NBER WORKING PAPER SERIES

THE VALUE OF URGENCY:
EVIDENCE FROM REAL-TIME CONGESTION PRICING

Antonio Bento
Kevin Roth
Andrew R. Waxman

Working Paper 26956
http://www.nber.org/papers/w26956

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2020, Revised February 2024

Previously circulated as “Avoiding Traffic Congestion Externalities? The Value of Urgency.” The authors thank Lawrence Goulder, Edson Severnini, Lowell Taylor, and Spencer Banzhaf for helpful comments and seminar participants at NOVA SBE, Arizona State University, Carnegie Mellon University, Cornell University, John Hopkins University, Simon Fraser University, Stanford University, University of California at Davis, University of California at Irvine, University of Maryland, Penn State University, University of Texas at Austin, ZEW, the ASSA/AEA Meetings, the Annual Meetings of the Association of Environmental and Resource Economists, the World Congress of Environmental and Resource Economists, the Kraks Fond Managing Congestion Seminar, and NBER Summer Institute. Finally, we thank Stephanie Wiggins and Kathleen McCune at LA Metro for providing access to ExpressLanes transaction data and Arthur T. Degaetano for supplying us with weather data. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Value of Urgency: Evidence from Real-Time Congestion Pricing
Antonio Bento, Kevin Roth, and Andrew R. Waxman
NBER Working Paper No. 26956
April 2020, Revised February 2024
JEL No. D47,D61,D62,H23,L51,L91,Q58,R41

ABSTRACT

Taking advantage of repeated appearances of drivers in Los Angeles's ExpressLanes under varied traffic conditions, we uncover the distribution of individuals' preferences for time savings in a novel application of a hedonic pricing model. In this setting, we introduce the concept of the value of urgency, defined by a component of willingness to pay to enter tolled lanes to avoid a congested alternative route. The value of urgency does not scale in the amount of time saved, reflecting discrete penalties for late arrival. We show that this value accounts for 87% of total willingness to pay to use the ExpressLanes, while the contributions to WTP from other widely used valuation measures, such as the values of time and reliability, are negligible. We suggest that quality-of-service pricing that varies in real time and removes uncertainty over travel times creates new markets for individuals to avoid lateness and reveal their preferences for urgency.

Antonio Bento
Sol Price School of Public Policy
and Department of Economics
University of Southern California
Los Angeles, CA 90089-0626
and NBER
abento@usc.edu

Kevin Roth
Laurits R. Christensen Associates
800 University Bay Dr., Suite 400
Madison, WI 53705-2299
kdroth@lrca.com

Andrew R. Waxman
University of Texas at Austin
awaxman@utexas.edu

A data appendix is available at http://www.nber.org/data-appendix/w26956

I. Introduction

Building on the classical theory of the allocation of time (Becker 1965; Johnson 1966; DeSerpa 1971; Becker 1993), a fundamental insight of microeconomics is that individuals' value of time (VOT) is a fraction of their hourly wage, reflecting the opportunity cost of time.[1] Studies that recover estimates of the VOT consider the time and money trade-offs that individuals face and implicitly assume that they value time on a minute-by-minute basis.[2] A related literature considers scheduling costs of arriving early or late relative to an ideal arrival time, which too scale up on a minute-by-minute basis (Vickrey 1969; Small 1982; Arnott, Palma, and Lindsey 1993; R. Noland and Small 1995; Hall 2018a). These two approaches for valuing time, however, are at odds with the idea that individuals are often willing to pay discrete amounts of money to meet critical schedule constraints, reflecting their preferences for urgency.[3]

Recently, in California, a new system of so-called ExpressLanes has been implemented that allows drivers to enter a freeway lane upon paying a toll that varies almost in real time, thereby securing a guaranteed minimum travel speed and removing uncertainly in travel times. As in any market where price adjusts to equate supply and demand, ExpressLanes tolls match changes in travel demand to the fixed capacity of the lanes. In such a setting, drivers have the ability to purchase ExpressLanes segments that map directly to a discrete amount of time savings needed to meet schedule constraints and reveal their willingness to pay to avoid discrete penalties for late arrival.[4] As a result, when observed on a minute-by-minute basis, ExpressLanes users' average willingness to pay (WTP) to save time, reflected by the toll paid, appears to be absurdly high, suggesting that discrete WTP to meet schedule constraints is of critical economic importance.[5] In this paper, we take advantage of a uniquely rich dataset of drivers who use ExpressLanes in California to conceptualize and provide first estimates of the *value of urgency*. We define the value of urgency as a discrete component of the WTP to save time that does not scale up with the amount

[1] Becker (1965) pioneered household production models in which individuals combine time with a market good to produce commodities. This basic framework served as a basis of a new field of home economics, with the basic model applied to issues ranging from the time spacing of childbirths to the division of labor among household members. For a review, see Greenwood et al. (2017).
[2] The VOT is also a fundamental parameter for benefit–cost analysis and project evaluation, including in areas such as the evaluation of road infrastructure and pricing (Yang, Purevjav, and Li 2020), recreational demand (Smith 1981), health improvements (Grossman 1972), and a variety of government regulations, such as seatbelt use provisions and speed limits (van Benthem 2015).
[3] Examples include premium payments for faster processing of visas and passports to assure the ability to travel on a certain date and the willingness to pay for an organ that would save one's life, especially the closer the individual is to life-threatening organ failure.
[4] We use the expression "time savings" to describe time reallocation that has positive WTP, as is common in the transportation literature and common parlance despite the fact that, in principle, time can only be reallocated, not saved.
[5] Importantly, we are able to show that there is no evidence that these results reflect heterogeneity in the valuation of time, lack of experience with ExpressLanes, deviations from complete rationality for rather small time savings, or drivers' misperceptions of time savings.

of time saved to meet critical schedule constraints. Our fundamental insight is that the value of urgency reflects the bulk of the welfare benefits from access to ExpressLanes: namely, 87% in our main specifications. With an average time-varying toll of $3.73 per trip and an average travel time savings of 4.30 minutes during the morning peak, our central estimate of the value of urgency is $3.24 per trip. Consistent with the literature (Brownstone and Small 2005; Small, Winston, and Yan 2005; Small 2012), we also recover estimates of the value of time and reliability of $8.19 and $17.61 per hour, respectively.[6] Overall our results have important implications for infrastructure policy. An ex-ante benefit–cost calculation that considers only these two components (value of time and reliability), as is typically done in evaluations of infrastructure projects, would fail to capture the bulk of the benefits from ExpressLanes. Even with a value of time two or three times higher than the standard, the resulting benefit estimate would still be off by more than 100%.[7]

We assemble a uniquely rich dataset on ExpressLanes users, including information on time of use, points of entry and exit, and toll charged. The total toll charged to a driver is the sum of tolls for sub-segments of the corridor, which is locked in and displayed to drivers at entry, eliminating any uncertainty about the total toll paid. These data are supplemented with data on real-time speeds from California's Freeway Performance Measurement System (PeMS), permitting counterfactual calculations of the differences in travel time and reliability between ExpressLanes and mainline lanes in freeways.

We estimate a hedonic price regression of a trip's total toll paid for a combination of segments on the trip's attributes, which include time savings and reliability difference between the ExpressLanes and the mainline lanes, and a constant corresponding to the value of urgency.[8] An appeal of the hedonic model is its simplicity: the derivative of a hedonic price function with respect to an attribute, the implicit price function, provides a consumer's MWTP for that attribute (Rosen 1974). We show that the choice of how many segments of the ExpressLanes to use reveals valuation of continuous attributes and that for those who use the ExpressLanes, the lane choice

[6] In most empirical settings, the analyst is unable to calculate the scheduling costs of early or late arrival because of the lack of detailed individual-level information on desired arrival times. To overcome this challenge, researchers have developed measures of reliability that capture uncertainty through variability in travel times. For a review of the value of reliability, see Fosegerau and Karlström (2010).

[7] When evaluating infrastructure projects, the US government uses VOT estimates of between 33% and 50% of the hourly wage rate. This value captures all major categories of trips, including personal, leisure and work trips (US DOT 2015). The average hourly wage in Los Angeles is $20, suggesting a VOT of at most $10 per hour.

[8] In this setting, the driver has full knowledge of the price, but obviously, like in any other hedonic setting (e..g, cancer clusters (Davis 2004), crime (Bishop and Murphy 2011; Linden and Rockoff 2008), or air quality (Chay and Greenstone 2005)), there could be uncertainty in the attribute of interest: here travel time savings. While uncertainty in attributes does not invalidate the hedonic approach it can, however, result in attenuation bias via measurement error. As is standard in the literature, we use instrumental variables to address this concern.

decision reveals an additional attribute that does not scale in time savings: the value of urgency. In our empirical setting, we observe individuals making repeated choices of different combinations of ExpressLanes segments that map into different levels of the attributes: time savings and reliability difference. As a result, it is possible to recover demand curves even without instrumental variables simply because we observe several points of the individual WTP function for an attribute, as shown more broadly in the hedonic literature (Bajari and Benkard 2005; Kuminoff and Pope 2014; Bishop and Timmins 2018).[9] The logic of this approach is that exogenous shocks to demand, including unobserved ones, produce a new hedonic equilibrium that allows us approximate demand by tracing out the optimized choices of households.[10]

We leverage the richness of our repeat-transaction data to better characterize the empirical preference heterogeneity of ExpressLanes drivers through hedonic regressions. Our central results come from highly parameterized hedonic price regressions derived from the segment choice decision of drivers purchasing time linked directly to classical models of scheduling and time allocation (Becker 1965; Small 1982; Vickrey 1969; Arnott, de Palma, and Lindsey 1993). We regress the total toll paid by commuters on the three key attributes of their valuation of time that reflect drivers' choice to enter ExpressLanes and purchase ExpressLanes segments: the travel time savings, measured in reference to the mainline lanes; the reliability differences between ExpressLanes and mainline lanes; and a constant that reflects the component of the WTP for time savings that does not scale up with the amount of time saved, reflecting discrete penalties for late arrival. In the spirit of Bishop and Timmins (2018) and Banzhaf (2020), we estimate individual-specific commuting preferences from hedonic regressions by taking advantage of the repeated appearances of drivers in the ExpressLanes under varied traffic conditions. These coefficients have a specific economic interpretation in a hedonic setting that corresponds to individual willingness-to-pay functions.[11] These estimates allow us to recover the full distribution of the underlying preference parameters for the values of urgency, time and reliability and demonstrate that a large source of their variation is individual-specific rather than corresponding to a particular time of day.

[9] In such contexts, Banzhaf (2020) shows that connecting different points of an individual's full willingness-to-pay function provides a good approximation of a true Hicksian welfare measure, regardless of the shape of the underlying demand curve.

[10] Prior concerns about the hedonic in the context of cross-sectional data required the use of instruments to recover welfare-relevant policy impacts (Bartik 1987), but have been subsequently resolved (Ekeland, Heckman, and Nesheim 2004; Heckman, Matzkin, and Nesheim 2010). In our setting, we observe repeat transactions allowing us to recover an approximation to the underlying demand curve and the empirical exercise is much simpler: recovering the valuation of the components of willingness-to-pay.

[11] These correspond to hedonic bid-rent functions, which have long been used in urban economics to value housing demand for amenities (Wheaton 1977). These prices also have a welfare interpretation in the context of corresponding differentiated product markets, even in the presence of endogenous supply (Rosen 1974).

One important caveat to our approach is that, like all such studies, we can observe transactions only for drivers who use the ExpressLanes. If the composition of drivers entering the ExpressLanes on a given morning changes with the distribution of travel demand, the set of tolls we observe may not be indicative of the average willingness to pay to save time commuting. This issue is present in all empirical work on local amenities and housing values as well as the literature on compensating differentials in the labor market (Bishop et al. 2019; Evans and Taylor 2020; Redfearn 2009). Unfortunately, without data on commuter demographics, we cannot examine this issue in our context. Further, we adopt several strategies to address potential identification concerns. A first concern common to hedonic regression is that attributes are endogenous. As an example, estimating the benefit of clean air from housing prices may be confounded if unobserved economic conditions are positively correlated with housing prices and pollution. In our setting, unobserved shocks to travel demand along the corridor, say from the closure of a transit line or an accident elsewhere in the transportation system, could congest the mainline lanes and result in greater time savings for ExpressLanes use. If these time savings increase demand for the ExpressLanes, the toll responds directly to the induced increase in lane occupancy. In other words, unobserved demand shocks create identifying variation in the covariates of the regression rather than acting as confounding factors, obviating the need for instruments to address them as omitted variables.

A second concern is either mismeasurement of the perceived time savings by the econometrician or misperception of the actual time savings by drivers. We address the first issue by applying an instrumental variable approach that follows the labor economics literature (Aizer et al. 2018) to use future realizations of time savings as instruments. To address potential misperception of time savings, we take advantage of a natural experiment: the installation in the middle of our sample period of lane entry-point signs that display in real time the predicted travel time savings for ExpressLanes users. We find that this information shock actually increased the value of urgency in our study, and we perform several other robustness tests to rule out the possibility that driver mistakes could explain our results.

A third identification challenge is unobservables, and we also show that our results are robust to inclusion of a large set of time, individual and lane-specific fixed effects. Last, we use a difference-in-difference–style strategy to separately identify lane-specific unobservable characteristics that may be present for more urgent trips by examining accounts with trips appearing during both the weekday morning peak and the weekend and using these accounts as a control group.

This study makes three main contributions to the literature. First, we provide the first credible estimates of the value of urgency and demonstrate its economic importance. In doing so, we follow a growing recent literature that relies on purely *revealed preference* settings leveraging big data on mobility where the *actual* behavior of individuals can easily be captured (Wichman and Cunningham 2023; Ang, Christensen, and Vieira 2020; Christensen and Osman 2020). Most earlier studies that recover drivers' preferences for travel time savings rely on stated preference surveys and recover estimates only of the value of time and reliability (Brownstone and Small 2005; Train and Wilson 2008).[12] Even if a survey could recover an individuals' likelihood of being late and the resulting value of urgency, one would be concerned about whether respondents in stated preference surveys can precisely reveal their valuations for rather small travel time savings like those observed for ExpressLanes users.[13] Part of this challenge is that respondents in stated preferences surveys may take a long-run view on the re-allocation of time, reflecting their underlying valuation of time savings. As shown by Peer, et al. (2015), there is reason to believe that the exigencies of individuals from time constraints on any given day may lead them to make quite different short-run decisions than even revealed preference choices for longer-run time allocations allow. Along with that paper, our study points to the importance of accounting for time constraints specific to exact times of day. This builds on a related literature in recreational demand (Smith, Desvousges, and McGivney 1983; Palmquist, Phaneuf, and Smith 2010; McConnell, Siikamäki, and Smith 2017) that demonstrates the limitations that individuals may have to reallocate time smoothly from one use (e.g., work or home production) to another (e.g., recreation), but does not directly measure the penalties from deviation of idealized activity start times. Admittedly, in our context, we are limited in the ability to describe empirically the potentially complicated, idiosyncratic and individual-specific nature of how time allocation results in particular penalties, but the nature of our repeat-transaction data allows us to provide overwhelming evidence that a major component of their cost in our setting can be represented by

[12] An exception is Small, Winston and Yan (2005), who combine stated with revealed preference data. However, this study covers a limited time period, consisting of just a few months, from which it is more difficult to estimates of urgency.

[13] An added dimension with respect to this issue is that stated preference surveys typically solicit drivers' WTP for travel time savings from choosing tolled over untolled lanes on an "average day." The value of urgency may not emerge on a hypothetical "average" day or "average" trip but rather depend on the purpose of a specific trip and the penalty associated with failing to meet specific schedule constraints. More broadly, Calfee, Winston and Stempski (2001) warn about the limitations of recovering value of time estimates from stated preference data, particularly using overly restrictive models such as ordered probit, where the recovered values of time may be implausibly small. A broader class of studies in environmental economics considers the limitations and challenges of stated preferences approaches, including survey designs (Kling, Phaneuf, and Zhao 2012; Johnston et al. 2017; Carson and Hanemann 2005).

a constant in a hedonic price regression. Moreover, we are able to document the extent to which the values of urgency, time, and reliability evolve over the morning peak in a manner consistent with schedule constraints for individuals on any given day associated with common work or appointment start times (e.g., 7:30 or 8:00 AM). While these estimates yield central values of urgency, time and reliability, there is no reason to believe that their values are constant, and we document substantial heterogeneity by individual, day and time during the AM peak.

Despite documenting the predominance of the value of urgency in explaining WTP for time savings, we nonetheless recover point estimates of the value of time and reliability that are aligned with those from these past studies and may help to better clarify the variation in the value of time recovered in recent work. Goldszmidt et al. (2020) conduct two large-scale natural field experiments with the ridesharing company Lyft and utilize random variation in both wait times and prices to recover estimates of the value of time of approximately $19 per hour and typically 75% or more of the hourly wage rate. The novelty of the study comes from the ability to estimate how the waiting time elasticity and the VOT vary over the length of the wait time. While the design of Goldszmidt et al. (2020) is not equipped to recover WTP that does not scale with time saved, it seems plausible that passengers' sense of urgency while they await a ride may help to explain the high values of time that they recover.

Second, we present a novel application of hedonic price regressions to recover preferences for product attributes in markets where these attributes scale up with quantity choices—here segments of the ExpressLanes. To establish the basis for this approach, we build on the three canonical approaches to recover preferences for time savings in the transportation literature and then demonstrate how by focusing on the intensive, rather than extensive, margin of choice estimates can be recovered in a new manner. Our focus on segment choice exploits the fact that a large proportion of drivers in ExpressLanes during peak periods are likely to be running late and so allows us to recover preferences that may not be as important in past work that has instead modeled lane choice (Small, Winston, and Yan 2005; Brownstone and Small 2005; Brownstone et al. 2003). Moreover, the use of segment choice with repeat-transaction data—observing the same consumers many times under different toll and congestion levels–motivates the use of a hedonic approach in lieu of discrete choice. We show that a hedonic approach still allows one to recover a distribution of preferences, but without the heavy data requirements needed to link these preferences back to individual characteristics or distributional assumptions on unobserved demand parameters. In addition, this methodology allows one, given the specific nature of our data and design of the ExpressLanes, to recover lower bounds on WTP in a manner that collapses a discrete-continuous estimation approach (lane choice and segment quantity) into a single regression equation. This approach trades the rich complexity of a structural model with the appeal of a simple reduced-form model applied to "big data" that can still speak to welfare effects by providing a "sufficient statistic" for their measurement in the spirit of Chetty (2009) and Banzhaf (2020).

Third, we shed light on the importance of quality-of-service pricing as a mechanism to create markets where individuals can reveal their value of urgency. This is not to say that the value of time and reliability are not essential for the evaluation of infrastructure projects. However, a novelty of the Los Angeles ExpressLanes is that their tolls vary almost in real time, are set to assure that speeds never fall below a minimum threshold and remove most of the travel time uncertainty. Thus, they allow drivers to assess whether the time savings from entering the ExpressLanes are sufficient to recover lost travel time and still meet a schedule constraint. While the recent literature on real-time pricing has highlighted its role as a strategy for shifting individuals away from peaks (Ito 2014; Ito, Ida, and Tanaka 2018), this form of pricing is fundamentally different from the quality-of-service type studied here. An insight of our description of the program is that quality-of-service pricing eliminates uncertainty regarding aggregate congestion, permitting individuals to precisely reoptimize their decisions when they are faced with schedule constraints. Given this behavior, we demonstrate that a fixed component of WTP that does not scale with the distance traveled in the ExpressLanes corresponds to the discrete value of on-time arrival. As a result, we can separately recover the valuation of discrete lateness penalties from that of time-varying lateness penalties, which correspond to reliability differences between ExpressLanes and mainline lanes in our setting.[14]

The rest of the paper is organized as follows: Section II presents information about the ExpressLanes program. Section III presents a conceptual framework to rationalize drivers' ExpressLanes choice, describes the data, and provides initial suggestive evidence of the relation of urgency with schedule constraints. Section IV outlines our empirical strategy, section V presents

[14] In this sense, our study is also closely related to Kreindler (2022), who uses a model of departure time choice to design a field experiment with congestion pricing in Bangalore, India. Like us, Kreindler (2022) relies on a revealed preference approach to study WTP to save travel time. In that study, tolled routes have lower congestion, which may explain the relatively low estimated costs of early and late arrival that are suggestive of schedule flexibility. However, in settings with higher levels of congestion and significant uncertainty in expected arrival times, even with flexibility in departure times, the quality-of-service pricing offered by ExpressLanes can provide another avenue for commuters to protect themselves from high peak-hour congestion.

the results and robustness checks, and section VI offers some concluding thoughts.

II. Program Background

On February 23rd, 2013, Los Angeles converted a 10.5 mile section of the high-occupancy vehicle (HOV) lanes on the I-10 into a high-occupancy toll (HOT) facility, as part of the ExpressLanes program.[15] This portion of the I-10 connects the suburb of El Monte, a dense residential area on the east side of the metropolitan area, to the major employment center in downtown LA. ExpressLanes users have to acquire a transponder, available at major supermarkets, public transit stations and by mail, which requires an initial balance payment of approximately $40.[16]

The ExpressLanes are divided into five segments, corresponding to distinct entry and exit points as shown in Appendix Figure F.1. Drivers using the I-10 mainline lanes see electronic signs at the entrance of each segment displaying the toll that they will pay for use of the ExpressLanes up to the final downtown exit and a midpoint exit as shown in Appendix Figure F.2. Given this information, a driver using the ExpressLanes decides the quantity of segments to purchase access to. The resulting total toll paid is the sum of the segment tolls. After October 20th, 2013, signs were set up at each ExpressLanes segment entrance indicating the exact travel time savings that drivers could expect. We leverage this feature in a natural experiment below and test whether this additional information about travel time savings altered driver behavior.

ExpressLanes Toll Algorithm.—As in any market where price adjusts to equate supply and demand, ExpressLanes tolls match changes in travel demand to the fixed capacity of the lanes. The tolls adjust to provide quality-of-service pricing, which is common in other regulated markets (De Vany 1976; Saving and De Vany 1977; Chao and Wilson 1987). Two features of the ExpressLanes are important for our analysis: first, the toll ensures that speeds usually stay above 62 MPH even when mainline lane speeds are much lower, as demonstrated by columns III–IV of Table 1.[17] Second, if toll increases are unable to rein in demand sufficiently and average five-minute speeds fall below 45 MPH, access to the lane closes to single-occupancy vehicles.

The tolls change every five minutes independently across the five segments as follows:

[15] This was the second of such conversions in Los Angeles. The first conversion occurred on the I-110 ExpressLanes, which opened on November 10th, 2012. An advantage of our focusing on the I-10 corridor is that drivers may have had more time to learn about the program because of this other corridor.
[16] Like solo drivers, carpools are required to have a transponder but are not charged when they have at least three passengers during peak times (5–9 AM and 4–8 PM) and at least two passengers during off-peak hours.
[17] ExpressLanes speeds do occasionally dip below 62 MPH in our sample: 12% of average ExpressLanes speeds over five-minute intervals are below 62 MPH.

$$(1) \qquad toll_{s,t} = toll_{s,t-1} + \kappa(o_{s,t} - o_{s,t-1}).$$

Here, $toll_{s,t}$ and $toll_{s,t-1}$ denote the toll for segment s at time t and in the preceding five-minute interval, respectively. $o_{s,t}$ represents the occupancy rate in cars per mile of segment s during time interval t, which reflects the demand for segment s. The parameter κ is always positive and is monotonically increasing with changes in occupancy, making the toll more responsive as capacity in the lane fills, which secures a minimum quality of service in the lane. Toll adjustment, mediated through κ in equation (1), happens smoothly so that there are no exogenous changes in the toll; rather, they respond in direct relation to changes in occupancy in the ExpressLanes and therefore to travel demand. As a result, ExpressLanes operate like any other market, with the market price always increasing in response to increases in demand and falling with decreases.

Calculation of Total Toll Paid by ExpressLanes Drivers.—Single-occupant vehicles (SOVs) are charged a per-mile toll ranging from \$0.10 to \$15.00, which varies in response to demand for the lane. The total toll paid, $Toll_{s,t}$, is the sum of the segment-level tolls, $toll_{s,t}$, corresponding to segments driven by a vehicle

$$Toll_{s,t} = \sum_{s \in S} toll_{s,t}.$$

Drivers approaching the ExpressLanes entrance see relevant total tolls that they would pay for a full trip along the ExpressLanes and a partial trip, exiting midway. This is because the bulk of trips occur between these points as shown in Appendix Table G.1. Once a vehicle enters the lane, the set of segment tolls for whatever segments it uses is locked in for all segment tolls based on the displayed prices for the duration of the trip.[18] Therefore, from a driver's perspective, there is no ex-ante uncertainty about the total toll paid for the trip. Once the maximum price is reached, the lane is closed to further SOV traffic until lane occupancy decreases sufficiently; this did not occur in our sample period but did occur on the other ExpressLanes corridor (I-110) during the same period.

Effect of Travel Demand Shocks on ExpressLanes.—Figure 1 illustrates the functioning of the market for travel demand, the interactions between mainline lanes and ExpressLanes, and the

[18] If a driver exits the lanes before the midpoint, then the expected toll is bounded from above by the midpoint toll, and the same is the case if a driver leaves the lanes before the final exit.

effect of mainline demand on the ExpressLanes toll. It simplifies the design of the ExpressLanes, by focusing on the price of a single segment's toll, ignoring, for the moment, segment-level variation in demand shocks and toll responses. Of course, when a demand shock increases a segment toll, it also increases the total toll for all trips that include that segment. During peak periods, demand in the mainline lanes rises as depicted in Panel A of Figure 1. This can occur because of regular cyclical travel demand such as common work start times or because of idiosyncratic shocks such as public transportation shutdowns, unexpected construction or accidents on parallel roads, all of which increase demand for the corridor and therefore also the potential for congestion in the transportation system. This demand shift induces an increase in occupancy in the mainline lanes indicated by ΔN, the change in the number of vehicles, and a corresponding increase in the mainline travel times associated with moving from point a to point b.[19]

Higher travel times in the mainline lanes induce demand for the ExpressLanes due to the corresponding rise in travel time savings from the gap in travel times between the two types of lanes. Importantly, the fixed relationship between throughput and the ExpressLanes toll as described in equation (1) means that the effect of substitution from mainline lanes to ExpressLanes is not necessarily to raise ExpressLanes travel times but rather to raise the toll. This can be seen in Figure 1, Panel B, which shows the relationship between tolls and traffic flows for a single segment. The figure plots an upward-sloping supply curve that becomes vertical when the 45 MPH minimum speed is binding at $\overline{N}$ and by a demand curve shift raising the price from P_1 to P_2 corresponding to points c and d. In other words, the ExpressLanes toll responds exclusively to the pattern of demand for ExpressLanes use, rising as entry to the lanes increases in the preceding five-minute interval. As Figure 1, Panel C, illustrates, ExpressLanes travel times also respond to demand increases with small, perhaps negligible increases as the lanes fill up, corresponding to the slightly declining speeds in Table 1. However, as the toll increases, demand for the ExpressLanes responds, bringing the travel times back to lower levels. Even with a large demand increase for the ExpressLanes, the number of vehicles will never exceed $\overline{N}$, the number of vehicles corresponding to a speed of 45 MPH. Turning to Figure 1, Panel D, we can see that the demand-induced increased in the toll for a single segment also increases the total toll paid for an entire

[19] Occupancy (vehicles per mile) is related to the horizontal axis of Figure 1 showing flow (vehicles per hour) through the fundamental diagram of traffic congestion.

ExpressLanes trip. A direct consequence of the ExpressLanes design, however, is that rising tolls filter out drivers with lower willingness to pay, changing the composition of driver preferences in our sample. We explore the implication of this for our findings in section V.D.

Panels A-C of Figure 1 only show segment-level responses of the ExpressLanes, since tolls vary by segment in response to demand. Any increase in the segment toll from panel B would translate into an increase in the total toll shown in panel D. It is useful, therefore to define a tolling equilibrium in the context of the ExpressLanes with segment-level demand and tolls.

Tolling Equilibrium in the ExpressLanes.—A tolling equilibrium is defined by a vector of five segment-level tolls over a five-minute period in-between adjustments in their level. Between five-minute intervals, the vector of tolls adjusts in response to demand conditions to maximize throughput. The equilibrium is therefore defined by the vector of segment-level tolls and a vector of segment-level demand. As we will illustrate in Section IV in the derivation of individual choice, the consumption of a quantity of segments is the margin we observe on each choice occasion in our data, and as result, the hedonic equilibrium that we derive will be defined across all segments for a five-minute interval rather than for any individual segment.

Segment-level demand includes the segment choices that drivers make for their entire journey along the ExpressLanes. These choices may be made after the set of tolls changes for other drivers entering upstream, but the toll paid by this driver reflects the set of rates they saw upon entry. Therefore, from the perspective of our empirical strategy below, there is no uncertainty in the toll paid from entry into the ExpressLanes through until the point at which they choose the last ExpressLanes segment. Specifically, drivers pay a total toll that is the sum of the segment tolls that they drive based on the entire set of segment-level toll rates locked in upon ExpressLanes entry. Drivers have information about the total toll that they pay based on information displayed upon entry to the ExpressLanes: the total toll for a drive along the entire length of the ExpressLanes, and for a partial trip exiting midway. As a result, this equilibrium follows the logic of hedonic equilibrium, namely the "law of one price," here a single vector of segment tolls over a five-minute interval.[20]

For illustrative purposes, consider a car entering the ExpressLanes towards the end of a five-

[20] In some instances, drivers may choose a combination of segments that is between these distances, in which case the tolls displayed bound the amount the drivers will pay for a partial trip. For example, a driver entering at WT01 will see total tolls for an exit at WT04 or WT06. They may choose to exit at WT03 or WT05, in which case, the displayed tolls bound the amount drivers will pay for these trips. Therefore, even in these case, uncertainty in the toll paid would be limited to a minimal approximation error.

minute interval, seeing a set of displayed tolls. They may choose their final segment after this interval has expired, but still pay the original toll they saw. A second vehicle, entering after the end of that interval, will face a new set of tolls reflecting the congestion imposed by the first vehicle. While both of these vehicles may be on the ExpressLanes at the same time, they reflect two separate equilibria. Interactions between these vehicles are unlikely to be meaningful for two reasons: the second vehicle is likely to be upstream of the first and so cannot impose congestion on it, and the contribution of any single vehicle to the congestion level of the ExpressLanes is likely to be small. As we discuss below, we cluster standard errors at the segment level reflecting the fact that these interactions may induce serial correlation.

Exogeneity of Mainline Congestion to ExpressLanes Tolls.—A useful question that helps inform our research design below is to what extent ExpressLanes tolls impact mainline lane congestion. ExpressLanes flows are substantially lower than mainline flows, with the median number of cars per hour in the ExpressLanes being 17.0% of that for mainline flows across all lanes. Nevertheless, as a test of this impact, in Appendix B and Appendix Table G.2, we report results from a regression discontinuity in time around the opening of the I-10 ExpressLanes to measure the impact on mainline speeds. These results show no statistically significant impact of the implementation of the ExpressLanes on mainline speeds across a range of bandwidths, suggesting that the availability of ExpressLanes has no discernable effect on mainline congestion. This is despite the fact that when we perform the same analysis on the ExpressLanes, comparing the speeds on the HOV lanes that preceded them to the speeds on the ExpressLanes, we find that speeds fall by a statistically significant 3.15%. This provides strong evidence that the difference in the sizes of the markets for ExpressLanes and for mainline lane traffic means that there is no observable effect of ExpressLanes tolling on mainline congestion.

III. Conceptual Framework and Data

Before describing the data used and how key variables are constructed, we begin with a framework to understand how drivers are likely to form expectations over uncertain travel time and the extent to which we are able to measure this with our rich data. In Section IV, we then present a choice framework for how this influences driver choices which allow us to recover the components of willingness-to-pay.

How Drivers Evaluate Time Savings.—A central component of the framework is to explain how drivers interpret both the deterministic and stochastic components of travel time savings in ExpressLanes. Three canonical models from the transportation literature provide the basis for how empirical work has sought to recover time preferences from commuting decisions: the time allocation problem deriving from Becker (1965) and extended by DeSerpa (1971) and Small (1982), the bottleneck model proposed by Vickrey (1969) and formalized by Arnott, et al. (1993), and the lane choice model that leverages a discrete choice approach as used by Small, Winston, and Yan (2005).

A standard approach is to show that the value of time arises from the ratio of the shadow cost of time and monetary constraints, but that schedule constraints (among other things) can affect this ratio directly through changes in marginal utility or through constraints directly. In these models, a commuter traveling to work during the morning peak along an untolled highway is typically subject to uncertain levels of congestion. A key point is that drivers can always reduce the likelihood of late arrival (and associated penalties) by departing earlier. As a result, the bottleneck model solves for a set of optimal departure times. In the case of the ExpressLanes, departure time has already been determined at the point of ExpressLanes entry, and much uncertainty about travel time has been resolved once drivers observe the level of congestion on the road, however, the scheduling choice framework will influence how we define the driver's problem in section IV and so proves useful.

The driver forms expectations about congestion levels that determine the expected arrival time, $E[T^A]$, relative to her desired arrival time, T^*. The driver then decides on a departure time, T^D, to minimize the opportunity cost of the time associated with commuting. Starting with Small (1982), empirical researchers have transformed a consumer utility maximization problem with budget, time and schedule constraint sets, as in Becker (1965), into an equivalent cost function formulation.[21] For example, Noland and Small (1995) model how the departure time decision trades off early and late arrival with commuting costs as follows:

[21] Direct estimation of travel demand is hampered by several conceptual and practical problems as articulated in Small (1982). An alternative approach which underlies that paper's approach as well as that laid out in this paper in Section IV follows Train and McFadden (1978) to estimate the dual problem that reflects monetary and time costs.

$$(2) \qquad \min_{T^D} Cost = \underbrace{\theta \cdot (E[T^A(T^D)] - T^D)}_{\text{Travel Time Cost}} + \underbrace{\beta \cdot \min(T^* - E[T^A(T^D)], 0)}_{\text{Continuous Early Arrival Cost}} + \underbrace{\gamma \cdot \min(E[T^A(T^D)] - T^*, 0)}_{\text{Continuous Late Arrival Cost}} + \underbrace{\delta \cdot \mathbf{1}\{E[T^A(T^D)] > T^*\}}_{\text{Discrete Late Arrival Cost}}.$$

These costs have four components, all of which depend upon the expected time of arrival, $E[T^A(T^D)]$, which is uncertain. Arrival times have a deterministic component which reflects free-flow travel times and a stochastic component which is determined by congestion levels and therefore also the commuter's choice of departure time, T^D, since departing earlier or later will change the state of congestion a driver faces and therefore travel times. The cost components are the time spent commuting, *Travel Time* (TT), the cost associated with arriving early, *Early Arrival Cost*, the cost associated with arriving late, *Continuous Late Arrival*, and a discrete cost of arriving late that does not scale with how late the individual is, *Discrete Late Arrival Cost.* θ is the cost per hour of time spent commuting, β is the cost per hour of arriving early, γ is the cost per hour of arriving late, and δ is a discrete lateness penalty that does not scale with time and provides the conceptual basis for the value of urgency. In practice, these preferences may vary by individual and could be captured depending upon the empirical specification, as is the case for our approach described in section IV.B.[22]

Treatment of Travel Time Uncertainty.—Two approaches are possible to recover the set of preferences in equation (2). We can think of a driver evaluating expected travel time, where from equation (2): $E[TT(T^D)] = E[T^A(T^D)] - T^D$, in terms of a deterministic, $\mu(T^D)$, and a mean zero stochastic component, $\sigma(T^D)$, which depends on departure time because of congestion:

$$E[TT(T^D)] = \mu(T^D) + \sigma(T^D).$$

This approach is analogous to portfolio theory approaches to incorporate mean-variance utility over investment decisions (Baumol 1963; Levy and Markowitz 1979) and has been used in transportation by Noland and Pollack (2002) and Hollander (2006), but ignores scheduling components of time preference. Alternatively, as shown in Bates, et al. (2001) and Fosgerau and Karlström (2010), one can take expectations over equation (2), and recover an expression for cost

[22] A general rule for these parameters based on several decades of empirical research is that, in most empirical settings, γ is four times β and θ is twice β (Small 2012). Units of time are specified here as hours so that they can be made comparable to hourly wage equivalents in the context of the value of time.

that approximates continuous late arrival costs with an empirical measure of the shape of the travel time distribution, $\widetilde{H}(\cdot)$:

$$(3) \qquad E[Cost(T^D)] = \theta \cdot \mu(T^D) + \widetilde{H}(T^A(T^D), T^*; \theta, \beta, \gamma) + \delta \cdot \Pr[T^A(T^D) > T^*].$$

Here, the first term on the right-hand side is the cost of deterministic travel time, the second are continuous early and late arrival times, and the last is the discrete penalty for uncertain late arrival. The second term, $\widetilde{H}(\cdot)$, captures not only the effect of uncertainty itself, but also the extent to which drivers dislike travel time uncertainty because it is more likely to make them early or late. The second term also approximates the components of stochastic travel times (and their cost valued by θ) from the mean variance approach, $\sigma(T^D)$.

Two necessary and defensible assumptions allow us to derive empirical approximations to $\widetilde{H}(\cdot)$, which we describe in the next subsection. One, with relatively few exceptions (Kreindler 2022), the econometrician does not observe desired arrival times, T^*, nor may she know drivers' predeparture expectations about expected travel times, $E[T^A(T^D)]$, and it may even be challenging to recover the true departure times, T^{D*}, from travel diary surveys. Specifying $\widetilde{H}(\cdot)$ based on past realizations of the distribution of travel times around a common departure time (e.g., drivers who leave for work between 8-9AM over the past month) allows us to evaluate how variability affects drivers' decisions, under the assumption that drivers use their experience from recent commutes to inform their expectations of uncertainty. Second, common sense and much of the transportation literature would indicate that drivers are likely to value unexpected lateness differently than unexpected early arrival. As a result, we follow this literature to allow $\widetilde{H}(\cdot)$ to be additively separable in terms corresponding to the left and right portions of the travel time distribution. This conceptualizes our measurement of travel time reliability, which we will define in the context of differences in travel costs between lanes in subsection III.B below.

A. Sources of Data

We combine data from two sources. Data on highways in California are collected by the Freeway Performance Measurement System (PeMS), a joint effort by the California Department of Transportation (Caltrans), the University of California, Berkeley, and the Partnership for Advanced Technology on the Highways (PATH). PeMS obtains real-time 30-second loop detector data on traffic flows and lane occupancy, which are then used to calculate speeds for the

ExpressLanes and mainline lanes over five-minute intervals.[23] Data on ExpressLanes users are collected by the Los Angeles Metropolitan Transportation Authority (LA Metro) based on each driver's transponder through a unique identifier. Every time the driver enters the ExpressLanes, information on the time of use, points of entry and exit, and toll charged is collected. In addition, these data report the primary vehicle associated with a transponder and the zip code from the driver's billing address.

We focus on the westbound direction of the 10.5-mile corridor of the I-10 where the ExpressLanes are located. During the morning peak hours, traffic in this direction predominately consists of morning commutes to work, with drivers potentially facing penalties for late arrival. Importantly, this corridor is one of the most congested morning weekday commuting corridors in the country. It also has one of the highest PeMS detector concentrations of any freeway in the US, with 3.5 detectors per mile in the mainline lanes and 2.73 per mile in the ExpressLanes, as shown in Appendix Figure F.1. This density ensures that the calculation of travel times is not overly dependent upon a small set of detectors, minimizing concerns over error in the measurement of travel times. Our full dataset contains 982,056 trips in the ExpressLanes on this route spanning the period from February 23rd, 2013, to December 31st, 2013.[24] We focus on the 433,623 trips that occur during the weekday AM peak on accounts registered to private households.[25]

To take advantage of the same individuals' repeated appearances in the ExpressLanes, we use a version of our model with heterogeneous preferences and restrict the sample to accounts that appear 5 or more times in our sample, corresponding to 9,054 accounts and 409,783 trips. We also examine accounts appearing in the lane during both the weekday AM peak hours and the same period during weekends. This subsample includes 758 accounts.[26]

B. *Travel Time Savings and Reliability*

In the context of the ExpressLanes, the perceived travel costs are relevant with direct reference

[23] Induction loop detectors are sensors buried in the roadway and measure vehicles passing above them. For more details, see Chen (2003).

[24] We do not include data past December 31st, 2013, to avoid contamination of our estimates from the Clean Air Vehicle Decals program, which granted access to low-emissions single-occupant vehicles and was implemented on January 1st, 2014. For a study of this program, see Bento et al. (2014).

[25] Private accounts correspond to 97.8% of the accounts and 89.8% of the trips in our data.

[26] For the additional robustness checks presented in the Appendix, we also use vehicle price data from Ward's Automotive Yearbooks (1945–2013), weather data for Los Angeles during our sample period from the National Climatic Data Center, and weekly retail regular reformulated gasoline prices for Los Angeles from the Energy Information Agency. For robustness checks, we use data from PeMS and the ExpressLanes during morning (5–9 AM) and afternoon (4–8 PM) peak periods for the I-10 W I-10 E, I-110 N, and I-110 S. AM peak speeds from PeMS are also used for the I-210 W. Appendix Table G.3 and Figure 5 use 2013 ACS Census data on income and commuting times. More details about data construction can be found in Appendix A.

to the untolled alternative, the mainline lanes. As a result, equation (3) should be interpreted as reflecting the cost differences in terms of the deterministic and stochastic components of *travel time differences* between the lanes. We operationalize equation (3) to reflect a commuter who decides how much of the ExpressLanes to use based upon the travel time they actually realize at the end of the trip reflecting drivers with perfect foresight. However, the value of the ExpressLanes may be augmented by the uncertain possibility of a slowdown in the mainline lanes, and so we also include a measure of reliability to reflect this. Any remaining uncertainty for the driver associated with time savings will be reflected as measurement error in our empirical formulation for which we describe our approach to address in section IV.C. We now explain how the deterministic component of travel time savings are measured as well as reliability.

Time Savings.—The differences in travel times between the mainline and ExpressLanes for each individual, segment, and time are computed as follows. For trips in the ExpressLanes, the travel time is simply the difference between the timestamps at the entry and exit segments. The travel time for the mainline lane trips is calculated as the distance traveled by an individual in the ExpressLanes divided by the average speed from PeMS detectors on a parallel stretch of the mainline lanes during the same 5-minute interval as that in which the trip began.[27] The trip-level travel time savings, which are the sum of segment travel times differences, are then the difference between the realized travel time in the ExpressLanes and that for a trip of the same distance taken at the same time along a parallel stretch of the mainline lanes of the I-10 W:

(4)
$$\Delta TT_{s,t} \equiv \sum_{\mathcal{s} \in s} \Delta TT_{\mathcal{s},t} = \sum_{\mathcal{s} \in s} \left(TT^{ML}_{\mathcal{s},t} - TT^{EL}_{\mathcal{s},t} \right).$$

Naturally, when commuting to work, individuals may use corridors other than the I-10 W. However, travel times in the mainline I-10 W lanes serve as a good proxy for the times associated with all these alternatives. This is because during the peak commuting period, the Nash equilibrium in routing serves to equalize average travel times between substitute commuting routes, so for a commuter taking a trip along the I-10 W corridor, travel times in the mainline lanes of that highway are consistent with the lowest possible travel times for any mainline (untolled) route commutes in the transportation system. In other words, there is little scope for saving time by taking an alternate

[27] That individuals would infer travel times in the mainline lanes based on contemporaneous speeds is consistent with the fact that the speed data from PeMS and other sources are widely available from news outlets. In addition, mobile technology such as Waze that tracks user speeds provides extremely accurate travel time predictions based on contemporaneous travel conditions.

freeway because congestion during the peak is high there, too. This also provides the rationale for our focus on the morning peak period, as during the afternoon peak period, drivers may, on average, have considerably more discretion over their departure times and commuting routes and modes than in the morning. In section V.E, we consider alternative ways to calculate trips and travel times.

Reliability.—As reflected above in equation (3), we can think of travel time savings as having a deterministic and a stochastic component. Moreover, as discussed with reference to that equation, when evaluating how drivers perceive travel costs, they are likely to value variability in travel time that is likely to make them late different from that which may make them early. As a result, empirical studies in transportation have often focused on a measure from the travel time distribution that reflects the likelihood that travel conditions will be much worse. These capture part of the underlying variance in travel times, specifically the part that is likely to make drivers late and are analogous to WTP for electricity reliability or a vehicle that is unlikely to break down (Borenstein, Bushnell, and Mansur 2023; Barber and Darrough 1996). A standard approach is to measure this form of reliability in time savings in terms of differences in the spread of the travel time distribution between the lanes, for example, the difference in the spread between higher quantiles and the median travel time savings. We measure $\Delta Reliab_{s,t}$ as the difference in reliability between the ExpressLanes and mainline lanes.[28] For a sum of given segments, *s*, day of the week and hour, reliability is the expectation difference in the travel times at the median and the 80th quantile:

$$\Delta Reliab_{s,t} \equiv \sum_{\mathcal{s} \in s} \Delta Reliab_{\mathcal{s},t} = \sum_{\mathcal{s} \in s} \left(Reliab_{\mathcal{s},t}^{ML} - Reliab_{\mathcal{s},t}^{EL} \right). \tag{5}$$

Each term is constructed as an average over a 30-day rolling window, that is, from the subsample of all previous ExpressLanes travel times at that given hour over the 30 days preceding date *t*. Here as in the rest of the paper, we subsume date and time into a single index *t*. As a result, $\Delta Reliab_{s,t}$ has considerable variation across observations by hour, date and segment where the trip is taken.

[28] As with the value of time, the value of reliability is commonly interpreted with reference to commuter's hourly wage equivalents in dollars per hour. As such, we measure $\Delta Reliab_{s,t}$, differences between lanes in the difference between the 80th and 50th quantiles of travel time in hours.

Appendix Figure F.3 gives an indication of the extent of this variation.[29] All prior work that we are aware of constructs this measure, in contrast, based on the value of travel times over a fixed period of time, implying that, in a given instance, reliability would be constant. Small, Winston and Yan (2005), for example, collect data on travel times over 11 days for the hours of 4–10 AM to model ExpressLanes use in Orange County, California. Given the small time window of the data collected, they use the entire distribution of travel times to construct estimates of reliability by hour of the day. As a result, their measure remains the same for drivers on any given day during that sample.

C. Trip-Level Summary Statistics

Table 1 displays key trip-level summary statistics by decile of travel time savings, where higher deciles imply higher travel time savings from taking the ExpressLanes. A comparison of columns III and IV underscores that the variation in travel time savings comes primarily from mainline lanes. As we move across deciles, speeds in the mainline lanes decline 7 times more than those in the ExpressLanes. Importantly, even throughout the morning peak hours, ExpressLanes speeds are always above 62 MPH. A standard metric for assessing the extent to which roadway capacity constraints are binding is whether speeds are above or below the free-flow level. Using a standard speed-flow estimation approach (Small and Verhoef 2007), we estimate the free-flow speed along the ExpressLanes corridor to be 62.3 MPH. While there is very limited variation in ExpressLanes speeds, Appendix Figure F.4 demonstrates that there is substantial variability in mainline speeds both within and across days, from which it is clear that the lion's share of variation in travel time differences between the mainline lanes and the ExpressLanes comes from the former.

D. Suggestive Evidence of Urgency and Schedule Constraints

Urgent Drivers.—Panel A of Figure 2 plots the kernel-smoothed estimates of the average toll per hour against the travel time saved for drivers who use ExpressLanes in California. The toll paid can be interpreted as a lower bound of average willingness to pay to use the ExpressLanes.[30] Seen on a per-hour basis, this resulting implied WTP to use the ExpressLanes appears to be absurdly large as the time saved decreases and is many orders of magnitude above even the most recent

[29] The measure is constructed independently of whether the driver in question was in the ExpressLanes on the prior 30 days in the data or not.

[30] For presentation purposes, the figure is truncated at $120 per hour but continues to substantially higher values. Table 1 shows the distribution of travel time savings by decile and key trip attributes such as tolls paid and distances traveled. Panel B of Figure 5 shows the evolution of the average toll per mile over the morning peak during the first, third, and sixth months of our sample, demonstrating considerable variation.

estimates of VOT in Goldszmidt et al. (2020), where the VOT is about 75% of the hourly wage.[31] Perhaps even more strikingly, Panel B of Figure 2 shows that these observations are not outliers—rather, they correspond to the bulk of all ExpressLanes uses.

One explanation for this high implied WTP for small time savings could be that drivers deviate from complete rationality by choosing to enter ExpressLanes for rather small time savings or that they simply misperceive these savings. In Panels C and D of Figure 2, we consider drivers who use the ExpressLanes only 1–3 times and those who use them 4–10 times during the weekday morning peak. There is little discernable difference in the average distributions of WTP per hour for the two groups, suggesting that how regularly drivers use the ExpressLanes has little impact on their revealed valuation of time savings. In Appendix Figure F.5, we also show results from individuals with 11–15 and 16–20 uses per month and observe little discernable difference. In Appendix Figure F.6, we compare WTP per hour by account for the first appearance of the account in the I-10 W ExpressLanes during the AM peak, the second to fifth appearances, the thirtieth to fiftieth appearances and hundredth appearance and up. Generally, experience should reduce mistakes, but in the figure, experienced drivers do not show qualitative differences in their behavior. Moreover, since drivers typically make morning commute decisions daily, it seems plausible that bounds on rationality due to a lack of experience may not be a valid explanation for this high WTP for small time savings.

Heterogeneity in the valuation of time may provide an alternative explanation for the hyperbolic shape of the implied willingness to pay per hour displayed in Panel A of Figure 2. Appendix Table G.3 shows that even drivers with presumably high implied WTP per hour use the lane infrequently at 8.8 times per month—a frequency that inexplicably increases as the average implied WTP per hour decreases across the deciles.[32] This infrequent use helps to explain the number of unique individuals in our data, 26,833. If those users in the first decile truly valued their time at this level, we would expect them to appear in the lane considerably more frequently and for nearly every commute. Further, if this were the case, the implied hourly wage that would explain this choice based on the value of time alone would be $640 per hour, corresponding to an annual wage of $1.3

[31] In Appendix Table G.3, we report the equivalent average hourly wage by zip code of individuals using the ExpressLanes, which is $19.63 on average in our sample.

[32] Appendix Table G.3 provides more detailed information about ExpressLanes use by decile of $\Delta TT_{i,s,t}$ including average implied WTP over different periods of our sample, and information about vehicles registered to account. Appendix Table G.4 shows that there is substantial variation in the frequency of use across deciles of $\Delta TT_{i,s,t}$. The bulk of use within any given decile is generally evenly distributed between 2-5, 6-10 or 11-20 uses per month.

million. This level of income would most likely be incongruous with the types of cars that we observe most frequently to be associated with the use of the lane: Honda Civics, Toyota Corollas and Toyota Camrys, as shown in Appendix Table G.5.[33]

Urgent Drivers Face Schedule Constraints.—Panel A of Figure 3 depicts the kernel-smoothed density of demand over the morning peak, where ExpressLanes use increases along the hours of the morning peak as expected. Vertical lines indicate the key hours of 7:00, 8:00, and 8:30 AM, when schedule constraints associated with work start times or morning appointments may be common. The distribution of ExpressLanes use in the figure rises 10–15 minutes before these times and then immediately falls afterwards illustrates a pattern of bunching which suggests that drivers may be using the ExpressLanes for on-time arrival at common work start times.

Further, in Panel B of Figure 3, we examine how tightly exit times cluster around the account's mean ExpressLanes exit time. First, we select accounts with an average exit time within a given 15-minute window, for example, 6:45–7:00 AM. We then plot the exit time of all trips associated with those accounts. The figures show that, in the morning, the exit times are closely clustered around the average window while, in the afternoon, there is a wider dispersion and relatively few exit times are within the focal 15-minute window. Morning exit times are less dispersed because commuters often face common fixed arrival times whereas they may have more discretion over afternoon peak arrival times when returning home or engaging in after-work activities. In addition, penalties for arriving after a desired time in the afternoon may be nonexistent or less severe.[34] Together, Figures 2, 3, F.5, F.6, and Table 1 provide initial convincing evidence of the existence of schedule constraints, penalties for lateness, and begin to create the basis for understanding the tolling market mechanisms that reveal individuals' preferences for urgency.

IV. Empirical Strategy

In this section, we lay out an empirical strategy to recover commuting preferences using ExpressLanes that uses a hedonic price regression of the total toll paid on differences in lane

[33] An implied hourly wage of $640 comes from dividing the average toll paid from the first decile in Table 1 by the average time savings in that decile, and multiplying it by two to reflect the fact that conventional estimates of the value of time roughly correspond to half the hourly wage. The actual average implied WTP per hour in the first decile is even larger as shown in Appendix Table G.3 at $1,977 per hour. An hourly wage of $640 becomes $1.3 million assuming an 8-hour workday and 255 working days per year.

[34] In Appendix Table G.6, we show that drivers leaving the ExpressLanes later during the morning weekday peak tend to choose to use the ExpressLanes for longer trips. This might reflect the fact that drivers consume ExpressLanes more the later that they expect to arrive at their destination.

attributes.[35] A key component to the revealed preference approach of our research design is that we take the choice of the ExpressLanes over the mainline lanes as given and, conditional on that choice, model the choice of the number of segments to take. This empirical approach is novel to the transportation literature and is made possible by two unique features of our data: One, the ability for drivers to choose combinations of segments, with different prices that respond, almost in real time, to demand, and which are locked in for all segments at the point of lane entry regardless of which segments they choose. In other words, there is no ex-ante uncertainty in the total toll paid for a trip. Two, our setting is unique due to the repeat transaction nature of our data which uses variation in segment choice by the same commuter on multiple appearances facing different levels of the toll, time savings and reliability difference.

However, this approach also has two important implications: first, we are working with a selected sample of only commuters who select into the ExpressLanes. We explore the implications of this fact in section V.F, but note that this feature is common to almost every other empirical application of hedonics (see Linden and Rockoff (2008) for an extended discussion). Second, in this section, we show how the optimal choice of a quantity of segments yields allows one to recover the components of WTP that reflect ExpressLanes attributes relative to the mainline lanes. However, this relies on continuity of ExpressLanes segments, which are actually discrete. We explain how this fact puts bounds on our estimates of the components of WTP, which we formalize in Appendix C. After laying out these features, we explain the equilibrium features of the hedonic price equilibrium of the ExpressLanes market. The rest of the section lays out the full details of our econometric strategy with particular attention to identification of parameters of the model.

A. Hedonic Approach for Recovering the Value of Urgency, Time and Reliability

In our empirical setting, we observe the total toll paid by drivers, which is the sum of segment-level tolls at the point of entry locked-in for the duration of their trip. We also observe the set of segments, s, they choose and the attributes of those segments. At this point, the commuter has already chosen a departure time conditional on expected roadway congestion as described from equation (2) in section III.

Segment-choice Decision.—At the point of ExpressLanes entry, much of the uncertainty about

[35] As an alternative, one could estimate a discrete choice model of travel demand. In our context, the additional predictive power of this approach would be entirely based on distributional assumptions given the lack of identifying demographic data.

congestion has been resolved. To connect commuter preferences with the market price explicitly, we write the conditional indirect utility, which imposes strong, but standard assumptions about quasi-linearity in income, as well as separability from other forms of consumption and time use. Drivers choose to consume a quantity of ExpressLanes segments $s = \bigcup_{\mathscr{s}=1}^{5} \mathcal{S}_{\mathscr{s}}$, where $\mathscr{s}$ indexes each segment and s indicates a set of segments. For each segment chosen $\mathscr{s}$, conditional indirect utility for of attributes varies by lane *l* and by individual *i* and time *t*:[36]

$$(6) \qquad U_{i,\mathscr{s},t}^{l} = \xi_{i,t}^{l} + \beta_i \boldsymbol{X}_{i,\mathscr{s},t}^{l} + \varepsilon_{i,\mathscr{s},t}^{l}\ , \quad l \in \{EL, ML\}.$$

$\xi_{i,\mathscr{s},t}^{l}$ captures lane-specific unobserved attributes, and $\boldsymbol{X}_{i,\mathscr{s},t}^{l}$ is a vector of lane-specific characteristics including travel time, $TT_{i,\mathscr{s},t}$, reliability, $Reliab_{i,\mathscr{s},t}$, and, in the case of the ExpressLanes, the segment-level toll based on the price locked-in and displayed at entry, $toll_{\mathscr{s},t}$. ξ_{it} does not vary by segment since it captures individual- and/or time-specific lane preferences, but there is no reason to qualitatively value a segment beyond its underlying time savings, though we will test this premise in our robustness checks by considering the role of time-invariant unobservables.

In contrast to equation (6), and for the rest of the paper, we will only consider the conditional indirect utility for combinations of segments, indexed as s, which is what we observe ExpressLanes drivers consuming, of which there are 11 possibilities given the set of entry and exit points.[37] As such, we rewrite lane-specific, conditional indirect utility for a quantity of segments, s, where the summation represents the sum of each attribute and the toll over the set of endogenously chosen segments:

$$(7) \qquad U_{i,s,t}^{l} \equiv \int_{\mathscr{s}=0}^{s} U_{i,t}^{l}(\mathscr{s})d\mathscr{s} = \xi_{i,t}^{l} +$$

$$\int_{\mathscr{s}=0}^{s} \left(\beta_i^0 toll_t(\mathscr{s}) + \beta_i^1 TT_t(\mathscr{s}) + \beta_i^2 Reliab_t(\mathscr{s}) + \varepsilon_{i,t}^{l}(\mathscr{s})\right) d\mathscr{s}\ , \quad l \in \{EL, ML\}.$$

[36] Here, and in subsequent equations, $l = EL$ is the ExpressLanes and $l = ML$ is the mainline lanes.
[37] To see the set of possible combinations of segments in ExpressLanes design, see Appendix Table G.1.

Here the first line shows that at the segment combination-level, we are summing the segment level conditional indirect utility over segments, which corresponds to a constant plus the sum of segment level tolls, travel time, reliability and idiosyncratic error.

The central choice modeled in our revealed preference approach is the selection of a set of segments s corresponding to quantities of each attribute that sum up across the set of segments used. We assume, for now, continuity of segments but, as we discuss below, our estimates become bounds on the parameters when we account for discrete segments.

Implicit in this formulation is that the lane choice problem from equation (6) has already been made and we are conditioning on it. Moreover, in the context of the tolling equilibrium defined in Section II, the ExpressLanes market clears with a single vector of combinations of segment-level tolls every five minutes. Because drivers choose a combination of segments and also observe the total tolls they will pay at the point of entry, the equilibrium corresponds to the "law of one price" necessary for a hedonic equilibrium to hold. This equilibrium, as described in detail in Section II, is defined over all segment combinations over a five-minute interval, where each combination of segments corresponds to a differentiated product in the sense of a classical hedonic model. We will now show how it is possible to recover components of the WTP for ExpressLanes use from the segment choices drivers make.

Revealing Valuation of Components of WTP from Segment Combination Choices.—Here, the indirect utility obtained becomes the difference between that for each lane from equation (6), $U_{i,s,t}^{EL} - U_{i,s,t}^{ML}$, resulting in:

$$\text{(8)} \qquad \max_{s} V(s) \equiv \Delta\xi_{i,t} - \beta_i^0 Toll_t(s) + \beta_i^1 \Delta TT_t(s) + \beta_i^2 \Delta Reliab_t(s) + e_{i,t}(s).$$

Here each term on the right-hand side of the equation except $\Delta\xi_{i,t}$ is the sum of segment-level components, and for the purpose of clarity, we write the total toll paid, which is the sum of segment-level tolls capitalized:

$$Toll_t(s) = \int_{\mathcal{s}=0}^{s} toll_t(\mathcal{s})d\mathcal{s}.$$

The choice variable s defines a quantity of location-specific ExpressLanes segments that the commuter has chosen. We have expanded the components of $\boldsymbol{X}_{i,t}(s)$ into $Toll_t(s)$, $\Delta TT_t(s)$ and

$\Delta Reliab_t(s)$, the total toll, difference in travel time and reliability between lanes, which are written as a function of the segments chosen. e_{it} is the difference in idiosyncratic error terms from equation (6).[38] $Toll_t(s)$ is the total toll paid to use all s segments, which is the sum of segment-specific, dynamically priced segments. Similarly, $\Delta TT_t(s)$ and $\Delta Reliab_t(s)$ are the total time savings and reliability difference for a trip across along the quantity of segments s. We index the parameters in equation (8) by i because the repeat-transaction nature of our data allows us to recover individual-specific preferences. However, this is done without having information about the source of the factors that create this variation. We will attempt to explain some of the potential sources of this variation in section V.B, but an important aspect of this approach is that it does not require the kind of information one would ideally have to recover heterogeneous preferences in discrete choice settings.

For equation (8) to have an interior solution would mean an optimal choice of $s^* > 0$, which implies that the lane choice problem from equation (8) has already been made and the driver has chosen to use the ExpressLanes. As a direct consequence of this fact, equation (8) has several useful properties. First, the quantity of the lane attributes and the total toll paid scales up directly with the number of segments traveled. Second, because drivers make choices across lanes, the relevant comparison is between a trip over a given set of segments in the ExpressLanes and the comparable trip in the mainline lane.[39] However, equation (8) imposes several assumptions that underlie the form of conditional indirect utility written: quasilinear preferences and separability from underlying consumption and time allocation outside of the morning commute. In the context of the decision over how many segments to choose after entering the ExpressLanes, it seems plausible that a driver has taken consumption and other time allocation as given and simply makes the choice modeled in equation (8). In addition, for parsimony, the attributes all enter linearly, although in our empirical specification, we allow for the possibility of non-linearity.

Importantly, there is nothing in equation (6) that directly reflects the value of avoiding discrete lateness penalties. To realize this, we need to compare the utility of the ExpressLanes to that of their unpriced alternative, which yields the difference $\Delta\xi_{i,t}$ in equation (8). How should this term be interpreted? All other attributes scale up monotonically with an increase in the number of segments, whereas $\Delta\xi_{i,t}$ does not. Thus, it captures the component of WTP that does not scale up

[38] We assume, for the purpose of this analysis that e_{it} is known to each driver, but not by the econometrician.
[39] Another important element of this comparison is that during rush hour, the mainline acts, on average, as a lower bound on alternative travel times, where trips on alternative parallel routes or even public transit therefore reflect higher comparable travel times and therefore greater disutility.

with expectations or the spread in time savings (i.e., reliability).

Recovering Components of WTP from Hedonic Price Function.—Two features of equation (8) help to demonstrate how we recover the components of willingness-to-pay from equation (8) using a hedonic price regression. First, taking the first order conditions and rearranging yields:

$$\text{(9)} \qquad \frac{\partial Toll_t(s^*)}{\partial s} = \frac{\beta_i^1}{\beta_i^0}\frac{\partial \Delta TT_t(s^*)}{\partial s} + \frac{\beta_i^2}{\beta_i^0}\frac{\partial \Delta Reliab_t(s^*)}{\partial s} + \frac{1}{\beta_i^0}\frac{\partial e_{i,t}(s^*)}{\partial s}.$$

How does uncertainty in time savings affect WTP? From section III, we can see that the component of uncertainty that corresponds to slowdowns in mainline lanes is captured by the reliability term. All other uncertainty in time savings reflects measurement error which we address econometrically through instrumental variables as described in section IV.C. By showing the derivative of the total toll paid from the demand equation with respect to the choice variable, equation (9) shows how the MWTP for use of the ExpressLanes responds to a marginal change in the quantity of segments used. Each of the lane attributes in equation (9) is valued by the ratio of its respective coefficient and β_i^0, which, since it corresponds to a dollar-denominated variable, is linked to the marginal utility of income. Indeed, the ratio of these coefficients reflects marginal willingness to pay for each of these differences in attributes, as is done in standard discrete choice applications. In a transportation context, they reflect the value of time and reliability. Another parallel with related discrete choice applications is that here, the ratio of coefficients is equivalent to the coefficients from a hedonic price regression of the toll on the same differences in attributes as we will now illustrate. [40]

However, for the first order condition of equation (8) to hold so that we can derive equation (9) requires that s must be continuous. As described in section II, ExpressLanes segments are discrete, but we proceed in this section assuming that s is a continuous variable for expositional ease. In Appendix D, we show formally that with discrete segments, a precise upper bound on the value of time can be derived and a precise lower bound on the value of urgency can be written that depends on the convexity of the toll function's response to demand. The intuition for each of these bounds comes from the fact that if drivers are rational, they will not consume more of the lanes than they

[40] The notation in this expression also reflects an underappreciated aspect of hedonic models: that it is possible to recover preference heterogeneity across individuals, *i*, from hedonic regression models provided one observes repeated choices (Bajari and Benkard 2005; Kuminoff and Pope 2014; Bishop and Timmins 2018).

would like (i.e., toll is less than or equal to WTP), and so the value of time will be at least smaller than the segment that bounds this choice from above. Similarly, the choice of the ExpressLanes over the mainline lanes means that the value of urgency has to be at least as big as the segment that bounds this choice from below.[41]

We lack intuition for valuation of $\Delta\xi_{i,t}$ in equation (9) since it does not directly affect marginal valuation of the change in segment. Instead, we have to rely on the lane choice decision that equation (8) is conditioned on, resulting in a non-negative value of $V(s)$. Imposing this condition and rearranging yields:

$$(10) \qquad Toll_t(s^*) \leq \frac{\Delta\xi_{it}}{\beta_i^0} + \frac{\beta_i^1}{\beta_i^0}\Delta TT_{s,t}(s^*) + \frac{\beta_i^2}{\beta_i^0}\Delta Reliab_{s,t}(s^*) + \frac{1}{\beta_i^0}e_{i,s,t}(s^*).$$

In this study, we focus our analysis on the willingness to pay that does not scale up in the observed attributes, which here corresponds to $\frac{\Delta\xi_{i,t}}{\beta_i^0}$. We interpret this component as the value of urgency. Equation (10) illustrates a pricing rule for how the value of urgency affects the valuation of lane attributes at the optimum: it acts as a shifter of the bid function reflecting WTP for use of the ExpressLanes. This pricing rule also makes clear that the toll incorporates a lower bound on this value corresponding to marginal individuals who are just indifferent between entering the ExpressLanes or not.

A Hedonic Approach.—In our empirical setting, certain features of the hedonic model make it more appealing than a discrete choice approach. First, our dataset provides rich repeat-transaction data for ExpressLanes drivers, but we have no information about mainline lane drivers other than an approximation of their number and speed. The appeal of the hedonic approach is its simplicity and transparency: the key assumption required is that mainline lane travel times can serve as a proxy for the next most desirable alternative.[42]

However, equations (9) and (10) differ from the approach typically leveraged in traditional hedonic models. In those models, the econometrician aims to use assumptions on preferences to recover an approximation to a Hicksian measure of welfare for exogenous changes of an amenity

[41] We ignore the effect of reliability on these bounds since its contribution to WTP is shown to be very small.
[42] Brownstone, et al. (2003) similarly recover preferences from ExpressLanes entry decisions along a corridor with dynamic pricing. However, that generation of ExpressLanes design did not allow for differentiated pricing between segments of the corridor, and so it is not clear that it would be possible to recover the bounds on valuation of WTP components from a hedonic model in that setting.

in a related market. In our present context, we observe commuters purchasing additional segments in a market where these choices translate directly into purchasing time: either directly as time savings, reduction in the spread of the distribution of time savings (i.e. reliability), or a discrete component of time savings that may ensure on-time arrival (i.e., urgency). As such, there is no need for an approximation as in Harberger (1964); we observe WTP directly from the revealed choices of individuals in the spirit of Varian (1982).[43] However, the goal of our exercise is not to recover WTP itself, but rather the contribution of different components to WTP.

Given the program background provided in section II, we use the properties of a hedonic price equilibrium to recover commuter preferences. A hedonic price function, $Toll_t = \tau_t(\boldsymbol{X}, \boldsymbol{\rho})$, describes the evolution of the total toll in five-minute equilibria, *t*, in response to changes in demand for the ExpressLanes given fixed capacity.[44] Importantly, this capacity is never reached in our sample as discussed in section II, since increases in segment-level tolls are sufficient to reduce demand and maintain speeds well above 45 MPH. As a result, the logic of the hedonic envelope embodied in equation (8) reflects the fact that toll filters out drivers with lower total WTP than the toll.[45] Changes in the composition of commuter preferences, $\boldsymbol{\rho} = \{\boldsymbol{\delta}, \boldsymbol{\theta}, \boldsymbol{\gamma}\}$, and the number of commuters translate into changes in demand and therefore lane attributes, $\boldsymbol{X}$ (time savings and reliability difference), and the toll adjusts separately for each segment, *s*, in response to match lane capacity to demand. $\tau_t(\boldsymbol{X}, \boldsymbol{\rho})$ is a vector of the set of total tolls possible for different combinations of segments, *s*.

Panel A of Figure 4 illustrates the hedonic price function for the ExpressLanes along a single dimension as it increases with greater travel time savings. The figure shows two commuters with different values of time, $\theta_2 > \theta_1$, which translate into different bid functions reflecting their MWTP for travel time savings.[46] As in a conventional hedonic setting, the marginal commuter with each of these valuations chooses a quantity of segments, $s_1 < s_2$, resulting in $\Delta TT_1(s_1) < \Delta TT_2(s_1)$ that clear the market at $Toll_1(s_1) < Toll_2(s_2)$.[47] Moreover, the tangency of bid

[43] Unlike in cases where one may wish to approximate WTP from a proximate market (e.g., Schlee & Smith (2019)), in our case the hedonic reveals valuation of the components of WTP of the product purchased itself, here ExpressLanes. As a result, our approach requires a much weaker set of identifying assumptions.

[44] $\boldsymbol{X}$ and $\boldsymbol{\rho}$ are matrices reflecting attributes and preferences. For parsimony, we suppress the subscript *t*, but both matrices change with each period.

[45] If the lane were capacity constrained, then there could be drivers with higher WTP who could not enter the ExpressLanes.

[46] We re-parameterize the model from equation (9) to reflect coefficients in a hedonic regression of the toll on attributes, where the new coefficients are those for the same attributes in (9) divided by β_i^0. Here, we also ignore the value of urgency for purposes of illustration, but these results hold into multiple dimensions with reliability held constant.

[47] Panel A also illustrates the well-known tangency condition between the hedonic price function and bid-rent curves that equates the implicit price function (the derivative of the hedonic price function) and an individual's willingness to pay for that attribute.

functions with respect to the hedonic price function at s_1 and s_2 corresponds directly to the MWTP relationship from equation (7).

Panel B of Figure 4 clarifies how the term $\frac{\Delta\xi_{i,t}}{\beta_i^0}$ is reflected in a hedonic price function by the parameter δ_i, corresponding to the value of urgency for three individuals, i = 1, 2, 3. As shown in equation (8), the value of urgency shifts the MWTP up and down relative to the hedonic price function. The value of the ExpressLanes relative to that of the relevant alternative, the mainline lanes, that does not scale up in the amount of time saved would shift the bid curve up or down for three individuals with the same value of time, θ_1. The individual with a value of urgency δ_2 amounts to the marginal individual lying tangent to the hedonic bid function. The individual with δ_3 lies above the function and is inframarginal, and therefore, estimates of δ_2 from a hedonic regression would be underestimates of the value of urgency. Lastly, the individual with preference δ_1 does not have sufficient willingness to pay and would not enter the ExpressLanes.

Our data also allow us to explore revealed preferences along another dimension by observing the same individual (here assuming a constant value of time θ_1) on different days. Each appears represents a distinct equilibrium and therefore hedonic price functions, corresponding to $\tau_1(\boldsymbol{X},\boldsymbol{\rho}), \tau_2(\boldsymbol{X},\boldsymbol{\rho})$ and $\tau_3(\boldsymbol{X},\boldsymbol{\rho})$ in Panel C of Figure 4. Each of these hedonic price functions reflect the attributes of the lane and the composition of preferences for commuters that clear the market at the corresponding equilibrium toll. For the individual plotted here, we can see that these three hedonic price functions lie tangent to the commuter's bid-rent curve corresponding to three distinct levels of travel time savings, $\Delta TT_3(s_3) < \Delta TT_1(s_1) < \Delta TT_2(s_2)$, reflecting different segment choices $s_3 < s_1 < s_2$ and resulting in different tolls paid $Toll_3 < Toll_1 < Toll_2$. Therefore, rather than a single value of this commuter's marginal willingness to pay for travel time savings, we can recover a full WTP curve for the characteristic.

Panel D of Figure 4 then illustrates how this willingness to pay translates into a commuter's demand curve. Here, the solid lines plot the derivative of the hedonic price function with respect to ΔTT at time t and k, called the implicit price function. The choices of ΔTT_3 made here correspond to intersections between the implicit price function and the (dashed) commuter's travel demand curve. Thus, under the implicit price function for time t, consumer 1 chooses segment s_1 resulting in $\Delta TT_1(s_1)$, while commuter 2 chooses segment s_2 giving $\Delta TT_2(s_2)$. Each new five-minute interval in our data results in a new equilibrium and therefore implicit price function, whether from cyclical patterns of demand or idiosyncratic shocks like weather or accidents. Thus, for consumer 2, we also observe her in period k choosing fewer segments corresponding to smaller travel time savings. These two choices trace out a linear approximation of the commuter's demand curve that can be made more flexible with more appearances in the lane.

Lastly, Panel D of Figure 4 helps illustrate a useful empirical property of hedonic price models: they rely on identification via *equilibrium market conditions*. As highlighted in Rosen (1974), the presence of endogenous supply of the product model does not confound estimation of hedonic models. Rather, the hedonic price function acts as an envelope responding to implicit prices for attributes based on bid functions from consumers reflecting their WTP for those attributes and offer functions from suppliers reflecting the cost to supply the same attributes.

B. Estimation of Hedonic Toll Regressions

Given the properties of a hedonic pricing equilibrium just explained and the fact that one can derive WTP as it relates to both changes in the differences of travel time and reliability between the ExpressLanes and the mainlines and a constant, equations (9) and (10) allow one to the express an empirical estimating equation in the form of a hedonic price regression. Our baseline results are estimated form the following highly parameterized hedonic pricing function:

$$Toll_{i,s,t} = \delta_{i,t} + \theta_{i,t}\Delta TT_{i,s,t} + \gamma_{i,t}\Delta Reliab_{s,t} + \varepsilon_{i,s,t}, \qquad (11)$$

Where $Toll_{i,s,t}$ is the total toll paid over all segments driven, locked in at their level at ExpressLanes entry, by individual i for access to a combination of segments s at entry time t in the ExpressLanes. As equation (9) shows, there is reason to believe that the error term may respond to changes in the number of segments used, so as a result we cluster the standard errors by segment.[48] We estimate equation (11) for the hours of 5–9 AM in the morning peak, when it is reasonable to expect that most trips are work commutes, where individuals have limited discretion over their departure time and potentially face lateness penalties. $\Delta TT_{i,s,t}$ is the difference in travel

[48] The standard errors in our account-level regressions are based on the bootstrapped standard deviation of coefficient estimates and reflect the statistical precision of the mean of the coefficient distribution as distinct from the standard errors in column 1, which reflect the statistical precision of our estimate of the conditional mean of the covariates on the dependent variable. Whether in the homogeneous or heterogeneous model, standard errors are clustered at the subsegment level (and bootstrapped in the case of the heterogeneous model). In Appendix Tables G.7 and G.8, we present other forms of clustering, including two-way clustering (Cameron, Gelbach, and Miller 2011) to address potential spatial and temporal correlation (Anderson 2014). Since the toll level in each period is a function of the previous period's toll, we account for autocorrelation in the toll across a variety of approaches in section V.E and find that clustering by segment produces the largest standard errors.

time between the mainline lanes and the ExpressLanes, and $\Delta Reliab_{s,t}$ is the difference in reliability in the mainline lanes and ExpressLanes capturing the uncertainty in travel time as function of the value of relatively certain travel time in the ExpressLanes relative to the mainline lanes.[49] While drivers may not be able to perfectly calculate $\Delta TT_{i,s,t}$ at the point of lane entry, this does not pose a problem for estimation of hedonic regression beyond traditional concerns of measurement error which we discuss in detail in the next subsection.[50] $\theta_{i,t}$ and $\gamma_{i,t}$ are estimates of the value of time and reliability, respectively. $\delta_{i,t}$ is the estimate of the value of urgency, the component of willingness to pay for travel time savings independent of the amount of time saved, which we allow to vary by individual or time period, depending on the specification.

Another feature of our data reflected in equation (11) is its panel nature, observing the same individual across different days and therefore toll, attribute and segment choices, helps recover preferences for continuous attributes as shown in equation (9). On the other hand, the cross-sectional nature of the data, observing some commuters entering the lane on a given day, provides identifying variation to recover the discrete value of urgency. As discussed above, the value of urgency reflects willingness to pay to avoid discrete lateness penalties. As conceptualized in section III.A, commuters can be thought of as having discrete and continuous lateness penalties, and as shown in section III.C., unlike the measures in many other empirical studies, our measure of reliability differences, $\Delta Reliab_{s,t}$, has considerable variation by date, hour and segment, making it likely that it recovers continuous lateness penalties as distinct from the discrete ones recovered, $\delta_{i,t}$. As an indication of this fact, when we estimate equation (11) without reliability in Appendix Table G.9, it has an effect on the coefficient on $\Delta TT_{i,s,t}$ but almost no meaningful effect on our estimates of $\delta_{i,t}$.

C. Identification

A common concern in estimating hedonic price regressions is the existence of endogenous

[49] Of course, there is no reason to believe that these attributes have to enter the hedonic price function linearly, an issue we explore in section V.E by allowing the hedonic to be estimated non-parametrically and with non-linear terms.

[50] Indeed, a review of the seminal empirical hedonic literature from the past several decades shows that most amenities being valued are usually subject to some form of substantial uncertainty. A formal treatment of this issue is presented in Kask and Maani (1992). Ma (2019) and Hausman and Stolper (2021) which show how variation in information about (dis)amenity levels affects choices and outcomes. A broader set of hedonic set of papers where uncertainty plays a major role include those considering the valuation of cancer clusters (Davis 2004), crime (Bishop and Murphy 2011; Linden and Rockoff 2008), air quality (Chay and Greenstone 2005), water quality (Muehlenbachs, Spiller, and Timmins 2015; Kuwayama, Olmstead, and Zheng 2022), extreme weather (Bin and Landry 2013; Hallstrom and Smith 2005), flooding (Bakkensen and Barrage 2022), and school quality (Black 1999; Figlio and Lucas 2004). The standard approach to address uncertainty, all of which we follow in section V, is to subject models to robustness testing of how amenities are measured, include additional controls for the extent of uncertainty, and address measurement error through instrumental variables.

attributes. For example, when measuring capitalization effects of improvements in air quality on housing prices, past studies have been concerned that both are positively correlated with improvements in unobserved local economic conditions (Chay and Greenstone 2005). In our setting, a potential concern may be the presence of unobserved shocks to travel demand—for example, from an accident somewhere else in the transportation system or a sudden increase in the number of drivers on the road before or after a sporting event. When we consider the determination of the total toll from equation (1) in section II, it is clear that shocks to demand for the corridor translate to changes in $\Delta TT_{i,s,t}$, which are, in turn, mediated into changes in occupancy, $o_{s,t}^{EL}$, which determines the total toll, resulting in a new hedonic equilibrium. This example makes clear that this kind of variation in demand—indeed nearly any variation in demand—allows us to recover variation in willingness to pay to recover individual preferences. Since the main purpose of our empirical exercise is to understand the relative contributions of the components of equation (8) to willingness to pay for the ExpressLanes, another potential effect of a demand shock could be to change the distribution of preferences of ExpressLanes users. This would not bias our estimates but rather has implications for their external validity, a concern that we discuss in section V.D. In summary, the context of dynamic tolls in the ExpressLanes obviates the need for instruments typical of other hedonic settings.

To put this in the context of a hedonic model, Panel D of Figure 4 illustrates how such a shock would result in a shift of the implicit price function for travel time savings. Considering two moments in time before, *t*, and after, *k*, this shock, the comparison of these two static equilibria allows us to recover two points along a given commuter's demand curve, an exercise that relies only on the assumption under the setting of a standard static hedonic equilibrium that commuters can choose a new set of segments regularly or can costlessly reoptimize.

A separate concern could be that because $\Delta TT_{i,s,t}$ and $\Delta Reliab_{s,t}$ are constructed as differences including values for mainline lanes, reverse causality could occur if ExpressLanes tolls affect mainline congestion. However, as discussed in section II, travel times in the ExpressLanes have a negligible impact on $\Delta TT_{i,s,t}$ given the size of mainline lane variation. In addition, mainline lane congestion, and therefore travel times, cannot be affected directly by the toll given the difference in the relative size of the mainline and ExpressLanes markets as illustrated by our regression discontinuity results. As a result, reverse causality is not a plausible explanation, and the same logic applies to differences in reliability.

Finally, interpretation of estimates for the value of reliability has been given increased scrutiny given a nascent literature which documents how drivers experience psychological distress from uncertain traffic conditions. If drivers simply dislike variability but do not face time-varying lateness penalties, this could affect the interpretation of reliability in our main specification. For example, if drivers value the tranquility in avoiding the stress of uncertainty (Steimetz 2008), this would lead us to overestimate the value of reliability.[51] To the extent that these preferences are correlated with expected congestion based on traveling at particular times of day, including time fixes effects, as described in section V.A, will capture their effect separately from the values of time and reliability. However, since this preference for tranquility responds to the same variation in traffic conditions used to recover the value of time and reliability, to the extent that it responds to idiosyncratic changes in travel conditions, it may not be possible to disentangle it from pure time preferences. This is an interpretation challenge shared more broadly in the estimation of time preferences from revealed preference data and an important avenue for future research.

Time-Invariant Unobservables.—A separate concern is whether the constant $\delta_{i,t}$ actually recovers the value of urgency or, instead, is contaminated by the presence of other unobserved time-invariant attributes of the ExpressLanes. If such an unobserved attribute exists, it cannot take the form of congestion, which would generate travel time savings.[52] Instead, it could simply reflect some drivers' belief that ExpressLanes are safer, perhaps due to differences in pavement quality, or that they provide a relatively smoother ride and more comfort than the mainline lanes. To address this concern, we take advantage of the fact that off-peak trips are less likely to be characterized by significant urgency because they may have a lower penalty for late arrival and so can be used as a control for weekday urgency.

We pool weekday morning and weekend ExpressLanes trips and decompose the constant into a weekday morning peak component, δ_i, and a component that is always present, μ, so that equation (11) now becomes:

[51] Because this valuation would nevertheless scale with the variance of travel time, it would only bias our estimates of the value of urgency if it resulted in misspecification of the regression, so test the effect of different functional forms and distribution moments in Appendix Tables G.10 and G.11 and find no effect on the value of urgency.

[52] In principle, there could be unobserved, travel time–varying attributes that scale with the level of congestion such as the "value of tranquility" from using the ExpressLanes. Unobserved attributes that scale with distance or travel times in only one of the lanes are likely to be highly correlated with $\Delta TT_{i,r,t}$, and so, to the extent that these exist, we may attribute some portion of the implied value of time to these attributes. Given the consistency of our value of time estimates with the those in literature, the valuation of these unobserved attributes seems likely to be very small and would be addressed by the approach of our instrumental variable estimates discussed above.

$$(12) \qquad Toll_{i,s,t} = \delta_{i,WE} + \delta_{i,MP} 1(MP_t) + \theta_{i,MP} \Delta TT_{i,s,t} 1(MP_t) + \theta_{i,WE} \Delta TT_{i,s,t} + \gamma_{i,MP} \Delta Reliab_{s,t} 1(MP_t) + \gamma_{i,WE} \Delta Reliab_{s,t} + \mu + \varsigma_{i,s,t}$$

where $1(MP_t)$ is an indicator for whether the time t when a trip taken is during the weekday morning peak and δ_i, θ_i and γ_i now vary between weekend and morning peak periods. The coefficient $\delta_{i,MP}$ measures the weekday morning peak urgency premium—how much extra punishment a failure to arrive on time may incur—over the weekend trip urgency premium. Unobserved features of the ExpressLanes (e.g., safety, comfort) that are always present are picked up by the parameter μ. To the extent that there is urgency during both time periods, the coefficient on this variable will be downwardly biased, and for this reason, we consider it a lower bound estimate of the value of urgency.

Accounting for Measurement Error.—Even if $\Delta TT_{i,s,t}$ is truly exogenous, it can be measured with error, leading to attenuation bias. Underestimates of the value of time and reliability would lead to an overestimate of the value of urgency. Measurement error in $\Delta TT_{i,s,t}$ could result from mismeasurement by the econometrician of drivers' perceived travel time savings or from misperceptions on the part of drivers. To address the first concern, in section V.E, we test alternative ways of calculating $\Delta TT_{i,s,t}$. To address driver misperceptions, we take two approaches: First, we take advantage of a natural experiment where drivers started to be presented with travel time savings at the ExpressLanes entry point. After October 20th, 2013, signs were set up at each ExpressLanes entrance indicating the travel time savings that drivers could expect based on real-time congestion data.

A second approach to addressing error in the measurement of $\Delta TT_{i,s,t}$ is to use instrumental variables (IVs). Following the approach of Aizer et al. (2018), we use moments in the distribution of the potentially mismeasured variable as an instrument since these are likely to be correlated with actual travel time perceptions. Our instrument set includes the average time savings one hour, one week and two weeks after the trip by hour of day, day of the week and road segment. We use leads, as opposed to lags, in travel time, because leads are likely to be highly correlated with contemporaneous time savings during a given hour, day of the week and road segment but unlikely to be affected by any unobserved contemporaneous factor affecting the driver when a particular trip is taken. More detail about the estimation of our IV model is provided in Appendix E.

V. Results

A. Estimates of the Value of Urgency, Value of Time, and Value of Reliability

Table 2 presents estimates of the values of urgency, time, and reliability that drivers reveal when paying the toll to enter the ExpressLanes.[53] In columns I and II, we first estimate a modified version of equation (11) where these values are assumed to be the same for all ExpressLanes users, using data pooled across the morning peak hours (5–9 AM). We term this the *homogeneous model*. In contrast, column III reports mean point estimates based on 9,054 different regressions of equation (11), one per individual who appears 5 or more times in the ExpressLanes during our sample period. We report the mean point estimate (averaged across all individuals and hours of the morning peak), bootstrapped standard errors, and interquartile range of the individually estimated coefficients. We term this the *heterogeneous model*. To provide a direct comparison between the homogeneous and heterogeneous models, column II reports the estimates of the homogeneous model for the same sample of individuals included in column III.[54]

Sources of Exogenous Variation.—While restrictive, the homogeneous model may be more representative of infrequent users since it includes individuals with fewer than five appearances during the study sample period.[55] Estimation of the homogeneous model exploits the cross-sectional variation in traffic conditions in the mainline lanes that pushes individuals into the ExpressLanes. In contrast, estimation of the heterogenous model in equation (11) exploits the within-individual time series variation in traffic conditions in the mainline lanes during the ten-month sample.[56] An intermediate approach could be to estimate the homogeneous model but include individual- or time-specific fixed effects that would capture heterogeneity in urgency. In Appendix Table G.12, we estimate such a model and show that the fixed effect estimates reflecting urgency show the bulk of the distribution to be close to our central estimates. One may, in principle,

[53] In our main specifications, we utilize realized travel time differences to measure time savings from the ExpressLanes in view of the fact that drivers are generally well informed about real-time driving conditions, as discussed in section III.D.

[54] As discussed in section IV.B, columns III-V report in parentheses the standard deviation of bootstrapped account-level regressions including segment-level clustering. The order of magnitude of these statistics is lower than that for the homogeneous model in columns I-II reflecting the fact that we are reporting a standard deviation across a distribution of coefficients rather than the standard error of a single coefficient pooling all accounts. Conceptually these statistics are different, reflecting the fact that the coefficient distribution has less variance. This also suggests that there is greater variance between accounts than for the same account over time.

[55] Of the 26,833 total users, 66% used the ExpressLanes fewer than five times during our entire study period, and therefore, we do not include them in the estimation of equation (11).

[56] The differences in coefficients between columns I and II could be the result of the reduction in sample size in the account-level regressions due to our decision to focus on accounts with 5 or more appearances in our sample data.

be concerned that certain sources of variation in demand, such as unobserved demand shocks, act as confounders. As discussed in section IV.C, any of these sources of variation act as plausibly exogenous changes in the pattern of tolls and time savings that help identify the distribution of preferences for ExpressLanes drivers.

Table 2 underscores the following key results. First, based on the homogeneous model in column I, the point estimates of the values of urgency, time, and reliability are, respectively, \$2.89 per trip, \$8.30 per hour, and \$22.67 per hour, and all statistically significant. These estimates of the value of time and reliability are broadly consistent with the estimates in prior studies (Small 2012), despite the fact that we are also able to account for the value of urgency at the same time. Second, a comparison of the results from the homogeneous models in columns I and II echoes our findings in Figure 2 that the distribution of WTP per hour is isomorphic to changes in ExpressLanes use. These regression results show that the value of urgency is not meaningfully different when we include (column I) or exclude (column II) infrequent ExpressLanes users.

Third, a comparison of the results from the homogeneous model in column II and the heterogeneous model in column III reveals that the mean point estimates of the heterogeneous model are broadly consistent with the estimates of the homogeneous model, with the mean point estimate for the value of urgency being only slightly higher at \$3.24 per trip, albeit not statistically different from \$2.89 based on a bootstrapped Hausman test of the coefficients. Taken together, these results suggest that differences in cross-sectional composition do not have discernable impacts on the mean of the parameter estimates. This does not imply that there is a single preference parameter structure for urgency, time savings, and reliability across individuals—only that compositional differences do not seem to meaningfully impact the mean point estimates.

Fourth, the table reveals a striking empirical finding—the value of urgency ranges between \$2.89 and \$3.24 in columns I–III of Table 2 and represents 77% to 81% of the average toll paid. In comparison, the contribution of time savings evaluated at its mean of 4.2 minutes is only \$0.57. Similarly, the contribution of reliability evaluated at its mean of 0.6 minutes is only \$0.18. The 0.6 minute (36 second) difference in reliability reflects the fact that the spread in lateness for the mainlines lanes is 0.6 minutes wider relative to the median than that for the ExpressLanes. The range of values of reliability can be seen in the bottom of Table 2. The combined contribution of the values of time and reliability to the average toll paid is no more than 19%.

The insight that the value of urgency represents the bulk of the willingness to pay to enter the

ExpressLanes has important implications for benefit-cost analysis of infrastructure projects. Typically, ex ante assessments of the benefit of ExpressLanes would predict the time saved by individuals from using ExpressLanes and then multiply this value by the estimate of the value of time over the program period studied, in our case, February 23rd, 2013, to December 31st, 2013. In this case, the projected benefits of the project would be \$241,399, which barely surpasses the infrastructure costs during that period of \$215,250.[57] In sharp contrast, the program *actually* generated \$1.31 million during that period, more than four times the ex ante benefits estimate of \$241,399.

By ignoring urgency, an ex-ante benefit–cost analysis of the project would underestimate the total benefits substantially—specifically, by an order of magnitude—during this time frame. Even with a value of time two or three times higher than the standard, the resulting benefit estimate would still be off by more than 100%.

Our estimates are comparable to those in the literature, which nonetheless tend to vary widely. Taking 2013 ACS incomes and converting them into hourly equivalent wages for zip codes registered to accounts in our sample, we find that the average hourly wage equivalent is \$19.63 as reported in Appendix Table G.3. This places our preferred value of time estimate at 41.7% of the wage and 89.7% for the value of reliability. The US Department of Transportation advises a VOT of 50% for evaluating transportation infrastructure projects (US DOT 2015). Small, Winston and Yan (2005) recover estimates of 93% and 83.8% for their median estimates of the VOT and VOR, respectively. There is consensus in the literature that commuters value their time 25-55% more under congested conditions (Small 2012), which could account for the difference relative to our estimates, however, this could also reflect an inability to disentangle the value of urgency from the value of time in past studies.[58]

B. ExpressLanes Market Conditions and Revealed Preferences

We now leverage variation over the AM peak in equilibrium tolls, travel times, and commuter decisions to decompose how quality-of-service pricing in the ExpressLanes reveals changes in the

[57] Source: Correspondence with LA Metro, 04/15/14. This corresponds to the operation and maintenance costs of the corridor including weekends and holidays across all hours of the day.

[58] The relative size of VOR to VOT is consistent with much of the literature, which tends to find estimates of the former are 1.5 to 2 times larger than the latter (Carrion and Levinson 2012). We recall, as described in section III.B, that our measure of reliability is somewhat different than that typically used in the literature. Traditionally the measure of reliability is constant over time and for the corridor. In our data, reliability varies by day, hour and segment since it is constructed using a rolling 30-day window by hour-segment. Given this additional variation, we are able to disentangle the value of reliability with urgency.

composition of preferences, as introduced in section III.E. In principle, the value of urgency could vary by individual, by time of day or both. However, since individuals likely sort into particular commute times, it seems plausible that individual and time variation in urgency is not independent. To explore this conjecture, we now examine the co-movement of tolls, travel demand and the recovered estimates of the value of urgency. A key conceptual question arises from these results: is urgency more specific to an individual or to a particular occasion? We explore this question using two sets of regressions in Table 3, those with account and hour-by-day of week fixed effects, which can be thought of as capturing individual- or time-specific urgency. The results show that there is more variation (as measured by the standard deviation of the fixed effects) across individuals ($0.99 per trip) than across hour-days of the week ($0.16).[59] These results point to the fact that the heterogeneity in the values of urgency and time over the AM peak is not usually statistically meaningful and is generally less pronounced than the heterogeneity by commuter, echoing the findings of Small, Winston and Yan (2005).[60]

Turning to Panel A of Figure 5, we plot the estimated values of urgency from separate regressions over five-minute intervals during the AM peak, showing that the recovered values of urgency drop after 8 AM. To explain this drop, we first note that this result corresponds to what one would expect to see on average based on a bottleneck model with urgency as elaborated in Appendix D: large discrete lateness costs move the mass of drivers to periods earlier than the desired arrival times, and the optimal toll falls substantially after periods corresponding to common desired arrival times. To further corroborate this interpretation, in Panel B of Figure 5, we plot the distribution of arrival times for employed workers in Los Angeles from ACS data. We can see from the plot that the bulk of the arrival times fall before 8 AM. In other words, tolls fall in the periods after the most common work start times even though utilization of the ExpressLanes remains high afterward, as shown in Panel A of Figure 3. Drivers exiting the lanes after 8 AM can be expected, on average, to already be incurring discrete lateness penalties or to not face them at all: hence, the recovered value of urgency in Panel A drops.

If lateness penalties are largely incurred for arrivals after 8 AM, why does ExpressLanes demand remain high? If we turn to Appendix Figure F.7, it becomes clear that after 8 AM, travel time

[59] These estimates also show that the value of the time during the 5–9 AM peak period is highest for drivers during the 7 AM hour, followed by drivers in the 6 AM, 8 AM and lastly 5 AM hours, although this heterogeneity is noteworthy only at 6 AM and the variation has no statistically meaningful effect on the value of urgency.

[60] This also echoes prior and more recent structural and theoretical work recovering heterogeneity in values of waiting time for ride-share services (Goldszmidt et al. 2020; Castillo 2019; Buchholz et al. 2020).

savings in the ExpressLanes remain relatively high. These higher time savings occur at times where, from Panel D of Figure 5, we can see that average hourly wages calculated from the ACS are rising, and so associated values of time are likely higher. Thus, after 8 AM, demand may be high because time savings and VOT are high.

C. *Identification Concerns*

We next present results from alternative approaches to addressing the identification concerns discussed in section IV.C. There, we discussed three potential threats to identification: simultaneity bias, measurement error, and omitted time-invariant unobservables. Given the design of the tolling algorithm described in section II, it is unlikely that our identifying assumption that $E\left[\varepsilon_{i,s,t}|\Delta TT_{i,s,t}\right] = 0$ is violated. This is because the tolls respond exclusively to the demand for the ExpressLanes to secure quality of service.

As a result, supply is, for all intents and purposes, fixed. If instead ExpressLanes speeds were to respond endogenously to the toll, our identifying assumption would be violated, since this would imply an elastic supply curve in Panel B of Figure 1 such that $toll_{i,s,t}$ and $\Delta TT_{i,s,t}$ are simultaneously determined. How large would the expected bias be? A simple back-of-the-envelope calculation suggests that such potential bias would have a trivial and statistically indistinguishable impact on the mean estimate of the value of urgency, lowering it from \$3.23 to \$3.13.[61]

Measurement Error in Travel Time Savings.—As discussed in section IV.C, there are two potential types of measurement error of time saved that may have implications for our estimates: error due to drivers' perception of the travel time saved and measurement error introduced by the researcher through the calculation of time savings. Regarding the former, in an ideal experiment, we would provide some drivers with additional information and test the value of this information. Fortuitously, LA Metro began including travel time differences on electronic signs that display tolls on October 20th, 2013, partway through our sample period. Restricting our sample to the subsequent period helps us determine the impact of the availability of salient, real-time congestion information on driver travel time savings perceptions.

[61] If we look down the deciles of Table 1, ExpressLanes speeds vary between 62 and 67.4 MPH, a difference of 5.4 MPH. The maximum effect of this variation on travel time savings over a 10.5 mile trip is 60 * 60 * (10.5 / 62) - (10.5 / 67.4) = 49 seconds. For the estimate of the value of urgency to be zero, travel time savings would have to be miscalculated by 24 minutes, so ExpressLanes speed variability would have an impact on the constant of (49 / 60) / 24 = 3.4%. We arrive at the value of 24 minutes by calculating the additional quantity of travel time savings at a value of time of \$8.19 per hour that would equate to the value of urgency in our main results, which is \$3.24.

In column IV of Table 2, we report mean point estimates of equation (11) while restricting the analysis to the subset of trips taken after October 20th, 2013. If nothing else, the mean of the point estimates, now at $3.62, suggest that the value of urgency only goes up when drivers have better information, indicating that measurement error is unlikely to be upwardly biasing our estimates.[62] Importantly, the value of reliability is now almost half that of column III, which may reflect the fact that with more precise information, uncertainty about the distribution of travel time differences, especially concerns over unexpected longer delays, is resolved. As a result, the overall contribution of urgency to the toll paid is now 90%. In section V.E, we present additional tests that drivers may be systematically making mistakes, and we find that such mistakes cannot explain the magnitude of our estimates of the value of urgency.

That said, systematic mismeasurement of what drivers perceive to be the time saved may still downwardly bias the coefficient on the value of travel time and therefore cause us to overstate the value of urgency. A standard approach to addressing concerns about attenuation bias from measurement error is to use IVs. As discussed in section IV.C, our instrument set consists of three variables: the average time savings one hour, one week and two weeks after the trip by hour of day, day of week and road segment. Our IV estimates using a heterogeneous model are reported in column V of Table 2 and do not meaningfully depart from the estimates of column III.[63] Below, in section V.E, we further explore additional potential measurement error that could arise through the researcher's calculation of travel time savings.

Time-Invariant Unobserved Attributes of ExpressLanes.—A separate concern may be that the constant captures not only the value of urgency but also any other time-invariant attribute in the ExpressLanes such as a perception of safety. To address this concern, we use trips taken during the weekend as a control group for the weekday AM peak trips following equation (12). If commuters experience no value of urgency during weekend trips, then an indicator variable for weekday morning peak trips would be an unbiased estimate of the value of urgency. However, to the extent that there is urgency during both time periods, the coefficient on this variable would be downwardly biased, and for this reason, we consider it a lower bound estimate of the value of urgency. In column I of Table 4, we present lower-bound estimates using the weekend as a control group of $2.41 on average without $\Delta Reliab_{s,t}$ and $2.44 with it.

[62] The difference between the coefficients in columns II and III is statistically significant when tested via a bootstrapped comparison of means.
[63] The first-stage estimates are reported in Appendix Table G.13. Alternative instrument sets are presented in columns II–V with second-stage estimates in Appendix Table G.14. Our preferred model with leads yields the largest first-stage Kleibergen–Paap F statistic: 2,342.

Sorting.—Another concern that is common to hedonic regression models is the extent to which preference heterogeneity results in sorting which biases estimates. In our setting, this would take the form of high-income individuals with high values of time potentially using shorter segments of the ExpressLanes. When we estimate individual preference parameters from the heterogenous model, we can examine their distribution for evidence of sorting. If these distributions were found to be pronouncedly asymmetric or bimodal, this might be evidence for sorting. As will be discussed in section V.D, we find these distributions to be symmetric. In Appendix Table G.15, we also show that the values of urgency and time are negatively correlated, suggesting that thigh value of time drivers tend to have lower values of urgency. Since this is the opposite of the pattern that the sorting described above would produce, we conclude that there is no clear evidence for sorting.

D. Heterogeneity in the Distributions of the Values of Urgency, Time, and Reliability

Heterogeneity Across Individuals.—A unique feature of our data is their repeated nature, which allows us to observe drivers in the ExpressLanes under different traffic congestion conditions. While estimating 9,054 individual-specific regressions based on equation (11), in columns III–IV of Table 2, we report in brackets the interquartile range of account-by-account estimates. Based on the point estimates of these individual bid-rent functions, in Figure 6, we also plot the distribution of the values of urgency, time, and reliability for the sample of repeated users of ExpressLanes. These underscore the substantial heterogeneity in the value of urgency, perhaps suggestive of underlying heterogeneity in commuters' underlying schedule constraints.[64] The value of urgency estimates for half of all account holders fall between \$2.30 and \$4.05. Importantly, the distributions of the value of urgency, time, and reliability in Figure 6 have a symmetric shape approximating a normal distribution, centered at mean point estimates relatively close to those found in the homogeneous model.[65]

Heterogeneity by Segment.—It is also conceivable that individuals with greater urgency might be more likely to start or end trips in particular locations (e.g., closer to downtown). In Appendix Table G.17, we estimate the homogeneous model by segments of the ExpressLanes. The results show that trips using longer stretches of the ExpressLanes and those closer to downtown (exit

[64] We also estimate the values of urgency and time during the afternoon peak and on other corridors of the ExpressLanes in Appendix Table G.16. We find that while there is heterogeneity, the qualitative results hold.

[65] It may seem puzzling that a subset of these accounts has negative valuations of reliability and a smaller subset of time and urgency, as can be seen in Figure 6. However, most of these negative estimates are not statistically different from zero. Moreover, heterogeneity in scheduling preferences may lead to variation in the value of reliability that we recover.

plazas WT05 and WT06) appear to have higher values of urgency. These results are consistent with those in Appendix Table G.6, which, as noted earlier, shows that drivers consume ExpressLanes more the later they expect to arrive at their destination.

Despite the substantial heterogeneity in the estimates of the values of time, reliability and urgency, our results underscore that the mean estimates from the heterogeneous model are largely indistinguishable from the estimates from the homogeneous model. In the absence of empirical evidence that a simpler model results in clear econometric bias, standard guidance suggests that a more parsimonious model is preferred. We argue that it is the simplicity of the homogeneous model that makes its application to recovering the primary welfare benefit of ExpressLanes most useful. We now turn to performing an assortment of robustness checks on the homogeneous agent model given its larger sample size and greater generalizability.

E. Robustness Checks

Nonparametric Hedonic.—First, theory does not guide the choice of functional form in hedonics (Cropper, Deck, and McConnell 1988; Bajari and Benkard 2005). Standard practice in hedonic models is to estimate an alternative version of equation (11) where the regressands enter nonparametrically. In columns I–II of Panel A of Table 5, we follow the approach proposed in Bajari and Benkard (2005) and estimate the hedonic price function using a local linear polynomial Epanechnikov kernel following Fan and Gijbels (1995). These results indicate that adding greater flexibility does not buy us substantially improved explanatory power or different coefficients.

Nonlinearity of Value of Time.—A separate concern is that equation (11) may be mis-specified if lateness penalties are highly nonlinear. In columns III-V of Table 5 and Appendix Table G.18, we estimate versions of the homogeneous model with higher-order terms for $\Delta TT_{i,s,t}$: quadratic, cubic, quartic, quintic and power models with and without a constant. Figure 7 summarizes these models in terms of their predicted values of WTP for time savings as measured by the toll. For the models with a constant, the results suggest that adding higher-order terms adds little additional explanatory power since the shape of the predicted WTP hardly changes, and the AIC and BIC also reflect this. In these nonlinear models, the point estimates for the value of urgency remain statistically indistinguishable from our central estimates from Table 2. In Panel B of Figure 7 and Appendix Table G.19, the models without the constant show greater variation, but this is because the model parameters must explain more variation in the absence of a constant. Many of these models result in predicted WTP that is lower or negative for longer travel time savings, which defies basic intuition and economic theory, suggesting that there is little reason to eschew the linear model used to derive our main results.

Accounting for Driver Mistakes.—Here, we explore additional tests for whether drivers are making mistakes in their perception of time savings. Ignoring reliability differences for the moment, only if all drivers systematically misjudged their time savings by 24 minutes, regardless of travel time savings that they actually experienced, would the value of urgency be reduced to zero.[66] In Panel B of Table 5, we account for potential mistakes that drivers may make in perceiving the travel time savings associated with the ExpressLanes. We do so by inflating the extent of time savings to determine whether driver misperception can explain the magnitude of the value of urgency that we recover. In column I, we double the observed time savings: the constant grows slightly larger and the value of time slightly smaller. In column II, we account for the fact that drivers may think the ExpressLanes segments to which they purchase access are longer than they actually are by increasing the length of each segment in the travel time calculation (difference in distance divided by speed) by 0.5 mile, and again, the constant is slightly larger and statistically significant.

In columns III and IV, we try to show how wrong drivers would have to be for the constant to be accounted for by mistakes, and while it is smaller, we still find a statistically significant constant of $1.66 when we add seven minutes to all travel time savings. It is only when we add fourteen minutes to all time savings that the constant becomes statistically insignificant. For drivers to misperceive time savings this radically seems outside the realm of possibility since, from Table 1, the average time savings is only 4 minutes and, even in the top time savings decile, it is only 11 minutes on average. This makes clear that if drivers are making mistakes, this can neither account for the value of urgency nor reduce it substantially.

Temporal and Spatial Autocorrelation.—Lastly, in Panel C of Table 5, we report estimates accounting for potential autocorrelation in the error term. Since the toll level in each period is adjusted from its level in preceding periods in response to changes in ExpressLanes demand, there is reason to believe that autocorrelation may be of concern. To account for this, in Appendix Table G.7, we calculate our standard errors using alternative cluster variables including two-way clustering by week and route following Cameron, Gelbach and Miller (2011). Our preferred cluster variable yields the largest standard errors and thus is the most conservative and hence preferred

[66] The explanation for the 24 minute calculation can be found in footnote 61.

approach. It may, however, be the case that autocorrelation occurs by other means, so in Panel C of Table 5, we report estimates where our standard errors are calculated as Conley (1999) standard errors that allow for correlation based on the distance between trip entry points (plazas) on the ExpressLanes, within a 1 km and 2 km spatial lag. We also vary the time lag by 5 and 10 minutes (columns I–III). Additionally, we allow the overlap between routes across road segments to dictate the form of autocorrelation. We account for this overlap in the calculation of our standard errors based on a dummy variable that indicates whether routes contain overlapping segments in column IV and how much distance those routes overlap for.[67] In all cases, our preferred estimates yield the largest standard errors, which is noteworthy since the estimates in columns IV and V use a substantially smaller sample.

Further Robustness Checks.—Beyond functional form, there may be other concerns over how we have chosen to define of $\Delta TT_{i,s,t}$. In the Appendix, we report further robustness checks on this variable and others and find no meaningful impact on our estimates of interest. First, one may be concerned that drivers respond to expected, not realized, travel time savings, so in Appendix Table G.20, we use past realizations of $\Delta TT_{i,s,t}$ instead. Alternatively, it may be that the mainline lane travel times do not offer the correct counterfactual, so instead, we use the I-210 W to construct a different measure of $\Delta TT_{i,s,t}$ in Appendix Table G.21 and find no meaningful difference in the estimates.

It could be that by restricting our sample to trips with positive $\Delta TT_{i,s,t}$ or nonnegative $\Delta Reliab_{s,t}$, we are biasing the results, so in Appendix Table G.10, we examine trips with negative values included. We also explore whether it is only small-$\Delta TT_{i,s,t}$ trips (e.g., those of less than 3 minutes) that drive our results in Appendix Table G.22 and find that they do not.

Alternatively, perhaps our choice of reliability measure biases the results, so in Appendix Table G.11, we use a variety of other moments and time windows to calculate the measure and find no meaningful effect. Last, to understand how determinants of travel demand affect our results, we control for gasoline prices and weather (Appendix Table G.23) and seasonality (Appendix Table G.24) and find that this has no meaningful impact on our results.

[67] For example, looking at Appendix Figure F.1, we can see that there is a route that uses segments 1, 2, and 3 and also a route that uses segments 2, 3, and 4. These routes therefore overlap across segments 2 and 3. Doing this requires us to collapse the data to the route–five-minute interval level; hence, the sample size decreases by an order of magnitude. The estimator allows the spatial (and temporal) lag to adjust, and in Appendix Table G.8, we report alternative choices of lags and alternative standard error calculations including Newey and West (1987) standard errors.

F. Implications for Road Pricing and Infrastructure Provision

Given the importance of urgency for ExpressLanes drivers, to what extent are these preferences as important in other settings, and how they might affect our understanding of optimal road pricing? When we add the value of urgency to the canonical model of dynamic pricing on a single tolled corridor (Vickrey 1969; Arnott, de Palma, and Lindsey 1993), two major differences appear, as shown formally in Appendix E. First, all that drivers can do to avoid higher lateness penalties is to leave earlier. As a result, with urgency, the rush hour period would shift earlier to reduce the mass of late drivers given this discrete penalty.

Second, under standard parametric assumptions, the optimal toll without urgency would be \$3.20.[68] If the underlying value of urgency (δ in equation (1)) is \$3, then for most of the rush hour (i.e., prior to a common desired arrival time), the optimal toll would be sixty cents higher at \$3.80. This increase of 19% seems surprisingly small given the relatively large magnitude of urgency in our findings. This discrepancy can be explained by several features of the ExpressLanes not embodied in a simple bottleneck model. One feature is that the simple bottleneck model does not incorporate stochastic congestion and therefore there is no uncertainty over travel times for quality-of-service pricing to help resolve. Accounting for this uncertainty can increase optimal tolls threefold, as shown in Hall and Savage (2019), and it seems plausible that urgency may also help account for this effect.[69]

A second feature is that a simple bottleneck model does not allow drivers to choose an untolled alternative such as the mainline lanes. Verhoef, Nijkamp and Rietveld (1996) derive the set of second-best tolls that maximize commuter welfare in a setting with an untolled alternative. Increasing these tolls from a low level reduces congestion and therefore improves welfare for drivers in the tolled lane but hurts those priced out of the tolled lane into congested untolled lanes. First-best tolling therefore dominates the second-best approach when drivers have homogeneous preferences. However, as shown in Verhoef and Small (2004), once a model accounts for heterogeneity in the value of time, the welfare benefits from differentiated pricing may be many

[68] These assumptions are that schedule delay late costs are four times schedule delay early costs, that the value of time is ten times the schedule delay early costs, that the value of time is \$10 per hour, and that the length of the rush hour is four hours. As mentioned earlier in section III.A, the standard approach in the transportation literature is to assume that the value of time is twice the schedule delay early costs, however Hall (2018b) demonstrates that doing so ignores the role of schedule constrained drivers which subsequently incorrectly predicts the pattern of equilibrium demand over the rush hour. Allowing for a mass of drivers with a value of time ten times the schedule delay early costs corrects this issue. We show that which assumption we use does not affect the resulting value of urgency from the model, but it does affect the optimal toll prediction, so we follow Hall (2018b), which results in an optimal toll closer in magnitude to what we observe on the ExpressLanes.

[69] In the previous example, this would correspond to a toll of \$9.60.

times higher and would likely imply higher optimal tolls. Adding a mass of drivers with high values of urgency would, as in the case with uncertainty, likely increase the optimal tolls substantially.

In contexts where drivers can choose between lanes, the difference in reliability is typically constructed as the difference in the spread between higher quantiles and the median of travel time savings. However, when we consider scheduling preferences, in addition to valuing being "less late," individuals also have preferences for on-time arrival. Markets that remove uncertainty over travel times, such as the ExpressLanes with their quality-of-service tolls, allow individuals to reveal their preferences under urgency, purchase the amount of travel time savings needed to arrive on time and avoid a discrete penalty for lateness.

A last question that arises is if the welfare benefits calculated here are credible, for whom are they calculated? In other words, can we expect to see comparable benefits from ExpressLanes programs if expanded to other locations of Los Angeles or in other cities? As travel demand and roadway capacity fluctuate over our sample, these result in toll and congestion levels that change the characteristics of the drivers in the ExpressLanes. For this reason, our estimates are less useful for a transfer of welfare benefits to other commuters on the highway who may not experience lateness to the same extent as ExpressLanes drivers. Instead, our results provide insights into the level of willingness-to-pay for drivers who are likely to be late, which may be a substantial number outside of the corridor we study. The range of estimates for the value of urgency recovered across varying times of day and corridor, between \$1.98 and \$3.58, seems consistent with our central estimate being appropriate for morning commuting along in the direction of a major employment center (downtown Los Angeles) based on prior studies. Considering the consequences of late arrival for important appointments, our central estimate of the value of urgency would seem plausible as a lower-bound estimate considering the cost of comparable convenience options such as short-term downtown parking lots and ridesharing services, such as Uber or Lyft.

VI. Conclusion

As noted in List (2004), the extent to which neoclassical microeconomic theory explains the *actual* behavior of economic agents in the marketplace remains under-researched. In an ideal setting where drivers are observed making *actual* choices about which lanes to use depending on varying tolls, the implied willingness to pay for travel time savings for relatively small travel time savings appear to be absurdly high when observed on a per-hour basis. To reconcile the existing literature on the VOT and scheduling (Small 1982; Small, Winston, and Yan 2005; R. Noland and Small 1995; Goldszmidt et al. 2020; Kreindler 2022) with the behavior of drivers in Californian ExpressLanes, in this study, we conceptualized and provided the first estimates of the value of urgency, defined as a discrete willingness to pay to meet a schedule constraint. Preferences for urgency result from the simple fact that, in their daily lives, individuals face penalties for being late and some of these penalties do not necessarily scale up with lateness. In the spirit of Muehlenbachs, Spiller, and Timmins (2015), Bishop and Timmins (2018) and Banzhaf (2020), we recovered individual-specific hedonic bid-rent functions and examined the heterogeneity in the underlying preference parameters for the values of urgency, time and reliability. An important insight is that the value of urgency, which has been neglected in the literature, is of first-order importance, implying that ex ante benefit–cost analysis that ignores this value may misguide road infrastructure investments.

Our findings have three broad implications. First, the welfare benefits associated with accounting for urgency can potentially alter the direction of cost–benefit analysis for infrastructure projects. Second, it is likely that amenities or externalities in contexts different from ours do not scale with quantity or that they have thresholds that are more relevant than just marginal improvements. For example, energy-efficient appliances for which WTP does not scale up with cost savings may create a large mass of inframarginal consumers (Boomhower and Davis 2014). Additionally, similar inframarginal behavior may be found for jurisdictions that need to comply with air quality standards relative to a threshold. As in this study, by ignoring the discrete nature of willingness to pay, the value of such amenities may be severely underestimated.

Third, our results suggest that there may be an economic justification for rethinking the structure of pricing to manage congested infrastructure. Economists' standard prescription is to consider Pigouvian tolls that fully correct for congestion externalities, and recent empirical work has built on this approach (Yang, Purevjav, and Li 2020; Bento, Hall, and Heilmann 2020). However, as noted in Peer et al. (2015), time savings may be more relevant for longer-term commuting decisions, whereas scheduling preferences, such as those that programs like ExpressLanes take advantage of, may be more important in the short run. Which of these dimensions is economically more important is ultimately an empirical question. However, given the political challenge of implementing anything approaching Pigouvian congestion pricing, differentiated pricing of a subset of the lanes of a freeway can offer two major benefits. First, its pricing mechanism can act as a filter on who enters the lane, which, in the presence of significant preference heterogeneity, can provide substantial benefits (van den Berg and Verhoef 2011; Hall 2018a). Second, the pricing mechanisms of ExpressLanes based on quality of service come with the additional advantage of resolving uncertainty over travel times. In this sense, ExpressLanes with quality-of-service pricing create new markets for drivers to avoid lateness and reveal their preferences for urgency. All in all, given the new technologies quickly disrupting how individuals interact in cities, the amount of real-time information increasingly available, and the opportunities for removing uncertainty in markets, we are only at the beginning of our understanding of the true value of urgency.

REFERENCES

Aizer, Anna, Janet Currie, Peter Simon, and Patrick Vivier. 2018. "Do Low Levels of Blood Lead Reduce Children's Future Test Scores?" *American Economic Journal: Applied Economics* 10 (1): 307–41.

Anderson, Michael L. 2014. "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion †." *American Economic Review* 104 (9): 2763–96. https://doi.org/10.1257/aer.104.9.2763.

Ang, Amanda, Peter Christensen, and Renato Vieira. 2020. "Should Congested Cities Reduce Their Speed Limits? Evidence from Brazil." *Journal of Public Economics* 184: 1041–55.

Angrist, Joshua D., and Alan B. Krueger. 1995. "Split-Sample Instrumental Variables Estimates of the Return to Schooling." *Journal of Business & Economic Statistics* 13 (2): 225–35.

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

Bajari, Patrick, and C. Lanier Benkard. 2005. "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach." *Journal of Political Economy* 113 (6): 1239–76. https://doi.org/10.1086/498586.

Bakkensen, Laura A, and Lint Barrage. 2022. "Going Underwater? Flood Risk Belief Heterogeneity and Coastal Home Price Dynamics." *The Review of Financial Studies* 35 (8): 3666–3709. https://doi.org/10.1093/rfs/hhab122.

Banzhaf, H. Spencer. 2020. "Panel Data Hedonics: Rosen's First Stage as a 'Sufficient Statistic.'" *International Economic Review* 61 (2): 973–1000. https://doi.org/10.1111/iere.12446.

Barber, Brad M., and Masako N. Darrough. 1996. "Product Reliability and Firm Value: The Experience of American and Japanese Automakers, 1973-1992." *Journal of Political Economy* 104 (5): 1084–99. https://doi.org/10.1086/262053.

Bartik, Timothy J. 1987. "Estimating Hedonic Demand Parameters with Single Market Data: The Problems Caused by Unobserved Tastes." *The Review of Economics and Statistics* 69 (1): 178–89.

Bates, John, John Polak, Peter Jones, and Andrew Cook. 2001. "The Valuation of Reliability for Personal Travel." *Transportation Research Part E: Logistics and Transportation Review* 37 (2): 191–229.

Baumol, William J. 1963. "An Expected Gain-Confidence Limit Criterion for Portfolio Selection." *Management Science* 10 (1): 174–82. https://doi.org/10.1287/mnsc.10.1.174.

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

———. 1993. "Nobel Lecture: The Economic Way of Looking at Behavior." *Journal of Political Economy* 101 (3): 385–409.

Benthem, Arthur van. 2015. "What Is the Optimal Speed Limit on Freeways?" *Journal of Public Economics* 124: 44–62.

Bento, Antonio, Jonathan Hall, and Kilian Heilmann. 2020. "Estimating Congestion Externalities Using Big Data." Working Paper.

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

Berg, Vincent van den, and Erik T. Verhoef. 2011. "Congestion Tolling in the Bottleneck Model with Heterogeneous Values of Time." *Transportation Research Part B: Methodological* 45 (1): 60–78. https://doi.org/10.1016/j.trb.2010.04.003.

Bin, Okmyung, and Craig E. Landry. 2013. "Changes in Implicit Flood Risk Premiums: Empirical Evidence from the Housing Market." *Journal of Environmental Economics and Management* 65 (3): 361–76. https://doi.org/10.1016/j.jeem.2012.12.002.

Bishop, Kelly C., Nicolai V. Kuminoff, H.Spencer Banzhaf, Kevin J. Boyle, Kathrine Gravenitz, Jaren C. Pope, V.Kerry Smith, and Christopher D. Timmins. 2019. "Best Practices in Using Hedonic Property Value Models for Welfare Measurement." *Review of Environmental Economics and Policy. Forthcoming*.

Bishop, Kelly C., and Alvin D. Murphy. 2011. "Estimating the Willingness to Pay to Avoid Violent Crime: A Dynamic Approach." *American Economic Review* 101 (3): 625–29. https://doi.org/10.1257/aer.101.3.625.

Bishop, Kelly C., and Christopher Timmins. 2018. "Using Panel Data to Easily Estimate Hedonic Demand Functions." *Journal of the Association of Environmental and Resource Economists* 5 (3): 517–43.

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics*, 577–99.

Boomhower, Judson, and Lucas W. Davis. 2014. "A Credible Approach for Measuring Inframarginal Participation in Energy Efficiency Programs." *Journal of Public Economics* 113 (May): 67–79. https://doi.org/10.1016/j.jpubeco.2014.03.009.

Borenstein, Severin, James Bushnell, and Erin Mansur. 2023. "The Economics of Electricity Reliability." *Journal of Economic Perspectives* 37 (4): 181–206. https://doi.org/10.1257/jep.37.4.181.

Brownstone, David, Arindam Ghosh, Thomas F. Golob, Camilla Kazimi, and Dirk Van Amelsfort. 2003. "Drivers' Willingness-to-Pay to Reduce Travel Time: Evidence from the San Diego I-15 Congestion Pricing Project." *Transportation Research Part A: Policy and Practice* 37 (4): 373–87.

Brownstone, David, and Kenneth A. Small. 2005. "Valuing Time and Reliability: Assessing the Evidence from Road Pricing Demonstrations." *Transportation Research Part A: Policy and Practice* 39 (4): 279–93. https://doi.org/10.1016/j.tra.2004.11.001.

Buchholz, Nicholas, Laura Doval, Jakub Kastl, Filip Matějka, and Tobias Salz. 2020. *The Value of Time: Evidence from Auctioned Cab Rides*. w27087. National Bureau of Economic Research.

Calfee, John, Clifford Winston, and Randolph Stempski. 2001. "Econometric Issues in Estimating Consumer Preferences from Stated Preference Data: A Case Study of the Value of Automobile Travel Time." *Review of Economics and Statistics* 83 (4): 699–707.

Cameron, A.Colin, Jonah B. Gelbach, and Douglas L. Miller. 2011. "Robust Inference with Multiway Clustering." *Journal of Business & Economic Statistics* 29 (2): 238–49.

Carrion, Carlos, and David Levinson. 2012. "Value of Travel Time Reliability: A Review of Current Evidence." *Transportation Research Part A: Policy and Practice* 46 (4): 720–41. https://doi.org/10.1016/j.tra.2012.01.003.

Carson, Richard T., and W.Michael Hanemann. 2005. "Contingent Valuation." *Handbook of Environmental Economics* 2: 821–936.

Castillo, Juan Camilo. 2019. "Who Benefits from Surge Pricing?"

Chao, Hung-po, and Robert Wilson. 1987. "Priority Service: Pricing, Investment, and Market

Organization.” *The American Economic Review*, 899–916.
Chay, Kenneth Y., and Michael Greenstone. 2005. “Does Air Quality Matter? Evidence from the Housing Market.” *Journal of Political Economy* 113 (2): 376–424. https://doi.org/10.1086/427462.
Chen, Chao. 2003. “Freeway Performance Measurement System (PeMS).” California PATH Research Report UCB-ITS-PRR-2003-22. Institute for Transportation Studies, University of California Berkeley. https://escholarship.org/content/qt6j93p90t/qt6j93p90t_noSplash_b3faa51086ad3c7d1a28cc92c23b920c.pdf?t=krnrxz.
Chetty, Raj. 2009. “Sufficient Statistics for Welfare Analysis: A Bridge between Structural and Reduced-Form Methods.” *Annual Review of Economics* 1: 451–88.
Christensen, Peter, and Adam Osman. 2020. “The Demand for Mobility: Evidence from an Experiment with Uber Riders.” In *IZA Discussion Paper Series # 14179*.
Colella, Fabrizio, Rafael Lalive, Seyhun Orcan Sakalli, and Mathias Thoenig. 2019. “Inference with Arbitrary Clustering.” SSRN Scholarly Paper. Rochester, NY. https://doi.org/10.2139/ssrn.3449578.
Conley, T. G. 1999. “GMM Estimation with Cross Sectional Dependence.” *Journal of Econometrics* 92 (1): 1–45. https://doi.org/10.1016/S0304-4076(98)00084-0.
Cropper, Maureen L., Leland B. Deck, and Kenenth E. McConnell. 1988. “On the Choice of Funtional Form for Hedonic Price Functions.” *The Review of Economics and Statistics* 70 (4): 668. https://doi.org/10.2307/1935831.
Davis, Lucas W. 2004. “The Effect of Health Risk on Housing Values: Evidence from a Cancer Cluster.” *American Economic Review* 94 (5): 1693–1704. https://doi.org/10.1257/0002828043052358.
De Vany, Arthur. 1976. “Uncertainty, Waiting Time, and Capacity Utilization: A Stochastic Theory of Product Quality.” *Journal of Political Economy* 84 (3): 523–41. https://doi.org/10.1086/260457.
DeSerpa, Allan C. 1971. “A Theory of the Economics of Time.” *Economic Journal*, 828–46.
Evans, Mary F., and Laura O. Taylor. 2020. “Using Revealed Preference Methods to Estimate the Value of Reduced Mortality Risk: Best Practice Recommendations for the Hedonic Wage Model.” *Review of Environmental Economics and Policy* 14 (2): 282–301. https://doi.org/10.1093/reep/reaa006.
Fan, Jianqing, and Irene Gijbels. 1995. “Data-Driven Bandwidth Selection in Local Polynomial Fitting: Variable Bandwidth and Spatial Adaptation.” *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (2): 371–94. https://doi.org/10.1111/j.2517-6161.1995.tb02034.x.
Figlio, David N., and Maurice E. Lucas. 2004. “What’s in a Grade? School Report Cards and the Housing Market.” *American Economic Review* 94 (3): 591–604.
Fosgerau, Mogens, and Anders Karlström. 2010. “The Value of Reliability.” *Transportation Research Part B: Methodological* 44 (1): 38–49. https://doi.org/10.1016/j.trb.2009.05.002.
Goldszmidt, Ariel, John A. List, Robert D. Metcalfe, V.Kerry Smith Ian Muir, and Jenny Wang. 2020. *The Value of Time in the United States: Estimates from Nationwide Natural Field Experiments*. w28208. National Bureau of Economic Research.
Greenwood, Jeremy, Nezih Guner, and Guillame Vandenbroucke. 2017. “Family Economics Writ Large.” *Journal of Economic Literature* 55 (4): 1346–1434.
Grossman, Michael. 1972. “On the Concept of Health Capital and the Demand for Health.”

Journal of Political Economy 80 (2): 223–55.
Hall, Jonathan D. 2018a. "Pareto Improvements from Lexus Lanes: The Effects of Pricing a Portion of the Lanes on Congested Highways." *Journal of Public Economics* 158: 113–25.
———. 2018b. "Pareto Improvements from Lexus Lanes: The Effects of Pricing a Portion of the Lanes on Congested Highways." *Journal of Public Economics* 158: 113–25.
Hall, Jonathan D., and Ian Savage. 2019. "Tolling Roads to Improve Reliability." *Journal of Urban Economics* 113: 103187.
Hallstrom, Daniel G., and V. Kerry Smith. 2005. "Market Responses to Hurricanes." *Journal of Environmental Economics and Management* 50 (3): 541–61. https://doi.org/10.1016/j.jeem.2005.05.002.
Harberger, Arnold C. 1964. "The Measurement of Waste." *The American Economic Review* 54 (3): 58–76.
Hausman, Catherine, and Samuel Stolper. 2021. "Inequality, Information Failures, and Air Pollution." *Journal of Environmental Economics and Management* 110 (October): 102552. https://doi.org/10.1016/j.jeem.2021.102552.
Hollander, Yaron. 2006. "Direct versus Indirect Models for the Effects of Unreliability." *Transportation Research Part A: Policy and Practice* 40 (9): 699–711. https://doi.org/10.1016/j.tra.2005.12.007.
Ito, Koichiro. 2014. "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing." *American Economic Review* 104 (2): 537–63.
Ito, Koichiro, Takanori Ida, and Makoto Tanaka. 2018. "Moral Suasion and Economic Incentives: Field Experimental Evidence and Energy Demand." *American Economic Journal: Economic Policy* 10 (1): 240–67.
Johnson, M.Bruce. 1966. "Travel Time and the Price of Leisure." *Western Economic Journal* 4 (2): 135–45.
Johnston, Robert J., Kevin J. Boyle, Wiktor Adamowicz, Jeff Bennett, Roy Brouwer, W.Michael Hanemann Trudy Ann Cameron, Nick Hanley, et al. 2017. "Contemporary Guidance for Stated Preference Studies." *Journal of the Association of Environmental and Resource Economists* 4 (2): 319–405.
Kask, S. B., and S. A. Maani. 1992. "Uncertainty, Information, and Hedonic Pricing." *Land Economics* 68 (2): 170–84. https://doi.org/10.2307/3146772.
Kling, Catherine L., Daniel J. Phaneuf, and Jinhua Zhao. 2012. "From Exxon to BP: Has Some Number Become Better than No Number?" *Journal of Economic Perspectives* 26 (4): 3–26.
Kreindler, Gabriel. 2022. "Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications." *Working Paper*.
Kuminoff, Nicolai V., and Jaren C. Pope. 2014. "Do 'Capitalization Effects' for Public Goods Reveal the Public's Willingness-to-Pay?" *International Economic Review* 55 (4): 1227–50.
Kuwayama, Yusuke, Sheila Olmstead, and Jiameng Zheng. 2022. "A More Comprehensive Estimate of the Value of Water Quality." *Journal of Public Economics* 207 (March): 104600. https://doi.org/10.1016/j.jpubeco.2022.104600.
Levy, H., and H. M. Markowitz. 1979. "Approximating Expected Utility by a Function of Mean and Variance." *The American Economic Review* 69 (3): 308–17.
Linden, Leigh, and Jonah E. Rockoff. 2008. "Estimates of the Impact of Crime Risk on Property Values from Megan's Laws." *American Economic Review* 98 (3): 1103–27.

https://doi.org/10.1257/aer.98.3.1103.
List, John. 2004. "Neoclassical Theory versus Prospect Theory: Evidence from the Marketplace." *Econometrica* 72 (2): 615–25.
Ma, Lala. 2019. "Learning in a Hedonic Framework: Valuing Brownfield Remediation." *International Economic Review* 60 (3): 1355–87. https://doi.org/10.1111/iere.12389.
McConnell, Kenneth E., Juha V. Siikamäki, and V.Kerry Smith. 2017. "A Strategy for Evaluating the Opportunity Cost of Time Estimates from New Choice Margins." *National Bureau of Economic Research*, no. w23349.
Muehlenbachs, Lucija, Elisheba Spiller, and Christopher Timmins. 2015. "The Housing Market Impacts of Shale Gas Development." *American Economic Review* 105 (12): 3633–59.
Newey, Whitney K., and Kenneth D. West. 1987. "Hypothesis Testing with Efficient Method of Moments Estimation." *International Economic Review* 28 (3): 777–87. https://doi.org/10.2307/2526578.
Noland, Robert B., and John W. Polak. 2002. "Travel Time Variability: A Review of Theoretical and Empirical Issues." *Transport Reviews* 22 (1): 39–54. https://doi.org/10.1080/01441640010022456.
Noland, Robert, and Kenneth A. Small. 1995. "Travel-time uncertainty, departure time choice, and the cost of morning commutes." *Transportation Research Record* 1493: 150–58.
Palmquist, Raymond B., Daniel J. Phaneuf, and V.Kerry Smith. 2010. "Short Run Constraints and the Increasing Marginal Value of Time in Recreation." *Environmental and Resource Economics* 46 (1): 19–41.
Peer, Stefanie, Erik Verhoef, Jasper Knockaert, Paul Koster, and Yin-Yen Tseng. 2015. "Long-Run versus Short-Run Perspectives on Consumer Scheduling: Evidence from a Revealed-Preference Experiment among Peak-Hour Road Commuters." *International Economic Review* 56 (1): 303–23.
Redfearn, Christian L. 2009. "How Informative Are Average Effects? Hedonic Regression and Amenity Capitalization in Complex Urban Housing Markets." *Regional Science and Urban Economics* 39 (3): 297–306. https://doi.org/10.1016/j.regsciurbeco.2008.11.001.
Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82 (1): 34–55. https://doi.org/10.1086/260169.
Saving, Thomas R, and Arthur De Vany. 1977. "Product Quality, Uncertainty, and Regulation: The Trucking Industry." *The American Economic Review* 67 (4): 583–94.
Schlee, Edward E., and V. Kerry Smith. 2019. "The Welfare Cost of Uncertainty in Policy Outcomes." *Journal of Environmental Economics and Management* 98 (November): 102264. https://doi.org/10.1016/j.jeem.2019.102264.
Small, Kenneth A. 1982. "The Scheduling of Consumer Activities: Work Trips." *American Economic Review* 72 (3): 467–79.
———. 2012. "Valuation of Travel Time." *Economics of Transportation* 1 (1–2): 2–14. https://doi.org/10.1016/j.ecotra.2012.09.002.
Small, Kenneth A, and Erik T. Verhoef. 2007. *The Economics of Urban Transportation*. Routledge Taylor & Francis Group.
Small, Kenneth A., Clifford Winston, and Jia Yan. 2005. "Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability." *Econometrica* 73 (4): 1367–82.
Smith, V.Kerry. 1981. "Congestion, travel cost recreational demand models, and benefit evaluation." *Journal of Environmental Economics and Management* 8 (1): 92–96.

Smith, V.Kerry, William H. Desvousges, and Matthew P. McGivney. 1983. “The Opportunity Cost of Travel Time in Recreation Demand Models.” *Land Economics* 59 (3): 259–78.

Steimetz, Seiji S. C. 2008. “Defensive Driving and the External Costs of Accidents and Travel Delays.” *Transportation Research Part B: Methodological* 42 (9): 703–24. https://doi.org/10.1016/j.trb.2008.01.007.

Train, Kenneth, and Daniel McFadden. 1978. “The Goods/Leisure Tradeoff and Disaggregate Work Trip Mode Choice Models.” *Transportation Research* 12 (5): 349–53. https://doi.org/10.1016/0041-1647(78)90011-4.

Train, Kenneth, and Wesley W. Wilson. 2008. “Estimation on Stated-Preference Experiments Constructed from Revealed-Preference Choices.” *Transportation Research Part B: Methodological* 42 (3): 191–203.

US DOT. 2015. “The Value of Time Savings: Departmental Guidance for Conducting Economic Evaluations.” US Department of Transportation. https://www.transportation.gov/administrations/office-policy/2015-value-travel-time-guidance.

Varian, Hal R. 1982. “The Nonparametric Approach to Demand Analysis.” *Econometrica* 50 (4): 945–73. https://doi.org/10.2307/1912771.

Verhoef, Erik, Peter Nijkamp, and Piet Rietveld. 1996. “Second-Best Congestion Pricing: The Case of an Untolled Alternative.” *Journal of Urban Economics* 40 (3): 279–302.

Verhoef, Erik T, and Kenneth A Small. 2004. “Product Differentiation on Roads.” *Journal of Transport Economics and Policy* 38: 30.

Vickrey, William S. 1969. “Congestion Theory and Transport Investment.” *American Economic Review* 59 (2): 251–60.

Wheaton, William C. 1977. “Income and Urban Residence: An Analysis of Consumer Demand for Location.” *The American Economic Review* 67 (4): 620–31.

Wichman, Casey J., and Brandon Cunningham. 2023. “Notching for Free: Do Cyclists Reveal the Opportunity Cost of Time?” *Journal of Environmental Economics and Management* 119 (May): 102817. https://doi.org/10.1016/j.jeem.2023.102817.

Yang, Jung, 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.

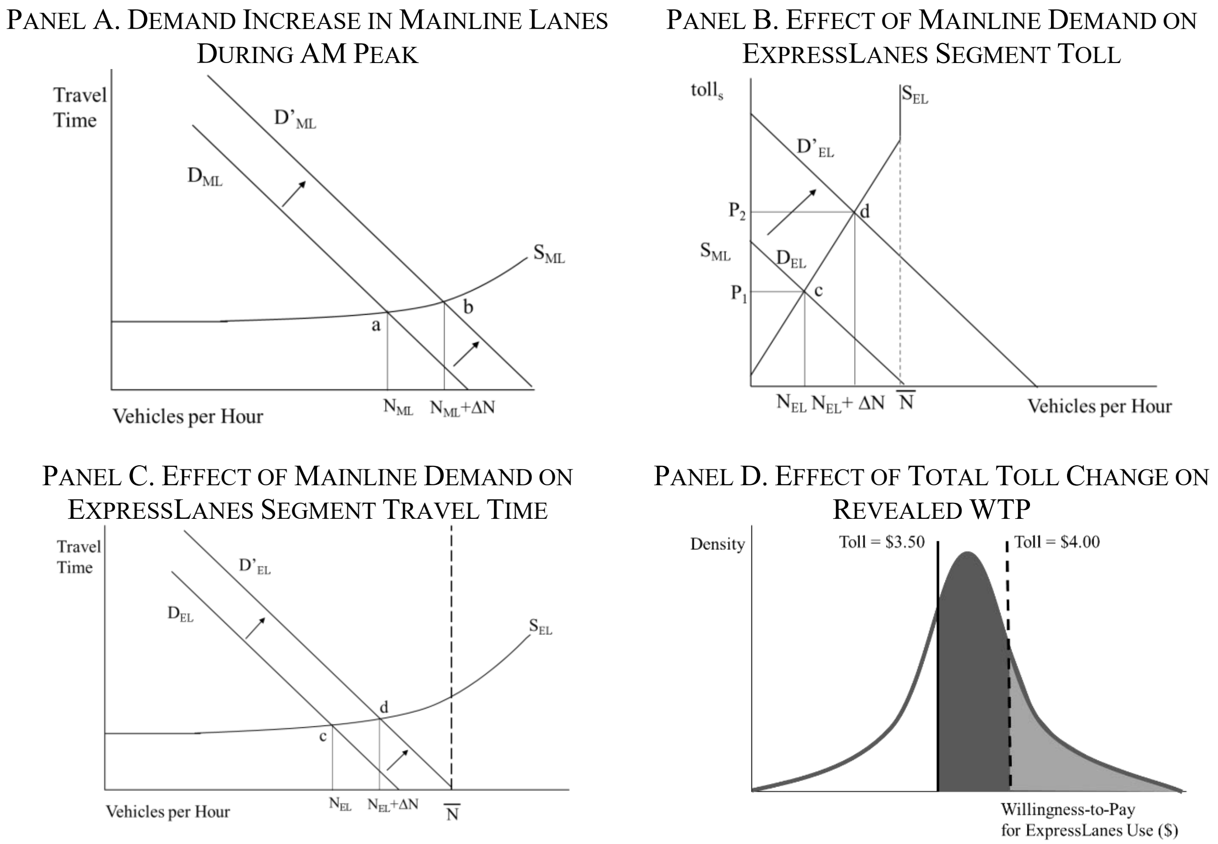


FIGURE 1. STRUCTURE OF PENALTY FOR LATENESS AND EFFECT OF MAINLINE DEMAND ON EXPRESSLANES AND WTP

Notes: The figures present the conceptual framework for understanding the theoretical basis of the penalty function and its link to willingness to pay for ExpressLanes Use. Commuting time saved and the penalty from late arrival, which are conceptually linked to the average WTP plotted in Figure 2. Panel A considers the effect of a regular rise in demand for trips in the mainline lanes as part of the peak of AM travel. Moving from point a to b, the travel times in the mainline lanes rise, but they remain constant in the ExpressLanes. Some portion of increased demand therefore enters ExpressLanes in response to the rise in travel time savings relative to the mainline lanes. Panel B demonstrates the effect on a single segment of the ExpressLanes. If they are operating near operational capacity (i.e., speeds are largely constant as shown in Table 1), then there is no impact on flow, but the segment-level ExpressLanes toll rises from P_1 to P_2. Panel C shows the same equilibria (c and d) in the ExpressLanes but with Travel Times rather than the Toll on the vertical axis. The increase in ExpressLanes demand induces a slight increase in Travel Times, but this is smaller than it might otherwise be because of the response of the toll from Panel B and the fact that the S_{EL} curve, which reflects the extent to which lane capacity constraints mean increased congestion, remains relatively flat. Once S_{EL} goes beyond $\overline{N}$, lane use corresponding to speeds at 45 MPH, travel times grow much faster. Panel D shows the effect of the toll increase from a single segment on the total toll for a hypothetical distribution of willingness-to-pay to use the ExpressLanes at a given time. As a result of a shock in any segment, the total toll increases. Here commuters with WTP greater than or equal to the toll enter the lanes, corresponding to the sum of the two grey areas. A toll increase as in Panel B reduces the mass of ExpressLanes users to the lighter colored grey area to the right of the new toll.

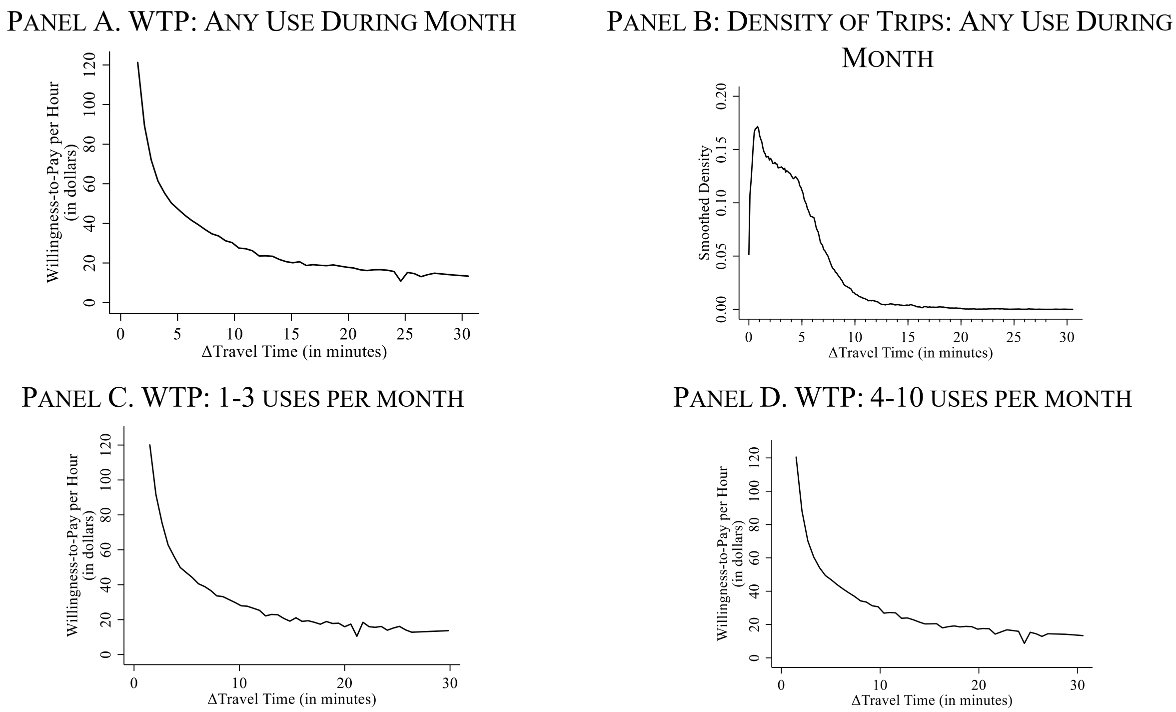
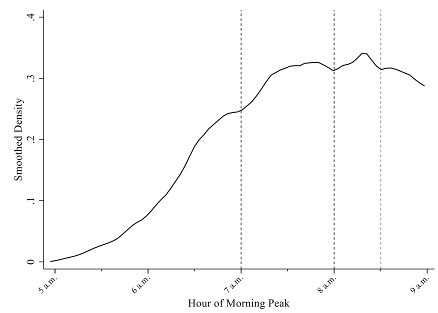


FIGURE 2. AVERAGE WTP PER HOUR AND DISTRIBUTION OF TRIPS IN THE EXPRESSLANES

Notes: Panel A displays the average willingness-to-pay (WTP) per hour for travel time savings in our sample the I-10 W ExpressLanes. The WTP per hour is calculated using kernel-weighted local polynomial smoothing for the ratio of the total toll paid for each trip over the travel time difference between the mainline lanes and the ExpressLanes. The vertical axis is truncated at \$120, although the actual values are much higher (See Table 1). Panel B displays the smoothed distribution of the trip-level travel time difference between the mainline lanes and the ExpressLanes. Panels C and D consider subsets of trips taken by individuals that appear in the ExpressLanes during the AM peak for the indicated number of uses during the same month. Panel A includes 65,526 trips, and Panel B 132,885. The smoother for all panels uses an Epanechnikov kernel with a bandwidth of 0.05. Travel times are calculated based on mainline speeds from PeMS and ExpressLanes time stamps and the actual distance traveled for each trip in the ExpressLanes. Both panels are generated using trip-level transponder data for the morning peak hours (5-9AM) of workdays in the first 10 months of the program, excluding holidays. Panel A considers (for illustrative purposes) only trips for travel time difference greater than 90 seconds, while panel B considers the entire travel time distribution. An unrestricted version of Panel A can be found in Appendix Table G.3. Trips with zero distance traveled and the 6.2% of observations with negative time saving are removed. Transponders registered to public sector, corporate or unknown individuals (2% of the entire sample) are dropped. Untolled HOV-3 trips (33% of sample) are removed. Observations from PeMS where any of the 30 second observations are missing are also dropped.

PANEL A. DISTRIBUTION OF TRIPS IN THE EXPRESSLANES BY HOUR OF THE MORNING PEAK



PANEL B. DISTRIBUTION OF TRIPS IN THE EXPRESSLANES FOR REPEATED USERS IN THE MORNING AND AFTERNOON PEAKS

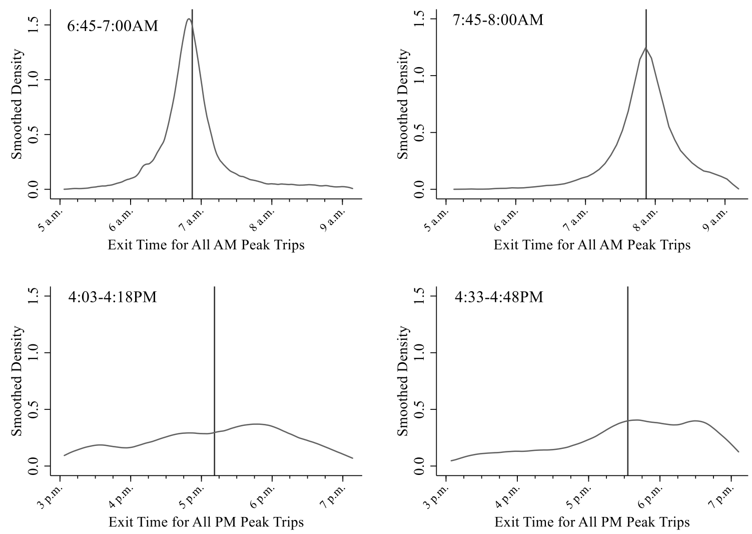


FIGURE 3. SCHEDULE CONSTRAINTS AND TRIP DENSITY DURING AM AND PM PEAKS

Notes: The figures plot the kernel smoothed density of trips on the I-10 W ExpressLanes over the morning and afternoon peaks. In Panel A, vertical lines correspond to times with a discernible trough in the distribution of trips, indicating potential “bunching” around preferred arrival times of 7:00AM, 8:00AM and 8:30AM at the exit point closest to downtown Los Angeles. These visible jumps occur 8-9 minutes before the hour for the 8:00AM and 8:30AM time windows, followed by a decline in trip density. In Panel B, we examine how tightly trip level exit time clusters to the individual-level average in the morning vs. afternoon peak. The figure selects individuals with average exit time within the 15-minute interval listed and then plots the kernel smoothed density of exit time for all trips associated with those individuals. The afternoon windows correspond to high demand times identified from a graph like Panel A but for the afternoon peak but provide less evidence for significant bunching around a common desired arrival time. Vertical lines indicate average exit time for trips in a given subsample. Only individuals with 10 or more ExpressLanes uses are included. Trips with zero distance traveled and the 6.2% of observations with negative time saving are removed. Transponders registered to public sector, corporate or unknown individuals (2% of the entire sample) are dropped. Untolled HOV-3 trips (33% of sample) are removed. Observations from PeMS where any of the 30 second observations are missing are also dropped.

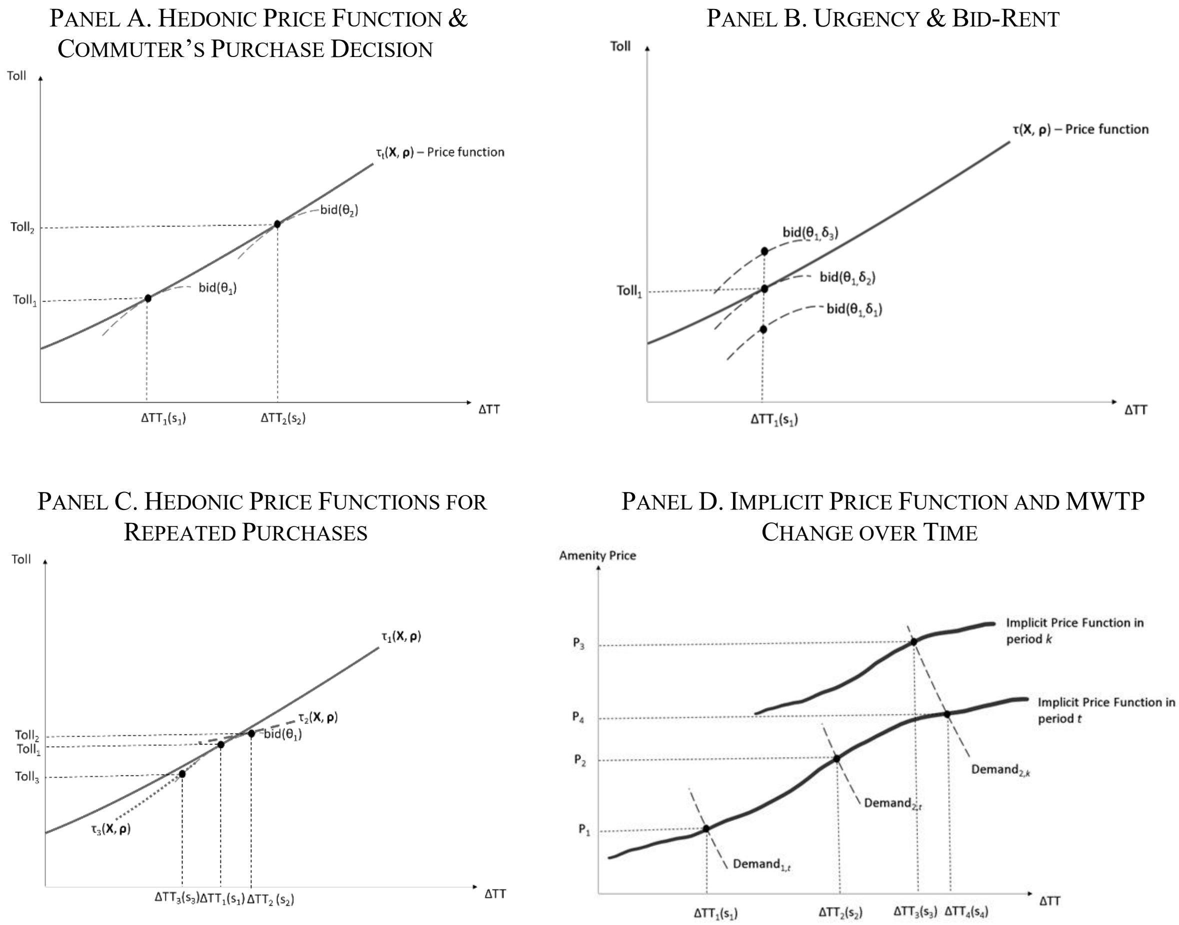
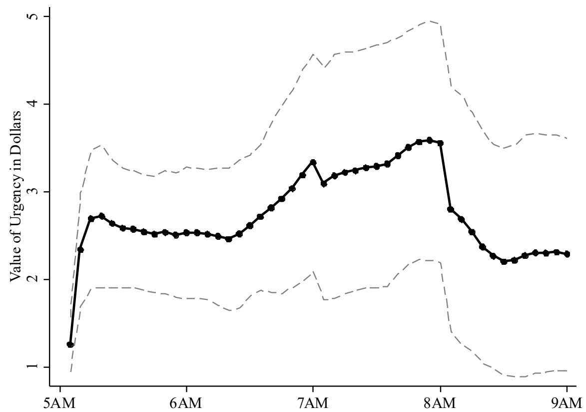


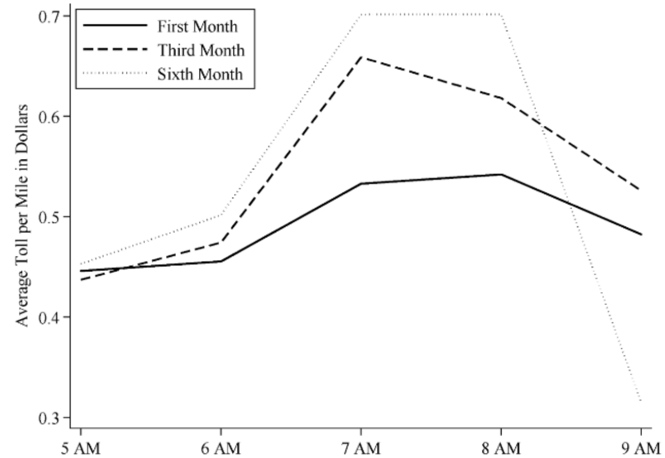
FIGURE 4. HOW THE HEDONIC PRICE FUNCTION REVEALS MWTP

Notes: Panel A shows the toll for the ExpressLanes increasing as the travel time savings relative to the mainline lanes increase. It also shows purchase decisions for two commuters. Commuter 2 purchases segments s_2, providing ΔTT_2 for total toll $Toll_2$ at the point where their bid curve lies tangent to the price function. Bid curves reflect WTP for the attribute and so are increasing and concave. This is also the point at which the commuter's demand curve intersects this amenity's implicit price function as shown in Panel D. Panel B shows how a commuter with the same value of time θ_1 can have different bid rent curves if their urgency varies. The commuter with urgency of δ_1 does not enter the ExpressLanes, with δ_2 enters and is marginal, and with δ_3 enters and is inframarginal. Panel C shows how the same commuter observed in the ExpressLanes on different days may face a different hedonic price function on each day (solid curves), and so the observed combinations of toll and ΔTT reveal the shape of their bid function. Panel D shows these same relationships with respect to the implicit price function, which is the derivative of the hedonic price function. Here again, we can see commuter 2 revealing different MWTP (equivalent to the Amenity Price in the figure) at different time periods because the implicit price function varies with the composition of demand.

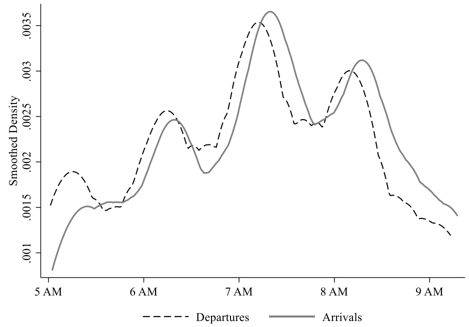
PANEL A. ESTIMATES OF THE VALUE OF URGENCY OVER THE AM PEAK



PANEL B. TOLL PER MILE BY HOUR & MONTH



PANEL C. DEPARTURE AND ARRIVAL TIMES FOR WORK, LA COUNTY, ACS



PANEL D. AVERAGE HOURLY WAGE BY ARRIVAL TIME TO WORK LA COUNTY, ACS

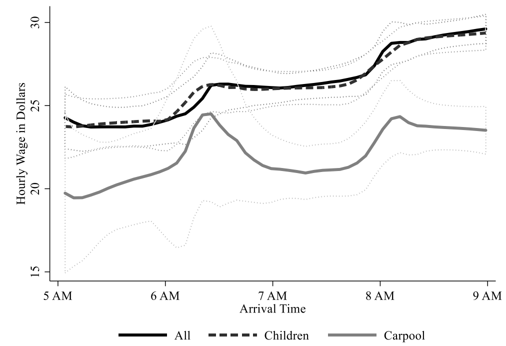
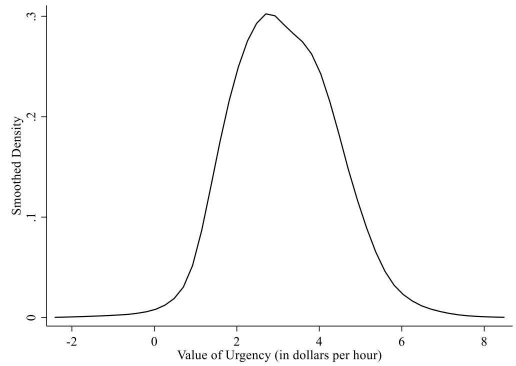
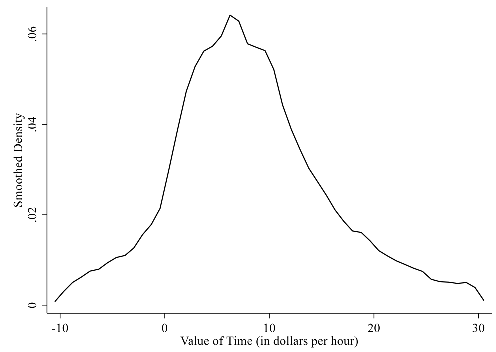


FIGURE 5. EVOLUTION OF THE DISTRIBUTION OF VALUES OF URGENCY OVER THE AM PEAK

Notes: Panel A plots estimates of the value of urgency in a homogeneous agent model from a single regression of the toll paid on ΔTravel Time, ΔReliability and dummy variables for each five-minute interval of the AM peak during workdays in the I-10 W ExpressLanes. The regressions do not include a constant. Each dot represents the coefficient for a five-minute interval dummy corresponding to the value of urgency for that time period. The dashed lines reflect the 95% confidence intervals for each value of urgency estimate based on standard errors clustered by road segment. Panel B displays the average hourly toll per mile in dollars paid during the morning peak (5-9 AM) for drivers on the I-10 W ExpressLanes during the first month, (February 25th, 2013 – March 31st, 2013), the third month (May 2013) and the sixth month (August 2013) of the program. Panel C plots the distribution of departure times for and arrival times at work as reported by employed workers that commute by private vehicle as reported in the 2013 American Community Survey (ACS). Panel D shows the kernel-weighted, locally smoothed estimates of hourly wages for commuters by the reported usual arrival time over the AM peak at work from the 2013 ACS for employed Los Angeles County residents earning more than $20,000 per year and driving to work. Hourly wages are based on reported annual wage income and assume 2,040 hours worked per year.

PANEL A. VALUE OF URGENCY DISTRIBUTION PANEL B. VALUE OF TIME DISTRIBUTION





PANEL B. VALUE OF RELIABILITY DISTRIBUTION

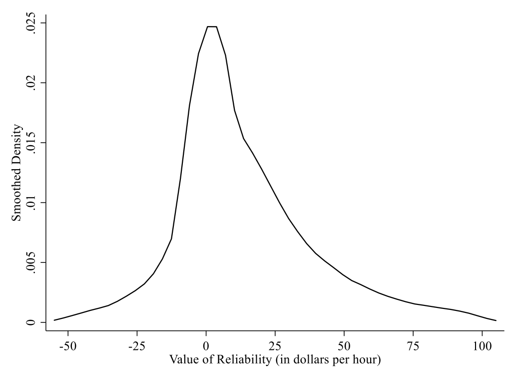
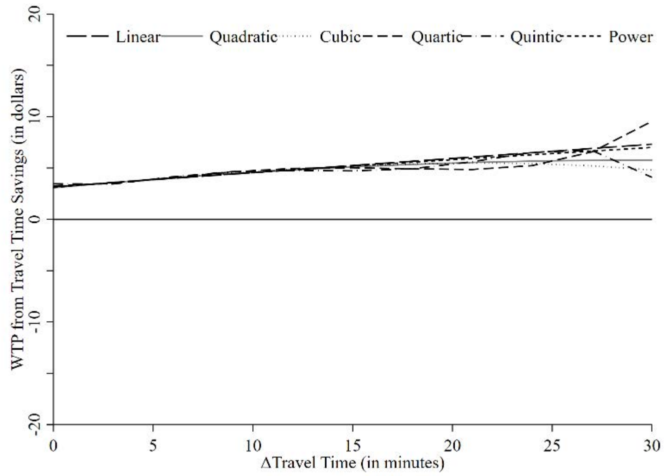


FIGURE 6. ESTIMATED DISTRIBUTION OF VALUES OF URGENCY, TIME, AND RELIABILITY

Notes: The figures depict smoothed kernel density estimates of the value of urgency, time and reliability from individual-specific hedonic regressions of the toll paid on a constant, travel time saved, and reliability following the heterogeneous individual bid curve model in equation (11). Reliability is the difference in the spread of the 80th and 50th quantiles of travel time savings between the mainline and ExpressLanes. The figures are restricted to values of urgency between -2 and 8 (99.6% of sample), values of time between -10 and 30 (90.1% of sample) and values of reliability between -50 and 100 (86.4% of sample). Only individuals with 5 or more I-10 W ExpressLanes uses are included. Reliability less than 0.01 hours (36 seconds) is set to zero in individual-specific regressions. Trips with zero distance traveled and the 6.2% of observations with negative time saving are removed. Transponders registered to public sector, corporate or unknown individuals (2% of the entire sample) are dropped. Untolled HOV-3 trips (33% of sample) are removed.

PANEL A: MODELS WITH A CONSTANT



PANEL B: MODELS WITHOUT A CONSTANT

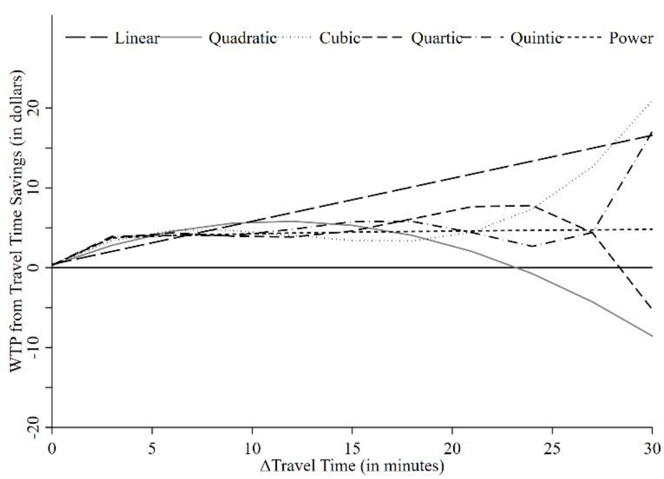


FIGURE 7. MODEL PREDICTED WTP FROM ALTERNATIVE FUNCTIONAL FORMS

Notes: The figure plots predicted WTP for variation in time savings between 0 and 30 minutes based on estimates from 12 different regressions in a homogeneous model during the morning peak (5-9 AM) on workdays in the I-10 W ExpressLanes of the toll paid on ΔTravel Time, ΔReliability and a constant. These correspond to the models estimated in Appendix Tables G.18 and G.19. Models vary by the polynomial order (1st through 5th) or as a power function, where all other terms enter linearly, but travel time savings include a multiplicative parameter and an exponent: $\beta_1(\Delta Travel\ Time)^{\beta_2}$. The regressions in Panel A include a constant while those in Panel B do not. Time, measured in hours, is the time saved by taking the ExpressLanes compared with mainline lanes, from mainline speeds reported by PeMS, for the chosen trip distance. ΔReliability, measured in hours, is the difference between lanes in the spread of travel times between the 80th and 50th quantiles. ΔReliability less than 0.01 hours (36 seconds) is set to zero. Trips with zero distance traveled and the 6.2% of observations with negative time saving are removed. Transponders registered to public sector, corporate or unknown individuals (2% of the entire sample) are dropped. Untolled HOV-3 trips (33% of sample) are removed. Observations from PeMS where any of the 30 second observations are missing are also dropped.

Table 1—Trip-Level Summary Statistics by Decile of Travel Time Savings

I	II	III	IV	V	VI
ΔTravel Time (decile)	ΔTravel Time in Minutes	Average Mainline Speed (in MPH)	Average ExpressLanes Speed (in MPH)	Average Distance Traveled (in Miles)	Average Toll Paid
1	0.39	60.3	65.3	5.8	$3.20
2	1.01	55.9	67.4	6.1	$3.10
3	1.66	50.0	66.6	6.2	$3.12
4	2.37	44.7	66.1	6.1	$3.17
5	3.11	40.6	66.0	6.1	$3.29
6	3.88	37.7	65.8	6.3	$3.57
7	4.69	34.6	65.5	6.3	$3.81
8	5.64	32.7	64.7	6.7	$4.15
9	6.95	30.9	63.8	7.3	$4.49
10	11.04	25.8	62.0	8.1	$4.95
Average	4.08	41.3	65.3	6.5	$3.69

Notes: Data cover workdays for the morning peak (5-9 AM) from February 25th, 2013 until December 30th, 2013 in the I-10 W ExpressLanes. ΔTravel Time is mainline travel time measured using PeMS speeds minus ExpressLanes travel time. Trips with zero distance traveled and the 6.2% of observations with negative time saving are removed. Transponders registered to public sector, corporate or unknown individuals (2% of the entire sample) are dropped. Untolled HOV-3 trips (33% of sample) are removed. Observations from PeMS where any of the 30 second observations are missing are also dropped. Each decile for the full time period contains 46,624 trips, for February and March contains 3,261 trips, for June contains 4,615 trips and for September contains 7,001 trips.

Table 2—Homogeneous Agent and Individual-Level Hedonic Price Function Estimates

	I	II	III	IV	V
Constant	2.89***	3.10***	3.24***	3.62***	2.98***
	(0.48)	(0.54)	(0.01)	(0.03)	(0.01)
			[2.30, 4.05]	[2.52, 4.65]	[2.13, 3.69]
ΔTravel Time	8.30**	7.99**	8.19***	7.16***	14.57***
	(2.88)	(2.83)	(0.12)	(0.27)	(0.22)
			[3.10, 12.79]	[-0.68, 14.32]	[7.74, 21.02]
ΔReliability	22.67***	20.79***	17.61***	9.41***	29.01***
	(4.61)	(5.69)	(0.38)	(0.62)	(0.66)
			[-1.43, 30.79]	[-7.59, 25.83]	[-0.16, 50.66]
≥ 5 uses of ExpressLanes		X	X	X	X
Heterogeneous Agent Model			X	X	X
After 10/20/13				X	
Split-Sample IV					X
Mean Toll Paid	$3.73	$3.98	$3.98	$4.02	$3.66
VOU % of WTP	77%	78%	81%	90%	81%
Mean ΔTravel Time (in Min.)	4.2	4.2	4.2	4.8	4.2
Mean ΔReliab. (in Min.)	0.6	0.6	0.6	1.20	0.6
Individuals	26,833	9,054	9,054	2,286	4,891

Notes: Values shown are the coefficients of regressions of the toll paid on the regressands. Columns I and II report a single regression of the toll on the travel time difference, while the remaining columns report parameter estimates from separate individual-level regressions from equation (11). Column II restricts the sample to match Column III. Column IV reports estimates for the period after signs displaying travel time differences were installed at ExpressLanes entrances. Column V reports estimates of individual-level split-sample instrumental variables using 1 hour, 1 week and 2 week leads of time savings by hour, day of week and segment following the approach of Angrist and Krueger (1995). Values shown in columns III-V are the average coefficient across regressions, with the interquartile range of values given in brackets. Data cover workdays in the I-10 W ExpressLanes during the morning peak (5-9AM). ΔTravel Time, measured in hours, is the time saved by taking the ExpressLanes compared with mainline lanes, from mainline speeds reported by PeMS, for the chosen trip distance. ΔReliability, measured in hours, is the difference between lanes in the spread of travel times between the 80^{th} and 50^{th} quantiles. ΔReliability less than 0.01 hours (36 seconds) is set to zero in individual-specific regressions. Trips with zero distance traveled and the 6.2% of observations with negative time saving are removed. Transponders registered to public sector, corporate or unknown individuals (2% of the entire sample) are dropped. Untolled HOV-3 trips (33% of sample) are removed. Observations from PeMS where any of the 30 second observations are missing are also dropped. Standard errors in columns I-II are clustered by road segment and are presented in parentheses. These are qualitatively different from the numbers in parentheses in columns III-V, which are the standard deviation of the coefficient mean, bootstrapped from a normal distribution from 500 iterations and are indicated in parentheses.
*** Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 3—Homogeneous Hedonic Price Function Estimates: Fixed Effects Models

	I	II	III	IV
Constant	2.887***	3.055***	2.733***	2.693***
	(0.481)	(0.005)	(0.016)	(0.018)
ΔTravel Time	8.345**	7.849***	7.393***	7.627***
	(2.868)	(0.061)	(0.058)	(0.061)
ΔReliability	22.742***	11.664***	15.929***	16.101***
	(4.624)	(0.195)	(0.207)	(0.205)
R^2	0.22	0.17	0.22	0.23
Observations	426,761	426,761	426,761	426,761
Implied Value of Urgency from Individual Fixed Effects in $ per Trip				
25th Quantile		2.13	1.80	1.76
Mean		3.06	2.73	2.69
75th Quantile		3.98	3.68	3.65
sd(Individual FEs)			0.99	0.99
sd(Time FEs)			0.38	0.16
Individual FE		X	X	X
Hour FE			X	X
Hour-DOW FE				X

Notes: This table is based on a model that assumes a common value of time and reliability across the sample, but an individual-specific value of urgency, using within variation to explain the latter and between to explain the former. Time-fixed effects (by hour or hour-by-day-of-week) also control for time-specific urgency across trip appearances in the sample. Values shown are the coefficients of 4 separate regressions of the toll paid during the AM peak period (5 – 9 AM) on the regressands. Travel Time, measured in hours, is the time saved by taking the ExpressLanes compared with mainline lanes, from mainline speeds reported by PeMS, for the chosen trip distance. Reliability, measured in hours, is the difference between lanes in the spread of travel times between the 80th and 50th quantiles. Three rows report the mean, 25th and 75th quantiles of the individual-level fixed effects corresponding to varying estimates of the value of urgency across individuals in our sample. Data cover workdays during the morning peak (5-9AM). Sd(.) reports the standard deviation of individual and time fixed effects. Reliability less than 0.01 hours (36 seconds) is set to zero. Trips with zero distance traveled and the 6.2% of observations with negative time saving are removed. Transponders registered to public sector, corporate or unknown accounts (2% of the entire sample) are dropped. Untolled HOV-3 trips (33% of sample) are removed. Observations from PeMS where any of the 30 second observations are missing are also dropped. Robust standard errors are in parentheses.

*** Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 4—Individual-Level Hedonic Price Function Estimates: Weekend Control Group

	I	II	III
	Full Sample	Full Sample	After 10/20/13
Constant	0.82***	0.90***	0.53***
	(0.01)	(0.02)	(0.04)
	[0.50, 1.05]	[0.49, 1.06]	[0.26, 0.69]
1(Morning Peak)	2.41***	2.44***	3.53***
	(0.02)	(0.03)	(0.12)
	[1.44, 3.30]	[1.48, 3.38]	[2.13, 4.67]
ΔTravel Time	3.31***	3.39***	5.32***
	(0.08)	(0.11)	(0.34)
	[1.27, 5.17]	[1.17, 5.26]	[2.85, 7.31]
ΔTravel Time x 1(Morning Peak)	7.24***	5.38***	1.48***
	(0.24)	(0.32)	(0.86)
	[1.58, 11.15]	[0.56, 9.24]	[-6.24, 8.04]
ΔReliability		2.58***	2.74***
		(0.46)	(0.44)
		[-4.69, 5.76]	[-0.04, 3.90]
ΔReliability x 1(Morning Peak)		12.41***	-0.33
		(0.85)	(1.97)
		[-4.50, 28.82]	[-15.72, 17.07]
Individuals	1,122	758	64

Notes: Values shown are the coefficients of regressions of the toll paid on the regressands following equation (9). Data cover workdays during the morning peak (5-9AM) and all day during weekends in the I-10 W ExpressLanes. Column II reports estimates for the period during which signs displaying travel time differences were installed at ExpressLanes entrances. Values shown are the average coefficient across regressions, with the interquartile range of values given in brackets. ΔTravel Time, measured in hours, is the time saved by taking the ExpressLanes compared with mainline lanes, from mainline speeds reported by PeMS, for the chosen trip distance. ΔReliability, measured in hours, is the difference between lanes in the spread of travel times between the 80th and 50th quantiles. Only individuals with 10 or more ExpressLanes uses are included. ΔReliability less than 0.01 hours (36 seconds) is set to zero in individual-specific regressions. Trips with zero distance traveled and the 6.2% of observations with negative time saving are removed. Transponders registered to public sector, corporate or unknown individuals (2% of the entire sample) are dropped. Untolled HOV-3 trips (33% of sample) are removed. Observations from PeMS where any of the 30 second observations are missing are also dropped. Standard errors presented in parentheses in all regressions are the standard deviation of the coefficient mean, bootstrapped from a normal distribution from 500 iterations.

*** Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 5—Hedonic Price Function Estimates: Robustness Checks

	I	II	III	VI	V
Panel A.	*Non-Param.*			*Functional Form Robustness*	
Constant	3.526***	3.998***	2.84***	3.00***	3.19***
	(0.006)	(0.007)	(0.31)	(0.30)	(0.30)
ΔTravel Time	5.599***	7.906***	9.27	-1.68	-18.61
	(0.151)	(0.360)	(13.63)	(15.64)	(13.55)
ΔTravel Time^2			3.89	162.27	514.40***
			(74.65)	(121.56)	(124.99)
ΔTravel Time^3			-31.31	-752.22*	-3,405.54***
			(110.98)	(372.86)	(610.69)
ΔTravel Time^4				969.18**	8,898.96***
				(389.22)	(1383.62)
ΔTravel Time^5					-7,973.01***
					(1147.69)
ΔReliability	61.743***	17.854***	22.19***	21.78***	21.59***
	(0.561)	(0.315)	(5.34)	(5.22)	(5.34)
AIC			1,511,412	1,509,235	1,506,824
Observations	51,590	51,381	433,623	433,623	433,623
	May 2013	*November 2013*			
Panel B. Mistakes in Travel Time Savings					
Constant	2.94***	2.97***	1.66***	0.36	2.39***
	(0.50)	(0.50)	(0.47)	(0.67)	(0.44)
ΔTravel Time	5.53***	9.70***	11.05***	11.05***	12.24***
	(1.51)	(2.85)	(3.03)	(3.03)	(1.26)
R^2	0.15	0.13	0.15	0.15	0.37
Observations	466,232	466,232	466,232	466,232	433,623
	Multiply time saved by 2	*Each segment is 0.5 miles longer than measured*	*Add 7 minutes to ALL time savings*	*Add 14 minutes to ALL time savings*	*Max. EL to min. ML speeds*
Panel C. Autocorrelation					
Constant	2.89***	2.89***	2.89***	2.89***	2.89***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
ΔTravel Time	8.30***	8.30***	8.30***	7.92***	7.92***
	(0.22)	(0.22)	(0.22)	(0.26)	(0.22)
ΔReliability	22.67***	22.67***	22.67***	23.76***	23.76***
	(0.63)	(0.64)	(0.63)	(0.81)	(0.59)
R^2	0.22	0.22	0.22	0.17	0.17
Observations	433,227	433,227	433,227	42,074	42,074
Correlation	Entry Plaza	Entry Plaza	Entry Plaza	Overlap	Distance
Spatial Lag	1 km	2 km	1 km	1 KM	1 KM
Time Lag	5 min	5 min	10 min	5 min	5 min

Notes: Values shown are the coefficients of 15 separate regressions of the toll paid on the regressands. Columns I-II of Panel A present estimates of our main hedonic regression model estimated using a nonparametric local-linear kernel regression following Cattaneo and Jansen (2017). The estimator uses an Epanechnikov kernel where the optimal bandwidth is chosen by minimizing the integrated means squared error. Columns III-V of panel A include higher order terms of time saved. In Panel B, column I, we replace the travel time saved with twice its realized value to reflect the fact that underestimating the time saved inflates the value of time. Column II assumes that each segment is half a mile longer than recorded, possibly because of transition zones. Columns III and IV add 7 and 14 minutes of time saved to all trips suggesting that only by uniformly adding travel time will the constant go to zero. Column V of Panel B compares the maximum speed in the ExpressLanes for the preceding month by hour and day of week to the minimum speed in the mainline lanes from PeMS detectors for the preceding month by hour and day of week. Panel C presents heteroskedasticity and autocorrelation consistent estimation of standard errors. In Panel C, columns I-III present Conley (1999) corrected standard errors based on the latitude and longitude of the entry plaza to the ExpressLanes. Columns IV-V allow for temporal and spatial autocorrelation where the latter accounts for overlap of ExpressLanes segments across routes either using dummy variables ("Overlap") or the length in miles of the overlap ("Distance") following Colella, et al.(2019). Travel Time, measured in hours, is the time saved by taking the ExpressLanes compared with mainline lanes, from mainline speeds reported by PeMS, for the chosen trip distance. Reliability, measured in hours, is the difference between lanes in the spread of travel times between the 80th and 50th quantiles. Data cover workdays during the morning peak (5-9AM). Reliability less than 0.01 hours (36 seconds) is set to zero. Trips with zero distance traveled and the 6.2% of observations with negative time saving are removed. Observations from PeMS where any of the 30 second observations are missing are also dropped. *** Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.