The Value of Time: Evidence from Traffic Congestion and Express Lanes

Andrea Mattia*

Please find the most recent version here

The value of time (VOT) determines the allocation of non-labor time to tasks and is crucial in travel demand and infrastructure, where congestion is a major source of loss. A large literature has estimated the mean VOT over a variety of subpopulations, but commuting choices and welfare effects of congestion policies depend on the individual VOT of the policy-relevant population. In this paper, I estimate the full VOT distribution for a population of drivers, using a rich new dataset in the unique context of highway Express Lanes (ELs), which offer time savings in exchange for a toll. A continuous function of traffic density sets the toll and rounds it to the nearest $0.25, creating 32 separate discontinuities which provide identifying variation. The analysis is divided into three parts. First, using an RDD, I show that EL drivers have a mean VOT of $66.56 per hour saved, substantially exceeding estimates from the literature. Second, the full VOT distribution for all drivers, which rationalizes EL aggregate traffic shares and RD results, shows wide heterogeneity: the median is $17.42 per hour and the 95th percentile is $166.05. Third, I build a structural model that endogenizes departure time (a key form of adjustment) to assess the welfare consequences of a range of counterfactual policies. I find that the EL is welfare-reducing because the value of the increase in travel times for non-users outweighs the benefits for users by $25.68 per year, more than what half of drivers spend on the EL in a year.

*University of Chicago. Email: amattia@uchicago.edu. I would like to thank Stéphane Bonhomme, Michael Greenstone and Alessandra Voena for their invaluable and kind advice, and Sergei Bazylik, Fiona Burlig, Neil Cholli, Thomas Covert, Francisco Del Villar, Manasi Deshpande, Michael Dinerstein, Eyal Frank, Dan Kashner, Ryan Kellogg, Koichiro Ito, Rafael Jiménez, Evan Rose, Lillian Rusk, Harshil Sahai, Laura Sale, Christopher Taber, and Alex Torgovitsky for their comments. I also thank participants in the Midwest Energyfest and the UEA 2021 European Meeting, and in the BFI Applied Economics Workshop, the EPIC Lunch, the EEE workshop and the Applied Microeconomics working group at the University of Chicago for their insightful comments. I thank Mike Solomonson for his help requesting the data and Vincenzo Mustich for his help collecting the data.

1 Introduction

The Value of Time (VOT) is a fundamental determinant of the allocation of non-working time to tasks, because time enters individuals' full income (Becker (1965)). For instance, the VOT determines the time devoted to education and human capital accumulation, the substitution between time and market goods, the intra-household division of labor, the amount of hours worked, the choice of pure leisure goods such as the restaurant or the theater and, in general, the choice regarding any human activity that requires time.

Commuting is a salient and time-consuming activity, and individuals' travel demand depends on their VOT.[1] Workers spend about 8% of their workday commuting, over 75% commute by car and traffic congestion is a major source of time loss.[2] The VOT is a key determinant of where workers choose to direct their job search and how they trade off residential amenities with wage offers that require commuting.[3] The VOT is also the basis for evaluating consumer benefit in infrastructure plans,[4] which often involve large government spending.[5] In recent decades, public policy has shifted from supply expansion to traffic management, which uses price signals to allocate commuters and minimize the time loss due to congestion. The distribution of VOT is crucial to determine how individuals respond to congestion policies and who benefits from them.

While a large literature has estimated the mean VOT, there is little evidence on individual-level, policy-relevant VOT. The scarcity of detailed data has led part of the literature to rely on stated preference (Small et al. (2005), Hall (2020)). When a revealed preference approach was possible, either heterogeneity was only expressed in terms of different VOT means for specific subsamples, or the data referred to a variety of subpopulations not immediately representative of the large body of car commuters (Goldszmidt et al. (2020), Buchholz et al. (2022)). The US Department of Transportation uses a mean VOT of $13.60 for all local personal travel.

In this paper, I estimate the full VOT distribution for a population of commuters, using a revealed preference approach. I consider an increasingly common congestion policy, Express Lanes (ELs), in the setting of Minneapolis-St Paul, where driving on the highway is the main commuting mode, as in the US in general.[6] The setting is especially suitable because there is wide variation in toll and time saved and drivers are making a clear trade-off choice between the EL toll payment and travel time savings. ELs also imply an aggregate trade-off between

[1] For an extensive review of travel demand models, see de Palma et al. (2011).

[2] See Redding and Turner (2015) and the American Community Survey (2016). It has also been estimated that, in 2018 in the largest 66 US metropolitan areas, the average commuter lost 97 hours due to traffic congestion on top of what the commute would take without traffic (INRIX Traffic Scorecard 2018).

[3] See, for instance, Manning and Petrongolo (2017), Monte et al. (2018) and Le Barbanchon et al. (2020).

[4] 60% of benefits in infrastructure evaluations are given by commuters' value of travel time (Hensher (2001)).

[5] For instance, in 2022, the USA introduced a $1 trillion infrastructure bill, the largest in their history.

[6] As of 2022, over 40 Express Lanes have been built by cities across the US as a congestion policy, and more are under construction. An example of EL is provided in *Figure 1*.

the travel time benefits of users and the extra congestion cost of non-users, who have one less free lane available.

I use a new, non-publicly available panel of over 45,000 EL drivers, which constitute about 5% of the relevant population.[7] I observe where drivers enter and exit the EL, what toll they pay, and how long they stay on the EL with precision at the seconds level. I matched this panel with uniquely fine aggregate traffic data observed every 30 seconds on both the EL and the standard lanes. The vast majority of commuters almost never uses the EL, whereas less than 1% of drivers uses it every day.[8]

The identifying variation is provided by the EL tolling algorithm. A continuous function of past traffic in the EL sets the toll every 3 minutes, rounds it to the nearest \$0.25 and caps it at \$8. This creates 32 discontinuity cutoffs, where past traffic on the EL changes continuously while the toll increases discontinuously by \$0.25. Commuters face a mutually-exclusive and exhaustive choice between the EL and the standard lanes. At each cutoff, on each side there are two otherwise equal drivers, but one of them sees a \$0.25 higher toll, which reduces aggregate demand for the EL and produces extra time savings.

I divide the analysis in three parts.

First, with an RDD, the mean VOT conditional on using the EL is \$66.56 per hour saved, about 2.5 times the average Minnesota wage rate in 2018. I identify the travel time savings that drivers are willing to accept in exchange for a \$0.25 toll increase. The RDD estimates the time savings effect off drivers who stay on the EL, and averages across the 32 cutoffs.[9] The VOT is identified as the ratio between \$0.25 and the estimated time savings effect. The RDD is advantageous because it isolates a change in the EL toll that is plausibly exogenous to both drivers' characteristics and unobservable road conditions on each side of the discontinuity.[10] The RD estimates refers to the particular set of EL drivers: how these agents' VOT compares to that of other drivers is crucial to study the welfare effects of the EL.

Second, the VOT distribution for all drivers, which rationalizes EL aggregate traffic shares and RD results, shows wide heterogeneity: the median is \$17.42 per hour saved, the 75th percentile is \$34.97 and the 95th is \$166.05. This estimation relies on the same identifying variation as the RD. The outcome variable is the difference in the aggregate EL choice probability at each cutoff, where the VOT is a random coefficient. The estimation follows the mixture approach by Fox et al. (2011) and Fox et al. (2012). I consider each day in the sample as

[7] In Minneapolis-St Paul there are about 1 million workers who commute from the suburbs to downtown.

[8] In the public debate, ELs have been dubbed "Lexus Lanes", a term also used in many opinion pieces, such as this Washington Post article. In a case study, Khoeini and Guensler (2014) find that EL users in Florida have newer or better cars and that two Lexus models are among the 20 most popular cars on the ELs.

[9] In this sense, this RDD is equivalent to a 2SLS design in which I control flexibly for the running variable and I impose comparisons only locally at the 32 cutoffs.

[10] A standard hedonic OLS of the toll on time saved would suffer from endogeneity. For instance, when agents are more of a hurry or road conditions are worse, they would accept a lower time saved for each level of the toll: the error would be negatively correlated with time saved, which would lead to downward bias.

a separate market, in which a population with the same VOT distribution chooses the same product (the EL) with different attributes (time saved and toll combinations), independent of the VOT. These estimates imply that EL drivers have VOT much higher than average. The VOT distribution by itself determines how agents choose among existing and counterfactual options that offer different travel times.

Third, I introduce a structural model where individuals choose when to commute, in order to conduct more realistic counterfactuals. The choice depends on a set of average departure time preferences parameters estimated as time-of-day fixed effects, and the main source of heterogeneity is the individual VOT. The model allows drivers to choose to travel during less-preferred times of the day if they expect more traffic during their more-preferred times. Counterfactual policies change expected travel times and the distribution of welfare effects is determined by the individual VOT, assuming that preference parameters and the VOT distribution remain stable. The model matches aggregate traffic in both the EL and the standard lanes, and the RD estimates of VOT without targeting them directly. The model predicts that EL use is strongly positively correlated with individual VOT.

The model results provide the basis for the fundamental counterfactual result: ELs are welfare-reducing. The benefits of EL users are outweighed by the extra congestion costs incurred by all the other drivers. In fact, reconverting the EL into a free lane, drivers' per-capita welfare increases by \$25.68 per year, more than what 52% of drivers pay for the EL over an entire year. Gains are concentrated among low-VOT drivers, who now travel at their more preferred times of the day because they expect less congestion. Even allowing for EL toll revenues to be rebated, reconverting the EL still results in a per-capita welfare increase of \$5.02.[11] In a second counterfactual, I make the composition of drivers more unequal by increasing the share of low-VOT individuals. Drivers' per-capita welfare increases by \$2.00 per year, but the policy is regressive: high-VOT drivers now face less competition for using the EL and reap the benefits. For similar reasons, in a third counterfactual, reducing the EL toll level increases drivers' welfare. Finally, assuming that all drivers have VOT equal to the mean results in a misallocation of drivers that reduces per-driver welfare by \$27.50 per year. The counterfactuals imply that the EL is welfare-improving when only a small group of high-VOT drivers use the EL, and all other drivers have very low VOT.[12] There can only be an integer number of ELs out of a small total number of lanes, which limits the ability to design a policy that targets the appropriate share of high-VOT drivers to produce a Pareto improvement.

This paper makes several contributions to the literature. First, I estimate the full distribution of VOT of a population of drivers, using a revealed preference approach. This distribution is representative of a general population who owns and commutes by car, which accounts for

[11] ELs across the US, including those in this paper, typically do not rebate toll revenues to drivers.

[12] This has some parallels with the phenomenon of "elite capture" studied in the development economics.

over 75% of all commuting trips in the US.[13] These results complement and expand on a large body of literature that has estimated the mean VOT, or expressed heterogeneity only in terms of different VOT means for specific subsamples. Among these papers are, especially, Small et al. (2005), Goldszmidt et al. (2020), Buchholz et al. (2022) and Kreindler (2022).[14]

Second, the finding that ELs are welfare-reducing is an important contribution to the literature about congestion pricing and to a public debate on a popular congestion policy across the US. This paper also sheds light on the size of the loss that stems from ignoring the VOT distribution, and on the distribution of the welfare effects of ELs. Because the model abstracts from other commuting modes and from long-term outcomes such as residential choice, the welfare results refer to the short-medium term.[15] This paper builds on a large body of literature that has established the theoretical framework and predictions for welfare analyses of traffic policies. Among these papers are Vickrey (1969), Arnott et al. (1993), Arnott et al. (1994), Braid (1996), van den Berg and Verhoef (2011) and Hall (2018).[16]

Third, this paper estimates the share of very-high-VOT commuters, who might not be well-described by the mean. This is essential for the welfare implications of congestion policies, because policy-makers need to know how many high-VOT agents they can target and how much revenue they can generate. High-VOT individuals are also relevant to models of job location search and urban amenities, where commuting is one of a sequence of choices and individuals' cost of time is a main source of heterogeneity.[17] The wide VOT heterogeneity implies that commuters who greatly value reductions in travel time are a small minority.

The remainder of the paper is organized as follows. Section 2 describes the institutional setting and the dataset used. Section 3 formalizes the identification strategy and presents the RD results. Section 4 presents the estimation of the VOT distribution, section 5 presents the structural model and section 6 the counterfactuals. Section 7 concludes.

[13] Even though demographics are not observed in the data, I can match EL drivers to zipcode-level Census data on the basis of where they enter the EL. A back-of-the-envelope calculation suggests that my RD VOT estimates are positively correlated with individual wage and with measures of extreme wealth.

[14] Other papers in this strand are Moses et al. (1963), Deacon and Sonstelie (1985), Chui and McFarland (1987), Small et al. (2006), Hall (2020) and Bento et al. (2020). For a review, see Hensher (2011). The substitution between time and some other good had been studied, for instance, by Gronau (1973), Aguiar and Hurst (2007), Miller and Urdinola (2010), Aguiar et al. (2013) and Nevo and Wong (2019).

[15] Duranton and Turner (2011) suggest that, for instance, in the long run the welfare gain of reconverting the EL into a standard lane would be dissipated by the entry of new drivers, who increase congestion.

[16] A review can be found in Small and Verhoef (2007). Some of these models rely on the existence of hypercongestion, which occurs when increasing demand for travel on a road reduces not just travel speed but also road capacity. My model does not rely on hypercongestion, a notion challenged by Anderson and Davis (2020). Other papers study the effects of congestion policies on voters behavior (De Borger and Proost (2012)), air pollution externality (Bento et al. (2014) and Gibson and Carnovale (2015)), accident externality (Green et al. (2016) and Romem and Shurtz (2016)), and optimal congestion charges (Yang et al. (2020)).

[17] Among these papers are McFadden (1974), Lucas and Rossi–Hansberg (2002), Graham (2007), Albouy and Lue (2015), Manning and Petrongolo (2017), Monte et al. (2018), and Le Barbanchon et al. (2020).

Figure 1: *Snapshot of highway I-394, Eastbound direction, just before the exit to General Mills Boulevard. The Express Lane is the leftmost lane, separated by the rest of the highway by a double white line on the ground. Source: Buckeye (2012).*

2 Institutional setting and descriptive analysis

In this section I first describe the institutional setting of ELs in this paper and especially the tolling algorithm, which is the key feature of these ELs and provides my source of identification. Then I describe the dataset constructed for this paper, which provides the richness of variation needed to estimate the VOT distribution; I show the main descriptive statistics of the data and explain how the measurement of travel time saved is constructed.

2.1 Description of Minnesota Express Lanes and tolling algorithm

This project uses data from 3 different Express Lanes operated by MnPASS and the Minnesota DOT in the area of Minneapolis-Saint Paul. In particular, the lanes are on I-394 west of Minneapolis, on I-35W south of Minneapolis and on I-35E north of Saint-Paul.[18] Each of these highways has one Express Lane for each direction: one goes into the city, which is typically tolled on weekdays during the morning peak between 6am and 10am and free at other times, and one comes out of the city, which is typically tolled on weekdays during the afternoon peak between 3pm and 7pm and free at other times.[19] The EL on I-394 is about

[18] A map of the Minneapolis-Saint Paul area that includes all three lanes is provided in the Appendix.

[19] In addition to these two peaks, some EL portions are also tolled at other times. In particular, on the I-394 the section between the Highway 100 and the I-94 intersections there is a reversible EL, which is tolled at all times of the day: it runs eastbound into Minneapolis from 6am to 1pm and westbound out of Minneapolis from 2pm to 5am. On I-35W, the EL sections closest to Minneapolis are tolled during both peaks: the

9 miles long, the EL on I-35W is about 16 miles long and the EL on I-35E is about 11 miles long. An example of what an Express Lane looks like is in *Figure 1.* The leftmost lane is the EL, and the sign displays the tolls needed to get to each intersection. For the majority of the EL length, the EL is separated from the rest of the highway by a double white line but there is no physical barrier. All ELs in both directions are divided into two sections (except I-35E southbound, which only has one section) that end in a major intersection, and each toll buys the right to use each section in its entirety.[20]

To use the ELs, drivers must place a transponder in their car.[21] Detectors placed at each EL entrance read the transponders and charge the appropriate toll. Carpools and public transport can use the ELs for free at all times, and transponders have a switch that, when turned on, signals to the detectors that the driver is carpooling. When on the highway, a few hundred feet before entering the EL, drivers see a first sign that lists the tolls and then a second sign at the entry point, where the detector is located.

The key feature of the MnPASS Express Lanes is that the toll changes dynamically every 3 minutes during peak times in response to traffic density on the EL. Density is measured as the number of vehicles per mile of road. The tolling algorithm works as follows.

First, all traffic sensors placed along the EL record a measurement of traffic density every 30 seconds. Then, each sensor computes the average density over the past 6 minutes. If there is more than one sensor in each downstream EL portion, only the highest average density among all those sensors is retained. The value obtained is then fed into the continuous function:

$$p = \alpha \cdot d^{\beta} \tag{1}$$

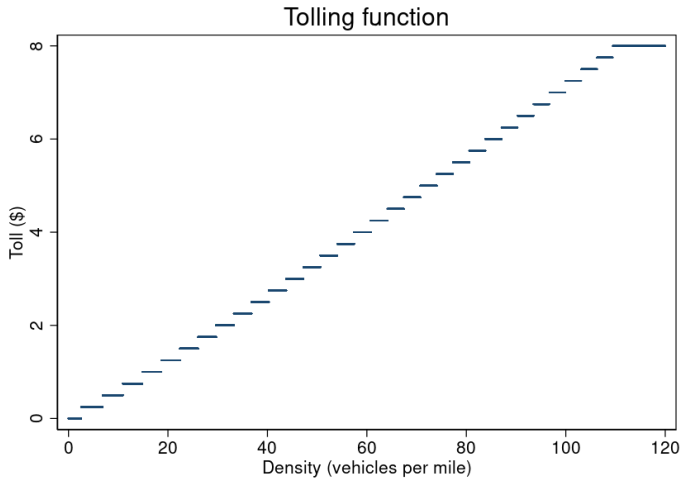
where p is the toll, d is the average density value computed in the aforementioned way, $\alpha = 0.045$, and $\beta = 1.1$. Finally, the toll that results from this function is rounded to the nearest \$0.25 and is capped at \$8. This final step in the tolling algorithm creates 32 discontinuities, where average traffic density on the EL during the past 6 minutes varies smoothly and the toll jumps discretely by \$0.25 at each discontinuity. Hence, in practice, the observed toll can be represented by the step function in *Figure 2.* At each discontinuity cutoff, the EL becomes relatively more expensive, which decreases aggregate demand for the EL and produces an increase in time saved by taking the EL.

northbound portion between the Highway 62 and the 26th Street intersections is also tolled from 3pm to 7pm; the southbound portion between the 42nd Street and the I-494 intersections is also tolled from 6am to 10am.

[20] Detailed maps of the ELs are provided in *Figures A.2, A.3 and A.4* in the Appendix.

[21] The transponder can also be used for electronic toll payments on other highways. Recently, the MnPASS became fully compatible with the E-ZPass multi-state network.

Figure 2: *Plot of the the toll (on the vertical axis) as a function of traffic density on the EL as measured by the tolling function (on the horizontal axis, from 0 to 120).*



2.2 Data, measurement of time saved, and descriptive evidence

The dataset I use is a panel of EL users matched with very fine traffic information contemporaneous to individuals' observed repeated EL choices. This detailed level of information, together with the richness of variation in toll levels and time saved in this EL setting, is a unique feature of my data relative to previous literature, and it is especially suitable for estimating VOT distribution.

The panel dimension of the data is a non-publicly available dataset of about 50,000 distinct MnPASS[22] EL users from March 2017 to April 2018, which covers 228 days when the ELs were tolled.[23] Drivers in the data are identified by their transponder tag number; thus, multiple EL uses by the same tag can be tracked over time. The dataset includes entry and exit locations, the date of entry and exit and their times precise to the second, plus the toll paid.[24] Moreover, for each EL use there is an indicator variable for whether the carpool switch was turned on. Using Google Maps together with the maps of each EL, I computed the length of each road segment between entries and added it as a variable to the dataset.

I matched the panel dataset with a record of contemporaneous aggregate traffic measurements taken at every sensor placed along the ELs and the corresponding highway lanes.[25]

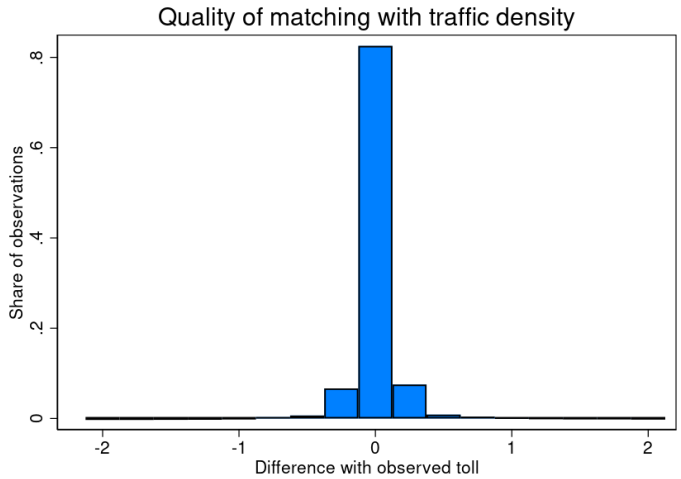
[22]MnPASS is the branch of the Minnesota Department of Transportation that managed the Express Lanes in 2018. It is currently known as E-ZPass Minnesota.

[23]50,000 users amounts to about 1.5% of the total population of the Minneapolis St. Paul metro area and about 5% of the population that lives in areas where drivers could realistically take the ELs to reach the city center.

[24]The exit time is recorded as the moment when the driver is last observed going through an EL detector. As a convention in my analysis, I assume that when a driver goes through a detector marked as her exit, she uses that segment until just before the next detector.

[25]This information can be downloaded from the website of the Minnesota DOT.

Figure 3: *Quality of match between traffic data and toll data. The x-axis is the difference between the toll implied by the matched traffic density and the observed toll; thus, 0 represents perfect matches.*



Each sensor record includes measurements of speed, traffic density and traffic volume, which is defined as the number of vehicles that go through a point on the road in one hour.[26] I then assembled the sensor records one by one and matched them with their correct location placement on the ELs. Importantly, this dataset is extremely precise, because all measurements are taken every 30 seconds, and it is complete because it includes both the ELs and all the standard free lanes.

To check match quality after the merge, I compute the toll that would be implied by the traffic density measured by the sensors and calculate the difference between it and the observed toll from the EL panel. As *Figure 3* shows, over 82% of the observations are matched to the exact density level because the difference is 0. Most of the remaining observations are within \$0.25 of the observed toll. The analysis that follows restricts attention to only the exact matches, but additional evidence provided in the Appendix confirms that results are robust to including imperfect matches.

As mentioned in the introduction, drivers in this dataset are only observed when they use the ELs. When they are not observed, they could be driving on the standard free lanes parallel to the ELs, they could be driving somewhere else, or they could not be driving at all. The next section of the paper describes how the identification strategy, under a set of mild assumptions, exploits this feature of the data and explains how to take it into account when interpreting the results.

In terms of aggregate trends, however, other data sources show that, even when drivers are not observed on the EL, they are likely to be driving on the same highway in the standard

[26] As a standard practice in traffic measurements, the definition of traffic volume is readily extended to time intervals different from one hour by linear proportion.

Table 1: *Summary table of the total number of observations in the sample, divided by road, time of day, and day of the week. There are three categories: all observations in the sample, only peak-time observations, and only peak-time excluding carpools. Both the absolute number and the percentage on the total are shown for each category.*

	All		Peak time		Peak time, no carpools	
	N	%	N	%	N	%
I-394 (morning)	1,103,391	22.52	806,136	25.18	653,949	26.30
I-394 (afternoon)	977,068	19.94	700,756	21.89	565,985	22.76
I-35W (morning)	1,159,826	23.67	691,468	21.60	513,616	20.65
I-35W (afternoon)	818,577	16.70	398,545	12.45	283,696	11.41
I-35E (morning)	409,658	8.36	303,409	9.48	238,196	9.58
I-35E (afternoon)	432,045	8.82	301,543	9.42	231,299	9.30
Monday	822,759	18.22	597,789	18.67	461,599	18.56
Tuesday	958,817	21.23	698,462	21.81	544,578	21.90
Wednesday	989,935	21.93	712,541	22.25	556,586	22.38
Thursday	960,678	21.27	681,541	21.29	532,201	21.40
Friday	784,229	17.36	511,524	15.98	391,777	15.75
Total	4,900,565		3,201,857		2,486,741	

lanes. This is so for three reasons. First, according to a customer survey carried out by MnPASS, which manages the ELs in this paper, 93% of drivers use the ELs for their work commute and 90% of drivers travel on the free lanes when they are not using the EL. Second, the ELs connect the suburbs of Minneapolis and Saint Paul with the city center, and checking travel times on Google Maps suggests that it is not plausible to achieve shorter travel times by taking alternative routes to the highways examined here. Third, there is a minor presence of alternative commuting modes and there has been no significant shift to these alternatives over time. In fact, between 2010 and 2018 over 70% of Minneapolis commuters traveled by car and only 11% used public transit.[27] Moreover, during the same period, the shifts between alternative commuting modes have all been within 2 percentage points.

The dataset serves the following three purposes. First, I use the traffic density measured on the EL to construct the running variable for the RD design, using the discontinuities created by the tolling function. Second, I use the information on traffic on the free lanes to construct a measure of how much time EL users are saving by taking the EL at each time. Third, although this is an aggregate dataset, the fact that measurements are taken every 30 seconds allows me to track traffic flows between the ELs and the free lanes quite precisely. Thus, I estimate the model so that the choice by agents to not use the ELs, even if these choices are not individually observed in the data, still match the precise aggregate traffic levels observed in the free lanes at different times of the day.

[27]Source: American Community Survey (2010 and 2018, 5-year estimates). Furthermore, 10% of commuters walked or biked and 6% worked from home. These workers might not be relevant for the analysis since their commutes are likely shorter than those taken by individuals in the dataset.

Time saved is computed as follows. In the EL panel I compute travel time as the difference between exit time and entry time, both of which I observe measured at the seconds level. For the general free lanes, I observe the aggregate traffic speed in the time interval that is contemporaneous to each driver's EL travel. I use this aggregate speed to infer what the travel time would have been on the general lane for each EL travel I observe in the panel. The difference between the two travel times, measured in minutes, gives the time saved, which I use as dependent variable in the reduced-form analysis.

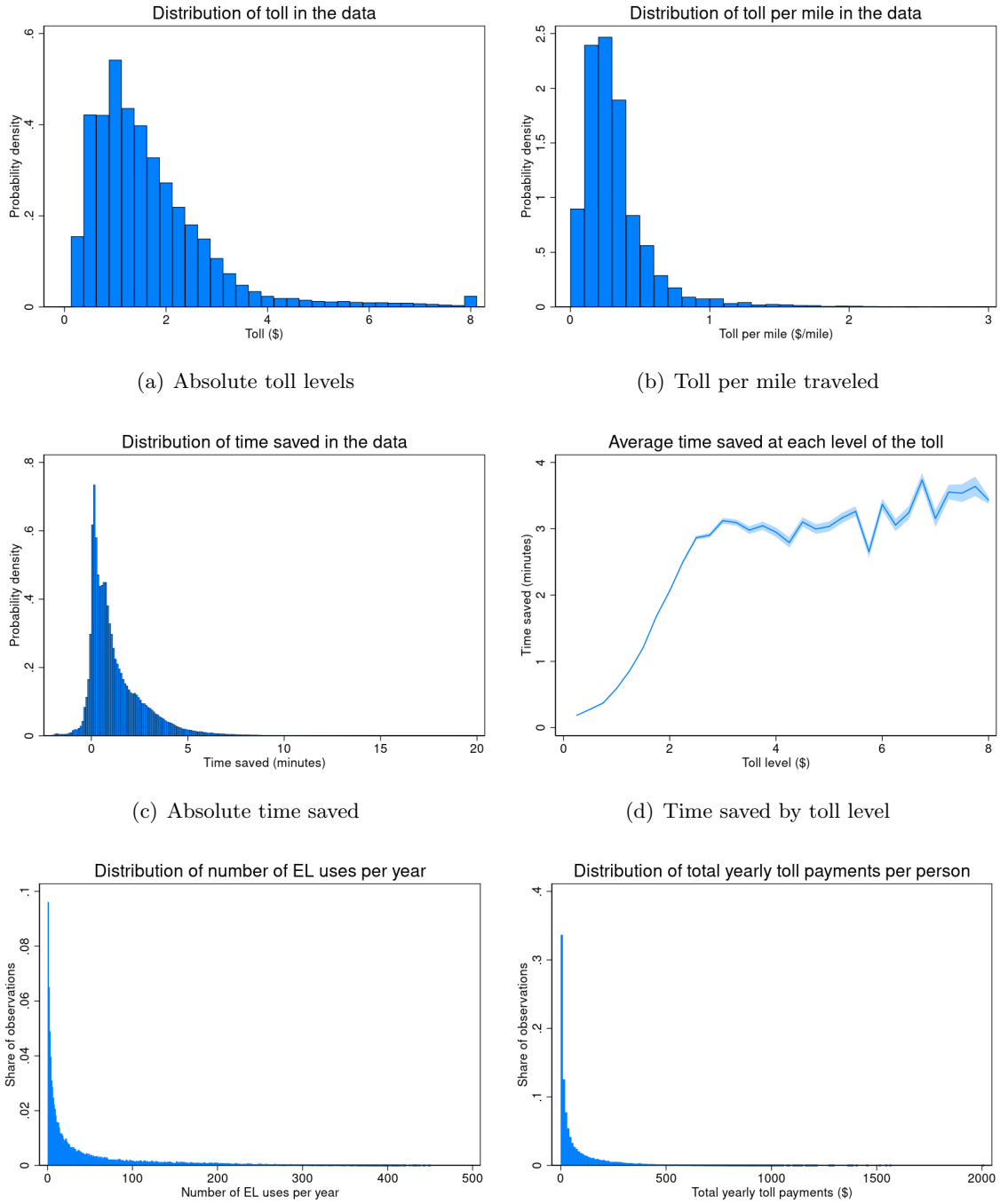
Table 1 provides an overview of the number of observations in the dataset. The EL panel includes repeated observations of 51,702 distinct drivers, which decrease slightly to 47,722 if I exclude off-peak observations and to 45,421 if I also exclude carpools. The analysis focuses on the 2,486,741 observations of single-driver vehicles during peak time, which correspond to just over half of all the observations in the sample. Road I-394 accounts for about 50% of the paper's subsample, road I-35W for about 30%, and road I-35E for the remaining 20%. In terms of days of the week, it seems that EL usage is stable from Tuesday to Thursday, and is about 3 percentage points lower on Monday and 6 percentage points lower on Friday.

Importantly, about 96% of drivers are always observed using the same highway, in the morning, the afternoon or both. For this reason, in both the reduced-form analysis and the model it is possible to consider each highway as a separate universe. Moreover, about 61% of drivers always enter the EL at the same entry location. While I will use entry location to study heterogeneity in the reduced-form results, I will abstract from this choice margin in the structural model.

Figure 4 shows the distribution of observations in the sample under relevant dimensions. Panel (a) shows the absolute toll levels, whereas Panel (b) shows the distribution of toll per mile traveled. To provide intuition of how traffic relates to tolls, a \$1.25 toll corresponds to a density of 20 vehicles per mile of road on the EL and a speed of 64.5 miles per hour, a \$3.25 toll to a density of 50 and a speed of 58mph, a \$7.25 toll to a density of 100 and a speed of 48mph. Most of the observations in the data are for tolls below \$4, but there is a long right tail and some minimal bunching at \$8. The majority of the observations are within \$1 per mile traveled, which is consistent with the fact that the average EL trip is about 6 miles long. Panel (c) shows the distribution of absolute time saved in the data: most observations are below 5 minutes, but there is a long right tail up to over 20 minutes, and about 5% of observations have negative time saved.[28] Panel (d) shows the mean time saved observed at each level of the toll, with a shaded area representing the 95% confidence interval. Mean time saved increases from 0 to 3 minutes for a toll level up to \$2.50; after that, time saved hits a plateau and remains slightly above 3 minutes for all other levels of the toll. Panel (e) shows the distribution of the number of yearly EL uses per driver and Panel (f) the distribution of yearly toll payments per driver. The median driver uses the EL 7 times and pays only \$23.50

[28] As will be explained in the reduced-form section, this is possible for a number of reasons.

Figure 4: *Distributions of observations in the data. Panel (a) shows the distribution of absolute toll levels observed in the data. Panel (b) shows the distribution of toll per mile traveled observed in the data. Panel (c) shows the distribution of absolute time saved in the data. Panel (d) shows the average time saved at each level of the toll, with a shaded area representing the 95% confidence interval. Panel (e) shows the number of yearly EL uses per driver. Panel (f) shows the yearly toll payments per driver.*



(a) Absolute toll levels

(b) Toll per mile traveled

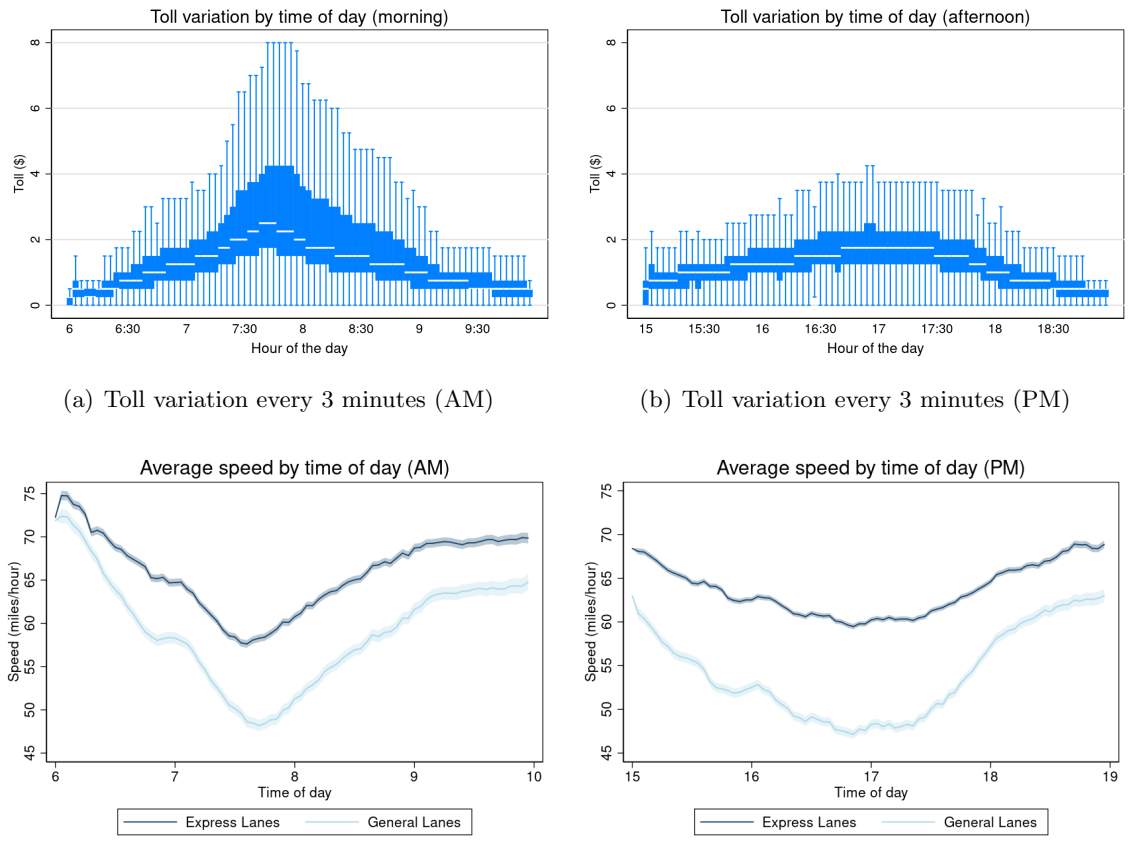
(c) Absolute time saved

(d) Time saved by toll level

(e) Yearly EL uses per driver

(f) Yearly toll payments per driver

Figure 5: *Variation of toll and average speed by time of day. Boxplots of the observed toll are shown for every 3-minute interval of the morning peak in Panel (a) and of the afternoon peak in Panel (b). Similarly, the average speed on the EL (dark blue) and on the general lanes (light blue) are shown in every 3-minute interval of the morning peak in Panel (c) and of the afternoon peak in Panel (d). The shaded areas in (c) and (d) represent the 95% confidence intervals.*



(a) Toll variation every 3 minutes (AM)

(b) Toll variation every 3 minutes (PM)

(c) Average speeds every 3 minutes (AM)

(d) Average speeds every 3 minutes (PM)

per year, whereas the top 1% of drivers uses 328 times and pays over $668.75 per year.

The plots in *Figure 5* display the wide variation of observations in the dataset. The two top panels show boxplots[29] of the observed toll in each 3-minute interval of the morning peak (a) and the afternoon peak (b). There is variation at all times but significantly more during the central hours of each peak, especially in the morning. The two bottom panels show the average speed in each 3-minute interval of the morning peak (c) and the afternoon peak (d), with shaded areas representing the 95% confidence intervals. The speed on the EL

[29]In these boxplots, the boxes cover from the 25th to the 75th percentile. The two extremes below and above are the lower and upper adjacent values, respectively. The lower adjacent value is defined as the smallest observations that is at least as large as the 25th percentile minus 1.5 times the interquartile distance. The upper adjacent value is defined symmetrically using the 75th percentile rather than the 25th.

is consistently higher than in the standard lanes, particularly so in the central hours of each peak. Since travel time is a function of the inverse of speed, the same difference between EL speed and general lane speed implies higher time savings at lower initial speed levels. To provide intuition, on a 10-mile road starting at 60 mph, going 5 mph faster would save about 46 seconds; at 40 mph, it would save about 1 minute and 40 seconds; at 20 mph faster, it would save 6 minutes.

The following sections explain how the rich variation in the data can be exploited to estimated first the VOT of EL drivers and, next, the VOT distribution for the full population of drivers.

3 Step I: Mean VOT estimates among EL users

This section shows the reduced-form evidence about the VOT of EL users. First, I introduce a theoretical framework that connects the data with the reduced-form quantities identified by the RD research design. Second, I provide a graphical intuition for the RD and show the formal regression equations. Finally, I show the RD results, including a comparison with a standard hedonic OLS and heterogeneity by time of day and location of EL entry.

3.1 Framework, identification and intuition

Some works in the literature have estimated VOT as the coefficient on time saved in a hedonic regression of EL toll on time saved.[30] An OLS regression of this type, however, suffers from endogeneity for a number of reasons, the principal one of which is that unobserved traffic and driving conditions and unobserved individual factors, like being in a hurry, are correlated with both time saved and toll paid.[31]

Instead, I isolate variation in tolls that is plausibly exogenous to traffic conditions and arguably unpredictable by drivers. In fact, as described above, the MnPASS tolling algorithm sets the toll on the basis of a continuous function of past traffic density in the EL, and then rounds it to the closest \$0.25 up to a \$8 cap, thus creating 32 discontinuities. At each discontinuity, traffic density varies continuously, whereas the toll jumps up discretely by \$0.25. Thus, a Regression Discontinuity Design could potentially estimate the corresponding reduced-form change in time saved at the cutoff and estimate the VOT as the ratio between \$0.25 and the time saved change.

I now introduce a theoretical framework that maps the individual agent's economic problem into the observed data. The framework highlights what the RDD identifies and guides

[30] See, in particular, Bento et al. (2020).

[31] Greenstone (2017) points out that omitted variable bias is common in estimations of hedonic price schedules in many settings.

the interpretation of results in a way that is consistent with the data. For simplicity, I assume in this example there is only one cutoff, where the toll increases discretely by \$0.25.

Suppose that individuals first face a choice between commuting by car on the highway and an outside option that includes both commuting by other means and not driving at all.[32] Conditional on commuting by car, a driver i chooses to use the EL if their expected utility u^{EL} is positive:

$$u_{it}^{EL} = \delta^{EL} + \beta_i^{VOT} \cdot \mathbb{E}[\tau_{it}|\Psi_{it}] - \pi_{it} + \varepsilon_{it}$$

where δ^{EL} is a general taste for the EL, τ_{it} is the time saved by taking the EL, π_{it} is the toll paid, and ε_{it} is an error term with cdf G. The expectation of time saved is taken on the basis of an information set Ψ_{it}, which can include surrounding traffic, the toll itself, other covariates, and unobservables. The individual EL choice probability is then:

$$P_{it}^{EL} = 1 - G(-\delta^{EL} - \beta_i^{VOT} \cdot \mathbb{E}[\tau_{it}|\Psi_{it}] + \pi_{it}) \tag{2}$$

Consequently, the aggregate demand for the EL, or the EL traffic share, is equal to $\frac{1}{I}\sum_i P_{it}^{EL}$, where I is the total number of drivers on the highway. This follows from the fact that, once drivers are on the highway, the EL and the standard lanes are mutually exclusive and exhaustive choices.

I now make two simple assumptions that allow me to bring this problem to the data.

Assumption 3.1: Rational Expectations. $\mathbb{E}[\tau_{it}|\Psi_{it}] = \tau_{it} + \nu_{it}, \quad \mathbb{E}[\nu_{it}] = 0$

The first assumption imposes that individuals have rational expectations about the time they could save by taking the EL. This means that individuals' expectation is equal to the ex-post realization of travel time savings, up to an error term ν_{it} with mean 0. Since most of the individuals in the dataset are commuters, it seems reasonable to assume that they are able to predict the duration of their commute accurately. This assumption ensures that the ex-post measurement of time saved observed in the data is a correct measurement of the expectation of time saved that individuals use in their choice process.

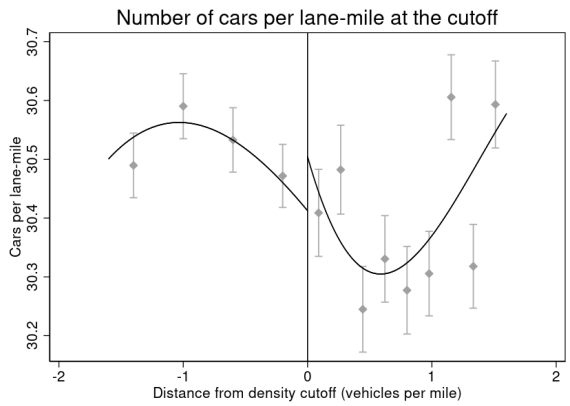
Assumption 3.2: Smoothness. $\mathbb{E}[\tau_{it}^0|R = r]$ *is continuous at* $r = c$

Here, τ^0 is the time saved if the toll did not change, R is the running variable and c the cutoff level. This is a re-statement of the standard RDD assumption that the untreated potential outcome is continuous at the cutoff: at each cutoff, individuals' expectation of time saved absent any change in the toll is smooth.

To see how the smoothness assumption applies to this context, I consider how time saved is determined. In general, time saved is a function of the relative speed on the EL and the

[32] As mentioned above, at the aggregate level both the aggregate amount of commuters and the share that chose each mode have been roughly constant from 2010 to 2018 in Minneapolis, with over 70% of commuters driving to work.

Figure 6: *Plot that shows how the total amount of cars on the highway, on all lanes including the EL, does not change at the cutoff, supporting the smoothness assumption. The running variable, in vehicles per mile, is the measurement of traffic density fed to the tolling function, and is re-scaled as distance from the discontinuity cutoff. The outcome variable is the number of cars per lane-mile. The regression controls flexibly for the running variable on each side of the cutoff using a third-degree polynomial. Standard errors on each bin follow Calonico et al. (2015).*



general lanes, which are, in turn, a function of the amount of traffic on each lane type. Thus, to simplify, time saved is equal to:

$$\tau_{it} = TT_{it}^{GL} - TT_{it}^{EL} = \varphi\left(I - \sum_{i=1}^{I} P_{it}^{EL}\right) - \varphi\left(\sum_{i=1}^{I} P_{it}^{EL}\right) \tag{3}$$

where TT^{GL} and TT^{EL} stand for travel time on the general lanes and on the EL, respectively, and $\varphi(\cdot)$ is the increasing function that links traffic on a lane to travel time on that lane, and P_{it}^{EL} is expressed in (2).

The smoothness assumption imposes that all the observable and unobservable variables that determine time saved, including those that enter the information set Ψ_{it}, are smooth at the cutoff. *Figure B.1* shows that observables covariates do not change at the cutoff, supporting the usual RDD argument that unobservables also change smoothly at the cutoff. Moreover, I can also check in the data that the total number I of drivers on the highway is smooth at the cutoff. *Figure 6* shows that the estimated change in I is a non-significant 0.079 cars per lane-mile. This implies that it is not necessary to observe the outside option, because the rate of never-takers and non-drivers does not change at the cutoff.

The smoothness assumption has the key implication that, locally at the cutoff, the \$0.25 toll increase perfectly predicts the decrease in aggregate demand for the EL. The rationality assumption implies that, consequently, individuals correctly expect the increase the time saved

by taking the EL.[33] Calling ΔP the change in aggregate EL demand, it follows from (3) that:

$$\tau_{it} = \underbrace{\varphi(\overbrace{I - \sum_{i=1}^{I} P_{it}^{EL}}^{\text{increases by } \Delta P}) - \varphi(\overbrace{\sum_{i=1}^{I} P_{it}^{EL}}^{\text{decreases by } \Delta P})}_{\text{increases by } \Delta\tau \text{ to be estimated}}$$

In this way, it is not necessary to observe individual demand, because the \$0.25 toll increase is directly connected to the time saved increase $\Delta\tau$. Thus, the RDD correctly identifies:

$$\mathbb{E}[\tau_{it}^{1} - \tau_{it}^{0} | R = c] = \Delta\tau$$

which represents the increase in time savings that drivers who stay on the EL are willing to accept in exchange for a \$0.25 toll increase.

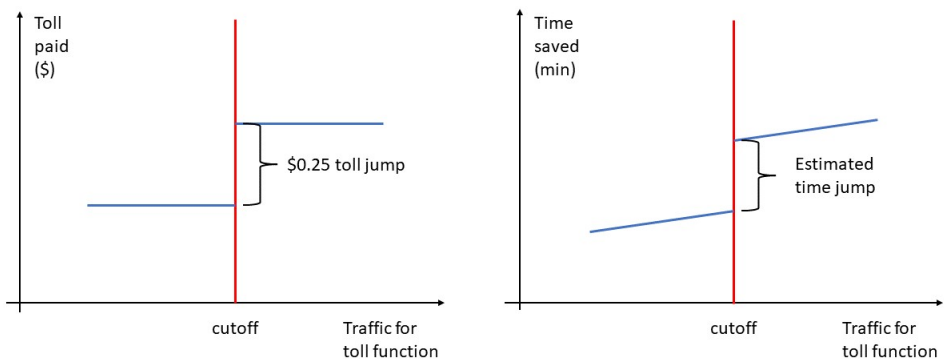
It is important to underline that all the assumptions need to be satisfied only at each cutoff. For instance, consider driver A who faces a \$1 toll and driver B who faces a \$7 toll. The RDD does not compare these two drivers and, consistently, the assumptions allow them to be observably and unobservably different. Instead, consider now driver C located virtually at the same level of the running variable as driver A but facing toll \$1.25. Thanks to the assumptions, these two drivers are assumed to be equal in terms of all unobservables and observables, with the exception of the toll. Driver C expects larger time saved from taking the EL just because the toll is higher, but otherwise does not use the toll any differently than driver A. The theoretical framework and the assumptions allow the RDD to correctly compare driver A to driver C.

Figure 7 provides a graphical intuition of the RDD estimation strategy. Panel (a) shows the \$0.25 toll increase at the cutoff and Panel (b) the corresponding time saved increase that can be estimated. The ratio between \$0.25 and the time saved increase gives an estimation of the VOT. The effect is estimated off drivers who use the EL; thus, the VOT is intended to be the value of travel time saved conditional on using the EL.[34] With this framework and interpretation in mind, I now introduce the formal regression strategy.

[33] In other words, the assumptions imply that, even without specifying the functional form of individuals' expectation of time saved, locally at the cutoff the only information that the toll increase conveys to individuals is that time saved changes because aggregate demand for the EL changes.

[34] In principle, these estimates should be lower-bound because drivers with very high VOT might be willing to accept even smaller travel time savings than the ones that appear in the data. However, Panel (c) in *Figure 4* shows that there are observations of very small and even negative time saved. Consequently, the lower-bound is only an issue if some high-VOT drivers are not faced with an EL choice when time saved is close to 0.

Figure 7: *Intuition for the RDD design that takes into account the theoretical framework and assumptions. Panel (a) shows the \$0.25 toll increase at the cutoff, Panel (b) the corresponding time saved increase that can be estimated. The ratio between \$0.25 and the time saved increase gives an estimation of the VOT saved conditional on using the EL.*



3.2 Estimation strategy

Intuitively, the estimation strategy entails repeating the exercise in *Figure 7* across all cutoffs. More formally, I run a regression of time saved on the toll paid, where I instrument the toll with 32 indicator variables, each equal to 1 when traffic density is above its relevant threshold, which triggers the toll jump. For individual i, at time t:

$$\text{FS:} \quad \pi_{it} = \alpha_z^{FS} + \beta_z^{FS} \cdot \mathbb{1}[d_{it} \geq c_z] + \gamma_z^{FS} \cdot f(d_{it} - c_z) + \delta_z^{FS} \cdot \mathbb{1}[d_{it} \geq c_z] \cdot f(d_{it} - c_z) + \eta_{it}$$

$$\text{SS:} \quad \tau_{it} = \alpha_z + \beta_z \pi_{it} + \gamma_z \cdot f(d_{it} - c_z) + \delta_z \cdot \mathbb{1}[d_{it} \geq c_z] \cdot f(d_{it} - c_z) + \nu_{it}$$

where π_{it} is the toll paid, τ_{it} is minutes of time saved, d_{it} is traffic density, c_z is the cutoff for discontinuity z and $f(\cdot)$ is a third-degree polynomial. This design is a fuzzy RD of time saved on the running variable of traffic density. The first line can be thought of as a first-stage RDD regression, in which the instruments are the indicators $\mathbb{1}[d_{it} \geq c_z]$. The second line is equivalent to second-stage regression of time saved on the instrumented toll, which keeps the same flexible controls on each side of the cutoff.[35]

A few points are worth mentioning about the appeal of the RDD in this setting. First, traffic conditions around each discontinuity are held fixed to the extent that traffic density over the past 6 minutes does not imply different present traffic conditions on each side of the discontinuity cutoffs. Second, the change in toll is arguably unpredictable to drivers because it depends on the average EL density over the past 6 minutes being above a certain threshold and measured in a number of different locations. It is impossible for drivers, who are moving,

[35] Alternatively, this design can be viewed as a 2SLS design where I force the comparisons to happen at each cutoff and where I flexibly control for traffic density around each cutoff. The indicators $\mathbb{1}[d_{it} \geq c_z]$ satisfy all the standard IV assumptions.

to observe all measuring locations simultaneously for 6 minutes. Third, the tolling function is such that cutoffs are about 4 density units (i.e., 4 vehicles per mile) apart from each other, which means the windows for any prediction are small in terms of density.

The RD design recovers the β_z effects on time saved for the \$0.25 toll increase at each cutoff. Since in principle there are 32 instruments and each cutoff has a different number of observations, the average treatment effect will be a weighted average of the β_z effects. I call this average $\tilde{\beta}$ and choose weights following Bertanha (2020).[36] Hence, the estimated mean VOT is given by:

$$VOT = \frac{1}{\tilde{\beta}}$$

In light of the previous discussion, this is the estimated value of travel time savings conditional on using the EL.

In terms of implementation, this strategy amounts to running 32 stacked regression discontinuity equations with density as the running variable and flexible controls on each side of each discontinuity. The estimation, including choosing the optimal bandwidth and the polynomial degree of $f(\cdot)$, follows Calonico et al. (2014), Calonico et al. (2015) and Calonico et al. (2020).[37] For most of the analysis, I restrict attention to the perfectly matched observations in the dataset, which means that the first-stage effects will mechanically be equal to \$0.25 and the first-stage polynomial coefficients will all be 0.[38] The computation of standard errors follows Bertanha (2020).

Before moving to the results, I perform a number of checks to support the assumptions of the research design. *Figure 6* in the previous section showed that the total number of cars on the road remains constant at the cutoff, which indicates that the share of non-drivers and never-takers is also constant. *Figure B.1* verifies that drivers on each side of the discontinuity cutoffs do not differ in terms of their observables (EL trip length, and EL entry time). Finally, Panel (a) in *Figure 8* shows that there is no bunching of observations on each side of the cutoffs in the distribution of the running variable. Panel (b) shows that the distributions of entry time at the seconds level are uniform on both sides of the cutoffs and almost perfectly overlap across each second of each 3-minute interval.[39]

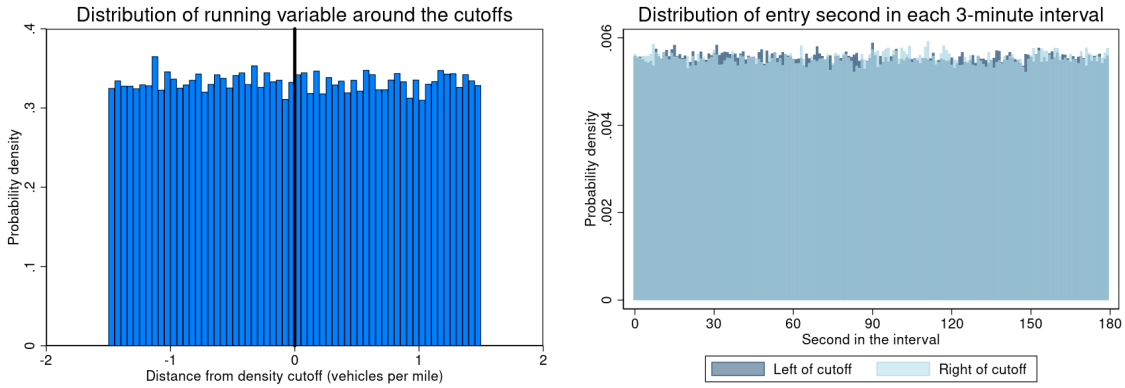
[36] The objective of the technique in Bertanha (2020) is to obtain an average treatment effect that is meaningful to a more general set of individuals than just the ones observed locally at each discontinuity cutoff.

[37] The minimum distance from one cutoff to the next is between the last two and is about 3.15 vehicles per mile. However, the optimal bandwidths are such that there is no overlap of treated and non-treated observations.

[38] I repeat the main analysis using imperfect matches as well as a robustness check in the Appendix.

[39] Remember that the toll changes every 3 minutes.

Figure 8: *Graphical support for the absence of selection in the distribution of observations around the cutoffs. Panel (a) shows that there is no bunching in the distribution of the running variable on each side of the cutoffs. Each 0.05-wide bar shows the probability density of the running variable (traffic density fed to the tolling function) within that 0.05 interval. The time unit of observation is each 3-minute interval, since the value of the running variable is changed every 3 minutes. Panel (b) shows that observations are uniformly distributed on both sides of the cutoffs and almost perfectly overlap across each second of each 3-minute interval.*



(a) No bunching around cutoffs

(b) No selection on entry second

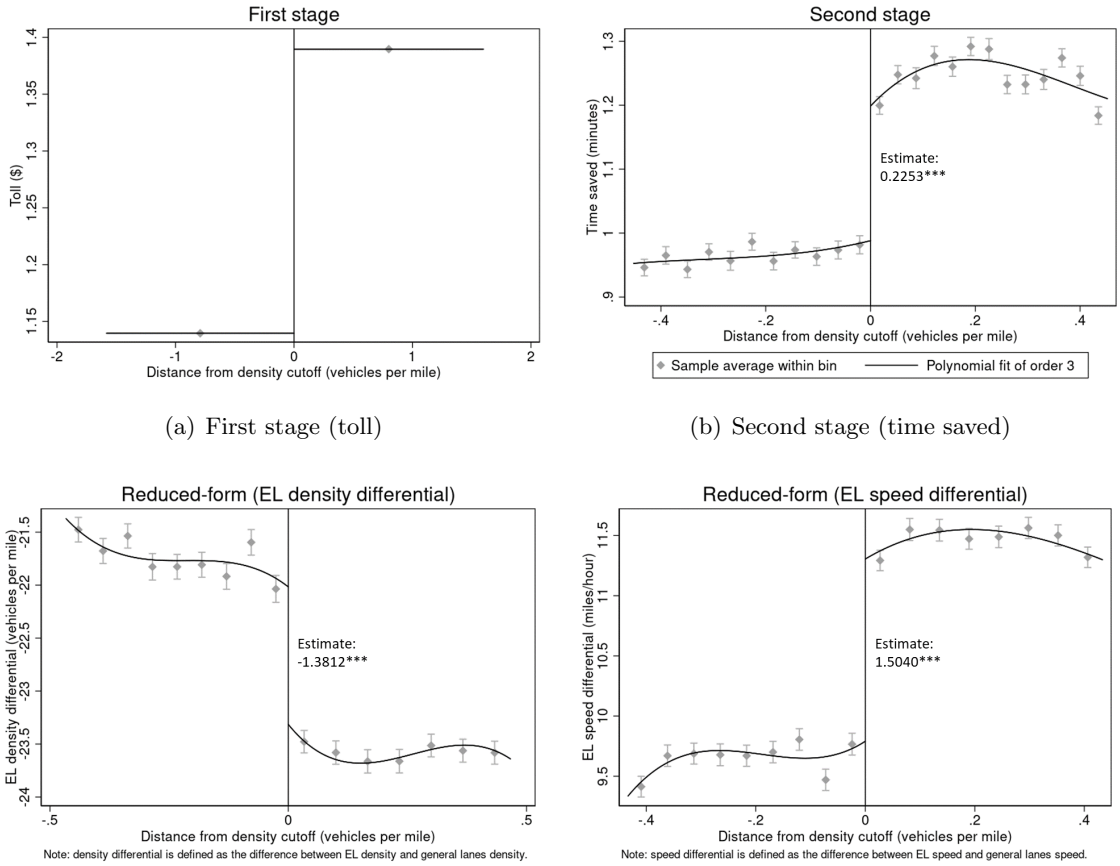
3.3 RD results

This subsection presents the reduced-form results, restricting attention to observations in the panel dataset that are perfectly matched to traffic density. First, I show the general results on time saved and the suggested mechanism for the results. Then, I show a comparison with OLS results to highlight the sign and magnitude of the endogeneity bias relative to the RDD. Finally, I look for heterogeneity in VOT in terms of entry location and time of day, based on whether drivers enter the EL in the first or the second section on each road and during either the morning or afternoon peak.[40] In these exercises I allow all coefficients in each regression to vary on the basis of the heterogeneity margin that is being analyzed.

Figure 9 shows the first-stage and second-stage results. Panel (a) illustrates the estimated jump in the toll at the discontinuity cutoff. Because the sample is restricted to exact matches, the jump is mechanically \$0.25, as expected. Panel (b) shows the second stage result on time saved, where the jump at the discontinuity is triggered by the \$0.25 toll increase. Hence, the ratio between the two jumps provides an estimate of the VOT saved from traffic conditional on choosing to use the EL. In *Appendix Figure B.2*, I repeat the exercise, including imperfect matches, and I show that the second-stage results are robust even if the estimated toll jump is less than \$0.25 because of the attenuation bias, which is due to imperfectly assigned toll-

[40]Heterogeneity with respect to time of day recognizes that this estimated VOT includes the value of not arriving late to one's destination, which might be more salient during the morning commute.

Figure 9: *First-stage (a) and second-stage (b) results of the RD regression for time saved (b), EL density premium (c) and EL speed premium (d). The EL density premium is defined as the difference between the EL density and the general lanes density. The EL speed premium is defined as the difference between the EL speed and the general lanes speed. The first stage shows the $0.25 toll increase triggered by the traffic density discontinuity. The second stage shows the RDD effects on time saved, the EL density premium, and the EL speed premium, which are triggered by the $0.25 toll increase. The data is fit using a third-degree polynomial. The gray whiskers are the 95% confidence intervals around each bin.*



(a) First stage (toll)

(b) Second stage (time saved)

(c) Second stage (density premium)

(d) Second stage (speed premium)

density pairs. In *Appendix Figure B.3*, I show the RD effects separately at each cutoff, which fall around the estimated mean RD effect.

Panels (c) and (d) shed some light on how the $0.25 toll increase triggers an increase in time saved. At the new higher toll to the right of the discontinuity, fewer drivers choose to enter the EL. Consequently, density on the EL decreases relative to the general lanes, while both EL speed and time saved increases. This mechanism is consistent with the theoretical framework, wherein the toll increases causes a shift in aggregate demand from the EL to the

Table 2: *Second-stage results of regression of time saved on traffic density as the running variable and comparison with OLS hedonic regression of the toll paid on time saved. Standard errors follow Bertanha (2020).*

PANEL 1: RDD	**Dependent: time saved (minutes)**			
	All roads	I-394	I-35W	I-35E
Estimated RDD effect	0.225***	0.215***	0.244***	0.254***
	(0.00667)	(0.00741)	(0.0128)	(0.0152)
Implied VOT ($/hour)	66.56	69.92	61.52	59.00
	(1.97)	(2.41)	(3.23)	(3.53)
PANEL 2: OLS	**Dependent: toll paid ($)**			
	All roads	I-394	I-35W	I-35E
Time saved (minutes)	0.281***	0.211***	0.331***	0.143***
	(0.00162)	(0.00121)	(0.00176)	(0.00159)
Implied VOT ($/hour)	16.83	12.66	19.85	8.58
	(0.10)	(0.07)	(0.11)	(0.09)
N	1,935,965	956,530	642,886	337,226

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

general lanes.

Table 2 shows the second-stage general results for all roads jointly, as does *Figure 9*, and for each road. As mentioned above, at this stage time saved (in minutes) is regressed on traffic density as the running variable, as in a sharp RD. The reciprocal of the coefficient of the estimated RD effect on time over the $0.25 toll increase gives the implied value of time saved from traffic, which the table reports in dollars per hour saved.[41] The overall mean VOT is $66.56 per hour saved, with some differences across roads: the implied mean VOT is $69.92 on I-394, $61.52 on I-35W and $59.00 on I-35E. All these estimates are 2-2.5 times as large as the average hourly wage rate in the Minneapolis-Saint Paul area, which is estimated to have been $28.52 between November 2016 and May 2019.[42]

In the second panel, the table shows the VOT results if the strategy had been to run an OLS hedonic regression of toll paid on time saved with no additional controls. The hedonic OLS estimates imply a VOT between $8.58 and $19.85, which is significantly lower than the values I obtain using the RD strategy.[43] Drivers who choose to use the ELs might be likely to be more in a hurry compared to other drivers, and, thus, should be willing to trade higher

[41]The reciprocal of the coefficient in the table would be measured in dollars per minute saved, and the table conventionally reports this value multiplied by 60. However, the maximum amount of minutes saved in the data is just above 15 minutes, so the linear extrapolation from minutes to hour has to be taken just as a convention, but it might not be supported by the data as a functional form assumption.

[42]Source: US Bureau of Labor Statistics.

[43]These estimates are in a range similar to Bento et al. (2020), who run the same OLS hedonic regression using data from ELs in California.

Table 3: *Second-stage results of regression of time saved on traffic density as the running variable. For each road, this yields two peaks and two sections for each peak (except for road I-35E in the morning). Standard errors follow Bertanha (2020). "Section 1" in the morning peak denotes the portion of ELs that is further away from the city center (either Minneapolis or Saint Paul), while "Section 2" denotes the portion that is closer to the city center. "Sections 1-2" denotes trips that originated in the suburbs and concluded in the city center. The opposite is true during the afternoon peak.*

	I-394	I-35W	I-35E
Implied VOT morning section 1 ($/hour)	**56.43**	**81.93**	**46.76**
	(3.53)	(6.19)	(2.53)
Implied VOT morning section 2 ($/hour)	**70.53**	**161.43**	
	(9.02)	(48.57)	
Implied VOT morning sections 1-2 ($/hour)	**97.99**	**100.4**	
	(16.27)	(12.63)	
Implied VOT afternoon section 1 ($/hour)	**44.87**	**73.87**	**64.73**
	(3.62)	(9.54)	(13.14)
implied VOT afternoon section 2 ($/hour)	**85.23**	**35.39**	**79.53**
	(13.49)	(2.19)	(20.35)
implied VOT afternoon sections 1-2 ($/hour)	**62.03**	**48.93**	**89.40**
	(6.9)	(3.61)	(25.12)
N	956,530	642,886	337,226

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

tolls for lower time savings. Hence, the error term would be inversely correlated with the endogenous regressor of time saved, which, in turn, implies that OLS would underestimate the true parameter. This issue should hold true for all levels of the toll, all locations and at all points in time, which would make it difficult to alleviate the endogeneity by using, for instance, location or time-of-day fixed effects. Instead, the RD strategy arguably holds this type of selection into the EL fixed, since it compares drivers who are using the EL just before and just after each density cutoff that induces a discrete toll increase.

Table 3 shows results broken down by time of day (morning or afternoon peak) and by travel location (section 1 or section 2) road by road. The results now range from $161.43 on road I-35W during the morning peak to $35.39 on road I-35W during the afternoon peak. The implied value of time seems to be higher during the morning peak relative to the afternoon peak, which could be due to the fact that the value of not being late is more salient in the morning commute.

These results are robust to the exclusion of trips taken outside of each driver's usual route, as *Appendix Table B.1* shows. For each driver, the usual route is defined as the most frequent combination of entry and exit section in the EL. Roughly between 70% and 80% of trips

Table 4: *Correlation between reduced-form results and aggregate zipcode level Census data. Average hourly wage uses Census Business Patterns 2018 data from morning destination zipcodes. All other variables use American Community Survey 2018 5-year data from morning origin zipcodes. The underlying assumption is that drivers use the ELs to go to work in the morning and to return home in the afternoon. Data by zipcode is matched to drivers based on their EL entry and exit location.*

	Correlation with estimated VOT
Average hourly wage	0.2324***
Median individual income	0.1721***
Median household income	0.3571***
Share households above $200k yearly income	0.3114***
Median owned property value	0.2172***
Share of properties over $1M	0.1343***

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

happen along drivers' usual routes.

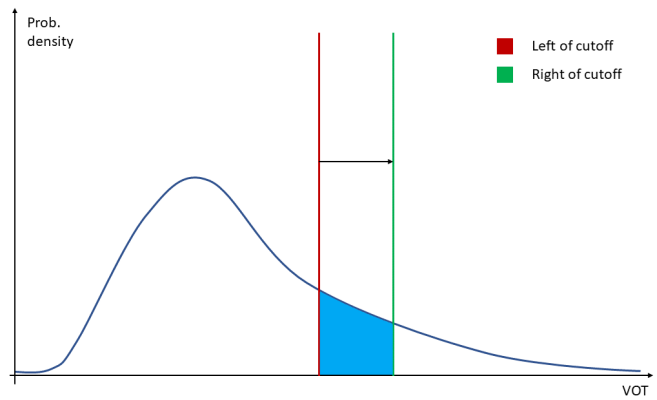
To further corroborate the result, I provide suggestive evidence using Census data at the zipcode level. Using drivers' usual entry and exit location to the EL, I can match them to their corresponding aggregate demographics in terms of income, property value and hourly wages. *Table 4* shows that the reduced-form VOT results are, as expected, positively correlated with average hourly wage and with median individual income. Moreover, results are also positively correlated with the share of households with income above $200,000 and the share of properties valued over $1 million, which might point to the importance of the share of individuals coming from the right tail of the VOT distribution in driving the results. *Table B.2* provides further details by breaking down these correlations by location and time of day.

Overall, the RD evidence in this section shows that EL users have much higher VOT than the average Minnesota wage and there is heterogeneity in VOT depending on travel location and time of day. The RD estimates are also substantially higher than OLS estimates, which are similar to the VOT used by the US government. However, these results are also conditional on using the EL and, as such, they refer to a policy-relevant but particular population. In the next section, I estimate the VOT distribution for this entire population of drivers, regardless of EL usage. The strategy exploits the same source of exogenous variation that comes from the tolling function.

4 Step II: estimation of VOT distribution among all drivers

In this section, relying on the same identification source as the RD, I estimate the distribution of VOT saved from traffic for the entire population of commuters who drive on the highway. I exploit the fact that I observe aggregate EL choice probabilities at all times and estimate the VOT as a random coefficient using a mixture approach. I provide the intuition

Figure 10: *Intuition behind the estimation of the VOT distribution. The blue curve is the underlying probability density function of the VOT distribution in the population of drivers. The two vertical lines are the value of the ratio between the toll paid and the time saved on each side of the cutoff. The light blue area under the curve is the part of the distribution identified by the RDD design in a simple one-cutoff example.*



behind the design, the link with the theoretical framework, the estimation strategy and finally the estimated distributions.

4.1 Intuition and estimation strategy

The intuition for the VOT distribution estimation follows directly from the same theoretical framework as in the previous RDD. The identifying variation for this estimation is also the same, but the outcome variable of interest is now the aggregate EL choice probability. Consider a market j as a subsample of observations where individuals face the same combination of time saved τ_j and toll π_j. Conditional on commuting by car, a driver i in market j chooses to use the EL if their latent utility u_{ij}^{EL} is positive:

$$u_{ij}^{EL} = \delta^{EL} + \beta_i^{VOT} \cdot \tau_j - \pi_j + \varepsilon_{ij}$$

where δ^{EL} is the pure preference for taking the EL and β_i^{VOT}. In each market j, the underlying VOT distribution in the population determines the aggregate demand P_j^{EL} for a product (the EL) with market-specific attributes (a combination of toll and time saved). Thus, the aggregate EL choice probability is directly linked to the VOT distribution.

To form intuition about how to estimate the VOT distribution from the aggregate EL choice probabilities, consider a simple example with only one market and one cutoff. A graphical intuition is given by *Figure 10*, where the blue curve is the underlying VOT distribution in the population. To the left of the cutoff, the time saved is τ^0 and the toll is π^0; their ratio

is VOT^0 is measured in dollars per unit of time and is represented by the red vertical line. At the cutoff, the toll increases by \$0.25 to π_j^1, the aggregate EL choice probability decreases and the time saved by taking the EL increases to τ_j^1. Thus, the ratio $VOT^1 = \frac{\pi^1}{\tau^1} > VOT^0$ is represented by the green vertical line. The difference between VOT^1 and VOT^0 depends on the change in aggregate EL choice probability, which in turn depends on how many drivers have a VOT between the red and the green line levels. Hence, the change in aggregate EL choice probability at the cutoff, which can be estimated by the RDD,[44] identifies the blue area of the VOT distribution. Repeating this exercise across different markets and cutoffs uncovers the entire VOT distribution.

With this intuition in mind, I now specify the model. Following Fox et al. (2011), the VOT distribution that β_i^{VOT} is drawn from can be approximated by a set of M mass points $\beta_i^{m,VOT}$ each having probability θ^m, with $\sum_m \theta^m = 1$. Hence, assuming that ε_{ij} follows a logistic distribution with mean 0 and scale parameter s to be estimated, the share of drivers who use the EL in market j is given by:

$$s_j^{EL} = \sum_{m=1}^{M} \theta^m \cdot s_{jm}^{EL} = \sum_{m=1}^{M} \theta^m \frac{exp\left(\frac{\delta^{EL}+\beta^{m,VOT}\cdot\tau_j-\pi_j}{s}\right)}{1+exp\left(\frac{\delta^{EL}+\beta^{m,VOT}\cdot\tau_j-\pi_j}{s}\right)} \tag{4}$$

Since what identifies the VOT distribution is the change in the EL share of traffic in each market j, I focus on $\Delta s_j^{EL} = s_j^{EL,1} - s_j^{EL,0}$, where the superscripts 0 and 1 denote left and right of the cutoff, respectively. The expression for Δs_j^{EL} is then:

$$\Delta s_j^{EL} = \sum_{m=1}^{M} \theta^m \left[\frac{exp\left(\frac{\delta^{EL}+\beta^{m,VOT}\cdot\tau_j^1-\pi_j^1}{s}\right)}{1+exp\left(\frac{\delta^{EL}+\beta^{m,VOT}\cdot\tau_j^1-\pi_j^1}{s}\right)} - \frac{exp\left(\frac{\delta^{EL}+\beta^{m,VOT}\cdot\tau_j^0-\pi_j^0}{s}\right)}{1+exp\left(\frac{\delta^{EL}+\beta^{m,VOT}\cdot\tau_j^0-\pi_j^0}{s}\right)} \right]$$

where τ_j^1, τ_j^0, π_j^1 and π_j^0 are all observed, and $\pi_j^1 = \pi_j^0 + 0.25$ and $\tau_j^1 = \tau_j^0 + \tau^{RDD,j}$, with $\tau^{RDD,j}$ the RDD effect on time saved in market j.

At this point, I make the following two assumptions to bring the model to the data.

Assumption 4.1: Independence. $\mathbb{E}[\theta^m|\tau_j, \pi_j] = \mathbb{E}[\theta^m] \quad \forall j = 1, \ldots, J, \quad \forall m = 1, \cdots, M$

This assumption states that all the probabilities θ^m, which approximate the VOT distribution, are independent across all J markets. Essentially, this imposes that the underlying VOT distribution in the population does not depend on any of the market j attributes that determine the aggregate EL choice probability. This assumption guides the choice of subsamples in the data that can serve as a market. I choose to consider each combination of day

[44] In fact, an example that shows the average effect across all cutoffs is in Panel (c) of *Figure 9*.

and cutoff as a separate subsample or market. There are 250 days available in the dataset[45] and at least 10 observed cutoffs per day. This choice seems reasonable because we expect the pool of commuters to be the same every day.[46] This particular market choice implies that the estimated VOT distribution refers to the average population of drivers who commute daily.[47]

The second assumption is the following. I denote the ratios between toll and time saved in each market j, on each side of the cutoff, by $VOT_j^0 = \frac{\pi_j^0}{\tau_j^0}$ and $VOT_j^1 = \frac{\pi_j^1}{\tau_j^1}$, respectively:

Assumption 4.2: Relevance. $\beta^{m,VOT} \in [VOT_j^0, VOT_j^1] \quad \forall m = 1, \cdots, M, \quad$ *for some* $j = 1, \ldots, J$

This assumption states that the market attributes of toll and time saved have to be such that there is support for all mass points $\beta^{m,VOT}$ of the VOT distribution, at least in some market j. This imposes a relevance condition on the dataset, and has two main implications. First, each subsample has to be large enough that there are marginal drivers who respond to the toll increase at the cutoff, so that $VOT_j^1 > VOT_j^0$ and $[VOT_j^0, VOT_j^1] \neq \emptyset$. Were this not the case, the design would not identify any area of the VOT distribution. Each day in the sample has between 5,000 and 8,000 observations, thus providing support for this condition. Second, I need to choose a support set for the $\beta^{m,VOT}$ points such that the variation across subsamples spans the entire set. I choose a VOT space from \$0 to \$200 per hour saved. *Figure 11* shows the variation in time saved and in the ratio of toll over time saved across these subsamples.[48] In both cases the distribution of observations right of the cutoff stochastically dominates the one left of the cutoff. Panel (b) provides suggestive support for the idea that this choice of subsamples has variation that spans the chosen VOT space.

These two assumptions guide the estimation of the model. Following Fox et al. (2011), what I actually observe in the data and estimate through the RDD is a noisy measure $\widehat{\Delta s}_j^{EL}$ of the true share change. The above expression can then be rewritten as:

$$\widehat{\Delta s}_j^{EL} = \sum_{m=1}^{M} \theta^m \left[\frac{exp\left(\frac{\delta^{EL} + \beta^{m,VOT} \cdot \tau_j^1 - \pi_j^1}{s}\right)}{1 + exp\left(\frac{\delta^{EL} + \beta^{m,VOT} \cdot \tau_j^1 - \pi_j^1}{s}\right)} - \frac{exp\left(\frac{\delta^{EL} + \beta^{m,VOT} \cdot \tau_j^0 - \pi_j^0}{s}\right)}{1 + exp\left(\frac{\delta^{EL} + \beta^{m,VOT} \cdot \tau_j^0 - \pi_j^0}{s}\right)} \right] + \zeta_j$$

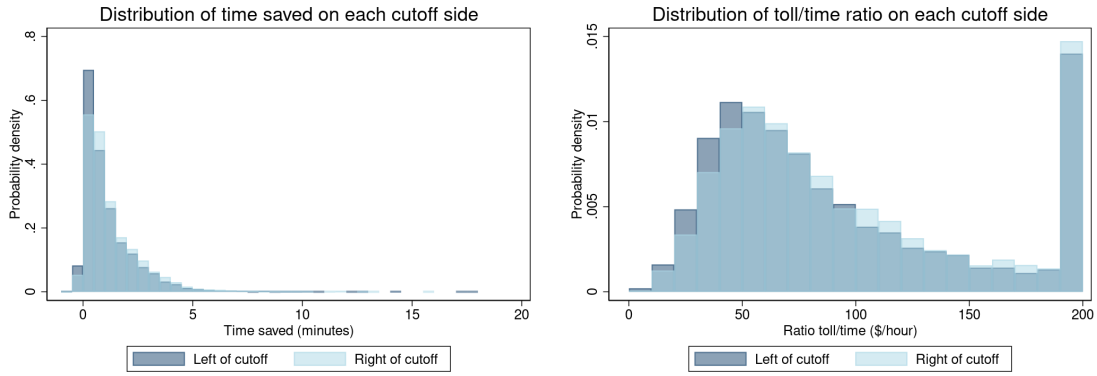
[45] 8 days, including Thanksgiving and Christmas, were dropped from the sample because they had few observations and extraordinarily little traffic.

[46] For example, a choice of markets that does not satisfy the independence assumption would be to divide the data in subsamples on the basis of travel location. It is likely that agents who travel and work in different locations have systematically different underlying VOT.

[47] Even though the individual choice of the non-commuting outside option is not observed, the estimated VOT distribution is the policy-relevant one.

[48] About 4.78% of observations left of the cutoff and about 2.99% right of the cutoff have negative time saved. The observations with negative time saved have been omitted for clarity from the toll/time ratio plot. About half of the observations where the ratio is larger than \$190 per hour in Panel (b) are due to time savings smaller than 10 seconds. The estimation rationalizes all these types of observations.

Figure 11: *Variation in the dataset across subsamples, where each subsample is a combination of day and cutoff level. Panel (a) shows the distribution of time saved on each side of the cutoff. About 4.78% of observations left of the cutoff and about 2.99% right of the cutoff have negative time saved. Panel (b) shows the distribution of the ratio between the toll paid and the time saved on each side of the cutoff. The observations with negative time saved have been omitted for clarity.*



(a) Time saved across subsamples

(b) Ratio of the toll and time saved across subsamples

where $\zeta_j = \widehat{\Delta s}_j^{EL} - \Delta s_j^{EL}$, given τ, π and $\beta^{m,VOT}$, is an error term with an expected value equal to 0. Thus, for every fixed couple (δ^{EL}, s), the probabilities θ^m can be estimated using inequality constrained least squares. This provides a non-parametric approximation of the VOT distribution of the population of drivers, conditional on commuting.

In terms of implementation, I divide the VOT space into \$10 bins, up to a maximum of \$200 per hour saved,[49] and consider the center points of each bin to be the mass points of the approximating distribution. I start with a guess for (δ^{EL}, s), then estimate the probabilities θ^m using inequality constrained least squares and finally I draw a simulated sample of individuals according to the implied VOT distribution. Using the simulated sample, the couple (δ^{EL}, s) is estimated by matching the average share of EL traffic[50] and the average share of trips where individuals would accept negative time saved.[51]

While this estimation procedure recovers the VOT distribution for all highway commuters, its design relies on aggregate observations because individual choice of the general lane is not observed. However, I can also characterize the individual VOT of drivers who are frequently observed using the EL. These drivers have the same latent EL utility as the general population,

[49] Because there might be drivers who have higher VOT, the amount of drivers in the top bin approximates the total density of the right tail of the distribution.

[50] Indeed, δ^{EL} is identified as $log(s^{EL}|\tau = 0) - log(1 - s^{EL}|\tau = 0)$, and an open set around $\tau = 0$ is observed in the data. What cannot be identified is the value of commuting, regardless of whether it happens on the EL or not, relative to the outside option, because the share that chooses the outside option is held constant by the design but is not observed.

[51] The fit of this moments is shown in *Figure C.1* in the Appendix.

but to estimate their individual VOT I employ a standard logit approach under a different set of assumptions.

In particular, I can relax the independence assumption because the individual VOT is within-person and one single driver does not affect the aggregate level of the toll and time saved. I can also relax the relevance assumption, because these drivers are frequently observed and I do not need specific variation at the discontinuity cutoffs. Instead, I make the following two additional assumptions. First, I assume that each individual is driving on the standard lanes when they are not on the EL. The credibility of this assumption relies on the fact that the focus is on frequent EL drivers, who are likely to commute every day. I focus on drivers who use the EL at least 10 times per year.[52] Second, I assume that frequent EL drivers have the same EL preference and scale parameter of the error distribution as the general population, which allows me to plug in the estimated δ^{EL} and s. At this point, I can estimate the individual VOT of frequent EL drivers using a standard likelihood function, where the choice of the standard lane is normalized to have 0 utility. For each driver, the estimation recovers the individual VOT parameter that maximizes the likelihood of observing that driver's set of EL choices.[53]

4.2 VOT distribution results

Figure 12 presents the resulting general VOT distribution. The median is equal to \$17.42 per hour saved, the 75th percentile is \$34.97, the 90th percentile is \$100.82, and the 95th percentile is \$166.05. About 3.25% of drivers have VOT larger than \$190 per hour saved, as the top bin shows. The mean is equal to \$35.03 per hour saved and is likely a lower bound to the true mean, given that the estimated distribution is truncated at \$200 per hour. In this case, the mean would not provide a good description of the average driver because the lower-bound mean already lies above the 75th percentile. Finally, as a useful comparison, the vertical red line in *Figure 12* shows the RDD result, which is close to the 85th percentile, suggesting that EL users tend to be high-VOT individuals.

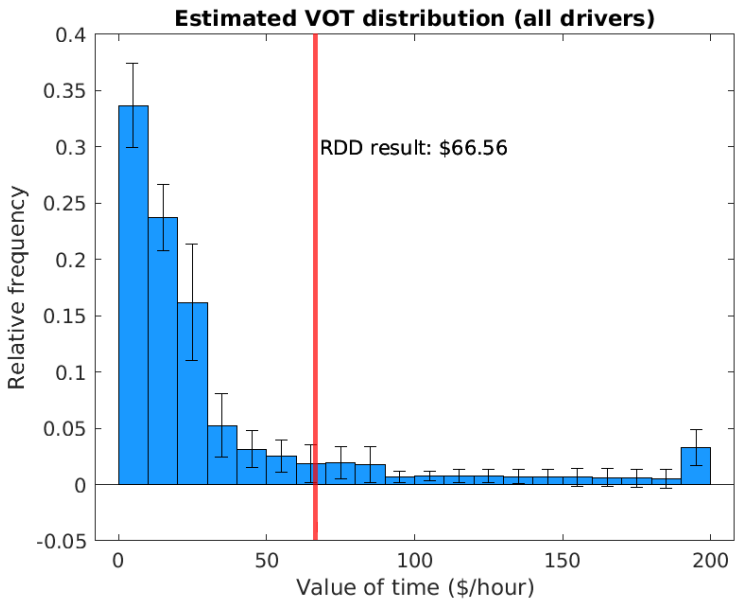
Figure 13 compares the general VOT distribution (in bright blue, as in the previous plot) and the one for frequent EL drivers only (in light blue). The red vertical line is the benchmark RDD result. The distribution for frequent EL drivers appears to first-order stochastically dominates the general distribution.[54] The RDD result is generally more consistent with the mean VOT implied by the VOT distribution for frequent EL users. This might be because the RDD result is driver primarily by frequent users.

[52] This excludes about 50% of the individuals in the panel. In what follows, I provide robustness checks for this choice.

[53] For about 2% of the frequent users the first order condition of the score function cannot be set equal to 0. In those cases the estimation does not return a valid individual VOT.

[54] In *Figure C.2* in the Appendix I show that using 10 observations per driver is not likely to introduce significant bias in the estimation.

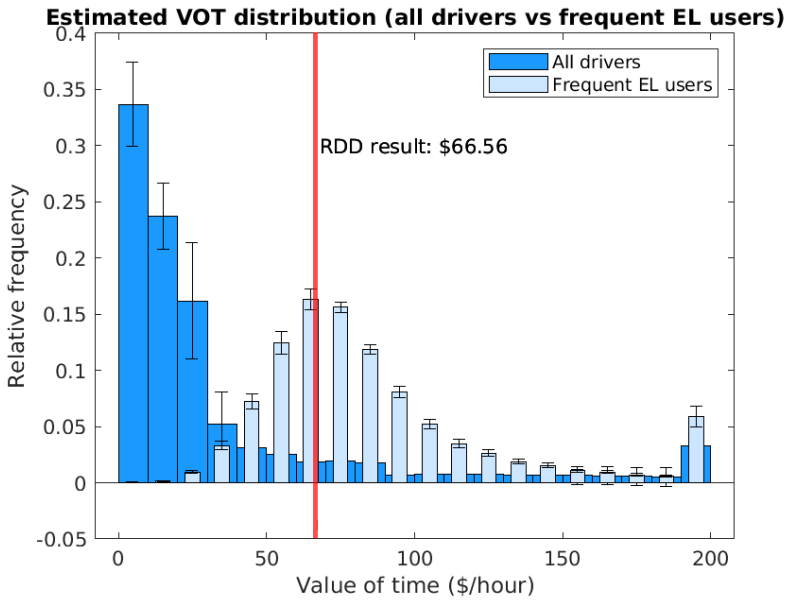
Figure 12: *Estimated distributions of VOT (in $ per hour) for all drivers. The width of each bar is $10 and the horizontal axis is cut at $200. The vertical axis reports the relative share of drivers with a certain VOT over the total number of drivers in that subsample. The median is equal to $17.42 per hour saved, the 75th percentile is $34.97, the 90th percentile is $100.82, and the 95th percentile is $166.05. The red vertical line represents the corresponding reduced-form VOT.*



Both estimations can be repeated for each road and time of day (morning or afternoon peak) separately. *Figure 14* shows the results, comparing both types of distributions. Although there are differences between roads, the same qualitative results found in the general distributions emerge. The plots also demonstrate that the estimation techniques are easily portable to other contexts whenever the data allows to measure the aggregate traffic shares and at least one of the two choices at the individual level, even when there is just one source of exogenous variation.

The estimation successfully recovers the VOT distribution for the entire population of drivers, conditional on choosing to commute. This is the policy-relevant population, and knowing their individual VOT already allows me to estimate the response to counterfactual congestion policies. In particular, the VOT determines individuals' valuation of travel time and the distributional effects of counterfactual policies. However, the fine and frequent measurements of aggregate traffic in the data allow me to characterize the driving choice further. In particular, again conditioning on choosing to commute by car, drivers are likely to choose their departure time before they decide whether or not to use the EL. In light of this, in the next section I present a structural model that includes the individual VOT as the key source of heterogeneity and introduces the departure time choice margin. The model allows for a better understanding and characterization of the individual responses to counterfactual

Figure 13: *Estimated distributions of VOT (in $ per hour) for all drivers (in bright blue) and for frequent EL users (in light blue). A driver is a frequent EL user if they use the EL more than 10 times per year. The width of each bar is $10 and the horizontal axis is cut at $200. The vertical axis reports the relative share of drivers with a certain VOT over the total number of drivers in that subsample. Frequent EL users account for only about 10% of the total number of drivers on the road each day. The red vertical line represents the corresponding reduced-form VOT.*



congestion policies.

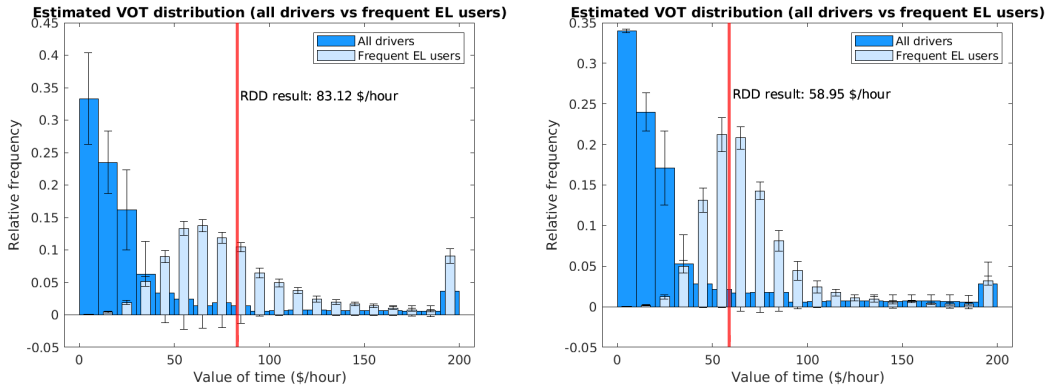
5 Step III: Structural model of departure time and EL choice

In this section, I introduce a structural model that uses the VOT distribution as the main source of heterogeneity and connects the previous results to a framework where individuals also choose when to commute. I focus on how the heterogeneity in VOT, together with departure time preferences, characterize individual behavior, which will be the basis for the analysis of counterfactual responses. First, I present the intuition, then a formal description of the model and finally the results. I present general results but, in principle, the model can also be estimated separately for each road in the sample.

5.1 Intuition and description of the model

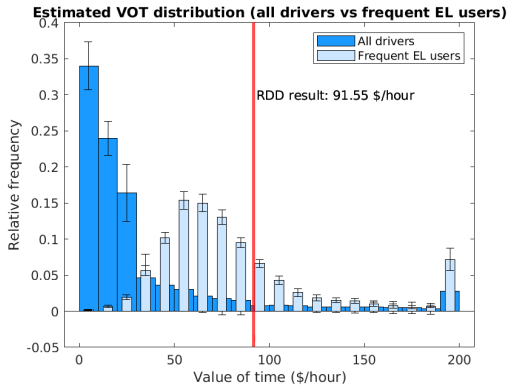
The distribution of VOT in the general population by itself allows me to study the individual response to counterfactual congestion policies because the VOT is sufficient to evaluate changes in travel times. Knowing the distribution is especially important to study the distributional effects of those policies. The purpose of the structural model is to further characterize

Figure 14: *Estimated distributions of VOT (in $ per hour) by road and peak for all drivers (in bright blue) and for frequent EL users (in light blue). A driver is a frequent EL user if they use the EL more than 10 times per year. The width of each bar is $10 and the horizontal axis is cut at $200. The vertical axis reports the relative share of drivers with a certain VOT over the total number of drivers. Frequent EL users account for only about 10% of the total number of drivers on the road each day. The red vertical line represents the corresponding reduced-form VOT for each road and time of day.*

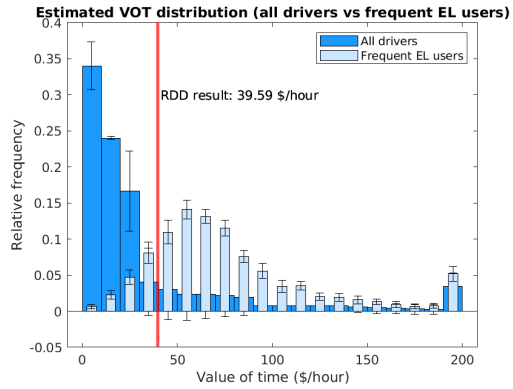


(a) Road I-394 (morning peak)

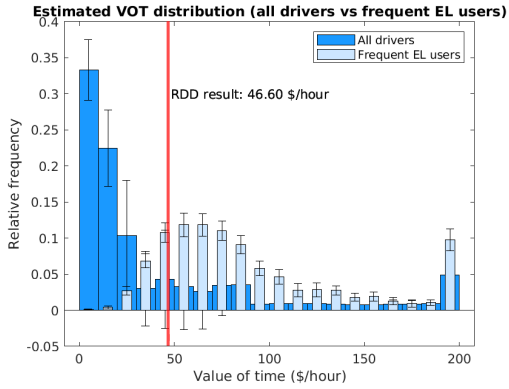
(b) Road I-394 (afternoon peak)



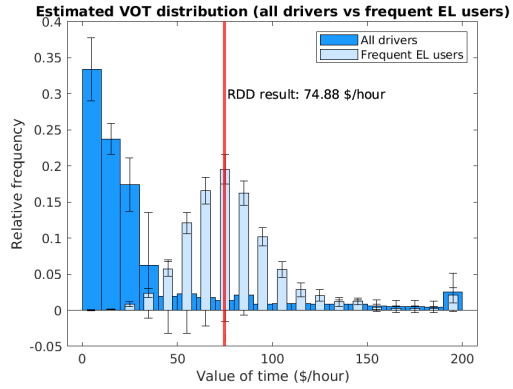
(c) Road I-35W (morning peak)



(d) Road I-35W (afternoon peak)



(e) Road I-35E (morning peak)



(f) Road I-35E (afternoon peak)

the commuting choice in order to provide additional understanding of individual responses beyond travel time evaluation. In particular, the model adds the choice of departure time, conditional on choosing to commute.[55]

The intuition for why this choice margin matters is the following. Suppose a driver is about to commute from home and, for any personal or work-related reason, has a certain pure preference for a departure time: for instance, 8am. Yet that driver also has an expectation of the travel time they will face on the highway (and the toll in case they use the EL). To pick an actual departure time, the driver balances these two factors. For instance, if at 8am the driver expects major traffic congestion, they might choose to leave home at 7:30am, even though that is not their purely preferred time. When policy changes are introduced to the highway setting, they can affect the expectations of both travel time and toll, but they do not affect the pure preferences, which remain stable. Hence, counterfactual policies might result in both a change in departure time choice and in EL choice once drivers are on the highway.

In this sense, the model provides a framework to evaluate individual drivers' choices conditional on commuting. In fact, the outside option of not commuting or using modes other than driving is still unobserved in the background and held fixed. Thus, the model refers to the policy-relevant population of usual car commuters. As mentioned in the data section, the existence of this stable policy-relevant population is supported by the fact that aggregate car volumes and aggregate transportation mode choice did not change from 2010 to 2018 in this study's setting. However, the model does not rely on exogenous variation in the extensive margin of commuting, mode choice, or residential location choice.[56] For this reason, the equilibrium of the model and the individual responses to counterfactual policies have to be intended as short-to-medium term.

Following the intuition, the model is formulated as an individual problem of minimizing commuting travel time. In this model, agents choose departure time and, conditional on departure, they choose whether to use the EL. Both the departure time choice and the EL choice are the result of a minimization of travel time. The choice process is divided into two stages. In the first stage, agents decide when to start their travel on the highway, knowing they have the option to use the EL and taking expectations of travel time. In the second stage, conditional on choosing the highway and on departure time, agents decide whether to use the EL in a general equilibrium framework. In other words, agents choose whether to use the EL depending on their VOT, their shock term, and how many other drivers are already using the EL. Each individual driver has a specific indifference point between the EL and the free lanes. The model assumes that preference parameters are stable and that drivers are

[55] *Figure C.3* in the Appendix provides graphical evidence to motivate why this might be a relevant choice margin. In fact, drivers seem to react to changes in expected travel times by slightly changing their entry time on the EL compared to their individual median entry time. The predictable changes in traffic and travel times are given by trips taken on Fridays, on snow days or on days close to national holidays.

[56] Each of these choice margins is interesting and deserving of future research.

rational in their travel time and toll expectations.

The first stage of the model happens before any travel begins, when drivers take expectations about travel time and the toll they might have to pay. The simplifying assumption here is that the drivers' travel begins directly on the highway.[57] In this stage, each driver i chooses, during each day d and peak p, the departure time t that maximizes expected utility $u(t_{dpi})$ knowing that they might take the Express Lane $EL(t_{dpi})$ or not.

$$u(t_{dpi}) = \beta_{pi} \cdot \alpha_{pt} + max\{EL(t_{dpi}), 0\} + \varepsilon_{dpti} \tag{5}$$

$$EL(t_{dpi}) = \delta^{EL} + \beta_{pi} \cdot \mathbb{E}[\tau_{pt}(t_{dpi})] - \mathbb{E}[\pi_{pt}(t_{dpi})] \tag{6}$$

The notation is as follows: α_{pt} are departure time FEs (at each peak p, morning or afternoon), β_{pi} is value of time at peak p, δ^{EL} is unobserved pure preference for the EL and the error term is ε_{dpti}. Travel time saved by taking the Express Lane is denoted by $\tau_{pt}(t_{dpi})$; toll on the EL is $\pi_{pt}(t_{dpi})$. In the first line, the departure time FEs are multiplied by the individual VOT so that preferences can be compared against the evaluation of expected travel time saved.[58] This means that individuals have the same shape of departure time FEs but at different absolute levels. The model divides each peak into 6-minute intervals, so each peak has 40 possible departure times.[59] The parameters β are non-parametric and drawn from the previous estimation of the VOT distribution.

Drivers are thus characterized by one fundamental type of heterogeneity in terms of the value of time through the parameter β_{pi}. The choice of departure time in each day and peak will depend on the trade-off between an average "pure preference" parameter α_{pt} and the disutility from expected travel time at each moment, up to an error term.[60] To reconnect with the intuition, for instance, a driver might have a pure preference for traveling at 8 am but, because 8 am is a very congested moment of the day, they might choose to travel at 7:30 am, when they expect to find faster traffic.

In the first stage, drivers are also only taking expectations of time saved and toll, because they are not traveling yet. I construct these variables in the following way. First, on each road and at each peak p, the commute of all drivers is as long as the corresponding average

[57] As mentioned in the data section, the highway is the only feasible route to reach downtown for commuters in the sample.

[58] If, instead, I measured the FEs directly in dollars, flexibility in terms of drivers' departure time preference would depend on the VOT by construction. In fact, the difference between any two FEs would be fixed across drivers, whereas the difference between any two expected travel times would be evaluated on the basis of individual VOT. Hence, high-VOT drivers would place more weight on travel time changes than would low-VOT drivers by construction.

[59] This choice makes the model computationally easier and it maintains a tight link with how the ELs work, given that the toll changes every 3-minute interval.

[60] Notice that the departure time choice does not depend on what drivers have chosen in the past days or on what they expect to choose in future days. In this sense, the model begins and ends with each day and peak, and repeats itself across all days in the simulation.

commute observed in the EL panel on each road at each peak p.[61] Second, drivers draw a value for traffic density (in vehicles per mile) on both the general lanes and the EL from a joint distribution that mimics the observed data.[62] Traffic density then has a one-to-one correspondence with toll and travel time saved, which enters drivers' utility.[63] The value of the expectation depends on the peak p and on the departure time t, but it is otherwise constant across drivers. Each drivers takes new expectations every day in the model, and since the error term ε_{dpti} has a different value every day, the optimal t_{dpti} can change every day, affecting the value of the expected $\tau_{pt}(t_{dpi})$ that enters the expression for $EL(t_{dpti})$.

Finally, I assume that the error term ε_{dpti} is logistic with location 0 and scale s^{ε} to be estimated. These shocks thus are drawn by each driver i, each day d, at each peak p, and for each potential departure time t. These terms could rationalize the fact that drivers on certain occasions might have a strong preference for a specific departure time (for example, if they have to arrive by a specific time). Moreover, the randomness introduced by the error term at this stage is also part of the error with mean 0 that drivers make while having rational expectations about travel time. Importantly, it is assumed that the errors at each stage of the model are independent of each other.[64]

After a driver chooses the departure time t^* that maximizes expected utility in the first stage, the problem moves to the second stage. At this point, the driver sees the realizations of travel times in each lane and the toll on the EL and chooses between the general lane and the EL. Thus, in the second stage, each driver i in day d and peak p chooses the EL if their latent utility u^{EL} is positive:

$$u_{dpi}^{EL} = \delta^{EL} + \beta_{pi} \cdot \tau(t_{dpi}^*) - \pi(t_{dpi}^*) + \eta_{dpi} \tag{7}$$

The notation is the same as in the first stage, except that there are no longer expectations of travel times and the toll, which are now dependent on t_{dpi}^*. The error term is η_{dpi}, which has mean 0 and a scale parameter plugged in from the VOT distribution estimation. The departure time FEs do not appear because this stage happens conditional on optimal departure time t^*, which was chosen in the previous stage. This second stage is essentially a re-statement of the VOT distribution estimation part, with the exception that here the model happens sequentially; thus, the realizations of toll and travel time depend on the simulated choices in the model during the previous 6-minute interval. In other words, the estimation of the

[61] The exact values can be found in the Appendix. The average travel length in the data is about 6 miles.

[62] For every 6-minute departure time interval, I assume the distribution of density on the general lanes and the EL is jointly normal and compute its mean and covariance from the traffic data.

[63] Regarding travel time, in principle, traffic density has a one-to-one correspondence with speed, as *Figure C.4* shows. From speeds on the EL and the standard lanes, I can then compute travel times, and their difference yields the time saved.

[64] If they were not independent, a driver who has a shock realization in the first stage that makes them prefer a certain departure time might also prefer the EL in the second stage. There is no *a priori* reason why such a preference pattern would occur but this possibility is excluded by assumption.

VOT distribution used data observations, whereas the model in this stage uses endogenously generated observations.

A few conditions are necessary to solve this stage of the model. First, for each day d and each peak p, this stage of the model has to be solved sequentially, starting from the first available departure time t^* and moving on to the last interval of the peak time. This is so because the toll at each t^* is set as a function of the EL density during the previous interval, with the function parameters set equal to the real EL tolling function in the data.[65]

Second, the solution needs to be found in a general equilibrium framework. In principle, each driver has an individual-specific tipping point after which they prefer the EL over the general lane. This tipping point depends on travel time in each lane, on the EL toll and on the realization of both error terms. However, travel time in each lane in turn depends on what all the other drivers traveling in the same interval are doing. Hence, given the total number of drivers traveling at each t^*, the solution allocates them between the EL and the general lane so that no drivers allocated to one option would prefer the other and vice-versa.[66]

5.2 Estimation strategy and results

The model is estimated using the Method of Simulated Moments. The model is simulated including both peaks for each road, and repeated over 250 days, as in the EL sample. In total, there are 82 parameters to be estimated: 40 FEs and 1 scale parameter for each peak (morning or afternoon). In principle, since drivers in the data always use the same road 96% of the time, the model can also be estimated separately for each road.[67]

The estimation targets a total of 164 moments from the data. The main group of moments is the traffic density in the general lanes or the ELs during each 6-minute interval of each peak. This group then totals 160 moments, 80 for each lane (EL or general) and 40 for each peak in each lane. In general there are 2-to-3 general lanes for each EL, and the traffic that goes through the general lanes together is about 8-to-9 times larger than the traffic that goes through the EL. I also target moments that exploit the panel dimension of the dataset and the exogenous variation. In particular, for each peak, I target the standard deviation of entry time on the EL and the average percentage change in traffic density in the general lanes at the cutoffs.

Thus, the identification of the parameters comes primarily from matching aggregate traffic data from both the ELs and the general lanes; the individual choices on aggregate rationalize the traffic patterns observed in the data. The other moments contribute to making the model consistent with the panel repeated observations and with the reduced-form evidence.

[65] Since the model uses 6-minute intervals, the toll in the model changes every 6 minutes instead of each 3 as in the real ELs. As in the real-world EL, the toll is set to equal \$0.25 during the first interval of each peak.

[66] This way of solving the model also guarantees that there is a unique allocation of drivers between the EL and the general lanes that solves the second stage in each 6-minute interval.

[67] I perform this exercise for validation and show results in the next subsection.

The estimation proceeds as follows. I start with a guess for the α_{pt} preference parameters and the scale s^{ε} for each peak p. Given these parameters, I need to solve for the equilibrium shares of the drivers population that choose to travel at each time of the day, i.e. the set of s_j^{EL} in (4). To do this, I start again with a guess for each of the s_{jm}^{EL} shares in (4) for each of the VOT distribution m mass points. Given these shares, drivers in each m VOT class form expectations about the time they could save by taking the EL at each departure time t and the toll they would have to pay.[68] These expectations are used to get a value for (6) for drivers in each m VOT class, which is then plugged into the utility function in (5). Given the guess for s^{ε}, this yields a set of probabilities that drivers in each m VOT class choose to leave at each departure time t, which are common knowledge across drivers, who have rational expectations. This process is iterated until the distance between these probabilities and the initial guess is below a set threshold.[69]

Notice that, intuitively, an equilibrium exists for the following reasons. First, the demand for EL time saved (hence EL traffic) in (6) is monotonically decreasing in the toll and monotonically increasing in individual VOT. The supply of EL time saved, instead, is inversely related to EL traffic. Hence, given any total number of drivers on the road, since the VOT distribution is continuous, there is only one equilibrium level of toll and time saved such that no driver is using the EL but has a negative utility from it and viceversa.[70] Second, given the VOT and the value of taking the EL for each m VOT class, each driver has a unique expected utility of choosing each departure time in (5), which depends on the scale s^{ε}. The aggregate traffic shares at each t need to be such that they minimize the distance from actual traffic moments.

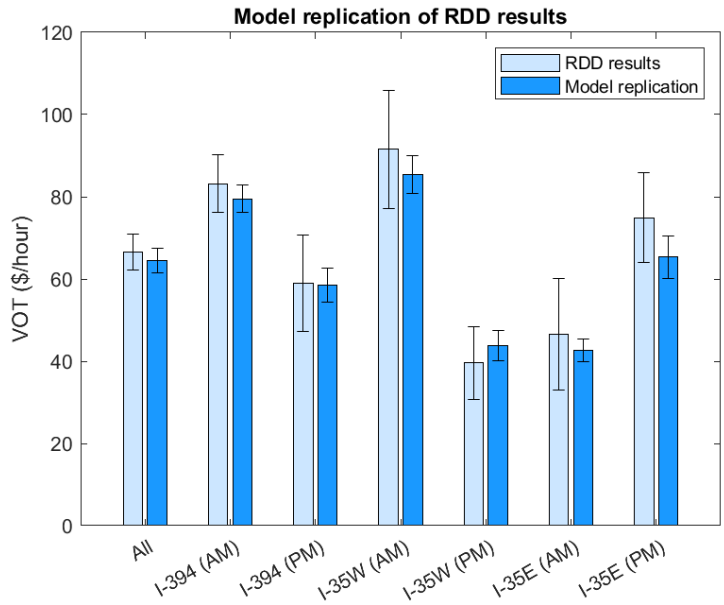
At this point, each driver in each day sees the realization of their own shock ε_{dpti} and chooses their preferred time t to commute as the one that yields the highest expected utility as defined by (5). Given these choices, the estimation moves to the second stage, represented by (7). In the earliest t, the toll is set to \$0.25, following the actual EL policy; at all other t times, the toll depends on the EL traffic share in the previous interval. Given the toll and which drivers chose to commute at each t, the model finds a general equilibrium EL traffic such that no driver is using the EL but has a negative utility in (7) and viceversa. An equilibrium exists here for the same reason as in the first stage of the model for (6). The estimation procedures continues until the simulated traffic shares match the traffic moments and until the simulated EL usage matches the standard deviation of entry time on the EL and the average percentage change in traffic density in the general lanes at the cutoffs.

[68] In particular, the expected toll results from the tolling function where traffic density is taken as the expected EL traffic in the previous 6-minute departure time interval. The toll in the first interval is set to \$0.25, following the actual EL policy.

[69] In the Matlab simulation, I set this threshold to 10^{-8}.

[70] In computational terms, since the VOT distribution is discretized by a set of mass points, the solution is found when the distance between the initial EL traffic share and the one implied by the solution is below a set threshold.

Figure 15: *Comparison of reduced-form VOT estimates (in light blue) and results from the model replication (in bright blue). 95% confidence intervals are delimited by the whiskers on top of each bar. The leftmost couple of bars include all roads and peaks, whereas the other couples are divided by road and time-of-day peak.*



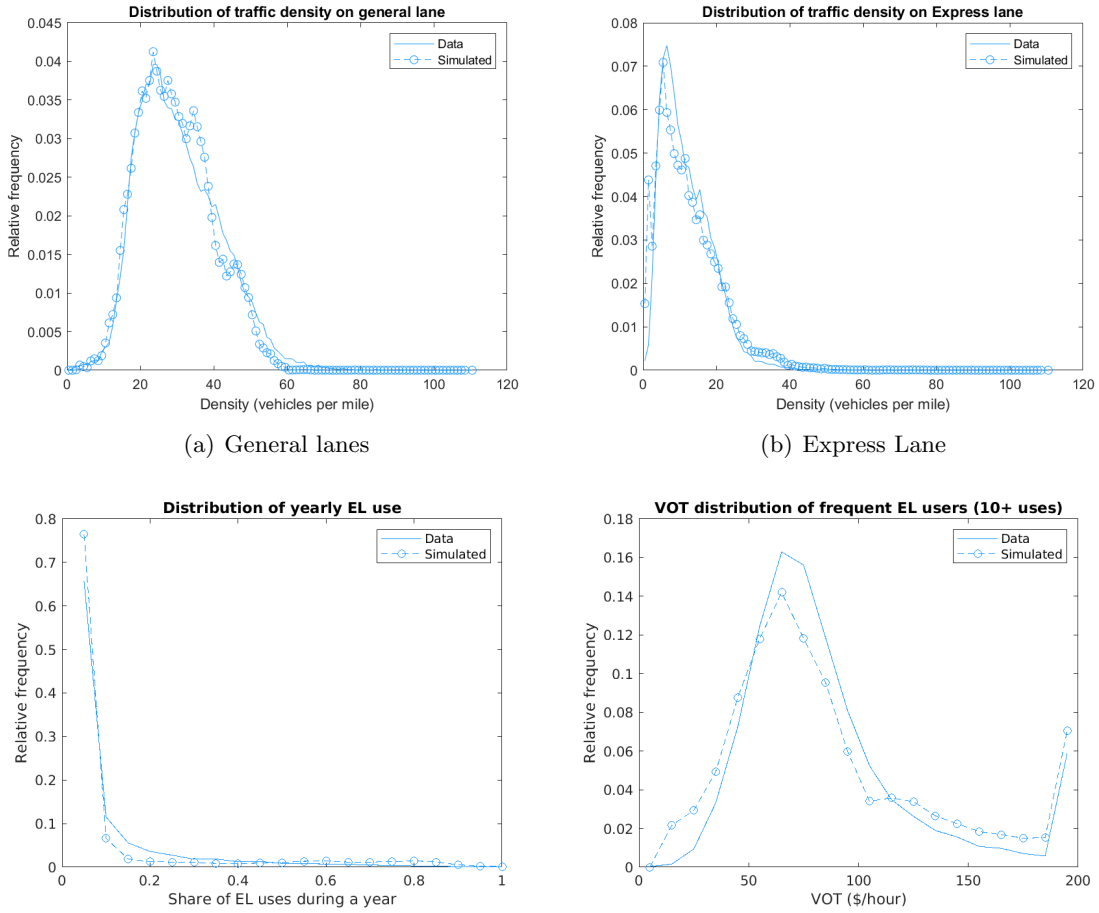
Importantly, the level of the departure time FEs has to be interpreted as relative to the lowest one, which is normalized to 0, plus an error term.[71] Ultimately, the estimated level of these FEs is pinned down by the total number of individuals used in the simulation, which is set to be equal to the average number of cars flowing on the highway each day and during each peak. Because this number includes all commuters on the highway and individual EL choices are endogenized, the model allows for the presence of never-takers of the EL.[72]

The model fits the targeted moments very well. Plots that show these are found in the Appendix in *Figures C.5, C.6 and C.7*. Note that the model also replicates the RDD results quite precisely, as *Figure 15* shows. Moreover, the model also reproduces three sets of untargeted moments. First, panels (a) and (b) in *Figure 16* show the data and simulated distribution of traffic in both the general lanes and the Express Lane. This indicates that the model is replicating both the average behavior of the data and the wide variation in traffic displayed by the data. Second, panel (c) shows the data and simulated share of drivers that use the EL from 5% to 100% of days in a year. Third, panel (d) shows the data estimated and simulated VOT distribution of frequent EL users (who use the EL at least 10 times per

[71]The lowest FEs have to be normalized because the outside option is not observed. Consequently, the model cannot estimate the relative value of the set of inside options as a whole.

[72]For completeness, I also add a baseline level of density on the EL that accounts for buses and carpools, which are not included in the model. This baseline level is below 10% of the EL traffic so it does not affect the dynamics of the model significantly.

Figure 16: *Comparison of the distribution of traffic density in the general lanes (Panel (a)) and in the Express Lanes (Panel (b)) as measured in the data (solid line) and as simulated in the model (dashed with round markers). The x-axis, measured in vehicles per mile, is divided into bins that are 1-unit wide. Relative frequency on the y-axis is the share of observations in each bin.*



(a) General lanes

(b) Express Lane

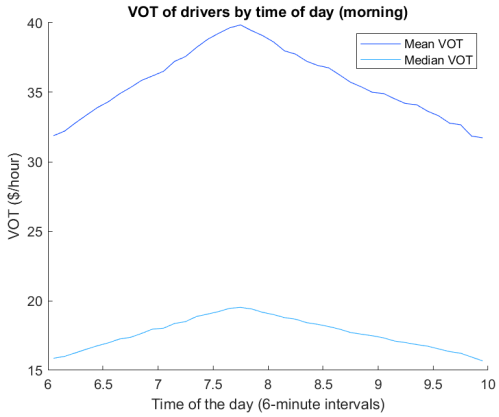
(c) Yearly EL use

(d) VOT distribution of frequent EL users

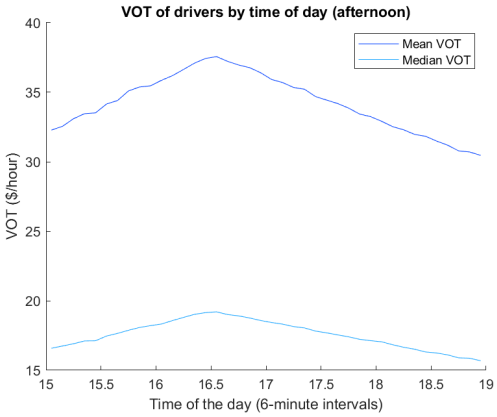
year). All these elements together confirm that this structural model provides a good and reliable representation of the data.

The model has implications about the characterization of individual choices that are especially relevant to the counterfactuals, as *Figure 17* shows. First, Panels (a) and (b) show the mean and median VOT of drivers by their departure time in the morning and in the afternoon, respectively. The general tendency is for higher-VOT individuals to travel in the hours of each peak that are more congested (6:30-8am and 3:30-5pm). This is due to the fact that high-VOT drivers have the option to use the EL to avoid congestion, whereas low-VOT drivers usually find the EL too expensive. These plots also suggest that, because of congestion and the inability to use the EL, low-VOT drivers decide to travel at less congested times, even

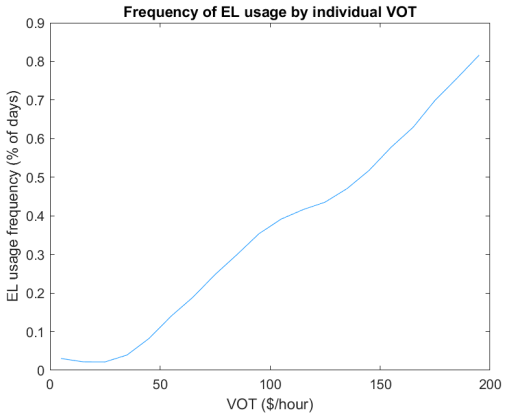
Figure 17: *Estimated correlation between EL usage frequency and VOT by road. The horizontal axis is VOT. The vertical axis reports the frequency of EL usage as the share of days in the simulation that a driver chooses the EL.*



(a) Mean and median VOT by departure (AM)



(b) Mean and median VOT by departure (PM)



(c) EL usage by individual VOT

if those are not their purely preferred departure times.

Second, Panel (c) in *Figure 17* shows how frequently drivers use the EL over the course of a year, depending on their individual VOT. There is a clear increasing trend, with individuals in the top bin using the EL over 80% of the days in the simulation. The very little EL usage by drivers who have VOT below $50 per hour indicates that the model allows for the presence of never-takers. However, since over 80% of drivers have VOT below $50 per hour, the model also shows that the EL is only used by a small fraction of the population of drivers.

The characterization of individual choices in the model highlights the importance of knowing the distribution of VOT. In fact, in both cases the differences in VOT result in behavioral differences. Knowing the departure time preferences jointly with the VOT distribution allows to estimate individual responses to and the welfare consequences of counterfactual policies,

with a particular focus on distributional effects.

6 Counterfactuals

The VOT distribution and the departure time preference parameters together allow me to estimate the response to and welfare effects of a wide range of counterfactual congestion policies. In particular, to assess the distributional effects of each policy, it is fundamental to know the VOT distribution, for two reasons. On the one hand, if the policy is meant to target a specific subset of drivers, knowledge of the individual VOT is necessary to judge whether the targeting was successful. On the other hand, if the policy-maker is concerned about any inequality generated by the policy, the individual VOT is again necessary to determine which drivers benefit from or are hurt by the policy.

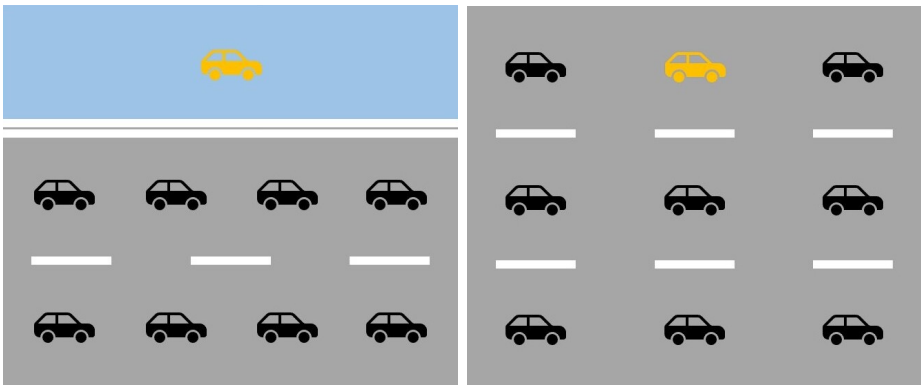
For these reasons, in this section I focus on four main counterfactual exercises. First, I reconvert the EL into a standard free lane, that all drivers can use at all times. Second, I change the composition of drivers to increase the share of low-VOT individuals but keep the EL in place. Third, I change the toll level from 0.1x to 2x the original level. Finally, I show what happens if the policy-maker tries to predict EL usage while ignoring the VOT distribution, assuming that all drivers have VOT equal to the mean.

In all counterfactuals, after the policy change I allow drivers to re-allocate to different departure times until aggregate traffic reaches a new equilibrium.[73] When making the new departure time choice, drivers face the usual trade-off between their pure preference over a certain departure time and their wish to have the lowest possible travel time. Drivers' utility is measured in dollars, and so it is straightforward to sum across individuals to obtain the aggregate change in welfare over the simulated year in the model and then express it in per-capita terms. I can further break down the welfare changes into the pure preference contribution and the travel time contribution.

In terms of outcomes, I am first interested in measuring the per-driver change in drivers' welfare compared to the baseline case of the structural model with the EL. Then, I compute the welfare changes after allowing the toll revenues generated by the EL to be rebated to all drivers. The intuition is that when individuals are stuck in traffic, everyone is paying the commute with their own time, which has an individual-specific value but is not transferable to others. Where an EL exists, some drivers use it and transform a time saving into money, which can be rebated to others. In this way, even if the EL causes the other lanes to become more congested, the rebates can serve as a form of compensation and, if they are high enough, they can produce a Pareto improvement. Finally, I show how the welfare changes are distributed across drivers depending on their individual VOT. This is a fundamental question shared by

[73] Because the simulation relies on the structural model presented in the previous section, the new equilibrium will also be a short-medium term response.

Figure 18: *Graphical representation of a counterfactual where the EL is reconverted into a standard lane that all drivers can use for free.*



(a) Baseline highway with EL (b) Highway with all standard lanes

policy-maker interested in the potential inequality generated by congestion policies, and it can be addressed only if the VOT distribution is known.

In each counterfactual exercise that follows, I first provide intuition about the policy change and then present the results and explain how they are relevant.

6.1 EL is converted into a standard lane

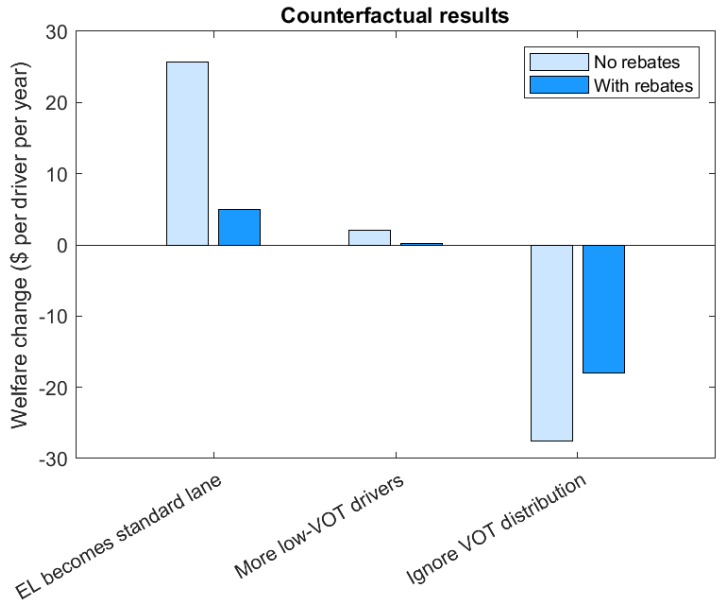
In the first counterfactual, I convert the EL into a general lane that all drivers can use for free. This decreases travel time on the other general lanes but it also prevents high-VOT drivers from using the EL as a time-saving option. The graphical intuition for this counterfactual is provided in *Figure 18*, where the left panel still shows the baseline from the structural model, with one EL in place. After the counterfactual policy is enacted, cars reallocate uniformly across all lanes, which are now all interchangeable.

In practical terms, I re-estimate the model with the EL transformed into a standard lane. This affects individuals' expectation of travel time at each time of day in the first stage of the model, and it lets drivers distribute themselves uniformly across lanes in the second stage. Hence, the counterfactual captures two kinds of responses. First, individuals might choose a different departure time based on their new expectation of travel time. Second, individuals no longer have the EL as an option to save time, so their final realization of travel time might change.

This counterfactual policy could be a response to the criticism that ELs are used only by a small portion of the population of drivers. *Figure 19*[74] shows that, in fact, reconverting the EL into a lane that all drivers can use for free increases per-capita welfare by \$25.68 per year. This number is higher than what 52% of drivers spend on the EL over the course of a year.

[74]The figure anticipates a summary of the other counterfactual results as well, which will be discussed later.

Figure 19: *Summary graph of per-driver welfare changes in response to counterfactual policies. The cases when rebates are allowed are shown in bright blue.*



The option to rebate toll revenues is interesting because this possibility is eliminated once the EL is reconverted; thus, these revenues might provide an argument in favor of keeping the EL. However, even with rebates, reconverting the EL increases per-capita welfare by $5.02, which is considerably smaller than the alternative, although still positive. In other words, high-VOT EL users do not benefit enough from the EL to compensate other drivers for the extra congestion the EL causes.

To study how welfare changes are distributed among drivers, I exploit the knowledge of the departure time preference parameters and, in particular, the VOT distribution. *Figure 20* shows the distribution of total changes and breaks it down into the contribution from pure preference and from travel time. It is evident that the gains from this counterfactual policy mostly go to low-VOT individuals, whereas welfare for high-VOT drivers actually decreases. Most of the gains for low-VOT drivers stem from the fact that they are now able to travel at the times of the day they prefer. This suggests that the presence of the EL forces these drivers to travel at times of the day they did not like. Conversely, because they no longer have the EL option, the losses for high-VOT drivers mostly come from increases in travel time. Hence, this counterfactual policy is not only welfare-increasing, it is also progressive.

6.2 Composition change: higher share of low-VOT drivers

In the second counterfactual, I change the composition of drivers to increase the share of low-VOT drivers. In particular, I substitute 20% of drivers drawn from anywhere in the VOT

Figure 20: *Plots of how counterfactual welfare change is distributed among drivers across the VOT distribution when the EL is converted into a standard lane. The horizontal axis reports the VOT and the vertical axis the share of total welfare change, broken down by the pure preference contribution and the travel time contribution.*

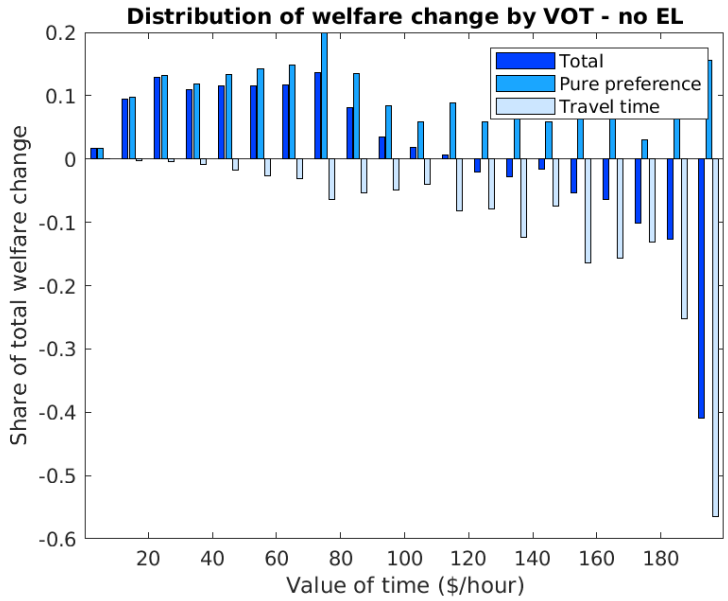
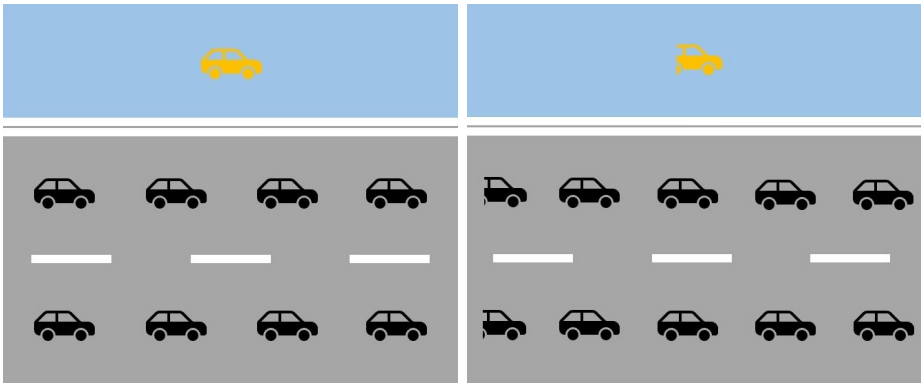


Figure 21: *Graphical representation of a counterfactual where the EL is reconverted into a standard lane that all drivers can use for free.*

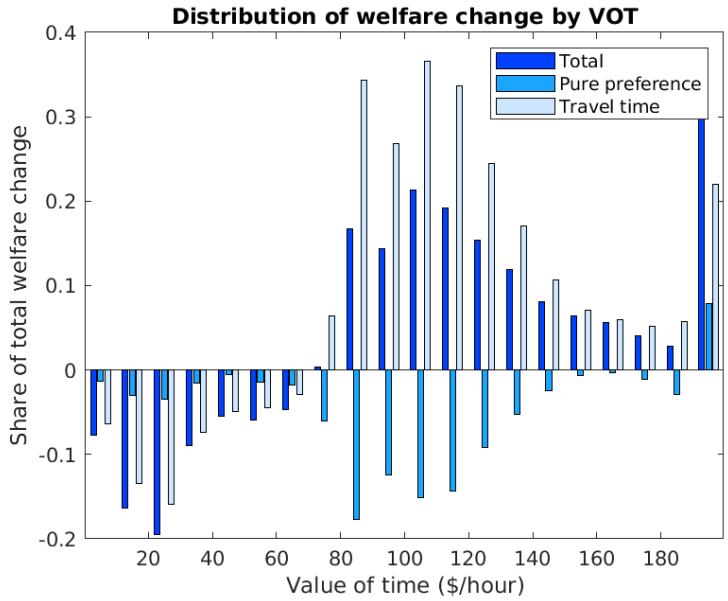


(a) Baseline highway with EL (b) Higher share of low-VOT drivers

distribution with 20% of drivers who have VOT lower than $20 per hour. Although arbitrary, the composition change is meant to represent the short-term consequences of shocks to the local population, such as labor market or health shocks.[75] The intuition for this counterfactual

[75]Bartik et al. (2019), for instance, find that the discovery of hydraulic fracturing opportunities led to both a population increase and a composition change in the workforce in parts of the US.

Figure 22: *Plots of how counterfactual welfare change is distributed among drivers across the VOT distribution when the share of low-VOT drivers in the population increases. The horizontal axis reports the VOT and the vertical axis the share of total welfare change, broken down by the pure preference contribution and the travel time contribution.*



is depicted in *Figure 21*. Given that the share of low-VOT is higher, this should result in less competition to use the EL and higher congestion on the general lanes.

As in the previous counterfactual, I re-estimate the model under the new conditions. In particular, individuals should slightly change their expectation of travel times in the first stage, given the new VOT composition in the population. However, the EL choices in the second stage respond to the likely change in competition to use the EL. *Figure 19* shows that this counterfactual modestly increases per-driver welfare by $2.00 per year. If rebates were allowed, the increase would be $0.20 because in this counterfactual there are relatively fewer high-VOT drivers, and so total EL revenues decrease.

However, the aggregate welfare increase masks the fact that this policy is regressive. In fact, *Figure 22* shows that 30% of the total gains go to drivers in the top bin of the VOT distribution, which accounts for less than 5% of the population. The gains are generally concentrated among high-VOT drivers. Low-VOT drivers bear extra costs even though they also face less competition to use the EL during the few times they find it optimal to use it.

Overall, this counterfactual policy disproportionately hurts low-VOT drivers, but it still increases aggregate welfare. This stems from the particular shape of the VOT distribution in the population and from the indivisibility of highway lanes. In fact, this counterfactual changes the VOT composition into one that is better suited for this particular EL setting, with

this tolling function and this population. In some sense, making the EL a more "elite" good is welfare-increasing, even if the increase occurs at the expense of low-VOT drivers.[76] Ideally, this suggests that keeping the EL in place is welfare-increasing when the underlying VOT distribution produces a sort of separating equilibrium: a small group of high-VOT individuals use the EL and all the other drivers use the general lanes and receive toll revenue rebates.

On the other hand, by construction, there can only be integer quantities of highway lanes. This limits the ability of the particular EL good to cater to the underlying VOT distribution in the population to increase aggregate welfare. There could be other settings in which individuals have their own valuations of a good that cannot be transferred but that can be divided more finely. For instance, waiting times for healthcare or hospital beds involve the same logic as this paper, but they pertain to goods that are more easily divisible. In those contexts, depending on how many high-valuation individuals there are and on how high their valuation is, allowing those individuals to cut waiting times and rebating their payments might have the potential to produce a Pareto improvement.

6.3 Toll level changes from 0.1x to 2x the original level

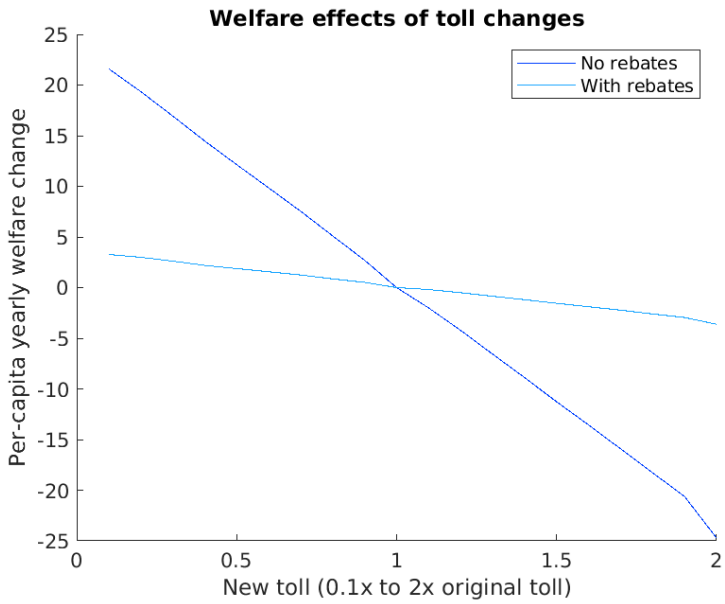
In the third counterfactual, I change the level of the toll so that it spans from 0.1x to 2x the original toll level, in 0.1x increments. I do this by changing the α parameter in the tolling function (1) from 0.0045 to 0.09, while holding the functional form fixed. Intuitively, lower levels of the toll make the EL more similar to a general lane and thus increase EL demand and traffic. Conversely, when the toll increases, less drivers use the EL and the general lanes become more congested.

At each new level of the toll, I re-estimate the model under the new conditions and compute the dollar welfare change per driver per year. I assume that drivers have the same average constant taste δ^{EL} for the EL as in the baseline model regardless of the new toll levels.

Figure 23 shows that, as the toll decreases, drivers' welfare per year increases, and vice-versa. When the toll is a tenth of the original level, drivers' welfare increases by \$21.60 per year; when the toll is twice the original level, drivers' welfare decreases by \$24.72. When the toll is the same as in the baseline model, the welfare change is 0 by construction. The shape of the welfare change function as the toll increases is roughly linearly decreasing and depicted by the dark blue line in the figure. If I allow for toll revenues to be rebated, the welfare effects of this counterfactual are qualitatively the same but they are almost entirely muted. In fact, in this case, when the toll is a tenth of the original level, drivers' welfare after rebates increases by \$3.37 per year; when the toll is twice the original level, drivers' welfare after rebates decreases by \$3.62. These welfare changes are depicted by the light blue line

[76] This result has some parallels with the phenomenon of elite capture studied in development economics, whereby a small group of individuals exploits public resources at the expense of the rest of the population.

Figure 23: *Plot of counterfactual welfare change per driver per year when the EL toll level is changed to span from 0.1x to 2x the original level. The horizontal axis reports the new level of the toll in proportion to the original one. Hence, at 1x the counterfactual welfare change is 0 by construction. The dark blue line and light blue line represent, respectively, the welfare change when rebates are not or are allowed.*



in the figure.[77] Overall, this result again points to the fact that, given the underlying VOT distribution, the current EL policy and pricing seem to be welfare-reducing for drivers.

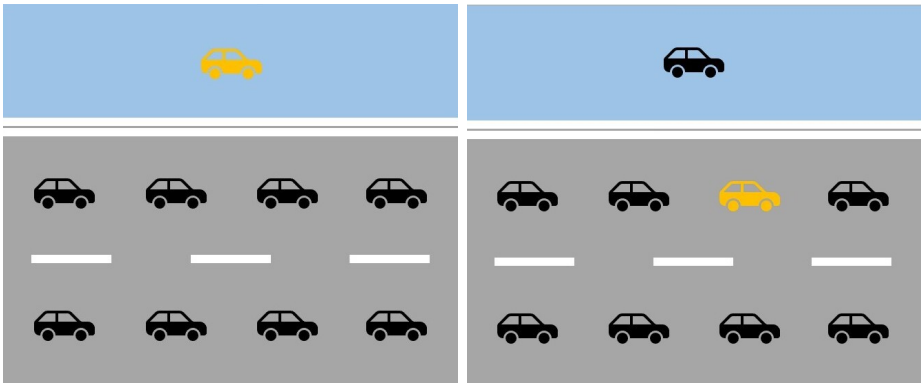
6.4 All drivers have VOT equal to the mean

In the fourth and final counterfactual, I suppose that the policy-maker wants to predict usage of the EL while ignoring the VOT distribution, and assumes that all drivers have VOT equal to the mean. If that is the case, then all drivers will be simultaneously indifferent between using the EL or not, and the equilibrium ratio of time saved and toll to be paid will be equal to the mean VOT. To achieve the equilibrium, an appropriate share of drivers needs to be allocated to the EL. Since drivers are supposed to all have the same VOT, this prediction by the policy-maker is equivalent to a random assignment of drivers to the EL.

The graphical intuition is provided in *Figure 24*. In the baseline case of a highway with one EL (Panel (a)), the yellow high-VOT car is on the EL. If drivers were mistakenly assumed to all have VOT equal to the mean, the policy-maker prediction of EL usage would look like Panel (b), where a low-VOT car finds itself on the EL and the high-VOT yellow one ends up

[77] Notice that extrapolating from the two welfare change lines to the 0 toll level yields slightly different results compared to the counterfactual where the EL is converted to a standard free lane. This is because, in that counterfactual, the newly converted free lane does not give the δ^{EL} constant utility anymore.

Figure 24: *Graphical representation of a counterfactual where the policy-maker wants to predict usage of the EL while ignoring the VOT distribution and assuming that all drivers have VOT equal to the mean. Compared to the baseline case with an EL (Panel (a)), this counterfactual amounts to randomly assigning drivers to the EL (Panel (b)).*



(a) Baseline highway with EL (b) Random assignment to EL

in the general lanes. However, drivers do still have their own individual-specific VOT, so the random assignment produces a different allocation than the one drivers would choose if they could do so. Thus, this counterfactual also suggests how much drivers value the possibility of being able to choose the EL or not, taking the existence of the EL as given.

In practical terms, I estimate this counterfactual by holding total highway traffic at each time of the day constant, and randomly assigning drivers to the EL to achieve an equilibrium. *Figure 19* shows the result: ignoring the VOT distribution implies a per-driver welfare loss of \$27.50 per year, which is more than what 53% of drivers pay for the EL over the course of a year. If rebates were allowed, the per-capita loss would be reduced to \$17.95.

The loss stems from two mirroring facts. On the one hand, high-VOT drivers would like to use the EL more often than the random assignment allows them to. On the other hand, low-VOT drivers are forced to use the EL more often than they would like to. Random assignment also prevents all drivers from responding to traffic conditions different from the mean and to individual shocks that make them want to use the EL more or less.

Hence, even without considering inequality concerns, ignoring the distribution of VOT implies a significant mischaracterization of individual EL choices and would result in a poorly designed congestion policy.

7 Concluding remarks and discussion

In this paper, I estimate the full distribution of value of time saved for a population of drivers, bridging a gap in the literature. I use a revealed preference approach and new data in

the setting of Express Lanes in Minnesota. The VOT distribution is a key source of heterogeneity in labor and urban models that account for commuting, and it is the policy-relevant object for traffic management and urban planning. The estimation exploits a particular feature of the EL tolling function, which allows me to isolate a travel time savings response to plausibly exogenous variation in the toll. The estimated VOT distribution has a median of \$17.42 per hour saved and a long right tail, with the 95th percentile at \$166.05 per hour. A small share of frequent EL drivers has a higher VOT than average and uses the EL substantially more than the rest of the population.

The commuting choice of the population of drivers is further characterized by a structural model that adds the departure time choice margin, with the individual VOT as the main source of heterogeneity. The departure time preferences together with the VOT distribution allow me to estimate a number of counterfactual congestion policies. Knowing the VOT distribution allows me to study how welfare effects are distributed across individuals, which is relevant both for inequality concerns and to design policies that target specific subsets of the population.

I find that the EL is welfare-reducing: reconverting it into a free lane increases per-capita welfare by \$25.68 per year because the benefits for high-VOT drivers do not compensate all other drivers of costs due to the extra congestion. In particular, this counterfactual policy is progressive because gains accrue mostly to low-VOT drivers. Moreover, I find that if the share of low-VOT drivers increased, keeping the EL in place would modestly increase per-capita welfare by \$2.00. However, this policy is regressive because most of the gains are appropriated by drivers in the top bin of the VOT distribution. For similar reasons, drivers' welfare is inversely related to toll changes: decreasing the toll increases drivers' welfare, since it makes the EL more similar to a standard free lane. In general, the EL is welfare-increasing when the underlying VOT implies a separating equilibrium: a small group of high-VOT individuals use the EL and all other drivers use the general lanes and receive toll revenue rebates. Finally, I find that ignoring the VOT distribution and assuming that all drivers have VOT equal to the mean results in a poor EL design that decreases per-driver welfare by \$27.50 per year. This is more than what half of drivers spend on the EL over the course of a year.

The reasoning of this paper extends to any context where agents have individual-specific valuations of a rival, non-transferable good. Depending on how many high-valuation individuals there are and on how high their valuation is, allowing these individuals to pay for the good and rebating their payments might have the potential to produce a Pareto improvement through a targeted policy. An example of these contexts could be waiting times for healthcare or hospital beds. Furthermore, the model and estimation procedure used in this paper can be easily replicated on samples from other cities, especially ones that have congestion policies in place.

Something that the model framework in this paper does not have exogenous variation

to study is, for instance, the extensive margin of commuting. This would be an interesting choice margin to analyze in future research, particularly focusing on how it affects the choice of commuting through other modes and intra-household daily activity planning.

Also worthy of future analyses is the change in car emissions due to changes in traffic congestion. Traveling at a constant speed allows drivers to consume less fuel. On the one hand, Express Lanes are usually less congested and more reliable, so they provide additional benefits to society in terms of reduced emissions. On the other hand, the introduction of Express Lanes might cause the general lanes to become more congested if drivers are not flexible enough to re-adjust, which, in turn, would increase traffic emissions. To fully assess the effect of congestion policies, future analyses should disentangle these effects and evaluate them against travel time savings and departure time preferences.

References

Aguiar, M., and E. Hurst (2007): "Measuring Trends in Leisure: The Allocation of Time Over Five Decades," *The Quarterly Journal of Economics*, 122(3), 969–1006.

Aguiar, M., E. Hurst, and L. Karabarbounis (2013): "Time Use during the Great Recession," *American Economic Review*, 103(5), 1664–96.

Albouy, D., and B. Lue (2015): "Driving to opportunity: Local rents, wages, commuting, and sub-metropolitan quality of life," *Journal of Urban Economics*, 89, 74–92.

Anderson, M. L., and L. W. Davis (2020): "An empirical test of hypercongestion in highway bottlenecks," *Journal of Public Economics*, 187, 1041–97.

Arnott, R., A. de Palma, and C. Lindsey (1993): "A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand," *American Economic Review*, 83(1), 161–79.

Arnott, R., A. de Palma, and R. Lindsey (1994): "The Welfare Effects of Congestion Tolls with Heterogeneous Commuters," *Journal of Transport Economics and Policy*, 28(2), 139–161.

Bartik, A. W., J. Currie, M. Greenstone, and C. R. Knittel (2019): "The Local Economic and Welfare Consequences of Hydraulic Fracturing," *American Economic Journal: Applied Economics*, 11(4), 105–55.

Becker, G. S. (1965): "A Theory of the Allocation of Time," *The Economic Journal*, 75(299), 493–517.

Bento, A., D. Kaffine, K. Roth, and M. Zaragoza-Watkins (2014): "The Effects of Regulation in the Presence of Multiple Unpriced Externalities: Evidence from the Transportation Sector," *American Economic Journal: Economic Policy*, 6(3), 1–29.

Bento, A., K. Roth, and A. R. Waxman (2020): "Avoiding Traffic Congestion Externalities? The Value of Urgency," Working Paper 26956, National Bureau of Economic Research.

Bertanha, M. (2020): "Regression discontinuity design with many thresholds," *Journal of Econometrics*, 218(1), 216–241.

Braid, R. M. (1996): "Peak-Load Pricing of a Transportation Route with an Unpriced Substitute," *Journal of Urban Economics*, 40(2), 179–197.

Buchholz, N., L. Doval, J. Kastl, F. Matějka, and T. Salz (2022): "The Value of Time: Evidence from Auctioned Cab Rides," *Working Paper*.

Buckeye, K. (2012): "Performance Evaluation of I-394 MnPASS Express Lanes in Minnesota," *Transportation Research Record: Journal of the Transportation Research Board*, 2278, 153–162.

Calonico, S., M. D. Cattaneo, and M. Farrell (2020): "Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression-Discontinuity Design," *Econometrics Journal*, 23, 192–210.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82(6), 2295–2326.

——— (2015): "Optimal Data-Driven Regression Discontinuity Plots," *Journal of the American Statistical Association*, 110(512), 1753–1769.

Chui, M., and W. F. McFarland (1987): "The value of travel time: new estimates using speed choice model," *Transportation Research Record*, 1116, 15–21.

De Borger, B., and S. Proost (2012): "A political economy model of road pricing," *Journal of Urban Economics*, 71(1), 79–92.

de Palma, A., R. Lindsey, E. Quinet, and R. Vickerman (2011): *A Handbook of Transport Economics*. Edward Elgar Publishing, Cheltenham, UK.

Deacon, R., and J. Sonstelie (1985): "Rationing by Waiting and the Value of Time: Results from a Natural Experiment," *Journal of Political Economy*, 93(4), 627–47.

Duranton, G., and M. A. Turner (2011): "The Fundamental Law of Road Congestion: Evidence from US Cities," *American Economic Review*, 101(6), 2616–52.

Fox, J. T., K. i. Kim, S. P. Ryan, and P. Bajari (2011): "A simple estimator for the distribution of random coefficients," *Quantitative Economics*, 2(3), 381–418.

——— (2012): "The random coefficients logit model is identified," *Journal of Econometrics*, 166(2), 204–212.

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

Goldszmidt, A., J. A. List, R. D. Metcalfe, I. Muir, V. K. Smith, and J. Wang (2020): "The Value of Time in the United States: Estimates from Nationwide Natural Field Experiments," Working Paper 28208, National Bureau of Economic Research.

Graham, D. (2007): "Variable returns to agglomeration and the effect of road traffic congestion," *Journal of Urban Economics*, 62(1), 103–120.

Green, C. P., J. S. Heywood, and M. Navarro (2016): "Traffic accidents and the London congestion charge," *Journal of Public Economics*, 133, 11–22.

Greenstone, M. (2017): "The Continuing Impact of Sherwin Rosen's "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition"," *Journal of Political Economy*, 125(6), 1891–1902.

Gronau, R. (1973): "The effect of children on the housewife's value of time," *Journal of Political Economy*, 81(2, Part 2), S168–S199.

Hall, J. D. (2018): "Pareto improvements from Lexus Lanes: The effects of pricing a portion of the lanes on congested highways," *Journal of Public Economics*, 158, 113 – 125.

Hall, J. D. (2020): "Can Tolling Help Everyone? Estimating the Aggregate and Distributional Consequences of Congestion Pricing," *Journal of the European Economic Association*.

Hensher, D. A. (2001): "Measurement of the Valuation of Travel Time Savings," *Journal of Transport Economics and Policy*, 35(1), 71–98.

——— (2011): "Valuation of Travel Time Savings," in *A Handbook of Transport Economics*, ed. by A. de Palma, R. Lindsey, E. Quinet, and R. Vickerman, Chapters, chap. 7. Edward Elgar Publishing.

Khoeini, S., and R. Guensler (2014): "Socioeconomic Assessment: Conversion of I-85 High-Occupancy Vehicle to High-Occupancy Toll in Atlanta, Georgia," *Transportation Research Record*, 2450(1), 52–61.

Kreindler, G. E. (2022): "Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications," *working paper*.

Le Barbanchon, T., R. Rathelot, and A. Roulet (2020): "Gender Differences in Job Search: Trading off Commute against Wage*," *The Quarterly Journal of Economics*, 136(1), 381–426.

Lucas, R. E., and E. Rossi–Hansberg (2002): "On the Internal Structure of Cities," *Econometrica*, 70(4), 1445–1476.

Manning, A., and B. Petrongolo (2017): "How Local Are Labor Markets? Evidence from a Spatial Job Search Model," *American Economic Review*, 107(10), 2877–2907.

McFadden, D. (1974): "The measurement of urban travel demand," *Journal of Public Economics*, 3(4), 303–328.

Miller, G., and B. P. Urdinola (2010): "Cyclicality, Mortality, and the Value of Time: The Case of Coffee Price Fluctuations and Child Survival in Colombia," *Journal of Political Economy*, 118(1), 113–155.

Monte, F., S. J. Redding, and E. Rossi-Hansberg (2018): "Commuting, Migration, and Local Employment Elasticities," *American Economic Review*, 108(12), 3855–90.

Moses, L. N., H. F. Williamson, and Jr. (1963): "Value of Time, Choice of Mode, and the Subsidy Issue in Urban Transportation," *Journal of Political Economy*, 71.

Nevo, A., and A. Wong (2019): "The Elasticity of Substitution Between Time and Market Goods: Evidence from the Great Recession," *International Economic Review*, 60(1), 25–51.

Redding, S., and M. Turner (2015): "Transportation Costs and the Spatial Organization of Economic Activity," in *Handbook of Regional and Urban Economics*, vol. 5, chap. Chapter 20, pp. 1339–1398. Elsevier.

Romem, I., and I. Shurtz (2016): "The accident externality of driving: Evidence from observance of the Jewish Sabbath in Israel," *Journal of Urban Economics*, 96, 36–54.

Small, K., and E. Verhoef (2007): "The Economics of Urban Transportation," *The Economics of Urban Transportation*, pp. 1–276.

Small, K. A., C. Winston, and J. Yan (2005): "Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability," *Econometrica*, 73(4), 1367–1382.

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

van den Berg, V., and E. T. Verhoef (2011): "Winning or losing from dynamic bottleneck congestion pricing?: The distributional effects of road pricing with heterogeneity in values of time and schedule delay," *Journal of Public Economics*, 95(7), 983 – 992.

Vickrey, W. (1969): "Congestion Theory and Transport Investment," *American Economic Review*, 59(2), 251–60.

Yang, J., A.-O. Purevjav, and S. Li (2020): "The Marginal Cost of Traffic Congestion and Road Pricing: Evidence from a Natural Experiment in Beijing," *American Economic Journal: Economic Policy*, 12(1), 418–53.

Appendix A. Additional figures and tables for institutional setting

Figure A.1: *Map of the Minneapolis-Saint Paul area showing the Express Lanes discussed in this paper. I-394 is west of Minneapolis; I-35W, south of Minneapolis, connects it to Bloomington; I-35E, north of Saint Paul, connects it to Vadnais Heights. Source: Google Maps.*

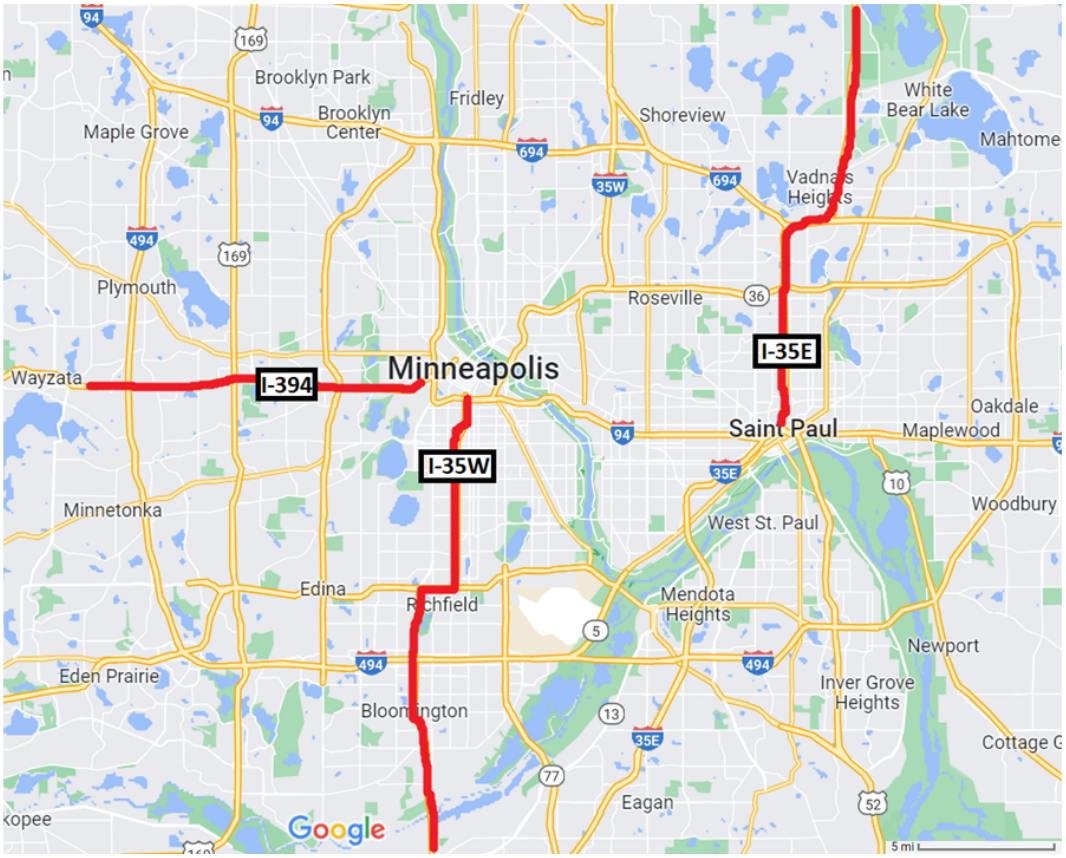


Figure A.2: *Map of I-394 Express Lane. West is located at the top of the map, and North to the right. Source: MnPASS website.*

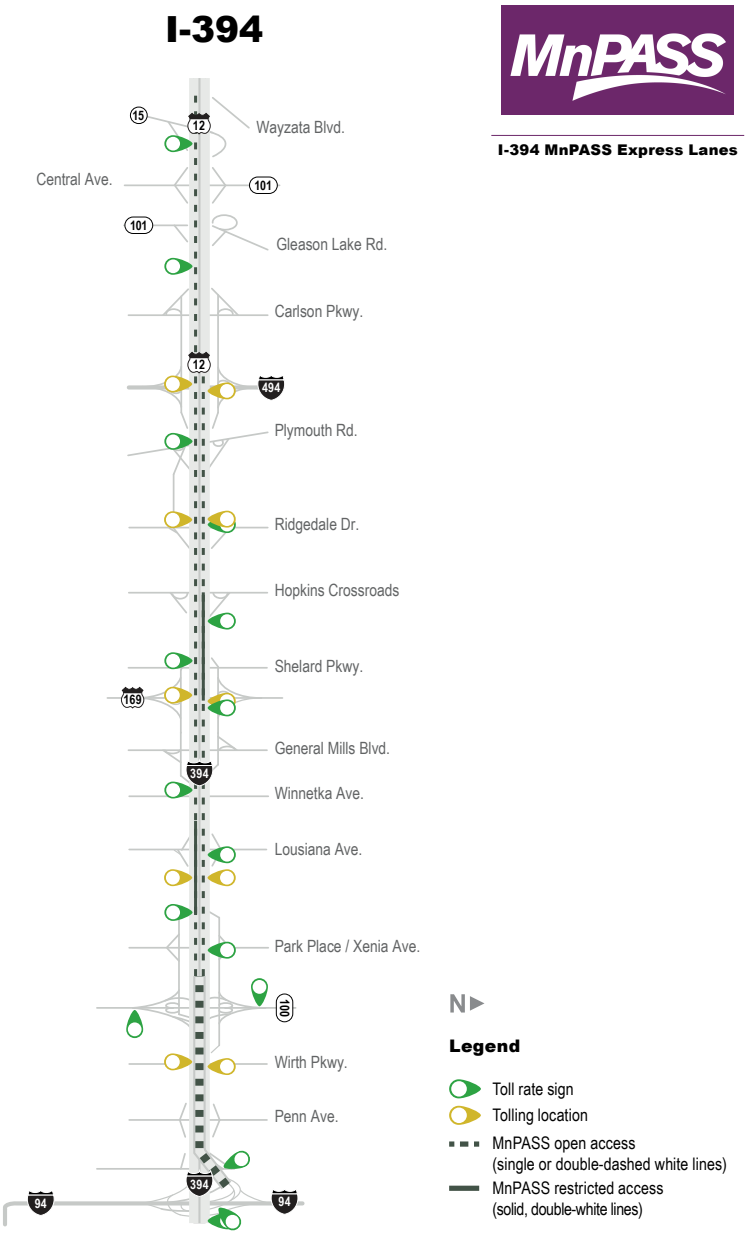


Figure A.3: *Map of I-35W Express Lane. North is located at the top of the map. Source: MnPASS website.*

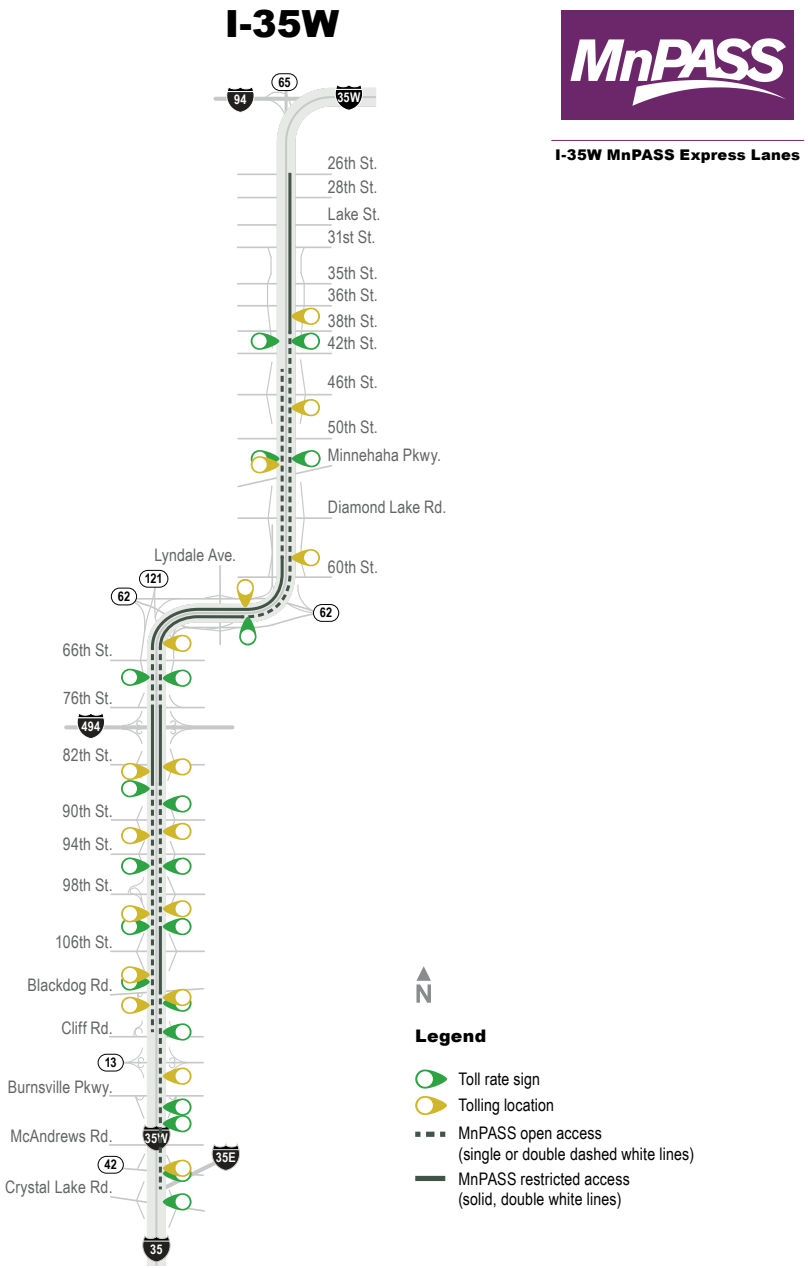
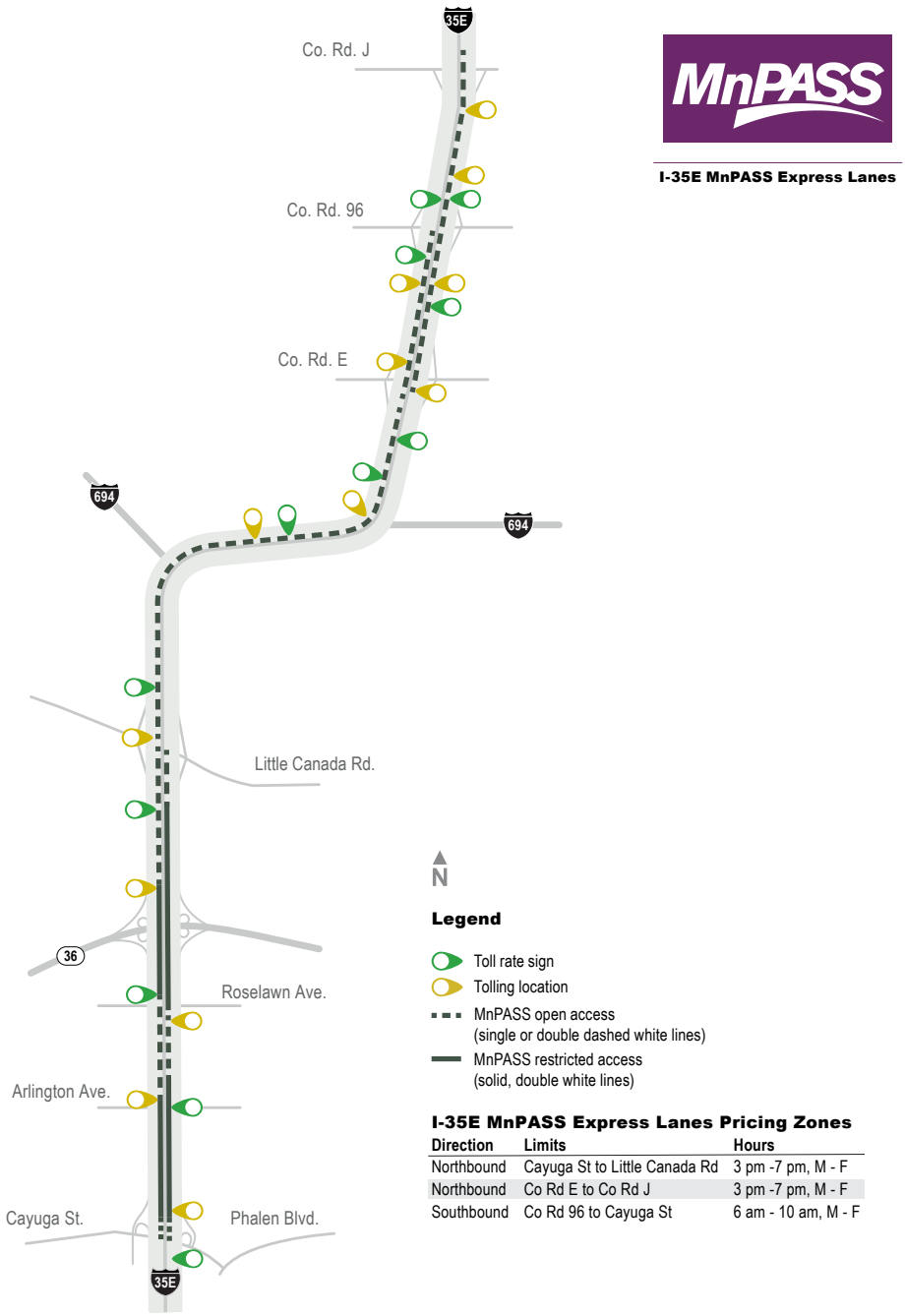


Figure A.4: *Map of I-35E Express Lane. North is located at the top of the map. Source: MnPASS website.*

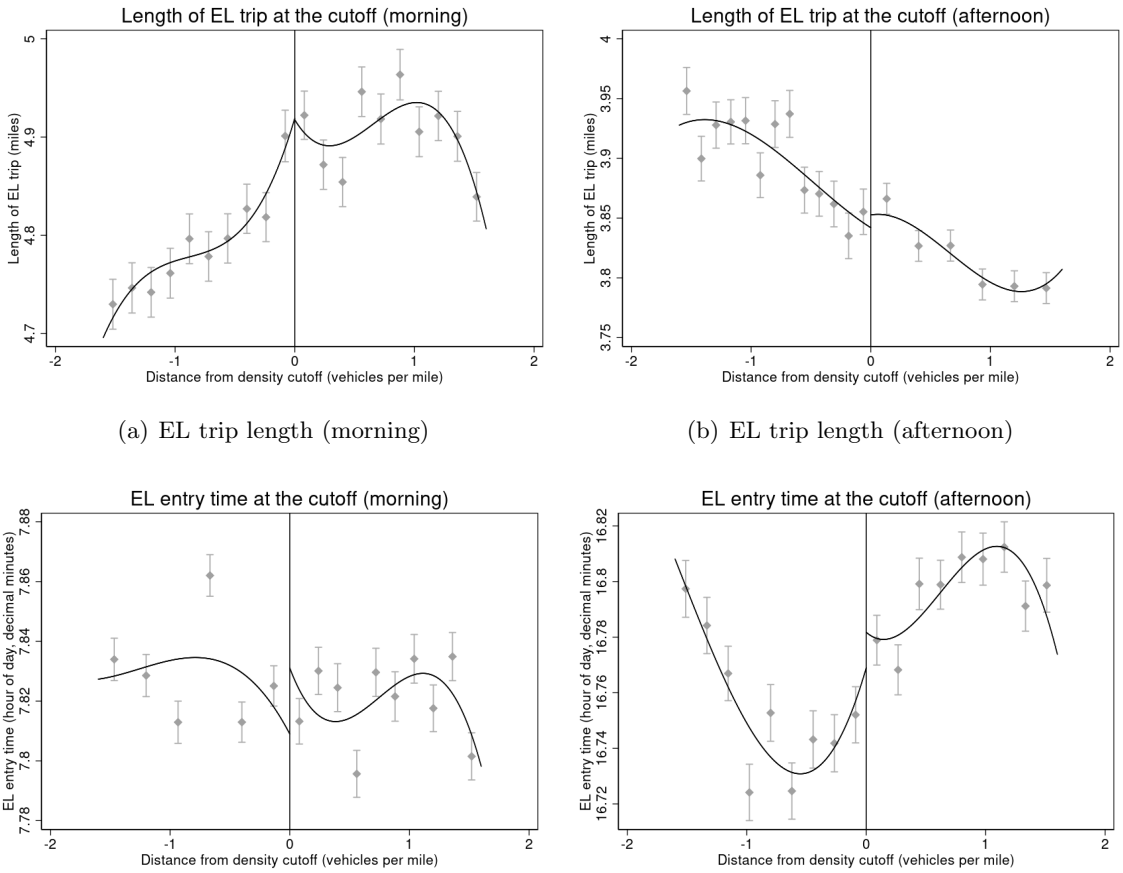


I-35E MnPASS Express Lanes Pricing Zones

Direction	Limits	Hours
Northbound	Cayuga St to Little Canada Rd	3 pm -7 pm, M - F
Northbound	Co Rd E to Co Rd J	3 pm -7 pm, M - F
Southbound	Co Rd 96 to Cayuga St	6 am - 10 am, M - F

Appendix B. Additional figures and tables for RDD analysis

Figure B.1: *Estimated behavior of observable covariates at the cutoff: EL trip length in the morning (a) and in the afternoon (b), and EL entry time in the morning (c) and in the afternoon (d). The x-axis is in traffic density (vehicles per mile). EL trip length is measured in mile. EL entry time is denoted in hour of the day, where the minutes portion is expressed in hundredths. All plots show that observables of drivers on each side of the cutoffs are not significantly different, which provides support for the RDD strategy. The x-axis is in traffic density (vehicles per mile).*



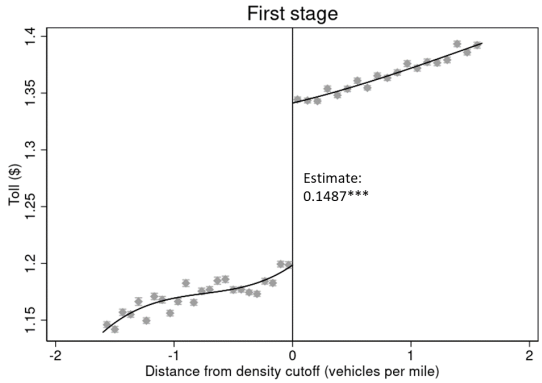
(a) EL trip length (morning)

(b) EL trip length (afternoon)

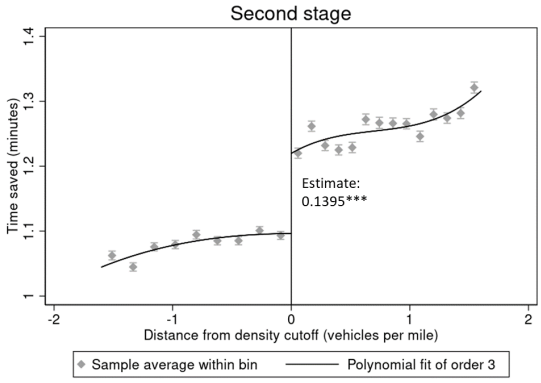
(c) EL entry time (morning)

(d) EL entry time (afternoon)

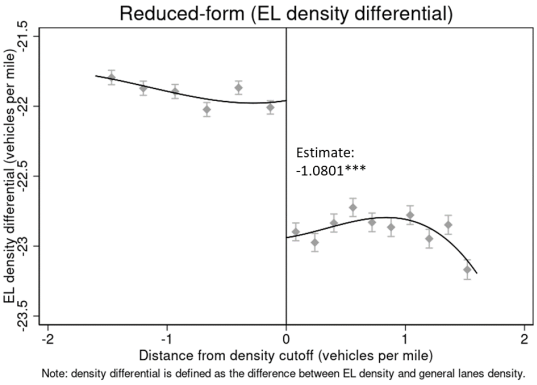
Figure B.2: *First-stage (a) and second-stage (b) results of the RD regression for time saved (b), the EL density premium (c) and the EL speed premium (d) when imperfect toll-density matches are included in the data. The EL density premium is defined as the difference between EL density and general lanes density. The EL speed premium is defined as the difference between the EL speed and the general lanes speed. The first stage shows the toll increase triggered by the traffic density discontinuity. The second stage shows the RDD effects on time saved, the EL density premium and the EL speed premium that are triggered by the first-stage toll increase. In both cases the data is fit using a third-degree polynomial. The gray whiskers are the 95% confidence intervals around each bin. The results imply a VOT of 63.96 $/hour with a standard error of 2.64, which is in line with the RDD results when only perfect toll-density matches are used for estimation. As the bottom panels show, the mechanism is the same as the one explained in the body of the paper.*



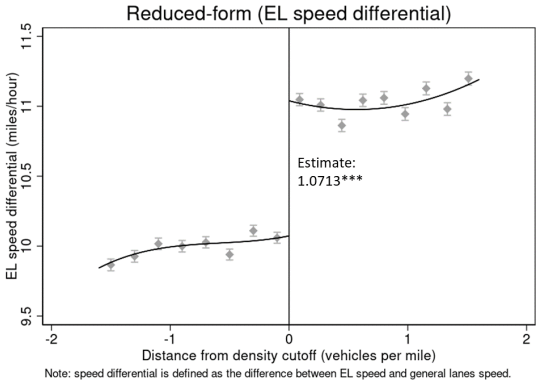
(a) First stage (toll)



(b) Second stage (time saved)



(c) Second stage (density premium)



(d) Second stage (speed premium)

Figure B.3: *Plot of estimated RD effect separately at each cutoff with time saved as the outcome variable. The x-axis represents the cutoffs numbered from 1 to 32. The left y-axis reports the RD effects in minutes. The red horizontal line is the benchmark estimated mean effect equal to 0.2253. The shaded area in bright blue represents the 95% confidence interval around the estimates. The light blue shaded area at the bottom of the plot represents the share of observations at each cutoff relative to the total number of observations in the sample (the values are reported on the right y-axis). The plot shows that most of the observations produce effects that fall around the benchmark estimated mean effect. Where there are few observations, the RD effects are not precisely estimated but they do not carry significant weight towards the computation of the average effect either.*

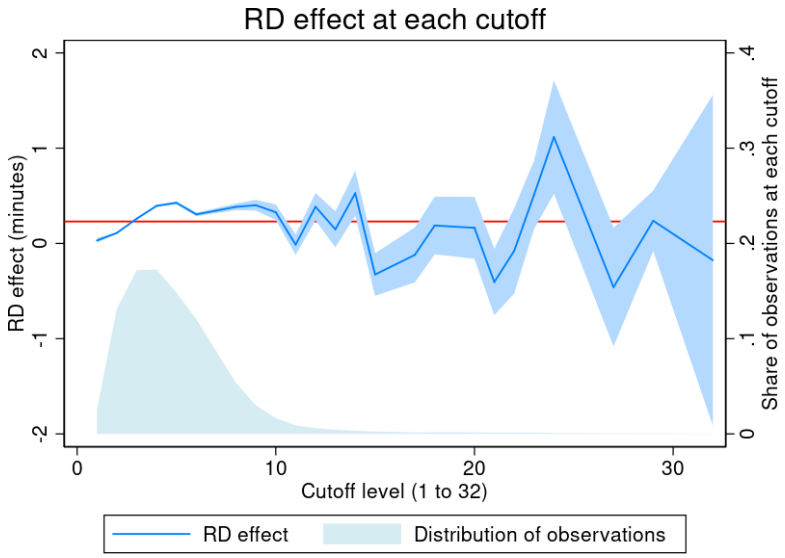


Table B.1: *Second-stage results of regression of time saved on traffic density as the running variable when the analysis is restricted to observations where drivers are on their usual routes. For each driver, the usual route is defined as the combination of their modal entry point and modal exit point. For each road, this yields two peaks and two sections for each peak (except for road I-35E in the morning). Standard errors follow Bertanha (2020). "Section 1" in the morning peak denotes the portion of ELs further away from the city center (either Minneapolis or Saint Paul) and "Section 2" denotes the portion closer to the city center. "Sections 1-2" denotes trips that originated in the suburbs and concluded in the city center. The opposite is true during the afternoon peak.*

	I-394	I-35W	I-35E
Implied VOT morning section 1 ($/hour)	**138.36**	**73.67**	**50.93**
	(23.18)	(6.51)	(3.64)
Implied VOT morning section 2 ($/hour)	**64.91**	**155.06**	
	(13.35)	(60.19)	
Implied VOT morning sections 1-2 ($/hour)	**94.49**	**96.73**	
	(27.3)	(15.72)	
Implied VOT afternoon section 1 ($/hour)	**43.53**	**67.93**	**74.95**
	(5.97)	(11.19)	(28.37)
Implied VOT afternoon section 2 ($/hour)	**66.57**	**34.7**	**77.28**
	(16.18)	(2.74)	(31.1)
Implied VOT afternoon sections 1-2 ($/hour)	**61.17**	**44.21**	**104.43**
	(11.79)	(3.7)	(55.14)
N	727,285	399,751	237,803

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.2: *Correlation between reduced-form VOT results and zipcode level Census data, broken down by location. Average hourly wage uses Census Business Patterns 2018 data from morning destination zipcodes. All other variables use American Community Survey 2018 5-year data from morning origin zipcodes. The underlying assumption is that drivers use the ELs to go to work in the morning and to return home in the afternoon. Data by zipcode is matched to drivers based on their EL entry and exit location.*

Estimated VOT	Average hourly earnings	Median individual income	Median household income	% households over $200k	Median property value	% properties over $1M	Home ownership rate
Road 1:							
56.43	36.29	67,822	104,573	23.43	426,293	3.86	76.72
70.53	48.99	50,643	60,416	8.79	219,760	1.21	69.46
97.99	48.99	52,118	76,569	12.94	320,559	2.06	55.47
44.87	48.99	50,643	60,416	8.79	219,760	1.21	69.46
85.23	36.29	67,822	104,573	23.43	426,293	3.86	76.72
62.03	48.99	67,822	104,573	23.43	426,293	3.86	76.72
Road 2:							
81.93	30.66	38,379	71,781	5.91	231,641	0	67.33
161.43	48.99	58,881	111,697	22.57	349,350	2.13	79.00
100.40	48.99	38,794	90,329	10.74	251,100	0	74.02
73.87	48.99	40,719	66,586	8.15	271,456	1.54	43.50
35.39	30.66	42,030	94,761	12.30	262,342	0.24	77.41
48.93	48.99	42,030	94,761	12.30	262,342	0.24	77.41
Road 3:							
46.76	34.83	37,670	66,413	6.00	238,700	0	65.00
64.73	34.83	41,727	77,708	9.13	244,060	0	73.50
79.53	23.98	41,872	98,888	11.00	266,700	0	89.00
89.40	34.83	41,872	98,888	11.00	266,700	0	89.00

Appendix C. Additional figures and tables for structural estimation

Figure C.1: *Moment fit by time-of-day peak for share of trips with negative time savings. The data moments are in light blue and the simulations by the VOT distribution estimator are in bright blue. The moment is measured in percentage terms.*

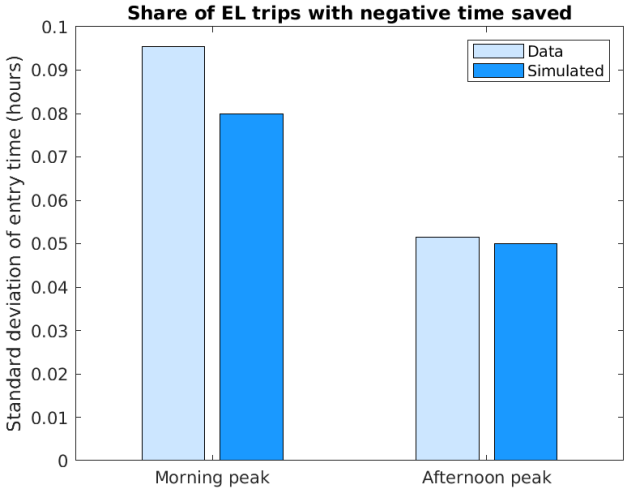
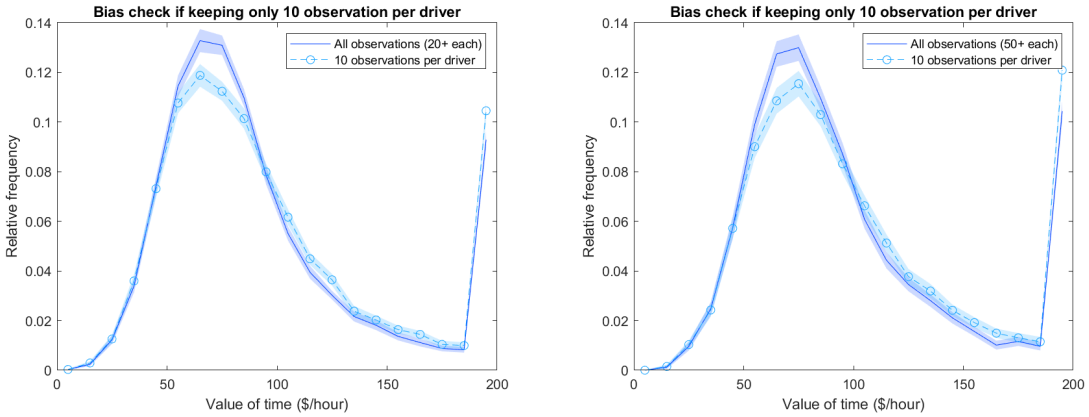


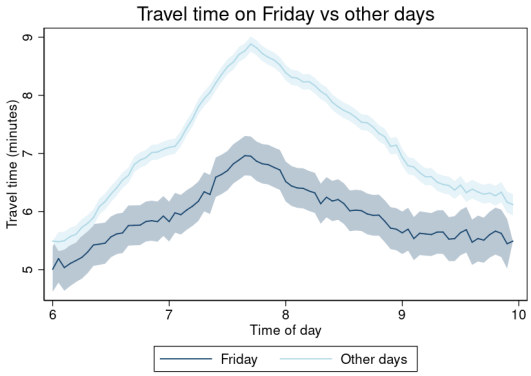
Figure C.2: *Bias check for estimated distributions of VOT for frequent EL drivers that use the EL at least 20 times per year (Panel (a)) and at least 50 times per year (Panel (b)). In both plots, I estimate the VOT distribution using all observations for these individuals (in dark blue) and using only a random subsample of 10 observations per individual (in light blue). The two distributions are similar, which alleviates concerns that using only 10 observations biases the estimation.*



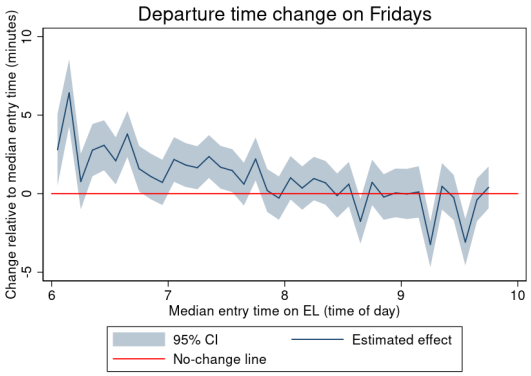
(a) 20+ uses per year

(b) 50+ uses per year

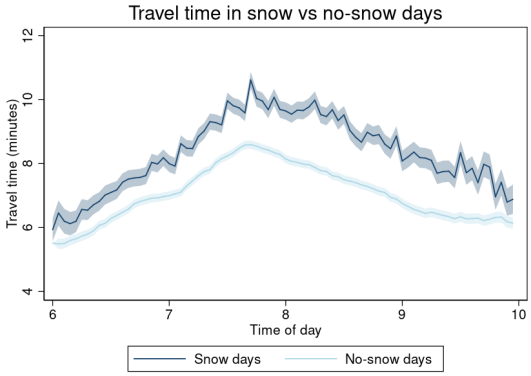
Figure C.3: *Departure time changes of EL users in response to expected changes in travel time. The left panels show the average expected travel time changes on Fridays, snow days and days around holiday weekends. The right panels show the average changes in entry time on the EL relative to the median, for each 3-minute level of the median entry time.*



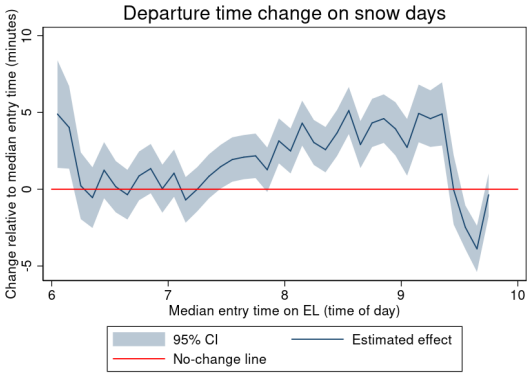
(a) Travel time change on Fridays



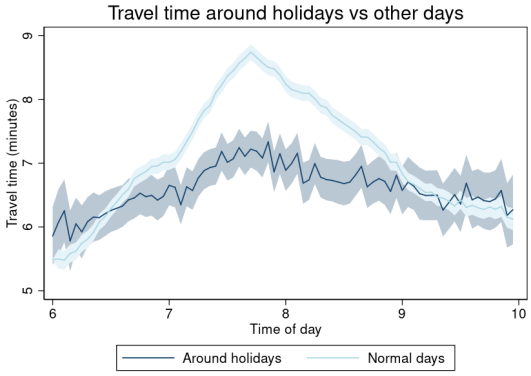
(b) Departure time change on Fridays



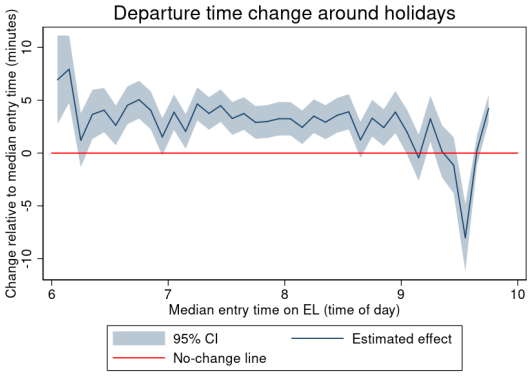
(c) Travel time change on snow days



(d) Departure time change on snow days

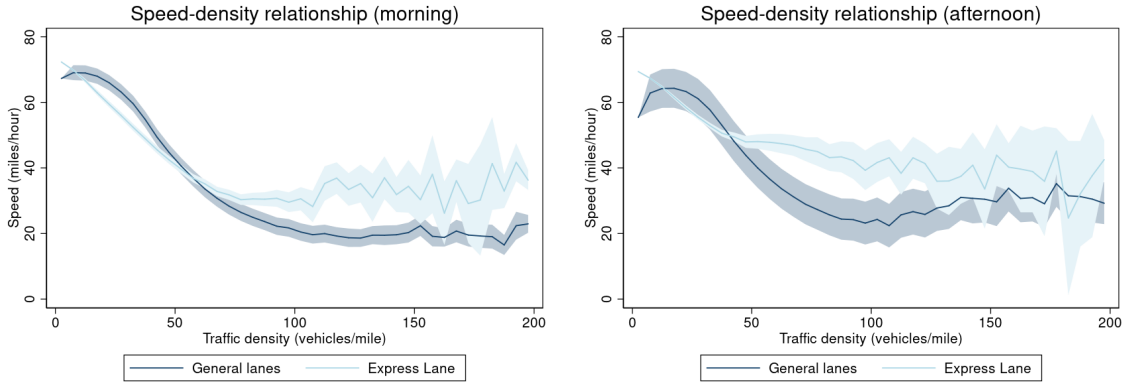


(e) Travel time change around holidays



(f) Departure time change around holidays

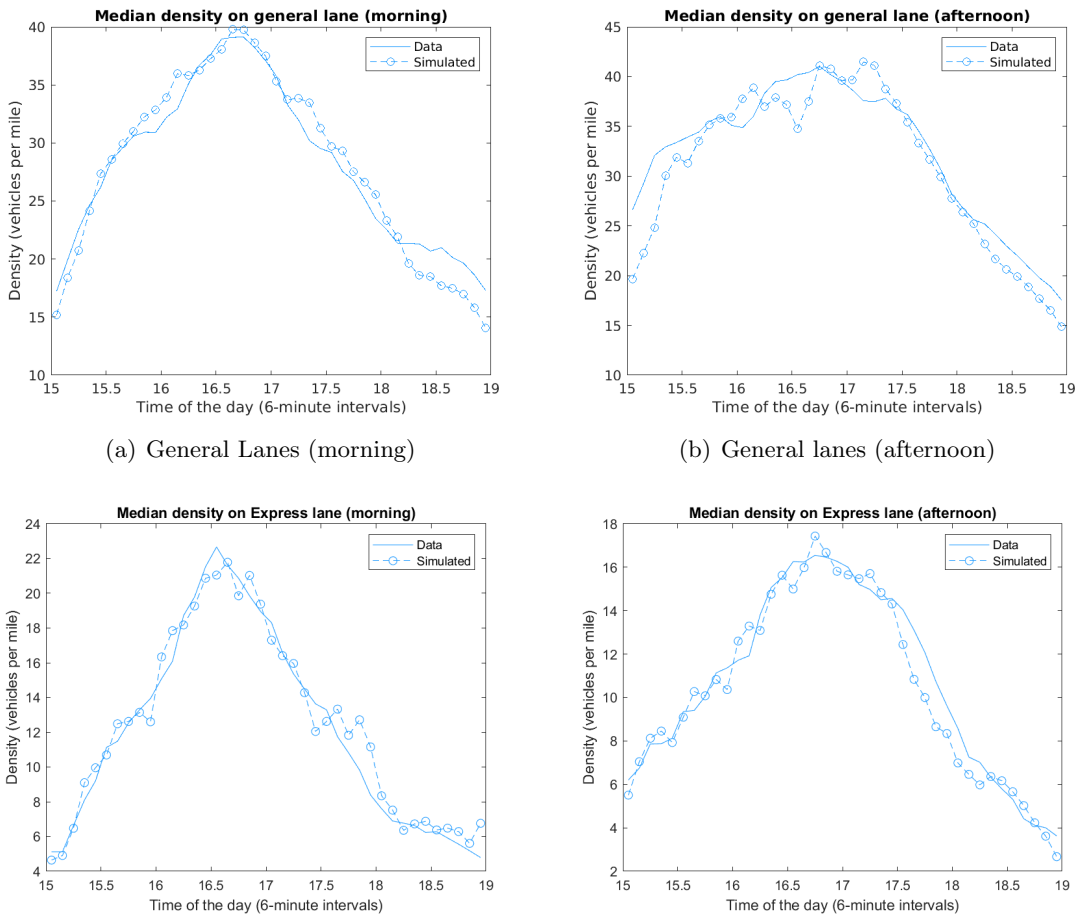
Figure C.4: *Plots of estimated relationship between speed and density by time of day for both the EL (in light blue) and the general lanes (in dark blue). The shaded areas represent the 95% confidence intervals. The relationships have a "flipped S" shape: for low levels of the density, speeds are roughly flat; then they decrease until the density is about equal to 100; finally they plateau for all higher density values.*



(a) Morning

(b) Afternoon

Figure C.5: *Moment fit by road and peak for median traffic density in the general lanes (Panels (a) and (b)) and the EL (Panels (c) and (d)). The horizontal axis is the time of day divided into 40 departure time intervals. The vertical axis reports the density measured in vehicles per mile.*



(a) General Lanes (morning)

(b) General lanes (afternoon)

(c) Express Lane (morning)

(d) Express Lane (afternoon)

Figure C.6: *Moment fit by time-of-day peak for standard deviation of entry time into the EL. The data moments are in light blue and the model simulations are in bright blue. The vertical axis reports the standard deviation of entry time measured in hours.*

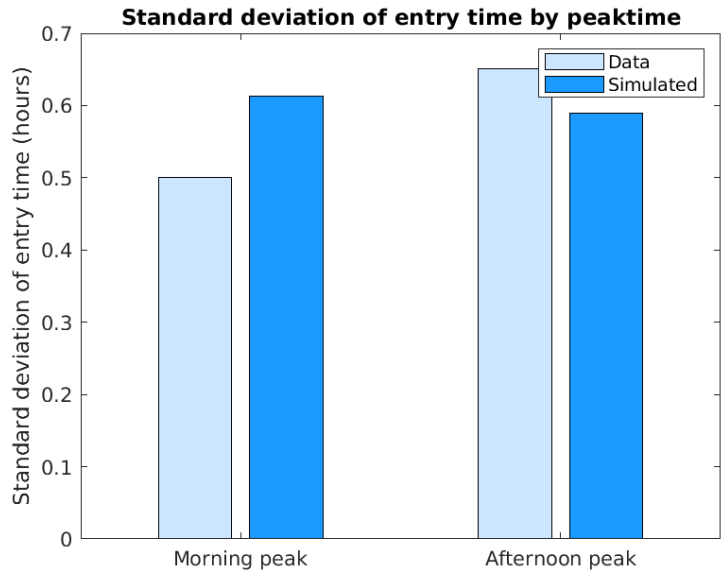


Figure C.7: *Moment fit by time-of-day peak for percentage change in traffic density in the general lanes at the cutoff. The data moments are in light blue and the model simulations are in bright blue.*