

NBER WORKING PAPER SERIES

AVOIDING TRAFFIC CONGESTION EXTERNALITIES? THE VALUE OF URGENCY

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

The authors thank Lawrence Goulder, Edson Severnini, Lowell Taylor, and Hendrik Wolff for helpful comments as well as seminar participants at AERE 2015, Arizona State University, Carnegie Mellon University, Cornell University, John Hopkins University, Kraks Fond Managing Congestion Seminar, NBER Summer Institute, Simon Fraser University, Stanford University, University of California at Davis, University of California at Irvine, University of Maryland, Penn State University, and the University of Texas at Austin, WCERE 2018, ZEW. Finally, we thank Stephanie Wiggins and Kathleen McCune at LA Metro for access to ExpressLanes transaction data and Arthur T. DeGaetano for 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.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by Antonio Bento, Kevin Roth, and Andrew R. Waxman. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Avoiding Traffic Congestion Externalities? The Value of Urgency

Antonio Bento, Kevin Roth, and Andrew R. Waxman

NBER Working Paper No. 26956

April 2020

JEL No. D47,D61,D62,H23,L51,L91,Q58,R41

**ABSTRACT**

In Becker (1965) and neoclassical microeconomic theory the value of time is a constant fraction of the hourly wage. When taken to data, however, this value departs from theoretical predictions, and appears to vary with the amount of time saved. By observing drivers on freeways opting to enter toll lanes with high-frequency, time-varying prices that secure a minimum level-of-service, we uncover a new and fundamental aspect of preferences for travel time savings related to urgency. The presence of preferences for urgency, which reflect the fact that individuals often face discrete penalties for being late, allows us to reconcile the pattern observed in the data with neoclassical theory. Using a rich, repeated-transaction data and individual-level hedonic estimation, we show that the value of urgency accounts for 87 percent of total willingness-to-pay for time savings. As a result, ignoring the value of urgency in cost-benefit analysis severely underestimates the true value of time savings that projects deliver, as such omission will typically ignore non-trivial welfare gains to a potentially large number of individuals.

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

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

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

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

# Avoiding Traffic Congestion Externalities? The Value of Urgency

By ANTONIO BENTO, KEVIN ROTH, ANDREW WAXMAN\*

*In Becker (1965) and neoclassical microeconomic theory the value of time is a constant fraction of the hourly wage. When taken to data, however, this value departs from theoretical predictions, and appears to vary with the amount of time saved. By observing drivers on freeways opting to enter toll lanes with high-frequency, time-varying prices that secure a minimum level-of-service, we uncover a new and fundamental aspect of preferences for travel time savings related to urgency. The presence of preferences for urgency, which reflects the fact that individuals often face discrete penalties for being late, allows us to reconcile the pattern observed in the data with neoclassical theory. Using a rich, repeated-transaction data and individual-level hedonic estimation, we show that the value of urgency accounts for 87 percent of total willingness-to-pay for time savings. As a result, ignoring the value of urgency in cost-benefit analysis severely underestimates the true value of time savings that projects deliver, as such omission will typically ignore non-trivial welfare gains to a potentially large number of individuals.*

*JEL Codes: D47, D61, D62, H23, L51, L91, Q58, R41*

\* Bento: University of Southern California and NBER (email: abento@usc.edu); Roth: Laurits R. Christensen Associates (email: kd Roth@lrca.com); Waxman: University of Texas at Austin (email: awaxman@utexas.edu). The authors thank Lawrence Goulder, Edson Severnini, Lowell Taylor, and Hendrik Wolff for helpful comments as well as seminar participants at AERE 2015, Arizona State University, Carnegie Mellon University, Cornell University, John Hopkins University, Kraks Fond Managing Congestion Seminar, NBER Summer Institute, Simon Fraser University, Stanford University, University of California at Davis, University of California at Irvine, University of Maryland, Penn State University, and the University of Texas at Austin, WCERE 2018, ZEW. Finally, we thank Stephanie Wiggins and Kathleen McCune at LA Metro for access to ExpressLanes transaction data and Arthur T. DeGaetano for weather data.

## I. Introduction

Since Becker (1965), the value of time has been defined as the opportunity cost of time, and it is a fraction of an individual's hourly wage (Johnson, 1966; DeSerpa, 1971; Gronau, 1973). The value of time is also a key parameter used in cost-benefit analysis and project evaluation, including in areas such as recreational demand (Smith, 1981; Bockstael and McConnell, 1981; Phaneuf, Kling and Herriges, 2000), health improvements (Grossman, 1972; Murphy and Topel, 2006), road infrastructure, pricing and the value of public transportation (Keeler and Small, 1977; Parry and Small, 2005; Anderson, 2014; Yang, Purevjav, and Li, 2020), and of a variety of other government regulations, including those related to the use of seatbelts and speed limits (Ashenfelter and Greenstone, 2004, Wolff, 2014; van Benthem, 2015, Fisher and Gallagher, 2020).

A direct consequence of Becker's (1965) framework is that the value of time is constant. Does this theoretical conjecture hold empirically when taken to the marketplace? Panel A of Figure 1 plots the kernel-smoothed estimates of the willingness-to-pay (WTP) per hour against travel time saved for drivers that use ExpressLanes in California. ExpressLanes allow for the possibility of entering a faster lane in a freeway that secures a minimum speed, upon the payment of a toll that varies almost in real-time.<sup>1</sup> The Becker framework suggests that this estimate of the WTP for time savings per hour should be a flat line, and equal to a fraction of the hourly wage. In sharp contrast, data from the 'marketplace' reveals a surprising hyperbolic shape. Consistent with estimates found in the literature, when

<sup>1</sup> In Panel A of Figure 1, the WTP per hour is calculated as the ratio of the toll paid and travel time saved (by taking the trip in the ExpressLanes as opposed to an unpriced regular lane in the freeway). For presentation purposes, the figure is truncated at \$120 per hour but continues to substantially higher values. Appendix Figure B.1 shows the evolution of average toll per mile over the AM peak during the first, third and sixth month of our sample, demonstrating considerable variation

individuals save substantial amounts of time, this WTP per hour tends towards fifty percent of the hourly wage, roughly \$10 per hour in the Los Angeles metropolitan area (Small, 2012). But as the time saved decreases, the value of the WTP per hour increases dramatically. Panel B of Figure 1 shows that these observations with unusually high WTP for time savings are not outliers – rather, they form the bulk of all uses of ExpressLanes. Seen on a per-hour basis, this WTP appears to be absurd, and many orders of magnitude above widely accepted estimates of the value of time (Small, 2012). For example, for the 10% of observations with time savings less than 0.39 minutes, the implied WTP per hour, is \$1,977.44, nearly 200 times the standard estimate of the value of time.<sup>2</sup>

For the shape of the WTP per hour displayed in Panel A of Figure 1 to be plausible, either the valuation of time changes with the amount of time saved, which is inconsistent with neoclassical microeconomic theory, or the simple Becker (1965) framework for recovering the value of time misses important aspects of the structure of individuals' preferences. Implicit in neoclassical microeconomic theory is the assumption that all time is fungible. There are at least two potential competing explanations that suggest otherwise: First, much of the recent behavioral economics literature has demonstrated that not all dollars are fungible (Thaler and Shefrin, 1981; Hastings and Shapiro, 2013; Sahm, Shapiro, and Slemrod, 2012), and time should be no different than money (Leclerc, Schmitt, and Dubé, 1995). Second, the presence of schedule constraints and resulting penalties for late arrival may also make the valuation of time depart from a constant, and depend on the context of trips (Small, 2012).

<sup>2</sup> Appendix Table C.1 shows the distribution of average willingness-to-pay per hour in the data by decile of travel time savings.

The purpose of this paper is twofold: First, we aim to provide a potentially plausible explanation for the observed pattern of the willingness to pay for travel time savings observed in Panel A of Figure 1, and to reconcile this pattern with the classical framework proposed by Becker (1965). Second, we aim to learn about the structure of individual preferences, and to decompose the determinants of their willingness to pay for time savings, while inferring their relative importance.

We demonstrate that the bulk of the willingness to pay for travel time savings when accessing ExpressLanes comes from what we term the *value of urgency*, as opposed to the value of time or the value of reliability widely studied in the literature (Small, 2012; Small, Winston and Yan, 2005; Brownstone and Small, 2005; Calfee and Winston, 2001; Calfee and Winston, 1998). As a consequence, when schedule constraints are prevalent, failure to account for the value of urgency in cost-benefit analysis severely underestimates the true value of time savings that projects deliver. As we will document, such omission will typically ignore non-trivial welfare gains to a potentially large number of drivers.

We define the value of urgency as a discrete WTP to meet a schedule constraint, recognizing that individuals often face penalties for being late that do not necessarily scale up with the amount time that they are late. In other words, the value of urgency reflects the valuation that drivers place in avoiding a penalty for failing to meet a schedule constraint. For example, individuals face large penalties when failing to arrive at the airport on time and, as a consequence, missing their flights. Intuitively, in this case, saving one minute has a substantially different value depending on whether the individual misses the plane or not. The value of that minute needed to meet the schedule constraint can be substantial. Similarly, because the structure of payments for daycare or lawyers' appointments are typically made on an hourly basis, individuals pay for the hour whether they arrive on time or late. Therefore, when faced with important schedule constraints and penalties, individuals are likely willing to pay a substantial toll to enter the

ExpressLanes as they help to ensure on-time arrival. Relative to an unpriced lane of a freeway, the payment of this toll effectively allows drivers to ‘jump a queue’ and purchase the discrete amount of time needed to meet the schedule constraint and avoid a penalty.<sup>3</sup>

That individuals face schedule constraints is easy to appreciate and well documented in the literature. Prior studies have attempted to estimate the value of early, late and on-time arrivals (Small, 2012; Noland and Small, 1995). In fact, Small (1982) notes that: “it can also be shown that the valuation of time savings will vary by time of day depending on the interaction between scheduling considerations and time-varying price or service quality” (p. 467). Still, we are the first to recover estimates for the *value of urgency* for at least four reasons.

First, most earlier studies that recover drivers’ preferences for travel time savings, including the value of time and reliability, rely on stated preference surveys, sometimes combined with revealed preference data covering limited time periods, usually just a few months (Small, Winston and Yan, 2005; Train and Wilson, 2008; Brownstone and Small, 2005). Stated preference surveys typically solicit drivers’ WTP for travel time savings from choosing tolled over untolled lanes on an ‘average day.’ The fundamental problem with these surveys is that 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.<sup>4</sup>

<sup>3</sup> The value of urgency does not emerge exclusively in the context of transportation. Yet, it seems to have been largely overlooked in the economics literature. Other examples where individuals may exhibit preferences for urgency include organ donation, where individuals would likely reveal the value of urgency to pay for an organ that would save their life, especially the closer they are to organ failure that threatens survival. Premium payments for faster processing of visas and passports also provide another mechanism for individuals to reveal preferences for urgency.

<sup>4</sup> More broadly Calfee, Winston and Stempki (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 and natural resource economics considers the

Second, even if a survey could recover an individuals' likelihood of being late and the value of avoiding lateness to meet a schedule constraint, one should be concerned whether respondents of stated preference surveys would be able to precisely reveal valuations for rather small travel time savings like those documented in Panel A of Figure 1.<sup>5</sup> We overcome these two fundamental survey design issues and data limitations of earlier studies by relying on a purely revealed preference setting where we observe the *actual* repeated behavior of drivers in the ExpressLanes with almost real-time varying tolls and different levels of congestion.

Third, most prior studies that examined the value of tolled lanes have relied on theoretical models often supplemented with simulations (Verhoef, et al., 1996; Light, 2009; Hall, 2018; Hall, 2020; Hall and Savage, 2019) or econometric studies of toll lanes on separate routes with limited variation in prices (Small, Winston and Yan, 2005; Brownstone and Small, 2005).<sup>6</sup> In both cases, tolls differ, at best, during peak and off-peak periods leaving aggregate demand, and resulting levels of congestion, unknown. Our concern is that, with such structure of tolls, drivers may not be able to take advantage of tolled lanes to meet schedule constraints.

The Los Angeles ExpressLanes program is novel because its tolls vary in almost real-time, and importantly are set to assure that speeds never fall below a minimum threshold. This toll structure resembles Vickrey's (1971) quality-of-service pricing (which is analogous to level-of-service pricing in transportation), and the approach laid out in De Vany (1976), De Vany and Saving (1977), Chao and Wilson (1987).

limitations and challenges of stated preferences approaches (Kling, Phaneuf and Zhao, 2012; Johnston, et al., 2017; Carson and Hanemann, 2005).

<sup>5</sup> Indeed, many surveys usually cap travel time savings to at least 10 minutes, making it easier for respondents to mentally process time savings and their valuations, but precluding valuation of short time savings which may be more common.

<sup>6</sup> ExpressLanes and other forms of high-occupancy tolling (HOT), called "managed lanes" by transportation planners, have been used in 11 states in the US in major metropolitan areas such as Atlanta, Minneapolis-St Paul, Washington, DC and the San Francisco Bay area. The programs vary in their design and implementation with some being administered by private contractor while others, like the LA ExpressLanes, managed by a metropolitan authority. These types of lanes have been studied more recently by Janson and Levinson (2014) and Brent and Gross (2016).



This type of optimal pricing design offers a menu of contingent contracts for the distribution of scarce space on the road. In addition, by securing a minimum level of speed, it allows drivers to assess whether the time savings from entering the ExpressLanes are sufficient to recover lost travel time, and still meet a schedule constraint. The flexibility of the price mechanism ensures that in instances where the value of urgency outweighs the cost of the toll, individuals are able to obtain a guaranteed minimum speed. At the level of the market, the price ensures that drivers with the highest willingness-to-pay effectively can move to the front of the traffic queue, which offers the potential for large welfare gains.

Fourth, earlier work using discrete choice models has recovered drivers' preferences, including the value of time and reliability, even under peak and off-peak varying prices (Small, Winston and Yan, 2005; Brownstone and Small, 2005). Recovering the value of urgency with these models would require explicitly modeling discrete penalty functions, which, for the same individual, are likely to vary depending on the circumstances of a particular trip. Moreover, time-varying prices are an endogenous function of contemporaneous congestion and so identifying their impact on demand at an individual level poses a substantial econometric challenge (Timmins and Murdock, 2007). Therefore, instead of explicitly modeling individual schedule constraints, our strategy is to rely on the equilibrium identifying assumptions of a hedonic-style price function (Rosen, 1974; Roback, 1982). Given the repeat-transaction nature of our data, with the same drivers observed with varying prices and congestion levels, we are able to recover lower-bound estimates of the value of urgency and this first-stage hedonic price regression constitutes a 'sufficient statistic' for welfare measurement (Banzhaf, 2019a). In fact, as noted in Bishop and Timmins (2018), who estimate the marginal willingness to pay (MWTP) for air quality improvements, observing individuals

repeatedly is sufficient to recover a linear demand function separately for each individual, with no need for additional functional form assumptions.<sup>7</sup>

Hedonic estimation, nevertheless, presents its own methodological challenges (Bishop, et al., 2019). However, given the nature of our data, we are able to address all these classical challenges, including potential confounding due to measurement error of key attributes, omitted variables, and functional form misspecification. We are also able to assess the extent to which drivers sorting by attributes in their use of the ExpressLanes biases our estimates and find no evidence of sorting.

Our baseline estimate of the value of urgency is \$3.24, with half of all account holders' estimates of the value of urgency falling between \$2.30 and \$4.05. The estimate of the value of time is \$8.19 per hour, only slightly lower than half the local hourly wage, and the estimate of value of reliability is \$17.61 per hour, consistent with the literature (Small, 2012). With an average time-varying toll of \$3.71 and an average travel time savings of 3.79 minutes, this implies that the value of urgency represents 87% of the average toll and, as a consequence, is the first-order missing parameter in a cost-benefit analysis of road infrastructure, such as ExpressLanes, with level-of-service real-time pricing.

<sup>7</sup> With the exception of Bishop and Timmins (2018), most other hedonic studies of housing markets rely on repeated sales of the same house, rather than repeated purchases by the same individual. Without this individual-level variation in prices and amenities, relatively strong assumptions need to be made so that the estimates of the MWTP can be used for welfare measurement (Abbott and Klaiber, 2011; Bajari, Fruehwirth and Timmins, 2012; Currie, Davis and Greenstone, 2015).

## II. Program Background and Data

### A. *The ExpressLanes Program*

On February 23<sup>rd</sup>, 2013, Los Angeles converted the High Occupancy Vehicle (HOV) lanes on the I-10 into a High Occupancy Toll (HOT) facility, as part of the ExpressLanes program.<sup>8</sup> The goal of the program was to increase the total throughput of these roads and to raise funds to maintain the corridors.<sup>9</sup> Maximum throughput is maintained along the ExpressLanes through a level-of-service system that adjusts prices every five minutes. The policy is designed to minimize costs to pre-existent carpoolers, who are ensured free-flow conditions by the mandated minimum speed of 45 mph and the continued ability to use the ExpressLanes free of charge.<sup>10</sup> Drivers may enter or exit the ExpressLanes at 6 separately-priced locations along the I-10 W.<sup>11</sup> At these entry points drivers see posted toll rates, ranging between \$0.55 and \$14.70 in our sample, and once a vehicle enters the lane the corresponding toll rate is locked in for the duration of its trip even if the price for subsequent vehicles changes.<sup>12</sup>

Compared with fixed- or peak-toll lanes, the ExpressLanes program has two unique features that present a unique opportunity for individuals to make a last-

<sup>8</sup> This was the second such conversion in Los Angeles, the first being the I-110 ExpressLanes, which opened on November 10<sup>th</sup>, 2012. We limit the scope of our study from the pre-policy expansion of the HOV lanes on December 1<sup>st</sup>, 2012 to December 31<sup>st</sup>, 2013. More details on the timing of ExpressLanes implementation can be found in Appendix A.

<sup>9</sup> The program opened the lanes to Single Occupant Vehicles (SOV) who were charged a per-mile toll ranging from \$0.10 to \$15.00, debited from a FasTrak<sup>®</sup> account linked to a required transponder in the vehicle. The ExpressLanes function such that once the maximum price is reached the lane is closed to further SOV traffic. The lane was never closed during the period considered on the I-10 W.

<sup>10</sup> Carpools are required to use a transponder but are not charged when it is set to HOV 3+ during peak times or HOV 2 during off-peak hours.

<sup>11</sup> These exit and entry points are indicated by arrows in Appendix Figure B.2.

<sup>12</sup> Between entry points the ExpressLanes are separated from the mainline lanes by a solid double white lane marker that drivers may not cross. Crossing this marker is a moving violation. The program funds cameras at entry and exit points that read license plates to toll vehicles without transponders and California Highway Patrol officers that patrol the road segment.

minute choice that offers the potential for on-time arrival in the face of a schedule constraint. First, the ExpressLanes adjust prices to maintain a minimum 45 mph speed. While other toll lanes may, generally, provide faster travel than an untolled alternative, the ExpressLanes can guarantee congestion-free driving. Second the ExpressLanes allow drivers multiple points of entry and exit. This allows them flexibility to change their decisions based on conditions and consume exactly the amount of distance desired. This is often only a few miles—considerably shorter distances than other toll roads where drivers must commit to a decision and are unable to opt out of the lane if conditions improve.<sup>13</sup> In practice, these two features help drivers solve much of the uncertainty in travel time, allowing them to purchase the time needed to meet schedule constraints with a high level of precision.

### *B. Data*

Our empirical demonstration of the value of urgency is conducted with a unique confidential dataset that reveals how individuals trade time for money. Our data combine transponder-level travel information of all ExpressLanes trips provided by Los Angeles Metropolitan Transportation Authority (Metro) with travel times observed in mainline lanes from the California Department of Transportation’s (CALTRANS) Freeway Performance Measurement System (PeMS) traffic detector data. This section describes each data source and presents some key summary statistics related to the value of urgency.

<sup>13</sup> Varied subsegment use by drivers is substantial, with a large proportion of observed trips exiting either at mid-way points along the corridor or at the end as documented in Appendix Table C.2.

*Sample Composition.*—The dataset of ExpressLanes trips along the I-10 westbound from LA Metro allows us to observe individual trips associated with the same transponder account.<sup>14</sup> Our data include information on times, points of entry and exit, the toll charged, the primary vehicle registered to it and the zip code for the billing address.<sup>15</sup> Our full dataset contains 982,056 observations on this route spanning the period from February 22<sup>nd</sup>, 2013 to December 31<sup>st</sup>, 2013.<sup>16</sup>

We focus on the 466,232 trips that occur during the AM peak on accounts registered to private households. We focus on the AM peak period (5-9 AM) for three reasons. First, this is the peak window of usage because the road is traveling towards downtown Los Angeles. Second these trips are likely work commutes with comparable penalty functions for late arrival, and finally the congestion levels allow drivers less opportunity for passing which increases measurement error in calculating the (hypothetical) mainline travel time had the driver not used the ExpressLanes and instead drove along the mainline of the freeway.

*Key Variable Construction.*—Sorting of individuals across roads and freeways implies that the travel time in the mainline lane of the I-10 can serve as a close approximation of the travel time it would take in any other travel option. We construct an estimate of the hypothetical mainline travel time from road segment average speeds reported by detectors in the I-10 W mainline lanes from the

<sup>14</sup> With the exception of the final set of regressions, all regressions can be replicated without account level information, therefore we remove this identifier to honor the confidentiality agreement signed with Metro and assure the anonymity of program users. We focus on the I-10 W corridor as it has one of the highest PeMS detector counts per mile, one every 0.18 miles on average, and the westbound direction is the predominant commuting direction during the AM peak. Our focus on the AM peak is motivated by the fact that drivers faced with congested roads during this period have little discretion to deviate from the average speed on the road, which is not true when it is in free flow.

<sup>15</sup> For ExpressLanes trips, we compute travel time based on the difference between the timestamp for entry and exit to the lanes.

<sup>16</sup> We do not include data past Dec 31<sup>st</sup>, 2013 to avoid contamination of our estimates from the Clean Air Vehicle Decals program that granted access to single occupant vehicles driving low emissions vehicles which was begun January 1<sup>st</sup>, 2014. Bento, et al. (2014) studied the impact of this program during an earlier period of operation (2004-2011).

California Department of Transportation's Freeway Performance Measurement System (PeMS).<sup>17</sup> Travel time in the mainline lanes is calculated as the distance traveled in the ExpressLanes divided by the average speed from PeMS speed detectors on a parallel stretch of the mainline lanes during the same 5 minute interval as the start of the trip.<sup>18</sup> This is done by matching each ExpressLanes trips observed in our transponder data to the average speeds observed in the mainline lanes from PeMS for the same starting time. Trip-level travel time savings in our analysis is then the difference between the realized travel times in the ExpressLanes and that for the hypothetical same distance trip taken at the same time along a parallel stretch of the mainline lanes of the I-10 W.<sup>19</sup> In most figures and regressions below, we omit the 6.2% of trips where the mainline speed implies negative time savings but include these observations in robustness checks.<sup>20</sup>

Reliability is a willingness to pay for reduced uncertainty in travel time when choosing between routes, and empirical evidence suggests that drivers particularly dislike longer delays (Brownstone and Small, 2005).<sup>21</sup> When arriving at the entrance to the ExpressLanes, the driver already perceives differences in travel

<sup>17</sup> PeMS generates real-time 5-minute speed and flow data for HOV and mainline lanes from loop-detectors embedded in all major California divided highways based on calibrated flow and occupancy observations taken every 30 seconds.

<sup>18</sup> That individuals would infer travel times in the mainline based on contemporaneous speeds is consistent with the fact that the speed data from PeMS as well as other sources is widely available from news outlets, and mobile technology like Waze that tracks the speed of users provides extremely accurate travel time predictions based on contemporaneous travel conditions.

<sup>19</sup> This is the appropriate comparison to make because during the peak commuting period, the Nash Equilibrium in routing serves to equalize average travel times between substitute commuting routes, so travel times in the mainline of the I-10 W are consistent with the lowest possible travel time commutes for any untolled route in the transportation system. We focus our analysis on accounts registered to private individuals for whom travel time savings likely correspond to trips to work.

<sup>20</sup> Trips with negative travel time savings appear to occur when mainline speeds are abnormally high, suggesting passing is possible and our measure of counterfactual mainline speed is subject to error. Median mainline speeds are in excess of 65 mph for these negative trips, while those with positive time savings have a median mainline speed less than 45 mph.

<sup>21</sup> A large literature in transportation has considered how commuters value reliability of travel time savings (Fosgerau and Karlström, 2010). Nevertheless, it is not immediately obvious how to measure preferences for over the cost of uncertain travel times in stated preference surveys for hypothetical trips, when the value of resolving uncertainty may depend on trip-specific factors (Small, 2012).

conditions suggesting there is limit uncertainty at that point in the trip. If this is truly the case, individuals' willingness to pay to enter the ExpressLanes is unlikely to reflect insurance or an option value in case conditions deteriorate since there is no longer any uncertainty. To the extent that some uncertainty remains, this will be the most onerous in the right tail of the travel time distribution and so we construct a measure following Brownstone and Small (2005) and Small, Winston and Yan (2005). This variable measures the difference in the difference between the 80<sup>th</sup> and 50<sup>th</sup> quantile of travel time savings between the mainline lanes and the ExpressLanes. This measure varies by time of day and is calculated based on a random sample of weekdays during the year when their study was conducted.

We note two important differences in our reliability measure. First, given the institutional features of the program, our measure of reliability varies by road segment, reflecting portions of the roadway where slowdowns in mainline lanes may be more likely. Second, unlike prior studies, we observe individuals making repeated choices over the course of a year and therefore construct reliability measures that reflect changing realizations of the travel time distribution. Therefore, we calculate this measure for each hour and day of the week, using a moving average of the same weekday hour for the preceding 30 days and segment of trips, reflecting the fact that expectations of travel time savings are likely formed based on recent performance.<sup>22</sup> An important feature of this variable is that, unlike prior literature, there is considerable variation across observations in reliability by hour, date and segment where the trip is taken as shown in Appendix Figure B.3.

<sup>22</sup> In order to remove the influence of infrequent times when travel times are longer in the ExpressLanes (times with fewer cars when driver discretion results in slower speeds), we restrict the sample for constructing reliability to positive travel time differences. In Appendix Table C.6, we include these infrequent occurrences of negative travel time savings and find no impact on our estimates. In Appendix Table C.7, we test the robustness of our estimates to changes in the calculation of reliability via the choice of window and quantile of travel time differences.

### C. Suggestive Evidence of Urgency and Schedule Constraints

*Urgent Drivers.*—We have argued that preferences for urgency may explain the unusually high WTP for travel time savings found in Panel A of Figure 1. Heterogeneity in the valuation of time may provide an alternative explanation for this pattern. However, Table 1 shows that even these high value users infrequently use the lane at 8.8 times per month—a frequency that inexplicably increases as WTP per hour decreases across the deciles.<sup>23</sup> If these users in the first decile truly valued their time at this level, we would expect them to appear in the lane a lot more frequently, and for nearly every commute.<sup>24</sup>

*Schedule Constraints.*—Panel A of Figure 3 depicts the kernel smoothed density of demand over the morning peak, when schedule constraints are likely to be relatively more binding. Vertical lines indicate the key hours of 7:00, 8:00, and 8:30 AM when work start times or morning appointments may be common. It is clear from the figure that demand for the ExpressLanes rises 10-15 minutes before these times, and then falls immediately afterwards. This evidence is suggestive that drivers may be using the ExpressLanes to ensure on-time arrival based on scheduling needs.<sup>25</sup>

Further, in Panel B of Figure 3, we confirm that exit time is narrowly centered around the average exit time (calculated based on the distribution of repeated users by account) in the morning peak in the upper two plots. The bottom two plots show

<sup>23</sup> As noted in Appendix Table C.3, the most common vehicles in this first decile are the Honda Accord, Honda Civic, and Toyota Camry—surprisingly inexpensive vehicles. Appendix Table C.4 shows that there is wide spread in monthly use frequency across the travel time distribution, with most usage in 2-5, 6-10 and 11-20 time per month bins.

<sup>24</sup> And, if this were the case, the implied hourly wage would have been \$640, corresponding to an annual wage of \$1.3 million. It is highly unlikely that an individual with this income would drive the types of cars we observe most frequently associated with the use of the lane.

<sup>25</sup> While it is useful to characterize demand patterns to understand the causes of urgency, the empirical framework laid out section III does not require us to estimate demand for ExpressLanes use explicitly, but rather infer the implied willingness to pay through a revealed preference framework.



this same distribution for times during the afternoon peak when drivers typically have more discretion over arrival times and when lateness penalties may be non-existent or less severe. For these plots the distribution of exit times is substantially wider suggesting less repeated use of the lanes around a common afternoon desired arrival time.<sup>26</sup> Together Figures 2 and 3 and Table 1 begin to create the basis for understanding the mechanisms that lead drivers to value of urgency.

### **III. Conceptual Framework and Econometric Methods**

#### *A. Conceptual Framework for Recovering Drivers' Preferences for Travel Time Savings*

Earlier studies have modeled drivers' behavioral decisions, including departure times, and the choice of routes and lanes in freeways. Schedule constraints have been incorporated in two strands of the literature: Bottleneck models (Vickrey, 1969) and generalizations of the value of time framework with schedule constraints (Noland and Small, 1995).

In the classical bottleneck model, first developed by Vickrey (1969), and formalized in Arnott, de Palma, and Lindsey (1993), individuals choose departure times, given a desired time of arrival. Because of its deterministic nature, there is no uncertainty in travel time in the bottleneck model. As a result, individuals can always adjust departure times to meet any underlying schedule constraint.

The argument that a failure to meet scheduling constraints can generate measurable welfare costs is made in Small (1982), who notes that cost-benefit

<sup>26</sup> In Table C.5 in the Appendix, we consider the correlation between exit times and the quantity of the ExpressLanes used suggesting that commuters consume more of the ExpressLanes the later they expect to arrive at their destination. This provides further suggestive evidence that urgency arises from schedule constraints.

analysis of infrastructure pricing policies may contain inaccuracies by omitting costs and/or benefits in the form of altered schedules. The framework developed in Noland and Small (1995) does explicitly define a penalty function for late arrival for a more general commuting cost function incorporating a discrete penalty for late arrival in the context of assessing optimal departure time.

Building on Noland and Small (1995), the impact of scheduling constraints can be introduced in a stylized conceptual framework that also highlights the role of penalties that do not scale up linearly in the amount of time. Consider a commuter traveling to work during the morning peak. The driver forms expectations about congestion levels and decides on a departure time. Once on the road where previously uncertain travel times become realized, the probability that the driver will arrive late and fail to meet a schedule constraint is largely resolved.<sup>27</sup> Presented with the option of purchasing the time needed to avoid late arrival, the driver decides whether or not to enter the ExpressLanes.<sup>28</sup> As discussed above, ExpressLanes assure them a minimum level-of-service (in the form of secured minimum speed), upon the payment of a toll. In other words, when faced with the likelihood of late arrival, drivers can prevent penalties from being late by taking advantage of the ExpressLanes.

In a general setting, a driver's utility can be represented as:

$$(1) \quad u(T) = \begin{cases} R_{OT} - \theta T & \text{if on time} \\ R_L - \theta T & \text{if late} \end{cases}$$

<sup>27</sup> While theoretically there may still be some uncertainty left, in practice, for ExpressLanes programs with a varying toll system that guarantee a minimum speed and where drivers obtain information via signage uncertainty is dramatically reduced.

<sup>28</sup> In principle, it would be possible for the driver to leave the house early in order to lower the probability of late arrival. This then reflects the trade-off between this benefit and the cost of potentially arriving at work earlier than desired. In practice, drivers may find themselves running late for a variety of reasons. The most straightforward case is where an individual faces an idiosyncratic shock (e.g., left briefcase at home) and now has to re-evaluate commuting decisions without the ability to leave the house earlier.

where  $T$  denotes time commuting,  $\theta$  the value of time, and  $R_{OT}$  and  $R_L$  the gross benefits of arriving on time and late, respectively, where  $R_{OT} > R_L$ .

Let  $P_R$  and  $P_E$  denote the probability of arriving on time if the driver chooses the regular, untolled lane and ExpressLanes, respectively. Given uncertainty in travel time, the driver's expected utility can be represented as:

$$(2) \quad E(u(T^E)) = P_E[R_{OT} - \theta \cdot E(T^E)] + (1 - P_E)[R_L - \theta \cdot E(T^E)] - C$$

when the driver takes the ExpressLanes, and

$$(3) \quad E(u(T^R)) = P_R[R_{OT} - \theta \cdot E(T^R)] + (1 - P_R)[R_L - \theta \cdot E(T^R)]$$

when the driver takes the conventional, untolled lanes. In this case, the driver pays a toll in the amount of  $C$ .<sup>29</sup>

Faced with this decision, the driver opts for the ExpressLanes if the expected utility of choosing the ExpressLanes is higher than that of the regular lane. Or, in other words, the driver enters the ExpressLanes and pays a toll  $C$  when:

$$(4) \quad C < (P_E - P_R)[R_{OT} - R_L] + \theta \cdot E[T^R - T^E]$$

where  $T^E$  and  $T^R$  denote the travel times in the ExpressLanes and the regular lane, respectively.

<sup>29</sup> This representation of uncertainty of on-time arrival represents a relatively strong simplification, and should only be considered for illustrative purposes. Here we assume that these probabilities are uncorrelated. We also assume that the toll  $C$  is uncorrelated with  $T$ , whereas in reality it scales with the level of congestion in the untolled lanes (and therefore demand for the tolled lanes). We assume that both probabilities, the valuation of travel time savings,  $\theta$ , and the gross benefits of on-time and late arrival are known to the driver.

Equation (4) provides an intuitive decision rule for a driver choosing to enter the ExpressLanes when the toll paid is less than the sum of the expected relative penalty of being late (first term in the right-hand side of equation (4)) and the value of travel time savings from using the ExpressLanes (relative to the regular lane) (second term in the right-hand side of equation (4)). Equation (4) also implies that the expected penalty of being late is not directly linked with the amount of travel time savings that the ExpressLanes generates, but rather by the disutility of being late, which even in this simple framework is given by a discrete amount, a constant.<sup>30</sup> In Panel A of Figure 2, we plot the right-hand side of equation (4) to further highlight the point that, even for negligible amounts of travel time savings drivers may exhibit high willingness to pay to enter the ExpressLanes if, by entering the ExpressLanes, they recover the time that is needed to secure on-time arrival. Figure 2 also suggests that, even when the researcher does not directly observe drivers' underlying preferences and the structure of penalty functions they face depending on trip purpose, treating the expected value of the penalty as a constant that does not scale up with time may allow the econometrician to recover the bulk of the structure of preferences, as well as estimates for the value of urgency.

### *B. Estimating the Value of Urgency*

We now describe an approach to empirically implementing equation (4) and to recover the first estimates of the components of willingness-to-pay for travel time savings that incorporate the value of urgency. Earlier work that estimates drivers' preferences for travel time savings relied on discrete choice models (Small, Winston, and Yan, 2005), and recovered estimates for the value of time and

<sup>30</sup> In a more general framework, one can allow the penalty function to also scale with the number of minutes late (or early) as well, but, as we will demonstrate later, to explain the observed distribution of willingness-to-pay, this per-minute late penalty would have to be highly non-linear so that it can be approximated by a constant.

reliability. Traditional discrete choice demand estimation becomes intractable in the context of dynamic prices, which make it difficult to exogenously determine the effect of prices on individual choices since they are simultaneously determined with congestion delays.<sup>31</sup> Further, it is not obvious how to incorporate schedule constraints and penalty functions that vary by individual and trip purpose.

An advantage of the hedonic approach as laid out in Rosen (1974) and Roback (1982) is that with individual-level data, the simultaneity between market prices and attributes (here the toll paid and travel time savings) does not confound identification of the marginal valuation of attributes.<sup>32</sup> However, as documented in the literature (Bishop, et al., 2019) other confounders present themselves when estimating hedonics: measurement error, omitted variable bias, functional form misspecification, and sorting as considered in sections III.C and III.D.<sup>33</sup>

In our data we observe driver  $a$  paying observed toll,  $toll_{a,s,t}$ , on ExpressLane segment  $s$  on date  $t$  to save expected travel time  $E[TT_{a,s,t}]$ , corresponding to the last term in equation (4). The simplest possible model of homogeneous preferences decomposes the determinants of willingness-to-pay into components that scale up with travel time savings and those that do not. This model is presented in equation

<sup>31</sup> Even the use of instrumental variables that determine the delay in the mainline lanes would not allow us to separate urgency from correlated, unobservable factors.

<sup>32</sup> Consider the effect of an unobserved shock (e.g., an accident on a parallel freeway, the I-210) that induces a larger number of drivers onto the I-10 and the ExpressLanes. This increase in demand increases ExpressLanes travel time savings and the toll, but it will not alter the revelation of the value of urgency and time for these drivers and all other drivers on the road because the quality of service of the ExpressLanes will be unaffected by this shock as the speed in the lanes never drops below 45 mph by design.

<sup>33</sup> One strand of the empirical hedonic literature has demonstrated the importance of various strategies in accounting for omitted variable bias (Chay and Greenstone, 2004; Davis, 2004; Bento, et al., 2015; Black 1999). At the same time, a separate strand has shown the limitations of using standard difference-in-difference approaches to obtain welfare-relevant measures of willingness-to-pay for non-market amenities (Kuminoff and Pope, 2014). In the rare cases where the data make it possible to estimate a second stage of the hedonic model, this would allow for consideration of a range of individual characteristics that may help to better understand the distribution of individual MWTP such as preferred departure times, schedule constraints, multi-modal travel among others. While these patterns are clearly an interesting research topic, as the hedonic literature has shown (Ekeland, et al., 2004), they are not necessary to credibly estimate the underlying hedonic price functions and thus demonstrate the value of urgency.

(5) where we regress a price, the toll paid, on the driver’s expected travel time savings at the time entering the ExpressLanes and a constant:

$$(5) \quad toll_{a,s,t} = \delta + \theta E[TT_{a,s,t}] + \gamma Reliab_{s,t} + \varepsilon_{a,s,t}.$$

Equation (5), by regressing a price on a vector of characteristics, is a first-stage hedonic price function where the coefficients represent the marginal willingness to pay for that characteristic.<sup>34</sup> Note that because travel time savings should be directly linked to the opportunity cost of wages, the simplest approach is to assume that they enter the regression linearly.<sup>35</sup> In our setting, the coefficient on travel time saved,  $\theta$ , represents the WTP that scales with time, while  $\delta$ , the constant, is WTP that is constant in the amount of time saved. Following Panel B of Figure 2, the value of urgency for those who choose to enter the lane, which does not scale with the amount of time saved, is captured by  $\delta$ .<sup>36</sup> A more restricted model would omit the constant, implicitly assuming it is zero.

<sup>34</sup> Here the relevant “market” as related to a standard hedonic model of housing is a five-minute interval on a particular day. The choice is between the ExpressLanes and the mainline lanes, and the alternative-specific attributes for which we will recover hedonic price functions are travel time differences between the lanes, and in later models, reliability. The attributes of the unchosen mainline lane alternative are constructed based on the hypothetical travel time the driver would have experienced in the lane with the same time of entry. Here our identifying assumption is that a driver observed in the ExpressLanes on a particular day during a particular five-minute interval would have been in the mainline lanes had we not observed them in the ExpressLanes. As noted above, during peak commuting periods the Nash equilibrium of route choice serves to make travel time in the mainline the lowest unpriced travel time alternative, though we do validate the robustness of this assumption in Appendix Tables C.8 and C.9 by comparing travel times in the ExpressLanes to the I-210 W instead. Note also that this requires no assumption about what lane or mode of transportation the driver might take on other days when we do not observe them in the ExpressLanes.

<sup>35</sup> We examine this assumption in detail later in the paper. A related issue is whether (5) should include realized travel times rather than expected. We find that, in practice, estimates are largely the same with either variable, and that given the sophistication of drivers in Los Angeles and the availability of real-time travel time prediction from apps like Google Maps and Waze, using realized travel time is likely more appropriate. Further robustness checks of our construction of travel time are discussed in section III.C.

<sup>36</sup> There is also an implicit assumption that drivers using the ExpressLanes are not early. If drivers were early, the implied value of time would be much lower than observed in our estimation. Traditional estimates of the value of time for early individuals are roughly one-quarter of the wage (Small, 2012), which would be roughly \$5 per hour in the Los Angeles area. As we show in section IV models with and without urgency during the morning peak have market-clearing prices that are well above this value. Furthermore, scheduling models suggest that the benefits from early arrival scale with time; our detection of a statistically significant constant is evidence against this possibility.

Given uncertain realization of travel time savings, drivers may also value higher moments of the time distribution; in particular, they may dislike the possibility longer delays. Following prior studies (Brownstone and Small, 2005; Small, Winston and Yan, 2005), (5) also includes,  $Reliab_{s,t}$ , a measure of the difference in the dispersion of travel time savings between the mainline lanes and the ExpressLanes that represents the reliability of travel time savings.<sup>37</sup>

It is important to appreciate the difference between preferences for urgency and reliability. By construction, in previous studies such as Small, Winston and Yan (2005), where a single value of reliability for a given hour is often constructed for the entire sample, there may be a concern that this measure is correlated with the constant in (5), which represents the value of urgency. As explained in section II.B., while in the spirit of prior literature, our reliability measure is calculated using a moving window and by sub-segments of the ExpressLanes. As a result, our measure varies across days in the sample as well as within hours of the day as demonstrated in Appendix Figure B.3, allowing us to disentangle preferences for urgency from preferences for reliability. Because we pair a segment with the nearest mainline detectors to measure speeds, we cluster the standard errors in all regressions by segment.

### *C. Threats to Measuring Travel Time Savings*

Even if we have adequately controlled for unobserved determinants of willingness-to-pay in equation (5), we may still be concerned that measurement

<sup>37</sup> This variable measures the difference in the difference between the 80<sup>th</sup> and 50<sup>th</sup> quantile of travel time savings between the mainline lanes and the ExpressLanes and is calculated for each hour, date and segment of trips in our data using a moving average of the same weekday hour for the preceding 30 days. Our estimates are robust to the choice of quantile and window as shown in Appendix Table C.7.

error in the construction of travel time savings may lead to biased estimates of both the value of time and the value of urgency. This fact is easily visualized in Panel B of Figure 2, where the dashed gray line reflects how underestimation of the slope, the estimate of the value of time, affects the magnitude of the penalty function. An underestimation of the slope means that the intercept, the estimate of the value of urgency, is overestimated. This could occur if drivers perceive travel time savings to be less than their realized amount so that their valuation is underestimated by our model.

There are two potential types of measurement error of time saved that may have implications for our estimates: researcher measurement error and error due to drivers' perception of travel time saved. While we have extremely detailed information on individual speeds in the ExpressLanes, there is somewhat less certainty about what speed the driver would have achieved outside of them. During low demand hours the 5-minute average speed in the mainline lanes may obscure the ability of drivers to pass and achieve travel times very different than the average. Our main estimates, therefore, are for the morning peak because during these hours the road is sufficiently congested to prevent substantial passing. This helps to guarantee that the measured speed is reflective of what drivers in that lane would have been forced to experience as they have less discretion over speed.<sup>38</sup>

We may also be concerned that drivers do not accurately perceive their time savings. First, we note that the ability of drivers to enter or exit the lanes at multiple

<sup>38</sup> Using a side street during the morning peak is less likely to offer an improvement over mainline travel. When congestion is high, a Nash equilibrium ensures that indirect routes will have faster speed but equal travel time to a direct but congested route. As further robustness, we examine alternative specifications of travel time difference: In Appendix Table C.10 Panel A, as a bounding exercise, we consider certain extreme driver miscalculations of travel times (e.g., twice the time savings, random guess) and find almost no effect on the value of urgency. In Appendix Table C.8, we control for periods when the I-210 W, an imperfect substitute for the I-10 W was traveling at or below its average speeds. In Appendix Table C.9 we construct travel time difference between the I-10 W ExpressLanes and the I-210 W mainline lanes (rather than those on the I-10), and in Appendix Table C.10 Panel B, we calculate travel time difference based on past realizations of mainline speeds. In all cases, our estimates of the value of urgency remain close to the \$3.24 baseline estimate.



points makes substantial errors unlikely. If a driver saves less time than anticipated it is possible to leave the lanes. Second this error must be substantial to explain the magnitude of the constant.<sup>39</sup> We can also eliminate trips with relatively small travel time savings, which may be occasions where the individual mistakenly took the ExpressLanes. Furthermore, we can examine the time period after October 20<sup>th</sup>, 2013. On this date, signs were posted giving the expected travel time savings from using the ExpressLanes alongside the value of the toll. This measure would help to resolve any ambiguity that may have existed. Studies that measure the capitalization of critical information on housing values find non-trivial effects (Davis, 2004; Figlio and Lucas, 2004). Therefore, if this information yields new information for drivers, we would expect drivers' estimates for the value of time and urgency to change.

Even with improved information, drivers might still misperceive actual travel time savings. So, we also consider an instrumental variables approach constructed in the spirit of Aizer, et al. (2018) where future realizations of travel time serve as an instrument for contemporaneous travel time. For our account-by-account regressions, this involves IV estimation following the split sample approach developed by Angrist and Krueger (1995).

#### *D. Heterogeneous Agent Model*

While equation (5) provides an intuitive framework for decomposing willingness-to-pay into its key components, it is somewhat limited in that it assumes a homogeneous agent model. Further, in (5), we have not yet taken full advantage of

<sup>39</sup> From our baseline estimates this misperception would have to be 23 minutes.

the variation that comes from observing drivers repeatedly in the ExpressLanes, typically with varying levels of congestion and tolls. By observing the same commuter repeatedly over different price and attribute levels, we are able to estimate separate individual hedonic bid functions for each consumer and, therefore, recover the entire distribution of the value of urgency and time for participants in the program. Unlike much of the hedonic literature that uses repeated housing transactions,<sup>40</sup> observing the same commuter repeatedly enables us to estimate the hedonic price function without assuming a constant average marginal willingness-to-pay for the population.<sup>41</sup>

We estimate individual bid curves with account level measures of the value of urgency and time by estimating the regression

$$(6) \quad toll_{a,s,t} = \delta_a + \theta_a TT_{a,s,t} + \gamma_a Reliab_{s,t} + \varepsilon_{a,s,t}$$

separately for each account with multiple trips.<sup>42</sup> An advantage of this approach relative to the homogeneous agent model from equation (5) is that if high-income individuals self-select into trips with small time savings while low-income individuals sort into trips with longer time savings, the assumption of a single preference structure may bias our estimates and render the homogeneous agent model of equation (5) useless for welfare analysis.

<sup>40</sup> The exception is Bishop and Timmins (2018), who utilize a preference inversion technique that imposes limited assumptions on the elasticity of demand for the measured amenity with the benefit of minimized endogeneity problems.

<sup>41</sup> While other studies have examined repeated sales data for hedonic estimation, they examine the repeat sale of the same item, such as a home, exposed to different levels of an amenity (for example pollution or school quality) to different individuals.

<sup>42</sup> Consistent standard errors are estimated using the bootstrap.

### *E. Correlation with Omitted Variables*

Consistent estimation of equations (5) and (6) is confounded to the extent that any unobserved factors may co-vary with the observed characteristics (Chay and Greenstone, 2004; Davis, 2004; Bento, et al, 2015). In our setting, the concern is rather that any other time-invariant amenity such as smoother pavement or a feeling of superiority of being in the lane would also be captured by the constant, potentially leading to an overestimate of the value of urgency. In section IV.C, we employ an empirical strategy to compare time invariant and time variant determinants of willingness-to-pay during the weekday morning peak to that during weekend morning for the same accounts to determine the extent to other factors may confound estimation of the value of urgency. Importantly, if what drivers value is related to the pavement of the lane, such valuation should be present both in weekdays and weekends independently of the purpose of the trip and schedule constraints.

## **IV. Results**

### *A. Mean Estimates of the Value of Urgency*

Table 2 presents account-by-account estimates of equation (6) based on individual bid functions, consistent with a model of heterogeneous preferences. It also reports in Column I estimates from a single regression using the homogeneous agent framework of equation (5). Each subsequent column gives the mean, bootstrapped standard error, and interquartile range of estimated coefficients from an account-by-account regression for each of the 9,053 accounts with more than 10 transactions. This assumes that each account holder has their own value of urgency, value of time, and value of reliability.

Comparisons of the first and second columns reveal that estimates are somewhat different depending on whether we use expected versus realized travel time savings, but that this difference is not substantive. In column III, we present our central estimates of the determinants of willingness-to-pay for the ExpressLanes. We find that the mean value of urgency in the population is \$3.24 with bootstrapped standard errors in parentheses. In brackets we give the interquartile range of estimates showing that across accounts there is substantial heterogeneity in the value of urgency, suggesting likely heterogeneity in penalty functions.<sup>43</sup> Interestingly, half of all account holders' estimates of the value of urgency fall between \$2.30 and \$4.05.

### *B. Relevance of the Value of Urgency*

The results provided above give estimates of the value of urgency, value of time, and value of reliability but they do not assess the relative contribution of each component to the welfare generated for these drivers. To decompose this effect, we can compare the value of urgency with the average toll, which is \$3.71. The value of urgency at \$3.24 represents 87% of the value for the average toll. In comparison, the value of time evaluated at the mean time savings of 3.79 minutes is \$0.51, while the value of reliability of \$17.61 per hour evaluated at the mean value of reliability of 2.1 minutes is \$0.62.

It is important, however, to stress again that while we obtain a mean estimate of the value of urgency of \$3.24, there is no reason to believe that the value of urgency is constant across our sample. Indeed, in Figure 4, we plot the value of urgency and the value of time across accounts in our sample. These demonstrate considerable

<sup>43</sup> We also estimate the values of urgency and time during the afternoon peak and on other corridors of the ExpressLanes in Appendix Table C.11. We find that while there is heterogeneity, the qualitative results remain.

heterogeneity across drivers. Interestingly, the distributions of the value of urgency and value of time have a symmetric shape approximating a normal distribution, centered around the estimates found in the homogeneous model, based on equation (5). Variation in the value of urgency over the course of the morning peak are also plotted in Appendix Figure B.4, with the highest values between 7 and 8AM.

### *C. Threats to Identification*

*Measurement Error in Travel Time Savings.*—One particular concern with the estimates from equation (6) is that if the econometrician incorrectly measures travel time savings perceived by drivers, then this error may bias both the estimates of the value of time and the value of urgency. This bias can be visualized as mismeasurement of the horizontal axis in Figure 2, resulting in a downwardly (upwardly) biased slope parameter for  $\hat{\theta}$ , consequently also biasing the estimate  $\hat{\delta}$  up (down).

Only if all drivers systematically mistook their time savings by 23 minutes, regardless of travel time savings would the constant be reduced to zero.<sup>44</sup> Therefore, it is unlikely that, even with some bias, the estimate of the value of urgency will disappear. Nonetheless, to address these concerns, in column III of Table 2, we restrict our analysis to the subset of trips taken after October 20<sup>th</sup>, 2013, when signs were set up at each ExpressLanes entrance indicating the travel time savings that drivers could expect based on real-time data in the lanes. If nothing, these estimates suggest that the value of urgency only goes up when drivers have better information, indicating that measurement error is unlikely to be upwardly biasing

<sup>44</sup> This is based upon 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: \$3.24.

our estimates. Also of note is that the value of reliability is now almost half as much, which may reflect the fact that with information about travel time differences at the point of ExpressLanes entry, an even greater amount of uncertainty about the distribution of travel time difference is resolved. In housing markets where the difficulty of measuring the quality of many amenities and the heterogeneity of access to information means that new information about amenities has a profound effect on prices (Davis, 2004; Figlio and Lucas, 2004). In contrast, here, perhaps through the use of real-time technologies such as Waze and Google Maps, there is less of a measurable effect, probably because information about quality is unambiguous and readily available.

As previously noted there may be concerns that systematic mismeasurement of what drivers perceive to be the time saved may downwardly bias the coefficient on the value of travel time and therefore cause us to overstate the value of urgency. A standard approach to address concerns about attenuation bias from measurement error is instrumental variables (IV). We follow the approach of Aizer, et al. (2018) to take advantage of repeated measurements of travel time savings for the same time of day and segment over time. The logic of this approach is that if our measure of travel time savings is noisy compared to what drivers actually perceive, measurement error between any two trips are likely to be independent across drivers driving at the same time of day.

Our instrument set consists of three variables that are the average time savings one hour, one week and two weeks after the trip by hour of day, day of week and road segment.<sup>45</sup> We use leads, as opposed to lags, in travel time, because these are likely to be highly correlated to contemporaneous time savings during a given hour,

<sup>45</sup> With three instruments and a single endogenous variable (travel time savings), we are able to estimate the split sample IV model using an overidentified system of equations.

day of week and road segment, but unlikely to be affected by any unobserved contemporaneous factor affecting the driver when a particular trip is taken. Our IV estimates using a heterogeneous agent model are reported in column IV of Table 2 and do not meaningfully depart from our central estimates.<sup>46</sup>

*Time-Invariant Unobserved Attributes of ExpressLanes.*—A separate concern may be that the constant will capture not only the value of urgency but also any other time-invariant amenity in the ExpressLanes. If such an amenity exists it cannot take the form of congestion, which would generate travel time savings, but could be a belief that the ExpressLanes are safer or a smoother ride than the mainline lanes.<sup>47</sup> To address this concern, we take advantage of the fact that trips taken in the off-peak likely have a lower penalty for late arrival. If this is the case, then comparing the determinants of willingness-to-pay for weekday morning peak trips to weekend morning trips can help to separately identify the value urgency from amenities of the ExpressLanes present during lower levels of congestion.<sup>48</sup> And, while trip purpose can differ between weekend and weekday morning, the value of time-

<sup>46</sup> To estimate our IV model in our account-level regressions, we follow the split-sample IV approach of Angrist and Krueger (1995). This is done by estimating the first stage regression of travel time savings on our instruments and reliability using all accounts except for one. Then in the second stage we regress the toll paid by the excluded account on a constant, reliability for the trips that account took, and the first stage predicted value of time savings. This is repeated for all accounts. As with our other account-level regressions, confidence intervals are constructed based on a 500 iteration bootstrap. In principle, since reliability is positively correlated with travel time savings, we would want to also instrument for reliability. In practice, however, given the limited explanatory power of reliability in our model and the fact that theory does not give a clear rationale for how drivers form expectations in this context, the case for an instrument is less clear.

<sup>47</sup> If such an unobserved amenity was valued by commuters, then estimation of  $\delta$  in (6) would be biased. Note that this travel time invariant amenity cannot be lower congestion as congestion in the mainline lanes is what generates a travel time differential.

<sup>48</sup> Formally, we decompose the constant into a component during the weekday morning peak,  $\delta_a$ , and a component that is always there,  $\mu$ , so that (6) now becomes

$$toll_{a,s,t} = \delta_{a,WE} + \delta_{a,MP}1(MP_t) + \theta_{a,MP}TT_{a,s,t}1(MP_t) + \theta_{a,WE}TT_{a,s,t} + \gamma_{a,MP}Reliab_{s,t}1(MP_t) + \gamma_{a,WE}Reliab_{s,t} + \mu + \epsilon_{a,s,t}$$

where  $1(MP_t)$  is an indicator for trip  $t$  taken during the weekday morning peak, and  $\delta_a$ ,  $\theta_a$  and  $\gamma_a$  now vary between weekend and morning peak periods. The coefficient  $\delta_{a,MP}$  measures the weekday morning peak urgency premium—how much extra punishment failure to achieve on time arrival has—compared with a weekend trip.

invariant road characteristics, such as pavement quality, are unlikely to be related to the trip purpose, making the weekend an appropriate control group. We pool trips during weekend mornings and the weekday morning peak to estimate separate coefficients for weekday morning peak and relative to the weekend.

If commuters experience no value of urgency during weekend trips, then an indicator variable for weekday morning peak trips will 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 will be downwardly biased, and for this reason we consider it a lower bound estimate of the value of urgency.

In Table 3, we present estimates of the model using the weekend as a control group. In column I, we use the subset of 1,121 accounts with 5 or more observations on the weekend and 5 in the morning peak to bound the value urgency of individual accounts from below. By introducing the weekend as a control group, the morning peak indicator gives the lower bound on morning peak urgency as a statistically significant \$2.42, on average. While this estimate will exclude many potentially confounding factors, we view this lower bound as overly conservative. In column II, we introduce reliability and note that our estimates do not change meaningfully. In column III, we restrict the sample to the period where signs displayed true travel time savings, and we note that, while the small sample of 62 accounts is too low for inference about all accounts, the value of urgency remains substantial.

*Sorting.*—It is well known that the hedonic envelope may obscure substantial heterogeneity in bid functions across individuals (Chay and Greenstone, 2004; Bento, et al., 2015; Timmins and Murdock, 2007; Kuminoff and Jarrah, 2010). The concern is that our findings may be the result of sorting by income where high income individuals use the road for short time savings, while low income individuals use the road for larger time savings. Such a pattern could give rise to



the hyperbolic shape and when estimated assuming homogeneity in the bid curve it would give rise to a statistically significant constant.

We address this concern through several ways. First, we note that the drivers who consume small time savings, and have the highest VOT in a model without urgency, use the lane infrequently. Returning to Table 1, we note that drivers in the lowest decile of time savings use the lane on average 8.8 times per month, less than any other group. Although not impossible, it seems unintuitive that agents with extremely high values of time would consume so little of the ExpressLanes and less frequently than groups with a lower valuation. While these lanes are often derided as ‘Lexus Lanes’ we find that the most common vehicles in this lowest decile, Toyota Corollas and Honda Accords, are not typically driven by people we would anticipate earning roughly \$8 million a year.<sup>49</sup>

Nevertheless, the repeated sales nature of our data allows us to run account-by-account regressions, which directly address this concern by estimating individual bid functions rather than assuming a uniform value of time and urgency. Comparing estimates from a model with heterogeneous preferences to that with homogeneous ones can offer suggestive evidence of whether our results are affected by sorting due to heterogeneity. If high value of time drivers sort into particular subsections of the roadway, we would expect to see estimates that are meaningfully different, but comparing columns I and II in Table 2 we find that they are statistically indistinguishable.

An ideal test of sorting might be to regress our estimates on demographic information of the drivers. In the absence of this kind of data, which is a common limitation of most hedonic studies, in Appendix Table C.12, we regress values of urgency and values of time on the list price of the car registered to each account in

<sup>49</sup> See Appendix Table C.3.

our data, with price of the car approximating (imperfectly) individuals' income. We find that drivers of inexpensive vehicles tend to have lower VOT, consistent with Becker (1965), but have higher value of urgency. One explanation of this pattern is that these drivers may be shift workers who face strict penalties for late arrivals, while drivers of more expensive vehicles have a higher value of time but are less likely to be penalties for only a few minutes of delay. We also note from the last row of column I in Appendix Table C.12 that higher values of urgency are actually negatively correlated with values of time. We believe this reflects the fact that, unlike the value of time, higher values of urgency in this context may be less well-explained by driver characteristics such as income, but may relate more to the circumstances of a particular trip and the penalty function. This is again suggestive that if sorting is going on among drivers, it is not of the form that is likely to bias our estimates from equation (6).

#### *D. Further Robustness Checks*

In the Appendix, we consider a variety of robustness checks. Misspecification of functional form may bias our hedonic estimates (Cropper, Deck, and McConnell, 1988), so we consider many variations of functional form in Appendix Table C.13.<sup>50</sup> This includes models without a constant, and we find that these often lead to results that fit the data less well and often do not align with intuition and basic theory. To gauge the extent to which certain high (or low) demand hours affect our estimates, we include quarter, month, day-of-week, quarter by day-of-week and month by day-of-week fixed effects in Appendix Table C.15 and find no

<sup>50</sup> In Appendix Table C.14, we estimate models with higher order measures of travel time savings without the constant and find that the models have AIC and BIC values indicating a worse fit to the data and which imply a reduction in willingness-to-pay as travel time savings get very large, which is counterintuitive.

meaningful effect on our estimates. In Appendix Table C.16, we estimate the homogeneous agent model by segment of the ExpressLanes and find that some segments exhibit higher values of urgency than others, ranging from \$1.37 to \$4.51. Finally, in Appendix Table C.17, we build on studies that have explored the connection between driving behavior and gasoline prices (Burger and Kaffine, 2009) and weather (Leard and Roth, 2019) to show that there is no meaningful difference in the value of urgency on days where gasoline prices are lower or weather is better.

Standard errors in our baseline regressions are clustered by road segment traveled for each trip observation. In Appendix Table C.18, we examine other levels of clustering including two-way clustering (Cameron, Gelbach, and Miller, 2011) to address the spatial and temporal correlation (Anderson, 2014). Of these, clustering at the segment level produces the largest standard errors.

#### *E. Relation to Previous Findings*

In Table 4, we present models that approximate the findings of previous studies to illustrate why we are able to recover preferences for urgency while others have not. Because the mean estimates from the heterogeneous model are identical to the point estimates from the homogeneous model, for simplicity, here we rely on the homogeneous model. And, to perform comparisons with the literature, we run models with and without the constant. We also run separate models, broken down by different minimum threshold values of time savings. Further sample restrictions to travel time savings are considered in Appendix Table C.19, which demonstrate only moderate variation in the value of urgency.

A direct comparison with Small, Winston and Yan (2005) can be made by contrasting the estimates in column III for the model without the constant with their

estimates based exclusively on the revealed preference model.<sup>51</sup> Our model yields estimates of the value of time and the value of reliability of \$21.56 and \$23.36 respectively, which are not statistically different than those reported in Small, Winston and Yan (2005). One should note, however, that the estimate of the value of time is substantially higher than half the local wage, the estimate generally found in the literature (Small, 2012). When we add back the constant, our estimates in column III, we recover estimates for the value of time ranging from \$8.02 to \$4.58 per hour, always at a minimum slightly lower than those in Small (2012).

While a closer comparison with earlier studies is not possible, given that the structure of tolls is fundamentally different across studies, two points are worth noting. First, removing the constant from the model appears to decrease the fit of the model using the AIC and BIC. Second, given the institutional features of previous programs like the ExpressLanes, travel time savings presented to respondents were capped at a minimum of 10 minutes. Based on Panel A of Figure 1, one can appreciate that visually it appears as if that part of the curve is relatively flatter. However, by contrasting Column III with and without the constant, we provide suggestive evidence that prior estimates of the value of time and reliability may be severely overestimated, since even at portions of the curve in Panel A, Figure 1 where time savings are greater than 10 minutes, drivers still exhibit preferences for urgency.

<sup>51</sup> Small, Winston and Yan (2005) estimate a joint revealed and stated preference model. Median estimates from the former are \$21.46 per hour for value of time and \$19.56 per hour for the value of reliability. Stated preference estimates for the value of time are lower \$11.92 per hour and for reliability are not comparable because of survey phrasing.

## **V. Implications for Cost-Benefit Analysis**

Preferences for urgency and the estimates of the value of urgency have fundamental implications for cost-benefit analysis of road infrastructure projects. Following prior literature, proper ex-ante assessments of the benefit of the ExpressLanes would require one to predict the time saved by agents using ExpressLanes and multiply this value by the estimate of the value of time over the program period studied here, February 22<sup>nd</sup>, 2013 to December 31<sup>st</sup>, 2013. In this case, the projected benefits of the project would be \$221,363, which barely surpasses the infrastructure costs during that time period of \$215,250.<sup>52</sup> In sharp contrast, the program actually generated \$1.31 million during that period.

Without the value of urgency, an ex-ante analysis of the project would underestimate the total benefits by an order of magnitude during this time frame. Even a value of time two or three times that of the standard, would be off by more than 100 percent.

## **VI. Conclusion**

In an ideal setting where drivers are observed making choices of lanes depending on varying tolls, level of congestion and schedule constraints, we study drivers' willingness to pay to avoid congestion, and provide the first estimates of the value of urgency. The value of urgency represents a discrete WTP to meet a schedule constraint, recognizing that individuals often face penalties for being late that do not necessarily scale up with the amount time that individuals are late. With unusually rich data that tracks the same drivers on multiple occasions with different

<sup>52</sup> Source: Correspondence with LA Metro, 04/15/14. This corresponds to the operation and maintenance costs of the corridor including weekends, holidays, across all hours of the day.

levels of congestion and time-varying tolls, we recover preferences for travel time savings in an hedonic price function, and overcome standard identification challenges related to measurement error of key attributes, omitted variables, and functional form misspecification (Bishop, et al., 2019; Chay and Greenstone, 2004; Bento, et al., 2015; Black, 1999). Even more important, by estimating individual-level bid functions, the implied marginal-willingness-to-pay estimates from the first-stage hedonic price function constitute a ‘sufficient statistic’ for welfare measurement (Banzhaf, 2019a).

Preferences for urgency and the value of urgency have three broad implications. First, they can alter the direction of cost-benefit analysis for road infrastructure with level-of-service pricing. In fact, we demonstrate that, because the value of urgency represents roughly 87% of the toll paid to enter the ExpressLanes, ex-ante cost-benefit analysis that ignore this value may misguide road infrastructure investments. More generally, it is likely that amenities or externalities in contexts different from ours do not scale with quantity or have thresholds that are more relevant than just marginal improvements. Like in our application, by ignoring the discrete nature of willingness to pay, studies of the value of such amenities may be severely underestimated.

Second, our results question the use of stated preferences surveys to infer drivers’ preferences for travel time savings. At a minimum, there is a need to reconsider the way researchers solicit stated preferences and demand for projects that generate time savings. Future survey work may improve prediction of the benefits of such projects by soliciting willingness to pay to avoid being late and the frequency at which individuals are late, as opposed to focusing on travel time saved from an average trip. At the same time, this implication also means that survey respondents will have to be trained to respond to questions related to specific, not average, trips, since penalties will likely vary with the purpose of different trips.

Third, our results suggest that the structure of pricing to regulate congested infrastructure matters, and was far from fully understood in prior literature (Small, Winston, and Yan, 2005). Recent studies have used rich datasets to examine congestion patterns in cities and derive the implied Pigouvian tolls that would internalize the congestion externality (Yang, Purvejav, and Li, 2000; Akbar et al., 2018). While these findings are insightful, in practice, anything approaching Pigouvian congestion pricing of entire road systems remains politically challenging.

In contrast to Pigouvian congestion pricing, the pricing structure here applies to only one lane of a freeway, and resembles real-time, level-of-service pricing. With a minimum speed guaranteed, drivers can re-optimize their decisions to still meet a schedule constraint, especially in situations where otherwise they would be late and would face a penalty. A potential advantage of this pricing scheme over Pigouvian pricing is that it solves the uncertainty related to travel time that comes from the drivers' inability to predict aggregate demand at any point in time. As documented in theoretical models that examined the role of pricing and service quality in regulated utilities and monopolies (De Vany, 1976; De Vany and Saving, 1977; Chao and Wilson, 1987), level-of-service pricing offers a menu of contingent contracts for the distribution of scarce space on the road, adjusting utilization of capacity to match supply and demand. Further, and importantly, the secured minimum speed, which is dependent on the real-time response of the toll price to changing congestion conditions, removes uncertainty in travel time. As a result, it creates the ideal setting for drivers to reveal their preferences for urgency.

More broadly, we note that others have implemented cleanly identified econometric methodologies, such as regression discontinuity designs or randomized control trials, to identify short-run effects to infrastructure investments to uncover underlying preferences for travel time savings (Anderson, 2014; Krindler, 2018). It may be less certain that the full benefits of congestible

infrastructure will be fully identified in studies where welfare calculations are based on an average estimated effect across a narrow time window. Indeed, the unusually rich data used in this study provides suggestive evidence that it may be relatively more important to observe variation in individual choices across different levels of congestion over time-varying prices to uncover the full distribution of preferences for meaningful welfare evaluation.



## REFERENCES

- Abbott, Joshua K., and H. Allen Klaiber.** 2011. "An embarrassment of riches: Confronting omitted variable bias and multi-scale capitalization in hedonic price models." *Review of Economics and Statistics*. 93(4): 1331-1342.
- 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.
- Akbar, Prottoy, Victor Couture, Gilles Duranton, and Adam Storeygard.** 2018. "Mobility and congestion in urban India." [http://real-faculty.wharton.upenn.edu/wp-content/uploads/~duranton/Duranton\\_Papers/Current\\_Research/CongestionIndia.pdf](http://real-faculty.wharton.upenn.edu/wp-content/uploads/~duranton/Duranton_Papers/Current_Research/CongestionIndia.pdf) (Last accessed 03/06/2020).
- Anderson, Michael L.** 2014. "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion." *American Economic Review* 104(9): 2763-96.
- 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): 225-235.
- Arnott, Richard, André de 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.
- Ashenfelter, Orley, and Greenstone, Michael.** 2004. "Using mandated speed limits to measure the value of a statistical life." *Journal of Political Economy*. 112(S1): S226-S267.
- Bajari, Patrick, Jane Cooley Fruehwirth, and Christopher Timmins.** 2012. "A rational expectations approach to hedonic price regressions with time-varying unobserved product attributes: The price of pollution." *American Economic Review*. 102(5): 1898-1926.

- Banzhaf, Spencer.** 2019a. "Panel-Data Hedonics: Rosen's First Stage as a 'Sufficient Statistic.'" *International Economic Review*. Forthcoming.
- Banzhaf, Spencer.** 2019b. "Difference-in-Difference Hedonics." <http://hsbanzhaf.gsucreate.org/Papers/Diff-in-Diff%20Hedonics%20Aug%202019.pdf> (accessed 12/27/19).
- Becker, Gary.** 1965. "A Theory of the Allocation of Time." *The Economic Journal* 75 (299): 493-517.
- Bento, Antonio, Matthew Freedman, and Corey Lang.** 2015 "Who benefits from environmental regulation? Evidence from the Clean Air Act Amendments." *Review of Economics and Statistics*. 97(3): 610-622.
- Bento, Antonio M., 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.
- Kelly C. Bishop, Nicolai V. Kuminoff, H. Spencer Banzhaf, Kevin J. Boyle, Kathrine von 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 Christopher Timmins.** 2018. "Using panel data to easily estimate hedonic demand functions." *Journal of the Association of Environmental and Resource Economists*. 5(3): 517-543.
- Black, Sandra E.** 1999. "Do better schools matter? Parental valuation of elementary education." *The Quarterly Journal of Economics*. 114(2): 577-599.
- Bockstael, Nancy E., and Kenneth E. McConnell.** 1981. "Theory and estimation of the household production function for wildlife recreation." *Journal of Environmental Economics and Management*. 8(3): 199-214.

- Brent, Daniel A., and Austin Gross.** 2018. "Dynamic road pricing and the value of time and reliability." *Journal of Regional Science*. 58(2): 330-349.
- Brownstone, David, and 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.
- Burger, Nicholas E., and Daniel T. Kaffine.** 2009. "Gas Prices, Traffic, and Freeway Speeds in Los Angeles." *Review of Economics and Statistics*. 91(3): 652-57.
- Calfee, John, and Clifford Winston.** 1998. "The Value of Automobile Travel Time: Implications for Congestion Policy." *Journal of Public Economics* 69(1): 83-102.
- Calfee, John, Clifford Winston, and Randolph Stempksi.** 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): 238-249.
- Carson, Richard T., and W. Michael Hanemann.** 2005. "Contingent valuation." *Handbook of Environmental Economics*. 2: 821-936.
- 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.
- Cropper, Maureen L., Leland B. Deck, and Kenneth E. McConnell.** 1988. "On the choice of functional form for hedonic price functions." *The Review of Economics and Statistics*: 668-675.

- Currie, Janet, Lucas Davis, Michael Greenstone, and Reed Walker.** 2015. "Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings." *American Economic Review*. 105(2): 678-709.
- Davis, Lucas W.** 2004. "The Effect of Health Risk on Housing Values: Evidence from a Cancer Cluster." *American Economic Review* 94 (5): 1693-1704.
- De Vany, Arthur.** 1976. "Uncertainty, waiting time, and capacity utilization: A stochastic theory of product quality." *Journal of Political Economy*. 84(3): 523-541.
- De Vany, Arthur, Thomas R Saving.** (1977). Product quality, uncertainty, and regulation: The trucking industry. *The American Economic Review*. 67(4): 583-594.
- DeSerpa, Allan C.** 1971. "A Theory of the Economics of Time." *Economic Journal*: 828-46.
- Ekeland, Ekeland, Ivar, James J. Heckman, and Lars Nesheim.** 2004. "Identification and Estimation of Hedonic Models." *Journal of Political Economy* 112 (1): 60-109.
- 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.
- Gallagher, Justin and Paul J. Fisher.** 2020. "Criminal Deterrence when there are Offsetting Risks: Traffic Cameras, Vehicular Accidents, and Public Safety." *American Economic Review*. Forthcoming.
- Gronau, Reuben.** 1973. "The intrafamily allocation of time: The value of the housewives' time." *The American Economic Review*. 63(4): 634-651.
- Grossman, M.** 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy*. 80(2): 223-255.

- Hall, Jonathan D.** 2018. "Pareto Improvements from Lexus Lanes: The Effects of Pricing a Portion of the Lanes on Congested Highways." *Journal of Public Economics* 158: 113-125.
- Hall, Jonathan D.** 2019. "Can tolling help everyone? Estimating the aggregate and distributional consequences of congestion pricing." *Journal of the European Economic Association*. Forthcoming.
- Hall, Jonathan D. and Ian Savage.** 2019. "Tolling roads to improve reliability." *Journal of Urban Economics*, 113: 103187.
- Hanna, Rema, Gabriel Kreindler, and Benjamin Olken.** 2017. "Citywide effects of high-occupancy vehicle restrictions: Evidence from 'three-in-one' in Jakarta." *Science*. 357(6346): 89-93.
- Hastings, Justine S., and Jesse M. Shapiro.** 2013. "Fungibility and consumer choice: Evidence from commodity price shocks." *The Quarterly Journal of Economics*. 128(4): 1449-1498.
- Janson, Michael, and David Levinson.** 2014. "HOT or not: Driver elasticity to price on the MnPASS HOT lanes." *Research in Transportation Economics* 44: 21-32.
- 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, Trudy Ann Cameron, W. Michael Hanemann, Nick Hanley, Mandy Ryan, Riccardo Scarpa, Roger Tourangeau, and Christian A. Vossler.** 2017. "Contemporary guidance for stated preference studies." *Journal of the Association of Environmental and Resource Economists*. 4(2): 319-405.
- Keeler, Theodore, and Kenneth A. Small.** 1977. "Optimal Peak-Load Pricing, Investment, and Service Levels on Urban Expressways." *Journal of Political Economy* 85 (1): 1-25.

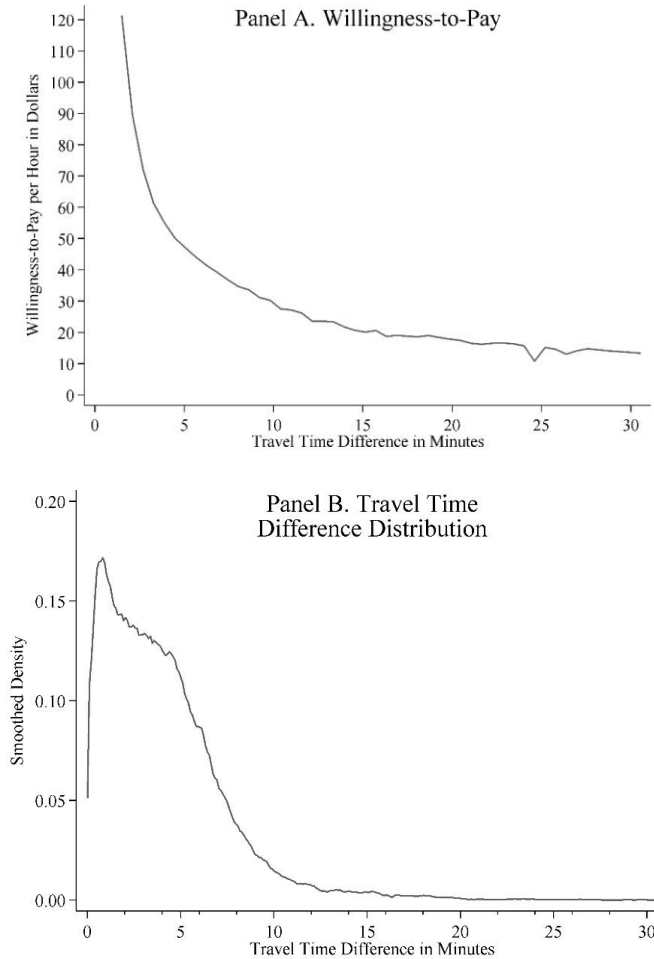
- Kling, Catherine L., Daniel J. Phaneuf, and Jinhua Zhao.** 2012. "From Exxon to BP: has some number become better than no number?" *Journal of Economic Perspectives*. 26(4): 3-26.
- Krindler, Gabriel.** 2019. "The Welfare Effect of Road Congestion Pricing: Experimental Evidence and Equilibrium Implications." Working Paper.
- Kuminoff, Nicolai V., and Abdul Salam Jarrah.** 2010. "A new approach to computing hedonic equilibria and investigating the properties of locational sorting models." *Journal of Urban Economics* 67(3): 322-335.
- 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): 1227-1250.
- Leard, Benjamin and Kevin Roth.** 2019. "Voluntary Exposure Benefits and the Costs of Climate Change." *Journal of the Association of Environmental and Resource Economists*. Forthcoming.
- Leclerc, France, Bernd H. Schmitt, and Laurette Dube.** 1995. "Waiting time and decision making: Is time like money?" *Journal of Consumer Research*. 22(1): 110-119.
- Light, Thomas.** 2009. "Optimal highway design and user welfare under value pricing." *Journal of Urban Economics*. 66(2): 116-124.
- Murphy, Kevin M., & Topel, Robert H.** 2006. "The value of health and longevity." *Journal of Political Economy*. 114(5): 871-904.
- Noland, Robert, and Small, Kenneth A.** 1995. "Travel-time uncertainty, departure time choice, and the cost of morning commutes." *Transportation Research Record*. 1493: 150-158.
- Parry, Ian W.H., and Antonio M. Bento.** 2002. "Estimating the Welfare Effect of Congestion Taxes: The Critical Importance of Other Distortions within the Transport System." *Journal of Urban Economics* 51 (2): 339-65.

- Parry, Ian W.H., and Kenneth A. Small.** 2005. "Does Britain or the United States Have the Right Gasoline Tax?" *American Economic Review*. 95(4): 1276-89.
- Phaneuf, D. J., Kling, C. L., & Herriges, J. A.** 2000. "Estimation and welfare calculations in a generalized corner solution model with an application to recreation demand." *Review of Economics and Statistics*. 82(1): 83-92.
- Roback, Jennifer.** 1982. "Wages, rents, and the quality of life." *Journal of Political Economy*. 90(6): 1257-1278.
- Rosen, Sherwin.** 1974. "Hedonic prices and implicit markets: product differentiation in pure competition." *Journal of Political Economy*. 82(1): 34-55.
- Sahm, Claudia R., Matthew D. Shapiro, and Joel Slemrod.** 2012. "Check in the mail or more in the paycheck: does the effectiveness of fiscal stimulus depend on how it is delivered?" *American Economic Journal: Economic Policy*. 4(3): 216-50.
- Small, Kenneth A.** 1982. "The Scheduling of Consumer Activities: Work Trips." *American Economic Review*. 72 (3): 467-79.
- Small, Kenneth A.** 2012. "Valuation of Travel Time." *Economics of Transportation*. 1(1-2): 2-14.
- 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.
- Thaler, Richard H., and Hersch M. Shefrin.** 1981. "An economic theory of self-control." *Journal of Political Economy*. 89(2): 392-406.
- Timmins, Christopher, and Jennifer Murdock.** 2007. "A revealed preference approach to the measurement of congestion in travel cost models." *Journal of Environmental Economics and Management*. 53(2): 230-249.

- 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.
- Van Benthem, Arthur.** 2015. "What is the optimal speed limit on freeways?" *Journal of Public Economics*. 124: 44-62.
- 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.
- Vickrey, William S.** 1969. "Congestion Theory and Transport Investment." *American Economic Review*. 59(2): 251-60.
- Vickrey, William S.** 1971. "Responsive pricing of public utility services." *The Bell Journal of Economics and Management Science*. 337-346.
- Wolff, Hendrik.** 2014. "Value of time: Speeding behavior and gasoline prices." *Journal of Environmental Economics and Management*. 67(1): 71-88.
- 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.

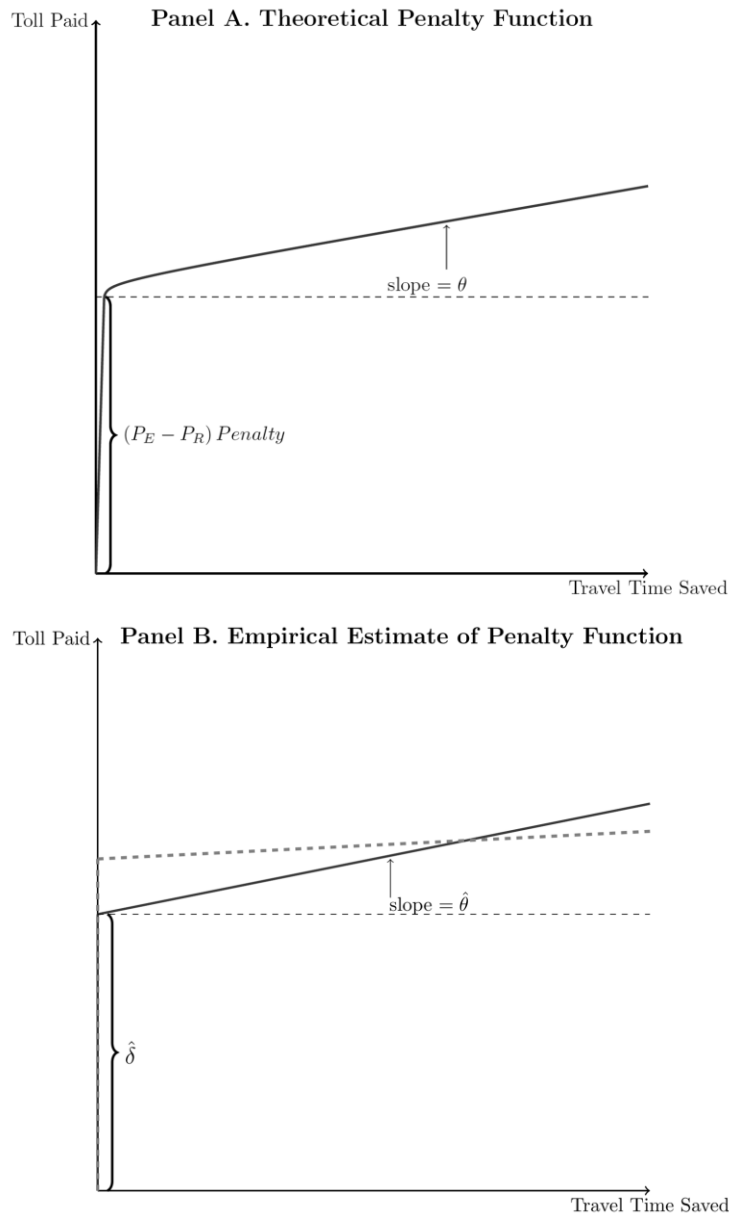


## FIGURES AND TABLES



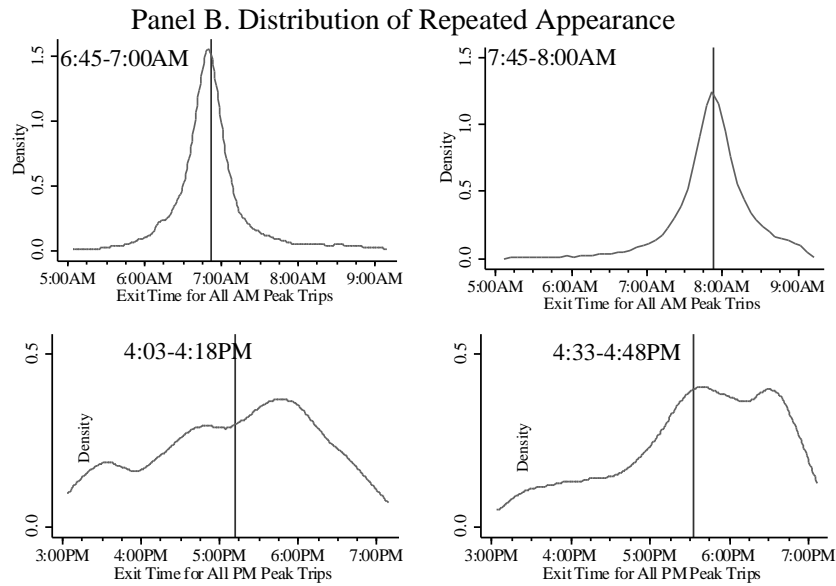
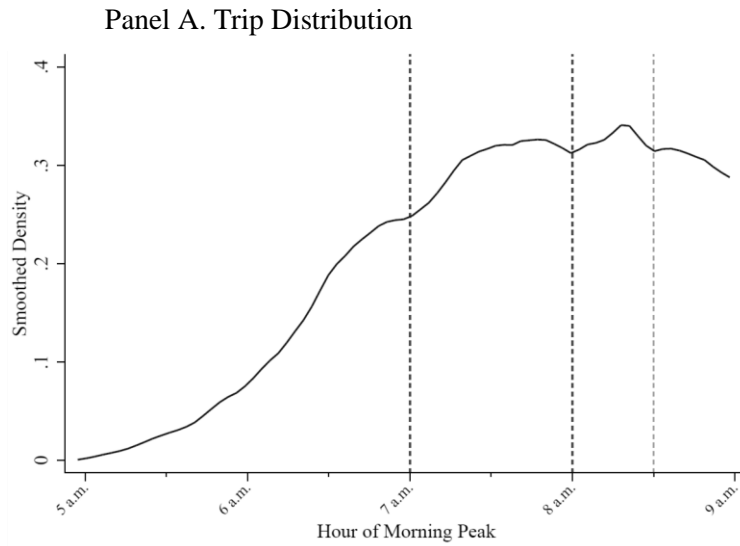
**FIGURE 1. WILLINGNESS-TO-PAY PER HOUR AND DEMAND FOR TRIPS IN THE EXPRESSLANES**

*Notes:* Panel A displays our lower bound estimate of willingness-to-pay for use of the ExpressLanes 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. Panel B displays the smoothed distribution of the trip-level travel time difference between the mainline lanes and the ExpressLanes. The smoother for both 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 of work days in the first 10 months of the policy, 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 C. 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 are dropped. Observations from PeMS where any of the 30 second observations are missing are also dropped.



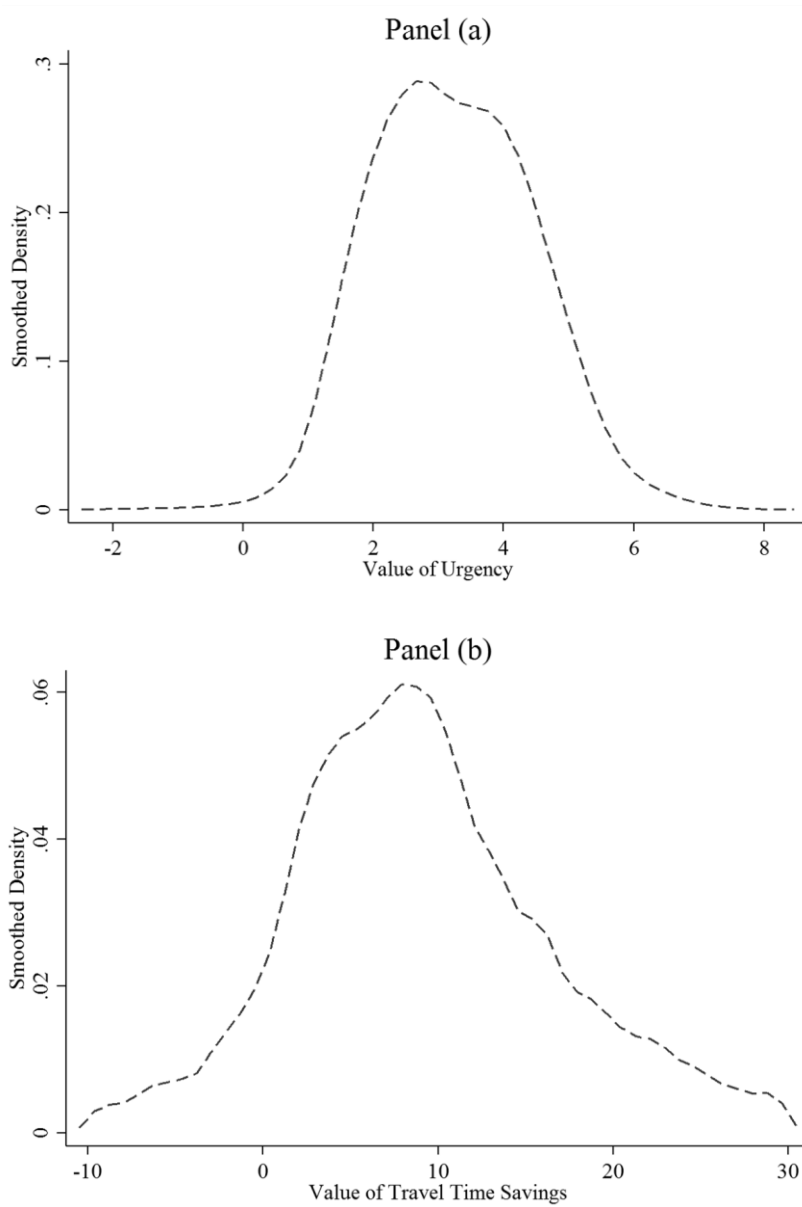
**FIGURE 2. CONCEPTUAL BASIS FOR THE VALUE OF URGENCY**

*Notes:* The figures present the relationship between willingness-to-pay as measured by the toll paid to the travel time saved by entering the ExpressLanes as compared to the conventional untolled lanes. Willingness-to-pay is reflected by a penalty function in Panel A, which grows dramatically initially for extremely small travel time savings, reflecting the high cost of being late, but then changes slope dramatically above the black horizontal line, reflecting the cost of longer travel times in terms of the value of time. Panel B, reflects the empirical analogue measuring the penalty function, where  $\hat{\delta}$  approximates the discrete cost of late arrival associated with urgency and  $\hat{\theta}$  measures the value of travel time savings. The dashed gray line illustrates an empirical estimate where measurement error in travel time savings would mean that the slope parameter is underestimated so that the constant is overestimated.



**FIGURE 3. I-10 W EXPRESSLANES TRIP DENSITY DURING AM PEAK**

*Notes:* The figures plot the kernel smoothed density of trips on the I-10 W ExpressLanes over the morning peak. In panel A, vertical lines correspond to times with a discernible trough in the distribution indicating potential “bunching” around preferred arrival times of 7:00AM, 8:00AM and 8:30AM. In panel B, the figures plot the kernel smoothed density of trip-level Exit Time for accounts with average Exit Time in 15 minute window before 7:00AM, 8:00AM, 4:03AM and 4:33PM, where the latter two times are times displaying bunching during the PM peak. Vertical lines indicate average exit time for trips in given subsample. Includes accounts with 10 or more ExpressLanes trips. 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 are dropped. Observations from PeMS where any of the 30 second observations are missing are also dropped.



**Figure 4. ESTIMATED DISTRIBUTION OF VALUE OF TIME AND URGENCY**  
*Notes:* The figures depict smoothed kernel density estimates of the value of urgency and travel time savings from account-specific regressions of the total toll on the travel time saved and a constant and are consistent with the heterogeneous individual bid curve model. Time, measured in hours, is the time saved by taking the ExpressLanes compared with mainline lanes, from mainline line speeds reported by PeMS, for the chosen trip distance. Observations from morning peak hours are included with weekends and holidays removed.

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

I	II	III	IV	V	VI
Panel A. Trip-Level Characteristics					
	Time Savings				
Decile of Time Savings	in Hours	in Minutes	Average ExpressLanes/ HOV Speed in MPH	Average Mainline Speed in MPH	Average Distance Traveled in Miles
1	0.01	0.39	65.3	60.3	5.8
2	0.02	1.01	67.4	55.9	6.1
3	0.03	1.66	66.6	50.0	6.2
4	0.04	2.37	66.1	44.7	6.1
5	0.05	3.11	66.0	40.6	6.1
6	0.06	3.88	65.8	37.7	6.3
7	0.08	4.69	65.5	34.6	6.3
8	0.09	5.64	64.7	32.7	6.7
9	0.12	6.95	63.8	30.9	7.3
10	0.18	11.04	62.0	25.8	8.1
Average	0.07	4.08	65.3	41.3	6.5
Panel B. Account-Level Characteristics					
Decile of Time Savings	Average Uses per Month	Average Hourly Wage in Zip Code	Average Toll Paid	Modal Vehicle Registered to Account	Average Vehicle Value
1	8.8	\$19.35	\$3.20	Honda Accord	\$9,543
2	9.5	\$19.40	\$3.10	Honda Accord	\$9,512
3	9.8	\$19.47	\$3.12	Honda Accord	\$9,553
4	9.9	\$19.47	\$3.17	Honda Accord	\$9,443
5	9.8	\$19.65	\$3.29	Toyota Camry	\$9,477
6	9.9	\$19.71	\$3.57	Honda Accord	\$9,476
7	9.8	\$19.73	\$3.81	Honda Accord	\$9,523
8	9.8	\$19.76	\$4.15	Honda Accord	\$9,607
9	9.8	\$19.79	\$4.49	Honda Accord	\$9,793
10	9.6	\$20.00	\$4.95	Honda Accord	\$9,943
Average	9.7	\$19.63	\$3.69	Honda Accord	\$9,587

*Notes:* Data cover work days for the morning peak (5-9 AM) from February 25th, 2013 until December 30th, 2013. "Time Savings" is travel time saved driving in the ExpressLanes, calculated as the difference between Metro transponder travel time compared to travel times in mainline lanes from PeMS. "Average Uses per Month" excludes the first month that a transponder appears in the data to control for learning behavior. "Average Hourly Wage in Zip Code" comes from the reported transponder zip code and 2008-12 ACS Census mean zip code data, assuming an assumed average household with two wage-earners and 2,040 working hours per year. "Modal Vehicle Registered to Account" reports the most common vehicle registered to accounts within each decile of time saving. The last row reports the modal vehicle for the whole sample. 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 are dropped. 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—Account Level Regressions I-10 West**

	I	II	III	IV
	Full Sample	Full Sample	After Oct. 20 <sup>th</sup> , 2013	Split Sample IV
Constant	2.84*** (0.48)	3.24*** (0.01) [2.30, 4.05]	3.62*** (0.03) [2.52, 4.65]	2.81*** (0.01) [2.03, 3.53]
Travel Time	8.02** (3.00)	8.19*** (0.12) [3.10, 12.79]	7.16*** (0.27) [-0.68, 14.32]	12.24*** (0.21) [4.40, 19.67]
Reliability	24.76*** (5.24)	17.61*** (0.38) [-1.43, 30.79]	9.41*** (0.62) [-7.59, 25.83]	22.34*** (7.40) [-1.59, 71.55]
Heterogeneous Agent Model		X	X	X
Mean Toll Paid	\$3.73	\$3.71	\$4.01	\$3.60
Urgency's Share of WTP	76%	87%	90%	78%
Mean Time Savings in Minutes	4.08	3.79	4.20	3.42
Mean Reliability in Minutes	0.83	2.1	1.02	0.38
Number of Accounts	28,075	9,053	2,286	6,422

*Notes:* Values shown are the coefficients of regressions of the toll paid on the regressands. Column I reports a single regression of the toll on the expected travel time savings for which there are 433,623 observations, while the remaining columns report statistics from account-level regressions. Column III reports estimates for the period during which signs displaying travel time differences were installed at ExpressLanes entrances. Column IV reports estimates of account-level split-sample IV using 1 hour, 1 week and 2 week leads of time savings by hour, day of week and segment. Values shown in columns II-IV are the average coefficient across regressions, with the interquartile range of values given in brackets. Standard errors of the mean, calculated by randomly sampling from the mean and standard error of individual coefficients 500 times, are given in parenthesis. 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 50<sup>th</sup> and 80<sup>th</sup> quantiles. Observations from morning peak hours (5-9AM) are included with weekends and holidays removed.

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

**Table 3—Account Level Regressions I-10 West: Weekend Control Group**

	I	II	III
	Full Sample	Full Sample	After 10/20/13
Constant	0.79*** (0.01) [0.50, 1.04]	0.79*** (0.02) [0.49, 1.07]	0.50*** (0.04) [0.26, 0.68]
1(Morning Peak)	2.42*** (0.02) [1.46, 3.35]	2.43*** (0.03) [1.46, 3.37]	3.53*** (0.12) [2.12, 4.69]
Travel Time	3.46*** (0.08) [1.29, 5.22]	3.47*** (0.11) [1.13, 5.24]	5.49*** (0.37) [3.69, 7.37]
Travel Time x Morning Peak	7.07*** (0.24) [1.54, 11.06]	5.45*** (0.33) [0.55, 9.39]	1.38*** (0.87) [-6.31, 8.07]
Reliability		2.42*** (0.44) [-5.05, 5.79]	2.83*** (0.43) [-0.09, 3.94]
Reliability x Morning Peak		12.60*** (0.84) [-4.67, 28.83]	-0.33 (1.91) [-16.31, 19.06]
Number of Accounts	1,121	761	62

*Notes:* Values shown are the coefficients of regressions of the toll paid on the regressands. The sample for this regression includes trips taken during weekdays between 5-9AM and trips during all hours of the weekend. 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. Standard errors of the mean, calculated by randomly sampling from the mean and standard error of individual coefficients 500 times, are given in parenthesis. 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 50<sup>th</sup> and 80<sup>th</sup> quantiles. \*\*\* Significant at the 1 percent level. \*\*Significant at the 5 percent level. \*Significant at the 10 percent level.

**Table 4—Homogeneous Agent Models for External Validity**

Panel A: Model with Constant			
	I	II	III
Constant	2.84*** (0.48)	3.21*** (0.98)	3.87** (1.36)
Time in hours	8.02** (3.00)	6.90*** (1.88)	4.58* (2.08)
Reliability	24.76*** (5.24)	17.70** (5.66)	9.81 (8.74)
Trip Restriction	> 0 minutes	> 4 minutes	> 8 minutes
Observations	433,623	188,369	42,277
AIC	1,510,246	695,318	153,230
BIC	1,510,279	695,349	153,255
Panel B: Model without Constant			
Time in hours	30.18*** (3.149)	28.43*** (2.915)	21.56*** (2.907)
Reliability	44.18*** (8.305)	29.55*** (2.829)	23.36*** (1.585)
Trip Restriction	> 0 minutes	> 4 minutes	> 8 minutes
Observations	433,623	188,369	42,277
AIC	1,923,834	792,574	171,281
BIC	1,923,856	792,594	171,298

*Notes:* Values shown are the coefficients of six regressions of the toll paid on the regressands. 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 50<sup>th</sup> and 80<sup>th</sup> quantiles. AIC and BIC are Akaike and Bayesian Information Criteria. Standard errors, clustered by road segment, are in parentheses. Observations from morning peak hours (5-9AM) are included with weekends and holidays removed.

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