

Value Pricing or Lexus Lanes?

The Distributional Effects of Dynamic Tolling*

Cody Cook[†] Pearl Z. Li[‡]

October 10, 2025

Abstract

This paper studies the welfare and distributional effects of dynamically priced highway toll lanes, the most common form of congestion pricing in the United States. To quantify the equilibrium effects of tolling, we develop and estimate a model of driver demand, the road technology, and the pricing algorithm. The demand model features heterogeneous drivers who choose their departure time under imperfect information about travel times and prices. Once the uncertainty is resolved, they then choose whether to take the priced (faster) or unpriced (slower) lanes. We estimate the model using data on toll transactions, historical traffic conditions, and driver characteristics for I-405 in Washington State. We find that tolling a subset of highway lanes increases welfare in aggregate, especially when the newly priced lanes were previously carpool-only. A key component of these gains is the “option value” of tolling: even drivers who infrequently take the priced lanes benefit from having the option to pay for speed when traffic is worse than expected. Moreover, the largest gains accrue to drivers in the bottom income quartile, primarily due to the spatial distribution of lower- and higher-income drivers rather than preference heterogeneity. Finally, we show how simple revisions to the pricing algorithm can increase aggregate welfare and help achieve redistributive goals.

*We are grateful to our advisors, Lanier Benkard, Rebecca Diamond, Liran Einav, Matthew Gentzkow, Paul Oyer, and Paulo Somaini, for their guidance and support. We thank Hunt Allcott, Claudia Allende, Nick Buchholz, Ignacio Cuesta, Gabriel Kreindler, Mark Hallenbeck, Neale Mahoney, Paul Milgrom, Michael Ostrovsky, Shosh Vasserman, Lawrence White, Ali Yurukoglu, and many seminar participants for their helpful comments and feedback. We are also thankful to James Carothers, Manouchehr Goudarzi, Eric Knigge, Mark Morse, Tyler Patterson, and Marie Rogers at the Washington State Department of Transportation, as well as Ryan Avery at WSP, for sharing their data and expertise with us. All errors are our own. Li gratefully acknowledges financial support from the George P. Shultz Dissertation Support Fund and the B.F. Haley and E.S. Shaw Fellowship for Economics via grants to the Stanford Institute for Economic Policy Research, as well as by the Institute for Research in the Social Sciences at Stanford University.

[†]Yale University. codyfcook@gmail.com.

[‡]Yale University. Corresponding author. pearlzli16@gmail.com.

1 Introduction

Traffic congestion has large human and environmental costs, resulting in substantial wasted time, wasted fuel, and air pollution. As early as [Pigou \(1920\)](#), economists have advocated congestion pricing as a way of correcting this negative externality. In the past 25 years, cities have begun charging congestion fees for driving in crowded downtown zones, including Singapore ([Phang and Toh, 2004](#)), London ([Leape, 2006](#)), and, most recently, New York City ([Cook et al., 2025](#)). However, the predominant form of congestion pricing in the United States takes place on highways, with dynamically priced toll lanes adjacent to existing unpriced lanes.¹ By updating prices in response to real-time traffic conditions, these toll lanes keep speeds high and offer paying drivers substantial time savings. Drivers have the option but not the obligation to pay for speed, as they retain the unpriced lanes as a (slower) alternative.

However, there is ongoing disagreement about the effectiveness and distributional impacts of highway toll lanes, which has hampered their expansion. On one side, policymakers refer to dynamic tolling as “value pricing” and emphasize that it provides choice to drivers ([Samdahl et al., 2013](#)). On the other side, opponents are concerned that “Lexus lanes” provide a fast lane for the wealthy at the expense of other drivers ([Astor, 2017](#); [Rosendorf, 2018](#)). Evaluation of these perspectives is an empirical question, and depends on the distribution of driver preferences and the relationship between traffic quantities and travel times, which we refer to as the “road technology.” When one lane becomes tolled, drivers substitute from the newly priced lane into the remaining unpriced lanes, increasing their travel times. High peak-hour prices may also induce drivers to substitute toward driving off-peak, which can reduce average travel times if travel time increases convexly in the number of cars. Finally, since tolling changes the *predictability* of travel times, having the option to take the priced lanes can serve as insurance against worse-than-expected traffic conditions.

In this paper, we study the aggregate and distributional effects of dynamic tolling. We bring together data on toll transactions, driver characteristics, and historical traffic conditions from I-405 in Washington State. We begin by presenting a series of stylized facts documenting heterogeneity in the potential value of the toll lanes for drivers. To quantify the welfare effects of tolling in equilibrium, we build and estimate a model of driver demand, the road technology, and the pricing algorithm. In particular, the demand model incorporates the features of dynamic tolling highlighted above: choices of where and when to drive, uncertainty about prices and travel times, and heterogeneous preferences that vary by income proxies.

Our empirical analysis uses data from the I-405 Express Toll Lanes in Washington State. The tolled part of this corridor features both unpriced general-purpose (GP) lanes and priced high-occupancy/toll (HOT) lanes. Prices are updated dynamically every five minutes and vary across “trip definitions”—where the driver enters and exits the toll lanes—motivating our definition of a market as a group of highway entry-exit pairs with the same trip definition. Drivers who are

¹[Wood et al. \(2021\)](#) report that at the end of 2019, there were 53 highways with toll lanes in the United States, totaling 1858 lane-miles across 16 states. (These numbers do not include toll roads, on which all lanes are priced.) Of the 53 toll lane facilities, 40 used dynamic pricing.

carpooling with others (“high-occupancy vehicle”, or HOV) can access the HOT lanes for free.² Our primary dataset comprises over 14 million I-405 toll transactions from 2019, which we link to proxies for driver income. We also construct a panel of historical traffic conditions using data from sensors embedded in the road, approximately every half mile, which record speeds and volumes at five-minute intervals. We use this data to estimate travel times along the highway. Travel times are highly variable; across days, the 25th-75th percentile range to travel the full length at 8 AM is 37-52 minutes in the GP lanes and 19-33 minutes in the HOT lanes.

We begin by showing suggestive evidence that differences across lower- and higher-income drivers arise primarily from where and when they drive, rather than underlying preference heterogeneity. We use two proxies for driver income: the median household income in her home Census tract and the retail price of her car. By both measures, lower-income drivers tend to belong to “longer” I-405 markets—that is, their travel involves greater distances on I-405. Drivers in these longer markets face an advantageous tradeoff between HOT prices and time savings, as prices increase more slowly in distance than the time saved by taking the HOT instead of GP lane. In contrast, patterns in the toll transaction data suggest that preference heterogeneity along observables is modest. Conditional on when and where they drive, HOT trips taken by higher-income drivers have only slightly lower time savings and higher prices. Moreover, using variation due to prices being rounded to the nearest 25 cents, we show that lower- and higher-income drivers respond similarly to prices being rounded up instead of down.

Since a full welfare analysis requires accounting for the equilibrium effects of dynamic tolling, our next step is to develop a model of driver demand, the road technology, and the pricing algorithm. The model provides a framework for decomposing gains and losses, both across drivers and across channels, and allows us to simulate outcomes under counterfactual policies. For our structural model, we restrict attention to the morning commute, a setting where drivers are likely to make repeated choices with stable preferences.

On the demand side, we extend the classical [Vickrey \(1969\)](#) trip scheduling model to incorporate additional preference heterogeneity and imperfect information on traffic conditions when choosing when to depart. Drivers in the model have heterogeneous preferences over prices, travel times, and how early or late they are to their destinations relative to their ideal arrival times. They make decisions in two stages. In the first stage, each driver chooses a departure time (or the non-405 outside option) given imperfect information about future prices and travel times. In the second stage, she observes the realized price and travel times at her chosen departure time and decides between the priced and unpriced routes. This empirical adaptation of the theoretical model in [Vickrey \(1969\)](#) is similar to that of [Kreindler \(2024\)](#), but the addition of imperfect information allows for a new welfare channel, which we refer to as the “option value” of dynamic tolling. Drivers value the option to reoptimize their choice of lane once they observe the road conditions,

²Carpooling is defined as having at least three occupants during peak hours (5 AM to 9 AM and 3 PM to 7 PM) or at least two occupants during off-peak hours (9 AM to 3 PM). Carpooling drivers must still have a transponder, which can be switched to HOV mode when carpooling. The HOT lanes are free for all drivers between 7 PM and 5 AM.

and even drivers who rarely choose the priced route may still derive value from having the option to do. For example, a driver can choose to leave later without risking arriving late to their destination, because she can pay for speed in the event of worse-than-expected traffic.

Equilibrium travel times and prices are determined by the corridor’s road technology and the pricing algorithm. Our road technology model embeds a static relationship between traffic density and traffic speed inside a discretized model of traffic dynamics in space and time. Dynamics are important because drivers from multiple markets and multiple departure times contribute to the traffic density of a given road segment during a given time interval. Previous research in this vein focuses on the estimation of static road “supply curves” that map quantities into travel times,³ which, when combined with demand curves mapping travel times into quantities, can capture the externalities imposed on other drivers traveling at the *same location* and at the *same time* as the focal driver. In our model, these static congestion costs are amplified in space and time: each driver additionally imposes externalities on others traveling at *other locations* (i.e., on both upstream and downstream road segments) and at *other times*. To incorporate these dynamics, we use a discretized “hydrodynamic” model, which treats the flow of vehicles analogously to the flow of fluids in physical models (Lighthill and Whitham, 1955; Richards, 1956). Finally, for the pricing algorithm, we estimate a flexible function from GP and HOT speeds into prices for each market. This is an approximation of the true algorithm, which we observe, but which is difficult to use in counterfactual simulations.⁴

We estimate the two stages of the demand model jointly using the simulated method of moments. The mean preference coefficients on price and travel time are identified using variation from price rounding and car crashes, which shift prices and travel times when the driver is on the road. We also use morning precipitation, which increases the variance of prices and travel times, as a beliefs shifter to identify drivers’ costs of being early and late to their destinations. Intuitively, if drivers’ disutility of being late is high, they will “buy more insurance” by shifting their departure times earlier on mornings when it rains or snows. Finally, we identify the parameters governing heterogeneous preferences for price and travel time by matching micro moments in the toll transaction data. To focus on commuters, we estimate the model for southbound morning peak hours.

We find that drivers have moderate preference heterogeneity and low scheduling costs. The average driver is willing to pay \$27.02 to avoid one hour of travel time, holding fixed her time early and late. This value of travel time (VOTT) increases in the driver’s tract income and slightly decreases in her car price. The 5th and 95th percentiles of the VOTT are \$19.47 and \$47.45, respectively.⁵ Scheduling costs are relatively low, with the average driver willing to pay \$3.37 to

³This approach dates back to at least Walters (1961). Recent papers have estimated these supply curves using rich cross-sectional (Couture, Duranton and Turner, 2018; Akbar et al., 2023) and time series (Mangrum and Molnar, 2018; Yang, Purevjav and Li, 2020; Russo et al., 2021; Kim, Moon and Yang, 2023) variation.

⁴We use the true algorithm and algorithm inputs to reconstruct the unrounded prices. These unrounded prices are then used to estimate both the price rounding regression discontinuity and the pricing algorithm approximation.

⁵A common benchmark for the value of travel time comes from the U.S. Department of Transportation, which estimates the value of time spent on local personal travel (including commuting) as 50 percent of the median hourly wage (Belenky, 2011). Our mean estimate is 82 percent of the \$32.91 median hourly wage in the Seattle-Tacoma-Bellevue metropolitan statistical area in May 2019 (Bureau of Labor Statistics, 2020).

avoid being one hour early and \$3.82 to avoid being one hour late relative to her ideal arrival time.

In our first set of results, we compare welfare and other equilibrium outcomes with and without tolling, varying whether the unpriced baseline highway has only GP lanes or a subset of HOV lanes, like on I-405 prior to the launch of the HOT lanes. We hold the total number of lanes constant in all equilibria. We also decompose the welfare changes into two channels. The first channel, the ex ante value, captures the value of differentiating the highway “product” and allowing drivers to sort into priced and free lanes based on their heterogeneous preferences. We compute the ex ante value as the welfare change when tolling is available, but drivers must commit to their lane choices in the first stage, under imperfect information. The second channel, the option value, represents the additional value drivers gain from choosing between priced and free lanes in the second stage, following the realization of uncertainty about prices and travel times. When uncertainty is higher, so is the option value of HOT lanes.

We find that tolling a subset of the highway lanes is welfare-improving, especially when tolling is added to previously HOV-only lanes. Converting the underutilized HOV lanes to HOT substantially expands the highway capacity available to non-carpooling drivers, increasing aggregate welfare by \$117 per driver-year, which includes \$46 in revenue per-driver year that we assume is redistributed uniformly. Average I-405 travel times decrease as drivers spread out among the GP and HOT lanes, even as overall highway traffic volume increases. Tolling a subset of the general-purpose lanes (without allowing for an HOV discount) increases aggregate welfare more modestly by \$14 per driver per year. I-405 travel times also decrease, this time because tolling functionally reduces the unpaid highway capacity available, inducing some marginal drivers to substitute to the non-405 outside option. The option value is \$11 to \$13 per driver per year, an important component of the gains from tolling that does not vary much with the choice of unpriced baseline. In contrast, the ex ante gains from tolling are much larger when tolling is used to correct the inefficient underutilization of the HOV lane (+\$105 per driver per year) than when tolling is applied to GP lanes (+\$0.12).

Moreover, these are not Lexus lanes: tolling increases welfare across the income distribution, and the greatest gains accrue to drivers in the lowest income quartile. Drivers in the bottom 25 percent of the tract income distribution account for 38 percent of the aggregate gains from the HOV-to-HOT conversion and 39 percent of the aggregate gains from the GP-to-toll conversion. Low-income drivers benefit from both high ex ante values and high option values of tolling. As suggested by the descriptive evidence, differences in the ex ante value are largely driven by the favorable trade-off between price and time savings faced by low-income drivers, who disproportionately travel the longest distances on I-405. Low-income drivers also have high option values of tolling due to differences in their preferences. Since low-income drivers are more responsive to both price and travel time, they benefit more from the ability to make their lane choices under full information.

In our last set of results, we simulate equilibrium outcomes under alternative pricing policies. Raising the price ceiling from \$10 to \$12 is modestly welfare-improving both in aggregate and across the income distribution.⁶ The higher price ceiling allows the algorithm more flexibility with which

⁶The existing algorithm generates unrounded prices between 50 cents and \$12. We consider a \$12 price ceiling

to manage high-traffic periods. While lower-income drivers in longer markets face slightly higher prices than under the status quo, these prices are still far from being proportional to the HOT time savings. We also explore two forms of income-based pricing that are under consideration by the tolling authority: a 50 percent proportional discount and a \$2 flat discount for low-income drivers (Washington State Transportation Commission, 2021). We find that both forms of income-based pricing reduce toll revenues and increase HOT travel times, reducing aggregate welfare relative to the status quo. However, to the extent that transportation costs (in both time and money) already fall most heavily on low-income drivers, policymakers may still value the redistributive potential of income-based pricing.

These results contribute novel empirical evidence on the efficiency and equity of highway toll lanes, the most common form of congestion pricing in the US. A rich theoretical literature has explored the heterogeneous welfare effects of highway tolling, extending the Vickrey (1969) model of endogenous trip scheduling to incorporate heterogeneous drivers (Arnott, de Palma and Lindsey, 1994), “value pricing” of a subset of the highway (Small and Yan, 2001; Verhoef and Small, 2004; Small, Winston and Yan, 2006), and pricing when there is ‘hypercongestion’ (Hall, 2018). Others use the variation from the highway toll lanes to estimate drivers’ value of time (Small, 2012; Mattia, 2022)⁷ and the related value of “reliability,” i.e., preferences for the *variance* of travel times (Lam and Small, 2001; Small, Winston and Yan, 2005; Brent and Gross, 2018; Bento, Roth and Waxman, 2024).⁸ The literature has historically treated reliability as a reduced-form product characteristic. Our two-stage demand model microfounds preferences for reliability from drivers’ first-stage departure time choice under imperfect information: drivers value reliability because it reduces their realized costs of being early/late in the second stage. Incorporating imperfect information also introduces an option value of tolling, which we estimate to be quantitatively important, both in aggregate and for low-income drivers. Separately, by linking the rich transaction-level data on driver decisions to characteristics of the same drivers, we identify who gains and loses from the introduction of HOT lanes.

A broader empirical literature studies the effectiveness and incidence of congestion pricing in urban settings, rather than on highways. Research in this area has evaluated real-world urban congestion pricing (Phang and Toh, 2004; Leape, 2006; Börjesson et al., 2012; Cook et al., 2025) and quantified the aggregate welfare effects of counterfactual cordon pricing (Tarduno, 2022), time-of-day pricing (Kreindler, 2024), distance-based and personalized pricing (Durrmeyer and Martinez, 2022; Almagro et al., 2024), and congestion surcharges on taxis and ridesharing (Arora, Zheng and Girotra, 2020; Rosaia, 2023). Two recent papers have also evaluated cordon pricing (Herzog, 2024) and distance-based pricing (Barwick et al., 2024) in longer-run equilibria incorporating both

in order to avoid extrapolating our approximation of the pricing algorithm beyond the data.

⁷Recent papers have also estimated the value of travel time in other settings, including residential choice (Su, 2022) and ridesharing (Goldszmidt et al., 2020; Buchholz et al., 2025).

⁸For more on the effects of highway toll lanes, see theoretical work by Dahlgren (1998); Yang and Huang (1999); Konishi and Mun (2010) and empirical evidence from surveys and calibrated traffic models by Brownstone and Golob (1992); Safirova et al. (2004); Hall (2021). Hallenbeck et al. (2019) also conduct a descriptive analysis of distributional effects using toll transaction data.

residential and mode (e.g., driving vs. public transit) choices. They find that low-income commuters gain the most from these other forms of congestion pricing, mirroring the results in our paper.⁹ Unlike with HOT lanes, cordon-based congestion pricing schemes apply to all entries into the priced region, and drivers generally do not have the option between saving time versus money.

2 The I-405 Express Toll Lanes

Our empirical setting is Interstate 405 (I-405) in Washington State, a well-trafficked highway which connects several Seattle-area suburbs. Its tolled section, mapped in Figure B.14, runs north-south for seventeen miles between Lynnwood and Bellevue, east of and across Lake Washington from Seattle proper. Along this corridor, there are typically one to two high-occupancy/toll (HOT) lanes and three to four unpriced general-purpose (GP) lanes in each direction. The priced lanes run directly adjacent to the unpriced lanes, separated by double white lines.

The highway is designed so that drivers have a relatively accurate picture of current prices and travel times (Figure A.1). Drivers may only enter or exit the toll lanes at designated access points, which are typically available between every few interchanges, roughly every few miles.¹⁰ As drivers approach each access point, they are shown the current prices on two sequential electronic signs. Some but not all HOT access points also have electronic signs which show estimated travel times in both the priced and unpriced lanes.

Prices vary dynamically across time, and are subject to both rounding and a floor and ceiling. Tolling is in effect from 5 AM to 7 PM on weekdays, excluding major federal holidays.¹¹ The pricing algorithm is designed to maintain HOT speeds of at least 45 miles per hour for 90 percent of tolled hours. That is, the objective of I-405 tolling is explicitly to reduce congestion rather than to raise revenue. New prices are computed every five minutes using data on speed and throughput from induction loops embedded in the pavement.¹² Importantly for our identification strategy, while the pricing algorithm initially calculates continuous toll rates, the prices actually faced by drivers are rounded to the nearest 25 cents. Prices are also subject to both a 75-cent floor and a \$10 ceiling; we explore raising the ceiling in one of our policy counterfactuals.

Within a five-minute interval, prices also vary across “trip definitions”—where the driver enters and exits the toll lanes—which form the basis of our market definition. The algorithm independently computes prices for each trip definition. In particular, prices are non-additive across road segments, so that the A to C toll is not necessarily equal to the sum of the A to B toll and the B to C toll.

⁹Other papers have estimated that a distance-based tax (i.e., a tax on vehicle miles traveled) would be regressive, but less regressive than existing gas taxes (Martin and Thornton, 2018; Glaeser, Gorback and Poterba, 2023).

¹⁰In practice, drivers sometimes illegally cross the double white lines separating the GP and HOT lanes rather than entering and exiting solely at the designated access points. Most of the access points are short segments on which the double lines are replaced with single dashed lines. There are also two ramps which provide direct access to the HOT lanes from an interchange.

¹¹Outside these hours, even single-occupancy vehicles can drive in the HOT lanes for free.

¹²Dynamic pricing is also used in other settings, including Uber surge pricing (Castillo, 2023; Castillo, Knoepfle and Weyl, 2025) and revenue management in the airline (Williams, 2022), railroad (D'Haultfoeuille et al., 2022), and hotel (Cho et al., 2018) industries.

We therefore define a market as a group of highway entry-exit pairs which share the same trip definition.¹³ In each direction, there are fourteen markets, of which three do not have a feasible HOT route. Appendix B.1 provides more details.

Paying vehicles have two ways to pay tolls, though we abstract away from this distinction in the model. Most drivers pay using online accounts which are linked to transponders inside their vehicles. However, drivers without transponders can still take the HOT lanes: their license plates are photographed and they receive bills in the mail. An additional \$2 fee is charged for “paying by plate” in this way. Non-paying vehicles include buses, motorcycles, and high-occupancy vehicles (HOV), which can drive in the HOT lanes for free. Drivers in this last group must have their transponders set to the HOV setting.

When they launched in September 2015, the I-405 HOT lanes replaced existing HOV-only lanes. However, the opening of the HOT lanes bundled together several policy changes. First, along the entire corridor from Lynnwood to Bellevue, previously HOV-only lanes were converted to HOV-plus-toll (i.e., HOT) lanes. Second, an entirely new HOT lane was built between Woodinville and Bellevue, on the higher-trafficked southern half of the newly tolled section. Third, occupancy requirements were increased. Vehicles with two occupants, which were previously permitted to access the HOV lanes for free, were now required to pay the same toll as single-occupancy vehicles during weekday peak hours. Vehicles with three or more occupants continued to drive for free at all times.¹⁴

Appendix C.3 documents how average speed and throughput changed in the years surrounding the opening of the HOT lanes. Aggregate speed and throughput increased in both the always-unpriced GP lanes and the newly priced HOT lanes in the years after the HOT opening. In particular, average throughput grew by a modest 5 percent in the GP lanes and a more substantial one-third in the HOT lanes, suggesting the previous HOV-only lanes were underutilized. However, because multiple policy changes were bundled together in the first year after the opening, it is difficult to isolate the effect of pricing. These limitations, in part, motivate the need for our structural model.

3 Data

We combine data on toll transactions, historical traffic conditions, and driver characteristics. Several datasets were obtained in partnership with the Washington State Department of Transportation (WSDOT) Tolling Division, which administers the I-405 HOT lanes. Our primary estimation sample is the calendar year 2019, the intersection of our data samples. Appendix B provides additional

¹³We assume that the trip definition for a given market is the one in which drivers take the HOT lanes for as long as possible between their highway entry and exit. Each trip definition is associated with a unique market, and most markets are associated with a single trip definition. There are two exceptions in which there is a nontrivial trade-off between two trip definitions in the same market, where one typically has both higher prices and higher HOT time savings. For the model, we say there are two HOT routes to choose from in these two markets.

¹⁴Initially, the HOT lanes were tolled at all hours, including nights, weekends, and holidays. The current tolled hours were introduced in March 2016.

details on data construction.

3.1 Toll transactions

We obtain data on the universe of I-405 HOT transactions in 2019 from WSDOT. For each transaction, we observe the trip definition, HOV status, price paid, and timestamps for each toll gantry driven under during the trip. Each transaction is linked to an account identifier; we match a subset of accounts to the Census tract of the account holder’s billing address and the make, model, and year of her vehicle.¹⁵

Table 1: Transaction summary statistics

	All			Peak hours			Peak hours + income proxies		
	p25	p50	p75	p25	p50	p75	p25	p50	p75
Paid trips									
Time saved (mins)	0.15	1.93	6.90	0.69	4.22	9.23	0.78	4.40	9.47
Price paid (\$)	0.75	0.75	4.75	0.75	2.75	6.50	0.75	3.00	6.50
Price per min saved	0.47	0.91	5.69	0.46	0.73	2.11	0.46	0.72	1.95
# trips	9,633,458			6,160,687			4,940,253		
# unique drivers	1,123,028			746,706			530,000		
HOV trips									
Time saved (mins)	0.11	1.58	6.76	0.65	4.32	9.46	0.64	4.29	9.44
# trips	4,486,472			2,786,645			2,358,260		
# unique drivers	264,779			196,511			160,902		
Totals									
# trips	14,119,930			8,947,332			7,298,513		
# unique drivers	1,223,449			833,373			598,915		

Note: Each observation is a toll transaction (paid or HOV). Southbound peak hours are 5 AM to 11 AM; northbound peak hours are from 1 PM to 7 PM. The third set of columns, labeled “Peak hours + income proxies,” restricts to paid, peak-hour transactions which are matched to at least one of a tract income or a car price.

Table 1 reports summary statistics of paid and HOV transactions. Our analysis restricts attention to peak hours, which we define as southbound 5 AM to 11 AM and northbound 1 PM to 7 PM. For paid trips, the sample covers 6.1 million transactions by about 747,000 unique drivers, the median time saved in the HOT lanes is 4.2 minutes, and the median price paid is \$2.75. We identify the preference heterogeneity parameters in the structural model by matching empirical covariances between income proxies and these trip attributes. Further restricting to paid transactions that are matched to a tract income or a car price results in 4.9 million transactions by 530,000 unique drivers. Approximately 32% of trips are made by HOV drivers, who can use the lanes for free. The median HOV trip saves less time than the median paid trip in the full sample, but this difference

¹⁵52 percent of HOT drivers are matched to a tract income; 39 percent are matched to a car price.

is largely absent for peak-hour trips. The total number of drivers is less than the sum of unique HOV and paid drivers, as drivers can toggle their HOV status across trips depending on whether they have passengers. About 22% of drivers ever use the lanes as an HOV driver.

3.2 Historical traffic conditions

We construct a panel of historical traffic conditions at the five-minute level and a panel of potential market sizes at the daily level. Together, these two datasets describe the inputs to driver decision-making. Table A.1 reports market lengths and median travel times and prices in each southbound market, separately for peak and off-peak hours. Figure A.2 shows how travel times and prices vary by time of day and across days in two example markets, one short and one long.

Prices and pricing algorithm inputs. We obtain historical rate cards and pricing algorithm inputs from September 2015 to March 2022 from WSDOT. The rate cards show realized prices which have been rounded to the nearest 25 cents. While WSDOT stores only the rounded prices, we recover the algorithm-generated *unrounded* prices—which we will use in a regression discontinuity design—by rebuilding the algorithm based on its source code, also shared by WSDOT. The algorithm inputs are a processed version of the induction loop data, which we describe next.

Traffic sensor data. We add I-405 traffic sensor data from January 2011 to December 2021, downloaded from the Washington State Transportation Center (TRAC)’s online TRACFLOW tool. The sensors are induction loops embedded in the each lane, spaced out roughly every half mile on the mainline, as well as on highway on- and off-ramps.

Each loop reports its average speed and throughput in each five-minute interval.¹⁶ The loop’s throughput is the number of vehicles passing over it per unit of time. We are also interested in traffic density, the number of vehicles per unit of distance, which the loops do not measure directly. We approximate density (cars per lane per mile) at each loop in each five-minute interval by dividing throughput (cars per lane per hour) by speed (miles per hour), following the transportation literature (Hall, 2005). Appendix B.2 validates this approximation empirically; Appendix D.2 discusses the theoretical relationships between speed, density, and throughput.

To turn speeds on road segments into market- and route-level travel times, we assume that travel speed is constant within each segment (between two adjacent loops) and five-minute interval. When computing HOT travel times, we additionally account for time spent in the GP lanes, for example between the highway on-ramp and the HOT access point. Since markets group multiple highway entry-exit pairs, we define the GP and HOT travel time in a market as the travel time between the first highway entry and the last highway exit in the market.

Lastly, to compute the quantity of drivers departing (i.e., entering I-405) in each five-minute interval in each *market origin* (not market), we sum the throughputs on all on-ramps for that market origin.¹⁷

¹⁶The loop speeds are top-coded at 60 miles per hour; we replace top-coded speeds with freeflow speeds estimated from the toll transaction data, which vary at the (loop, day of week, hour of day) level.

¹⁷An added complication is that each on-ramp contains both metered (general-purpose) and unmetered (carpool-

Tract-to-tract travel flows. We use historical travel flows between pairs of Census tracts aggregated from cell phone GPS pings in 2019 to construct potential market sizes. These flows include all forms of travel between a given tract pair—not only travel on I-405 or even only travel by car. We use GraphHopper’s routing API to group together pairs of tracts for which a particular entry-exit pair on I-405 is one of the top three suggested routes. This procedure collapses *realized* flows between Census tracts into *potential* flows between I-405 entries and exits, which become our estimates of potential market sizes. Appendix B.4 discusses the procedure in more detail. The resulting market sizes are at the (market, date) level.

Other price and travel time shifters. We use car crashes as additional shifters of *realized* prices and travel times and precipitation as a shifter of drivers’ *beliefs* about prices and travel times. The data on car crashes, obtained via a public disclosure request from WSDOT, contain the date and time, the milepost, the location of first impact (i.e., which lane or shoulder, which we use to determine whether the crash happened in the GP or HOT lanes), and other characteristics for each crash in 2019. The precipitation data were collected by Automated Surface Observing Systems (ASOS) and downloaded from Iowa State University’s Environmental Mesonet. We use hourly precipitation at the Everett weather station in 2019.¹⁸

3.3 Driver characteristics

Our primary driver characteristics are Census tract median household incomes and car prices, two proxies for income. We take tract-level demographics, including income distributions, from the 2019 American Community Surveys (ACS). We estimate the manufacturer-suggested retail price (MSRP) for each combination of car make, model, and year using data on vehicle registrations from March 2017 to December 2022 from the Washington State Department of Licensing (WSDOL). Appendix B.3 describes our procedure for estimating vehicle MSRPs.

We construct two samples of drivers, the HOT sample and the unconditional sample of *potential* I-405 drivers. The HOT sample contains real-world drivers who are observed taking the I-405 toll lanes at least once in 2019. These real-world drivers are linked to their tract incomes and car prices via their billing addresses and car make-model-years. The unconditional sample contains simulated drivers, drawn from the joint distribution of tract incomes and car prices in the population of potential I-405 drivers. To construct this joint distribution, we sample Census tracts, weighting by the tract-to-tract travel flows inferred from the GPS data, then sample vehicle registrations conditional on tract; Appendix B.4 describes this procedure in more detail.

only) lanes, and the metered throughputs reflect rationed demand for departure times. Within each day and ramp, we reallocate the metered throughputs so that they match the profile of unmetered throughputs. That is, we assume that single-occupancy drivers demand the same distribution of departure times as carpools drivers. Appendix B.5 discusses this procedure and assumption in more detail.

¹⁸ASOS is jointly operated by the National Weather Service, the Federal Aviation Administration, and the Department of Defense. The Everett weather station is about 11 miles north of I-405’s northern terminus in Lynnwood.

4 Descriptive evidence

In this section, we present a series of stylized facts about our empirical setting and discuss how they motivate the key features of our model.

Fact 1: Travel times and prices are highly variable

For the full-length southbound market, the median HOT travel time is half the median GP travel time during peak hours, and the median price is at the \$10 cap (Figure 1). However, there is a lot of day-to-day variation. Between 8-8:05 AM, the 25th–75th percentile range of travel times is 37–52 minutes for GP lanes and 19–33 minutes for HOT lanes. While the HOT price in this full-length market hit the cap at 8 AM on the majority of days, the 25th percentile of prices is \$7.75.

This variation is not unique to the full-length market, and drivers in all markets face uncertainty about prices and travel times when choosing when to get on the road, especially at peak hours. Even conditional on season, day of week, and 5-minute interval, the coefficient of variation—i.e., the ratio of the standard deviation to the mean—for the average market is 0.23 for GP travel times during peak hours. For HOT lanes, dynamic pricing helps reduce some of the variation in travel times—the average coefficient of variation is 0.18—but prices themselves are highly variable, with an average coefficient of variation of 0.41.¹⁹ While some drivers may use modern tools such as Google Maps to monitor traffic conditions, route requests for departure times in the future show a range of likely travel times rather than a precise estimate. (For example, Figure A.6 shows a screenshot of a Google Maps search at 8:30 AM for a full-length trip departing at 9:30 AM with an estimated travel time of “typically 30 min to 1 hr”.) This variation motivates the design of our structural model, in which commuters choose when to drive with imperfect information on traffic conditions and can then re-optimize between HOT and GP lanes once entering the road and observing traffic conditions and HOT prices.

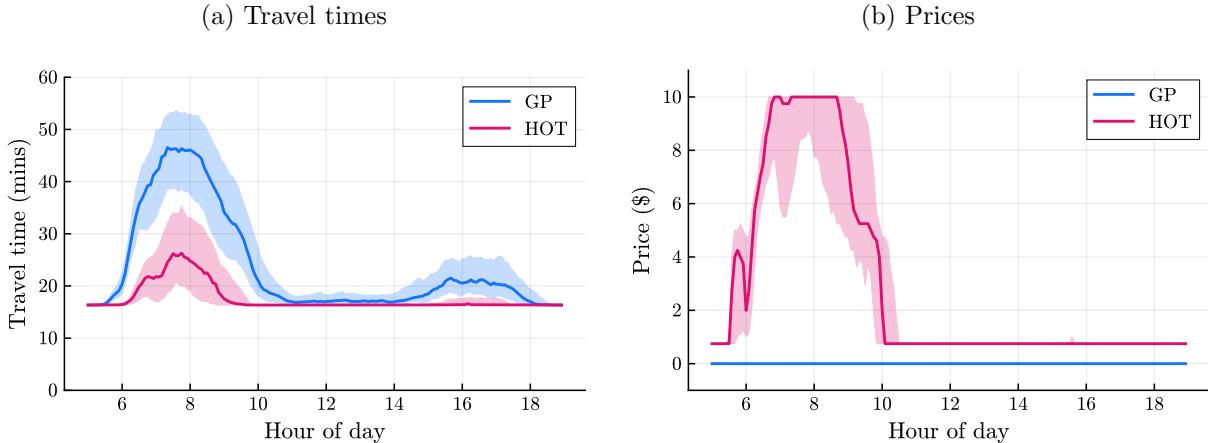
Fact 2: The potential value of HOT lanes is greater in longer markets, which disproportionately serve lower-income drivers

HOT lanes are better deals for commuters in markets that traverse longer stretches of I-405. Relative travel times increase more slowly with distance in the HOT lanes than in the GP lanes, such that HOT time *savings* are greatest in the longest markets.²⁰ Figure 2a shows that from 7–8 AM—the height of the southbound morning peak—the median time savings are less than two minutes in the four shortest southbound markets with feasible HOT routes, compared to nearly twenty minutes in the full-length market. However, drivers in the longest markets do not pay proportionally higher tolls. In Figure 2b, we show that the price per minute saved *decreases* in market length, akin to a bulk discount. The median price per minute saved is \$14.65 in one of the shortest markets and only \$0.48 in the full-length market.

¹⁹Figure A.5 plots the average coefficients of variation by market-hour. The coefficients of variation *unconditional* on season, day of week, and 5-minute interval are each about twice as large as those presented here.

²⁰Unsurprisingly, travel times in both the GP and HOT lanes are the greatest in these long markets, with both means and variances of travel times increasing in length of highway traveled (Figure A.4, panels a and b).

Figure 1: Prices and travel times by time of day (full-length SB market)



Note: This figure documents travel times and prices for the full-length southbound market. Thick lines indicate across-day medians and shaded areas are between the 25th and 75th percentiles. Each underlying observation is a (route, 5 min, day) from 5 AM to 7 PM (tolled hours) in 2019 for trips from Lynnwood to Bellevue, the longest southbound market.

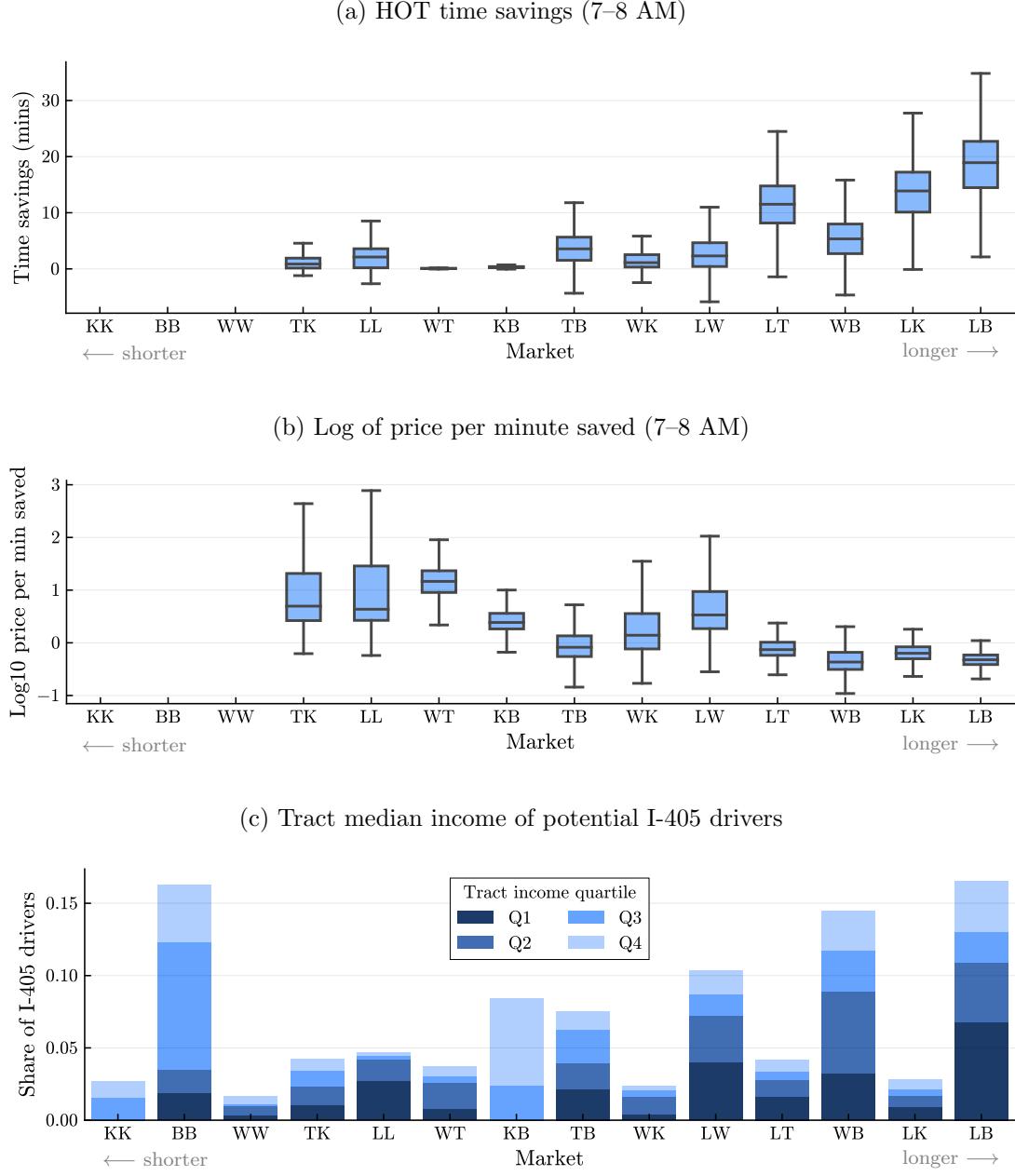
Not only do the longest markets offer greater potential value, but lower-income drivers disproportionately belong to these markets. Figure 2c plots the share of potential I-405 drivers in each market, split by the quartile of their home tract median income. 61 percent of drivers from below median income tracts are in one of the five longest markets, compared to just 36 percent of drivers from above median income tracts. This pattern is qualitatively similar, though less extreme, when drivers are instead divided into quartiles of car price (Figure A.3). This suggests that the spatial distribution of drivers can act as a “tag” in the spirit of [Akerlof \(1978\)](#), i.e. an immutable (or, at least, difficult to change) characteristic correlated with income that a social planner can use to target assistance. Whether intentional or not, the current design of WSDOT’s pricing algorithm effectively provides a better deal to the disproportionately lower-income drivers that travel longer distances on I-405.²¹

Fact 3: Conditional on market, lower- and higher-income drivers choose HOT lanes at similar rates and under similar conditions

Differences in aggregate conditions by geography are especially important because heterogeneity by observable income measures appears to be modest. Figures 3a and 3b show that the tract incomes and car prices of drivers taking the HOT lanes are similar to those in the broader population of potential 405 drivers, suggesting lower- and higher-income drivers take the lanes at similar rates. Figure 3c investigates how the attributes of HOT trips vary by income by regressing time saved and

²¹The algorithm sets prices according to contemporaneous traffic conditions, but using worst-case (across road segments traversed by the market) rather than average traffic conditions. The highest prices are found in markets that traverse the highly congested northern half of the corridor, which has only one HOT lane compared to two in the southern half. In these markets, the median price from 7–8 AM is at the \$10 ceiling.

Figure 2: HOT time savings, price per minute saved, and tract income by market



Note: Figures show the distributions of HOT time savings from 7–8 AM (top panel), the log of price per minute saved from 7–8 AM (middle panel), and potential drivers’ tract income quartiles (bottom panel) in each southbound market. In the top two panels, the boxes indicate quartiles and the whiskers extend to the nearest observed data point within a distance of 1.5 times the interquartile range (IQR) from the quartile. The bottom panel uses the median tract income for the simulated sample of potential drivers in the market (whether or not they actually choose I-405), which does not change hour-to-hour, to rank drivers from lowest income quartile (Q1) to highest income quartile (Q4). Markets are ordered from shortest to longest. The three shortest markets have no feasible HOT route.

price paid on driver tract income and car value.²² Comparing HOT trips unconditional on market or hour (in blue), those taken by higher-income drivers tend to have higher prices and time savings. These patterns are driven by drivers' choices of when to take the toll lanes, and reverse when adding fixed effects for hour of day (in pink). Comparing trips within the same market (in orange) or same market-hour (in purple), those taken by higher-income drivers are nearly identical to those taken by lower income drivers. The remaining differences are small in magnitude: conditional on the market and hour, each one standard increase in a driver's home tract income is associated with an increase in average time saved by 3.5 seconds and a decrease in average price paid by 0.4 cents. These relationships between income proxies and HOT trip characteristics will be important moments for estimating our model.

Fact 4: Lower- and higher-income drivers respond similarly to exogenous price and travel time changes

We use price discontinuities at the 25-cent rounding thresholds and car crashes as two sources of plausibly exogenous variation in prices and travel times, which will each be used for estimating our model. For prices, unrounded prices on opposite sides of a rounding threshold likely reflect similar underlying market conditions, but are almost-randomly assigned to realized prices that differ by \$0.25. Figure 4 shows results for a series of regression discontinuities, estimated with rounding threshold and interacted trip definition-hour fixed effects.²³ The number of paid HOT trips falls by about 5 percent when the underlying prices are rounded up rather than down (Figure 4a). Away from the threshold, average HOT quantities increase in price, reflecting price endogeneity: higher unrounded prices are associated with higher latent HOT demand. However, the average tract income and vehicle MSRP of drivers in the HOT lane is smooth through the threshold (Figure 4b), suggesting that the price-marginal driver is similar in their income proxies to the inframarginal drivers.

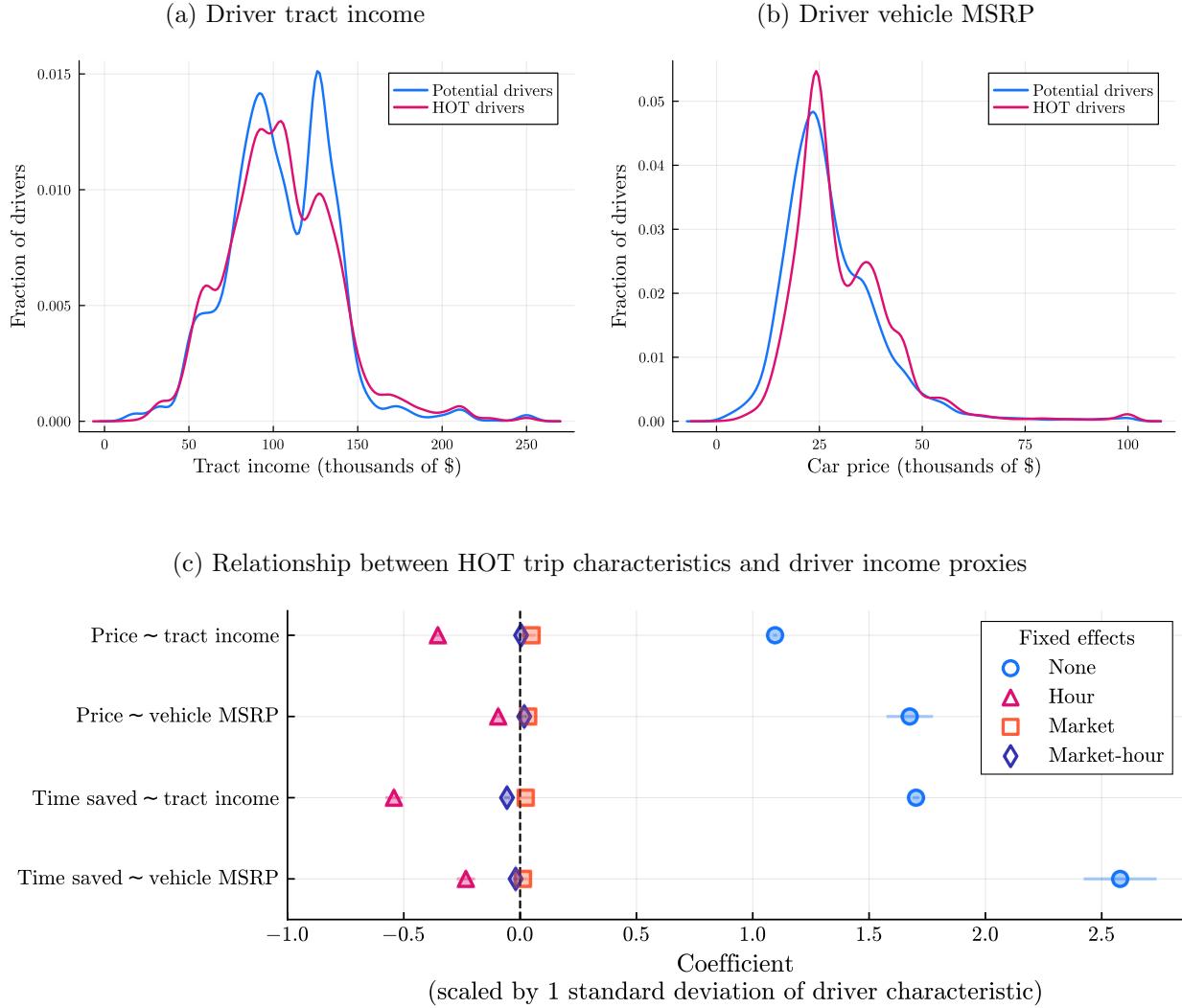
Our second source of variation is car crashes, which affect both travel times and, indirectly, prices. We measure the effects of a crash during peak hours using a dynamic difference-in-differences specification with interacted trip definition and time of day fixed effects. We compare "treated" (HOT trip definition, 5-minute interval) pairs that have a crash within 60 minutes and within 8 miles downstream of the trip definition's entry point to "control" observations with *no* crashes during the same day's peak period.²⁴ In the hour after a crash, prices for HOT lanes and travel times in both lanes increase (Figure C.22), and the number of HOT trips drops by about 6 percent (Figure 4c). As with variation due to price rounding, however, the average tract income and

²²Figure A.7 shows these same relationships as binscatters. Including HOV trips, Figures A.7e and A.7f shows that trips by drivers from higher income tracts are more likely to be HOV trips, with no relationship between HOV status and car value.

²³We exclude observations where the unrounded price is below the 75-cent price floor or above the \$10 price ceiling, as well as observations where rounding our reconstructed unrounded price does not yield the observed rounded price. These dropped observations respectively account for 42 percent, 10 percent, and 2 percent of the original observations.

²⁴For the southbound morning peak, we use just observations around the first southbound crash each morning, excluding any further crashes. Likewise, for the northbound afternoon peak, we only use observations around the first northbound crash each afternoon. We also exclude crashes so severe that pricing for the HOT lanes is turned off. The final sample includes 206 unique crashes.

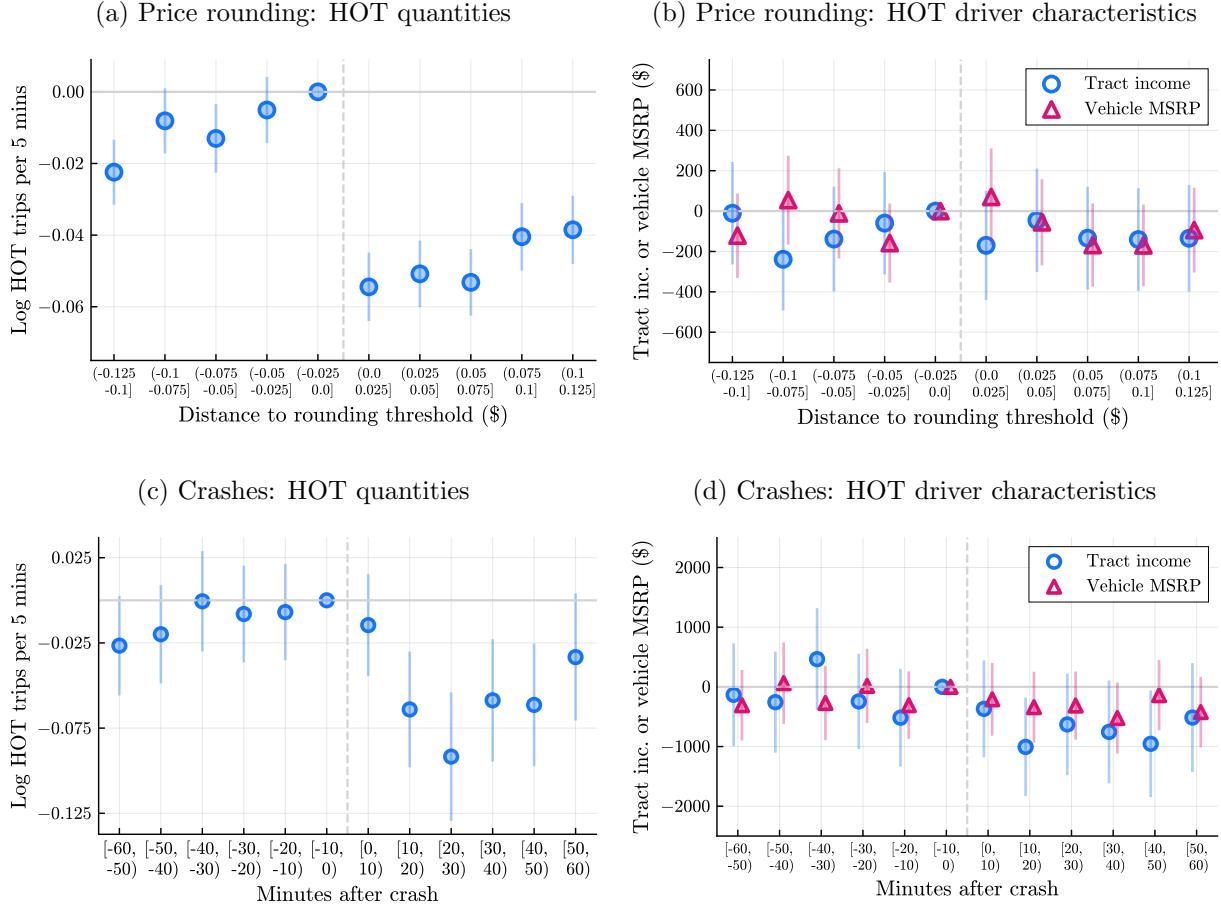
Figure 3: Heterogeneity in HOT usage



Note: The density plots in the top two panels show marginal distributions of income proxies in the sample of potential I-405 drivers (blue) vs. the sample of HOT drivers (pink, weighted by number of HOT trips taken). The bottom two panels plot coefficients from a series of regressions of HOT price or minutes saved on driver tract income or vehicle MSRP. The coefficients are scaled by the standard deviation of either tract income (\$31,274) or vehicle MSRP (\$16,558). Standard errors are clustered at the driver-level, and horizontal lines represent 95 percent confidence intervals.

vehicle MSRP of drivers in the HOT lane is smooth through the timing of a crash (Figure 4d), with at most a small decrease in the median tract income of the average HOT driver. The muted heterogeneity in response to price rounding and car crashes—and the modest relationships between income proxies and HOT trip characteristics documented in Figure 3c—foreshadow later results showing only limited preference heterogeneity along these two proxies for income.

Figure 4: Responses around price rounding thresholds and car crashes



Note: These figures plot coefficients from regression of HOT quantities or characteristics of HOT drivers on indicators for either bins of the unrounded price's distance to the rounding threshold or time since a crash occurred. Observations are at the (trip definition, five-minute interval, day) level and are subset to peak hours. The top two panels are estimated with threshold and interacted trip definition-hour fixed effects. For the bottom two panels, observations are ‘treated’ if they have a crash with 60 minutes and within 8 miles of the start point for the trip, and we exclude observations from crashes after the first crash in each day-direction and observations with a crash that occurs outside of the ‘treated’ definition limits. The bottom two panels are estimated with interacted trip definition and five-minute interval fixed effects. Vertical lines represent 95% confidence intervals.

5 Model

Motivated by these descriptive facts, we develop an equilibrium model of highway traffic, with the goal of creating a unified framework in which to conduct welfare analyses and simulate alternative policies. In the model, drivers with heterogeneous preferences choose departure times in the first stage and (priced or unpriced) routes in the second stage. Travel times and prices, which adjust in equilibrium to clear the market, are determined respectively by the road technology and the pricing algorithm.

The basic setup of the model is as follows. A **market**, indexed by $m \in \mathcal{M}$, is a highway entry-exit pair. It is possible that a given highway segment will be traversed by drivers from multiple markets, but each driver belongs to exactly one market, which she takes as given. Drivers choose between the unpriced and priced **routes**, indexed by $j \in \{0, 1\}$, in each market. Each day t is independent; time of day h is discretized into five-minute intervals. Travel times, prices, and quantities of cars on the road all vary at the (route j , market m , departure time h , day t) level. Prices may additionally vary at the driver i level, such as in the case of discounts for carpooling or low-income drivers, but we assume that driver i 's personal price is always a known function of the market-wide price.

5.1 Demand

In the demand model, drivers make sequential departure time and route choices each day. In the first stage, departure times are chosen under imperfect information about traffic conditions. In the second stage, this uncertainty is resolved and routes are chosen under full information. The stages are described below in reverse order. This two-stage structure allows drivers to benefit from tolling in two ways. Drivers who take the priced lanes benefit from realized time savings, which they value more than the price they have to pay. Drivers who don't take the priced lanes ex post nonetheless derive option value from their existence ex ante, as they make their choice of departure time knowing that they can reoptimize in the second stage, after the uncertainty about prices and travel times is resolved.

5.1.1 Stage 2: route choice

In the second stage, each driver observes the realizations of price and travel times at her chosen departure time and optimally chooses between the unpriced and priced routes. Drivers have heterogeneous preferences over prices, travel times, and time early or late to their destinations.

Consider a driver i in market m . At this stage, she has already chosen departure time h on day t . Now that she is on the road, she observes the realized market-wide price p_{jhmt} and travel time d_{jhmt} (measured in hours) on each route $j \in \{0, 1\}$. She therefore knows her personalized price, which we denote (abusing notation) by $p_{ijhmt} = \mathcal{P}_i(p_{jhmt})$, a known function $\mathcal{P}_i(\cdot)$ of the market-wide price.²⁵ Driver i is additionally endowed with an **ideal arrival time** $\eta_i \in [0, 24]$.

²⁵For example, in a counterfactual exercise we simulate cases where drivers from low-income tracts receive a 50%

This ideal arrival time combines with her chosen departure time and route to determine her time early or late to her destination.²⁶

Driver i chooses a route $j \in \{0, 1\}$ to maximize stage 2 utility

$$u_{i,j|h,m,t} = U_{i,j|h,m,t} + \varepsilon_{i,j|h,m,t} \quad (1a)$$

$$U_{i,j|h,m,t} = \alpha_{ijm}^0 + \underbrace{\alpha_i^P p_{jhmt}}_{\text{price}} + \underbrace{\alpha_i^D d_{jhmt}}_{\text{travel time}} + \underbrace{\alpha_i^E (h + d_{jhmt} - \eta_i)_-}_{\text{time early}} + \underbrace{\alpha_i^L (h + d_{jhmt} - \eta_i)_+}_{\text{time late}} \quad (1b)$$

where for any scalar x , we denote the negative part by $x_- = -\min(x, 0)$ and the positive part by $x_+ = \max(x, 0)$. The intercept term α_{ijm}^0 captures the baseline attractiveness of route j in market m . The coefficients $(\alpha_i^D, \alpha_i^E, \alpha_i^L)$ are the opportunity costs of time driving, time early, and time late, respectively.²⁷ The stage 2 utility shocks $(\varepsilon_{i,j|h,m,t})_{j \in \{0,1\}}$ are independently and identically distributed type 1 extreme value.

Given this utility specification, driver i chooses route j with logit choice probability

$$\pi_{i,j|h,m,t} = \frac{\exp U_{i,j|h,m,t}}{\exp U_{i,0|h,m,t} + \exp U_{i,1|h,m,t}}$$

To compute the aggregate market share of route j conditional on departure time h , we integrate over the continuum of drivers who have chosen to depart at h :

$$s_{j|h,m,t} = \int \pi_{i,j|h,m,t} dF_{hmt}(i) \quad (2)$$

Finally, the expected value of departure time h conditional on the realizations $\mathbf{p}_{hmt} = (p_{jhmt})_{j \in \{0,1\}}$ and $\mathbf{d}_{hmt} = (d_{jhmt})_{j \in \{0,1\}}$ is

$$EU_{ihmt}(\mathbf{p}_{hmt}, \mathbf{d}_{hmt}) = \log [\exp U_{i,0|h,m,t}(p_{i0hmt}, d_{0hmt}) + \exp U_{i,1|h,m,t}(p_{i1hmt}, d_{1hmt})]$$

5.1.2 Stage 1: departure time choice

In the first stage, each driver chooses a highway departure time or the non-highway outside option given imperfect information about future prices and travel times. Drivers are forward-looking, comparing their expected values of the different departure times.

Let \mathcal{H} be a discrete set of highway departure times, and let $h = \emptyset$ denote the non-highway outside option. On day t , drivers in market m have common beliefs $G_{hmt}(\mathbf{p}_{hmt}, \mathbf{d}_{hmt})$ about the joint distribution of prices and travel times at each departure time $h \in \mathcal{H}$. There are additionally time-varying shocks ξ_{hmt} to demand for different departure times, which are common across drivers

discount on tolls, so $\mathcal{P}_i(p_{jhmt}) = p_{jhmt}/2$ for these drivers.

²⁶We only model travel on the highway of interest. Strictly speaking, the chosen departure time h is the time at which the driver enters the highway, rather than the time she leaves home. Similarly, η_i is driver i 's ideal time at which she exits the highway, rather than when she arrives at her destination.

²⁷Appendix D.1 discusses the microfoundation of this utility specification.

in the market.

Driver i chooses $h \in \mathcal{H} \cup \{\emptyset\}$ to maximize stage 1 utility u_{ihmt} . The utility from the non-highway outside option is normalized to $u_{i\emptyset mt} = \varepsilon_{i\emptyset mt}$. The utility from departure time h is

$$u_{ihmt} = U_{ihmt} + \xi_{hmt} + \varepsilon_{ihmt} \quad (3a)$$

$$U_{ihmt} = \beta_{im}^0 + \beta_{im}^1 \int EU_{ihmt}(\mathbf{p}_{hmt}, \mathbf{d}_{hmt}) dG_{hmt}(\mathbf{p}_{hmt}, \mathbf{d}_{hmt}) \quad (3b)$$

The primary component of utility from departure time h is the expected value of departing at h , where the expectation is now additionally taken over drivers' beliefs about prices and travel times. The stage 1 utility shocks $(\varepsilon_{iht})_{h \in \mathcal{H} \cup \{\emptyset\}}$ are independently and identically distributed type 1 extreme value.

Driver i chooses departure time h with logit probability

$$\pi_{ihmt} = \frac{\exp\{U_{ihmt} + \xi_{hmt}\}}{1 + \sum_{h' \in \mathcal{H}} \exp\{U_{ih'mt} + \xi_{h'mt}\}}$$

and the overall market share of departure time h is

$$s_{hmt} = \int \pi_{ihmt} dF_m(i) \quad (4)$$

5.2 Road technology

Our road technology model embeds a static relationship between traffic density and traffic speed inside a model of traffic dynamics in space and time. Highway travel is a congestible good: as the density of vehicles on the road increases, vehicles are forced to slow down in order to maintain safe following distances, resulting in lower speeds. By additionally incorporating dynamics, we allow drivers to impose congestion externalities on other drivers traveling not only at the same location and time, but also at other locations and times. The road technology, which maps quantities of drivers on the road to travel times, encodes the technological constraints on the supply of highway travel.

We model the highway as a sequence of **links** indexed by $l \in \{1, \dots, L\}$, each divided into (GP and HOT) routes $j \in \{0, 1\}$. Each link l has length λ_l in miles, and each pair of route j and link l has width κ_{jl} , representing the number of lanes. Time of day is discretized into five-minute intervals of the form $[h, h + \Delta h]$; dates are indexed by t .

The “static” speed-density relationship holds at every discrete point in space and time.²⁸ **Speed** v_{jlht} , in miles per hour, and **density** ρ_{jlht} , in cars per lane-mile, are constant within each route j , link l , five-minute interval h , and date t . Speed is a function of contemporaneous density and a

²⁸ Appendix D.2 discusses the theoretical relationships between speed, density, and throughput.

speed shock ψ_{jlht} :

$$v_{jlht} = V(\rho_{jlht}, \psi_{jlht})$$

We assume that the mapping $V(\cdot, \psi_{jlht})$ is decreasing in density and does not vary across space or time on our highway of interest. The speed shocks are exogenous and capture idiosyncratic deviations from speeds predicted by density alone. For example, drivers on a particular segment at a particular time might be comfortable traveling at higher speeds or with shorter following distances.

To close the road technology model, we describe how speeds and densities on *road segments* are related to travel times and quantities in *markets*. Borrowing from the physics literature, we use a discretized “hydrodynamic” model that treats flows of vehicles analogously to the flow of fluids (Lighthill and Whitham, 1955; Richards, 1956). As Figure A.8 illustrates, travel times depend on speeds along the entire length of highway traversed and along the entire time interval taken.²⁹ Similarly, drivers contribute to traffic densities along their entire trajectories. For each driver who departs on route j in market m at time h' , let $\underline{h}_{jlmh't}$ and $\bar{h}_{jlmh't}$ denote the (continuous) times at which she respectively enters and exits link l . The density in route j on link l is obtained by summing the cars there from different markets and departure times, then dividing by the total lane-miles:

$$\rho_{jlht} = \frac{1}{\kappa_{jl} \times \lambda_l} \sum_{m \in \mathcal{M}} \sum_{h' \in \mathcal{H}} \underbrace{\frac{|[\underline{h}_{jlmh't}, \bar{h}_{jlmh't}) \cap [h, h + \Delta h]|}{\Delta h}}_{\text{fraction of interval } [h, h + \Delta h) \\ \text{in which drivers departing at } h' \\ \text{in market } m \text{ are on link } l} \times \underbrace{q_{jh'mt}}_{\text{mass of} \\ \text{departures at } h' \\ \text{in market } m}$$

5.3 Pricing algorithm

Finally, the pricing algorithm maps current and past traffic conditions into prices. In principle, traffic conditions on any link can affect the market m price—for example, if prices in a given market depend on traffic downstream of the segment traversed in that market.

In our empirical setting, the primary pricing algorithm inputs are contemporaneous densities and speeds in both the unpriced ($j = 0$) and priced ($j = 1$) routes. Let $(\boldsymbol{\rho}_{ht}, \mathbf{v}_{ht}) = (\rho_{jlht}, v_{jlht})_{j \in \{0,1\}, l \in \{1, \dots, L\}}$ denote the vectors of densities and speeds in each route and on each link at time h . The pricing algorithm computes HOT prices using market m -specific functions P_m :

$$p_{0hmt} \equiv 0 \tag{5a}$$

$$p_{1hmt} = P_m(\boldsymbol{\rho}_{ht}, \mathbf{v}_{ht}) \tag{5b}$$

²⁹We model speeds and densities for an additional two hours after the end of the departure time choice set \mathcal{H} , assuming there are no departures in these two hours. If a trip still has not concluded by then, we assume the remaining distance is traveled at freeflow speed.

6 Estimation

The demand, road technology, and pricing algorithm components of the equilibrium model are each estimated separately using data from the southbound morning commute in 2019. First, on the demand side, we estimate drivers' heterogeneous preferences for highway travel via the simulated method of moments. Second, for the road technology, we estimate an asymmetric logistic relationship between density and speed. Finally, we approximate the pricing algorithm using market-specific cubic polynomials of GP and HOT travel times into prices. Our data sample throughout is the southbound morning peak, 5–11 AM, in 2019.

6.1 Demand

We estimate the demand model using the simulated method of moments. The primary estimands are drivers' mean preferences for prices, travel times, and time early and late to their destinations—which are identified using plausibly exogenous shifters of price and travel time—and parameters governing preference heterogeneity—which are identified by matching micro moments in the toll transaction data.

Two features of our setting preclude us from directly applying the [Berry, Levinsohn and Pakes \(1995\)](#) method of estimating random-coefficients logit demand models. First, the two stages of the demand model—departure time choice and route choice—are closely linked, requiring joint estimation to account for drivers selecting into departure times based on unobservables. Second, we lack full data on market shares in the two stages: we observe departure time quantities at the *market origin*—not market—level from road sensors, and we observe HOT but not GP route quantities at the market level in the transactions data. Our approach addresses these data limitations by aggregating demand shocks up to the market origin level and using route *quantities* rather than route *shares* in the moment conditions.

6.1.1 Parameterization

We begin by augmenting the demand model with additional parametric assumptions on drivers' choice sets, preferences, and beliefs. Our empirical model is of the morning commute for drivers traveling southbound on I-405. Drivers choose from departure times $h \in \mathcal{H}$ spaced out every five minutes from 5 AM to 10:55 AM.

In stage 2, drivers have heterogeneous preferences over price and travel time, as well as heterogeneous ideal arrival times. Let \mathbf{x}_i be a vector containing driver i 's tract income and car price. Drivers' price and travel time coefficients vary observably with their characteristics and with an unobservable normally distributed component, while their time early and time late coefficients are homogeneous:

$$\begin{bmatrix} \alpha_i^P \\ \alpha_i^D \end{bmatrix} \sim N \left(\begin{bmatrix} \bar{\alpha}^P + \mu^{\alpha,P} \cdot \mathbf{x}_i \\ \bar{\alpha}^D + \mu^{\alpha,D} \cdot \mathbf{x}_i \end{bmatrix}, \Sigma^{\alpha,PD} \right) \quad \begin{aligned} \alpha_i^E &\equiv \bar{\alpha}^E \\ \alpha_i^L &\equiv \bar{\alpha}^L \end{aligned}$$

Driver characteristics in \mathbf{x}_i are expressed in deviations from the population mean, so that the common coefficients $\bar{\alpha}$ represent population means. Drivers share common route intercepts $\alpha_{ijm}^0 \equiv \alpha_{jm}^0$; without loss of generality, we normalize the GP route intercepts in each market to zero.³⁰ Ideal arrival times η_i are normally distributed with mean $\bar{\eta}$ calibrated to 8:30 AM and estimated standard deviation σ^η , truncated to the interval between 5 AM and 12 PM.

In stage 1, the primary estimands are the “inside good” intercepts, which control how attractive drivers in each market find I-405 relative to the non-405 outside option. We estimate different inside good intercepts depending on driver i ’s carpool status $c(i) \in \{\text{SOV}, \text{HOV}\}$.³¹ (These stand for “single-occupancy vehicle” and “high-occupancy vehicle”, respectively.) Motivated by data limitations, we assume that SOV drivers’ inside good intercepts vary only at the *market origin* level, while HOV drivers’ inside good intercepts continue to vary at the market level:

$$\beta_{im}^0 = \beta_{c(i),m}^0 = \begin{cases} \beta_{\text{SOV}, \text{orig}(m)}^0, & c(i) = \text{SOV} \\ \beta_{\text{HOV}, m}^0, & c(i) = \text{HOV} \end{cases}$$

We impose that demand shocks ξ_{hmt} also vary only at the level of market m ’s origin: $\xi_{hmt} \equiv \xi_{h, \text{orig}(m), t}$. This dimension reduction is necessary because we only observe overall departure time quantities (summing across carpool statuses) at the *market origin* level from the loop data, while we observe HOV drivers’ departure time quantities at the *market* level in the transaction data. Finally, we fix the stage 1 coefficients β_{im}^1 on the stage 2 expected value at one for all drivers in all markets. This assumption, which is not without loss of generality, amounts to imposing that the stage 1 and stage 2 logit shocks have the same scale parameter.

For the purpose of estimation, when drivers form beliefs about prices and travel times in stage 1, they observe the quarter of the year (i.e., the season), the day of week, and the presence or absence of morning precipitation. In counterfactual simulations, we model drivers as having rational expectations about prices and travel times. We estimate the joint distribution of prices and travel times in each market, conditional on quarter, day of week, and precipitation, in a first offline step. Appendix E.1 describes this procedure.

Putting it all together, the estimands of the demand model are collected in the vector $\theta = (\bar{\alpha}^P, \bar{\alpha}^D, \bar{\alpha}^E, \bar{\alpha}^L, \mu^{\alpha, P}, \mu^{\alpha, D}, \Sigma^{\alpha, PD}, (\alpha_{1m}^0)_{m \in \mathcal{M}}, (\beta_{\text{SOV}, a}^0)_{a \in \mathcal{A}}, (\beta_{\text{HOV}, m}^0)_{m \in \mathcal{M}}, \sigma^\eta)$, where market origins are indexed by $a \in \mathcal{A}$.

6.1.2 Price and travel time shifters

Next, we describe three sources of plausibly exogenous variation in prices and travel times, which we use to identify drivers’ average preferences for prices and travel times and their scheduling costs.

³⁰The HOT intercepts capture across-market differences in the convenience of the HOT route. For example, one of the HOT access points is a direct access ramp to the EvergreenHealth medical complex in Totem Lake. GP drivers exiting at Totem Lake must take a different off-ramp, which requires additional driving on suburban roads to get to the medical center.

³¹We draw carpool statuses for our simulated drivers jointly with their income proxies. Appendix B.4 provides more detail.

We think of the first two, price rounding and car crashes, as shifting *realized* prices and travel times in stage 2, when the driver is already on the road and is choosing between the GP and HOT routes. We think of the third, precipitation, as shifting drivers' *beliefs* about prices and travel times in stage 1, when she is choosing when to drive.

Price rounding We leverage the price discontinuities at the 25-cent rounding thresholds to identify drivers' mean price responsiveness. As shown above in Figure 4a, average paid HOT transactions drop discontinuously by 5 percent at the price rounding threshold. In principle, the price discontinuity can also affect realized travel times via the discontinuous effect on HOT demand. However, this effect is damped because travel times depend on demand in both the current five-minute interval (which does respond to this period's price rounding) and future five-minute intervals (which does not). Figure C.20 shows that realized HOT time savings do not exhibit a similar discontinuity at the price rounding threshold.

Moreover, we argue that the discontinuous drop in HOT quantity is not due to drivers updating their beliefs about imperfectly observed travel times. A subset of HOT access points have additional electronic signs displaying estimated travel times in the GP and HOT routes. While the sign-reported HOT time savings are smooth around the price rounding threshold, the HOT quantity discontinuity persists after restricting to trip definitions which have these signs (Figure C.21).

Car crashes We identify drivers' disutility from travel time using variation from car crashes. As discussed in Section 4, crashes are negative technology shocks, effectively temporarily reducing highway capacity. This in turn results in lower speeds and higher travel times (Figure C.22a), and in equilibrium, also higher prices (Figure C.22b). Figure 4c shows that the number of paid HOT trips drops by about 6 percent in the 60 minutes following a peak-hour crash.

Precipitation Finally, we use precipitation, which increases the *variance* of travel times (Appendix C.4), to identify drivers' costs of arriving early or late to their destinations. The direction and extent to which drivers shift their departure times in response is informative about the direction and extent of asymmetry in these scheduling costs. If drivers find it much more costly to be late than to be early, then they will choose to depart much earlier on rainy days to avoid potentially incurring those costs.

We find limited departure time responses to this increase in variance, suggesting that drivers have relatively low and slightly asymmetric scheduling costs, with time late slightly more costly than time early. On days with positive precipitation during the morning peak (5–11 AM), the shares of early morning departures increase and the shares of peak-hour departures decrease (Figure C.27a). However, the magnitudes of these differences are extremely small (Figure C.27b).

6.1.3 Estimation and identification

The two stages of the demand model are estimated jointly using the simulated method of moments. The population mean preference parameters are identified using variation from the price and travel time shifters—price rounding, crashes, and precipitation—described above in Section 6.1.2. The preference heterogeneity parameters are identified by matching moments from the toll transaction data, which we first saw in the descriptive evidence in Section 4.

Joint estimation of the two stages is necessary because of the tight link between the two stages in the demand model. The second-stage parameters are needed to construct the expected value of each departure time in the first-stage utility (3). However, the second stage can't be estimated alone because drivers may select on unobservables into different departure times: for example, drivers who select into departing during peak hours may have lower values of travel time and higher scheduling costs.³²

We adapt the approach of [Berry, Levinsohn and Pakes \(1995\)](#), henceforth BLP) to evaluate the moment conditions at each candidate parameter vector θ . At each candidate θ , we first compute $U_{ihmt}(\theta)$, the expected value of departure time h , for each simulated driver, market, and date. Let \mathcal{M}_a denote the set of markets with origin a . For each origin a and date t , we solve a system of nonlinear equations for the vector of demand shocks $(\hat{\xi}_{hat}(\theta))_{h \in \mathcal{H}}$ that rationalize the observed departure time quantities q_{hat} :

$$\underbrace{q_{hat}}_{\substack{\text{departure time } h \\ \text{quantity in} \\ \text{market origin } a}} = \sum_{m \in \mathcal{M}_a} \underbrace{q_{mt}}_{\substack{\text{market} \\ m \text{ size}}} \times \underbrace{\left(\frac{1}{I} \sum_{i=1}^I \frac{\exp \{U_{ihmt}(\theta) + \xi_{hat}\}}{1 + \sum_{h' \in \mathcal{H}} \exp \{U_{ih'at}(\theta) + \xi_{h'at}\}} \right)}_{\substack{\text{departure time } h \\ \text{share in market } m}} \quad (6)$$

This is done by iterating a BLP-like contraction mapping. Then we use the candidate parameters θ and the implied demand shocks $\hat{\xi}_{hat}(\theta)$ to predict demand.³³

A key challenge is that we only observe departure time quantities at the *market origin*—not market—level, from data on throughputs on highway on-ramps. This data limitation affects our estimation approach in both stages of the demand model. In stage 1, during the BLP inversion (6) of market shares to demand shocks, we are only able to recover demand shocks $\hat{\xi}_{hat}$ at the market origin a level. In stage 2, since we additionally do not observe GP quantities at the market level, we are unable to compute GP and HOT route *shares* conditional on departure times. Instead, we match moments of HOT route *quantities* in each market, which we do observe in the toll transaction data.

We summarize the following primary moment conditions. In the first stage, we use precipitation as a traditional method of moments instrument, imposing that it is independent of unobserved de-

³²This selection on unobservables can also be seen in equation (2) for the market share of route j in departure time h , where the choice probabilities are integrated over the driver distribution $F_{hmt}(i)$ *conditional on choosing departure time h* .

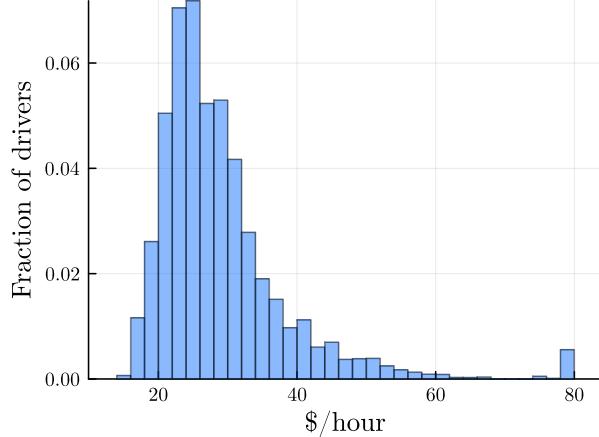
³³We approximate the distribution of driver types using $I = 100$ simulated drivers in each market. Driver characteristics are sampled from the unconditional distribution described in Section 3.3.

mand for departure times. In the second stage, we match reduced-form coefficients from regressions of HOT quantity on indicators for price rounding and crashes. To identify the parameters governing preference heterogeneity, we match micro moments of driver characteristics and HOT trip attributes (price paid and time saved) in the toll transaction data. For example, the covariances between a driver's income proxies and the price she pays in her toll transactions are informative about the slope coefficients $\mu^{\alpha, P}$ governing how drivers' price coefficients vary with their incomes. Appendix E.2 describes the full set of moment conditions in greater detail.

6.1.4 Estimated preferences

Tables A.2 and A.3 report the estimated demand parameters. The average driver is willing to pay $\bar{\alpha}^D/\bar{\alpha}^P = \27.02 to avoid one hour of travel time. Figure 5 shows there is modest heterogeneity in the value of travel time (VOTT), α_i^D/α_i^P . Drivers in the 5th and 95th percentiles of VOTT are willing to pay \$19.47 and \$47.45, respectively, to avoid an hour of travel time. Drivers' VOTTs are positively correlated with tract incomes and car prices (Figure A.9). Scheduling costs are relatively low: the average driver is willing to pay $\bar{\alpha}^E/\bar{\alpha}^P = \3.37 to avoid being one hour early to her destination and $\bar{\alpha}^L/\bar{\alpha}^P = \3.82 to avoid being one hour late. The standard deviation of ideal arrival times, σ^η , is 52 minutes.

Figure 5: Estimated value of travel time distribution



Note: This figure shows the estimated distribution of drivers' value of travel time, α_i^D/α_i^P . Each underlying observation is a simulated driver from the unconditional distribution of driver characteristics.

6.1.5 Discussion

Our empirical demand model comes with important caveats and limitations. First, we assume drivers have perfect information about prices and travel times when they are on the road. Prices are directly observable, as they are displayed on the electronic signs displayed at each HOT access point (Figure A.1). Travel times may be more difficult for drivers to infer based on the traffic at

their current location. However, in Section 6.1.2, we showed that the discontinuous drop in HOT quantities at the price rounding threshold is similar when restricting attention to trips where there is an electronic sign displaying estimated travel times near the entry. In practice, however, there may still be drivers who misread the signs or are otherwise confused about how tolling works on I-405. To the extent that such drivers are common, our revealed preference approach will tend to underestimate drivers' responsiveness to prices and travel times.

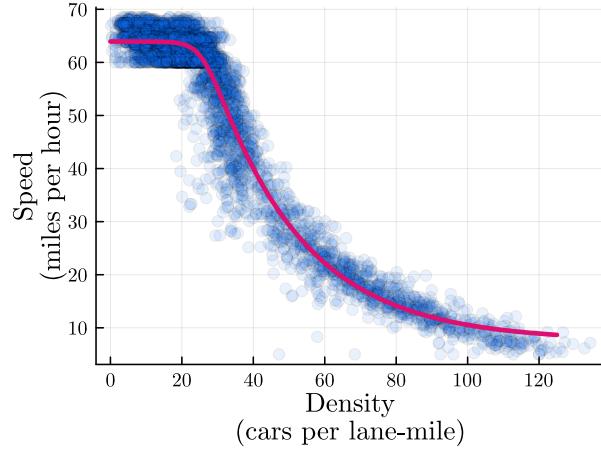
There are also potentially important dimensions of heterogeneity that are not included or not estimated in the current version of the model. We estimate heterogeneous preferences for price and travel time, which vary both observably along income proxies and unobservably across drivers. However, we do not estimate heterogeneity in scheduling costs or in the distribution of ideal arrival times, and allow for very limited preference heterogeneity across carpool statuses.³⁴ Additional heterogeneity in scheduling costs and ideal arrival times can in principle be identified using additional data on how departure times (and departure time responses to beliefs shifters like precipitation) vary by driver income proxies.³⁵ True preferences may also vary *within drivers*, across days. For example, a morning trip to the airport is likely to involve both a different ideal arrival time and higher scheduling costs. If drivers' true ideal arrival times are very different day-to-day even when subset to morning commutes, then we will tend to underestimate scheduling costs and the option value of tolling.

Finally, we model a medium-run choice problem in which agents decide only when and where they drive. In the long run, households have more margins of adjustment. Workers may adjust their morning schedules in response to traffic conditions, which could lead us to underestimate the true costs of early and late arrivals for drivers. Additionally, households may change where they live and work in response to changes in transportation policy. Our utility specification includes market-specific HOT intercepts α_{1m}^0 , which absorb across-market differences in tastes for HOT lanes. However, our demand estimates may be biased if preferences for price, travel time, and time early and late also vary systematically across markets in a way that is not captured by preference heterogeneity along income proxies.

³⁴Using survey data from California SR 91 and the National Household Travel Survey, Hall (2021) estimates that drivers' values of travel time are negatively correlated with their scheduling costs. That is, higher-income drivers, who have higher values of time, also have more flexible schedules. Limiting preferences by carpool status does not present a problem for estimation, so long as drivers take their carpools as exogenously given and HOV drivers' preferences over non-price attributes are not systematically different from those of single-occupancy vehicle (SOV) drivers with the same income proxies.

³⁵The Puget Sound Regional Council (PSRC) Household Travel Surveys are a promising source of these data, with detailed travel diaries linked to household characteristics. The mean morning departure time in the 2019 travel diaries does not differ substantially by household income in the Seattle metro area. However, the *variance* of morning departure times is higher for low-income households. Though realized departure times are not the same as ideal arrival times, this suggests that the first-order heterogeneity in ideal arrival times is likely in the variance rather than the mean.

Figure 6: Density vs. speed



Note: The figure shows a random sample of 5000 southbound observations, where each observation is a (loop, five-minute interval, date) from 5–11 AM in 2019. Top-coded speeds have been replaced with estimated freeflow speeds. The pink curve is the estimated asymmetric logistic speed-density relationship.

6.2 Road technology

To take the road technology model to the data, we estimate the static speed-density relationship specific to I-405. Additionally, in our model of traffic dynamics, we discretize I-405 into $L = 2$ links, corresponding to the congested northern half, which has a single HOT lane, and the more free-flowing southern half, which has two HOT lanes.³⁶

The primary empirical object is the speed-density relationship, which is different on every road. It depends on physical factors, including the hilliness of the terrain, the pavement materials and quality, and the presence of medians or other dividers. At each density, we expect speeds to be lower on single-lane, back-country roads than on multilane highways. It also depends on the geometry of the road network: highways, which feature relatively uninterrupted traffic flows, are able to support higher freeflow speeds than dense urban road networks.

We assume the following asymmetric logistic functional form (Wang et al., 2011) mapping traffic density ρ_{jhlt} into traffic speed v_{jhlt} :

$$v_{jhlt} = \underline{v} + \frac{\bar{v} - \underline{v}}{\left[1 + \exp\left(\frac{\rho_{jhlt} - \rho^*}{\delta_1}\right)\right]^{\delta_2}} + \psi_{jhlt} \quad (7)$$

The estimands are the jam speed \underline{v} and the freeflow speed \bar{v} (the lower and upper asymptotes, respectively), a scale parameter δ_1 , a parameter δ_2 which controls the degree of asymmetry, and a critical density ρ^* at which traffic transitions from freeflow to congested.

Speed is flat at very low density levels, then decreases convexly in density before finally leveling

³⁶The first, northern link is 8.1 miles long and averages 2.4 GP lanes. The second, southern link is 8.2 miles long and averages 3.2 GP lanes.

off.³⁷ We estimate equation (7) on 2019 loop data from the southbound morning peak. Each observation is a (route j , loop l , five-minute interval h , day t). Figure 6 shows the best-fit curve in pink over a random sample of speeds and densities in blue; Table A.4 reports the estimated parameters. The freeflow speed is $\bar{v} = 63.9$ miles per hour, slightly above the speed limit of 60 miles per hour. Speed begins to fall at the critical density, $\rho^* = 26$ cars per lane-mile, which corresponds to a car-following distance of 2.15 seconds at freeflow speed. The jam speed is $\underline{v} = 7.6$ miles per hour.

6.3 Pricing algorithm

We approximate WSDOT’s pricing algorithm using a flexible function of speeds in each market. While we observe the true pricing algorithm, its inputs are not easily generated in equilibrium simulations. In each market, we regress the unrounded HOT price p_{1hmt} on a third-order polynomial of contemporaneous speeds $\mathbf{v}_{jt} = (v_{jlht})_{j \in \{0,1\}, l \in \{1,2\}}$ in the GP and HOT lanes and on each of the two links indexed by l .³⁸ We construct link (road segment)-level speeds from the loop data by averaging speeds across loops on each link.

Since our estimated polynomials are an approximation of the true pricing algorithm, we introduce **price shocks** ϕ_{jhmt} that capture the difference between our predicted unrounded price and the observed unrounded price. We rewrite the pricing equation (5) for our empirical analysis as:

$$p_{0hmt} \equiv 0 \tag{8a}$$

$$p_{1hmt} = \tilde{P}_m(\mathbf{v}_{jt}) + \phi_{1hmt} \tag{8b}$$

where \tilde{P}_m denotes the estimated polynomial approximation in market m , which maps contemporaneous highway speeds \mathbf{v}_{jt} into predicted unrounded prices.

7 Welfare analysis

Armed with our estimated parameters, we now turn to quantifying the equilibrium effects of tolling and evaluating alternative pricing policies. First, we show that low-income drivers indeed gain the most from tolling, especially when tolling is applied to previously HOV-only lanes. Second, we quantify the option value of tolling, which is an important component of welfare gains both in aggregate and for low-income drivers. Finally, we compare status quo tolling to two types of policy-relevant alternative designs: raising the \$10 price ceiling and instituting income-based pricing.

³⁷The initial flatness is partially an artifact of replacing the top-coded speeds, as the freeflow speeds replacing them are estimated at the less granular (loop, day of week, hour of day) level. However, it is unlikely that average speeds are much higher than 70 miles per hour even when densities are extremely low.

³⁸The true pricing algorithm is a function of both contemporaneous speed and traffic density, but adding polynomial terms for traffic density did not improve the fit of the approximation.

7.1 Aggregate and distributional effects of tolling

We begin by evaluating the aggregate and distributional effects of tolling a subset of lanes on the highway. We find that tolling increases aggregate welfare, but the magnitude of the increase depends on whether the newly tolled lanes were previously HOV or previously GP. Furthermore, tolling does not create Lexus lanes: tolling is welfare-improving for drivers across the income distribution, but it particularly benefits drivers in the lowest quartile of tract income, who tend to belong to the longest markets.

To show this, we simulate equilibria under four regimes. In each equilibrium, drivers choose between two routes, the first of which of which is always GP. The four regimes differ in how the second route is priced: it is either an HOV lane, an HOT lane, a GP lane, or a toll lane. When the second route is priced (i.e., when it is an HOT lane or a toll lane), we use the approximation of the pricing algorithm estimated in Section 6.3.³⁹ We simulate equilibria using the demand shocks that rationalize the observed departure time shares—that is, by performing the BLP-like inversion (6) at the estimated demand parameters $\hat{\theta}$.⁴⁰ Our primary outcome of interest is net welfare, which we define as the sum of driver surplus and uniformly redistributed toll revenue. We include utility from all drivers, including those who choose the non-405 outside option; the utility of these non-405 drivers is normalized to zero. We hold the number of potential I-405 drivers and drivers' carpooling behavior fixed in counterfactual equilibria.⁴¹

We show that the effects of tolling differ depending on whether we begin tolling previously HOV lanes or previously GP lanes. We evaluate the following two types of lane conversions:

1. **HOV to HOT:** The primary means of introducing highway congestion pricing in the US is by converting existing HOV lanes to HOT lanes. This comparison tells us about the effects of the policy change we are most likely to see in the real world.⁴²
2. **GP to toll:** Alternatively, we can convert a subset of existing GP lanes to a toll lane (for which HOV drivers must also pay). This comparison captures the “pure” effect of pricing in the absence of discounts for carpooling.

Table 2 reports aggregate outcomes under the four regimes, as well as changes in aggregate outcomes under each of the two types of lane conversions.

³⁹In the all-GP equilibrium, we model the choice between two identical routes to ensure that our welfare changes are not driven by the difference between maximizing over one vs. two draws from the extreme value distribution.

⁴⁰We also simulate new draws of speed shocks $\psi_{j|h}$ and price shocks $\phi_{j|h|m}$ by bootstrapping estimated residuals of the speed-density equation (7) and the pricing equation (8), respectively, conditional on the route j , time of day h , and link l or market m . These draws of speed and price shocks are held constant across the counterfactuals.

⁴¹HOV lanes were originally conceived as a means of reducing congestion by incentivizing carpooling, so in the GP-or-HOV counterfactual, we should expect to see more HOV drivers, fewer SOV drivers, and fewer total cars on the road than in the status quo. Our estimates may overstate the true gains from converting HOV lanes to HOT lanes, if this conversion leads to some unmodeled reduction in carpooling. However, we expect this upward bias to be small, as the empirical literature finds that HOV lanes induce only small changes in carpooling behavior (Giuliano, Levine and Teal, 1990; Small, Winston and Yan, 2006).

⁴²As discussed in Section 2, on I-405, the initial HOV-to-HOT conversion was bundled with the construction of an entirely new lane and a change in HOV occupancy requirements.

Table 2: Aggregate equilibrium outcomes

	(a) Levels			
	(1) GP or HOV	(2) GP or HOT	(3) All GP	(4) GP or toll
Market shares				
Stage 1 I-405 share	0.463	0.517	0.549	0.516
Stage 2 HOT share	0.085	0.302	0.000	0.278
Drivers' average I-405 outcomes				
Price paid (\$)	0.000	0.391	0.000	0.551
Travel time (mins)	12.621	12.191	12.786	12.044
Time early (mins)	30.757	30.539	31.039	30.460
Time late (mins)	18.054	17.487	17.641	17.731
Welfare				
Driver surplus (\$/driver/year)	445.810	516.498	568.227	521.908
Toll revenue (\$/driver/year)	0.000	45.850	0.000	59.871
Net welfare (\$/driver/year)	445.810	562.338	568.227	581.766
	(b) Changes			
	(2) – (1) HOV to HOT	(4) – (3) GP to toll		
Change in market shares				
Stage 1 I-405 share	0.054	–0.032		
Stage 2 HOT share	0.217	0.278		
Change in drivers' average I-405 outcomes				
Price paid (\$)	0.391	0.551		
Travel time (mins)	–0.431	–0.742		
Time early (mins)	–0.218	–0.579		
Time late (mins)	–0.568	0.090		
Change in welfare				
Driver surplus (\$/driver/year)	70.688	–46.318		
Toll revenue (\$/driver/year)	45.850	59.871		
Net welfare (\$/driver/year)	116.528	13.540		
Welfare decomposition				
Ex ante value (\$/driver/year)	105.057	0.122		
Option value (\$/driver/year)	11.471	13.418		

Note: The stage 2 shares and average I-405 outcomes are conditional on choosing I-405—that is, conditional on *not* choosing the non-405 outside option. Average time early averages zeros (if the driver is late) and positive values (if the driver is early); the same is true for average time late. The welfare values sum across all drivers, including those who choose the outside option.

We find that tolling increases aggregate welfare in both cases, but the increase is much larger when the newly tolled lanes were previously HOV than when they were GP. We estimate that tolling the HOV lanes increases aggregate welfare in the southbound morning peak by \$117 per driver per year. In contrast, tolling the same number of GP lanes increases welfare more modestly by \$14 per driver per year. Summing over all potential I-405 drivers, these welfare changes amount to \$19.3 million per year and \$2.2 million per year, respectively. The HOV-to-HOT conversion is welfare-improving even before redistributing the toll revenue, increasing driver surplus by \$71 per driver per year, while the GP-to-toll conversion *reduces* driver surplus by \$46 per driver per year. On the other hand, HOT lanes raise less revenue (\$46 per driver per year) than do toll lanes without an HOV discount (\$60).

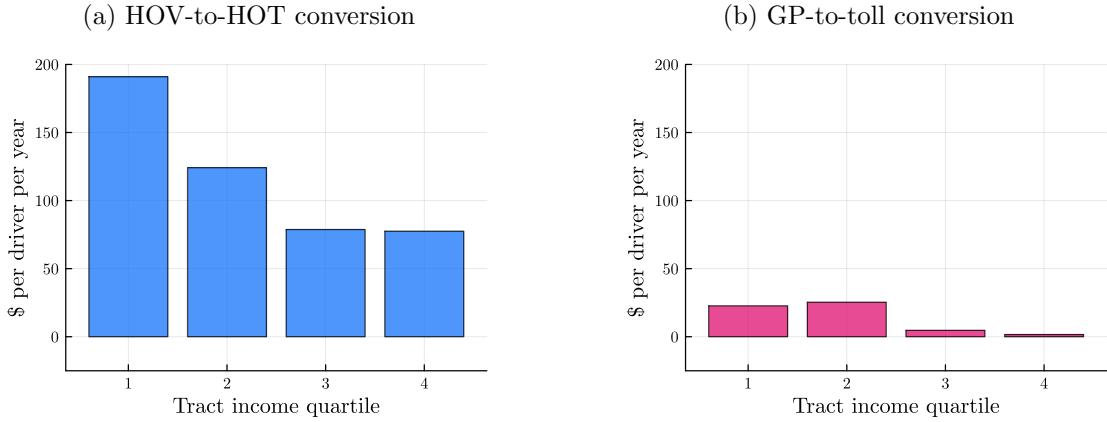
These differences across the two versions of tolling arise via opposite effects on I-405's effective capacity. The HOV-to-HOT conversion increases the effective highway capacity available to SOV drivers. Pre-tolling, the HOV lanes are severely underutilized: of the drivers who choose I-405 over the outside option, only 8.5 percent take the HOV lanes in stage 2, with the remaining 91.5 percent all crowding into the GP lanes. Post-tolling, drivers spread out more evenly across the lane types, with 30.2 percent now choosing HOT. Average travel times decrease even as the share of drivers taking I-405 increases by 5.4 percentage points.⁴³ Each SOV driver gains in net welfare (+\$162 per driver per year) approximately as much as each HOV driver loses (-\$164), but SOV drivers make up 86.1 percent of population. In contrast, the GP-to-toll conversion effectively *reduces* (unpaid) highway capacity. The outside option share increases by 3.2 percentage points, making those marginal drivers worse off, but inframarginal drivers who continue to take I-405 enjoy lower average travel times.⁴⁴

Moreover, the new HOT lanes are not Lexus lanes: in both comparisons, we find that tolling benefits lower-income drivers more, though drivers across the income distribution are made better off. The HOV-to-HOT conversion increases welfare in the lowest tract income quartile by \$191 per driver per year, accounting for 38 percent of the aggregate welfare gain (Figure 7a). They also substitute the most away from the outside option toward I-405 (6 percentage points). The pattern is qualitatively similar but quantitatively smaller for the GP-to-toll conversion. Net welfare in the bottom half of the tract income distribution increases by \$24 per driver per year, compared to a \$3 per driver per year increase for drivers in the top half (Figure 7b). In both cases, conditional on choosing I-405, lower-income drivers pay higher average tolls but save more in travel time and scheduling costs relative to the untolled equilibrium, and they take the priced lanes at higher rates.

⁴³Since the HOT conversion functions as an expansion of (SOV) highway capacity, the long-run welfare improvements may be smaller than those estimated here if the 'induced demand' response offsets any speed gains (Duranton and Turner, 2011). Other empirical papers, which model joint commuting, employment, and residential choice, study the effects of tolling previously general-purpose roads, effectively *reducing* (unpaid) highway capacity. These find that congestion pricing incentivizes households to reduce their commuting distances and substitute away from commutes through highly congested areas (Herzog, 2024; Barwick et al., 2024).

⁴⁴Our simulations hold the value of the non-405 outside option fixed across counterfactual equilibria. We expect that, because moving from GP-or-HOV to the status quo GP-or-HOT regime increases the share taking I-405, congestion on substitute roads and highways will decrease. This would improve the outside option, leading us to underestimate the welfare gains from converting the HOV lanes to HOT.

Figure 7: Welfare changes by tract income quartile



Note: These figures show changes in welfare by drivers' tract income quartile. Figure A.12 shows these differences by market.

Lower-income drivers substitute more toward I-405 in the HOV-to-HOT conversion (than higher-income drivers) and away from I-405 less in the the GP-to-toll conversion. Table A.5 and Table A.6 report changes in the full set of outcome variables by tract income quartile for HOV-to-HOT and GP-to-toll, respectively.⁴⁵

Both the HOV-to-HOT and GP-to-toll conversions benefit some markets more than others, and tend to benefit longer markets more than shorter markets. The gains from tolling previously HOV lanes range from \$53 per driver per year in one of the shortest markets to \$302 per driver per year in one of the longer markets (Figure A.12a). In contrast, tolling previously GP lanes transfers substantial surplus, ranging from a \$150 per driver per year loss in one of the shortest markets to a \$238 per driver per year gain in one of the longer markets. Drivers in longer markets experience the greatest reductions in travel time and scheduling costs, and they are disproportionately likely to take I-405 and the priced lanes. Across-market spillovers due to multiple markets traversing the same segments of I-405 also play an important role, particularly for drivers in the three shortest markets who have no feasible HOT route. These drivers nonetheless benefit from the HOV-to-HOT conversion as traffic in overlapping markets shifts out of the congested GP lanes and into the HOT lanes. They are likewise worse off under the GP-to-toll conversion as tolling induces traffic to shift into the GP lanes.

7.2 Ex ante and option values of tolling

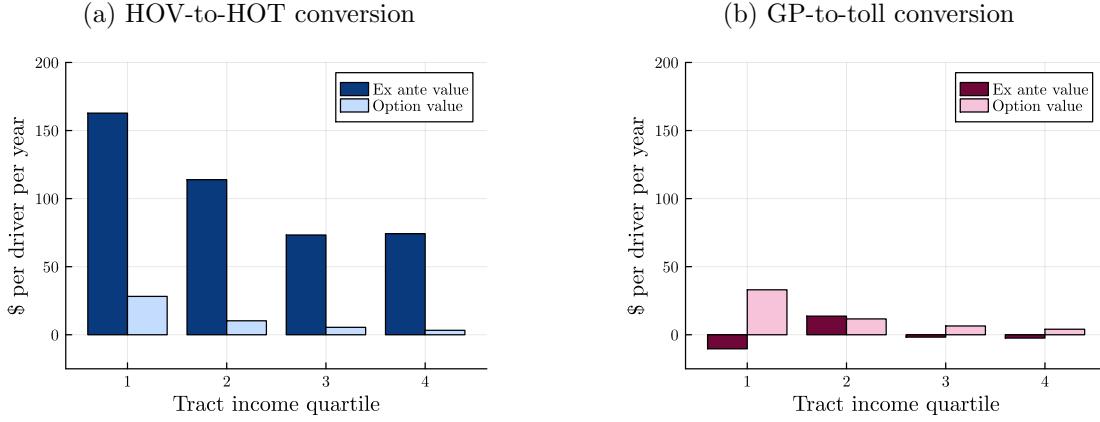
Next, we decompose the welfare changes from tolling into two channels, the ex ante value and the option value of tolling. Since drivers make departure time choices in the first stage of the demand model under imperfect information, they derive option value from being able to reoptimize their route choices in the second stage, after the uncertainty is resolved. The ex ante value captures

⁴⁵Figure A.10 shows that welfare changes do not systematically vary with drivers' vehicle MSRP.

the value of tolling in the absence of this ability to reoptimize. Since it also accounts for the re-equilibration of GP travel times after the introduction of tolling, it can be positive or negative, while the option value is nonnegative by construction.

To quantify these two welfare channels, we compute welfare in an intermediate scenario with full first-stage commitment. In this intermediate step, drivers continue to face status-quo prices and travel times, but they must commit to both their departure time and route choices in stage 1, before they observe the realizations of prices and travel times in stage 2. For SOV drivers, the option value from tolling is the difference between status-quo welfare and welfare under commitment. The ex ante value is the remaining difference between welfare under commitment and welfare in the no-toll counterfactual.⁴⁶ In the HOV-to-HOT conversion, we define the option value for HOV drivers—who already had access to the HOV lanes pre-tolling—to be zero.

Figure 8: Welfare decomposition by tract income quartile



Note: These figures show the decomposition of welfare gains into ex ante and option value, broken out by drivers' tract income quartile. Figure A.13 shows the decompositions by market.

The option value is an important component of the aggregate gains from tolling, and its magnitude is relatively stable across the two types of lane conversions. The option value due to the HOV-to-HOT conversion is \$11 per driver per year, similar to the \$13 due to the GP-to-toll conversion. The ex ante values, however, are very different depending on source of the newly tolled lanes. The ex ante value is the dominant component of the gains from tolling HOV lanes, contributing another \$105 per driver per year. When tolling GP lanes instead, after averaging over ex ante gains in some markets and ex ante losses in others, the ex ante value is only \$0.12 per driver per year.

Low-income drivers also experience the greatest option value from tolling. Figure 8a shows that in the HOV-to-HOT conversion, drivers in the lowest quartile have both the highest option value (+\$28 per driver per year) and the highest ex ante value (+\$163 per driver per year). The

⁴⁶Note that this is not a counterfactual *equilibrium*: we do not recompute equilibrium prices and travel times in this commitment world. Instead, the thought experiment is that one driver at a time is forced to go about her highway travel under full first-stage commitment. Since each driver is small and can't affect aggregate traffic conditions by herself, the equilibrium prices and travel times are unchanged.

bottom quartile's option value also accounts for the greatest share (21 percent) of their overall welfare increase. Likewise, Figure 8b shows that in the GP-to-HOV conversion, bottom-quartile drivers again have the highest option value (+\$33 per driver per year), but the ex ante value is nonmonotone in tract income. Low-income drivers are more sensitive to both price and travel time (Table A.2a) even though they're less willing to trade off price for travel time (Figure A.9). These drivers are more marginal when deciding between the unpriced and priced routes, so they benefit more from being able to make that choice under full information about the trade-offs.

7.3 Alternative pricing policies

Finally, we compare status quo tolling (i.e., including HOV discounts) to two types of alternative pricing policies that are currently under consideration by WSDOT:

1. **Higher price ceiling:** As of October 2023, WSDOT is considering raising the \$10 price ceiling to \$15 or \$18 in order to control increasing HOT congestion (Lindblom, 2023). We simulate a counterfactual equilibrium in which the price ceiling is \$12. We consider this more modest increase to avoid extrapolating our pricing algorithm approximation beyond the historical unrounded prices, which are capped at \$12.
2. **Low-income discount:** In 2021, the Washington State Transportation Commission (WSTC) conducted a study exploring a potential Low-Income Tolling Program (Washington State Transportation Commission, 2021). We evaluate two versions of income-based tolling that they consider in that study: a 50 percent discount and a \$2 per trip discount.⁴⁷ In our simulations, drivers in the bottom quartile of tract income are eligible for this discount.⁴⁸

Table 3 reports how departure time and route market shares, prices, travel times, and welfare change under these alternative policies relative to the status quo. We report both aggregate changes and changes for drivers in the bottom tract income quartile, who are targeted for the low-income discounts.

Raising the price ceiling to \$12 is very modestly welfare-improving, both in aggregate and across all quartiles of tract income and car price. The greater flexibility afforded by this higher ceiling allows the pricing algorithm to more efficiently manage peak congestion without substantially increasing average prices paid. Under the higher ceiling, drivers pay very slightly higher average tolls (+0.2 cents per day) but obtain modest reductions in average travel time (-0.013 minutes), time early (-0.088 minutes), and time late (-0.007 minutes). Aggregate welfare is higher by \$0.99

⁴⁷The WSTC study considers three additional options: a fixed toll credit per month, a fixed number of free toll trips per month, and a lower price ceiling for low-income drivers only. In April 2023, the San Francisco Bay Area Metropolitan Transportation Commission began piloting its own income-based tolling program, Express Lanes START, which offers low-income drivers a 50 percent discount (Metropolitan Transportation Commission, 2023).

⁴⁸The most common eligibility criterion, both considered in the WSTC study and actually in place in the Bay Area, is that the driver's household income is less than 200 percent of the federal poverty line. We do not directly observe household incomes, so we choose this 25th percentile cutoff for tract median household income based on ACS estimates that 23.3 percent of Washington State households had incomes below 200 percent of the federal poverty line in 2019 (KFF, 2023).

Table 3: Changes in outcomes under alternative pricing policies

	Raise price ceiling to \$12		Low-income 50% discount		Low-income \$2 discount	
	Agg	Q1	Agg	Q1	Agg	Q1
Change in market shares						
Stage 1 I-405 share	+0.000	+0.001	+0.002	+0.018	+0.003	+0.026
Stage 2 HOT share	+0.000	0.000	+0.001	+0.067	+0.004	+0.074
Change in drivers' average I-405 outcomes						
Price paid (\$)	+0.002	+0.003	-0.087	-0.093	-0.084	-0.172
Travel time (mins)	-0.013	-0.031	+0.085	+0.059	+0.067	+0.010
Time early (mins)	-0.088	-0.143	+0.976	+1.283	+0.595	+0.739
Time late (mins)	-0.007	-0.004	+0.053	-0.316	+0.022	-0.112
Change in welfare						
Driver surplus (\$/driver/year)	+0.698	+2.690	+2.409	+17.681	+5.848	+31.328
Toll revenue (\$/driver/year)	+0.295	+0.467	-8.402	-11.127	-10.019	-23.029
Net welfare (\$/driver/year)	+0.994	+2.986	-5.991	+9.281	-4.168	+21.311

Note: This table reports changes in outcomes under alternative pricing policies relative to the status quo, computed both in aggregate and for the bottom tract income quartile.

per driver per year. Drivers in the bottom tract income quartile, who drive the longest distances on I-405, gain even more relative to the status quo (+\$2.99). They also pay very slightly higher average tolls (+0.3 cents per day) but enjoy larger reductions in average travel time (-0.031 minutes) and time early (-0.143 minutes).

Neither version of the low-income discount improves on the status quo in aggregate, though they do benefit the bottom tract income quartile drivers who they target. Under these discounts, low-income drivers still impose the same externality on other drivers as their higher-income peers do, but they pay less, incentivizing them to take the toll lanes inefficiently often. Under the 50 percent discount, the HOT share for bottom-quartile drivers increases (+6.7 percentage points). These HOT trips also shift toward the most congested part of the morning, when prices and HOT time savings are high and the 50 percent discount is large in absolute terms. Aggregate welfare decreases by \$5.99 per driver per year relative to the status quo. Under the \$2 per trip discount, the HOT share for bottom-quartile drivers increases even more (+7.4 percentage points), but these new HOT trips are more uniformly distributed throughout the morning, resulting in a smaller aggregate welfare loss (-\$4.17 per driver per year). Both forms of income-based tolling result in higher travel times and lower toll revenue to be redistributed.

8 Conclusion

This paper studies the aggregate and distributional effects of highway congestion pricing, which is often thought to be regressive ([Ecola and Light, 2009](#); [Taylor, 2010](#)). We show that in our empirical

setting, it is actually *low-income* drivers who gain the most from status-quo tolling. Low-income drivers have high ex ante values of tolling because they drive the longest distances on our focal highway, in markets where time savings in the toll lanes are high and prices are relatively low. At the same time, these drivers, who are less willing or able to pay to avoid travel time, also derive considerable value from having the option but not the obligation to pay for speed.

While discussion of the potential regressiveness of congestion pricing often focuses on how low-income drivers must spend a greater fraction of their incomes on congestion fees, our results highlight that low-income drivers often also bear the greatest costs of unpriced traffic congestion. In many urban and suburban areas around the United States, high housing costs have pushed low-income households to live increasingly far away from job centers, where they must both drive longer distances to work and drive more often due to limited public transit. While in our setting, low-income drivers do pay higher tolls on average than other drivers—even in dollars, not only as a share of their incomes—we nonetheless find that congestion pricing is less regressive than completely unpriced congestion.

Our findings also point to the importance of the option value channel, which is relevant in potentially many other settings. The trade-off between paying with time versus paying with money is not unique to highway tolling. Both Uber and Lyft offer riders the choice between a more expensive trip with a faster pickup and a cheaper trip that requires additional waiting time. At airports, travelers can pay for CLEAR Plus in order to skip long airport security lines. Even at Disney theme parks, the Genie+ program allows paying customers access separate, faster “Lightning Lanes” for rides. In any setting where there is uncertainty about future prices or travel or wait times, even market participants who are unlikely to pay ex ante may still value having the option to pay if it turns out to be necessary ex post.

References

- Akbar, Prottoy A., Victor Couture, Gilles Duranton, and Adam Storeygard.** 2023. "The fast, the slow, and the congested: Urban transportation in rich and poor countries." <https://www.nber.org/papers/w31642>.
- Akerlof, George A.** 1978. "The Economics of "Tagging" as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning." *The American Economic Review*, 68(1): 8–19. Publisher: American Economic Association.
- Almagro, Milena, Felipe Barbieri, Juan Camilo Castillo, Nathaniel Hickok, and Tobias Salz.** 2024. "Optimal Urban Transportation Policy: Evidence from Chicago."
- Anderson, Michael L., and Lucas W. Davis.** 2020. "An empirical test of hypercongestion in highway bottlenecks." *Journal of Public Economics*, 187: 104–197.
- Arnott, Richard, André de Palma, and Robin Lindsey.** 1994. "The welfare effects of congestion tolls with heterogeneous commuters." *Journal of Transport Economics and Policy*, 28(2): 139–161.
- Arora, Kashish, Fanyin Zheng, and Karan Girotra.** 2020. "Private vs. pooled transportation: Customer preference and congestion management." <https://papers.ssrn.com/abstract=3701056>.
- Astor, Maggie.** 2017. "A \$40 toll to drive 10 miles? It happened on Virginia's I-66." *The New York Times*.
- Barwick, Panle Jia, Shanjun Li, Andrew R. Waxman, Jing Wu, and Tianli Xia.** 2024. "Efficiency and equity impacts of urban transportation policies with equilibrium sorting." *American Economic Review*, 114(10).
- Belenky, Peter.** 2011. "Revised departmental guidance on valuation of travel time in economic analysis." U.S. Department of Transportation.
- Bento, Antonio, Kevin Roth, and Andrew Waxman.** 2024. "The Value of Urgency: Evidence from Real-Time Congestion Pricing." *Journal of Political Economy Microeconomics*, 2(4): 786–851. Publisher: The University of Chicago Press.
- Berry, Steven T., James Levinsohn, and Ariel Pakes.** 1995. "Automobile prices in market equilibrium." *Econometrica*, 63(4): 841–890.
- Börjesson, Maria, Jonas Eliasson, Muriel B. Hugosson, and Karin Brundell-Freij.** 2012. "The Stockholm congestion charges-5 years on. Effects, acceptability and lessons learnt." *Transport Policy*, 20: 1–12.
- Brent, Daniel A., and Austin Gross.** 2018. "Dynamic road pricing and the value of time and reliability." *Journal of Regional Science*, 58(2): 330–349.
- Brownstone, David, and Thomas F. Golob.** 1992. "The effectiveness of ridesharing incentives: Discrete-choice models of commuting in Southern California." *Regional Science and Urban Economics*, 22(1): 5–24.
- Buchholz, Nicholas, Laura Doval, Jakub Kastl, Filip Matějka, and Tobias Salz.** 2025. "Personalized pricing and the value of time: Evidence from auctioned cab rides." *Econometrica*, 93(3).
- Bureau of Labor Statistics.** 2020. "Occupational employment and wages in Seattle-Tacoma-Bellevue — May 2019."
- Castillo, Juan Camilo.** 2023. "Who Benefits from Surge Pricing?" <https://papers.ssrn.com/abstract=3245533>.
- Castillo, Juan Camilo, Dan Knoepfle, and E Glen Weyl.** 2025. "Matching and pricing in ride hailing: Wild goose chases and how to solve them." *Management Science*, 71(5): 4377–4395.
- Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng.** 2024. "On

- Binscatter.” *American Economic Review*, 114(5): 1488–1514.
- Cho, Sungjin, Gong Lee, John Rust, and Mengkai Yu.** 2018. “Optimal dynamic hotel pricing.” https://economics.sas.upenn.edu/index.php/system/files/2018-04/hp_final_update.pdf.
- Cook, Cody, Aboudy Kreidieh, Shoshana Vasserman, Hunt Allcott, Neha Arora, Freek van Sambeek, Andrew Tomkins, and Eray Turkel.** 2025. “The Short-Run Effects of Congestion Pricing in New York City.” <https://www.nber.org/papers/w33584>.
- Couture, Victor, Gilles Duranton, and Matthew A. Turner.** 2018. “Speed.” *Review of Economics and Statistics*, 100(4): 725–739.
- Dahlgren, Joy.** 1998. “High occupancy vehicle lanes: Not always more effective than general purpose lanes.” *Transportation Research Part A: Policy and Practice*, 32(2): 99–114.
- D'Haultfœuille, Xavier, Ao Wang, Philippe Février, and Lionel Wilner.** 2022. “Estimating the gains (and losses) of revenue management.” <https://arxiv.org/abs/2206.04424v3>.
- Duranton, Gilles, and Matthew A. Turner.** 2011. “The fundamental law of road congestion: Evidence from US cities.” *American Economic Review*, 101(6): 2616–2652.
- Durrmeyer, Isis, and Nicolas Martinez.** 2022. “The welfare consequences of urban traffic regulations.” https://www.tse-fr.eu/sites/default/files/TSE/documents/doc/wp/2022/wp_tse_1378.pdf.
- Ecola, Liisa, and Thomas Light.** 2009. “Equity and congestion pricing.” *Rand Corporation*, 1–45.
- Giuliano, Genevieve, Douglas W. Levine, and Roger F. Teal.** 1990. “Impact of High Occupancy Vehicle Lanes on Carpooling Behavior.” *Transportation*, 17(2): 159–177.
- Glaeser, Edward L, Caitlin S Gorback, and James M Poterba.** 2023. “How regressive are mobility-related user fees and gasoline taxes?” *Tax Policy and the Economy*, 37(1): 1–56.
- Goldszmidt, Ariel, John A. List, Ian Muir, Robert D. Metcalfe, V. Kerry Smith, and Jenny Wang.** 2020. “The value of time in the United States: Estimates from nationwide natural field experiments.” <https://papers.ssrn.com/abstract=3748629>.
- Hallenbeck, Mark, Shirley Leung, Cory McCartan, CJ Robinson, Kiana Roshan Zamir, and Vaughn Iverson.** 2019. “I-405 Express Toll Lanes analysis: Usage, benefits, and equity.” <https://depts.washington.edu/trac/research-news/freeway-and-arterial-management/i-405-express-toll-lanes-analysis-usage-benefits-and-equity/>.
- Hall, Fred L.** 2005. “Traffic stream characteristics.” In *Revised Monograph on Traffic Flow Theory*. Ed. Nathan Garner, Carroll J. Messer and Ajay K. Rathi, Chapter 2, 1–32. Federal Highway Administration.
- Hall, Jonathan 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, Jonathan D.** 2021. “Can Tolling Help Everyone? Estimating the Aggregate and Distributional Consequences of Congestion Pricing.” *Journal of the European Economic Association*, 19(1): 441–474.
- Herzog, Ian.** 2024. “The city-wide effects of tolling downtown drivers: Evidence from London’s congestion charge.” *Journal of Urban Economics*, 144: 103714.
- KFF.** 2023. “Distribution of total population by federal poverty level.” <https://www.kff.org/other/state-indicator/distribution-by-fpl/?currentTimeframe=1>.
- Kim, Jinwon, Jucheol Moon, and Dong-Yun Yang.** 2023. “The deadweight loss of congestion: Estimation for California freeways.”
- Konishi, Hideo, and Se-il Mun.** 2010. “Carpooling and congestion pricing: HOV and HOT lanes.” *Regional Science and Urban Economics*, 40(4): 173–186.
- Kreindler, Gabriel.** 2024. “Peak-Hour Road Congestion Pricing: Experimental Ev-

- idence and Equilibrium Implications.” *Econometrica*, 92(4): 1233–1268. [eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA18422](https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA18422).
- Lam, Terence C., and Kenneth A. Small.** 2001. “The value of time and reliability: measurement from a value pricing experiment.” *Transportation Research Part E: Logistics and Transportation Review*, 37(2-3): 231–251.
- Leape, Jonathan.** 2006. “The London congestion charge.” *Journal of Economic Perspectives*, 20(4): 157–176.
- Lighthill, Michael James, and Gerald Beresford Whitham.** 1955. “On kinematic waves II. A theory of traffic flow on long crowded roads.” *Proceedings of the Royal Society of London Series A*, 229(1178): 317–345.
- Lindblom, Mike.** 2023. “A \$15 toll? How about \$18? WSDOT may blow the lid off I-405 express-lane prices.” *The Seattle Times*.
- Mangrum, Daniel, and Alejandro Molnar.** 2018. “The marginal congestion of a taxi in New York City.” https://www.danielmangrum.com/docs/Boro_current.pdf.
- Martin, Leslie A., and Sam Thornton.** 2018. “The margins of response to road use prices.”
- Mattia, Andrea.** 2022. “The value of time: Evidence from traffic congestion and Express Lanes.” https://www.dropbox.com/s/jp78r9agui91173/Mattia_JMP_VOT.pdf.
- Metropolitan Transportation Commission.** 2023. “I-880 Express Lanes now more affordable for low-income travelers.” <https://mtc.ca.gov/news/i-880-express-lanes-now-more-affordable-low-income-travelers>.
- Phang, Sock-Yong, and Rex S. Toh.** 2004. “Road congestion pricing in Singapore: 1975 to 2003.” *Transportation Journal*, 43(2): 16–25.
- Pigou, Arthur Cecil.** 1920. *The Economics of Welfare*. Palgrave MacMillan.
- Richards, Paul I.** 1956. “Shock waves on the highway.” *Operations Research*, 4(1): 42–51.
- Rosaia, Nicola.** 2023. “Competing platforms and transport equilibrium.” <https://scholar.harvard.edu/files/rosaia/files/draft.pdf>.
- Rosendorf, Linda.** 2018. “Virginia’s ‘Lexus lanes’ deserve their nickname.” *The Washington Post*.
- Russo, Antonio, Martin W. Adler, Federica Liberini, and Jos N. van Ommeren.** 2021. “Welfare losses of road congestion: Evidence from Rome.” *Regional Science and Urban Economics*, 89: 103692.
- Safirova, Elena, Kenneth Gillingham, Ian Parry, Peter Nelson, Winston Harrington, and David Mason.** 2004. “Welfare and Distributional Effects of Road Pricing Schemes for Metropolitan Washington DC.” *Research in Transportation Economics*, 9: 179–206.
- Samdahl, Donald, Myron Swisher, Jennifer Symoun, and Will Lisska.** 2013. “Congestion pricing: A primer on institutional issues.” Federal Highway Administration.
- Small, Kenneth A.** 2012. “Valuation of travel time.” *Economics of Transportation*, 1(1-2): 2–14.
- Small, Kenneth A., and Jia Yan.** 2001. “The value of ‘value pricing’ of roads: Second-best pricing and product differentiation.” *Journal of Urban Economics*, 49(2): 310–336.
- Small, Kenneth A., Clifford Winston, and Jia Yan.** 2005. “Uncovering the distribution of motorists’ preferences for travel time and reliability.” *Econometrica*, 73(4): 1367–1382.
- Small, Kenneth A., Clifford Winston, and Jia Yan.** 2006. “Differentiated road pricing, Express Lanes, and carpools: Exploiting heterogeneous preferences in policy design.” *Brookings-Wharton Papers on Urban Affairs*, 53–96.
- Sound Transit.** 2023. “Paying for regional transit.” <https://www.soundtransit.org/get-to-know-us/paying-regional-transit/car-tab-motor-vehicle-excise-tax-mvet>.
- Su, Yichen.** 2022. “The Rising Value of Time and the Origin of Urban Gentrification.” *American Economic Journal: Economic Policy*, 14(1): 402–439.
- Tarduno, Matthew.** 2022. “For whom the bridge tolls: Congestion, air pollution, and second-best

- road pricing.” <https://matt-tarduno.github.io/jmp/>.
- Taylor, Brian D.** 2010. “How fair is road pricing? Evaluating equity in transportation pricing and finance.” National Transportation Policy Project.
- Verhoef, Erik T., and Kenneth A. Small.** 2004. “Product differentiation on roads: Constrained congestion pricing with heterogeneous users.” *Journal of Transport Economics and Policy*, 38(1): 127–156.
- Vickrey, William S.** 1969. “Congestion theory and transport investment.” *American Economic Review: Papers & Proceedings*, 59(2): 251–260.
- Walters, Alan A.** 1961. “The theory and measurement of private and social cost of highway congestion.” *Econometrica*, 29(4): 676.
- Wang, Haizhong, Jia Li, Qian Yong Chen, and Daiheng Ni.** 2011. “Logistic modeling of the equilibrium speed–density relationship.” *Transportation Research Part A: Policy and Practice*, 45(6): 554–566.
- Washington State Department of Transportation.** 2023. “Ramp meters.” <https://wsdot.wa.gov/travel/operations-services/ramp-meters>.
- Washington State Transportation Commission.** 2021. “Low-income toll program study for I-405 & SR 167 Express Toll Lanes.”
- Williams, Kevin R.** 2022. “The Welfare Effects of Dynamic Pricing: Evidence From Airline Markets.” *Econometrica*, 90(2): 831–858. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA16180>.
- Wood, Nick, Vivek Gupta, James P. Cardenas, Jinuk Hwang, and Deepak Raghunathan.** 2021. “National inventory of specialty lanes and highways: Technical report.” Federal Highway Administration.
- Yang, Hai, and Hai-Jun Huang.** 1999. “Carpooling and congestion pricing in a multilane highway with high-occupancy-vehicle lanes.” *Transportation Research Part A: Policy and Practice*, 33(2): 139–155.
- Yang, Jun, Avralt Od Purevjav, and Shanjun 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.

A Additional figures and tables

Figure A.1: Example HOT access point



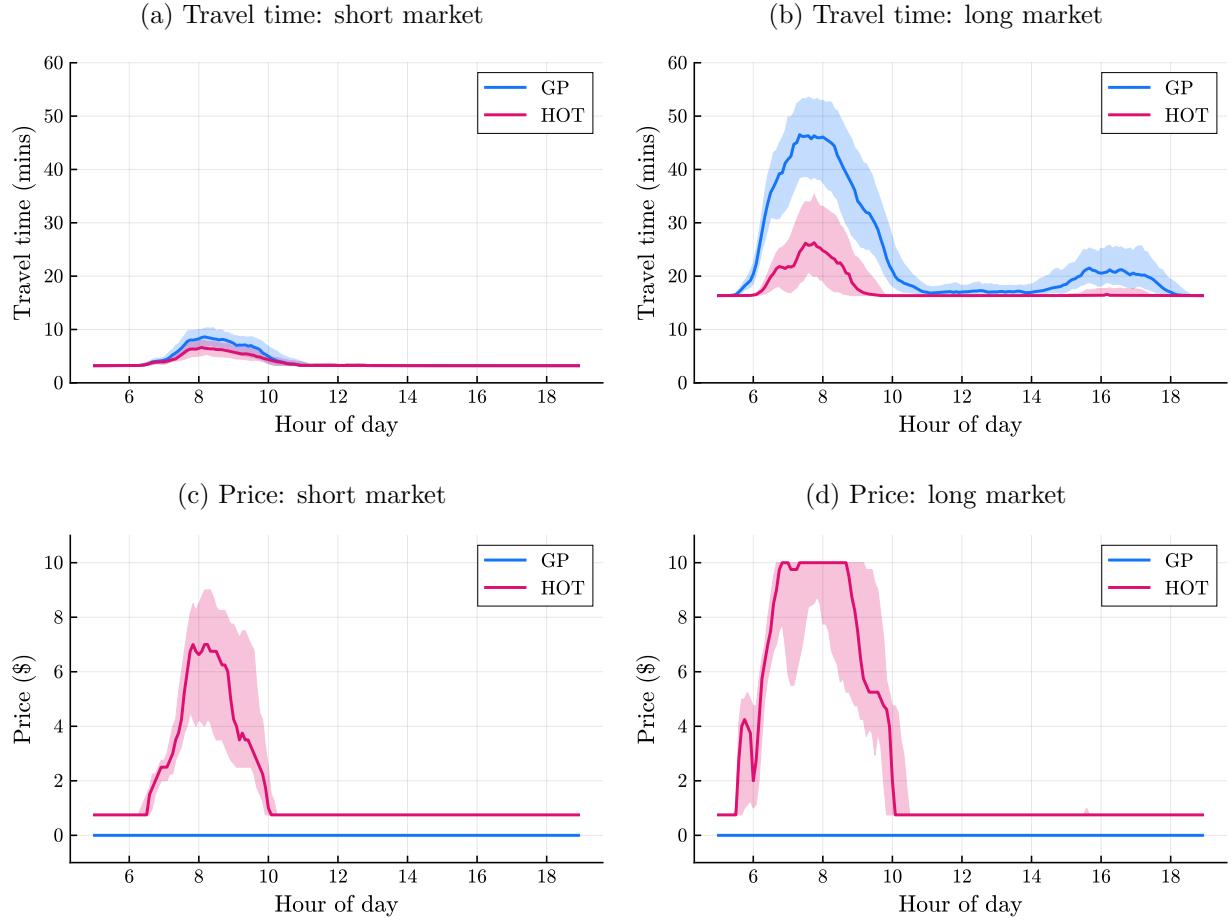
Note: This photograph shows an HOT access point on southbound I-405 just after the State Route 527 interchange. There are two GP lanes on the right and one HOT lane on the left. The GP and HOT lanes are separated at the bottom of the frame by double white lines; the access point is toward the top of the frame, where the double white lines become a single dashed line where drivers may legally cross. Above this access point, an electronic sign displays prices for travel to each downstream HOT access point (i.e., prices for each trip definition that begins at this access point). Source: Ken Lambert, *The Seattle Times*, <https://www.seattletimes.com/seattle-news/transportation/a-15-toll-how-about-18-wsdot-may-blow-the-lid-off-i-405-express-lane-prices/>

Table A.1: Southbound median traffic conditions by market

Market	Length (miles)	GP travel time (mins)		HOT travel time (mins)		Time savings (mins)		Price (\$)	
		Peak	Off-peak	Peak	Off-peak	Peak	Off-peak	Peak	Off-peak
KK	0.7	0.7	0.7	—	—	—	—	—	—
BB	1.1	1.2	1.5	—	—	—	—	—	—
WW	1.9	5.2	1.9	—	—	—	—	—	—
TK	3.2	4.3	3.2	4.0	3.2	0.1	0.0	1.75	0.75
LL	3.3	4.0	3.3	3.8	3.3	0.0	0.0	5.25	0.75
WT	4.4	8.5	4.5	8.3	4.5	0.0	0.0	0.75	0.75
KB	4.4	5.2	5.0	4.8	4.4	0.3	0.4	0.75	0.75
TB 1	6.9	9.1	7.6	7.8	6.9	0.9	0.6	0.75	0.75
TB 2	6.9	9.1	7.6	6.9	6.9	2.1	0.7	1.75	0.75
WK	7.0	12.2	7.1	11.5	7.1	0.3	0.0	0.75	0.75
LW	7.4	14.6	7.6	13.0	7.6	0.8	0.0	5.25	0.75
LT	10.0	18.2	10.2	10.7	10.0	6.5	0.2	5.00	0.75
WB	10.7	16.8	11.7	14.3	10.8	2.3	0.7	0.75	0.75
LK	12.6	22.5	12.9	14.0	12.6	7.6	0.2	5.25	0.75
LB	16.3	27.6	17.6	16.5	16.3	10.3	1.2	5.25	0.75

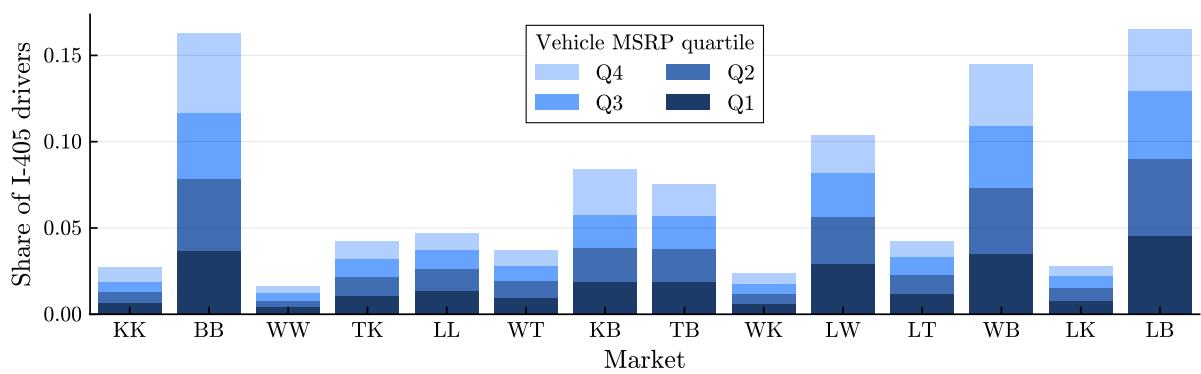
Note: Each observation is a (market, route, five-minute interval, day). Southbound peak observations are from 5 AM to 11 AM; off-peak observations are from 11 AM to 7 PM. Markets are ordered from shortest to longest. The three shortest southbound markets do not have a feasible HOT route. The Totem Lake to Bellevue market appears in two rows, labeled TB 1 and TB 2, because it has two HOT routes. The length in miles is equal to the travel time in minutes when traveling at the 60 mph speed limit.

Figure A.2: Travel time and price variation in two markets



Note: Figures show variation in travel time and price by route and time of day. In each five-minute interval, the thick line indicates the across-day median and the shaded area is between the 25th and 75th percentiles. Each underlying observation is a (market, route, five-minute interval, day) from 5 AM to 7 PM (tolled hours) in 2019. The short market is Totem Lake to Kirkland, the shortest southbound market with a feasible HOT route. The long market is Lynnwood to Bellevue, the longest southbound market.

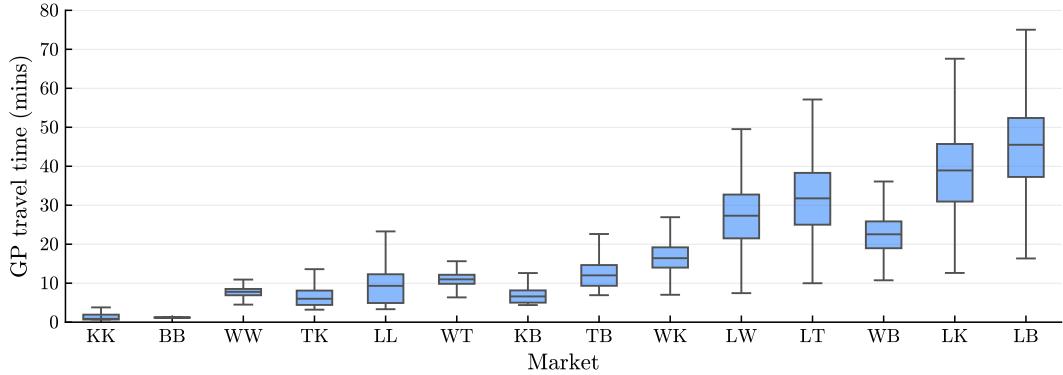
Figure A.3: Driver markets by car price quartile



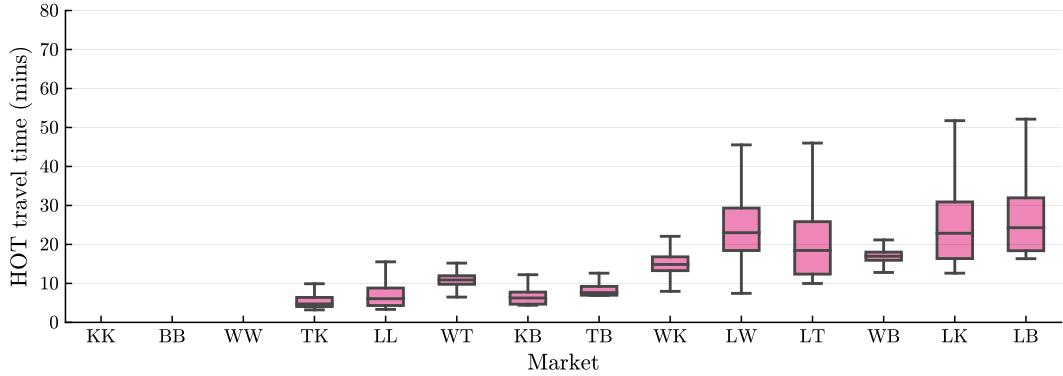
Note: Each underlying observation is a simulated driver in the unconditional sample of potential I-405 drivers. Markets are ordered from shortest to longest.

Figure A.4: Travel times and prices by market, 7–8 AM

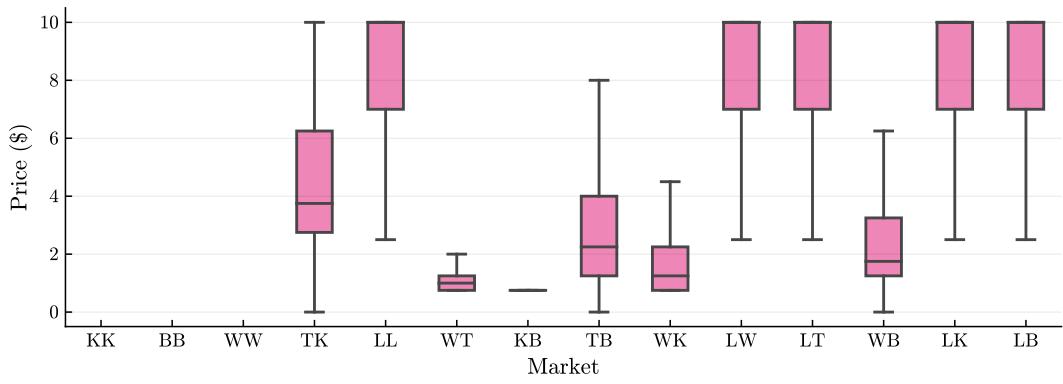
(a) GP travel time



(b) HOT travel time

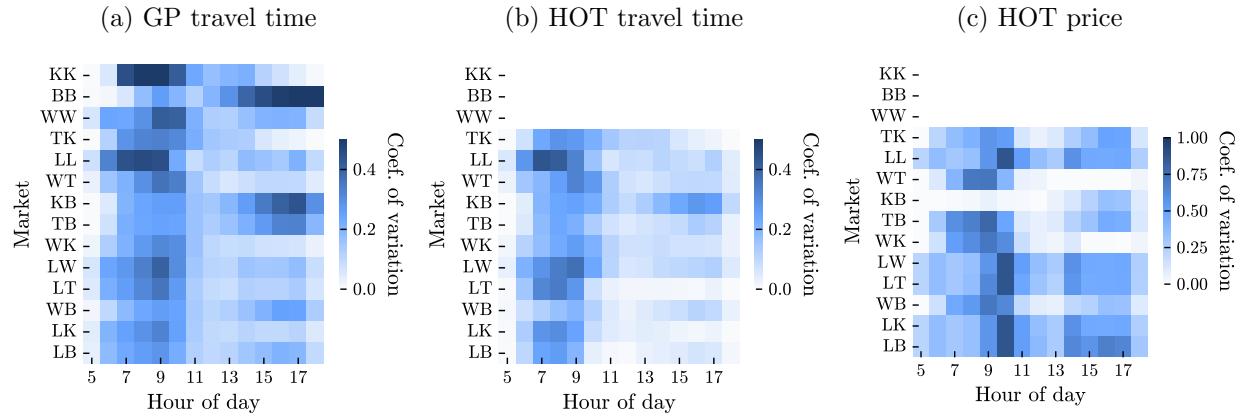


(c) HOT price



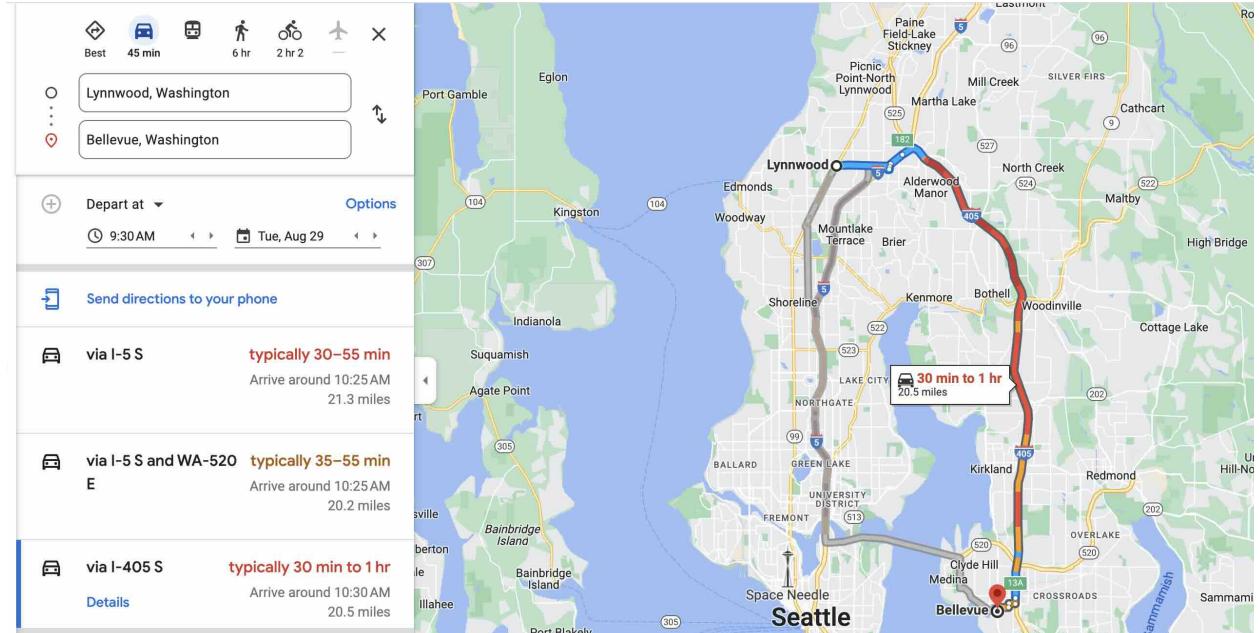
Note: Figures show the distributions of GP travel times (top panel), HOT travel times (middle panel), and HOT prices (bottom panel) from 7–8 AM in each southbound market. Boxes indicate the 25th percentile, median, and 75th percentile. Lower whiskers extend to the lowest observed data point that is within a distance of 1.5 times the interquartile range (IQR) from the 25th percentile. Likewise, upper whiskers extend to the highest observed data point within 1.5 times the IQR from the 75th percentile. Markets are ordered from shortest to longest. The first three markets have no feasible HOT route.

Figure A.5: Coefficient of variation by market-hour



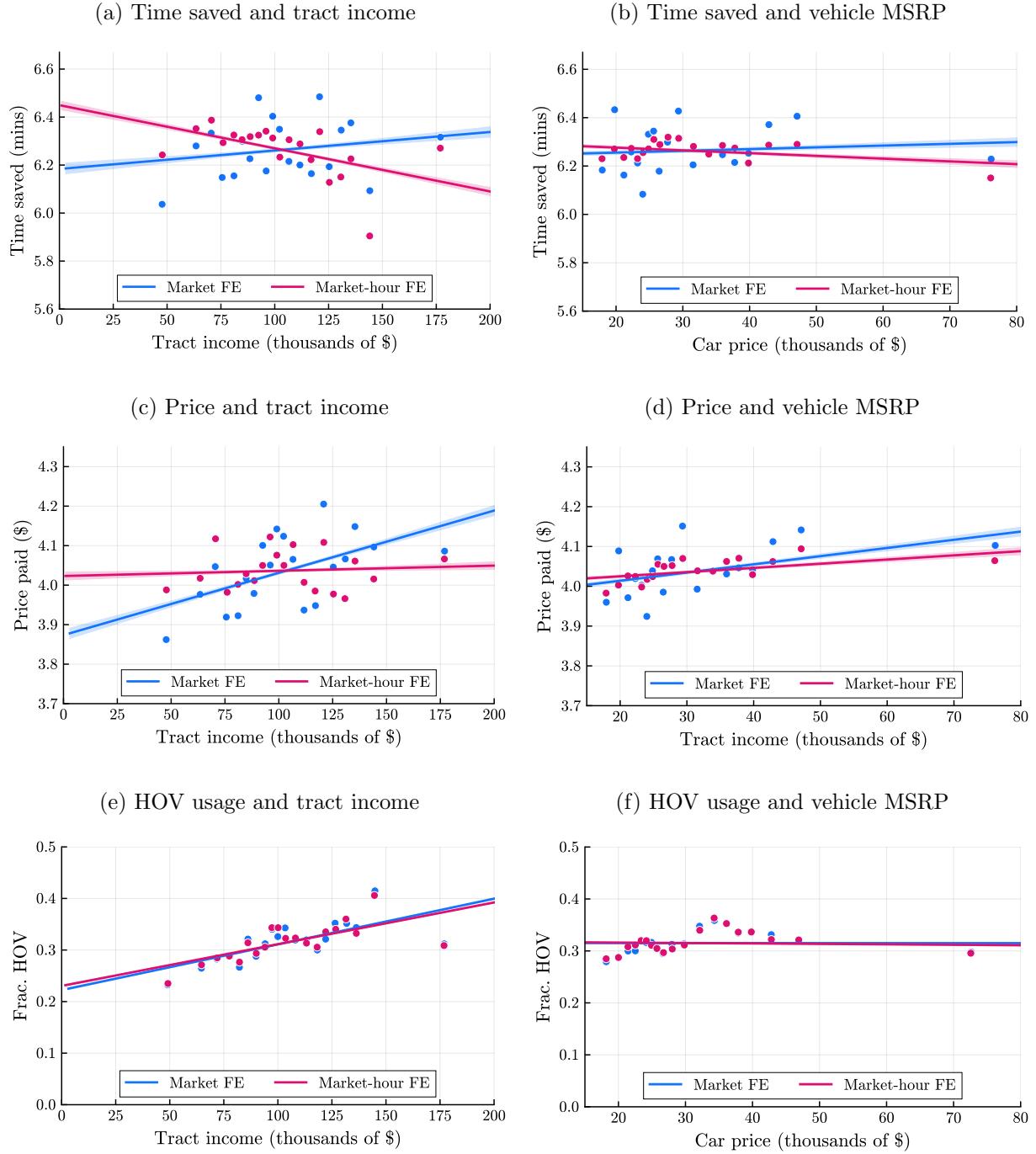
Note: Figures show the average coefficient of variation—i.e., the standard deviation divided by the mean—for travel times and prices by market-hour, conditional on the calendar quarter and day of week. Markets are ordered by shortest (at the top) to longest (at the bottom). The three shortest markets have no feasible HOT route.

Figure A.6: Sample Google Maps query



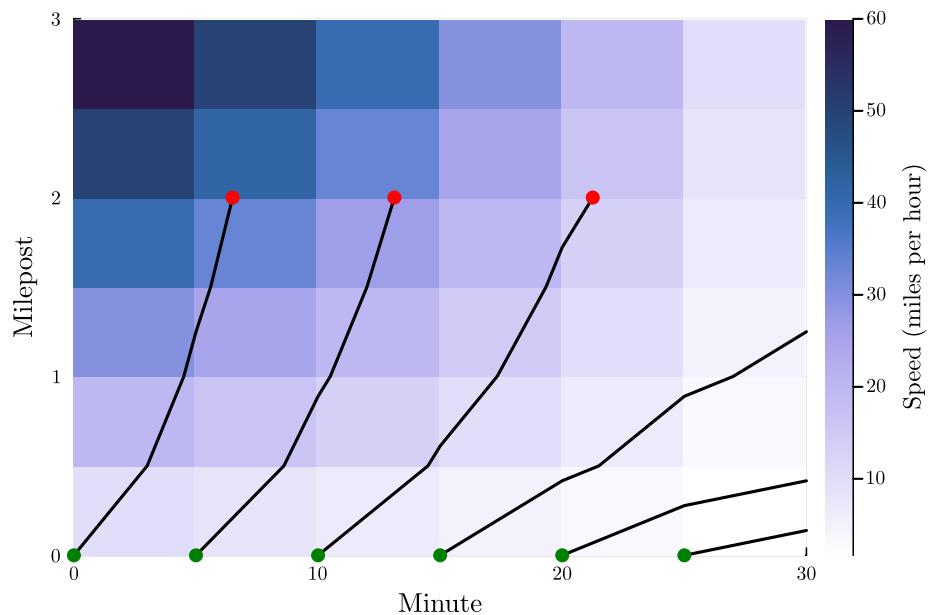
Note: Figure shows an example of the range of travel times Google provides when requesting routing information for one hour later the same morning. The shown route traverses the full length of I-405 with HOT lanes. Google Maps travel times are for the GP lanes. The image was taken at 8:30AM on August 29th, 2023 for a trip leaving at 9:30AM.

Figure A.7: Binscatters of HOT trip characteristics and driver income proxies



Note: These figures document the relationship between drivers' income proxies and HOT trip characteristics. For the first four panels, each underlying observation is a paid peak-hour toll transaction. The final two panels include all peak-hour transactions, including HOV. The shaded areas correspond to 95 percent confidence intervals estimated using Cattaneo et al. (2024).

Figure A.8: Example trajectories in spacetime



Note: This figure depicts example trajectories for a driver traveling from milepost 0 to milepost 2 at different starting times. Each trajectory begins at a green dot and ends at a red dot. In each shaded rectangle, the slope of the driver's trajectory equals the speed at that discretized location and time.

Table A.2: Demand parameter estimates: preference coefficients and ideal arrival times

(a) Stage 2 coefficients

	Income heterogeneity (μ^α)		Covariance (Σ^α) of unobserved heterogeneity		
	Mean ($\bar{\alpha}$)	Tract income	Car price	Price	Travel time
Price	-0.937 [-2.294, -0.699]	-0.0038 [-0.0055, 0.0080]	0.0023 [-0.0049, 0.0178]	0.0642 [0.0000, 1.7990]	0.3622 [-3.1214, 3.2502]
Travel time	-25.331 [-33.995, -0.967]	-0.1201 [-0.1988, 0.2619]	-0.1644 [-0.3574, 0.2212]	0.3622 [-3.1214, 3.2502]	2.0660 [0.0914, 64.2496]
Time early	-3.161 [-6.513, -1.928]	-	-	-	-
Time late	-3.579 [-9.707, -1.783]	-	-	-	-

(b) Ideal arrival time distribution

Mean ($\bar{\eta}$, not estimated)	8:30 AM
Standard deviation (σ^η)	0.873 hours [0.660, 1.585]

Note: These tables report demand parameter estimates. In panel a, the time variables—travel time, time early, and time late—are measured in hours. Tract income and car price are expressed in deviations from the population mean, in thousands of dollars.

Table A.3: Demand parameter estimates: intercepts

(a) Stage 1 intercepts

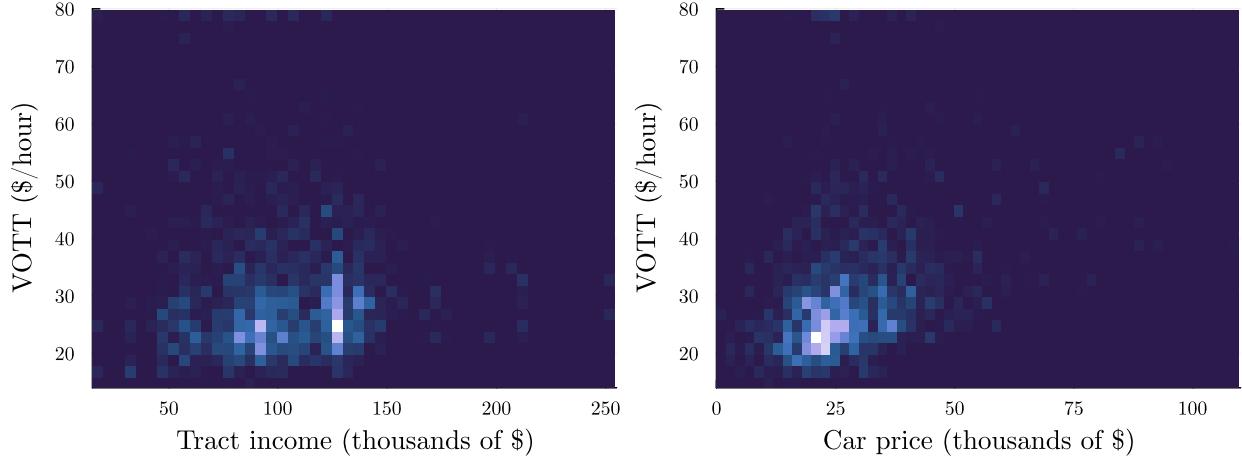
	SOV drivers (β_{SOV}^0)	HOV drivers (β_{HOV}^0)				
		L	W	T	K	B
Lynnwood	8.412 [-1.057, 11.331]	-3.029 [-3.780, 0.707]	4.228 [-2.833, 12.298]	4.564 [-1.347, 5.830]	5.713 [-3.126, 18.994]	4.570 [-2.833, 10.133]
Woodinville	4.430 [-2.509, 6.774]	-	-	14.706 [1.700, 34.674]	2.315 [-4.352, 9.845]	3.746 [-1.322, 15.080]
Totem Lake	3.538 [-1.085, 5.672]	-	-	-	9.382 [-0.134, 23.705]	2.238 [0.022, 26.483]
Kirkland	-0.606 [-3.914, 0.437]	-	-	-	-	-0.408 [-3.427, 3.429]
Bellevue	-0.607 [-2.349, 0.070]	-	-	-	-	-

 (b) Stage 2 intercepts (α^0)

	L	W	T	K	B
Lynnwood	-0.006 [-0.394, 0.196]	-2.139 [-2.667, -0.893]	0.007 [-0.259, 0.376]	1.359 [1.217, 3.563]	0.751 [0.012, 2.667]
Woodinville	-	-	-1.966 [-2.072, -0.983]	1.733 [1.395, 3.367]	1.443 [0.589, 2.384]
Totem Lake 1	-	-	-	-2.138 [-2.504, -1.428]	0.061 [-0.239, 0.331]
Totem Lake 2	-	-	-	-	0.072 [-0.276, 0.232]
Kirkland	-	-	-	-	-0.418 [-0.786, 0.337]
Bellevue	-	-	-	-	-

Note: These tables report demand parameter estimates. In each panel, the market origins (rows) and destinations (columns) are ordered from northernmost to southernmost. Totem Lake appears twice in panel b because the Totem Lake to Bellevue market has two HOT routes.

Figure A.9: Value of travel time vs. income proxies



Note: Figures show the joint distributions of value of travel time with tract income (left panel) and car price (right panel). Each underlying observation is a simulated driver from the unconditional distribution of driver characteristics.

Table A.4: Speed-density relationship parameter estimates

Param	Description	Estimate
\underline{v}	Jam speed	7.567 (0.017)
\bar{v}	Freeflow speed	63.932 (0.006)
δ_1	Scale	2.495 (0.012)
δ_2	Asymmetry	0.100 (0.001)
ρ^*	Critical density	26.194 (0.015)

Note: This table reports estimates of the parameters of the asymmetric logistic speed-density relationship (7), estimated via nonlinear least squares. Heteroskedasticity-robust standard errors are in parentheses.

Table A.5: HOV-to-HOT changes in outcomes by tract income quartile

	Q1	Q2	Q3	Q4
Share of potential drivers	0.233	0.262	0.254	0.250
Change in market shares				
Stage 1 I-405 share	+0.065	+0.054	+0.048	+0.048
Stage 2 HOT share	+0.256	+0.238	+0.185	+0.193
Change in drivers' average I-405 outcomes				
Price paid (\$)	+0.632	+0.393	+0.314	+0.244
Travel time (mins)	-0.562	-0.361	-0.330	-0.484
Time early (mins)	-0.594	+0.044	-0.632	+0.276
Time late (mins)	-0.805	+0.052	-0.931	-0.627
Change in welfare				
Driver surplus (\$/driver/year)	+145.137	+78.327	+32.895	+31.663
Toll revenue (\$/driver/year)	+91.054	+51.738	+26.242	+17.460
Net welfare (\$/driver/year)	+190.977	+124.167	+78.735	+77.503

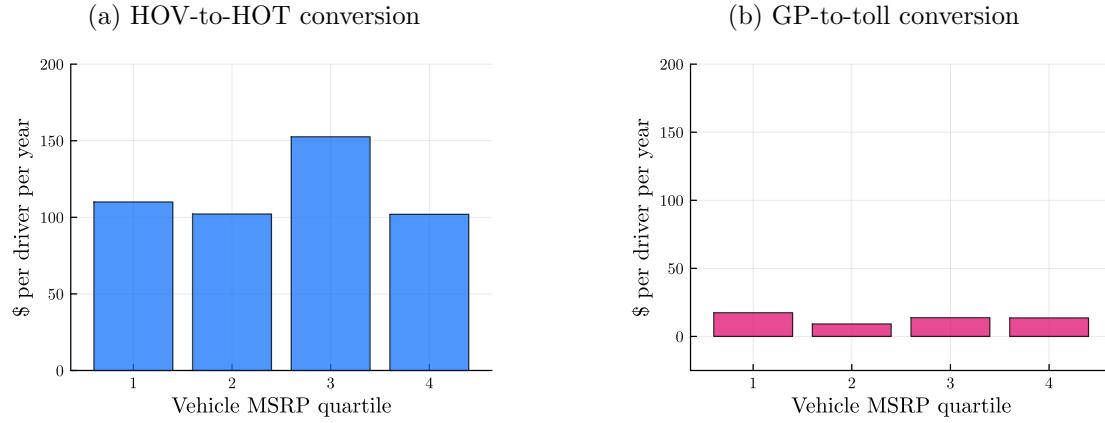
Note: This table reports differences in outcomes between the GP-or-HOV counterfactual and status quo tolling separately by quartiles of tract income. The stage 2 shares and average I-405 outcomes are conditional on choosing I-405—that is, conditional on *not* choosing the non-405 outside option. “Toll revenue” refers to revenue collected from drivers of each type, not revenue uniformly redistributed to drivers.

Table A.6: GP-to-toll changes in outcomes by tract income quartile

	Q1	Q2	Q3	Q4
Share of potential drivers	0.233	0.262	0.254	0.250
Change in market shares				
Stage 1 I-405 share	-0.007	-0.005	-0.053	-0.064
Stage 2 HOT share	+0.316	+0.319	+0.241	+0.236
Change in drivers' average I-405 outcomes				
Price paid (\$)	+0.892	+0.563	+0.430	+0.344
Travel time (mins)	-1.249	-0.805	-0.490	-0.460
Time early (mins)	-1.402	-0.245	-0.436	-0.306
Time late (mins)	+0.357	+0.172	-0.211	+0.060
Change in welfare				
Driver surplus (\$/driver/year)	-37.148	-34.528	-55.154	-58.243
Toll revenue (\$/driver/year)	+117.447	+67.611	+36.446	+21.884
Net welfare (\$/driver/year)	+22.710	+25.330	+4.704	+1.615

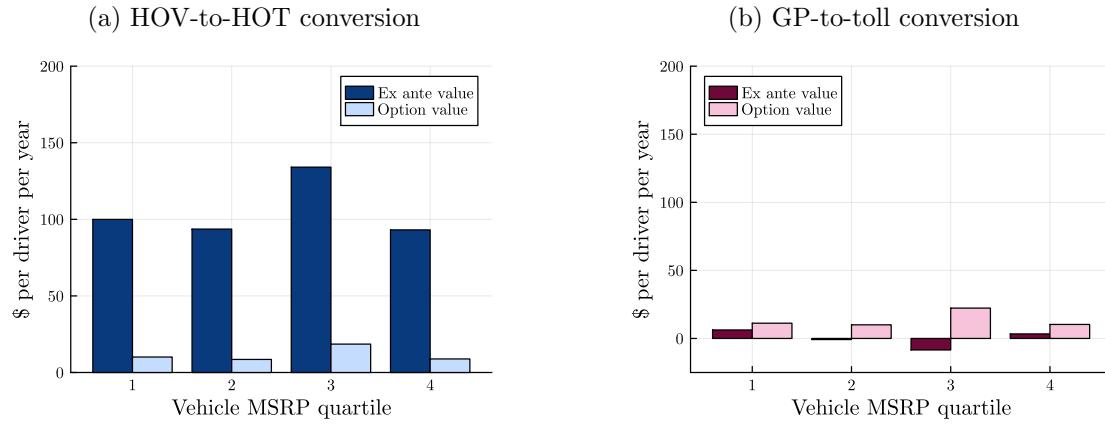
Note: This table reports differences in outcomes between the GP-or-HOV counterfactual and status quo tolling separately by quartiles of tract income. The stage 2 shares and average I-405 outcomes are conditional on choosing I-405—that is, conditional on *not* choosing the non-405 outside option. “Toll revenue” refers to revenue collected from drivers of each type, not revenue uniformly redistributed to drivers.

Figure A.10: Welfare changes by vehicle MSRP quartile



Note: These figures show changes in welfare by drivers' vehicle MSRP quartile.

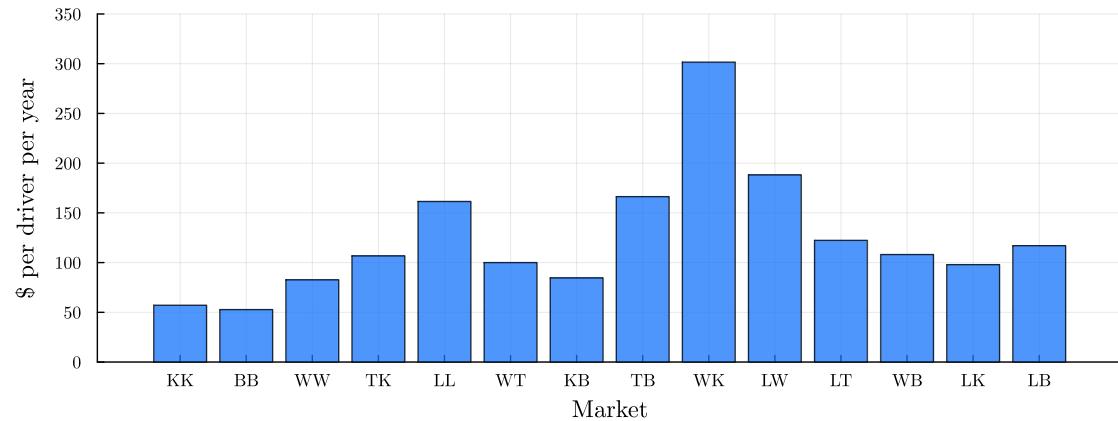
Figure A.11: Welfare decomposition by vehicle MSRP quartile



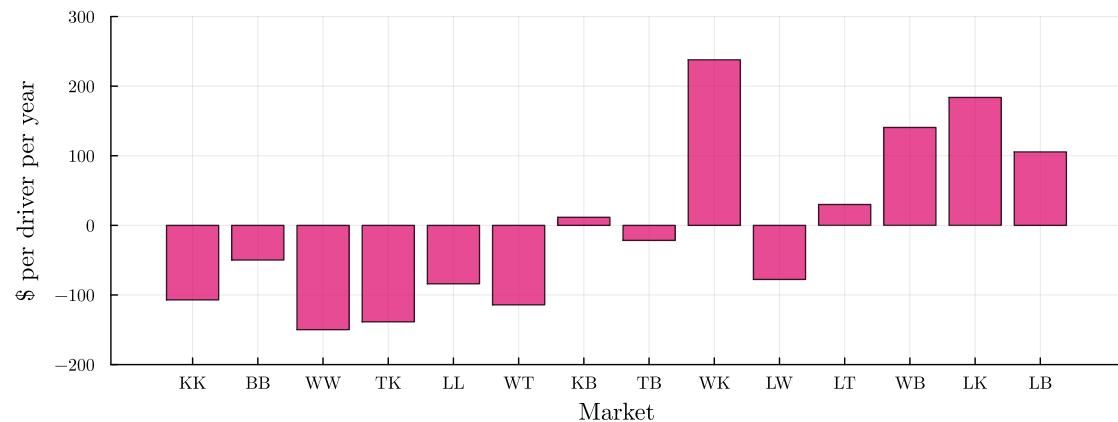
Note: These figures show the decomposition of welfare gains into ex ante and option value, broken out by drivers' vehicle MSRP quartile.

Figure A.12: Welfare changes by market

(a) HOV-to-HOT conversion



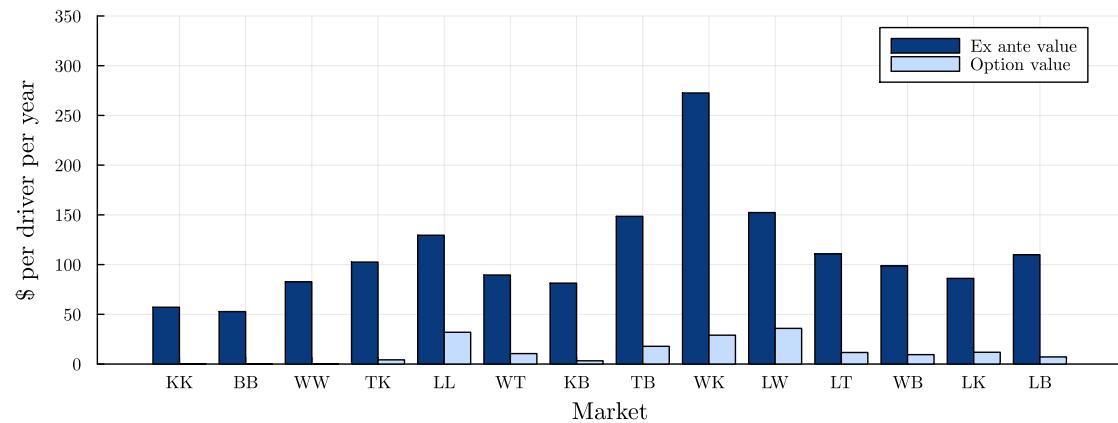
(b) GP-to-toll conversion



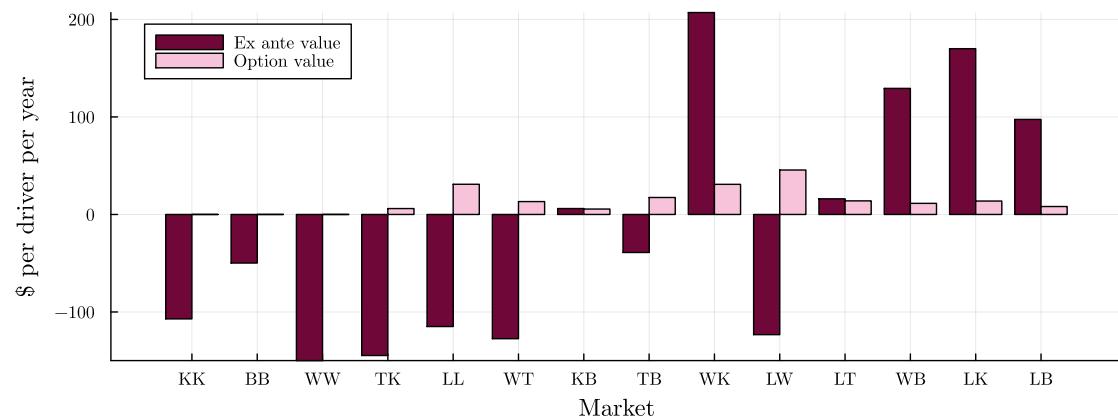
Note: These figures show changes in welfare for each southbound market. Markets are ordered from shortest to longest.

Figure A.13: Welfare decomposition by market

(a) HOV-to-HOT conversion



(b) GP-to-toll conversion



Note: These figures show the decomposition of welfare gains into ex ante and option value separately for each southbound market. Markets are ordered from shortest to longest.

B Data appendix

B.1 Market definitions

In order to define our set of markets, we introduce some additional terminology.

An **interchange** is a location on the highway where drivers can travel between I-405 and other roads or highways. Moving from north to south in Figure B.14, the interchanges on the tolled section of I-405 are: I-5, SR 527, NE 195th St, SR 522, NE 160th St, NE 128th St, NE 124th St, NE 116th St, NE 85th Pl, NE 70th Pl, SR 520, NE 8th St, NE 6th St, and NE 4th St. Each interchange may include multiple on- and off-ramps. For example, the SR 520 interchange has on- and off-ramps to and from both eastbound and westbound SR 520.

We call a group of interchanges a **town**. We define the following five towns, again moving from south to north:

- Lynnwood: all points north of the I-5 interchange on SR 525, I-5, and SR 527
- Woodinville: NE 195th St, SR 522, and NE 160th St
- Totem Lake: NE 128th St, NE 124th St, and NE 116th St
- Kirkland: NE 85th Pl and NE 70th Pl
- Bellevue: SR 520, NE 8th St, NE 6th St, NE 4th St, and all points south of NE 4th St on the untolled section of I-405

Finally, a **market** combines a direction of travel with an entry town and an exit town. The entry town and exit town may be the same: for example, travel from I-5 to SR 527 occurs in the southbound Lynnwood to Lynnwood market. The exception is travel from Totem Lake to Totem Lake, which is infeasible in both directions. There are therefore fourteen markets in each direction.

Taking the HOT route is not feasible in every market. We say that the HOT route is **feasible** in a given market if there is a way for drivers to enter the highway at an on-ramp, enter the HOT lanes at a legal access point (indicated by a white triangle in Figure B.14), exit them at a legal access point (indicated by a black triangle in the same figure), and finally exit the highway at an off-ramp. The HOT route is feasible in the southbound Lynnwood to Lynnwood market: drivers can enter I-405 at the I-5 interchange, enter the HOT lanes at the topmost white triangle, exit them at the topmost black triangle, and finally exit I-405 at the SR 527 interchange. It is not feasible in the southbound Bellevue to Bellevue market, as there are no legal HOT entry points south of SR 520.

B.2 Approximating density using speed and throughput

While the induction loops do not record densities, we approximate density at each loop in each five-minute interval by dividing throughput by speed. Identity (10) specifies an *instantaneous* relationship between the speed, density, and throughput at each point in space-time. Our density calculation is an approximation because we are using time-averaged measurements of speed and throughput. This approximation is very accurate at higher speeds, when vehicles are nearly uniformly distributed throughout both space and time, and slightly worse at lower speeds (Hall, 2005).

Figure B.14: I-405 Express Lanes map

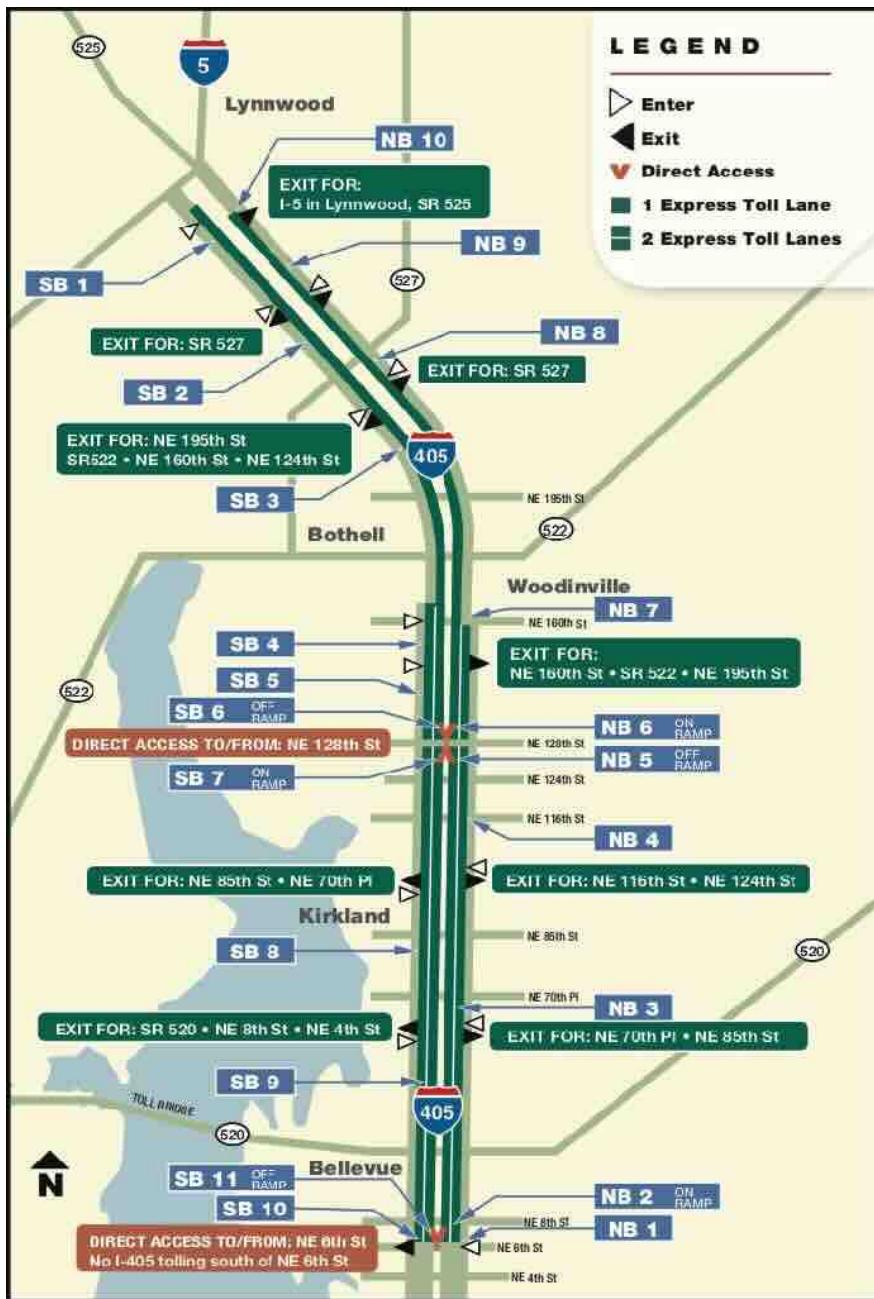
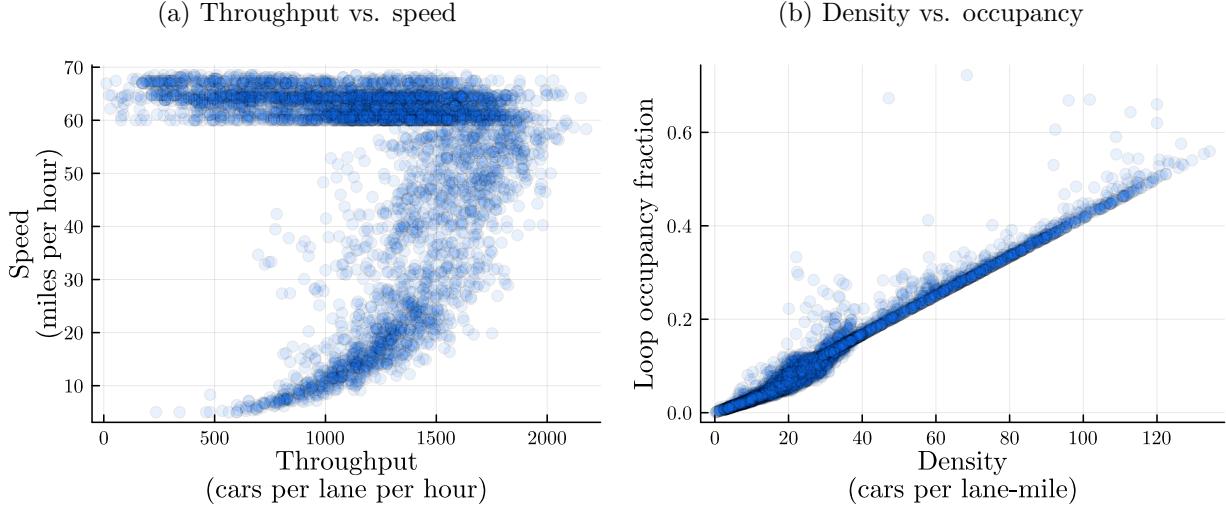


Figure B.15a shows the underlying speed and throughput data.

To validate our approximation, we compare our estimated densities to observed loop occupancy rates, which are measured directly by the induction loops. The occupancy rate is the fraction of time the loop has a vehicle passing over it. In theory, density equals occupancy multiplied by the average vehicle length. Figure B.15b shows that our estimated density is indeed linear in the observed occupancy rate. The R^2 from regressing observed occupancy on estimated density is 0.975.

Figure B.15: Relationships between loop traffic variables



Note: Figures show a random sample of 5000 southbound observations, where each observation is a (loop, five-minute interval, date) from 5–11 AM in 2019. Top-coded speeds have been replaced with estimated freeflow speeds.

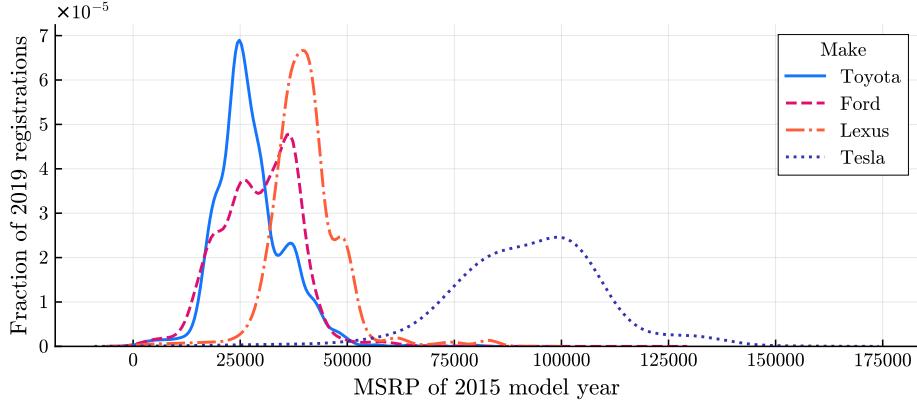
B.3 Estimating vehicle MSRPs

We use transaction-level data on vehicle registrations in Washington State, obtained via a public disclosure request from the Washington State Department of Licensing. For each transaction, we observe attributes including the date of registration; the make (e.g., Toyota), model (e.g., Prius), and model year of the vehicle being registered; the home Census tract of the registrant; and the amount of motor vehicle excise tax (MVET) paid. Our sample contains the universe of over 38 million vehicle registrations in the state from January 2017 to December 2022. Vehicle owners are required to register their vehicles when they move to the state and to renew their registrations annually; both types of transactions require MVET payments.

For each Seattle-area registration, we compute the manufacturer-suggested retail price (MSRP) implied by the amount of MVET paid. The MVET is levied on vehicle registrants living within the Sound Transit District. This district covers parts of three Seattle-area counties served by the Central Puget Sound Regional Transit Authority; it contains the I-405 Express Toll Lanes that we study in this paper. The MVET is a fraction of the vehicle's depreciated value, which is in turn

computed from the vehicle’s MSRP and a depreciation schedule (Sound Transit, 2023). We use the 13 million registrations with positive MVETs (i.e., registrations occurring in the Sound Transit District) beginning in March 2017, when the excise tax rate increased to 1.1 percent. At this stage, we discard the 0.0025 percent of registrations where the estimated MSRP is more than \$500,000; these observations likely reflect data errors or mistakes made by the registrant. Figure B.16 shows the distributions of estimated MSRPs for four sample makes.

Figure B.16: Estimated MSRPs: vehicles with model year 2015 registered in 2019



Finally, we match each I-405 HOT driver’s car to an MSRP estimate, which we use as our measure of the car price. If possible, we match the car to the median MSRP estimated for vehicles with the same make, model, and year. If not, we match it to the median MSRP among registered vehicles with the same make and model. If still not, we use the median MSRP among registered vehicles with the same make. Among HOT drivers with vehicle information, 43 percent are matched to an MSRP using the make, model, and year; 12 percent are matched using the make and model only; 33 percent are matched to the make only; and the remaining 12 percent are unmatched.

B.4 Market sizes and characteristics

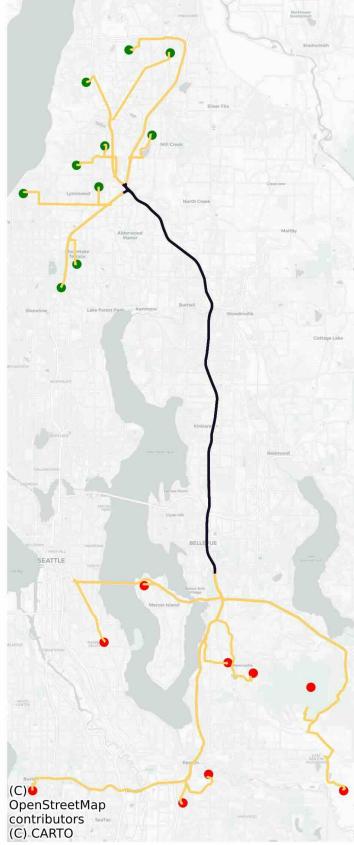
Let $r \rightarrow s$ denote an ordered pair of Census tracts. We define tract pair $r \rightarrow s$ as belonging to market m if one of the top three driving routes from tract r to tract s (according to the OpenStreetMaps router) involves taking I-405 in market m . Each tract pair belongs to at most one market.

First, we compute the size of each market m on each date t using the Replica GPS data. Slightly abusing notation, let $q_{r \rightarrow s, H, t}$ denote the historical travel flows from tract r to tract s in hour H on date t . Replica estimates these quantities based on trips taken by a sample of GPS devices. We compute q_{mt} , the size of market m on date t , by summing over tract pairs in the market and over hours $H \in \{5, \dots, 10\}$ in the morning peak:

$$q_{mt} = \sum_{(r \rightarrow s) \in m} \sum_H q_{r \rightarrow s, H, t}$$

Figure B.17 illustrates this procedure for $m =$ Lynnwood to Bellevue.

Figure B.17: Example market size construction: Lynnwood to Bellevue



Note: Each green dot is the centroid of an origin Census tract. Each red dot is the centroid of a destination Census tract. The figure shows a sample of the tract pairs for which taking I-405 southbound from Lynnwood to Bellevue (the full length of the tolled section) is one of the top three suggested routes. The black path indicates the part of the route on I-405.

Next, we construct the joint distribution of carpool status C , tract income X , and car price Y in each market m .⁴⁹ Each origin tract r is associated with a unique HOV share \bar{c}_r , a unique tract income value x_r , and a distribution $G_r^Y(y)$ of car prices from registrations in that tract. Let $G_r^{C,X,Y}(c,x,y)$ denote the joint distribution of HOV status, tract income, and car price in tract r . We assume that carpool status in tract r is distributed Bernoulli with probability \bar{c}_r and independent of car price in that tract.

Let ω_{rm} denote the share of total market m travel that originates in tract r :

$$\omega_{rm} = \frac{\sum_s \sum_H \sum_t q_{r \rightarrow s, H, t}}{\sum_t q_{mt}}$$

⁴⁹This distribution can't be obtained directly from the toll transaction data because that sample includes only drivers who have taken the toll lanes at least once.

The joint distribution $F_m^{C,X,Y}(c, x, y)$ of carpool status, tract income, and car price in market m is the weighted average of that joint distribution in each origin tract:

$$F_m^{C,X,Y}(c, x, y) = \sum_r \omega_{rm} G_r^{C,X,Y}(c, x, y)$$

B.5 On-ramp metering

As discussed in Section 3.2, each on-ramp contains both metered (general-purpose) and unmetered (carpool-only) lanes, and the metered throughputs reflect rationed demand for I-405 departure times.⁵⁰ True demand for 8 AM departures may be higher than observed 8 AM throughput because non-carpooling drivers must queue in order to exit the off-ramp and enter the highway. This queueing represents a loss of information from the researcher's perspective: a single queue exit rate profile can be rationalized by potentially many queue entry rate profiles.

Figure B.18 shows intraday variation in the distributions of speeds and throughputs in the metered and unmetered lanes of an example southbound on-ramp. In the unmetered lane, speed is always constant at (the topcoded value of) 60 miles per hour. Throughput varies smoothly throughout the day, peaking in the morning between 7–8 AM and largely paralleling the profile of intraday travel time variation in Figure A.2. In contrast, in the metered lane, speed drops discontinuously to 10 miles per hour from 6–10 AM. Throughput increases from 5–6 AM, but is flat from 6–10 AM while metering is in place. Throughput then remains elevated after 10 AM, due to a combination of queue emptying and true demand for post-10 AM departure times.

Our approach is to reallocate the metered quantities so that they match the profile of unmetered quantities within each day and ramp. To formalize this, we first define some notation which will be used in this appendix only. Let $\tilde{\mathcal{H}}$ denote the set of five-minute intervals during tolled hours between 5 AM and 7 PM. Let \tilde{q}_{jht} denote the raw throughput in *lane type* (unmetered or metered) $j \in \{0, 1\}$ at time h on ramp r . We use the raw unmetered quantities q_{0hrt} and impute the metered quantities q_{1hrt} as follows:

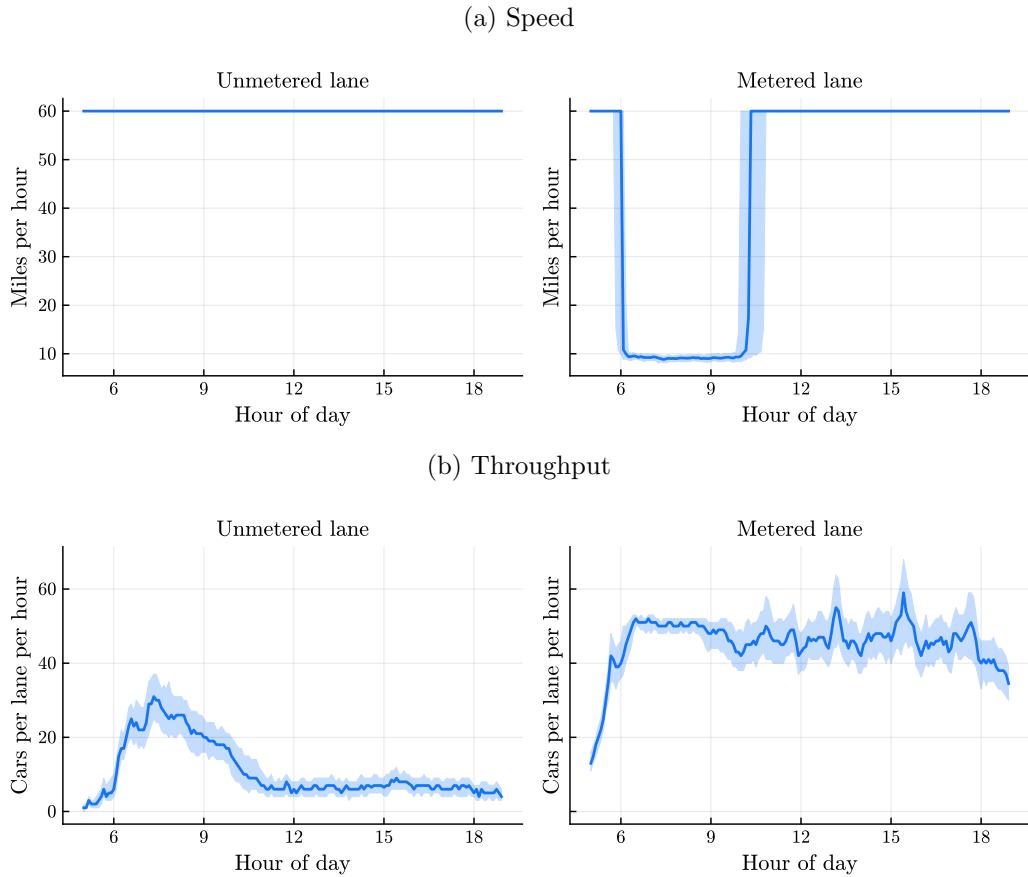
$$q_{0hrt} = \tilde{q}_{0hrt}$$

$$q_{1hrt} = \underbrace{\frac{\tilde{q}_{0hrt}}{\sum_{h' \in \tilde{\mathcal{H}}} \tilde{q}_{0h'rt}}}_{\text{share of unmetered departures at time } h \text{ on day } t} \times \underbrace{\sum_{h' \in \tilde{\mathcal{H}}} \tilde{q}_{1h'rt}}_{\text{total metered departures on day } t}$$

Figure B.19 shows the distributions of raw and imputed quantities in each market origin (summing across lane types and on-ramps).

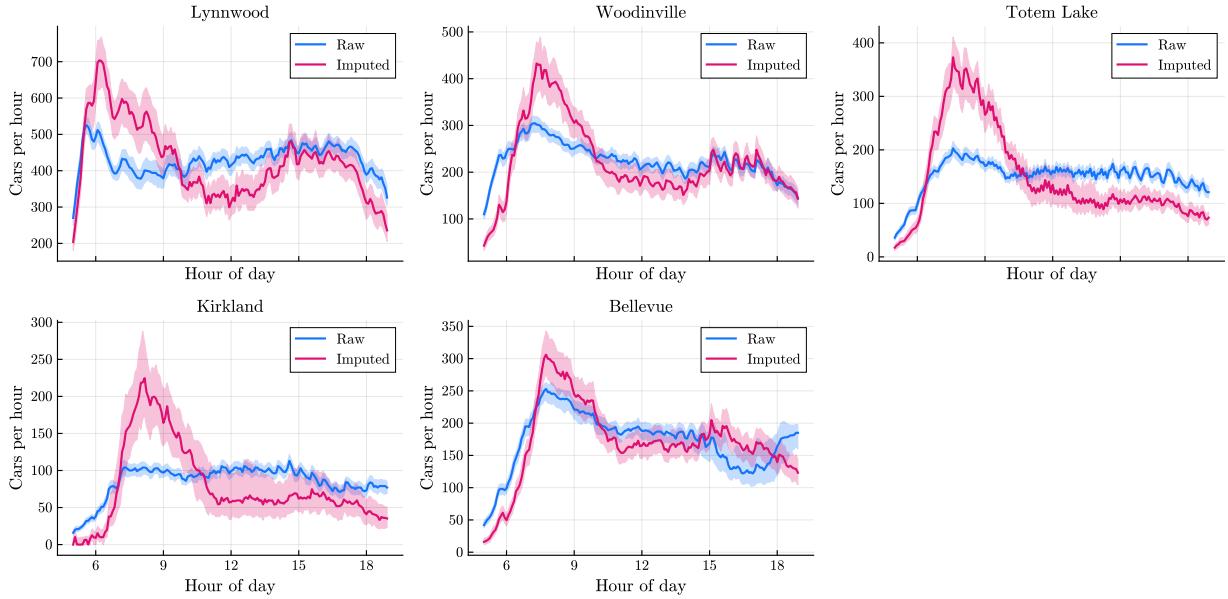
⁵⁰The WSDOT website writes of the design and goals of ramp metering: "Ramp meters are a specific type of traffic signal used to control how quickly vehicles enter traffic flow on a freeway, and are a freeway operation strategy designed to reduce collisions and decrease travel times. Ramp meters function by controlling the rate (metering) at which vehicles enter the freeway... Ramp meters typically operate during peak congestion times: 6 AM to 9 AM, and 3 PM to 7 PM. Meters may still be operated outside these hours, as their operation depends on freeway traffic

Figure B.18: Ramp metering example: southbound SR 522 on-ramp



Note: Figures show variation in speed and throughput by (unmetered or metered) lane and time of day. In each five-minute interval, the thick line indicates the across-day median and the shaded area is between the 25th and 75th percentiles. Each underlying observation is a (loop, five-minute interval, day) from 5 AM to 7 PM (tolled hours) in 2019. This ramp, which takes drivers from eastbound SR 522 to southbound I-405, has an unmetered carpool-only lane and a metered general-purpose lane.

Figure B.19: Raw vs. imputed departure time quantities



Note: Figures show variation in raw and imputed departure time quantities in each southbound market origin. In each five-minute interval, the thick line indicates the across-day median and the shaded area is between the 25th and 75th percentiles. Each underlying observation is a (market origin, five-minute interval, day) from 5 AM to 7 PM (tolled hours) in 2019.

This approach amounts to assuming that single-occupancy drivers demand the same distribution of departure times as carpoolsing drivers. This assumption is potentially violated if carpoolsing drivers, who face reduced-price or free travel in the HOT lanes (depending on whether they have two or three-plus people in the car), have greater demand for peak-hour travel than single-occupancy drivers. In this case, we will tend to underestimate the average disutilities of price and travel time in the population.

C Descriptives appendix

C.1 Regression discontinuity: price rounding

Prices are rounded to the nearest 25 cents, which creates almost-random variation in realized prices for unrounded values close to the threshold. We recover the unrounded prices by reconstructing the algorithm based on its source code, which was shared by WSDOT. To evaluate the effects around the discontinuity, we regress outcomes for a given trip definition, 5-minute interval, and day t on an indicator for binned distance to threshold, with fixed effects for trip definition times hour of day. We subset data to peak hours in each direction, i.e. southbound 5–11 AM and northbound 1–7 PM. We also exclude observations where our recovered unrounded price does not round to the

speeds and volumes, and not on time of day” ([Washington State Department of Transportation, 2023](#)).

observed price. This occurs for about 1% of observations, and is likely due to some “final check” rules that are applied before displaying a price but are outside of the algorithm source code shared with us (e.g., that a trip from A to B must cost less than a longer trip from A to C). When we turn to structural estimation, we further restrict the data to southbound AM peak hours (so that we can estimate on the model-simulated data) and estimate the relationships using market-level data.

Figure C.20a plots the effects on the trip fares paid. In practice, we occasionally observe trips that do not pay the new, rounded price. At the discontinuity, the realized price paid increases by about \$0.2 instead of \$0.25. Figure C.20b plots the effects on time savings for HOT trips. The time saved is unchanged at the threshold itself, although some points beyond the threshold show higher time savings.

Some HOT entry points have road signs displaying the expected travel times in GP and HOT lanes, offering drivers better information. Figure C.21 shows that the sign-reported HOT time savings are continuous through the threshold. Drivers entering at points with and without signs are similarly responsive to the price rounding, with perhaps a slightly larger effect at entry points with signs.

C.2 Differences-in-differences: car crashes

Our second source of variation comes from car crashes in the GP lanes, which, conditional on time of day and location, are plausibly exogenous. To evaluate the impacts of a crash, we use data at the trip definition and 5-minute level and define an observation as “treated” if there is a crash within 60 minutes and within 8 miles of the start of the trip. We restrict the sample to southbound and northbound peak hours and drop observations with multiple crashes in the same direction that day and observations with a crash that is outside of our definition of “treated.” For extreme crashes that shut down multiple lanes for an extended period, the normal pricing algorithm will be overrode so that the HOT lanes are free for all drivers; we also exclude these observations from the sample.

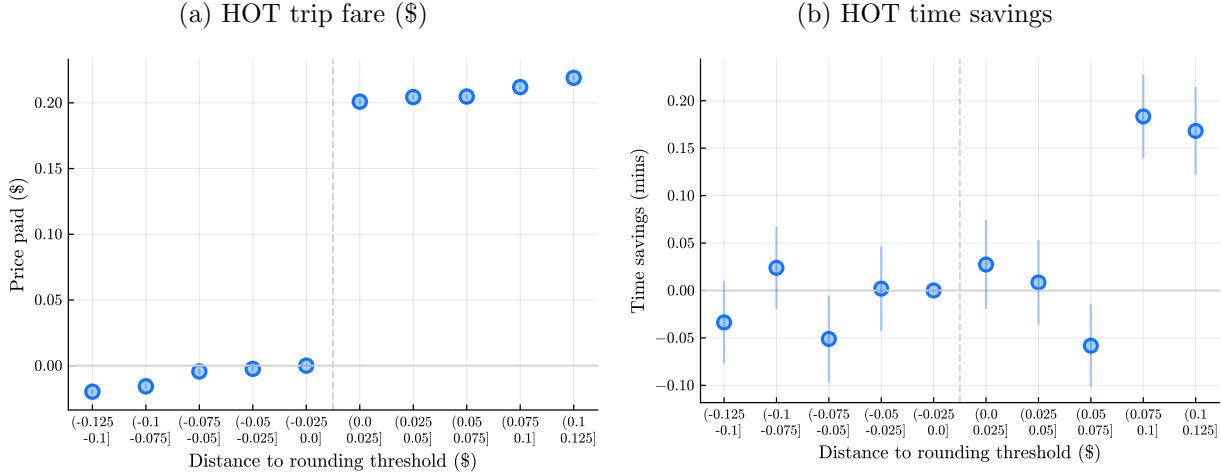
Figure C.22 documents the effects of crashes on HOT prices and travel times in both lanes. We regress each outcome on indicators for bins of time since crash with interacted (trip definition, five-minute interval) fixed effects. After a crash, prices increase by about \$0.30, although there is some evidence of a pre-trend. This slight pre-trend could be due, in part, to the misreporting of the crash timing. Average travel times in each lane also increase by about a minute, with a marginally greater effect on GP lane travel times than HOT travel times.

As with the price rounding variation, when we turn to the structural estimation, we further restrict the sample to SB peak hours (so we can match the simulated data) and estimate on market level instead of trip definition level data.

C.3 Event study: long-run speed and throughput

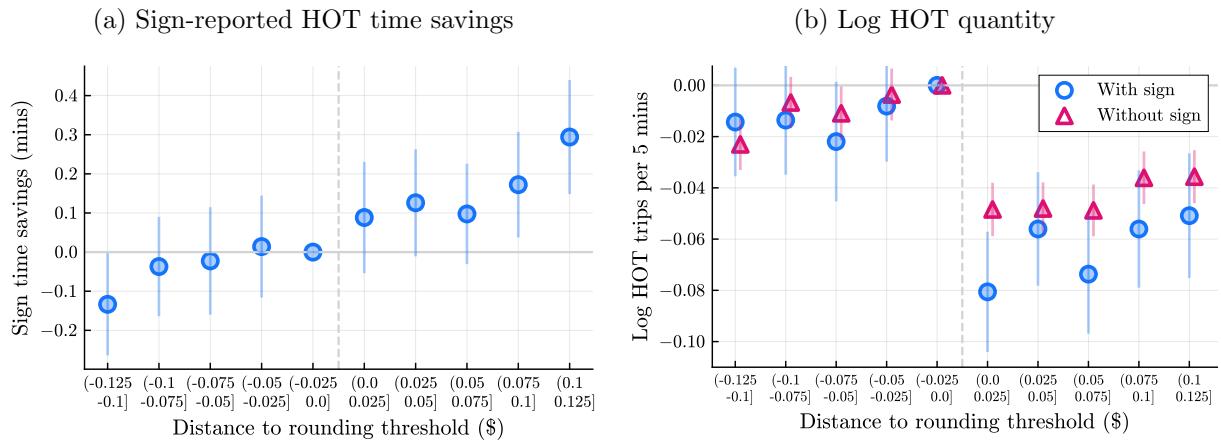
The HOT lanes opened in September 2015. While our transaction data does not start until 2019, we can observe speed and throughput of each lane before and after opening. We estimate changes in average speed and throughput by year relative to HOT opening and (GP or HOV/HOT) route.

Figure C.20: HOT time savings and fare by distance to price rounding threshold



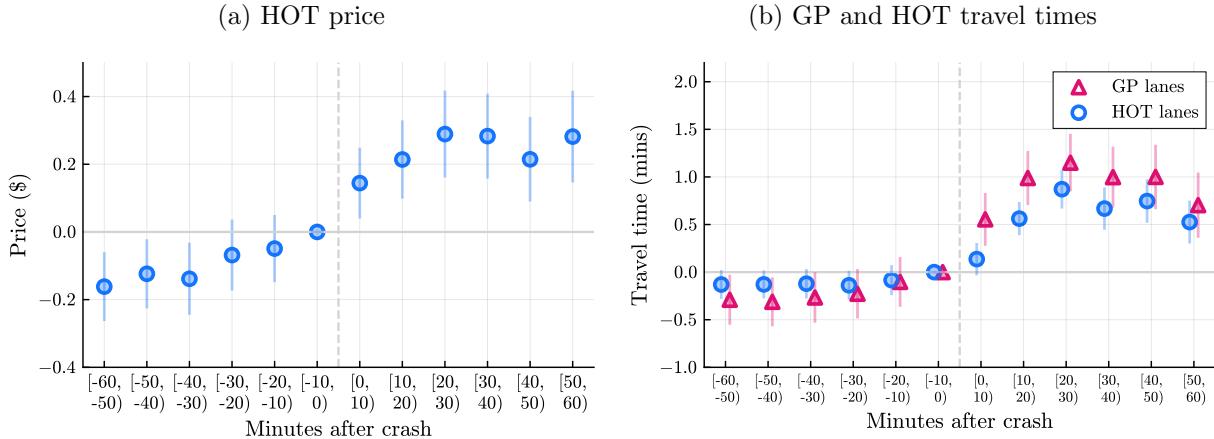
Note: This figure plots coefficients from a regression of HOT fares and time savings on indicators for bins of the unrounded price's distance to the rounding threshold, estimated with threshold and (trip definition, hour) fixed effects. Observations are at the (trip definition, five-minute interval, day) level. Data are subset to southbound 5–11 AM and northbound 1–7 PM.

Figure C.21: Price rounding discontinuity: trip definitions with travel time signs



Note: These figures plot coefficients from regressions of sign-reported HOT time savings (left panel) and log HOT quantity (right panel) on indicators for bins of the unrounded price's distance to the rounding threshold, estimated with threshold and (trip definition, hour) fixed effects. The top panel shows separate estimates for trip definitions with and without travel time signs. The bottom panel is estimated only on trip definitions with travel time signs. Observations are at the (trip definition, five-minute interval, day) level. Data are subset to southbound 5–11 AM and northbound 1–7 PM.

Figure C.22: Price and HOT time savings before and after crashes



Note: These figures plot coefficients from regressions of price (left panel) and travel times (right panel) on indicators for bins of time since crash, estimated with interacted (trip definition, five-minute interval) fixed effects. Observations are at the (trip definition, five-minute interval, day) level. Observations are “treated” if they have a crash with 60 minutes and within 8 miles of the start point for the trip. We exclude observations on days with multiple crashes in the same direction and observations with a crash that occurs outside of the “treated” definition limits. Data are subset to southbound 5–11 AM and northbound 1–7 PM. Vertical lines represent 95 percent confidence intervals.

We start by averaging loop speeds and throughputs at the (loop, route, hour, year relative to HOT open) level. The sample is peak hours, southbound 5–11 AM and northbound 1–7 PM, on weekdays from 2011–2019. We then estimate

$$y_{ijht} \sim \text{route}_j \times \text{year}_t + \text{loop}_i + \text{hour}_h \quad (9)$$

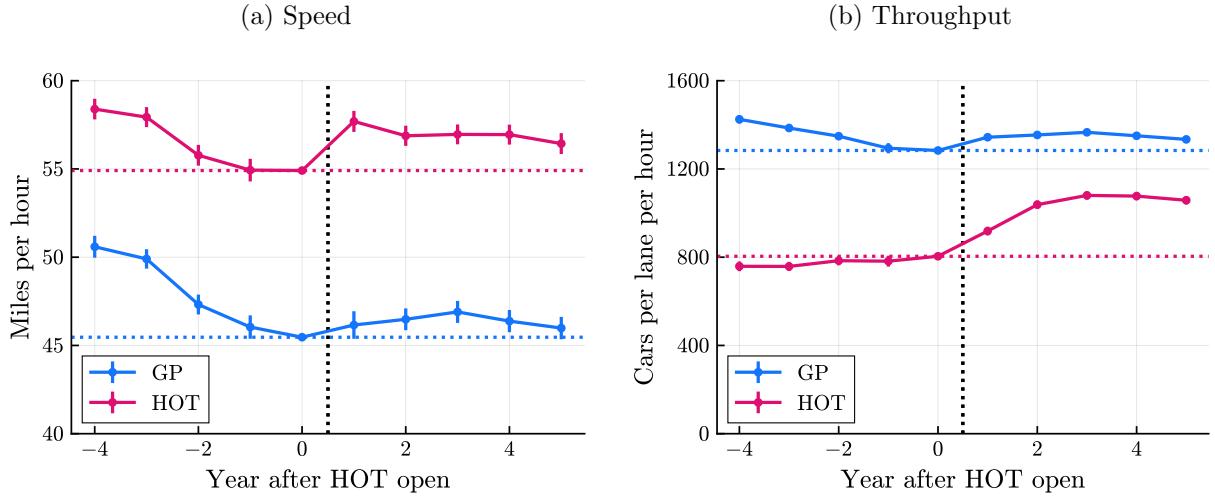
where the two outcomes y_{ijht} are speed and throughput for loop i embedded in route $j \in \{\text{HOT}, \text{GP}\}$, 5-minute interval h , and day t . Figure C.23 shows the estimated coefficients on the route-year interactions, where the omitted base level is the GP route in the last (non-calendar) year before the HOT lanes open.

Before the HOT lanes open, the GP and then-HOV lanes appear to be moving along different parts of the backward-bending speed-throughput curve.⁵¹ In the GP lanes, speed and throughput are both decreasing, indicating that these lanes are becoming more and more congested over time. In contrast, in the HOV lanes, speed is decreasing but throughput is *increasing*, suggesting instead that the then-carpool-only lanes are underutilized. HOV speeds average 10 miles per hour faster than GP speeds during this period.

After the HOT lanes open, speed and throughput increase in the short run and begin to decline again in the longer run. The short-run increases are due to the expansion of highway capacity

⁵¹Throughput is low both when speed is very low and when speed is very high (Figure D.29b). Recall that throughput is the number of vehicles passing over a given point in space per unit of time. When speed is low, then vehicles take a long time to pass over that point. When speed is high, car-following distances are also high, so that few cars pass over that point per unit of time. Throughput is maximized at an intermediate speed, the max-throughput speed. We can think of the road as being overutilized at lower speeds and underutilized at higher speeds.

Figure C.23: Long-run changes in aggregate speed and throughput

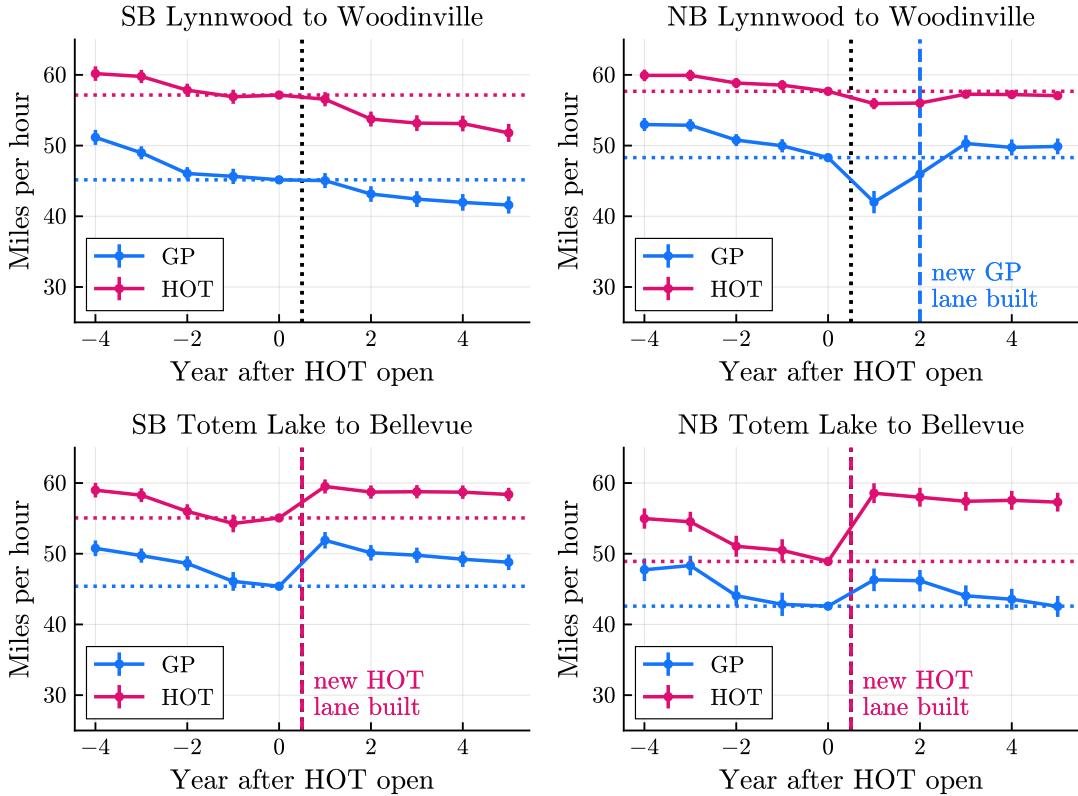


Note: These figures report changes in average speed and throughput by year relative to HOT opening and (GP or HOV/HOT) route. Each point is an estimated coefficient on a route-year interaction in equation (9), with the level normalized to the GP speed or throughput in the last year before the HOT lanes opened. The error bars show 95 percent confidence intervals. The sample is peak hours, southbound 5–11 AM and northbound 1–7 PM, on weekdays from 2011–2019.

along certain highway segments: a newly constructed HOT lane in September 2015, paired with the introduction of tolling, and a new GP peak-use shoulder lane in April 2017. Estimating equation (9) separately by road segment, we find that HOT throughput increases on all segments (Figure C.25), but speed increases are concentrated on the road segments with the new construction (Figure C.24). Starting about three years after the HOT opening, speed and throughput begin to fall in both lane types. However, the presence of pricing dampens demand increases, so that speed falls more slowly than before tolling was introduced.

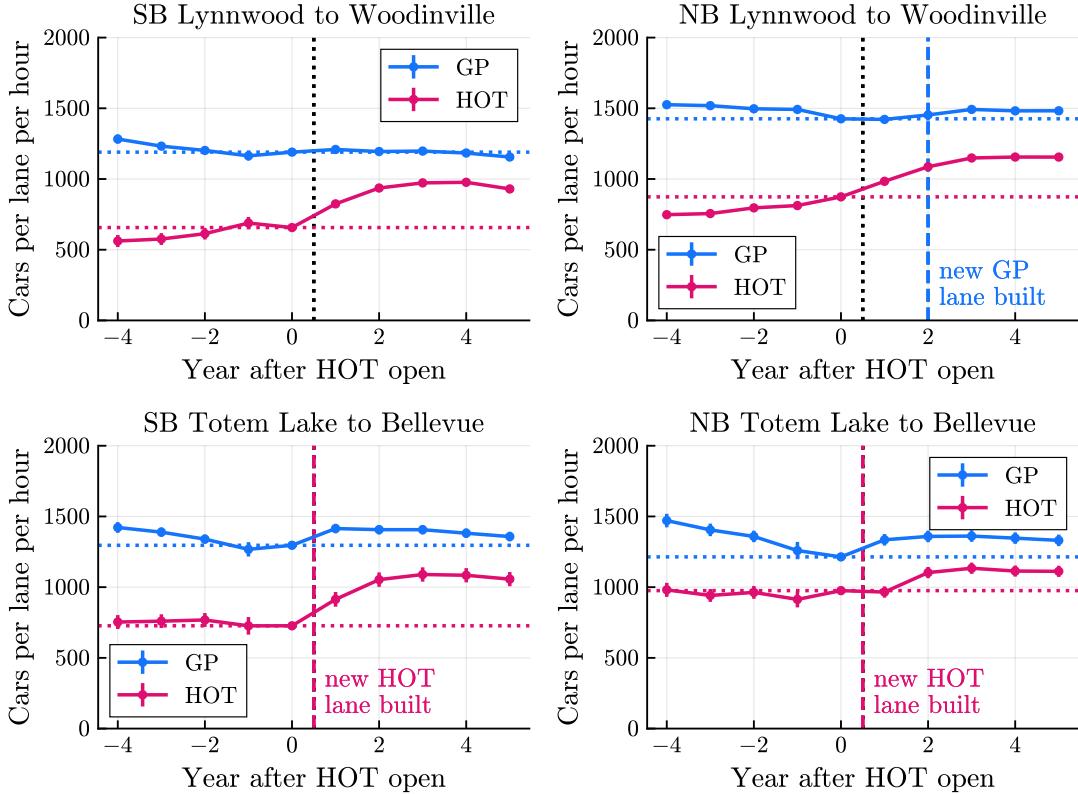
This analysis has two main limitations, which together motivate the need for our structural model. First, since the HOT lane opening bundled together several policy changes (detailed at the end of Section 2), it is challenging to satisfactorily separate the effects of the introduction of pricing versus the construction of additional HOT lanes. Even analysis at the road segment level is imperfect, since traffic on a given segment has spillover effects onto traffic on nearby segments. Second, since the 2015 HOT opening falls outside the 2019 sample period of our toll transaction data, this long-run analysis cannot speak to heterogeneous effects by driver characteristics. To overcome these limitations, in Section 5, we introduce an equilibrium model of highway travel, which we estimate and use to simulate welfare under counterfactual equilibria.

Figure C.24: Long-run changes in speed by road segment



Note: These figures report changes in average speed by year relative to HOT opening, (GP or HOV/HOT) route, and road segment. Each point is an estimated coefficient on a route-year interaction in equation (9), with the level normalized to the GP speed in the last year before the HOT lanes opened. The error bars show 95 percent confidence intervals. The pink dashed lines indicate the September 2015 construction of an additional HOT lane on the southern half of the corridor. The bottom two panels correspond to the segment with new construction. The blue dashed lines indicate the April 2017 opening of a GP peak-use shoulder lane on a northbound segment in the northern part of the corridor. The sample is peak hours, southbound 5–11 AM and northbound 1–7 PM, on weekdays from 2011–2019.

Figure C.25: Long-run changes in throughput by road segment



Note: These figures report changes in average throughput by year relative to HOT opening, (GP or HOV/HOT) route, and road segment. Each point is an estimated coefficient on a route-year interaction in equation (9), with the level normalized to the GP throughput in the last year before the HOT lanes opened. The error bars show 95 percent confidence intervals. The pink dashed lines indicate the September 2015 construction of an additional HOT lane on the southern half of the corridor. The bottom two panels correspond to the segment with new construction. The blue dashed lines indicate the April 2017 opening of a GP peak-use shoulder lane on a northbound segment in the northern part of the corridor. The sample is peak hours, southbound 5–11 AM and northbound 1–7 PM, on weekdays from 2011–2019.

C.4 Effects of precipitation

We use data from the Paine Field weather station, which is near the start of the southbound I-405 HOT lanes, which we download from the Iowa Environmental Mesonet. We define the morning peak hours (5–11 AM) as having precipitation if there is at least 0.01 inches of recorded precipitation, which is the threshold the National Weather Service uses for reporting a “chance of rain” in its forecasts. By this definition, 18 percent of mornings have precipitation and the average precipitation on such mornings is 0.08 inches.

Figure C.26 plots the distributions of GP travel times, HOT travel times, and prices for each market on days with and without precipitation. On average, days with precipitation have more variable traffic conditions in most markets than days without. Averaging across markets, prices

and GP and HOT travel times are all about 15-18% higher on days when it rains, and the variance of travel times increases substantially (Table C.7).

Figure C.26 documents the effect of precipitation on departure time shares. The underlying data is at the (origin, 5-minute interval, date) level for southbound AM peak hours. The outcome is the share of trips on a given day—both GP and HOT—that enter the highway at a specific origin and 5-minute interval. We regress these departure time shares on indicators for the hour of day interacted with whether there is precipitation, including both origin and month fixed effects. Figure C.27a shows that there are fewer departures from 7-9 AM on days with precipitation, and more departures earlier in the morning, although the magnitudes are small.

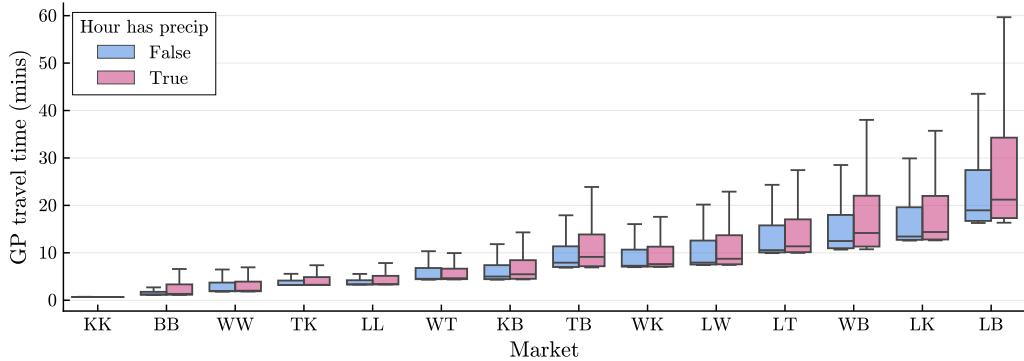
Table C.7: Travel times and prices with/without precipitation

Precip.	GP travel time		HOT travel time		Price	
	Mean	Std	Mean	Std	Mean	Std
False	14.63	6.89	11.19	4.23	3.32	3.40
True	16.77	8.75	12.90	6.19	3.90	3.03

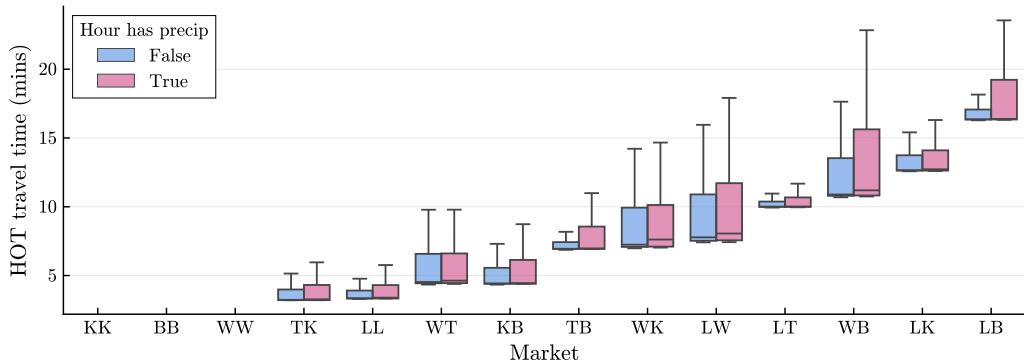
Note: This table documents the mean and standard deviation of travel times and prices with/without precipitation for the average southbound market during morning peak hours (5–11 AM). The mean and standard deviation are computed within market for days with and without precipitation, and then averaged across markets.

Figure C.26: Travel times and prices with and without precipitation

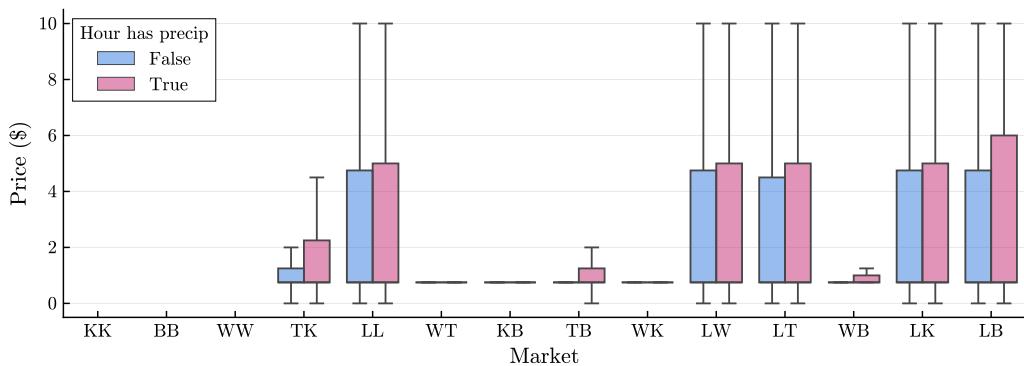
(a) GP travel time



(b) HOT travel time

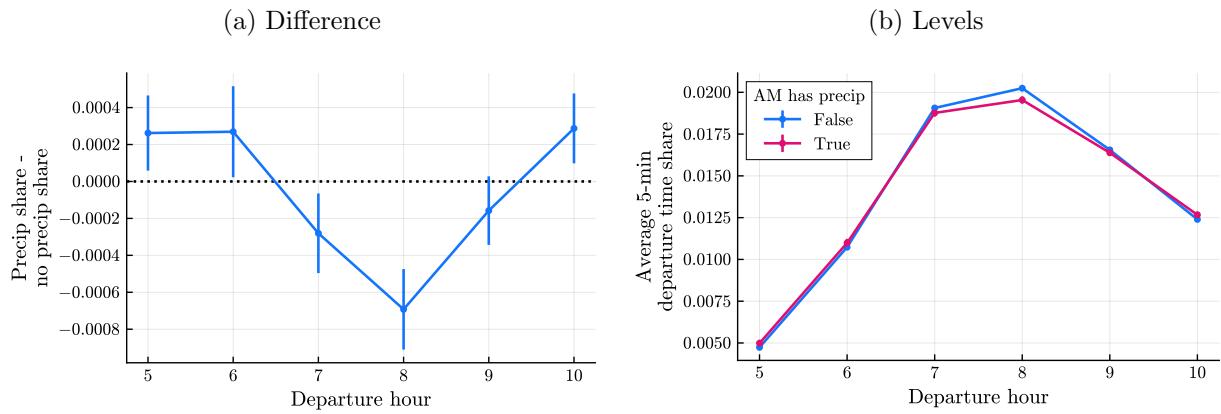


(c) Price



Note: These figures show the distributions of GP travel times (top panel), HOT travel times (middle panel) and prices (bottom panel) during the morning peak (5–11 AM) in each southbound market, comparing hours with zero precipitation to hours with positive precipitation. Boxes indicate the 25th percentile, median, and 75th percentile. Lower whiskers extend to the lowest observed data point that is within a distance of 1.5 times the interquartile range (IQR) from the 25th percentile. Likewise, upper whiskers extend to the highest observed data point within 1.5 times the IQR from the 75th percentile. Markets are ordered from shortest to longest. The first three markets have no feasible HOT route.

Figure C.27: Five-minute departure time shares with and without precipitation



Note: These figures plot coefficients from a regression of departure time shares on hour indicators and hour \times AM precipitation interactions, estimated including market origin fixed effects. Each observation is a (market origin, 5-minute interval, date).

D Model appendix

D.1 Utility microfoundation

The microfoundation for the second-stage utility specification (1) comes from Vickrey (1969), which posits that drivers have four different flow values of time. Figure D.28 illustrates these flow values, which depend both on where the driver is—at her origin, on the road, or at her destination—and on the time of day. Regardless of the time of day, driver i derives value u_i^{orig} per hour spent at her origin and u_i^{road} per hour spent on the road. However, her value per hour spent at her destination depends on whether that time is spent before or after ideal arrival time η_i : she gets $\underline{u}_i^{\text{dest}}$ per hour before η_i and \bar{u}_i^{dest} per hour after η_i .

The utility formulation in equation (1) follows from setting:

$$\begin{aligned}\alpha_i^D &= u_i^{\text{road}} - u_i^{\text{orig}} \\ \alpha_i^E &= \underline{u}_i^{\text{dest}} - u_i^{\text{orig}} \\ \alpha_i^L &= u_i^{\text{orig}} - \bar{u}_i^{\text{dest}}\end{aligned}$$

D.2 Speed, density, and throughput

In Section 5.2, we refer to the speed-density relationship as “the” road technology, but in fact, the road technology can be specified as a relationship between any pair of the three variables: speed, density, and throughput. This is because throughput, the number of cars crossing a point in space per unit of time, is identically the product of speed and density, which can be seen in the variables’ units of measure:

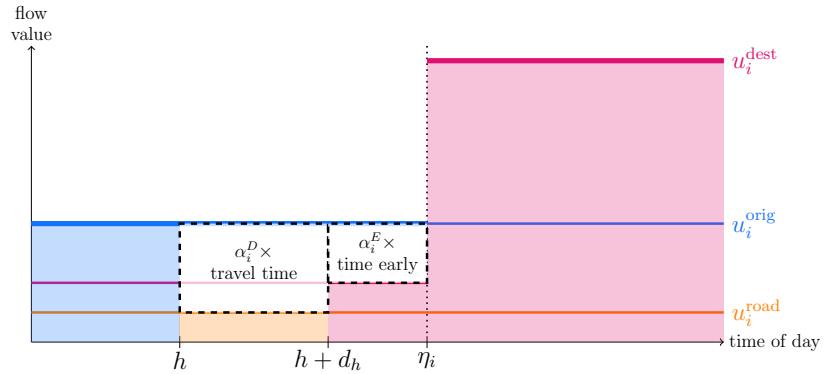
$$(\text{throughput in cars/lane/hour}) = (\text{speed in miles/hour}) \times (\text{density in cars/lane/mile}) \quad (10)$$

Thus, given a speed-density relationship, the relationships between the remaining pairs are determined by identity (10).

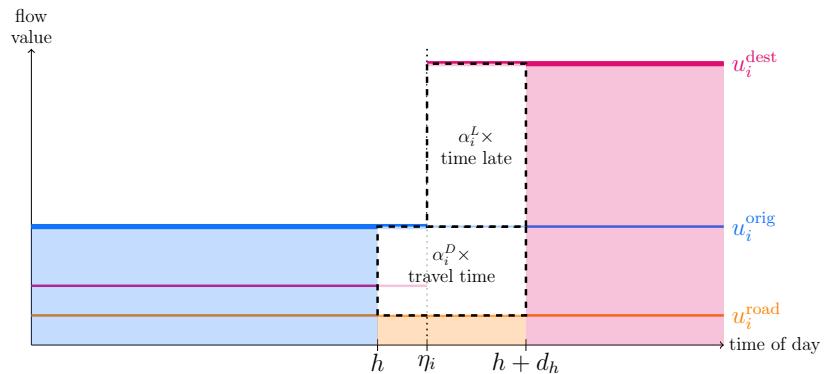
Figure D.29 illustrates two versions of the road technology which are common in the transportation literature. The speed-density curve (panel a) is monotone, with speed decreasing in density. The speed-throughput curve (panel b) is backward-bending. Start at the top left of the curve, where high speed (and low density, by panel a) is associated with low throughput. As speed decreases, throughput first increases due to increasing density. At some point, however, lower speeds reduce throughput, a phenomenon sometimes referred to as hypercongestion (Hall, 2018; Anderson and Davis, 2020). Both the speed-density and speed-throughput relationships are estimated in the empirical transportation literature, reviewed by Hall (2005).

Figure D.28: Utility microfoundation

(a) Case 1: depart early, arrive early



(b) Case 2: depart early, arrive late



(c) Case 3: depart late, arrive late

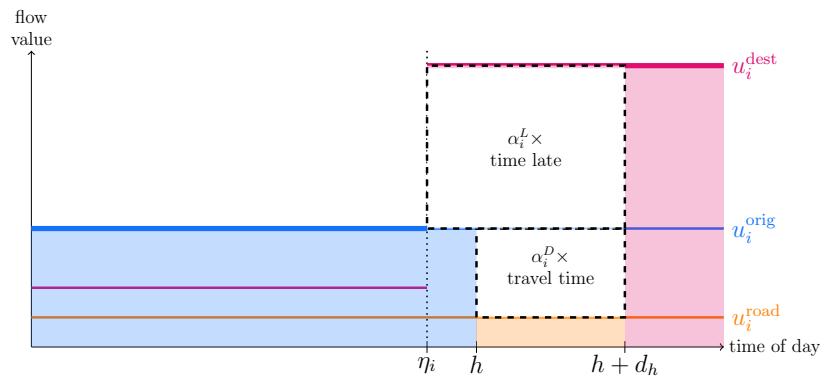
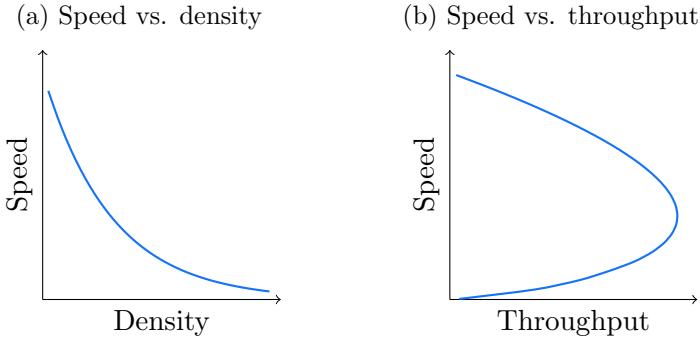


Figure D.29: Road technology relationships



E Estimation appendix

E.1 Estimating driver beliefs

In each market m , drivers have common beliefs about the joint distribution of prices and travel times on each day t . Let $(\mathbf{p}_{mt}, \mathbf{d}_{mt}) = (p_{jhmt}, d_{jhmt})_{j \in \{0,1\}, h \in \mathcal{H}}$ denote the $(3|\mathcal{H}|)$ -dimensional vector of prices and travel times in all routes j and all departure times h . This appendix describes how we parameterize and estimate the joint distribution $G_{mt}(\mathbf{p}_{mt}, \mathbf{d}_{mt})$.

We estimate a truncated joint normal distribution for each market m . The mean and variance-covariance matrix depend on the quarter (i.e., the season), day of week, and presence of absence of precipitation on day t . Estimation proceeds in three steps:

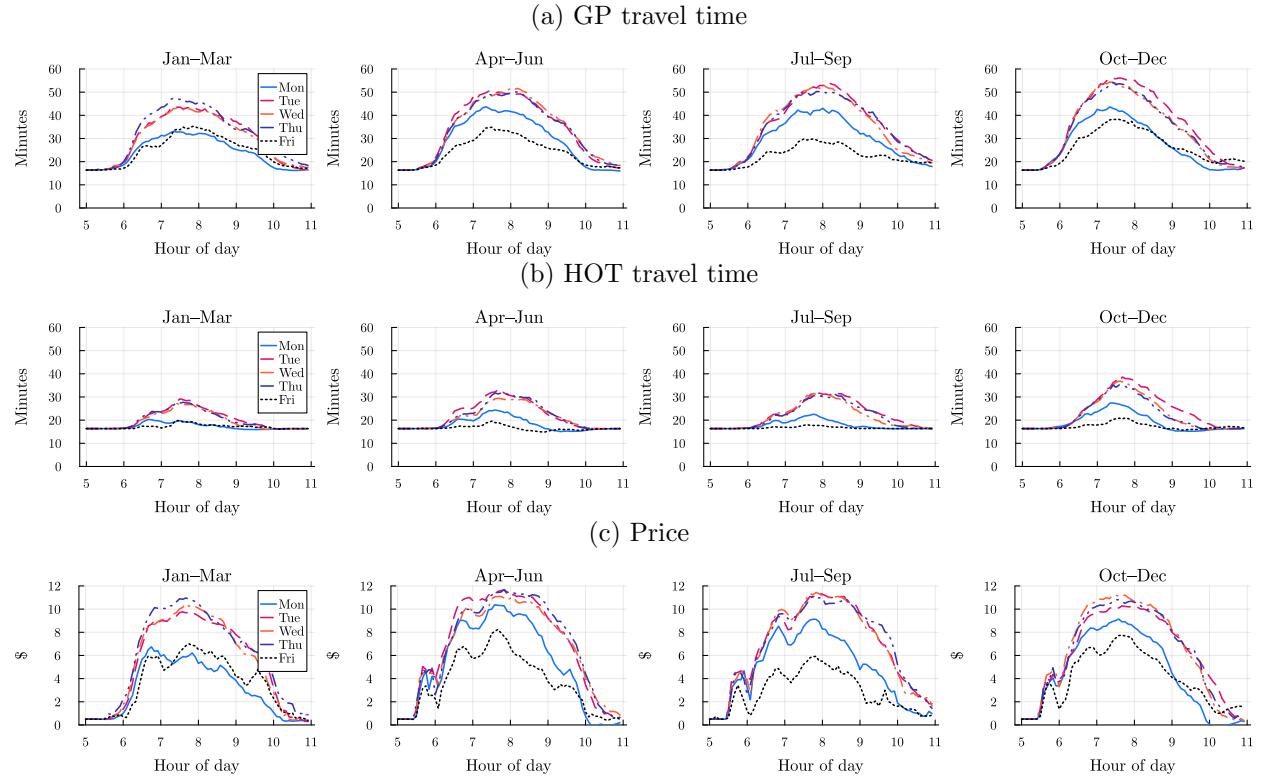
1. Obtain the mean vector by regressing

$$y_{jhmt} \sim \text{departureTime}_h \times \text{dayOfWeek}_t \times \text{quarter}_t + \text{departureTime}_h \times \text{precip}_t \quad (11)$$

On the left-hand side, y_{jhmt} is the *unrounded* price or the travel time. On the right-hand side, precip_t is an indicator for positive precipitation at the Everett weather station (about 11 miles north of I-405's northern terminus) between 5–11 AM on day t . Figure E.30 shows the estimated means on days without morning precipitation (i.e., the coefficients on the first right-hand side term) for the southbound full-length market.

2. Separately for days with and without precipitation, estimate the joint variance-covariance matrix for $(\mathbf{p}_{mt}, \mathbf{d}_{mt})$ using the sample variance-covariance of the residuals of equation (11). Each variance-covariance matrix has dimension $3|\mathcal{H}| \times 3|\mathcal{H}|$; Figure E.31 plots the $|\mathcal{H}| \times |\mathcal{H}|$ -dimensional submatrices corresponding to the variance-covariance matrices of GP travel times, HOT travel times, and prices for the southbound full-length market.
3. Truncate prices at the \$0.75 floor and \$10 ceiling. Truncate travel times from below at the minimum observed travel time in that market (across all departure times and dates).

Figure E.30: Driver beliefs: mean travel times and prices without precipitation



Note: Figure plots estimated coefficients on the (departure time, day of week, quarter) interactions in equation (11) for the southbound full-length market.

E.2 Demand moment conditions

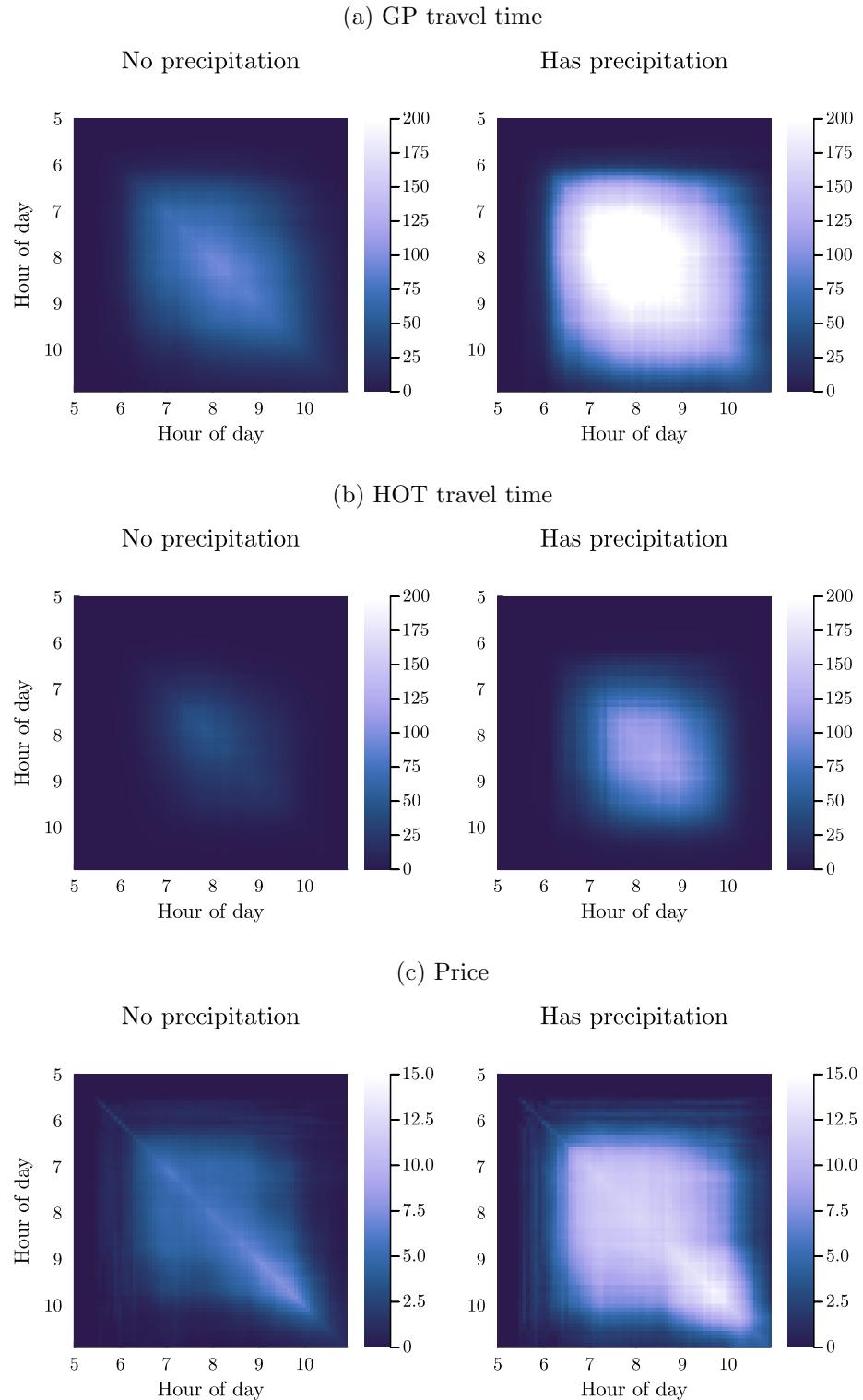
In this appendix, we describe the full set of moment conditions used to jointly estimate the two stages of the demand model. Each set of moment conditions is presented under the heading of the demand parameters about which they are most informative. Throughout this appendix, the subscripts underneath an expectation symbol indicate the dimensions over which the expectation is being taken.

E.2.1 Stage 1 moment conditions

Let $\hat{\xi}_{hat}(\theta)$ denote the demand shocks in market origin a at time h on date t that rationalize the observed departure time shares conditional on the candidate parameters θ . Recall that we can only recover demand shocks at the *market origin*—not market—level because we only observe departure time quantities at the market origin level, from highway on-ramp throughputs.

Furthermore, let $\tilde{\xi}_{hat}(\theta)$ denote the same demand shock after residualizing with respect to date t fixed effects. These date-demeaned demand shocks affect only substitution across departure times, not substitution toward the non-405 outside option.

Figure E.31: Driver beliefs: variance-covariance matrices of travel times and prices



Note: Figures show estimated variance-covariance matrices of the residuals of GP travel times (panel a), HOT travel times (panel b), and prices (panel c) from equation (11) for the southbound full-length market.

SOV inside good intercepts The demand shocks have mean zero in each market origin a :

$$\forall a : \quad \mathbb{E}_{ht} \left[\widehat{\xi}_{hat}(\theta) \right] = 0$$

These moment conditions are informative about non-carpoolers' *origin-level* inside good intercepts $\beta_{SOV,a}^0$, which control substitution to the non-405 outside option.

HOV inside good intercepts We match mean HOV quantities in each market. Let $q_{HOV,1hmt}$ and $\widehat{q}_{HOV,1hmt}(\theta)$ denote the observed and model-predicted HOV quantities, respectively, in departure time h in market m on date t . These moment conditions take the form:

$$\forall m : \quad \mathbb{E}_{ht} \left[\frac{q_{HOV,1hmt} - \widehat{q}_{HOV,1hmt}(\theta)}{q_{HOV,1hmt}} \right] = 0$$

They are informative about carpoolers' *market-level* inside good intercepts $\beta_{HOV,m}^0$.

Ideal arrival time distribution After demeaning by date, the demand shocks have mean zero in each hour H :

$$\forall H : \quad \mathbb{E}_{hmt} \left[\widetilde{\xi}_{hmt}(\theta) \mid \text{hour}(h) = H \right] = 0$$

These moment conditions are informative about the standard deviation σ^η of ideal arrival times. They impose that systematic differences in departure time shares across hours must be due to the distribution of drivers' ideal arrival times, rather than rationalized solely by demand shocks.

Time early and time late coefficients After demeaning by date, the demand shocks are independent of morning precipitation indicators in each hour H :

$$\forall H : \quad \mathbb{E}_{hmt} \left[\widetilde{\xi}_{hmt}(\theta) \times \text{precip}_t \mid \text{hour}(h) = H \right] = 0$$

As discussed in Section 6.1.2, precipitation increases the *variance* of prices and travel times. If drivers find it very costly to be late to their destinations, they will "buy more insurance" on rainy days by shifting their departure times earlier in the day. These moment conditions impose that there are no systematic differences in unobserved demand for different *departure times* on rainy versus sunny days. However, since we use the date-demeaned demand shocks here, we allow for the possibility that unobserved demand for the *outside option* is systematically different on rainy days.

E.2.2 Stage 2 moment conditions

In the second stage, we observe HOT ($j = 1$) but not GP ($j = 0$) route quantities. Since we can't construct route market shares, we instead match moments of HOT route *quantities*.

HOT intercepts We match mean paid HOT quantities in each market. Let $q_{SOV,1hmt}$ and $\hat{q}_{SOV,1hmt}(\theta)$ denote the observed and model-predicted paid HOT quantities, respectively, in departure time h in market m on date t . These moment conditions take the form:

$$\forall m : \quad \mathbb{E}_{ht} \left[\frac{q_{SOV,1hmt} - \hat{q}_{SOV,1hmt}(\theta)}{q_{SOV,1hmt}} \right] = 0$$

They are informative about the market-specific HOT intercepts α_{1m}^0 .

Mean price and travel time coefficients We match reduced-form coefficients of log paid HOT quantity on the second-stage price and travel time shifters from Section 6.1.2. First, we match the coefficient on the rounded-up indicator in the price rounding regression discontinuity. Second, we match the coefficient on the interaction of the treated and post-treatment indicators in the crash differences-in-differences regression. Appendices C.1 and C.2 describe the two regression specifications in more detail. For each coefficient φ estimated from the data, let $\hat{\varphi}(\theta)$ denote the analogous coefficient estimated from the model-predicted HOT quantities. These moment conditions impose that $[\varphi - \hat{\varphi}(\theta)] / \varphi = 0$ for each φ . They are informative about the mean price coefficient $\bar{\alpha}^P$ and the mean travel time coefficient $\bar{\alpha}^D$.

Heterogeneity by income We match covariances of HOT driver characteristics \mathbf{x}_{im} with attributes of HOT trips taken (price paid and time saved) conditional on the hour and market:

$$\begin{aligned} \text{Cov}_{ihmt} (\mathbf{x}_{im}, & p_{1hmt} \mid q_{i1hmt} > 0, \text{hour}(h), m) \\ \text{Cov}_{ihmt} (\mathbf{x}_{im}, d_{0hmt} - d_{1hmt} \mid q_{i1hmt} > 0, \text{hour}(h), m) \end{aligned}$$

The data covariances are computed from the toll transaction data, using drivers whom we observe choosing the HOT route at time h on day t . To compute the model covariances, we weight by drivers' predicted probabilities of choosing the HOT route. Matching these moments is informative about the parameters $\mu^{\alpha,P}$ and $\mu^{\alpha,D}$ specifying how preferences vary with driver characteristics. As with the rounding and crash moments, we impose that $[\varphi - \hat{\varphi}(\theta)] / \varphi = 0$ for each covariance φ .

Unobserved heterogeneity We match variances and covariances of the same attributes of HOT trips taken (price paid and time saved) conditional on the hour and market:

$$\begin{aligned} \text{Var}_{ihmt} (& p_{1hmt} \mid q_{i1hmt} > 0, \text{hour}(h), m) \\ \text{Var}_{ihmt} (d_{0hmt} - d_{1hmt} \mid q_{i1hmt} > 0, \text{hour}(h), m) \\ \text{Cov}_{ihmt} (p_{1hmt}, & d_{0hmt} - d_{1hmt} \mid q_{i1hmt} > 0, \text{hour}(h), m) \end{aligned}$$

Matching these moments is informative about the variance-covariance matrix $\Sigma^{\alpha,PD}$ of the unobserved component of drivers' preferences for price and travel time. Again, we impose that $[\varphi - \hat{\varphi}(\theta)] / \varphi = 0$ for each statistic φ .