The Value of Urgency:
Evidence from Congestion Pricing Experiments

By ANTONIO M. BENTO, KEVIN ROTH, AND ANDREW WAXMAN*

Taking advantage of a program that allows solo-drivers to enter ExpressLanes upon a payment of a toll, we provide the first estimates of commuters' value of urgency, defined as a discrete amount to avoid failing on-time arrival. We provide evidence that, because commuters are schedule constrained, preferences for urgency explain about 70 percent of drivers' willingness to pay to access these ExpressLanes. Earlier theoretical models that ignore preferences for urgency fail to fit the data and explain important empirical regularities. While the value of time and value of reliability have been commonly used for infrastructure project evaluation, our results show that the value of urgency is the critical parameter for evaluation of congestible infrastructure projects where pricing is possible.

First Draft, October 27th 2014
This Draft: December 30th 2014

* Bento: Cornell University and NBER, Charles H Dyson School of Applied Economics and Management. 424 Warren Hall, Ithaca, NY 14853 (email: amb396@cornell.edu); Roth: University of California, Irvine, Department of Economics (email: kroth1@uci.edu); Waxman: Cornell University, Charles H Dyson School of Applied Economics and Management. (emaill: arw227@cornell.edu); We thank seminar participants at Arizona State University, Cornell University, John Hopkins University, and the University of California, Irvine for useful comments.

I. Introduction

There are many instances where individuals are likely to exhibit preferences for urgency. We define the value of urgency as a discrete willingness to pay to jump a queue, and avoid a penalty for failing to meet an important schedule constraint. Examples include the willingness to pay to find a donor of an organ critical for survival, willingness to pay for expedited passport processing, and an automated trading company's willingness to pay to be the first to receive proprietary data feeds from the New York Stock Exchange.[1] While in all these instances the value of urgency must be relatively high, the economics profession has to date ignored urgency preferences in their inquiry to understand key features of human behavior.

Urgency preferences are likely to be particularly prevalent in the transportation sector, where commuters face large congestion costs in their daily journey to work. Recently, in an attempt to improve travel patterns and reduce congestion, policymakers have converted existing High Occupancy Vehicle Lanes (HOV) into ExpressLanes, allowing solo drivers access to them upon the payment of a toll. In this paper, we take advantage of the introduction of this program in Los Angeles, California, to recover the first estimates of commuters' value of urgency. We demonstrate the first-order importance of preferences for urgency, and show using hedonic regression that commuters are willing to pay a fixed $3 dollars per trip to access the ExpressLanes, suggesting that the value of urgency per trip is roughly 15% of local wages. Since individuals are schedule constrained and face discrete penalties for late arrival, a large willingness to pay arises because the

[1] In an effort to be the first to receive news and execute trades based on computer algorithms, firms will pay to have close proximity to trading servers. Tieman, Ross. 2008. "When Microseconds Really Count," *Financial Times*, March 19, 2008.

program allows individuals to purchase the exact time savings necessary for on-time arrival.

The fundamental insight here is that, because the bulk of the trips in the ExpressLanes have surprisingly small time savings (relative to the mainline lanes) and are quite infrequent, the behavior of these agents who also have rather large implied willingness to pay per hour for these trips seems absurd in light of earlier theoretical models. These models fall into two categories: models that are grounded in the concept of value of time, first outlined by Becker (1965), and adapted to measure the value of travel time savings. And scheduling models, first introduced by Vickery (1969), and formalized in Small (1982) and Arnott, de Palma, and Lindsey (1993), where individuals have preferences over schedule delays, and are willing to pay to avoid either early or late arrivals. In these models, the preference parameters that reflect the value of time, and costs of scheduling delays are measured on a per-hour basis. In contrast, the concept of the value of urgency proposed here is a discrete amount, not directly related to the opportunity cost of time or the wage rate, but rather a reflection of a potential serious penalty for being late at all.

When taken to the data, earlier models are rejected as they fail to correctly explain the behavior of the majority of drivers. In fact, more than 30% of drivers have willingness to pay to access the lane above $100 per hour, which is 5 times the hourly local wage of $20. In sharp contrast, a model with a scheduling that includes urgency fits the data extremely well, confirming that what drivers value is on-time arrival not minute-by-minute early or late arrival savings. In this sense, the concept of the value of urgency is more than just a simple generalization of earlier models, and it has fundamental implications for public policy. It is the critical parameter currently missing in any congestible infrastructure project evaluation where part of the infrastructure can be priced. In our case, urgency alone accounts for 81% of the toll revenues generated by the program during the

morning peak, while the portion attributed to travel time savings is less than 19%, and schedule delay preferences play at best a rather minor role. Therefore had a congestible infrastructure project been evaluated ignoring urgency, it may have failed to pass a cost benefit analysis.

To recover commuters' urgency preferences, and illustrate the importance of urgency in cost-benefit analysis of infrastructure projects we have assembled a rich dataset that combines individual level information on ExpressLanes trip speed, distance, and tolls that are linked to transponder users, matched with real time data from the Freeway Performance Measurement System (PeMS), which reports flow and speed for the HOV and mainline lanes. We use these data to calculate the per hour willingness to pay to access the ExpressLanes of trips with different time savings, and recover the implied value of urgency. Guided by the data, we then adapt the scheduling model presented in Arnott et al. (1993) to include urgency. The introduction of urgency in theoretical models allows us to reconcile two key empirical regularities found in the data. First, that the willingness to pay per hour to access the ExpressLanes declines with travel time differentials across lanes. Second, the percent of individuals late is aligned with the data and much smaller than in models without urgency. We then use hedonic regression to estimate the magnitude of urgency and the value of time and use account information to control for potentially confounding unobservable factors. Finally, we illustrate the importance of urgency for the overall welfare and distributional impacts of congestible infrastructure projects by measuring broader multi-market effects that the program can potential generate. The effects estimated allow us to demonstrate that by allowing ExpressLanes users to stay on time, they avoid urgency costs that are one of the most substantial elements in a welfare analysis and are large enough to suggest the optimal ordering of agents across lanes may be different than is currently the case.

II. The ExpressLanes Program

On February 23rd, 2013, Los Angeles converted the High Occupancy Vehicle (HOV) lanes on the I-10 into a High Occupancy Toll (HOT) facility, as part the ExpressLanes program.[2] This was the second such conversion in Los Angeles, the first being the I-110 ExpressLanes, which opened on November 10th, 2012.[3] The goal of the program was to increase the total throughput of these roads and to raise funds to maintain the corridors.

Our study examines the opening of the westbound ExpressLanes on the I-10 running from El Monte to downtown Los Angeles as shown in Map 1. 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® account linked to a required transponder in the vehicle.[4] The ExpressLanes program is a level-of-service pricing system that adjusts prices every five minutes to maintain maximum throughput. Concerned that excessive traffic would reduce incentives to carpool, policymakers mandated a minimum speed of 45 mph and carpools were allowed to continue using the ExpressLanes free of charge.[5] Our central

[2] High Occupancy Toll lanes are also found in Orange County, CA as well as the metropolitan areas of Atlanta, Miami, Minneapolis, Denver, Salt Lake City, Santa Clara, Oakland, San Diego, Seattle, Houston, Dallas and Washington, DC (TTI, 2014).

[3] We limit the scope of our study from the pre-policy expansion of the HOV lanes on December 1st, 2012 to December 31st, 2013 when drivers of Advanced Technology Partial Zero Emissions Vehicles (AT PZEV) were allowed to enter the lane without paying tolls as well. As of July 1st, 2011, the yellow Clean Air Vehicle Sticker (CAVS) program was discontinued, which allowed hybrid vehicles to drive as SOV in HOV lanes in California.

[4] The ExpressLanes function such that once the maximum price is reached the lane is closed to further SOV traffic. The maximum price of $15.00 was established out of concerns about pricing out low income commuters. This maximum was not attained during the study period. Initially the minimum toll was $0.15 but was lowered to $0.10 per mile again based on distributional concerns. Transponder cost involves a roughly $40 deposit for the transponder, a minimum account balance of about $10, and a $1 monthly account maintenance fee, with some variation depending on the payment method. Transponders can be moved between vehicles as long as each vehicle is registered with the program.

[5] 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. FasTrak® transponders associated with the LA ExpressLanes include 3 settings: SOV, HOV-2, and HOV-3+. The first two settings are tolled the same rate during peak hours (5-9 AM, 4-8 PM), while only SOV is tolled during off-peak hours (8 PM-5 AM, 9 AM-4 PM). Because SOV and HOV-2 are tolled the same amount, we interpret the

results focus on the AM peak of the I-10W from 5-9 AM. We restrict most of our attention to this period and direction of travel as the I-10 W has historically had high demand for carpooling during peak hours and requires 3 or more occupants per carpool from 5-9 AM and 4-8 PM.

Drivers may enter or exit the ExpressLanes at 6 locations along the I-10 W, indicated with arrows on Map 1. At these entry points drivers see posted toll rates and once a vehicle enters the lane, the corresponding toll rate for the vehicle is locked in for the duration of its trip even if the price for subsequent vehicles changes.[6] A central component to the design of the ExpressLanes on the I-10 W was the expansion of a second HOV lane along the freeway, allowing for vehicles desiring faster speeds to pass slower ones.[7] This second lane was opened almost three months prior to the start of the ExpressLanes program on that corridor.

III. Data

We have assembled an unusually rich dataset of repeated transaction lane use at both the individual and aggregate level. Roadway flow and speed data by lane over 5 minute increments is matched to information on individual trip speed, distance, and tolls linked by accounts from all transponders used in the ExpressLanes for the period of study. This revealed-preference dataset of all purchases for ExpressLanes access, matched with behavior of drivers in both lanes, provides an unprecedented level of detail for how consumers trade off time

---

smaller share of HOV-2 drivers observed in the ExpressLanes during peak hours (11.7%) as potentially being the result of vehicles with two occupants leaving the transponder in the SOV setting.

[6] 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 the California Highway Patrol officers that patrol the road segment.

[7] These capacity expansions are often coupled with toll lane introductions to ensure that there is a passing lane for drivers operating at faster speeds. While policymakers understandably include this capacity expansion facilitated by the ExpressLanes revenues as part of the welfare created by the project we isolate our study to the conversion of these lanes to an ExpressLanes and discuss the broader impact of this expansion for welfare in the appendix.

and money. These data allow us the ability to observe trips associated with the same transponder account for different price levels, levels of congestion, sub-segments of the ExpressLanes, and control for unobservables specific to an individual commuter, time-of-day or segment of the corridor. While our analysis using these data covers all hours of the day, we focus our study on the morning peak, 5 to 9 AM, when drivers faced with congested roads have little discretion to deviate from the average speed on the road, which is not true when it is in free flow.

A. PeMS Data

We obtain speed and flow data publicly provided by the California Department of Transportation's Freeway Performance Measurement System (PeMS). PeMS generates 5-minute speed and flow data for HOV and mainline lanes from 30 second loop-detector vehicle counts and occupancy.[8] We use data for the 52 detectors along the 10.5-mile road segment of the I-10 W that track the ExpressLanes.[9] One of reasons we analyze the I-10 W is that it has one of the highest detector counts per mile with detectors on average every 0.18 miles. Because missing data is occasionally imputed in the PeMS system, we delete any observations with imputation.[10] To match these observations to each transponder-level transaction, we average speeds from 5-minute observations across detectors by sub-segments of the ExpressLanes. While some regressions limit the sample of observations considered, our full dataset contains 982,056 observations, of which

[8] Lane occupancy is the fraction of time the detector is 'on' due to the presence of a vehicle. Based on average vehicle length and this lane occupancy measure, the speed of traffic is computed. See PeMS FAQ for more information: http://pems.eecs.berkeley.edu/?dnode=Help&content=help_faq. To generate speeds we average across the four mainline lanes and the two HOV lanes.

[9] We exclude all on- and off-ramp detectors.

[10] This is particularly true for the period before the second HOV lane was opened before all data was imputed due to construction.

164,744 occur during the AM peak. We also delete weekends and holidays when travel demand is substantially different than normal work days.[11]

Figure 1 plots average speeds over the hours of the day from December 1st, 2012 until February 22, 2013. While mean speed in the mainline lanes is in excess of 65 mph for most hours of the day it decreases during the morning peak to speeds as low as 45 mph at 7 AM, which is the dominant commuting time for this corridor traveling downtown. The HOV lanes are slower during off peak hours, because passing is more difficult with fewer lanes, but they maintain a relatively fast average speed in excess of 55 mph during the morning peak. The figure also displays the 20th quantile of speed, which shows another benefit of the HOV lanes during the morning peak. While the 20th quantile of speed in the mainline lanes is 25 mph, the HOV lanes only drop to 45 mph. A program that opens the HOV lanes to some mainline traffic using a price may allow individuals with a high cost of travel time to enter the lane to reduce their commute time.

B. Transponder Data

Repeated transponder transactions data on individual trips is collected by Los Angeles Metropolitan Transportation Authority (Metro) and operated by a contract through Xerox. These transponders require drivers to choose the number of occupants and records information on times, points of entry and exit, as well as the toll charged. Our transponder data spanning the period from February 23rd, 2013 to December 31st, 2013, provides information on 7,208,821 trips, which are linked to 285,169 transponder accounts.[12] Travel time savings in the data are

[11] In the appendix, we report key outcomes related to welfare obtained via regression discontinuity estimation of the program's effect on average flow and several moments of the speed distribution, which uses similar data for the same dates and lanes on the I-210 W, a competing route north of the I-10, to control for substitution effects in our regression analysis.

[12] 2,859,808 of these trips are taken on the I-10 W with 1,373,901 trips requiring payment.

calculated as the difference between the distance traveled in the Expresslanes divided by the mainline speed from the PeMS for the corresponding 5-minute interval, minus travel time spent in the ExpressLanes as measured to the millisecond by the transponder readers.[13] The observed total toll ranges from $0.55 to $14.70. This variation in price is due to two key features of the program: tolls adjust to maximize throughput every 5 minutes and drivers can choose to take sub-segments of the total ExpressLanes. This is different than many toll roads with single entry and exit points, and we observe distances ranging from 2.7 to 9.7 miles.

Figure 2 plots the evolution of the price per mile during the peak hours of the morning commute (Keeler and Small, 1977). This figure shows that prices build from $0.45 per mile at 5 AM to $0.55-$0.80 per mile at 7 AM and dropping again to below $0.50 per mile by 9 AM allowing for considerable variation in prices.

Figure 3 plots the total flow data during peak periods and cumulative transponder adoptions from December 2012 through July 2013. Transponder adoption continues over time, rising right around the opening of each corridor, but leveling out eventually. The flow before the opening of the ExpressLanes (from PeMS) is 2,500 3+ carpools, while under the program there are both 3+ carpools and paying SOVs.[14] Of these, roughly 1,500 SOV drivers entered the lanes on February 23rd, 2013, and 1,000 3+ carpoolers remained.[15] Figure 3 also shows that 3+ carpools and SOVs continued to increase over time.[16] Although the total flow

[13] Because the implied distance traveled is the same for the ExpressLanes and mainline travel time calculations, these measures are fully consistent.

[14] For simplicity, we use SOV to include 2+ carpools, which are required to pay the full fare during peak hours.

[15] To calculate aggregate flow from individual transponder data requires some processing. Because most trips are not the full 10.5 miles we sum the total number of miles and divide by 10.5. To verify the data we can compare the average flows in the first 20 work days of the program. While the PeMS and transponder data are recorded by separate regulatory institutions that count cars in very different ways, we find an exceptional degree of alignment between the two sources in the ExpressLanes in the period after the program began. PeMS records a daily average of 2,503 vehicles during the peak, while the transponder data implies 2,575 vehicles. See Appendix Table C.3 for further details.

[16] In discussion with authorities they suggest that some of pre-program HOV vehicles that did not return were violators, which are easier to detect with the ExpressLanes monitoring equipment.

in the ExpressLanes increases, the relative ratio of carpools to SOV drivers remains relatively constant over time. The number of carpools did increase by 500 over the next few months but some carpoolers did not return. This decrease in demand, coupled with the capacity expansion in December 2012 suggests that the ExpressLanes would be in free-flow without the addition of SOVs and well above the minimum speed of 45 mph. [17]

.

IV. Implied Willingness-to-Pay per Hour to Access the ExpressLanes

In the spirit of Becker (1965), suppose for the time being that drivers pay to access this lane to purchase time savings with an opportunity cost of lost wages. We assume that commuters using the ExpressLanes value time at α, and by choosing the ExpressLanes they gain τ hours. Given an observed toll p, by revealed preference, we will observe agents choosing the ExpressLanes when:

$$\alpha \cdot \tau > p \quad (1)$$

To calculate τ we take the difference of the total time recorded by the transponder and the time that would have been required to traverse the same distance at the prevailing mainline speed. We omit from our analysis 6.2% of observations where this time differential was negative. The infrequency of negative time savings is remarkable and not only speaks to the accuracy of our data, but also to how sophisticated these drivers are. Commuters in Los Angeles invest considerable energy in optimizing their commutes and there are many resources available to help them to predict travel times accurately.[18]

[17] In this way it is more representative of the average HOV lane in the U.S., which experiences very low demand compared with most HOV lanes in LA which are near or beyond optimal capacity during peak hours.

[18] There is a long history of real time traffic information in Lost Angeles dating back to the 1940s when crashes would be announced over the radio by Loyd Sigmon; these reports are still known today as Sig Alerts. 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 to provide extremely accurate travel time predictions.

In Table 1, we divide trips into deciles of time savings of 46,624 trips per decile, with mean hourly time savings, in column II.[19] Converted to minutes in column III, they range from 0.39 minutes in decile 1 to 11.04 minutes in decile 10. These represent substantial time savings given that it will take 9 minutes to these distances at speeds as low as 45 mph. To better understand what factors influence the generation of trips, we present statistics detailing the types of trips and drivers in Panel A for each decile. Columns IV and V reveal the source of variation in time savings: while the HOV lanes maintain a relatively constant speed of 62-67 mph, slower mainline speeds generate large time savings. There is relatively little variation across deciles in terms of average trip length in miles, column VI, average uses per month, column VII, and average wage in the account holder's zipcode, column VIII. To the extent that sorting does occur, it seems that compared with other deciles, drivers in decile 1 use the ExpressLanes for the least distance per trip, 5.8 miles, less frequently, 8.8 times per month, and come from zip codes with the lowest implied wage $19.35.[20] To the extent that heterogeneity exists, decile 1 drivers would be assumed to place the lowest value on the ExpressLanes.

In Panel B we use the revealed preference approach to examine the implied Willingness-to-Pay (WTP) per hour of travel time savings. The average total toll is reported in column IV which is $3.69. The implied WTP per hour is reported by decile of time saving in column V. The variation is remarkable. For the lowest decile of time savings, 0.39 minutes, the implied WTP per hour, assuming people are purchasing time savings, is $1,977.44. By contrast those for the longest time

[19] These trips are from the full time period February 23rd, 2013 to December 30th, 2013. One may be concerned that classical supply and demand endogeneity will arise with an analysis of prices in this market. We note that there is generally very little ability of supply to adjust to prices as road capacity is fixed. Transponder adoption by initially displaced HOVs is also unlikely to be influenced by toll prices. The one potential channel of adjustment is that some SOV drivers may switch to carpooling as prices rise however we note that carpooling has generally been found to be extraordinarily inelastic (Bento, Hughes, Kaffine, 2013) and drivers are not able to make this decision once on the road.

[20] The wage is calculated using 2008-12 ACS Census Data information on household income, assuming income is from two individuals working 2,040 hours annually.

differential of 11 min approach $28.47 per hour, a value larger than the local median wage. These extraordinary values for small time savings are not a temporary aberration and occur in the first month of the program at $1,730 per hour as well as six months later in September at $1,220 per hour. This pattern is also echoed in the hyperbolic shape of Figure 4 panel A. At first glance it may seem possible that this very high WTP per hour is the result of a small subset of peculiar individuals who clear the market at unusual levels. In Figure 4 panel B we plot the quantity of trips at each time savings level. We find that not only are these small time savings trips common but they form the bulk of all uses. In particular the behavior of agents saving small amounts of time seems absurd given most value of time measures. This pattern is puzzling because, if nothing else, we would expect the opposite pattern: that drivers with a high WTP per-hour would use the road for more time, longer distances, and more frequently. Indeed, a major distributional concern surrounding these lanes is that they are 'Lexus Lanes', only used by the rich. Given this striking pattern in the data we turn to theories of time use and scheduling to reconcile these findings with alternative theories that model the behavior of commuters. In doing so, we are the first to provide an empirical validation of these models that relies on data that captures the actual behavior of drivers as opposed to behavior solicited by surveys.

V. Theory of Value of Urgency

When measuring the benefits of road infrastructure projects, the literature relies on two classes of models. The first are models that quantify the value of travel time savings, which depend on the concept of value of time, first introduced by Becker (1965). The second are models of the journey to work that, in addition to travel costs, explicit consider scheduling costs (Small 1982, Arnott de Palma, and Lindsey 1990).

Broadly speaking, there are two potential ways of thinking about scheduling costs. First, scheduling costs of the form of a *schedule delay cost*, defined as per hour losses, and directly linked with the opportunity cost of time measured by the wage rate. For example, a loss in hourly wage depending on the number of minutes an individual arrives late or early at work. Second, scheduling costs of the form of a *schedule constraint cost*, defined as a 'one-time' cost of late arrival, a cost not necessarily linked with the wage rate and lost time.[21] The fundamental difference between a schedule delay cost and a schedule constraint cost is that the later does not increase with the total delay. *A priori*, there may be little reason to suspect that scheduling costs take a particular structure. However, to date, the literature has focused on schedule delay costs, and ignored schedule constraint costs. As we shall see below, in general, models that only consider schedule delay costs fit our data poorly. In contrast, a schedule constraint model allow individual to have preferences for urgency, and when presented to pricing options have the opportunity to 'purchase' just the time they need to avoid a penalty for failing to arrive on time.

In this section we provide an overview of competing models that formalize the behavior of commuters, with the goal of testing their ability to explain key patterns found in the data. We begin with models that consider the behavior of a representative agent, given the lack of heterogeneity across the different deciles in demographic, vehicle and trip characteristics documented in Table 1.[22] Below, we also discuss how heterogeneity can affect our central findings.

[21] Other examples would include billing for services, like time with a lawyer or doctor, that began at a given time irrespective of actual arrival time, or cases where fines for late pickup of an item or child from daycare were perfectly prorated.

[22] Further detail, including three most common vehicle models by decile, is given in Appendix Table C.2.

A. Value of Time and Travel Time Savings

The concept of value of time is credited to Becker (1965), with later important refinements applied to a transportation setting by Johnson (1966) and DeSerpa (1971). This concept has been at the heart of infrastructure project evaluation. In the classical Becker model, consumers face both a time and monetary constraint. The key insight of the Becker model is that, when the agent optimizes her choice of work and leisure hours, the shadow value of time becomes the hourly wage. Johnson (1966) notes that this must be discounted by the disutility of work. There is a fairly broad agreement that the value of time of time is roughly half of the wage (Small, 2012).[23]

Consider a route with a free flow HOV lane that allows for a travel time differential relative to a mainline of τ hours. When the HOV lanes are converted into ExpressLanes, mainline drivers may enter the ExpressLanes at a toll p. A mainline driver would start using the ExpressLane whenever the travel time differential between the two lanes satisfies:

$$\alpha > p/\tau \tag{2}$$

Where α denotes the value of time. Although trivial, the implications of using the value of time as the central statistic for infrastructure project evaluation are surprisingly restrictive. First, the marginal willingness to pay per hour to access the ExpressLanes is constant in the time differential between the lanes. Second, if there is heterogeneity in wages, higher value of time individuals are expected to use the lane nearly every day for the full commute length. As we shall see below, neither of these two implications hold in the data, seriously questioning the credibility of the model.

[23] This regularity was found as early as Lave (1969). Other notable estimates in this range include Small (1982) and Deacon and Sonstelie (1985), while Calfee and Winston (1998) found values closer to 19% of the local wage.

B. Scheduling Models

The second type of model, which explicitly incorporates scheduling, builds on the work by Vickery (1969), first formalized by Small (1982) and Arnott et al. (1990, 1993, 1994). Here we briefly review the Vickrey bottleneck model using the framework of Arnott et al. (1990, 1993, and 1994). We focus exclusively on the essential features of the model needed to understand the key results that guide much of the discussion in later sections.

Basic Assumptions—N identical individuals travel from home to work. N is assumed to be fixed, and trip demand is completely inelastic. Travel is uncongested except at a bottleneck with a capacity of s cars per unit of time. If the arrival rate at the bottleneck exceeds s, a queue develops. Travel time from home to work is:[24]

(3) $$T(t) = T^f + T^v(t)$$

Where T^f is free-flow travel time, $T^v(t)$ is variable travel time and t is departure time from home. Let $D(t)$ be the queue length (i.e, number of cars). Then, a driver that departs at time *t* faces a queuing time equals queue length divided by bottleneck capacity:

(4) $$T^v(t) = \frac{D(t)}{s}$$

With $r(t)$ denoting the departure rate function from home, and $\hat{t}$ the most recent time at which there was no queuing, then:

(5) $$D(t) = \int_{\hat{t}}^{t} r(u)du - s(t - \hat{t})$$

[24] Without loss of generality, we assume that T^f equals zero. Thus, an individual arrives at the bottleneck as soon as he leaves home and arrives at work immediately upon leaving the bottleneck.

All individuals have preferred arrival time t^*. The private travel cost function is taken to be linear in travel time and schedule delay, measured by time early or time late[25]:

(6) $$C(t) = \alpha T^v(t) + \beta(time\ early) + \gamma(time\ late)$$

Where α is, as before, the value of time, β is the per-hour unit cost of arriving early at work, and γ is the per-hour unit cost of arriving late at work. Consistent with empirical literature (Small, 1982), we assume that $\gamma > \alpha > \beta$. We refer to $\beta(time\ early) + \gamma(time\ late)$ as the value of *schedule delay costs*.[26]

Each individual decides when to leave home. In doing so, (6) implies that the individual trades off travel time and schedule delay. In addition, individuals are assumed to have full information about the departure time distribution.[27] Equilibrium in the bottleneck model is achieved when no individual can reduce her travel costs by altering her departure time, taking all other drivers' departure times as fixed.

Graphical Representation of the Bottleneck Equilibrium—The equilibrium is depicted in Figure 5. The beginning of the rush hour is denoted by t_q (that is, the departure time of the first individual), and $t_{q'}$ the end of the rush hour. Let $\tilde{t}$ represent the departure time of the individual that arrives just on-time (at t^*). Agents who depart after $\tilde{t}$ arrive late. Conversely, agents who depart before $\tilde{t}$ arrive early. Therefore, the individual who departs at $\tilde{t}$ is the only individual who faces no scheduling costs. The vertical distance between the cumulative departures schedule and the cumulative arrivals schedule is queue length in cars and the horizontal distance is travel time (denoted as $D(t')$ and $T^v(t')$,

[25] Consistent with the literature, we assume that the travel cost function is linear for analytical exposition. In the empirical section below, we generalize this function.

[26] Note that time early equals $Max\ [0, t^* - t - T^v(t)]$, and time late equals $Max\ [0, t + T^v(t) - t^*]$

[27] While this assumption may not always be realistic where traveling to an unfamiliar location ours is a setting where drivers commute regularly, have a wide range of traffic information and are making a decision having already observed congestion.

respectively, in the figure). Cumulative departures for agents who arrive before $t^*$are shown in segment AB (with slope $\frac{\alpha s}{\alpha-\beta}$).[28] For agents who will arrive after t^*, cumulative departures are given by BC (with slope $\frac{\alpha s}{\alpha+\gamma}$). In turn, cumulative arrivals are displayed by AC, which rise with slope s. The maximum travel occurs for the agent who departs at $\tilde{t}$, and arrives exactly at t^*. The queue builds up at a constant rate from t_q, when the first individual leaves, until $\tilde{t}$. The queue then dissipates, again at a constant rate, reaching zero at $t_{q'}$when the last person departs.

Since the first individual to depart at t_q and the last individual to depart at $t_{q'}$incur only schedule delay costs, the following must hold in equilibrium:

(7) $$\beta\left(t^* - t_q\right) = \gamma\left(t_{q'} - t^*\right)$$

Further, since the bottleneck operates at capacity throughout the rush hour, and the length of the rush hour is $\frac{N}{s}$:

(8) $$t_{q'} = t_q + \frac{N}{S}$$

These imply that the first person leaves home at:

(9) $$t_q = t^* - \frac{\gamma}{\beta+\gamma}\frac{N}{s}$$

And the last individual leaves at:

(10) $$t_{q'} = t^* + \frac{\beta}{\beta+\gamma}\frac{N}{s}$$

The peak individual, who arrives at exactly t^* , leaves home at $\tilde{t}$:

(11) $$\tilde{t} = t^* - \frac{\beta}{\alpha}\frac{\gamma}{(\beta+\gamma)}\frac{N}{s}$$

And the resulting fraction of late individuals in this model is given by:

(12) $$\frac{\beta}{\beta+\gamma}$$

[28] To calculate the slope of segment AB note that, the cost of an early arrival trip is $\alpha T^v(t) + \beta[t^* - t - T^v(t)]$. Total differentiation of (4) and (5) with respect to t and using (4), it follows that $r(t) = \frac{\alpha s}{\alpha-\beta}$

which with the standard ratio of parameters from the literature $\beta:\gamma = 1:4$ would imply that twenty percent of individuals would be late.

Implications of the Bottleneck Model with Schedule Delays—So far we have only considered the possibility that the road is a single lane that is congested during the rush hour. We now allow for the possibility that the road also has free flow ExpressLanes, and consider the case of a solo driver who can pay a toll. If it is always the case that $\pi > \alpha T^v(t) + \beta(t^* - t)$ for $t \in [t_q, \tilde{t}]$, then no early drivers are willing to pay the toll. In contrast, for an individual who is late and arrives at time $\bar{\bar{t}}$, the willingness to pay to access the ExpressLanes and arrive on time is $(\alpha + \gamma)(\bar{\bar{t}} - t^*)$. Therefore, and contrary to the pattern found in Figure 4, the willingness to pay per hour to access the ExpressLanes would simply be $\alpha + \gamma$, a constant that at best can only approximate the behavior of individuals for which the time differential between the mainline and HOV lane is relatively high.

C. Bottleneck Models with Scheduled Constraint and the Value of Urgency

We now generalize the Arnott et al. (1993) model to explicitly consider a schedule constraint, which in turn allows individuals to reveal preferences for urgency. In the presence of a schedule constraint, the private costs of a trip become:

$$C(t) = \alpha T^v(t) + \beta(early\ time) + \gamma(late\ time) + \delta(being\ late) \tag{13}$$

We refer to δ as the value of urgency. As before, we can proceed to find the first and last individual in the rush hour, and the peak individual who arrives just on time. Similarly to (7), the first and last drivers must be indifferent, leading to:

$$\beta(t^* - t_q) = \gamma(t_{q'} - t^*) - \delta \tag{14}$$

And (8), (9), and (10) become:

(15) $$t_q = t^* - \frac{\gamma}{\beta+\gamma}\frac{N}{s} - \frac{\delta}{\beta+\gamma}$$

(16) $$t_{q'} = t^* + \frac{\beta}{\beta+\gamma}\frac{N}{s} - \frac{\delta}{\beta+\gamma}$$

And

(17) $$\tilde{t} = t^* - \frac{\beta}{\alpha}\frac{\gamma}{\beta+\gamma}\frac{N}{s} - \frac{\beta}{\alpha}\frac{\delta}{\beta+\gamma}$$

The introduction of a scheduling constraint alters the equilibrium in several important ways. First, rush hour starts and ends earlier by $\frac{\delta}{\beta+\gamma}$. The individual that arrives just on time also leaves earlier in a schedule constraint model, but only by $\frac{\beta}{\alpha}\frac{\delta}{\beta+\gamma}$. As a result, the cumulative departures up to $\tilde{t}$ are substantially higher than in a model without a schedule constraint. Perhaps more interestingly, the presence of a discrete penalty for being late causes the queue to immediately dissipate after $\tilde{t}$. In fact, by virtue of the Nash equilibrium there will be a time period immediately after $\tilde{t}$ for which no new drivers enter the queue. Consider hypothetically a driver that could have chosen to depart at $\tilde{t} + \varepsilon$, as ε converges to zero in the limit one can ignore schedule delay costs. It is easy to demonstrate that:

(18) $$\alpha(t^* - \tilde{t}) \neq \alpha\big(t^* - (\tilde{t} + \varepsilon)\big) + \delta$$

precisely because of the presence of the discrete penalty for being late, the next individual to depart after $\tilde{t}$ will only depart at:

(19) $$\breve{t} = \tilde{t} + \frac{\delta}{\alpha}.$$

After $\breve{t}$ the queue starts building again. The intuition is rather simple. Since individuals that fail to depart by $\tilde{t}$ will be late and incur a cost of δ, it becomes optimal for some of them to actually depart later, creating a discontinuity in the second segment of the peak. The equilibrium is depicted in Figure 5 panel B.

We also note that the introduction of δ fundamentally alters the prediction of the fraction of individuals that are late in the model. This becomes:

(20) $\frac{\beta}{\beta+\gamma} - \frac{\delta/(\beta+\gamma)}{N/s}$

As discussed in the next section, with our estimate of $\delta = \$3$, lower values of β and γ and a rush hour of 4 hours, the percent of late individuals will decrease to about 7%.

Implications of the Bottleneck Model with a Schedule Constraint—Now consider a road with free flow ExpressLanes. Assuming that the toll is higher than $\alpha - \beta$, no early drivers are willing to pay the toll and late drivers continue to use the mainline lanes until the last possible second that switching to the ExpressLanes will get them to their destination at time t^*. An agent who leaves at time t will be willing to pay $(\alpha + \gamma) \cdot [(t + D(t)s) - t^*] + \delta$ to avoid mainline travel of $[(t + D(t)s) - t^*]$. This individual will use the mainline lanes until time almost t^* and then pay the toll to arrive at t^*. That is, if this individual saves τ minutes, her willingness to pay is $\delta + (\alpha + \gamma)\tau$ implying a WTP per hour of $\delta/\tau + (\alpha + \gamma)$.

A major insight of including urgency in the bottleneck model is that the resulting willingness to pay her hour is declining in τ, the time differential between lanes giving rise to the shape of the distribution of willingness to pay found in Figure 4 panel A.

VI. Empirical Evidence for the Value of Urgency

A. Empirical Strategy

Total toll paid is regressed on a constant and a function, *f*, of the expected travel time difference between the mainline lanes and the ExpressLanes:

(21) $TOLL_{i,s,t} = \theta_{0,t} + \theta_{1,t} f\left(TravelTimeSaved_{s,t}\right) + \theta_2 X_{i,s,t} + \mu + \xi_i + \varepsilon_{i,s,t}.$

Here *i* indexes individual commuters, *t* indexes time of day by peak, off-peak, and weekends, and *s* indexes sub-segments of the ExpressLanes corridor. The travel time saved on segments *s* is calculated as discussed in the previous section. The coefficient of interest, $\theta_{0,t}$, will be the estimate of the value of urgency, which may vary by time of day.. In our baseline specification, *TravelTimeSaved* enters the regression linearly, multiplied by a parameter $\theta_{1,t}$ which will include value of time, α, and schedule delay late costs, γ.[29] Theory does not dictate the shape of f and we also examine the fit of higher order terms (Cropper, Deck, and McConnell, 1988). Other trip characteristics are included in vector X. Because we pair a segment with the nearest mainline detector to measure these speeds we cluster the standard errors by segment.[30]

Least-squares estimation is consistent if the explanatory variables are exogenous conditional on μ and ξ_i, that is

$$E\{\varepsilon_{i,s,t}|X_{i,s,t}, TravelTimeSaved_{s,t}, \mu, \xi_i\} = 0. \quad (22)$$

After including covariates there may remain unexplained variation in the total toll paid because many of the characteristics that determine the value of a trip could be unobserved. We allow this unobserved variation to take the form of lane-specific unobservables μ, common to all drivers, as well as individual-specific unobservables, ξ_i. An appealing feature of panel data in hedonic models is that it is possible to link individuals across different sales to control for time-invariant unobservables (Davis, 2004; Figlio and Lucas, 2004; Brown, 1980). [31] However, an individual fixed effect will eliminate individual-specific, time-invariant

[29] Because early drivers are unlikely to use the ExpressLanes we cannot identify β.

[30] In Appendix Table C.9 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.

[31] In hedonic applications there is often the concern for unobservable characteristic of both the product and the buyer. Applications in the housing market often make use of house- or neighborhood-level fixed effects to remove unobserved covariates of the product. Unobservable characteristics of the buyer tend to be a larger concern in recovering the value of a statistical life, where employees who accept risky jobs may differ from those choosing safer jobs (Viscusi and Aldy, 2003) which can be removed with an individual fixed effect.

urgency. We can, however bound the minimum time-varying urgency at the morning peak when urgency is likely to be highest by introducing an account fixed effect. Punishment for late arrival can occur outside of the morning peak[32] but peak commuting is likeliest to consist of work commutes with set start times and a large potential punishment for late arrival. Restricting our sample to individuals who use the ExpressLanes during the morning peak and the weekend we use the weekend trips as a control to remove the influence of μ and ξ_i on our estimated coefficients. For any remaining omitted variable to explain our findings it would need to exist during peak hours irrespective of road speeds and be absent during trips during the weekend.

In the vector of observables, $X_{i,s,t}$, we also introduce a measure of reliability of travel time. Reliability is often highlighted in the transportation literature as a willingness to pay for certainty in travel time. Purchasing access to the ExpressLanes may act as a form of insurance against the possibility of extremely long travel times as measured by the difference between the median and the 20th quantile of speed over the segment in that month (Brownstone and Small, 2005).

B. Results

Initially we focus on the morning peak when we would expect drivers to be the most schedule constrained. Table 2 reports the least-squares estimation of the value of urgency during the morning peak.[33] In column I, our specification that

[32] Examples include missing the beginning of shift work, airline flights, picking up children from school or daycare, restaurant or entertainment reservations, and business meetings.

[33] The morning peak is when 51.8% of all ExpressLanes trips occur and 69.3% of all revenue is collected. While this period of time is the most congested, it represents 26.1% of the total daily flow as other times of day carry large amounts of traffic at faster speeds. This congestion is also helpful in assuring measurement accuracy because average mainline speed may not reflect the actual speed of the driver if traffic is sufficiently low. During off peak times drivers may be able to pass and have a lower travel time in the mainline lanes than would be implied by the average speed.

most closely follows the bottleneck model,[34] both the constant and time coefficient are statically significant at the 1% level. The estimate of the constant shows that commuters are willing to pay a fixed $2.94 for the use of the ExpressLanes, regardless of how much time is actually saved. With an average toll of $3.65, urgency represents 81% of the total toll. The estimated coefficient on the time saved is $11.05 in column I.[35]

In column II we allow for a quadratic in time savings to examine the possibility that our estimated constant is due to the assumption of linearity in travel time savings.[36] Using this more flexible functional form, the estimated constant changes only slightly to a statistically significant $2.82.[37] As can be noted from the AIC and BIC, this does improve the fit but to a very small extent.[38] We compare this result to a model estimated in column III where the constant is restricted to be zero. This regression would suggest drivers value travel time savings at $37.59 and the fit is substantially worse than the model estimated in column I that includes the constant.[39] The fit is improved with the addition of a

[34] Another confirmation that individuals are optimizing over trip length as discussed in the bottleneck model is that as drivers' exit times grow later, they chose longer segments of the ExpressLanes. In Appendix Table C.18 we regress distance on exit time and other covariates finding that their average distance grows by 0.25 miles for each hour later they are. While imperfect this regression demonstrates that their choice seems dependent upon on their schedule. Other statistics on segment choice are given in Appendix Tables C.6 and C.15.

[35] In Appendix Table C.14 we examine the robustness of our results to possible design and measurement errors. We find that the most problematic errors are if all travel time savings are consistently under-measured by 14 minutes. We observe no trips of even 7 minutes in lost travel time. If the measured mainline lanes speed is randomly generated we will find a statistically significant constant; however, the explanatory power will be much lower than we find in our regressions. In Appendix Table C.6 we present an entry-exit matrix of trip use, which suggests a large number of drivers enter at the first possible toll entry. In Appendix Table C.15 we display segment-level regressions clustered at the week level. Five of the seven sub-segments with a substantial number of trips display coefficients that are similar to the full regression. The two outliers (segments 10 and 17) display much lower travel time savings coefficients with very low R-squared indicating mainline lanes speeds may not be an accurate reflection of travel time savings for these particular sub-segments.

[36] In Appendix Table C.11 we examine other nonlinear models that recover similar estimates of the constant.

[37] While the time saving terms are not individually significant, they are jointly significant at the <.1% level.

[38] In Appendix Table C.16 we present evidence that there may be heterogeneity in urgency by examining off-peak times of day, weekends, and other routes. Off-peak times of day generally display a lower value of urgency, although the parameter remains above $0.70 for all regressions.

[39] There is an additional conceptual issue with assuming a model without urgency. In Table 1 panel B, the 11,216 accounts that generate these 46,624 trips would have a VOT of $1,977 per hour. The wage required to induce this VOT would imply an annual income of nearly $8 million per year. Reducing this WTP per hour by two-thirds, assuming that this value is inflated by γ, would only reduce the VOT to $659.12 per hour implying an annual income of $2.69 million. This

quadratic term, in column IV, but it is still inferior to the models that include a constant.[40]

In columns IV through VI, we introduce a measure of reliability (Brownstone and Small, 2005; Small, Winston, and Yan, 2005) to capture this preference for certainty in arrival time. The minimum speed restriction of the ExpressLanes may allow drivers to avoid the unreliability of the mainline lanes. We find that this decreases our estimate of the constant to \$2.60 and the estimated coefficient on travel time savings to \$7.30 but these estimates are statistically indistinguishable from our base specification in column I. These estimates show that consumers place a relatively high valuation on reliability, \$32.09 for each hour of difference between the 50th and 80th quantile of travel time in column I; however, with the mean reliability of 0.018 hours, this accounts for only \$0.58 of the WTP for access to the ExpressLanes for these trips. Because reliability may motivate a driver to use the ExpressLanes even when they are slower than the mainline lanes, in column VI we retain the trips with negative travel time. We find that including these trips does not change the estimated parameters.

C. Other Unobserved Attributes

To estimate the value of urgency requires that there be no other attribute that influences demand on the road that is invariant to the amount of travel time saved. For example one may be concerned that demand for the ExpressLanes is due to

result seems at odds with the fact that these drivers come from lower average income zip codes than other deciles and that the car makes and models most common in this decile are the Honda Accord, Honda Civic and Toyota Camry.

[40] There is an additional concern with this model that the estimate of the squared term on travel time saved indicates that in the range of travel time savings we observe, drivers face lower total costs from longer delays than shorter ones. It is difficult to justify costs that decline in travel time in a theoretical model. While models with a declining, but positive marginal cost function as in column II seem possible, it is very difficult to find a setting where the penalty for late arrival would follow this negative marginal cost structure. In Appendix Table C.11 we examine polynomials up to the 5th order, which always contain regions where total cost is decreasing as the delay increases. While models with a constant such as that in column VI also have negative terms, the decreasing portion of the cost function falls outside of the range of time savings we observe in our data.

some other characteristic like road quality, or safety rather than urgency.[41] In this section we attempt to bound the size of these unobserved time-invariant factors using trips that are less likely to be schedule constrained as a control group.

In Table 3 reports least squares and fixed effects regressions of the urgency premium at different times of day and on different roads. Panel A shows that urgency on the I-10 W during the afternoon peak, is $2.35, $1.52 during off-peak weekday hours, and $0.70 on the weekends.[42] We also find that urgency is comparatively high on other routes during the dominant commuting time. We find that urgency is $2.26 on the I-10 East in the afternoon, $3.58 during the morning peak on the I-110 N, and $2.38 on the I-110 S in the afternoon peak.

In Panel B we estimate models that use weekend trips as controls for weekday trips allowing the introduction of account and transponder fixed effects regressions. [43] Because weekend trips may also have urgency, this estimate will be downward biased by the degree of urgency on weekend trips. Column I gives the baseline estimate that the premium on the morning peak is $2.07. Introducing account fixed effects reduces the estimated coefficient to $1.89. Columns III through VI examine the robustness by introducing transponder fixed effects and measures of reliability. The observed morning peak premium is stable ranging from $1.88 to $1.68.

Finally if what the constant was capturing was road quality or safety, or any other non-time base amenity, this would generate a WTP for access to the

[41] Safety, in particular, seems like an unlikely candidate because prior empirical work shows that safety is lower in HOV lanes (Golob et al., 1989). Lower safety is a natural consequence of two of the primary attractions to the lane, faster speeds and a large time differential with neighboring mainline lanes.

[42] In Appendix Table C.17 we examine the eastbound direction of the I-10 and both north and southbound directions of the I-110. There is also a literature that uses convenience goods to reveal the value of time (Phaneuf, 2011). Drivers might make a long run decision that using the lanes repeatedly gives them a larger daily time budget, rather than to meet a particular scheduling goal as assumed here. If this were true, drivers would be expected to use the lane nearly every day in both directions. We find that 62% of all drivers who use the lane in the morning do not use it in the afternoon to return that same day, and more than 64% of all trips are from divers that use the lanes less than 10 times per month on average.

[43] There may be multiple transponders linked to a single account and not all accounts list a transponder. Trips without a listed transponder are omitted from this specification.

ExpressLanes that was independent of the travel time savings. Valuing these types of amenities we would expect to see demand for the ExpressLanes even when the time savings was negative, which Figure 1 shows occurs often at 5 and 6 A.M. If the $3 we attribute to urgency was some other non-time base amenity, a driver with a $11.05 VOT would be willing to tolerate an 16 minute delay in the ExpressLanes.[44] Only 6.2% of all morning peak observations display negative time savings. The largest delay anyone is willing to endure is 7 minutes and 24 seconds and the average is delay is 28 seconds. The longest delay valued at the $11.05 per hour suggests that any trip invariant attribute that is not urgency is worth at most $1.36 suggesting that the lower bound on urgency at this time of day is $1.58.

D. Recovering Drivers Preferences for Schedule Delay

The estimates above show that the value of urgency, δ, is roughly $3, but it is unclear what values of α and γ are most appropriate. In Section V we derived the number of late individuals using the bottleneck model based on the length of the rush hour and the baseline parameters of β, γ and δ. Prior to the policy there were 22,343 drivers per morning peak in the mainline lanes and 1,626 shifted to the ExpressLanes once the program began. Under the assumption that these drivers are late this implies that 7% of drivers have late arrival, which is roughly 1.5 late arrivals per worker per month. With a 4 hour peak, $3 urgency cost from the estimates above and the traditional 1:4 ratio of β:γ, solving for γ gives a value of $4.50. [45]This would imply that the value of time is $6.55 at least. With a local wage of $19.63, this is 33% of the local wage. We can also bound γ from below

[44] The time savings of trips with negative time savings are far below this level. 87% of trips with travel time losses are less than one minute and the largest loss recorded is 8 minutes.

[45] Note that this calculation is not dependent upon the value of time.

by solving for the value that results in no late individuals. In this case γ would be $3 and the value of time would be at most $8.00 or 42% of the local wage.

Prior literature (Small, 2012) has generally found that α is roughly half the local wage and γ is twice α, with δ assumed to be zero. Given the local wage this would imply α would be $10 and γ $20. These values are considerably higher than what emerges from our estimates. Following these we would have expected the coefficient on travel time would be nearly $30. One common feature of many studies is that they have used stated choice surveys where drivers are given travel time improvements of 5 to 10 minutes.[46] For example Small, Winston, and Yan (2005) allow drivers the choice between a free road with a travel time of 25 minutes and a fully separated Express Lane road with a travel time of 15 minutes and a fixed toll of $3.75 where drivers must commit to the full length of the road.[47]

In Table 4 we estimate models that assume δ is zero and impose a minimum time savings of more than 5 or 10 minutes. When limited to trips of more than 5 minutes in column I the coefficient on travel time becomes $31.22 and when limited to more than 10 minutes in column III, the coefficient decreases to $21.68. These results suggest that limiting the range of time savings to more than 5 minutes, may have resulted in a measure of schedule delay that absorbed urgency and biased this measure upward.[48] Columns II and IV show that such limiting the time differential in a model with urgency gives results that are statistically indistinguishable from those of Table 2 column I.

[46] These preferences are usually elicited with a survey presenting various toll levels allowing access to a separate road. On many toll roads drivers gain amounts of time that are larger than we often observe here because they must commit to the full length of the toll road.

[47] This reflects the conditions on prior toll roads that are completely separated from neighboring free lanes inducing larger travel time savings and price schedules that may have peak pricing at a fixed, predetermined rate.

[48] This result can also be seen in Panel A of Figure 4. If limited to trips of more than 5 or 10 minutes the WTP per hour appears to be largely flat and in the range of $20-$30 per hour.

E. Further Discussion

Using the values estimated here for welfare analysis requires assumptions on the underlying heterogeneity in the value of urgency, and value of time.[49] Above we assumed that these values were identical for all agents or identical within a time of day. In Table 1 we observed very little observable heterogeneity across the various deciles but if heterogeneity exists and sorting occurs, this may bias estimates above. It may also result in selection on drivers with unusual values that do not apply to the broader population.

One may be concerned that there is a correlation between the price that clears the market and the types of drivers we observe individuals in the lane based on commuting times. As noted in Table 1, a large time differential will arise from high mainline demand and this is likely to coincide with high carpool demand. This will result in a large pool of mainline drivers to draw from and relatively little capacity in the ExpressLanes. However, this would imply the highest WTP per hour, with or without the inclusion of γ, would clear the market when the time differential is large, the opposite of what we observe here.[50]

One benefit of our ability to link multiple trips by transponder is that we are able to link trips of different time savings to a single vehicle by transponder, which reduces the possibility that our effects are due to heterogeneity affecting our estimated results. Figure 6 shows the kernel density estimates of coefficients for the constant and travel time savings, respectively, from separate regression following equation 22 for each transponder account. We can see that there is some

[49] There could also be heterogeneity in preferred arrival time and segment choice however these should not affect our result because if individuals are unable to recover enough time to avoid paying delta our regression should not estimate a statistically significant constant.

[50] A theoretical argument also undermines this as a legitimate concern. When choosing departure times, drivers who are the most sensitive to early and late arrival, that is high β and γ, should travel at the peak, as in Arnott, et al. (1994). Peak times are generally associated with a large time differential implying that drivers with the highest values of γ should disproportionately show up when the time differential is largest, again inducing positive correlation, which is not what we observe.

heterogeneity in the implied value of urgency and value of time, but the central tendencies of these distributions coincide with what we report in Table 2.[51]

We do however note that considerable care must be used in applying these estimates more broadly because of heterogeneity. Our framework suggests that the presence of urgency will affect all drivers departure times and commuting patterns, only a subset will be late on any given day and this fraction is likely to be largest during times of day with a substantial schedule constraint.

The fact that most prior literature has been unable to differentiate urgency from other types of costs is important beyond accurately forecasting toll revenues. It has substantial implications for cost benefit analysis. In the next section we attempt to show how these estimates of urgency affect the costs and benefits of infrastructure projects like the ExpressLanes, paying particular attention to which drivers are urgent and which are not.

VII. Measuring the Implications of Urgency for Welfare

A. Conceptual Framework

In the spirit of the literature on taxation in a second-best setting (Harberger, 1974; Bovenberg and Goulder, 1996; Parry and Small, 2005; and Parry and Small, 2009), here we outline the implications of preferences for urgency for evaluating the welfare effects of the ExpressLanes program with the aid of simple diagrams. We provide more details and mathematical formulas used to calculate

[51] An underlying rationale for sorting may be heterogeneity in income. As the value of time is tied to a worker's hourly wage and costs for late arrival consistent with urgency may also scale with income, this may generate heterogeneity in in the marginal willingness-to-pay. While we do not observe the household income of ExpressLanes drivers, we are able to tie them to the average income in the zip code to which they reside from 2008-12 ACS Census Data. If consumers are paying for the time they save, wages should correlate with WTP expressed on a per hour basis. If instead they are paying to avoid urgency costs, wages should correlate with the total toll. When we regress WTP per hour on zip code income we find a marginally significant coefficient of *negative* 11.58. If we regress the average total toll on the wage in the account holder's zip code we find a highly statistically significant coefficient of 0.01.

these effects in Appendix B. Following the model outlined in section V, consider a transportation network with a fixed number of agents who select between the mainline and the HOV lanes (Vickrey, 1969). In both markets, distortions stem from the failure of agents to consider external congestion and air pollution costs generating a wedge between the marginal private cost and the marginal social cost of traveling. By allowing solo drivers into the HOV lane, the ExpressLanes program generates a tension between congestion relief benefits in the mainline lanes and potential congestion costs in the HOV lane.[52]

Figure 7 panel A depicts the demand for SOV travel as a function of the toll. Panels B and C depict the equilibrium in the HOV and mainline lanes, which we assume to be the only distorted markets in the economy. Suppose a SOV enters the HOV lane. The marginal welfare effect of allowing solo drivers into the HOV lane equals the sum of the shaded areas in each of the three panels of Figure 7 (see Appendix B for a mathematical derivation: First, area ***abcd*** denotes the primary welfare gain of the program. It equals the willingness to pay to access the lane. Because we only observe the tolls, the ExpressLanes toll revenues serve as a lower bound of the primary welfare gain of the program. Second, the area ***efg*** denotes the *direct congestion interaction effect* of the program. This is the potential welfare loss to existing carpoolers that results from lower speeds when solo drivers are allowed in the HOV lanes. Panel C represents the welfare gain from the ***secondary congestion interaction effect***. This is depicted by the area ***hijk***, and equals the wedge between the social and private costs of traveling in the mainline multiplied by the number of drivers that moved from the mainline to the HOV lane.

[52] To alleviate these costs in the HOV lane, the social planner expanded HOV lane capacity prior to the beginning of the program (as described in section II), and regulates the infrastructure with a real-time varying toll that secures the speed in the HOV lane not to fall below 45 mph.

In reality, the number of agents is not fixed, and the congestion relief in the mainline may induce demand from other transportation options.[53] Therefore, the area ***hijk*** should be interpreted as a hypothetical *partial equilibrium congestion interaction effect*, which we argue represents the upper bound of the congestion relief benefits for all other drivers in the freeway system. This hypothetical partial equilibrium welfare effect is exclusively calculated from the number of solo drivers that leave the mainline for the HOV lane, without allowing the flow in the mainline to re-adjust. That is, the change in the number of drivers that leave the mainline for the HOV are linked to the change in speed in the mainline, following standard speed-flow relationships (Burger and Kaffine, 2009; Bento et al. 2014)[54]. Because at peak periods the external costs of congestion are rather large, even the removal of a relatively small number of drivers can increase speed, and generate a large welfare gain (Anderson, 2014). In Appendix B, we demonstrate the conditions under which the partial equilibrium congestion interaction effect will upper-bound the general equilibrium system-wide congestion interaction effect[55].

The welfare effects of the program that stem from changes in pollution are slightly harder to calculate. Creating congestion relief benefits within the freeway system leads to reallocation of agents across freeways and potentially new trips and vehicle miles traveled (Hymel, Small, and Van Dender, 2010; Duranton and Turner, 2011). To the extent that the program induces new vehicle miles traveled, it may generate an additional negative source of welfare corresponding to the value of the external costs of emissions associated with the trips. At the same

[53] For example, drivers on less congested alternative routes will now replace the existing solo drivers on the congested travel route, dissipating the congestion relief for the original drivers. In turn, they may be replaced by drivers from backroads, or even new vehicle miles traveled. As in Bento et al. (2014), we assume new VMT accounts for 15 percent of these trips.

[54] See Appendix B for the estimation of the speed-flow relationship. The elasticity of speed with respect to flow is -0.6, implying that, for each vehicle removed from the mainline, the social welfare gain is 11 cents per mile.

[55] The intuition is simple. The larger the hypothetical congestion benefit is larger, the greater the external costs of congestion. Therefore, if other freeways or travel options are not as congested as the mainline, the potential benefit cannot be as large as the benefit in the mainline and planners generally place HOV lanes on the most congested routes.

time, this effect may be partially alleviated if, by moving to the HOV lane, SOVs drive closer to optimal speed (Currie and Walker, 2011; Knittel, Miller, and Sanders, 2011). While acknowledging these two sources of pollution effects, in the results below, we still do not include welfare effects that occur due to pollution changes. This will be incorporated in the next version of the paper.

B. Welfare Effects

Table 5 displays estimates of the welfare effects of the program, broken down by the first month and full program period.[56] The primary welfare effects of the program are calculated using the estimates reported in Table 2, columns I and V. The interaction effects are calculated based on the range of estimates for the value of time derived in the previous section. For the interaction effects, in the current version of the paper, we ignore any welfare that may occur due to changes in reliability in the HOV and mainlines.

The table underscores the following key results: First, abstracting from preferences for urgency substantially underestimates the magnitude of the primary welfare gain. Urgency alone accounts to about 70 to 80 percent of the primary welfare gain.[57] Second, simple ex-ante cost-benefit analysis of ExpressLanes programs that only considers the time saving benefits would imply that the project would barely pass a cost-benefit analysis test,[58] when, in reality, the primary welfare effect of the program substantially dominates the operating costs of the corridor.

[56] The full program period encompasses the full time period of data we have, 10.25 months, although the ExpressLanes are still in operation.

[57] In Appendix Table C.27, we further break down the primary welfare effect by the time differential between the HOV and mainlines. Even for travel time differences in excess of 5 minutes, urgency accounts for 65% of the total revenue.

[58] As shown in Appendix Table C.26 accounting only for trips in excess of 5 minutes would predict only $14,389 in benefits from value of time, which would not justify the upkeep cost of the corridor.

Interestingly enough, the direct congestion interaction effect is negligible, while the system wide interaction effect is rather large. Two features of the program led to a negligible direct congestion interaction effect. First, the real-time toll structure assures that the speed in the HOV lane is never below 45 mph. Therefore, even if the direct congestion interaction welfare effect would have translated into a welfare loss, the magnitude of the loss would be controlled by the price mechanism. Second, and perhaps more importantly, the project financed the adding of a lane to increase capacity. Since the elasticity of carpool formation with respect to capacity is rather small, even after the addition of the SOV, the HOV speed remained at free flow. In contrast, the system-wide congestion interaction effect results in a non-trivial welfare gain, ranging from $207,548 to $350,158. Because congestion is not priced in the mainline lanes, and even though the ExpressLanes program only moves a relatively small number of SOVs into the HOV, a large welfare gain is created. In fact, at peak periods, removing a SOV from the mainline lanes causes a social welfare gain of $1.20.

We also calculate the system-wide benefit per dollar transferred to a SOV, and find this to range from 0.17 to 0.24. In addition to ExpressLanes, several other programs aim to alleviate congestion in mainline lanes. Another option, for example, consists of increasing public transit availability (Anderson, 2014) or subsidizing public transit fares (Parry and Small, 2009). While such options may also generate similar system-wide congestion interaction benefits, unlike the ExpressLanes program, these options do not generate revenues but rather require financing.

Finally, while at a first glance one may think that it is desirable to replace carpoolers by urgent SOV in the HOV, the significance of the system-wide congestion interaction effect demonstrates the shortcomings of such proposal. Replacing a carpool without urgency with an urgent SOV improves total welfare by $1.65, but only if the carpool does not break into multiple mainline vehicles.

Splitting a carpool will result in welfare losses ranging from $1.45 to $2.43. These external benefits suggest that compared with the average $5.90 paid by a representative SOV traveling the full length of the ExpressLanes[59], a representative carpool should only pay a toll of $0.61 to $2.08.

VIII. Conclusion

This paper presents substantial evidence demonstrating that drivers scheduling decisions are largely determined by a discrete cost of late arrival we term urgency. Unlike prior literature that evaluates these costs on a per hour basis, we hypothesize that schedule constrained commuters place substantially more value on avoiding late arrival than they do arriving late but by a smaller margin. We estimate that urgency cost is $3, roughly equal to 15% of the local wage, and additional minutes late are valued at little more than half the local wage. Given the relatively small time savings generated by urban infrastructure improvements, the error introduced by assuming urgency scales with time is substantial.

Initially we find a striking pattern in the data that a multitude of agents have a willingness-to-pay for small time savings that does not seem well grounded in prior literature. Using a bottleneck model of queuing, we introduce a discrete lateness cost that allows us to match outcomes with surprising precision. We also show that incorporating urgency allows for more reasonable numbers of late agents.

Returning to the data hedonic regression shows that the only model that predicts WTP across the entire range of commuters is one that incorporates urgency. We examine the possibility that our estimated urgency of $3 is due to reliability, unobserved route characteristics, or individual preference and find that these

[59] As noted in Table 1 the average SOV trip length in the ExpressLanes is generally closer to half the total length.

cannot explain the estimated parameters. Failing to incorporate urgency dramatically under predicts total toll revenue with a simple value of time model at half the local wage accounting for less than 19% of the revenue. We also show that the assumptions necessary to attribute this benefit to heterogeneity in wages or other route benefits are unrealistic.

Our welfare analysis uses these estimates to demonstrate the critical importance of this parameter for infrastructure evaluation. Accounting for changes to travel time and the distribution of travel time we find that urgency benefits to SOV drivers using the ExpressLanes are the dominant welfare component and, raise substantially more revenue than the cost of the upkeep of the corridor. Moreover we find that on a per driver basis, schedule constrained drivers gain more from the ExpressLanes than carpoolers lose if forced into the mainline lanes, and subject to carpools remaining intact, it may be possible to increase total welfare by a reordering of agents across lanes.

REFERENCES

Anderson, Michael L. 2014. "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion." *American Economic Review* 104(9): 2763-96.

Arnott, Richard, André de Palma, and Robin Lindsey. 1990. "Economics of a Bottleneck." *Journal of Urban Economics* 27 (1): 111–30.

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.

Arnott, Richard, André de Palma, and Robin Lindsey. 1994. "The Welfare Effects of Congestion Tolls with Heterogeneous Commuters." *Journal of Transport Economics and Policy* 28 (2): 139-61.

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

Bento, Antonio M., Johnathan Hughes, and Daniel Kaffine. 2013. "Carpooling and Driver Responses to Fuel Price Changes: Evidence from Traffic Flows in Los Angeles." *Journal of Urban Economics* 77: 41-56.

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.

Bovenberg, A. Lans, and Lawrence H. Goulder. 1996. "Optimal Environmental Taxation in the Presence of Other Taxes: General-Equilibrium Analyses." *American Economic Review* 86 (4): 985–1000.

Brown, Charles. 1980. "Equalizing Differences in the Labor Market," *Quarterly Journal of Economics* 94(1), 113-134.

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.

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.

Cropper, Maureen L., Leland B. Deck, and Kenenth E. McConnell. 1988 "On the Choice of Functional Form for Hedonic Price Functions." *The Review of Economics and Statistics* 70 (4): 668-675.

Currie, Janet, and Reed Walker. 2011. "Traffic Congestion and Infant Health: Evidence from E-ZPass." *American Economic Journal: Applied Economics* 3 (1): 65-90.

Davis, Lucas W. 2004. "The Effect of Health Risk on Housing Values: Evidence from a Cancer Cluster." *American Economic Review* 94 (5): 1693-1704.

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

DeSerpa, Allan C. 1971. "A Theory of the Economics of Time." *Economic Journal*: 828-46.

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

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.

Golob, Thomas F., Wilfred W. Recker, and Douglas W. Levine. 1989. "Safety of High Occupancy Lanes without Physical Separation." *Journal of Transportation Engineering* 115 (6): 591-607.

Harberger, Arnold C. 1974. *Taxation and Welfare*. Chicago: University of Chicago Press.

Hymel, Kent M., Kenneth A. Small, and Kurt Van Dender. 2010. "Induced Demand and Rebound Effects in Road Transport." *Transportation Research: Part B: Methodological* 44 (10): 1220–41.

Johnson, M. Bruce. 1966. "Travel Time and the Price of Leisure." *Western Economic Journal* 4 (2): 135-45.

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.

Knittel, Christopher R., Douglas L. Miller, and Nicholas J. Sanders. 2011. "Caution Drivers! Children Present: Traffic, Pollution, and Infant Health." Mimeo.

Lave, Charles A., 1969. "A Behavioral Approach to Modal Split Forecasting." *Transportation Research* 3: 463-480.

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, Daniel J. 2011. "Can Consumption of Convenience Products Reveal the Opportunity Cost of Time?" *Economics Letters* 113 (1): 92-95.

Small, Kenneth A. 1982. "The Scheduling of Consumer Activities: Work Trips." *American Economic Review* 72 (3): 467–79.

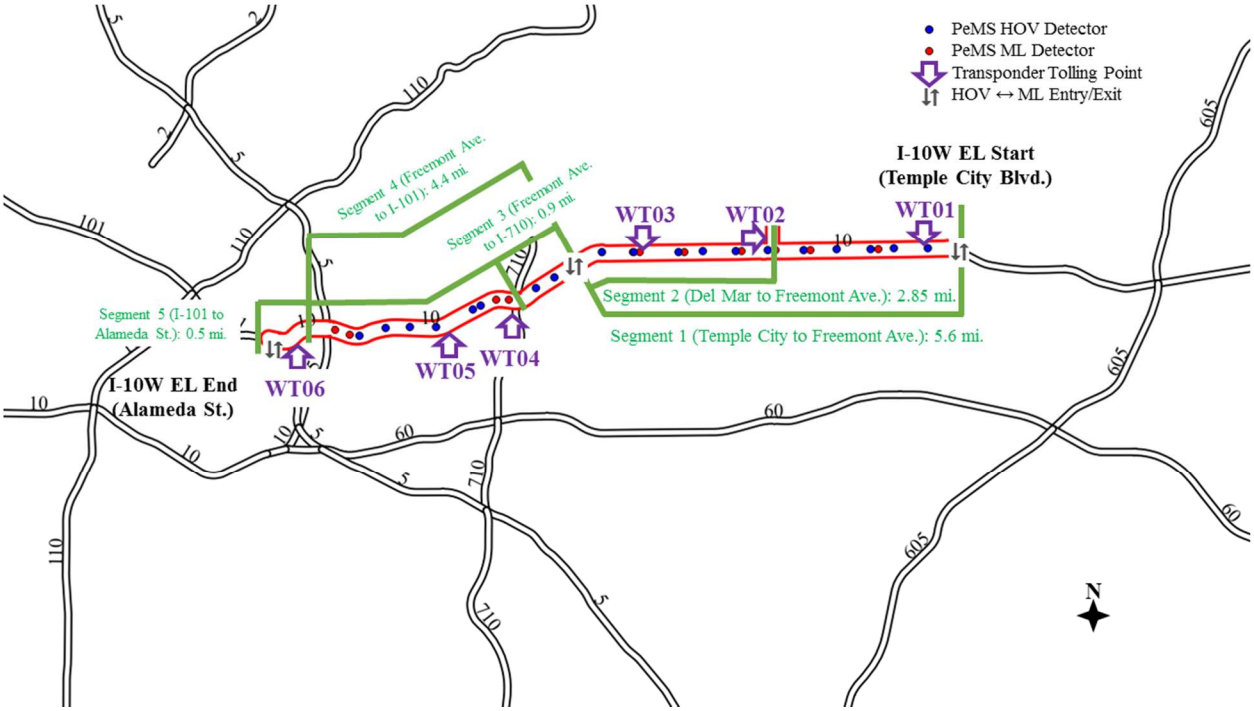
Small, Kenneth A. 2012. "Valuation of Travel Time." *Economics of Transportation* 1 (1-2): 2-14.

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.

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

Viscusi, W. Kip, and Joseph E. Aldy. 2003. "The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World." *Journal of Risk and Uncertainty* 27 (1): 5-76.

FIGURES AND TABLES



MAP 1. I-10W EXPRESSLANES DESIGN

Notes: The I-10W ExpressLanes design includes 5 separately tolled segments along its 10.5 mile stretch West of Downtown Los Angeles. The beginning and end of each segment is defined by a transponder detector and license plate scanner at each tolling plaza (indicated in the map with a purple arrow) that identifies vehicles entering and exiting the ExpressLanes.

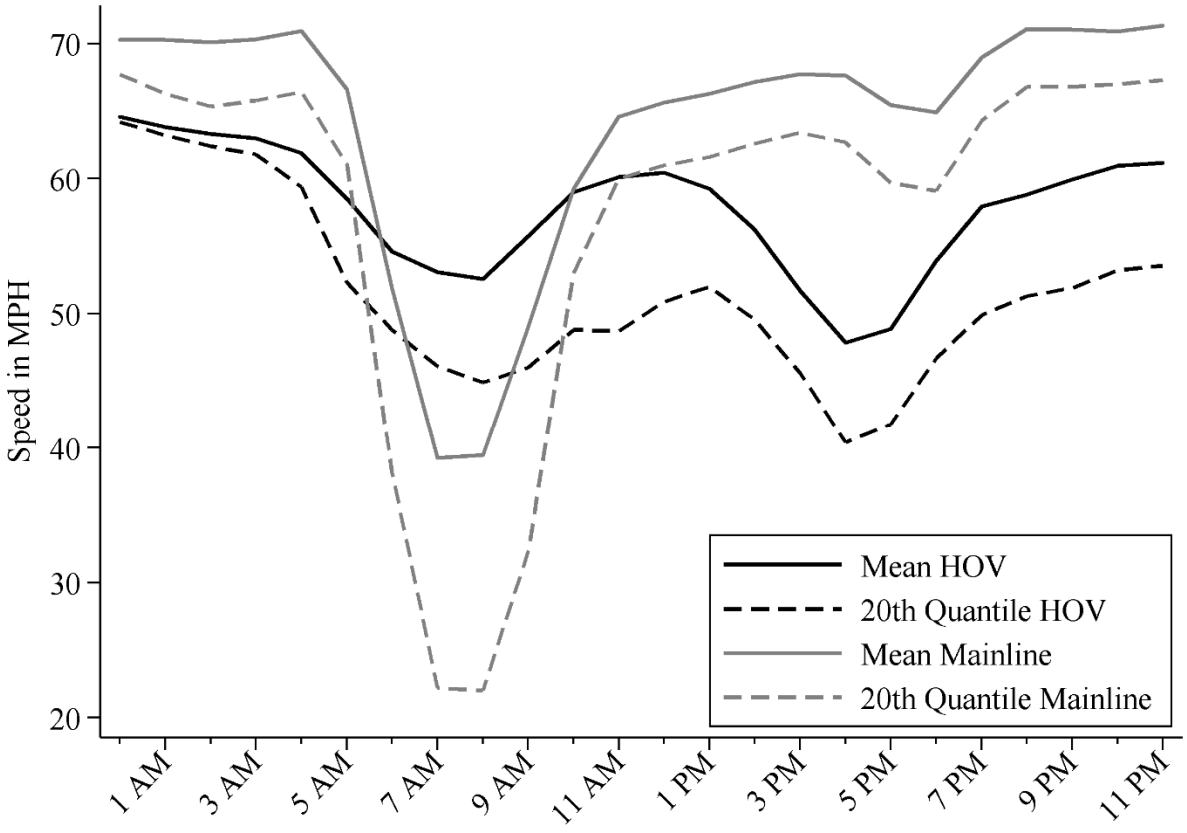


FIGURE 1. MEAN AND 20TH QUANTILE OF SPEED BY HOUR

Notes: The figure displays the average hourly pre-policy speed detected by PeMS from September 3rd, 2012 until February 22nd, 2013 in the indicated lane for each hour of the day on the I-10W in the HOV and mainline lanes. Weekends, holidays and observations where any of the 30 second observations are missing are dropped.

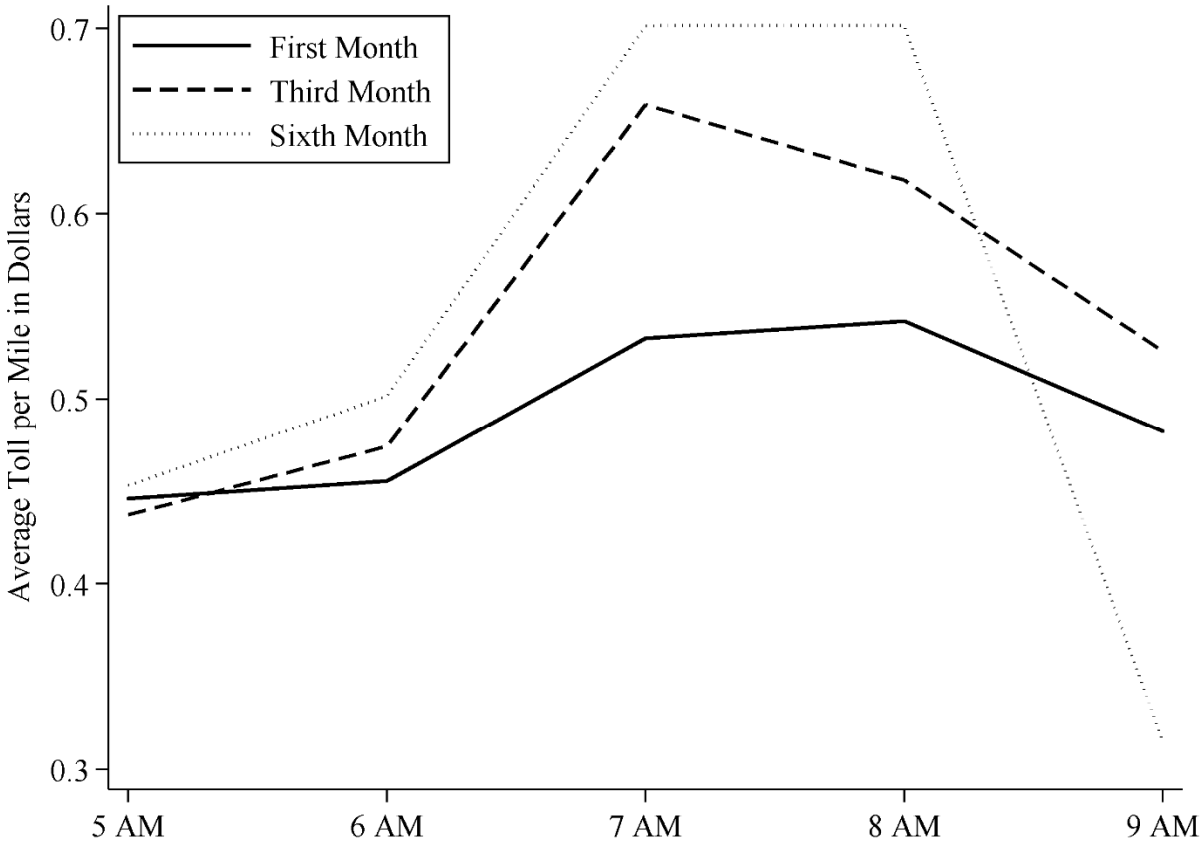


FIGURE 2. EVOLUTION OVER TIME OF TOLL PER MILE BY HOUR

Notes: The figure displays the average hourly toll per mile in dollars paid during the morning peak for drivers on the I-10W ExpressLanes during the first month, (February 25th, 2013 – March 31st, 2013), the third month (May 2013) and the sixth month (August 2013). Trips during weekends and holidays are removed as well as those for vehicles linked to public sector, corporate or unknown accounts.

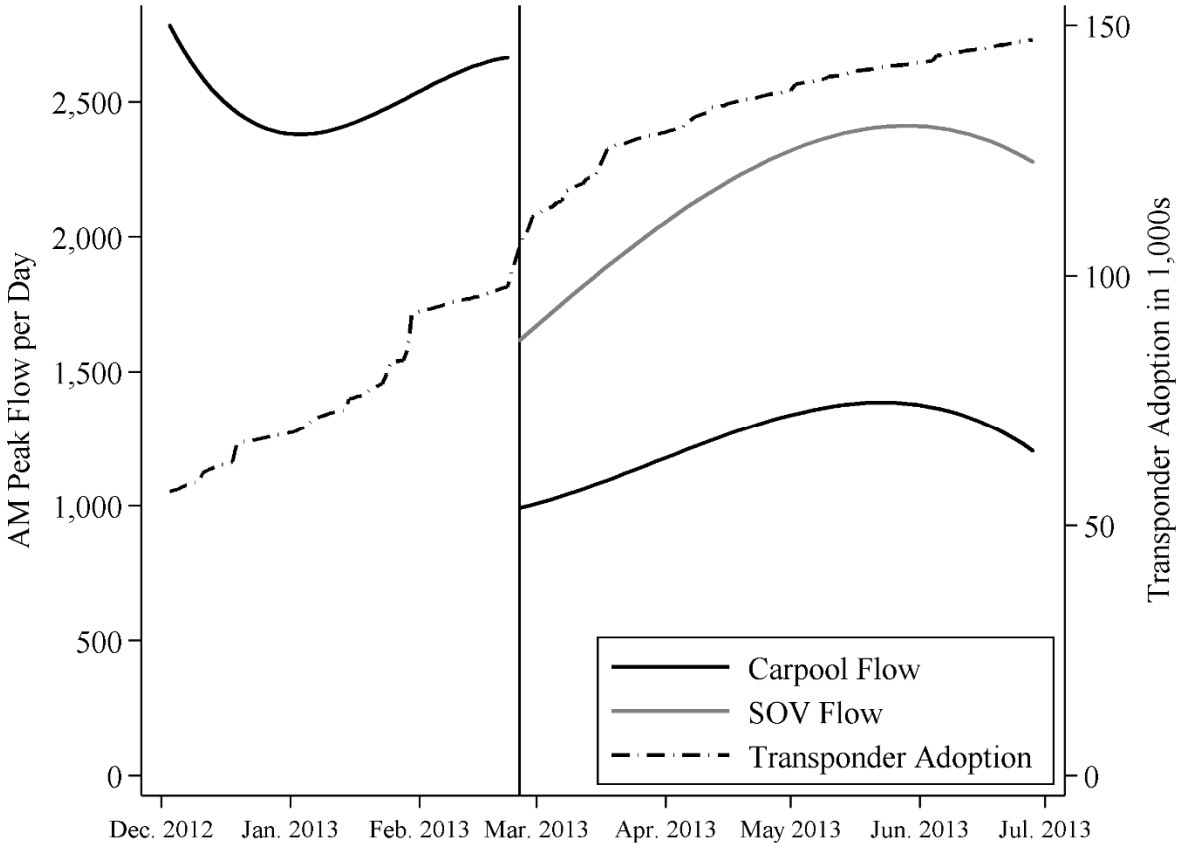


FIGURE 3. EXPRESSLANES FLOWS AND TRANSPONDER ADOPTION

Notes: The dashed line displays cumulative transponder adoption for the entire ExpressLanes program on the I-10 and I-110 in both directions. Flow is the number of cars passing the average detector. Carpool and SOV flows on the I-10W ExpressLanes are estimated using third order kernel-weighted polynomial smoothing. The vertical line denotes the policy implementation date, February 23rd, 2013. 'SOV Flow' includes both SOV and HOV-2 flows, while 'Carpool Flow' corresponds to vehicles with an occupancy of three persons or more. Flow data cover the morning peak hours of work days in the first 10 months of the policy. Trips during holidays are dropped as well as those for vehicles linked to public sector, corporate or unknown accounts. Observations from PeMS where any of the 30 second observations are missing are also dropped.

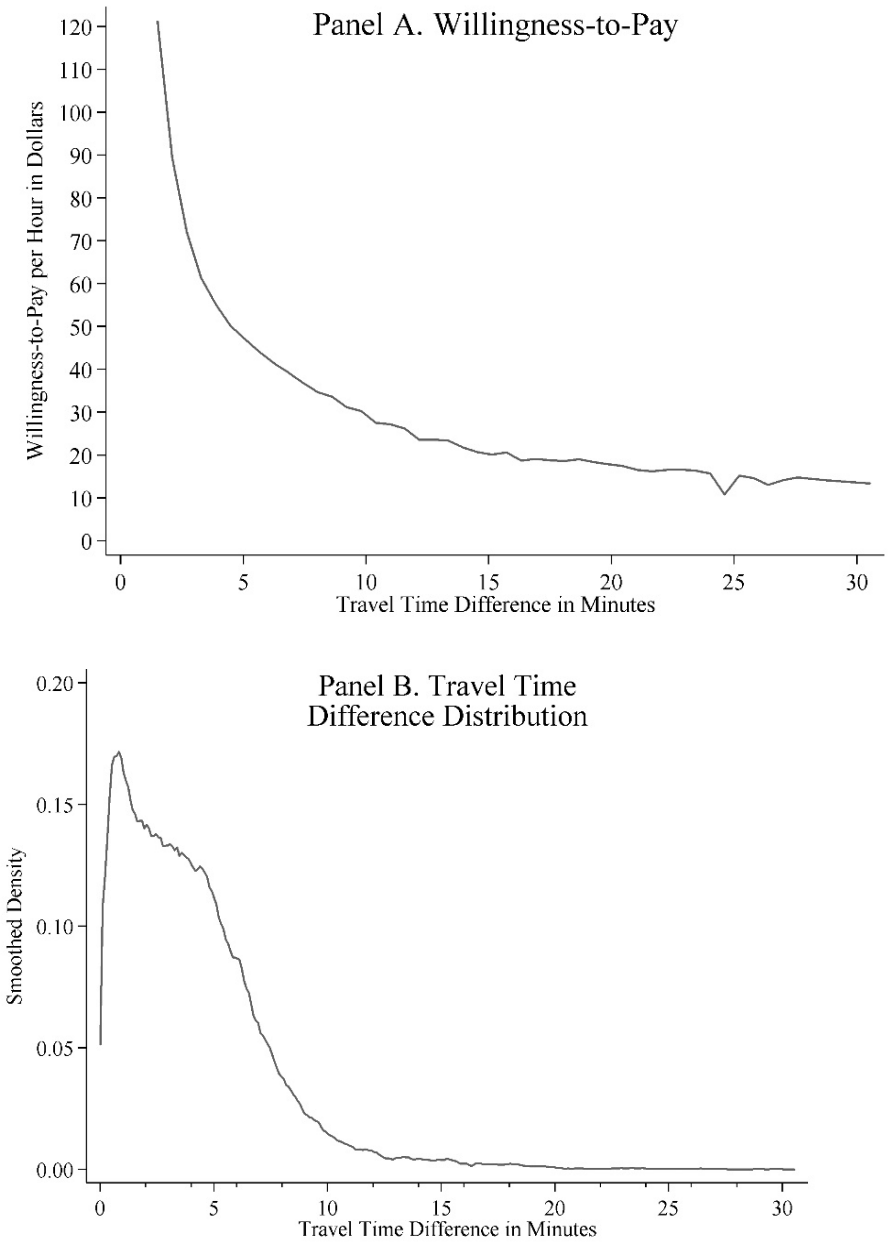


FIGURE 4. WILLINGNESS-TO-PAY PER HOUR AND DEMAND FOR TRIPS IN THE EXPRESSLANES

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.

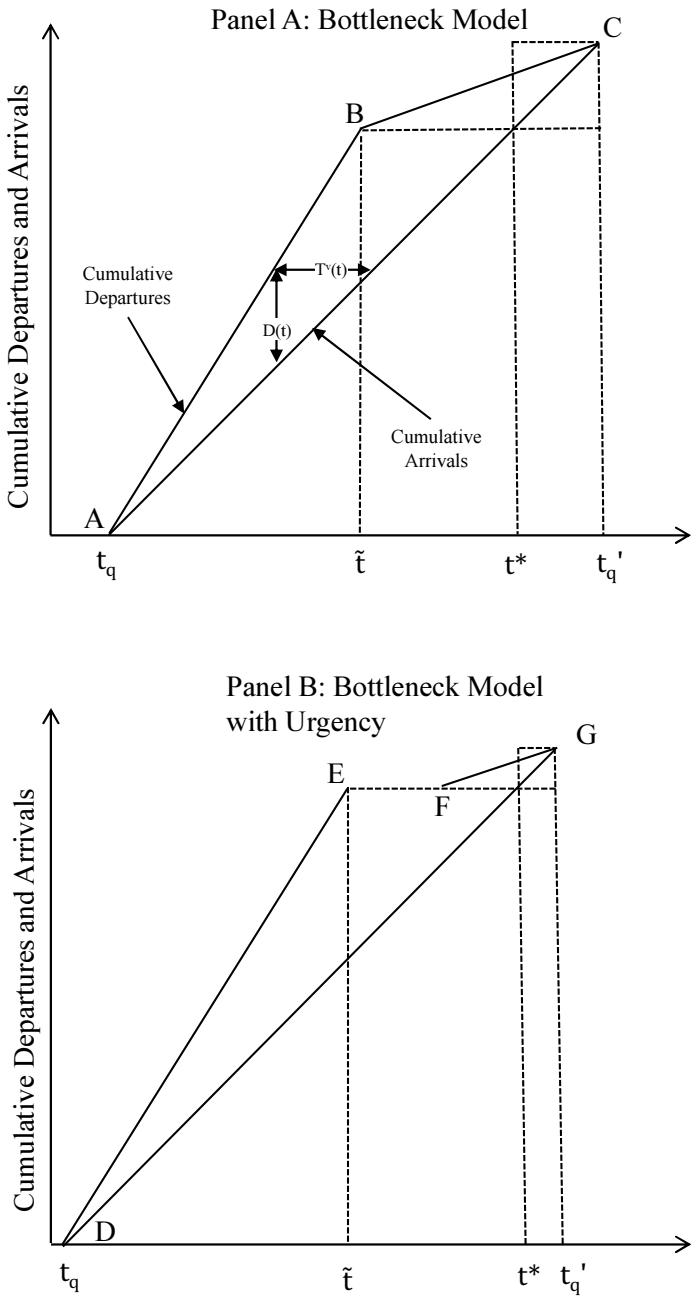


FIGURE 5. CONCEPTUAL FRAMEWORK FOR RELIABILITY FROM BOTTLENECK MODEL, WITH AND WITHOUT URGENCY

Notes: The figures depict the behavior of individual commuters during a daily commute in the bottleneck model. Panel A describes the bottleneck model with only per-hour penalties for being early or late, while Panel B describes the behavior of individuals with indirect utility that includes a discrete cost for being late associated with urgency. Solid lines along the triangle refer to the boundaries of the queue formed at various points of the peak. The vertical dashed line refers to the preferred arrival time t*, while t_q and t_q' refer to the beginning and end of the bottleneck, respectively. Horizontal distances between the solid lines, denoted by $T^v(t)$ in Panel A, refer to time spent in the queue, while vertical distances, denoted by D(t), refer to the mass of individuals in the queue at a given time. The distance EF in Panel B refers to the later shift in the mass of departures as a result of urgency.

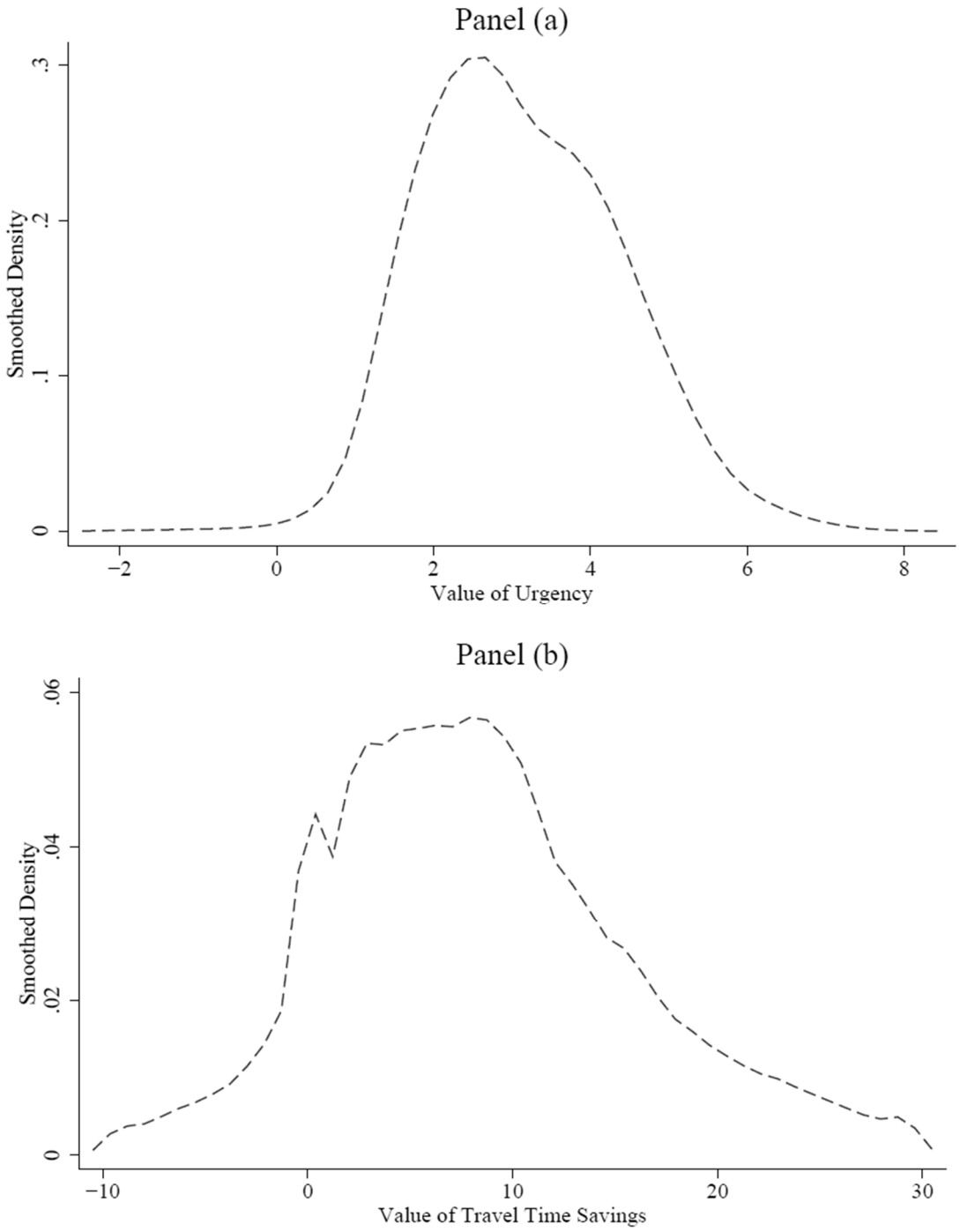


FIGURE 6. ESTIMATED DISTRIBUTION OF VALUE OF TIME AND URGENCY

Notes: This figure depicts 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. 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.

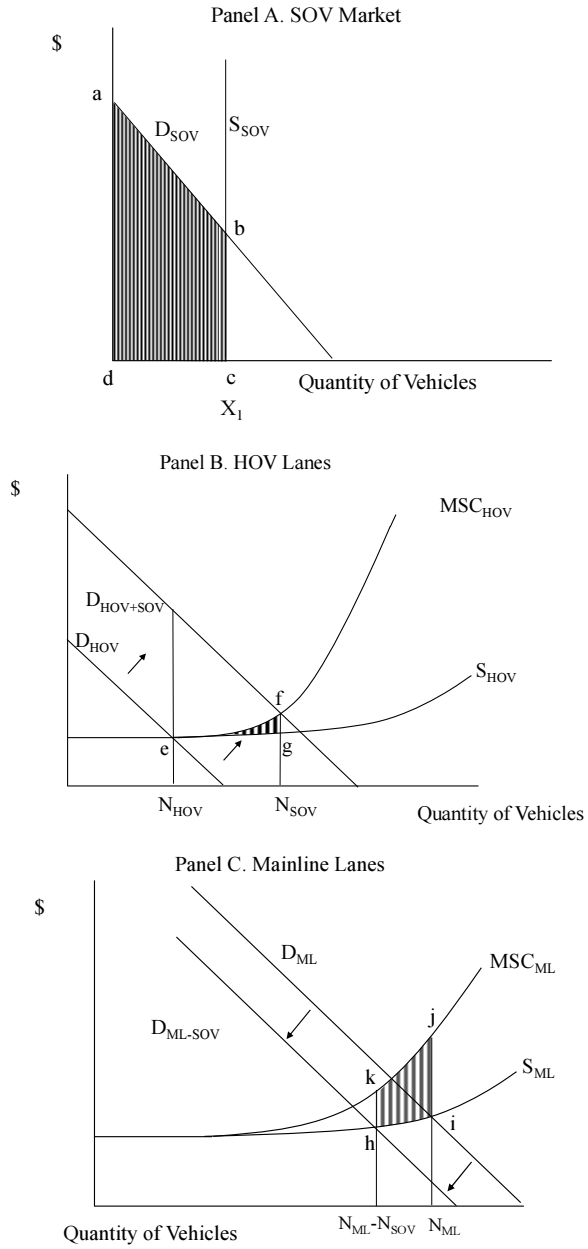


FIGURE 7. WELFARE EFFECTS IN HOV LANES AND MAINLINE LANES MARKETS

Notes: The figures depict the welfare effects of moving single-occupant vehicles from the mainline lanes to the HOV lanes. The curves S_{HOV} and S_{ML} refer to the supply of available spaces in the HOV and mainline, respectively. The curve MSC_{HOV} shows the additional marginal costs from congestion incurred in the HOV lanes, while MSC_{ML} shows these marginal costs in the mainline. The larger slope of the MSC_{HOV} curve is due to the 3+ passenger requirement of the lane, which causes a larger externality for each additional car. The demand curve, D_{HOV}, increases in the HOV lanes to $D_{HOV+SOV}$ due to the policy, while mainline lane demand, D_{ML}, decreases in the mainline lanes as ExpressLanes users are removed.

TABLE 1— TRIP-LEVEL STATISTICS BY DECILE OF TRAVEL TIME SAVINGS

I	II	III	IV	V	VI	VII	VIII
Panel A. Summary Statistics							
	Time Savings						
Decile of Time Savings	in Hours	in Minutes	Average Express-Lanes\HOV Speed in MPH	Average Mainline Speed in MPH	Average Distance Traveled in Miles	Average Uses per Month	Average Hourly Wage in Zip Code
1	0.01	0.39	65.3	60.3	5.8	8.8	$19.35
2	0.02	1.01	67.4	55.9	6.1	9.5	$19.40
3	0.03	1.66	66.6	50	6.2	9.8	$19.47
4	0.04	2.37	66.1	44.7	6.1	9.9	$19.47
5	0.05	3.11	66	40.6	6.1	9.8	$19.65
6	0.06	3.88	65.8	37.7	6.3	9.9	$19.71
7	0.08	4.69	65.5	34.6	6.3	9.8	$19.73
8	0.09	5.64	64.7	32.7	6.7	9.8	$19.76
9	0.12	6.95	63.8	30.9	7.3	9.8	$19.79
10	0.18	11.04	62	25.8	8.1	9.6	$20.00
Average	0.07	4.08	65.3	41.3	6.5	9.7	$19.63

Panel B. Willingness-to-Pay

	Time Savings			Average WTP per Hour			
Decile of Time Savings	in Hours	in Minutes	Average Toll Paid	Full Time Period	February & March	June	September
1	0.01	0.39	$3.20	$1,977	$1,730	$1,910	$1,220
2	0.02	1.01	$3.10	$190	$147	$242	$134
3	0.03	1.66	$3.12	$115	$94	$158	$86
4	0.04	2.37	$3.17	$81	$72	$116	$66
5	0.05	3.11	$3.29	$64	$55	$85	$55
6	0.06	3.88	$3.57	$55	$45	$70	$48
7	0.08	4.69	$3.81	$49	$39	$62	$44
8	0.09	5.64	$4.15	$44	$34	$56	$41
9	0.12	6.95	$4.49	$39	$29	$46	$38
10	0.18	11.04	$4.95	$28	$25	$40	$28
Average	0.07	4.08	$3.69	$264	$227	$278	$176

Notes: Unless otherwise indicated, all data cover work days for the morning peak (5-9 AM) from February 25th, 2013 until December 30th, 2013. 'Time Savings' is the travel time saved by driving in the ExpressLanes over the mainline lanes, calculated from Metro transponder information on vehicle distance traveled and speed compared with the speed recorded by PeMS in the mainline lanes. 'Average Hourly Wage in Zip Code' is calculated based on the reported zip code for each transponder and 2008-12 ACS Census average zip code data, assuming an assumed average household with two wage-earners and 2,040 working hours per year. 'Average Uses per Month' excludes the first month that a transponder appears in the data to control for learning behavior. 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—REGRESSION OF TOTAL TOLL ON TIME DIFFERENTIALS

	I	II	III	IV	V	VI
Constant	2.94***	2.82***			2.60***	2.66***
	(0.50)	(0.36)			(0.46)	(0.40)
Time in hours	11.05***	14.49	37.59***	62.27***	7.30**	6.50*
	(3.03)	(9.32)	(3.94)	(9.12)	(2.78)	(3.34)
Time in hours^2		-15.07		-158.39***		
		(27.65)		(18.82)		
Reliability					32.09***	32.71***
					(6.65)	(7.00)
Trip Restriction	>0 Minutes	>0 Minutes	>0 Minutes	>0 Minutes	>0 Minutes	None
Number of Obs.	466,232	466,232	466,232	466,232	466,232	496,839
AIC	1,655,287	1,653,423	2,106,127	1,951,494	1,598,105	-
BIC	1,655,310	1,653,456	2,106,138	1,951,516	1,598,138	-

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 lane speeds reported by PeMS, for the chosen trip distance. Standard errors, clustered by road segment, are in parentheses. Observations from morning peak hours 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— REGRESSION OF TOTAL TOLL ON TIME DIFFERENTIALS

	I	II	III	IV	V	VI
Panel A. Temporal Sensitivity						
	Other Times of Day on I-10W			I-10E	I-110N	I-110S
	Afternoon Peak	Weekday Off-Peak	Weekend	Afternoon Peak	Morning Peak	Afternoon Peak
Constant	2.35***	1.52***	0.70***	2.26***	3.58***	2.38***
	(0.28)	(0.22)	(0.13)	(0.53)	(0.34)	(0.25)
Time in hours	19.37**	8.80***	5.07**	23.42	21.16***	11.23**
	(7.10)	(1.91)	(1.17)	(15.91)	(3.47)	(4.87)
Included Times	4-8 PM	9 AM-4 PM, 9 PM-4 AM	All Sat. & Sun.	4-8 PM	5-9 AM	4-8 PM
Number of Obs.	37,208	140,638	9,178	320,666	474,762	646,562
Average Toll	$2.61	$1.91	$1.29	$2.43	$4.45	$2.79
Ratio of Urgency to Total Toll	0.90	0.80	0.54	0.93	0.80	0.85
Average Time Savings (Hrs.)	0.014	0.044	0.115	0.007	0.041	0.037
Average Time Savings (Min.)	0.816	2.659	6.895	0.436	2.476	2.196
Panel B. Weekend Control Group						
Morning Peak Indicator	2.07***	1.89***	1.88***	1.76***	1.73***	1.68***
	(0.50)	(0.35)	(0.37)	(0.45)	(0.37)	(0.39)
Time in Hours	3.11***	0.76	0.66	2.85***	1.10	0.79
	(0.96)	(1.31)	(1.38)	(0.88)	(1.12)	(1.21)
Time in Hours*AM Peak	8.01***	8.22***	8.17***	4.48*	6.65***	6.90***
	(2.27)	(1.76)	(1.77)	(2.20)	(1.65)	(1.66)
Reliability				7.95***	2.46	2.09
				(1.10)	(2.45)	(2.18)
Reliability*AM Peak				23.92***	14.21**	13.86**
				(5.75)	(5.24)	(5.34)
Number of Obs.	302,251	302,251	284,247	302,251	302,251	284,247
Account Fixed Effects	N	Y	N	N	Y	N
Transponder Fixed Effects	N	N	Y	N	N	Y

Notes: Values shown are the coefficients of twelve 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 lane speeds reported by PeMS, for the chosen trip distance. Reliability is calculated as the travel time difference between the 50^{th} and 80^{th} quantile of travel time for the month of the trip in that hour over the measured distance of the segment. Standard errors, clustered by road segment, are in parentheses. Observations from morning peak hours 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 4—REGRESSION OF TOTAL TOLL ON TIME DIFFERENTIALS

	I	II	III	IV
		Limited Time Variation		
Constant		3.57***		3.92**
		(1.10)		(1.25)
Time in Hours	31.22***	7.24**	21.68***	5.38*
	(3.81)	(2.58)	(2.53)	(2.45)
Trip Restriction	> 5 minutes	> 5 minutes	> 10 minutes	> 10 minutes
Number of Obs.	146,365	146,365	21,830	21,830

Notes: Values shown are the coefficients of four 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 lane speeds reported by PeMS, for the chosen trip distance. Standard errors, clustered by road segment, are in parentheses. Observations from morning peak hours 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 5—WELFARE EFFECTS OF PROGRAM

Panel A. Welfare Effect (Partial Equilibrium)		
	Monthly[1]	Full Program[1]
Private SOV Drivers[2]	$101,293	$1,718,492
Time Benefits at $11.05	$24,314	$350,070
VOU at $2.94	$95,710	$1,368,422
Time Benefits at $7.30	$16,049	$231,075
VOR at $32.09	$24,814	$275,416
VOU at $2.60	$84,769	$1,212,002
Other SOV drivers[2]	$53,274	$590,466
All SOVs	$154,567	$2,308,958
Monthly Operating Cost of Corridor[3]	$21,000	$215,250
Panel B. Interaction Effects		
Direct Interaction Effect		
HOV Market		
Short Run Effect[4]		$0
System Wide Interaction Effect[5]		
Using Partial Equilibrium Speed-Flow Relationship		
VOT $11.05	$39,122	$350,138
VOT $10.00	$35,404	$316,867
VOT $8.00	$28,323	$253,493
VOT $6.55	$23,190	$207,548
Panel C. Distributional Effects		
SOV Drivers[6]		1,626
Private SOV drivers		858
HOV Drivers[6]		948
Mainline Lane Drivers[7]		18,180
System Benefit per $ Transferred to SOV Driver[8]		$0.17 to $0.24
Benefit of Replacing a Carpool with an 'Urgent SOV'[9]		$1.65
if Carpool Splits		-$2.43 to -$1.45
Optimal Toll for New Carpool[10]		$0.61 to $2.08

[1]Excludes weekends and holidays.

[2]'Private SOVs' exclude accounts registered to government and business. The revenues from other accounts is included in 'Other SOV drivers'.

[3]Source: Correspondence with LA Metro, 04/15/14.

[4]Because the lane is observed to be in free flow, these costs for both travel time and reliability are zero. See Appendix B for more details and discussion.

[5]These effects include changes to travel time for mainline drivers, which are calculated by removing 1,626 vehicles from the mainline lanes and using the speed-flow elasticity of -0.6 to generate the partial equilibrium change in travel times. VOT of $11.05 assumes all costs that scale with time based on estimates from Table 2 are VOT, $10.00 is taken as the

VOT implied by half the local wage, $8.00 and $6.55 are the VOT values implied by the bottleneck model using the calculations in section IV.

[6]'Private SOV Drivers' is the average daily number of vehicles passing over a detector that are registered to a private vehicle. This number is calculated as the total number of miles driven by private SOV vehicles daily divided by 10.5 miles.

[7]Mainline driver counts are taken as the pre-policy flow less the daily number of SOV agents observed in the ExpressLanes.

[8] Calculated as the ratio of the System Wide Interaction Effect using $11.05 through $6.55 VOT to the total toll revenue.

[9] Calculated as average toll revenue per SOV divided by the private cost for 3 individuals with travel times in the mainline lanes as opposed to the ExpressLanes. Private cost of a 3-person carpool is calculated assuming that the travel time experienced would be the increased travel time experienced in the mainline lane times a VOT of $11.05. Split carpool assumes 2 additional vehicles generate external costs of congestion at $0.11 per mile and pollution costs at $0.08 per mile (Parry and Small, 2005).

[10]Assumes that the prevailing toll is $5.90 as observed for a representative agent using the full 10.5 mile ExpressLanes. External benefits of carpool assumed to be removed pollution and congestion costs of two average vehicles in the mainline lanes. Assumes no congestion costs in ExpressLanes.