during 2011. The implications for different pretreatment trajectories will be discussed in the methodology section (3.3). The impacts in of the toll are clearly observable in the aggregate data. After the imposition of the toll there is a large drop in traffic on SR520 in both directions, and I90 experiences a corresponding, but smaller, increase in average traffic volumes. These aggregate effects are consistent with our empirical results.

Figure 3.4: Monthly Mean of Bridge Volume



Moving from the aggregated volumes to the distribution of traffic throughout the day we display the mean and standard deviation of volume in every five minute interval pre- and post-toll in Figure 3.5. In the pre-toll period (2011) SR520 exhibits less daily variation than I90, maintaining high volumes between the commute peaks in both directions. The profile of volume on SR520 throughout the day is markedly different post-toll (2012), exhibiting distinct peaks around the morning and evening commute. The eastbound evening commute on SR520 also appears to have been reduced. I90 exhibits corresponding increases in midday volume between 2011 and 2012. It also appears that the standard deviation of volume decreases between the two years.

Figure 3.5: Mean and Standard Deviation of Volume



3.3 Methodology

We can exploit two identification strategies in estimating the causal effect of tolling on SR520: Difference-in-Differences (DD) and Regression Discontinuity (RD). The introduction of tolling on 29 December 2011 brought with it variation at two levels: tolled and untolled periods within each day as well as rate variation within the tolled period. We exploit observation of pre- and post-treatment outcomes on the tolled and untolled periods for the DD estimation. The between hour variation in toll rate is exploited for RD estimation.

Long a staple of policy analysis, DD estimation exploits divisions in populations between treated and untreated. We include dummy variables to capture both time-specific and treatment-specific effects, and their interaction captures the difference between the treatment group and the control group after treatment is administered. In our setting we compare the untolled hours to the tolled hours both before and after tolling was imposed.

Regression discontinuity has grown in popularity in recent years and has many attractive properties for causal inference [Calonico et al., 2014a]. This approach is appropriate when treatment status is assigned according to the value of an observable covariate at a known cutoff. The general idea is to examine outcomes on either side of the cutoff (the difference between treated and untreated), avoiding bias from unobserved confounding variables. A graph of the outcome variable on either side of the cutoff can provide visual evidence of the importance of the cutoff. Unfortunately for our analysis the discrete nature of the observations (five minute intervals) violates the RD condition requiring arbitrarily close observations to the cutoff [Calonico et al., 2014a]. We therefore follow Lee and Card [2008] and model the difference just above and just below the cutoff using a parametric formulation with random specification errors.

We begin with the DD strategy and then present the RD methodology.

3.3.1 Difference-in-Difference

There are two key identifying assumptions in DD estimation [Cameron and Trivedi, 2005]:

1. time effects are common across treated and untreated

2. and the composition of repeated cross sections is stable before and after treatment.

Since we expect there to be route switching between bridges as a result of tolling SR520 it would be inappropriate to hold I90 as the untreated group. Furthermore, we want to explicitly test for route shifting by analyzing traffic volume on I90 during SR520's tolled periods. Instead we hold the untolled hours of the day as the untreated group for each bridge and estimate DD equations separately. To graphically make the case we present the monthly means of volume by commute conditions for SR520 in Figure 3.6. Despite the seasonal regularity of the monthly averages the overall trends across the hours appear stable both pre- and post-treatment. We therefore conclude that, except in the case of eastbound I90 (see Figure 3.4) there is no reason to doubt the common trends assumption. To test the effect of eastbound I90 might have we run both a specification where east and westbound directions are pooled together and another where they are separately estimated. Results are presented in Section 3.4.

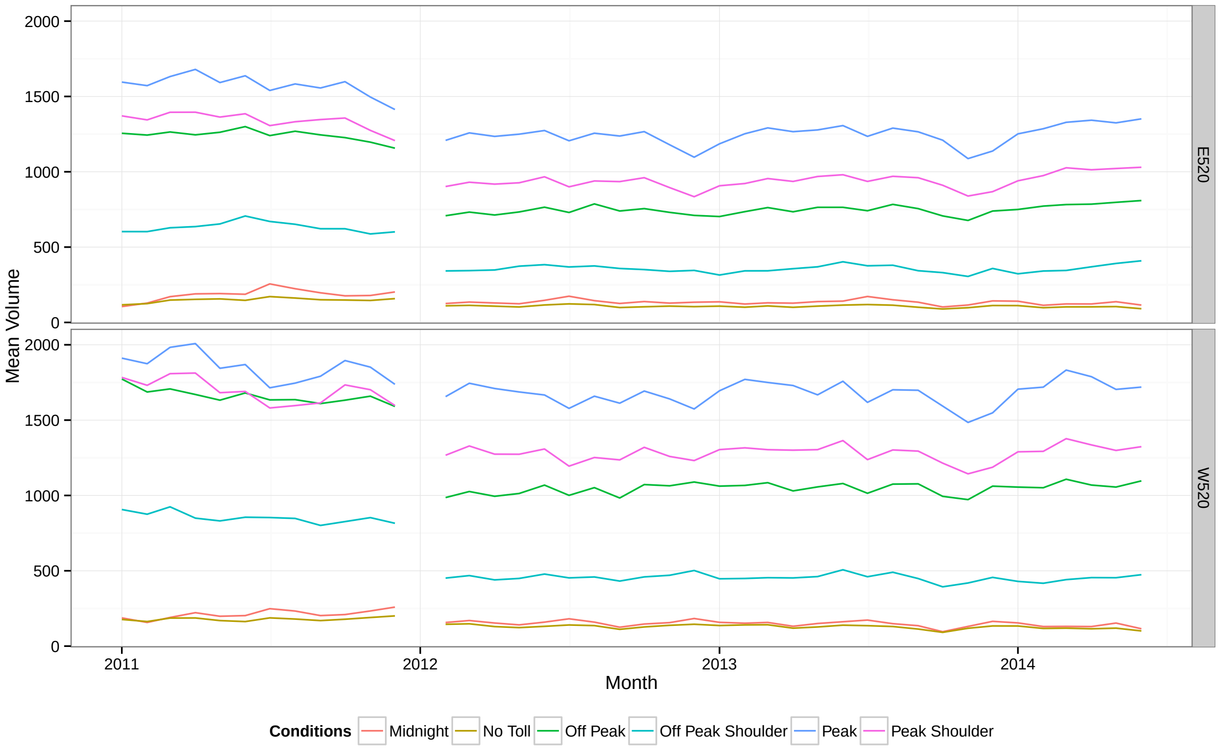
Regarding the composition of the cross section pre- and post-treatment, we assume that the housing and workplace decisions of most commuters are sticky and not easily adjusted. Considering the size of the service area on either size of the bridges it is impossible for us to accurately characterize the composition of the cross sections.

Equation (3.1) is the DD specification,

$$y_t = \alpha_0 + \alpha_1 treat + \beta_0 post + \beta_1 post * treat + day + week + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2) \tag{3.1}$$

where

Figure 3.6: Monthly Mean Volume by Traffic Conditions



$y_t =$ Volume of vehicles in five minute intervals

$\alpha_0 =$ Intercept

$treat =$ Indicator for tolled periods

$post =$ Indicator for after imposition of toll

$day =$ Day of week fixed effects

$week =$ Week of year fixed effects

The coefficient of interest is β_1, capturing the effect of tolling on either bridge respectively with the untolled period held as the reference group.

Given that the toll itself is different throughout the day we can also employ a DD where the treatment indicator is a multi-leveled factor variable for each five-minute interval, leading to the equation

$$y_t = \alpha_0 + \alpha_1 5min + \beta_0 post + \beta_1 post * 5min + FE + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2) \tag{3.2}$$

where the categorical variable $5min$ represents each five minute interval of the day. Midnight is held as the control.

3.3.2 Regression Discontinuity - Discrete Forcing Variable

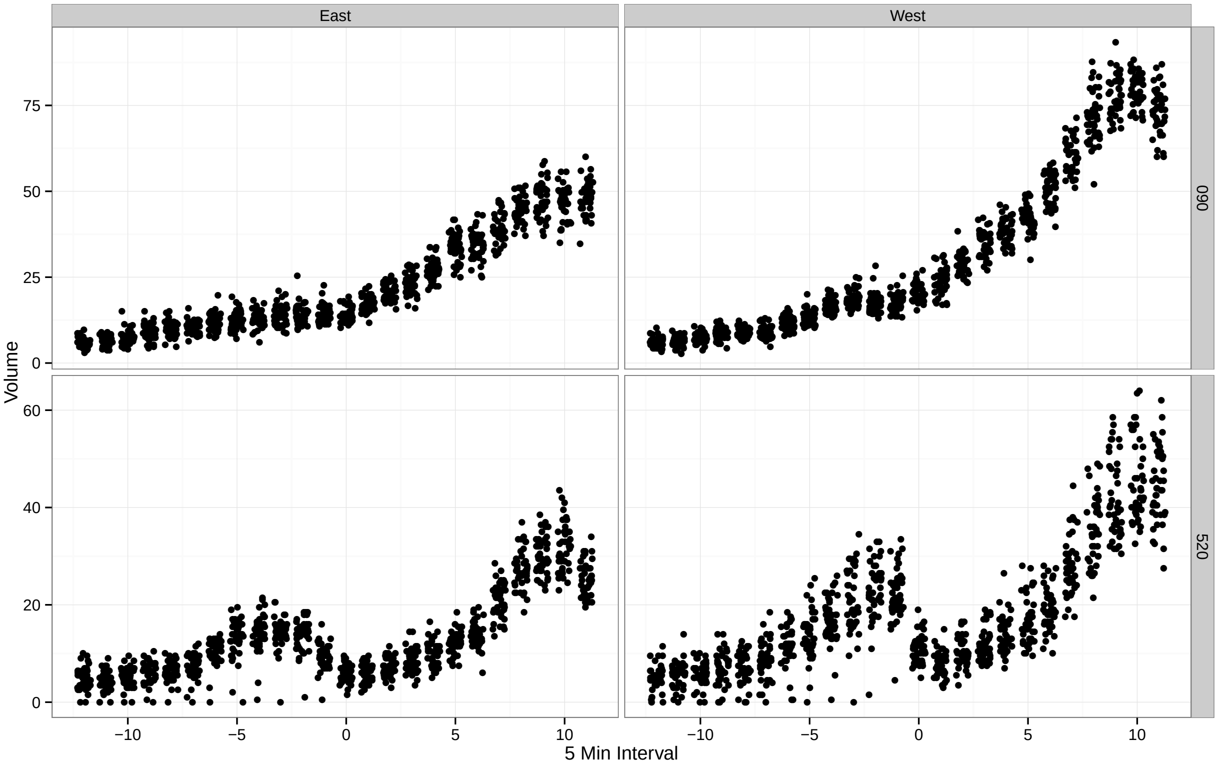
We'd like to estimate $E[Y_1 - Y_0|X = x_0]$ where x is a forcing variable and $x \geq x_0$ receives treatment. Using Lee and Card [2008]'s notation, we can show that if both $E[Y_1|X = x]$ and $E[Y_0|X = x]$ are continuous in x at x_0

$$\begin{aligned} & E[Y|X = x_0] - \lim_{e \to 0^+} E[Y|X = x_0 - e] \\ & \quad = E[Y_1|X = x_0] - \lim_{e \to 0^+} E[Y_0|X = x_0 - e] \\ & \quad = E[Y_1 - Y_0|X = x_0] \end{aligned}$$

Identification relies on observations arbitrarily close to the cutoff. The data in our project setting is, however, reported at five minute intervals. The violation of continuity is illustrated by Figure 3.7. The points have been jittered to better display the density of observations. While it may be clear that there is structure to the underlying process around the cutoff, we cannot rely on the strict RD definition for proper estimation. Rather, the previous transformation is instead[3]

$$E[Y|X = x_0] - \lim_{e \to 0^+} E[Y|X = x_0 - e]$$
$$= E[Y|X = 0] - E[Y|X = x_k - 1].$$

Figure 3.7: Jitter Plot Around 5am Price Change



Note: x-axis is 5 minute intervals before and after price change (x=0), y-axis is volume. Data is from 1 February 2012 through 3 April 2012 and includes 492 observations on either side of the price change. The points have been jittered to give a better sense of the density.

---

[3] For a complete treatment please see Lee and Card [2008].

While nonparametric estimation is not available in this setting, some simple and flexible parametric assumptions will still allow us to estimate a model. We thus reformulate the RD estimation equation using $h(\cdot)$ as a continuous function

$$\begin{aligned} E[Y|X = x_j] &= D_j\beta_0 + h(x_j) \\ \implies Y_{ij} &= D_j\beta_0 + h(x_j) + \epsilon_{ij} \\ \implies Y_{ij} &= \alpha_0 + D_j\beta_0 + X_j\gamma_0 + a_j + \epsilon_{ij} \end{aligned} \tag{3.3}$$

where $h(0) = E[Y_0|X = 0]$. Equation 3.3 is then the reduced form estimation equation with α_0 as the intercept, $D_j = 1[x_j \geq 0]$ as the treatment dummy, X_j representing polynomial terms in x_j, $a_j = h(x_j) - X_j\gamma_0$ and $\epsilon_{ij} \equiv Y_{ij} - E[Y_{ij}|X = x_j]$.

We address the within-group correlation introduced by this specification by computing heteroskedasticity and clustered standard errors [Lee and Card, 2008]. A final point to consider regarding the errors is the error in specification changes according to treatment. We have no reason to expect the specification error to be sensitive to the cutoff.

3.4 Results

3.4.1 Difference-in-Differences

We restrict the sample period for DD estimation to 1 January 2011 through 31 December 2012. This captures the year before and the year following the introduction of tolling, imposed on 29 December 2011. This focuses the analysis on the adjustment around the imposition of the tolls.

Table 3.1 presents pooled results for Equation 3.1. The imposition of the toll leads to a change of -30.44 vehicles per five minute interval on SR520. This is measured at the mean for both directions during the tolled hours (8am - 10pm) and represents a 29% decline in volume from pretreatment levels. Conversely I90 experiences a change of 14.04 vehicles per five minute interval during tolled hours (17% increase). While the signs of the treatment effect are as expected, it is clear that not everyone who switched away from SR520 substituted to I90. Some of these drivers may have switched modes, such as carpooling or public transport, while others may drive the long way around the lake.[4]

Table 3.1: Pooled Results

	SR520	I90
Toll Period * Post Treatment	-30.44^{***}	14.04^{***}
	(0.83)	(0.75)
Post Treatment	-3.71^{***}	1.93^{***}
	(0.36)	(0.27)
Toll Period	105.02^{***}	81.32^{***}
	(0.60)	(0.55)
Intercept	1.79^{*}	1.60^{*}
	(3.48)	(3.38)
R^2	0.58	0.60
Adj. R^2	0.58	0.60
Num. obs.	278744	278744

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Note: Dependent variable is 5 min traffic volume. Models include week of year and day of week controls. Standard errors are Newey-West with 26 lags.

Pretreatment averages during the periods to be tolled are captured by the 'Toll Period'

---

[4] Travel from downtown Bellevue to downtown Seattle takes 15 minutes in free flow using SR520 and 45 minutes taking I405 and SR522 according to Google Maps. The time difference will be shorter for those who start their commute north of Bellevue.

variable. SR520 was a popular choice during the hours that were eventually tolled (5am-11pm), with an average of 105.02 vehicles compared to 81.32 on I90. Finally, the pretreatment average volume across all five-minute intervals is shown by the intercept for 'SR520' to be 1.79 and 1.60 for 'I90'. Given that SR520 has two fewer lanes in each direction and a 10mph slower speed limit we are confident that this represents a strong demand signal.

The 'Post Treatment' variable displays the average change across all five minute periods after the toll was introduced. Again, the displacement from SR520 of -3.71 vehicles does not exactly match the average rise in vehicles on I90 of 1.93. This supports the intuition that while both bridges' catchment areas overlap, they are not perfect substitutes for one another.

Eastbound I90 displayed different pre and post treatment trends (Figure 3.4) leading to potential bias in the DD estimation. We present results from estimation of each each direction separately in Table 3.2. The coefficients on SR520 are virtually unchanged but the results indicate that the violation of the common trend assumption for eastbound I90 is biasing the treatment effect variable upwards (19.13 eastbound compared to 8.94 westbound). Interpreting the magnitudes as percent changes reveals a more nuanced picture. SR520 eastbound experiences a 32.8% decrease in volume with a 26% decrease for westbound. Conversely eastbound I90's volume increases by 26.1% while for westbound I90 it is only 10%. Given the differences in the stability of results between directions for SR520 and I90 we next examine time-of-day specific effects.

Figure 3.8 compares the treatment effect between SR520 and I90, presenting the DD ('Post Treatment * 5 Minute Interval') coefficient estimates of the treatment effect for each bridge with the untolled period as the control. Error bars represent the 95% confidence interval on the coefficient. Results for SR520 are presented in the upper panel, I90 in the lower. Each bridge was estimated separately pooling data from both directions.

Here evidence of spreading around rate changes is clear. The upper graph shows increases

Table 3.2: Direction Specific Results

	SR520 East	SR520 West	I90 East	I90 West
Toll Period * Post Treatment	-30.35^{***}	-30.52^{***}	19.13^{***}	8.94^{***}
	(0.98)	(1.20)	(1.06)	(1.01)
Post Treatment	-3.56^{***}	-3.86^{***}	2.43^{***}	1.42^{***}
	(0.47)	(0.55)	(0.45)	(0.35)
Toll Period	92.53^{***}	117.51^{***}	73.18^{***}	89.47^{***}
	(0.73)	(0.84)	(0.73)	(0.76)
Intercept	1.76	1.82	1.08	2.12
	(3.86)	(5.04)	(5.03)	(4.43)
R^2	0.61	0.62	0.59	0.63
Adj. R^2	0.61	0.62	0.59	0.63
Num. obs.	139372	139372	139372	139372

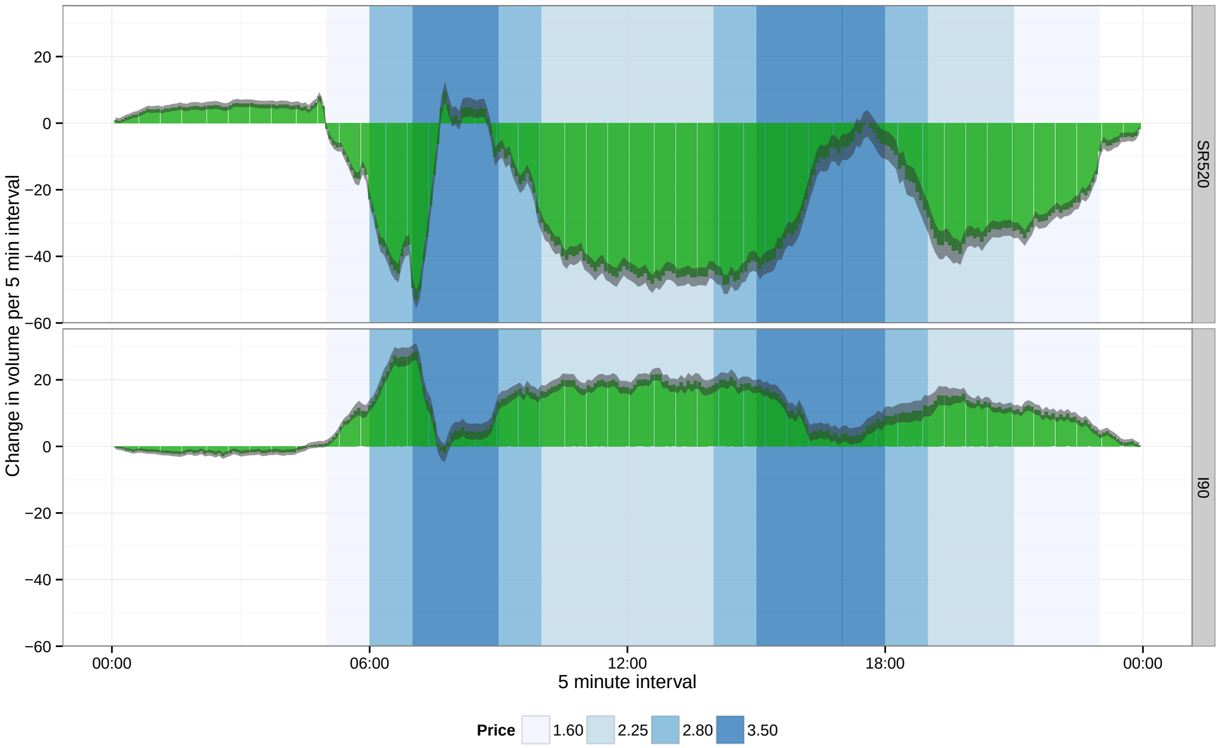
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Note: Dependent variable is 5 min traffic volume. Models include week of year and day of week controls. Standard errors are Newey-West with 26 lags.

in hourly volume on SR520 before tolling begins at 5am and decreases immediately after the toll period begins. During the morning commute the treatment effect on SR520 increases in magnitude as the hour progresses, decreasing before the next price increase and then resuming its increase. This pattern is mirrored on I90: while the magnitudes do not match, the overall adjustment of traffic between the bridges is clear.

Unsurprisingly the largest decrease in volume on SR520 corresponds to the beginning of the highest toll at 7am. The second greatest period of decreased volume on SR520 is during the middle of the day, when the toll has been reduced to \$2.25 from a peak of \$3.50. This

Figure 3.8: 5 Minute Interval Effect of Toll Introduction, including 95% CI



Note: Dependent variable is five-minute traffic volume. Dummies for every five-minute interval are compared to a reference level of 11pm to 4am. Models include week of year and day of week controls. Standard errors are Newey-West with 26 lags. The height of the bars indicate the treatment effect in each five-minute interval. The grey shaded region indicates the 95% CI. Tolled periods are shaded by toll level.

suggests that for midday trips, which may be less time-sensitive, the substitution to untolled routes is perceived as less costly. These results compare well with the I90 graph, showing increases in hourly volume on I90 corresponding to decreases on SR520.

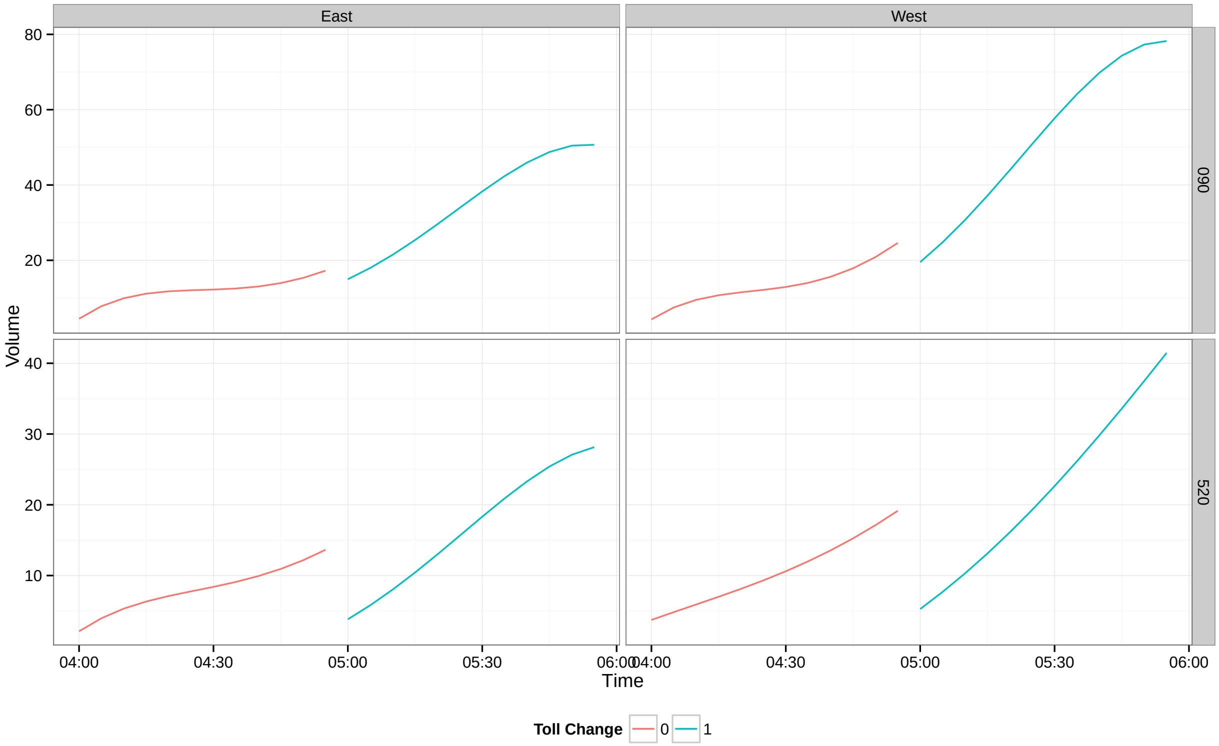
There is an interesting increase in volume between 7:30am and roughly 8:30am on SR520, although of moderate magnitude. A simple interpretation is that the reduction in congestion allows more vehicles to cross the bridge during those intervals. It might also be that the reduced travel time allows commuters to start their commute later based on reduced travel time estimates due to lower congestion. Lastly, this time period may consist of drivers that do not have any flexibility in their commute; for example they may need to drop children off at school and arrive at work at a specific time. This pattern is also evident in the evening commute, where the treatment effect is insignificant at 5:30pm on SR520.

3.4.2 Regression Discontinuity

In this section we explore behavior around toll change cutoffs and therefore restrict the dataset to the period after toll implementation (2012-01-01). Figure 3.9 presents the fitted results of Equation 3.3 where models were estimated around each cutoff for each bridge in each direction separately. The models included fourth-order polynomials in the forcing variable (chosen using methods from Calonico et al. [2014a], see appendix for more details) with standard errors corrected for heteroskedasticity and group-wise correlation [Lee and Card, 2008]. The fitted values tell a very similar story to Figure 3.7. There is a clear pattern of commuters pushing themselves to marginally earlier commute times to avoid paying higher tolls. The pattern in this figure matches the overall pattern captured by the DD treatment effect graph in Figure 3.8. The treatment effects recovered from discrete RD estimation are shown in Figure 3.10.

These results represent only slight ‘spreading’ of commuters into the lower tolled periods of commuting. When compared to the DD results above it is clear that the adjustment most commuters are making is away from tolled routes entirely. The RD results suggest that those who continue to use the tolled routes do marginally adjust the timing of their trips to avoid higher tolls. They also serve as a robustness check on the validity of the DD findings. And while they are not directly comparable to the DD results we will nevertheless be able to construct elasticities of substitution which we present in the next section.

Figure 3.9: RD Fitted Values



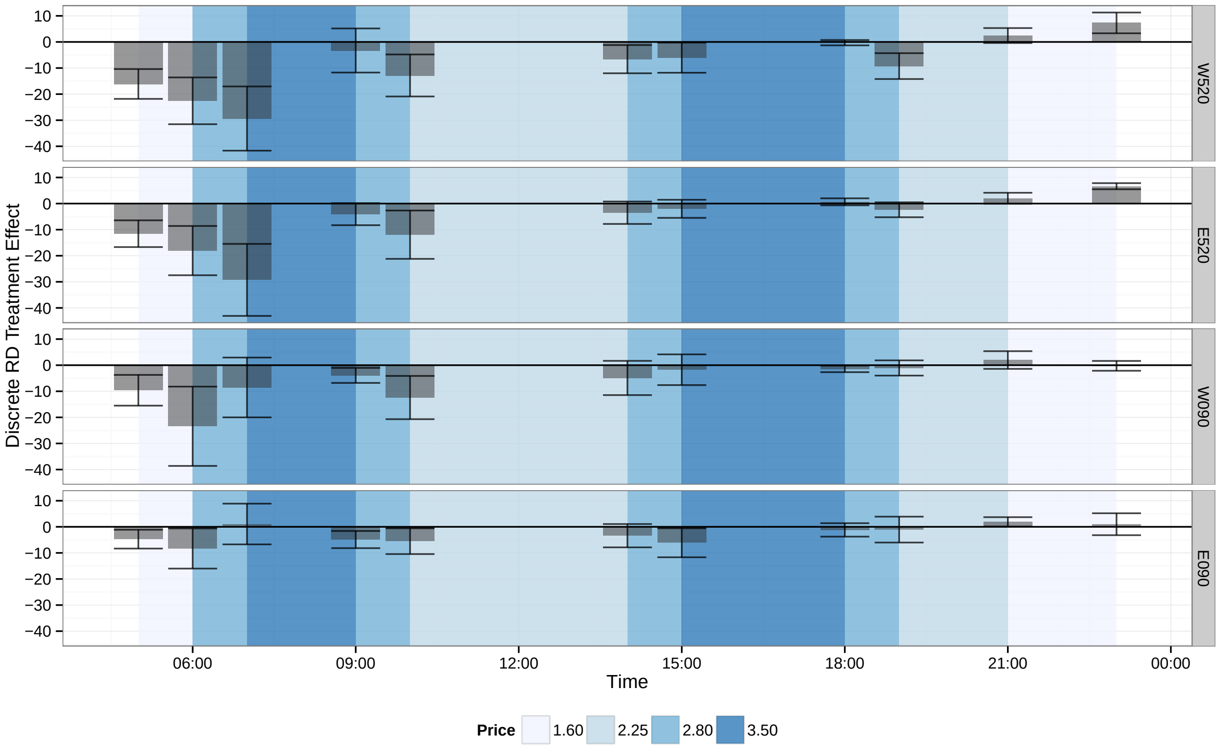
Note: Fitted values from a 4th order polynomial in the forcing variable including a treatment dummy. Standard errors are heteroskedasticity and cluster corrected.

3.4.3 Price Elasticity

In this section we present elasticities calculated from the DD and RD results. For the DD estimation we will be able to calculate own- and cross-price elasticities. For the RD estimation we will calculate 'spreading' elasticities. We begin with the DD results.

To recover price elasticities from the DD estimation we must change the data. Instead of holding the base period to be the untolled portion of the day, making the change in price infinite, we instead hold the beginning of the tolled period as the base. Specifically, we

Figure 3.10: RD Treatment Effect

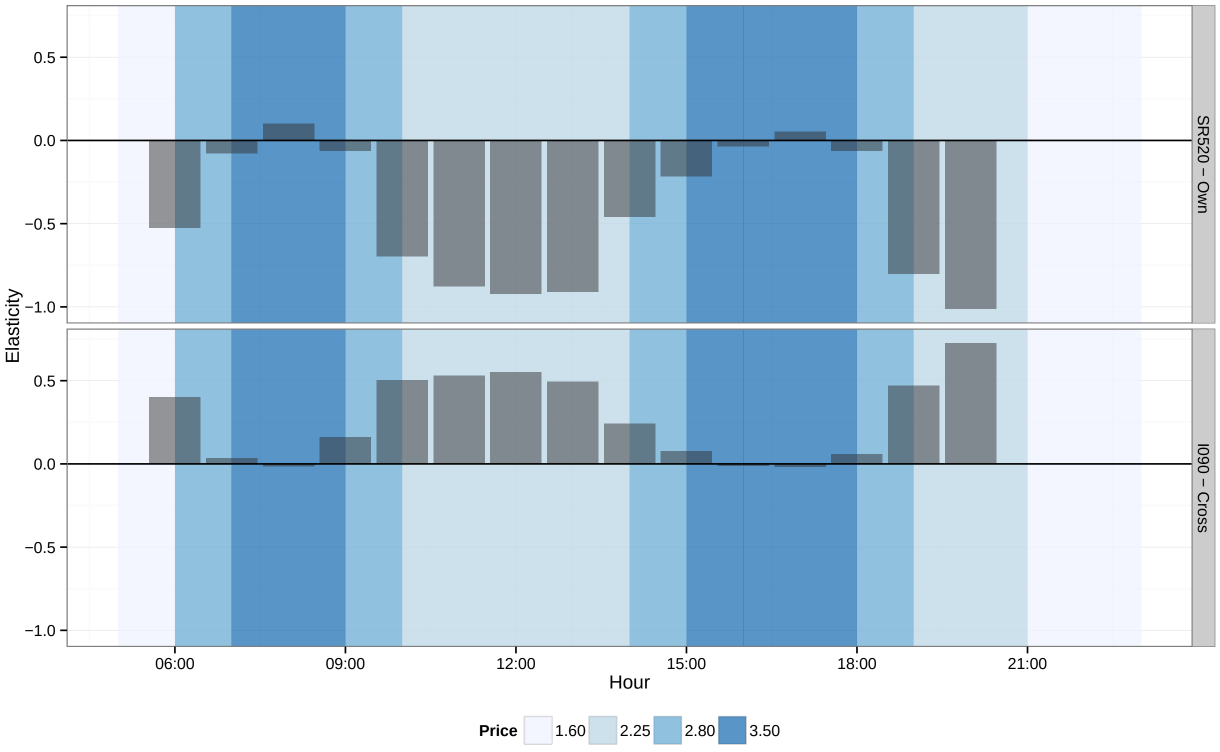


Note: Treatment effect estimated using parametric RD for SR520. Estimation equation includes a fourth-order polynomial in the forcing variable. Standard errors are corrected for heteroskedasticity and clusters.

estimate Equation (3.1) restricting the data to the tolled period (5am - 10pm). The 5am hour is now the base. Day of week and week of year fixed effects are included. All coefficients are significant at the 1% level except for 6pm. The model was estimated separately for each bridge in each direction.

Figure 3.11 presents the own- and cross-price elasticities for SR520 and I90. The own-price elasticities span positive and negative directions, having a minimum of -1.01 at 8pm and a maximum of 0.10 at 8am. While it is surprising to find a positive elasticity, as was

Figure 3.11: Own and Cross Price Elasticities



Note: Models were run separately for each bridge, including only treated hours with 5am as the base level. Day of week and week of year fixed effects were included. All coefficients were significant at the 1% level except for 6pm.

presented in the DD treatment results above (Figure 3.8), there is an increase in volume during the peak commute period after the introduction of tolling. The increase is likely due to the relaxing of congestion and the increased throughput of cars. Regardless, that the elasticities surrounding the periods of highest toll are either slightly negative/positive suggests that the step-sizes in the toll are too shallow. We will return to this point in Section 3.5.

The cross-price elasticities, while smaller in magnitude than their own-price counterparts, represent a significant shift in route choice behavior. With a minimum of −0.02 at 5pm and a maximum of 0.72 at 8pm.

These elasticites are larger than reported in previous research [Foreman, 2012, Brent and Gross, 2017, Finkelstein, 2009]. That people respond so strongly to pricing during the shoulder commute times is perhaps less surprising upon reflection: it is precisely these times during which people have the most leeway in their route choices. They are potentially less constrained in the their start times and therefore the most able to make substitutions to avoid tolls. In terms of the midday reductions it may be that the nature of the trip (not commuting to a job) increases the salience of the toll.

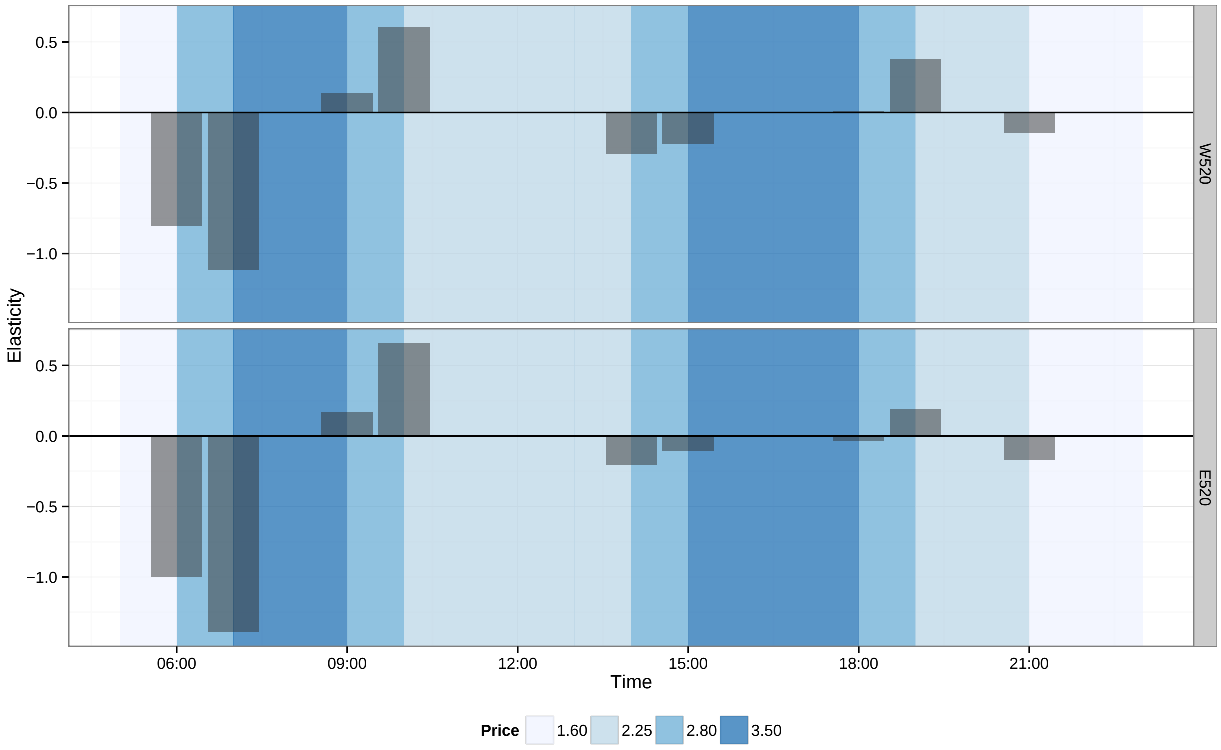
The spreading elasticities display the appropriate direction given the price changes they represent. For instance, the largest in magnitude is −1.39 at 6am when price increases from \$1.60 to \$2.25. Conversely the smallest in magnitude is 0.01 at 6pm when the price is being reduced from \$3.50 to \$2.80. We must be careful, however, to interpret the RD results in the correct light. These elasticities represent the behavior of drivers who find themselves in the vicinity of the cutoff. There are two characteristics of this group worth mentioning: that they may have self-selected by adjust their departure times to avoid the toll, and/or that they may have adjusted their speed to arrive on a specific side of the cutoff. In either sense these results bolster the findings of the DD approach above, in that they capture the same pattern of adjustment around the toll changes.

3.4.4 Consumer Welfare

What, then, is the benefit to society?

To estimate the degree to which consumers benefited from tolling on SR520 examine the effect on travel-times. Given the complexity of routing before and after bridge choice we restrict analysis to crossing times on either SR520 or I90. Re-estimating Equation 3.2 with travel-time as the dependent variable produces results displayed in Figure 3.13.

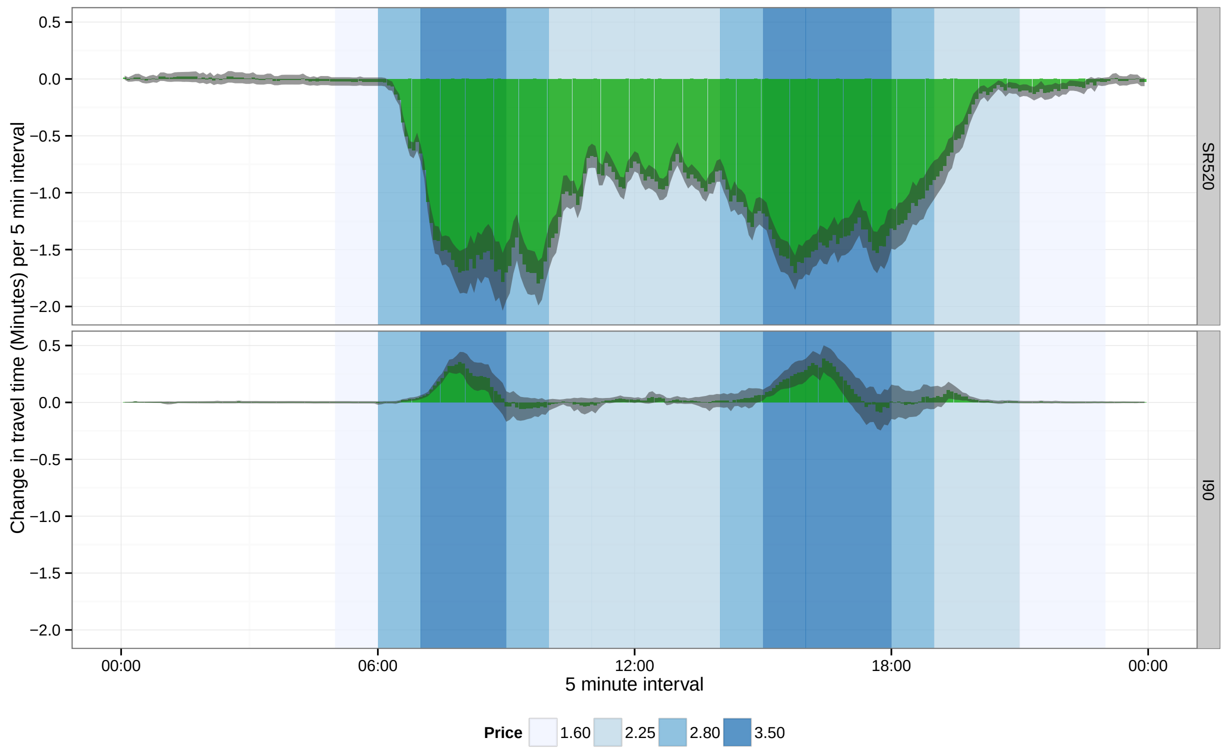
Figure 3.12: RD Spreading Elasticities - SR520



The travel times constructed lack data from sensors on the middle span of SR520, a section which can exhibit some of the worst congestion. We thus view the travel times, and the resulting estimates of the effect of tolling on travel time, to be lower bounds. That is, without information on the most congested portion of the bridge we will be estimating travel times that are likely below those realized by commuters. This is of course true in both the pre- and post-toll sample. Given the changes in volume that we presented above we believe the benefits in congestion to be in excess of those captured by the travel time analysis.

Total estimated time saved across both bridges is 309.16 hours a day. We arrived at this figure by multiplying the results presented in Figure 3.13 by the average volume in each five

Figure 3.13: Effect of Tolling on Travel Times



Note: Dependent variable is travel time to cross either bridge. Dummies for every five-minute interval are compared to a reference level from 11pm to 4am. Models include week of year and day of week controls. Standard errors are Newey-West with 26 lags. The height of the bars indicate the treatment effect in each five-minute interval. The grey shaded region indicates the 95% CI. Tolled periods are shaded by toll level.

minute period post-toll. Using Value of Time estimates (between \$5 and \$30 to construct the total value of time saved leads to estimates between \$1,546 and \$9,2784 of time savings every day.

3.5 Conclusion

We analyze the effect of introducing Time-of-day congestion pricing on one of two possible routes crossing Lake Washington in the Seattle Metropolitan area. We find that imposing tolls that vary according to historical traffic volume causes commuters in different time periods to vary their response to the toll. Our findings support the hypothesis that drivers respond to the imposition of tolls on SR520 by spreading out their commute around the higher toll periods and also by switching their commute to the free alternative of I90.

We recover elasticities around each toll level and also around each price change. We also use estimates of travel time savings to compute the magnitude of consumer welfare from the creation of tolls, finding daily savings to consumers through decreased travel times between $\$1,546$ and $\$9,2784$ on a daily basis.

Future work around the dynamics of adjustment, where people substitute away from the tolled option and then potentially return, is also of interest.

BIBLIOGRAPHY

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

Alan Altshuler. Equity, pricing, and surface transportation politics. *Urban Affairs Review*, 2010.

Alex Anas and Hyok-Joo Rhee. Curbing excess sprawl with congestion tolls and urban boundaries. *Regional Science and Urban Economics*, 36(4):510–541, 2006. ISSN 0166-0462. doi: 10.1016/j.regsciurbeco.2006.03.003.

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

Richard Arnott, André de Palma, and Robin Lindsey. Route choice with heterogeneous drivers and group-specific congestion costs. *Regional Science and Urban Economics*, 22 (1):71–102, 1992.

Richard Arnott, André De Palma, and Robin Lindsey. The welfare effects of congestion tolls with heterogeneous commuters. *Journal of Transport Economics and Policy*, pages 139–161, 1994.

Patrick Athol. Interdependence of certain operational characteristics within a moving traffic stream. Technical report, 1965. URL `http://trid.trb.org/view.aspx?id=117031`.

Siddhartha Banerjee, Ramesh Johari, and Carlos Riquelme. Dynamic pricing in ridesharing platforms. *ACM SIGecom Exchanges*, 15(1):65–70, 2016.

William J Baumol and Wallace E Oates. *The theory of environmental policy*. Cambridge University Press, 1988.

Antonio M. Bento, Kevin Roth, and Andrew Waxman. The value of urgency: Evidence from congestion pricing experiments. 2014.

Anthony E Boardman and Lester B Lave. Highway congestion and congestion tolls. *Journal of Urban Economics*, 4(3):340–359, 1977. URL `http://www.sciencedirect.com/science/article/pii/009411907790016X`.

Daniel A. Brent and Austin Gross. Dynamic road pricing and the value of time and reliability. *Journal of Regional Science*, pages n/a–n/a, 2017. ISSN 1467-9787. doi: 10.1111/jors.12362. URL `http://dx.doi.org/10.1111/jors.12362`.

David Brownstone, 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*, 37(4):373–387, 2003. URL `http://www.sciencedirect.com/science/article/pii/S0965856402000216`.

Bureau of Labor Statistics. Occupational employment and wages,. Technical report, U.S. Department of Labor, 2014.

Mark Burris, 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*, (2297):47–55, 2012.

Sebastian Calonico, Matias D Cattaneo, and Rocio Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014a. ISSN 0012-9682. doi: 10.3982/ecta11757. URL `http://dx.doi.org/10.3982/ECTA11757`.

Sebastian Calonico, Matias D Cattaneo, and Rocio Titiunik. Optimal data-driven regression discontinuity plots. 2014b.

A Colin Cameron and Pravin K Trivedi. *Microeconometrics: methods and applications*. Cambridge university press, 2005.

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

Carlos Carrion and David Levinson. Valuation of travel time reliability from a gps-based experimental design. *Transportation Research Part C: Emerging Technologies*, 35:305–323, 2013.

CDM Smith. Sr 520 bridge: Investment grade traffic and revenue study update. Technical report, 2017.

Chih-Lin Chung and Will Recker. State-of-the-art assessment of toll rates for high-occupancy and toll lanes. In *Transportation Research Board 90th Annual Meeting*, number 11-1331, 2011.

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

Janet Currie and Reed Walker. Traffic congestion and infant health: Evidence from e-zpass. *American Economic Journal. Applied Economics*, 3(1):65, 2011.

Joy Dahlgren. High-occupancy/toll lanes: where should they be implemented? *Transportation Research Part A: Policy and Practice*, 36(3):239–255, 2002. ISSN 0965-8564. doi: 10.1016/s0965-8564(00)00047-1.

Michel De Lara, André De Palma, Moez Kilani, and Serge Piperno. Congestion pricing and long term urban form: Application to paris region. *Regional Science and Urban Economics*, 43(2):282–295, 2013.

Prem Chand Devarasetty, 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*, 46(8):1227–1240, 2012.

Chengri Ding and Shunfeng Song. Traffic paradoxes and economic solutions. *Journal of Urban Management*, 1:63–76, 2012.

Anthony Downs. The law of peak-hour expressway congestion. *Traffic Quarterly*, 16(3), 1962. URL `http://trid.trb.org/view.aspx?id=694596`.

Anthony Downs. *Still Stuck in Traffic: Coping with Peak-Hour Traffic Congestion.* Brookings Institution Press, 2004. URL `http://books.google.com/books?hl=en&lr=&id=ckLcxEb5tM8C&oi=fnd&pg=PR7&dq=downs+2004+traffic&ots=fFt9FRzI0_&sig=PWuGYCPjj-0qfVqCO49ccuwM2tw`.

Gilles Duranton and Matthew A Turner. The fundamental law of road congestion: Evidence from us cities. *American Economic Review*, 101(6):2616–2652, 2011. ISSN 0002-8282. doi: 10.1257/aer.101.6.2616.

Joakim Ekström, Leonid Engelson, and Clas Rydergren. Optimal toll locations and toll levels in congestion pricing schemes: a case study of stockholm. *Transportation Planning and Technology*, 37(4):333–353, 2014.

Federal Highway Administration. Road pricing as a demand management tool, a. URL `https://www.fhwa.dot.gov/ipd/tolling_and_pricing/defined/demand_mgmt_tool.aspx`. Accessed: 2018-02-10.

Federal Highway Administration. Toll facilities in the united states, b. URL `https://www.fhwa.dot.gov/policyinformation/tollpage/history.cfm`. Accessed: 2017-05-20.

Federal Highway Administration, 2015. URL `https://www.fhwa.dot.gov/interstate/faq.cfm`.

Amy Finkelstein. E-ztax: Tax salience and tax rates. *Quarterly Journal of Economics*, 124 (3), 2009.

Kate Foreman. Crossing the bridge: The effectiveness of time-varying tolls on curbing congestion in the bay area. Technical report, Working paper, Berkeley, CA, 2012.

Matthew Gibson and Maria Carnovale. The effects of road pricing on driver behavior and air pollution. *Journal of Urban Economics*, 89:62 – 73, 2015. ISSN 0094-1190. doi: http://dx.doi.org/10.1016/j.jue.2015.06.005. URL `http://www.sciencedirect.com/science/article/pii/S0094119015000467`.

Moshe Givoni. Re-assessing the results of the london congestion charging scheme. *Urban Studies*, 49(5):1089–1105, 2012.

Douglas Holtz-Eakin. Testing for individual effects in autoregressive models. *Journal of Econometrics*, 39(3):297–307, 1988. URL `http://www.sciencedirect.com/science/article/pii/0304407688900607`.

Douglas Holtz-Eakin, Whitney Newey, and Harvey S Rosen. Estimating vector autoregressions with panel data. *Econometrica*, pages 1371–1395, 1988.

Erik Hysing. Citizen participation or representative government–building legitimacy for the gothenburg congestion tax. *Transport policy*, 39:1–8, 2015.

Kyung So Im, M.Hashem Pesaran, and Yongcheol Shin. Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115(1):53–74, 2003. ISSN 0304-4076. doi: 10.1016/s0304-4076(03)00092-7.

Guido W. Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635, 2008. ISSN 0304-4076. doi: 10.1016/j.jeconom.2007.05.001. URL `http://dx.doi.org/10.1016/j.jeconom.2007.05.001`.

Karolina Isaksson, Hans Antonson, and Linnea Eriksson. Layering and parallel policy making–complementary concepts for understanding implementation challenges related to sustainable mobility. *Transport Policy*, 53:50–57, 2017.

Michael Janson and David Levinson. Hot or not. *Research in Transportation Economics*, 44:21–32, 2014. ISSN 0739-8859. doi: 10.1016/j.retrec.2014.04.008.

Robert S Kirk and William J Mallett. Funding and financing highways and public transportation. *Congressional Research Service*, page 23, 2013.

Christopher R Knittel and Ryan Sandler. Carbon prices and automobile greenhouse gas emissions: the extensive and intensive margins. Technical report, National Bureau of Economic Research, 2010.

Hideo Konishi and Se-il Mun. Carpooling and congestion pricing: Hov and hot lanes. *Regional Science and Urban Economics*, 40(4):173–186, 2010. ISSN 0166-0462. doi: 10.1016/j.regsciurbeco.2010.03.009.

David S Lee and David Card. Regression discontinuity inference with specification error. *Journal of Econometrics*, 142(2):655–674, 2008. ISSN 0304-4076. doi: 10.1016/j.jeconom.2007.05.003. URL `http://dx.doi.org/10.1016/j.jeconom.2007.05.003`.

David Levinson. Equity effects of road pricing: A review. *Transport Reviews*, 30(1):33–57, 2010.

Jianling Li. Explaining high-occupancy-toll lane use. *Transportation Research Part D: Transport and Environment*, 6(1):61–74, 2001. ISSN 1361-9209. doi: 10.1016/s1361-9209(00)00013-4.

Zheng Li, 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*, 46(3):384–403, 2010.

Robin Lindsey. Do economists reach a conclusion? *Econ Journal Watch*, 3(2):292–379, 2006.

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

Wei Liu, Hai Yang, and Yafeng Yin. Traffic rationing and pricing in a linear monocentric city. *Journal of Advanced Transportation*, 48(6):655–672, 2014. ISSN 2042-3195. doi: 10.1002/atr.1219. URL `http://dx.doi.org/10.1002/atr.1219`.

Xiaoyue Liu, Guohui Zhang, 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*, 2229(-1): 102–109, 2011a. ISSN 0361-1981. doi: 10.3141/2229-12.

Xiaoyue Liu, Guohui Zhang, Yao-Jan Wu, and Yinhai Wang. Measuring the quality of service for high occupancy toll lanes operations. *Procedia - Social and Behavioral Sciences*, 16: 15–25, 2011b. ISSN 1877-0428. doi: 10.1016/j.sbspro.2011.04.425.

David Lupton. A practical rule for a benefit maximizing toll. 2017.

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

Daniel McFadden. The measurement of urban travel demand. *Journal of public economics*, 3(4):303–328, 1974. URL `http://www.sciencedirect.com/science/article/pii/0047272774900036`.

Diana Vonk Noordegraaf, Jan Anne Annema, and Bert van Wee. Policy implementation lessons from six road pricing cases. *Transportation Research Part A: Policy and Practice*, 59:172–191, 2014.

Ian W.H. Parry and Antonio Bento. Estimating the welfare effect of congestion taxes: The critical importance of other distortions within the transport system. *Journal of Urban Economics*, 51(2):339–365, 2002. ISSN 0094-1190. doi: 10.1006/juec.2001.2248.

Ian WH Parry and Kenneth A Small. Does britain or the united states have the right gasoline tax? *The American Economic Review*, 95(4):1276–1289, 2005.

I.W.H. Parry. Comparing the efficiency of alternative policies for reducing traffic congestion. *Journal of Public Economics*, 85(3):333–362, 2002. ISSN 0047-2727. doi: 10.1016/s0047-2727(00)00163-8.

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

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

M. Hashem Pesaran. A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, 22(2):265–312, 2007. ISSN 1099-1255. doi: 10.1002/jae.951.

Arthur Cecil Pigou. *The economics of welfare.* Macmillan, 1920.

Robert D Plotnik, Jennifer Lynn Romich, and Jennifer Thacker. Analyzing the impact of highway tolls on low-income persons: An application to the puget sound region of washington state. Technical report, 2010.

Farideh Ramjerdi and Nils Fearnley. Risk and irreversibility of transport interventions. *Transportation Research Part A: Policy and Practice*, 60:31–39, 2014.

Elena Safirova, 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*, 9:179–206, 2004. ISSN 0739-8859. doi: 10.1016/s0739-8859(04)09008-0.

David Schrank, 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.

Kenneth A Small. The incidence of congestion tolls on urban highways. *Journal of urban economics*, 13(1):90–111, 1983. URL `http://www.sciencedirect.com/science/article/pii/0094119083900475`.

Kenneth A Small. Economics and urban transportation policy in the united states. *Regional Science and Urban Economics*, 27(6):671–691, 1997.

Kenneth A. Small and Jia Yan. The value of 'value pricing' of roads: Second-best pricing and product differentiation. *Journal of Urban Economics*, 49(2):310–336, 2001. ISSN 0094-1190. doi: 10.1006/juec.2000.2195.

Kenneth A. Small, Clifford Winston, and Jia Yan. Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica*, 73(4):1367–1382, 2005. ISSN 1468-0262. doi: 10.1111/j.1468-0262.2005.00619.x.

Adam Smith. *An Inquiry into the Nature and Causes of the Wealth of Nations*, volume 2. 1817.

Claus Hedegaard Sørensen, Karolina Isaksson, James Macmillen, Jonas Åkerman, and Florian Kressler. Strategies to manage barriers in policy formation and implementation of road pricing packages. *Transportation Research Part A: Policy and Practice*, 60:40–52, 2014.

Andrew H Spencer and Chia Lin Sien. National policy towards cars: Singapore. *Transport Reviews*, 5(4):301–323, 1985.

Craig J. Stone. Sr 520: Toll operations and traffic performance summary report. Technical report, Washington State Department of Transportation, 2012.

United States Department of Transportation. State transportation statistics 2015. Technical report, 2015.

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.

Vincent van den Berg and Erik T. Verhoef. Winning or losing from dynamic bottleneck congestion pricing? *Journal of Public Economics*, 95(7-8):983–992, 2011. ISSN 0047-2727. doi: 10.1016/j.jpubeco.2010.12.003.

Erik Verhoef, Peter Nijkamp, and Piet Rietveld. Second-best congestion pricing: the case of an untolled alternative. *Journal of Urban Economics*, 40(3):279–302, 1996.

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

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

John Glen Wardrop. Road paper. some theoretical aspects of road traffic research. *Proceedings of the institution of civil engineers*, 1(3):325–362, 1952.

Washington Joint Transportation Committee. I-405 traffic data and corridor performance analysis. Technical report, 2018.

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

Wilbur Smith Associates. Sr167 hot lanes pilot project traffic and revenue analysis. Technical report, WSDOT, 2006.

Wilbur Smith Associates. Sr 520 bridge investment grade traffic and revenue study. Technical report, Washington State Department of Transportation, 2011.

Clifford Winston. Conceptual developments in the economics of transportation: an interpretive survey. *Journal of Economic Literature*, pages 57–94, 1985.

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

JM Wooldridge. *Econometric Analysis of Cross Section and Panel Data.* MIT press, 2nd edition, 2010. URL `http://books.google.com/books?hl=en&lr=&id=yov6AQAAQBAJ&oi=fnd&pg=PP1&dq=woolridge+cross+section&ots=iVm_yGHCTW&sig=VmTNzOZSwdoDbAL9pFqV2mQNyss`.

WSDOT. Sr 167 hot lanes pilot project fourth annual performance summary. Technical Report May 2008, WSDOT, 2012.

WSDOT. Sr 167 hot lanes - project library. Technical report, 2013. URL `http://www.wsdot.wa.gov/Tolling/SR167HotLanes/publications.htm`.

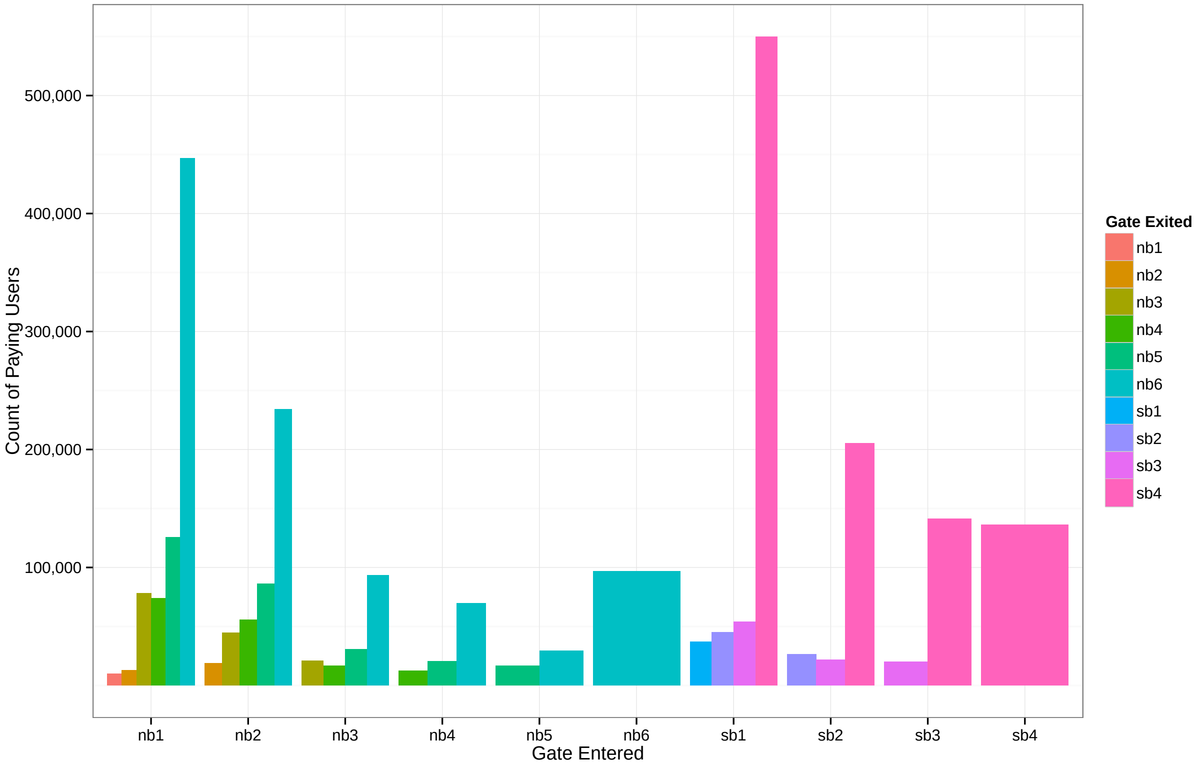
Carol Zimmerman, Rachel Klein, Jeremy Schroeder, Brian Pessaro, Mark Burris, Katie Turnbull, Barbara Joy, and Eric Schreffler. Contemporary approaches to congestion pricing: Lessons learned from the national evaluation of congestion pricing strategies at six sites. Technical report, 2015.

Appendix A

DYNAMIC CONGESTION PRICING

A.1 Trip combinations

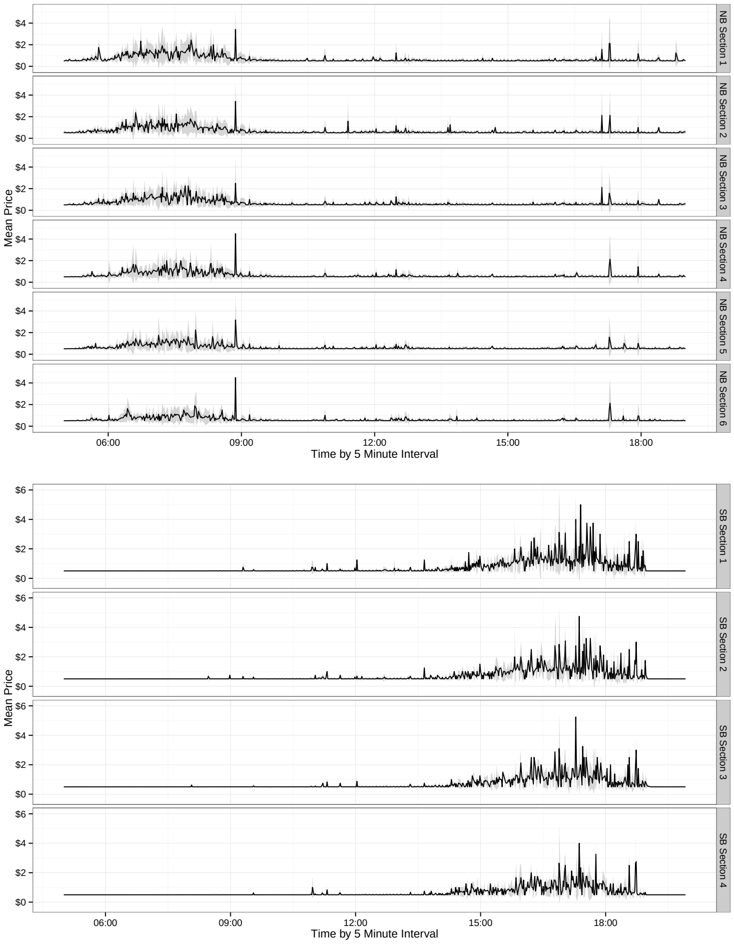
Figure A.1: SR167 trip combinations



Note: 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



Note: The above figures show the mean price surrounded by a shaded region representing one standard deviation for each five minute interval at each entry gate. There is a noticeably larger variation in prices in the southbound direction during the evening compared to the northbound direction in the morning.

A.3 Gate locations

Gate locations were determined using this tool.

Northbound mile posts:

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

Southbound mile posts:

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

A.4 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

Note: 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 obserations 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:

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

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

$$\Delta y_{it} = \alpha_i + \beta_i y_{i,t-1} + \sum_{j}^{p_i} \rho_{ij} \Delta y_{i,t-j} + \epsilon_{it} \tag{A.2}$$

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

$$t\text{-bar} = \frac{1}{N} \sum_{i=1}^{N} t_{\rho_i} \tag{A.3}$$

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

$$\Delta y_{it} = \alpha_i + \beta_i y_{it-1} + \gamma_i \bar{y}_{t-1} + \delta_i \Delta \bar{y}_t + \epsilon_{it} \tag{A.4}$$

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

$$\text{CIPS} = t\text{-bar} \frac{1}{N} \sum_{i=1}^{N} t_{\beta_i} \tag{A.5}$$

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

Note: 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 $1,282,349$.

Appendix B

TIME OF DAY CONGESTION PRICING

B.1 Regression Discontinuity - Continuous Forcing Variable

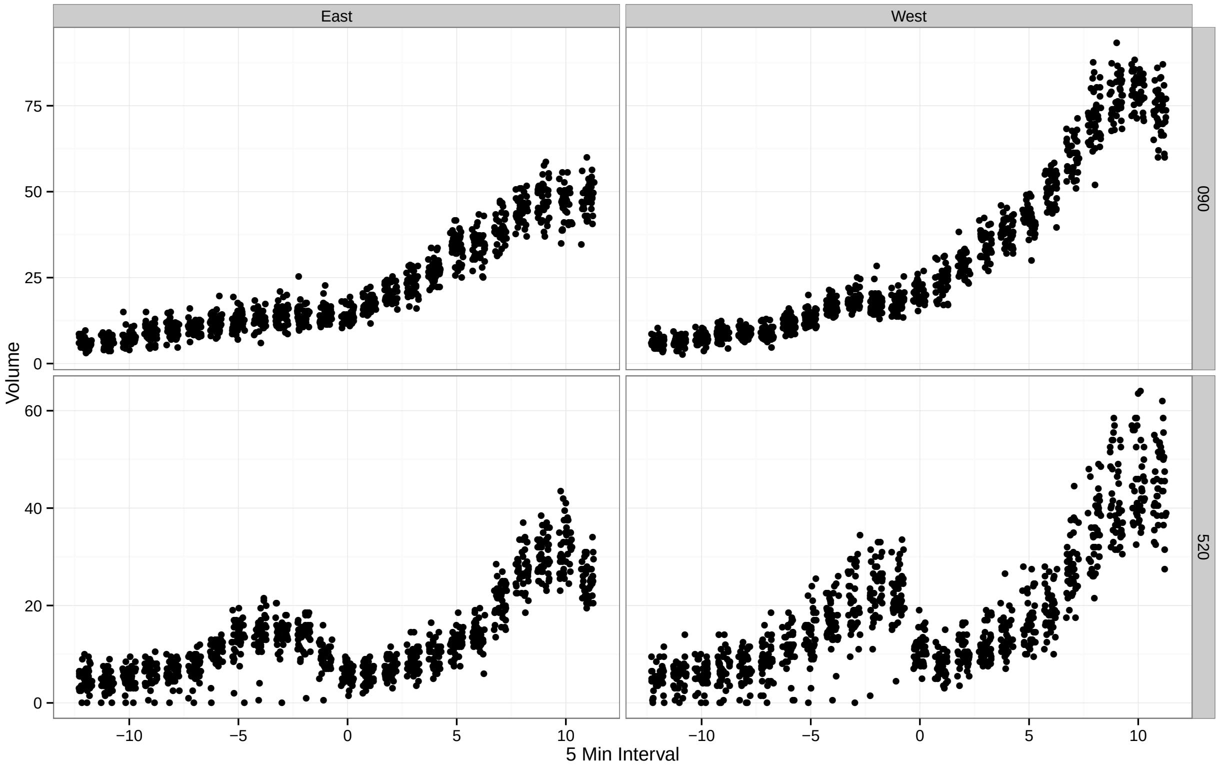
The most crucial assumption in RD is that there is no systematic variation in unobservable characteristics on either side of the cutoff. Given that drivers have very little control over which five-minute traffic bin they fall into we feel certain in asserting that this assumption is satisfied. Three further assumptions underlay RD estimation:

1. that the covariate determining treatment not be discretely valued around the cutoff
2. that the underlying regressions satisfy standard smoothness conditions and
3. that the variance of the forcing variable on either side of the cutoff is bounded away from zero and continuous.

Since RD is very much motivated by graphic analysis we begin confirming these assumptions using Figure B.1. The graph displays traffic volume in 5 minute intervals on either side of the 5am price change. The data is from 1 February 2012 through 3 April 2012 and includes 492 observations on either side of the cutoff. While I90 (upper two panels) exhibits smooth variation in volume in either direction, SR520 has a clear ‘break’ in the evolution of volume around the cutoff representing the 5am toll. There are no other discontinuities on either side of the cutoff (the drop in volume at the 10+ mark is the next price change).

The formal statement of RD is quite simple. A forcing variable X_i determines the treatment T_i of a population unit $i = 1, 2, \ldots, n$ drawn at random from a large population. Units

Figure B.1: Jitter Plot Around 5am Price Change



Note: x-axis is 5 minute intervals before and after price change (x=0), y-axis is volume. Data is from 1 February 2012 through 3 April 2012 and includes 492 observations on either side of the price change. The points have been jittered to give a better sense of the density.

with a X_i greater than a threshold $\bar{x}$ are assigned treatment ($T_i = 1$) and we observe outcome $Y_i(1)$, otherwise treatment is not assigned ($T_i = 0$) and we observe outcome $Y_i(0)$. Sharp RD, which we here employ, assumes perfect compliance and thus treatment is performed according to $1(X_i \geq \bar{x})$, where $\bar{x}$ is the time of a toll rate change.

Drawing on the notation and setup from Imbens and Lemieux [2008] the observed outcome is

$$Y_i = (1 - T_i) * Y_i(0) + T_i * Y_i(1) = \begin{cases} Y_i(0) & \text{if } T_i = 0 \\ Y_i(1) & \text{if } T_i = 1 \end{cases}$$

which makes the average treatment effect in the area around the cutoff

$$\tau_{RD} = E[Y(1) - Y(0)|X = \bar{x}] = E[Y(1)|X = \bar{x}] - E[Y(0)|X = \bar{x}].$$

Although we do not observe $Y(0)$ at $\bar{x}$ we can extrapolate from arbitrarily close observations. This estimation requires smoothness in the forcing variable so that we might estimate

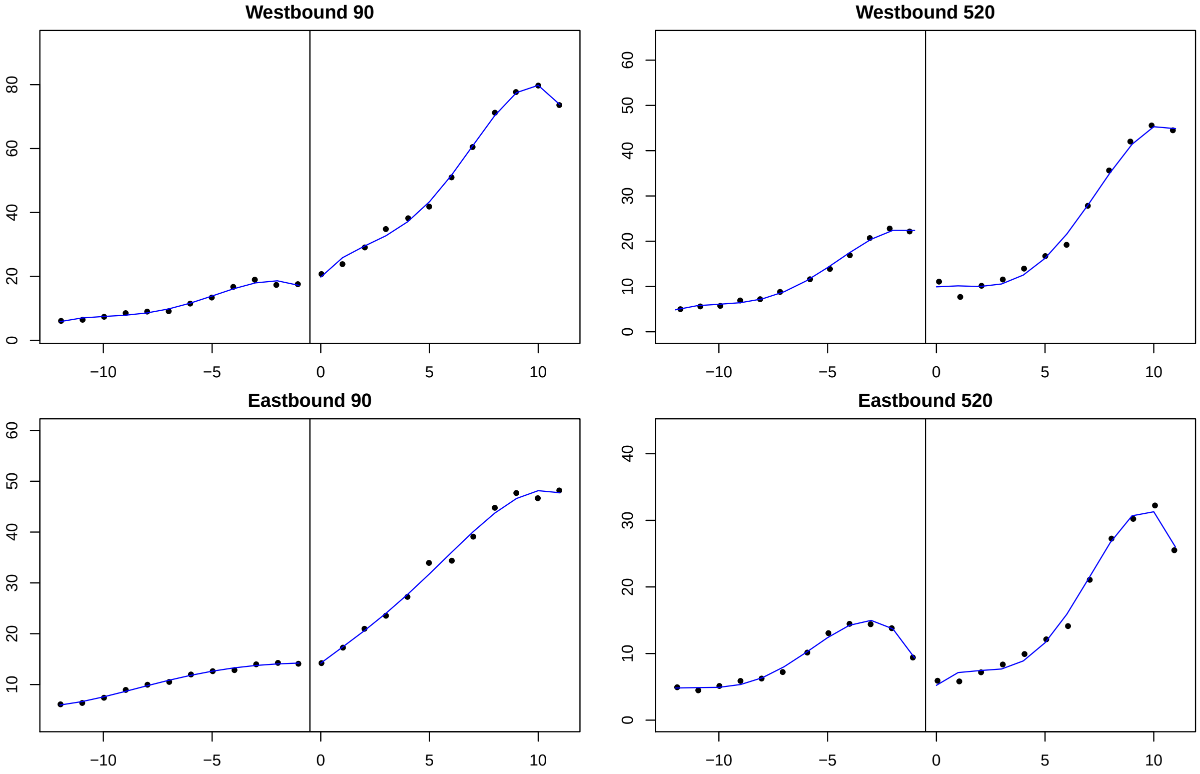
$$\tau_{RD} = \lim_{X \downarrow \bar{x}} E[Y|X = \bar{x}] - \lim_{X \uparrow \bar{x}} E[Y|X = \bar{x}].$$

While we volume at five minute intervals we believe that there are enough mass points for reliable estimation [Calonico et al., 2014a].

B.2 Regression Discontinuity Results

The summary of an RD design can generally be found in a plot of the forcing variable on a regression function estimated for some statistic of the outcome variable. Figure B.2 plots volume sample means of non overlapping evenly spaced bins along with a polynomial regression fitted on either side the 5am price change. The x-axis is five minute intervals, with 0 representing the 5am price change. The cutoff used in estimation is -0.5 (2.5 minutes before the price change). We chose a cutoff slightly to the left of zero based on the logic that drivers avoiding paying the higher toll must plan their trip with a buffer for conditions. Therefore, if drivers are interested in paying a lower toll they should try to arrive at the toll point early enough where they have minimized their chance of arriving after the toll change. This can be observed in the two right panels of Figure B.2, where the five minute period preceding the price change already exhibits a decrease in the binned sample mean.

Figure B.2: RD Plots Around 5am Price Change



Note: x-axis is 5 minute intervals, y-axis is volume. Dots are sample means of non-overlapping bins. Line is 4th order polynomial estimated on either side of the cutoff. The cutoff is at -0.5 (2.5 minutes before price change). The data is from 1 February 2012 through 3 April 2012. See Table B.1 for more information on plot construction.

As shown by the graphs there is no discontinuity in the left two panels representing I90, while a jump before and after the cutoff is evident for the right two panels representing SR520. It is also clear that there are no other candidates for a discontinuity on either side of the cutoff.[1]

Table B.1 presents the results of the data-driven process in the number of bins, scale and

---

[1] Plots for all cutoffs are in the Appendix.

order of the polynomial shown in Figure B.2.[2] Data is presented for each plot, for either side of the cutoff. 'Mimicking Variance bins' shows the optimal number of bins on either side of the cutoff. This implies an optimal 'Bin Length' of between 4% (East 90) and 21% (West 520). The 'IMSE-optimal' (Integrated Mean Squared Error) bins are also reported, along with the 'Scale'.

Table B.1: RD Plot Characteristics

	West 90		West 520		East 90		East 520	
	Left	Right	Left	Right	Left	Right	Left	Right
Bin Length	0.11	0.06	0.46	0.24	0.06	0.06	0.16	0.11
Chosen Scale	1	1	1	1	1	1	1	1
IMSE-optimal bins	18	22	10	13	17	19	21	21
Implied scale	5.56	8.27	2.40	3.46	11	9.37	3.24	4.71
Mimicking Variance bins	100	182	24	45	187	178	68	99
Number of Obs.	492	492	492	492	492	492	492	492
Polynomial Order	4	4	4	4	4	4	4	4
Selected bins	100	182	24	45	187	178	68	99
WIMSE bias weight	0.99	1	0.93	0.98	1	1	0.97	0.99
WIMSE variance weight	0.01	0	0.07	0.02	0	0	0.03	0.01

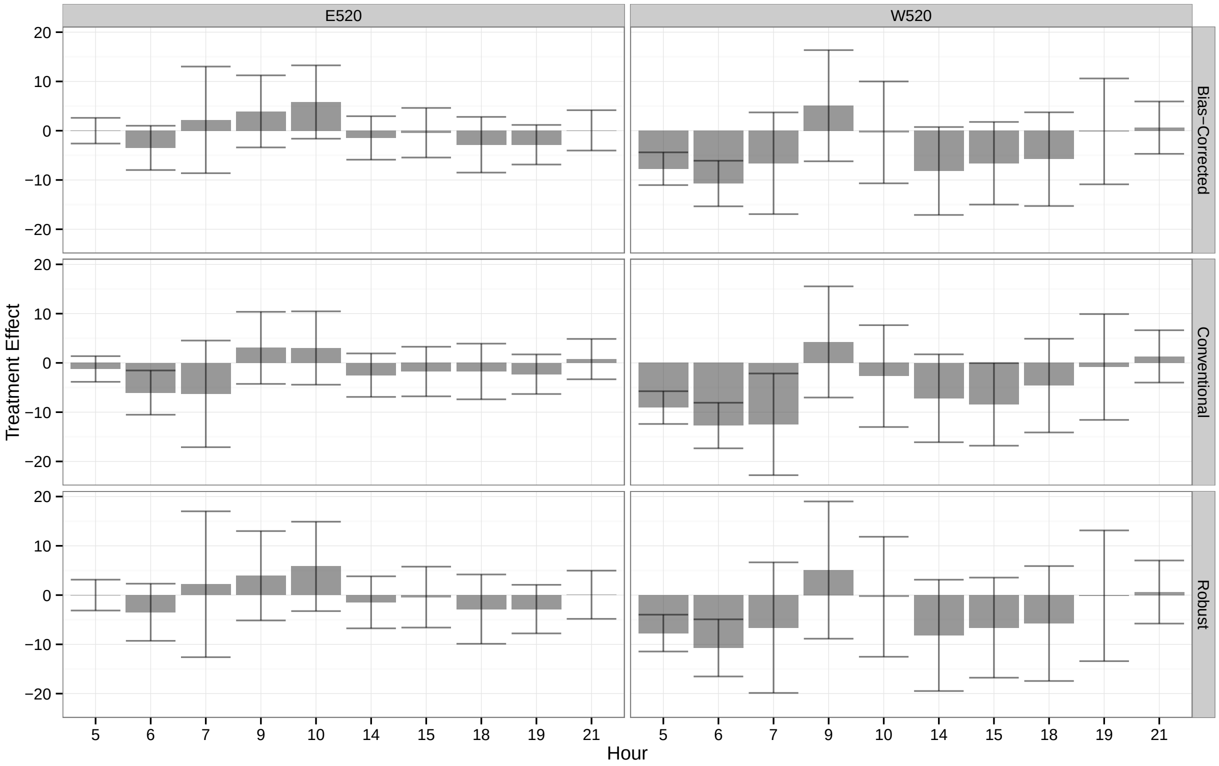
Estimation of the treatment effect for each direction of SR520 at the time of every price change is presented in Figure B.3. The hours are shown along the x-axis, with the left and right panels separating the direction (eastbound and westbound respectively). The magnitude of the effect is shown by the gray bars with the 95% confidence interval overlaid as

---

[2] For more information on the formulation of the underlying regression functions and the selection of the optimal bin size see [Calonico et al., 2014b].

error bars. The three rows in the figure demonstrate the effect of constructing the coefficients and standard errors using different methodologies [Calonico et al., 2014a].

Figure B.3: RD Treatment Effect with Robust Confidence Intervals



Note: x-axis is the hour of the price change. The solid bars represent the estimated treatment effect. The coefficients and error bars are constructed using either bias-corrected, conventional or robust standard methods [Calonico et al., 2014a]. We could not estimate a coefficient for 5am westbound SR520.

While the magnitudes are generally in the correct direction (negative for price increases, positive for price reductions), almost none of the effects are significant. Indeed, only the 5am and 6am price increases on westbound SR520 are consistently significant. This should not be interpreted as an invalidation of the DD results but rather an exploration of the phenomenon of ‘spreading’. By constructing the cutoff around the price change after tolls have been imposed we are visualizing how those who have chosen to pay the toll are adjusting the timing of their trip to compensate for the changing toll rate.