UNCOVERING THE DISTRIBUTION OF MOTORISTS’ PREFERENCES FOR TRAVEL TIME AND RELIABILITY: IMPLICATIONS FOR ROAD PRICING

by Kenneth A. Small, Clifford Winston, and Jia Yan

August 6, 2004

Abstract

Recent econometric advances have made it possible to identify empirically the varied nature of consumers’ preferences. We apply these advances to study commuters’ preferences for speedy and reliable highway travel, with the objective of exploring the efficiency and distributional effects of road pricing designed specifically to account for users’ heterogeneity. Our analysis combines revealed and stated commuter choices of whether to pay a toll for congestion-free express travel. We find that motorists exhibit substantial heterogeneity in their values of travel time and reliability. Moreover, we show that road pricing policies designed to cater to varying preferences can improve efficiency and reduce the disparity of welfare impacts compared with recent pricing experiments. By recognizing heterogeneity, policymakers have a long-awaited opportunity to break the impasse in efforts to relieve highway congestion.

Keywords: mixed logit, stated preference, congestion pricing, product differentiation, value of time

Kenneth A. Small
Department of Economics
University of California
Irvine, CA 92697-5100
USA
ksmall@uci.edu

Clifford Winston
Brookings Institution
1775 Mass. Ave., N.W.
Washington, DC 20036
USA
cwinston@brook.edu

Jia Yan
Department of Logistics
Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong
lgtjiay@polyu.edu.hk

*Corresponding author
Acknowledgment

The authors are grateful to the Brookings Center on Urban and Metropolitan Policy and the University of California Transportation Center for financial support. We thank Edward Sullivan for access to data collected by California Polytechnic State University at San Luis Obispo, with financial support from the California Department of Transportation and the U.S. Federal Highway Administration’s Value Pricing Demonstration Program. We also are grateful for comments from David Brownstone, Jerry Hausman, Charles Lave, Steven Morrison, and Randy Pozdena; from participants in seminars at UC Irvine, Northwestern University, the American Economic Association, and the University of Maryland at College Park; and from the referees and a co-editor of this journal.
1. Introduction

On a given weekday, roughly two hundred million people in the United States use a vehicle for work or personal trips. Highway authorities have relied primarily on the gasoline tax to charge road users, with the unfortunate result that congestion on urban and intercity highways continues to worsen. But what if policymakers recognized that motorists’ attitudes toward congestion range from loathing to indifference and accordingly offered motorists differentiated road prices that catered to their preferences? Could such a policy increase efficiency while being more palatable to users than pricing policies designed for a representative traveler?

Recent pricing experiments in the Los Angeles, San Diego, and Houston areas give highway travelers the option to travel free on regular lanes or to pay a time-varying price for congestion-free express travel on a limited part of their journey. These experiments, often called “value pricing,” provide rare opportunities to study motorists’ responses to pricing in automobile-dominated environments where real money is at stake.

This paper identifies motorists’ preferences by analyzing newly collected data concerning commuters’ choices of whether to pay a toll to use the express lanes in the Los Angeles-area pricing experiment. The key methodological features of our analysis are a random-parameters mixed logit model to account for unobserved heterogeneity; combining revealed preference and stated preference data to enable us to estimate key parameters; and developing plausible characterizations of travel-time reliability that capture users’ ability to know in advance the time it will take them to travel.

Based on their lane choices, we find that commuters vary substantially in how they value travel time and travel-time reliability. Our policy simulations using these results indicate that differentiated road prices have significant efficiency and distributional advantages. For instance, recent
experiments have set high prices on one part of a roadway. We find that setting more moderate but
differentiated prices on both parts can increase welfare gains and reduce the disparity in consumer
surplus losses among different users. We suggest that such advantages greatly enhance the political
viability of road pricing and may spur broader adoption.

2. Empirical Setting and Overview of the Samples

The commuter route of interest is California State Route 91 (SR91) in the greater Los Ange-
les region. It connects rapidly growing residential areas in Riverside and San Bernardino Coun-
ties—the so-called Inland Empire—to job centers in Orange and Los Angeles Counties to the west.
A ten-mile portion of the route in eastern Orange County includes four regular freeway lanes (91F)
and two express lanes (91X) in each direction. Motorists who wish to use the express lanes must set
up an account and carry an electronic transponder to pay a toll that varies hourly according to a pre-
set schedule. The toll is set to manage aggregate traffic levels; thus, its level is not influenced by
an individual motorist’s lane choice or unobserved characteristics.¹

Previous research has measured the value of motorists’ travel time using either revealed
preference (RP) data based on the choice between travel by car and public transit, or stated prefer-
ence (SP) data based on hypothetical commuting situations.² Use of RP data is often hindered by
strong correlations among travel cost, time, and reliability; whereas SP data are tainted by doubt
whether the behavior exhibited in hypothetical situations applies to actual choices. We draw on
well-developed methodologies to combine both types of data to help overcome the limitations of
each (Ben-Akiva and Morikawa (1990), Bhat and Castelar (2002)).

¹ Unlike the regular lanes, the express lanes on SR91 have no entrances or exits between their end points. Because SR91
traverses a mountain pass, few motorists have viable alternatives to this route with the exception of a fixed-toll express-
way that can be used by those who travel to south Orange County (Sullivan et al. (2000)).

² Goodwin (1992) and Wardman (2001) review this literature.
To enrich our analysis, we draw on two samples of people traveling on the SR91 corridor. The surveys generating the data contain sufficiently similar questions and were conducted at nearly the same times, so it has proven feasible and fruitful to combine them. The first is a telephone RP survey composed of SR91 commuters obtained by random-digit dialing and observed license plates on the corridor; it was conducted by researchers at California Polytechnic State University at San Luis Obispo (Cal Poly), under the leadership of Edward Sullivan with our participation. The Cal Poly data, collected in November 1999, asked participants about their most recent trip on a weekday (excluding Fridays) during the morning peak (4-10 a.m.). It included questions concerning lane choice (91X or 91F), trip distance, time of travel, vehicle occupancy, mode (drive alone or carpool), and whether they had a flexible arrival time. Participants also provided various personal and household characteristics. The sample we use consists of 438 respondents. 3

The second sample is a two-stage mail survey collected by us through the Brookings Institution, including both RP and SP elements. For the Brookings sample, a market research firm, Allison-Fisher, Inc., custom-designed a survey to our specifications and mailed it to SR91 commuters who were members of either the National Family Opinion or the Market Facts nationwide household panel. A screener was first used to identify motorists who made work trips covering the entire 10-mile segment and thus had the option of using either roadway (91F or 91X). Survey respondents reported on their daily commute for an entire five-day workweek, providing information on the same items as mentioned above. The same people were subsequently sent an SP survey containing eight hypothetical commuting scenarios based on SR91. For each scenario, they were given hypothetical tolls, travel times, and probabilities of delay on the two routes (one always being tolled, the

3 For more details about the Cal Poly sample see Sullivan et al. (2000). The sample also included some people who traveled on just a part of route 91F and then exited onto a new toll expressway going to Irvine and southern Orange County; we have not included these people in our analysis.
other always free), and asked which they would choose. The values presented in the scenarios were roughly aligned with a respondent’s normal commute. An illustrative scenario is shown in Appendix A.

Because of overestimates of how many respondents would actually face a choice between 91F or 91X, we had to survey three waves of potential respondents—in December 1999, July 2000, and September 2000—to assemble an adequate sample. The final Brookings sample consists of 110 respondents: 84 people providing 377 daily observations on actual behavior (RP) and 81 people providing 633 separate observations on hypothetical behavior (SP), with 55 people answering both surveys.²

3. Econometric Framework

Our basic model specifies commuters’ choice of express or regular lanes as conditional on related choices such as residential location, travel mode (car or public transportation), time of day, and car occupancy. Integrating all these decisions with lane choice would add theoretical richness to the analysis.⁵ However, it would do so at the cost of considerable complexity while, in our case, adding little empirical insight. For instance, mode choice is unlikely to be important because public transportation has a very small share of travelers in the corridor that we are studying, while location choice is typically a long-run decision and the express lanes had been open only a few years. Time-of-day choice is more likely to be relevant, and we discuss our treatment of it at the end of this section.

---

⁴ Everyone who answered the SP survey had previously returned an RP survey, thereby providing key socio-demographic data; but 26 of them did not complete the RP questions about lane use and so appear only in the SP sample.

⁵ Models of several simultaneous discrete choices such as route, mode, and car occupancy are exposit, for example, in Train (1986). Theories incorporating traveler choice of time of day are provided by Henderson (1981) and Small (1982). Wilson (1992) considers simultaneous choice of residential location and time of day for work trips. An applied model for London by Bates et al. (1996) simultaneously predicts trip generation, destination, mode, time of day, and route.
We did find it tractable to analyze two of the other choices jointly with lane choice: car occupancy and whether to acquire a transponder. But, as we discuss later, we find that analyzing these choices jointly does not materially affect our findings. So for simplicity, our base model considers only lane choice. We specify it to be conditional on car occupancy and simultaneously determined with transponder acquisition, with neither of these decisions explicitly modeled.

Formally, a motorist $i$, facing an actual or hypothetical instance $t$ of choice between commuting lanes, chooses the option that maximizes a random utility function. Let

$$U_{it} = \theta_i + \beta_i X_{it} + \varepsilon_{it}$$

be the utility difference, so that the express lanes are chosen whenever $U_{it} > 0$. Variables included in $X_{it}$ measure the toll difference $C_{it}$, travel-time difference $T_{it}$, and (un)reliability difference $R_{it}$ between the two alternatives. The values of travel time and reliability are defined as:

$$VOT_i = \frac{\partial U_{it}}{\partial T_{it}}; \quad VOR_i = \frac{\partial U_{it}}{\partial C_{it}}.$$

As the notation indicates, the models are specified so that $VOT$ and $VOR$ depend on the individual traveler $i$ but not on the separate instance $t$ that a choice is made. (However, they may depend on whether a given individual is answering an RP or an SP question, a distinction we will add to the notation as appropriate.)

To capture heterogeneity, we specify parameters $\theta_i$ and $\beta_i$ in (1) as follows:

$$\theta_i = \bar{\theta} + \phi W_i + \xi_i$$

$$\beta_i = \bar{\beta} + \gamma Z_i + \zeta_i.$$

Observed heterogeneity is captured by the effects of measured variables $W_i$ and $Z_i$, while unobserved heterogeneity is captured by the random terms $\xi_i$ and $\zeta_i$. The scalar $\xi_i$ indicates an indi-
individual’s unobserved alternative-specific preferences, whereas the vector $\zeta_i$ represents an individual’s unobserved preferences regarding travel characteristics.\textsuperscript{6} Specifying individual alternative-specific preferences, in addition to $\varepsilon_{it}$ in (1), accommodates a panel-type data structure arising from an individual’s repeated observations $t$. Note that the derivatives in (2) depend on variables $Z_i$ and also contain components of $\zeta_i$, causing both observable and stochastic variability in the coefficient ratios $VOT$ and $VOR$.

We assume $\xi_i$ and all the components of $\zeta_i$ are distributed normally, independently of each other, and independently of $\varepsilon_{it}$:

$$\xi_i \sim N(0, \sigma_\xi^2), \quad \zeta_i \sim N(0, \Omega)$$

with $\Omega$ diagonal.\textsuperscript{7}

We denote our two data sets by superscripts $B$ (for Brookings) and $C$ (for Cal Poly). As noted, the Brookings data contain both RP and SP responses (further denoted with superscripts $R$ and $S$) and contain multiple responses from the same individual. The Cal Poly data are RP only and purely cross-sectional. As described later, in the RP portion of the Brookings data we form one choice variable from a motorist’s multiple-day observations; so all the RP observations have a single choice instance $t$, and we can write $\varepsilon_{it}^{BR} = \varepsilon_{iR}^{BR}$ and $\varepsilon_{it}^C = \eta_i^C$.

One would expect correlation between the RP and SP error terms for those individuals who answered both the Brookings RP and SP surveys. We account for this by splitting the corresponding error terms in (1) into two independent parts:

\textsuperscript{6} These two sources of randomness are called “preference” and “response” heterogeneity, respectively, by Bhat and Castelar (2002).

\textsuperscript{7} Normally distributed random parameters may cause a given coefficient to take the “wrong” sign with some probability, which is not a serious problem if that probability is small. We tried log-normal or truncated normal distributions for the random coefficients, but were unable to reach convergence—a common problem with mixed logit noted by Train (2001) among others.
where $\rho$ captures the correlation between the error terms for a given individual and the random terms $\eta_{i}^{BR}$ and $\eta_{i}^{BS}$ are assumed independent of each other. We assume $v_{i}^{BR} \sim N(0,1)$.

The full joint model is then represented by the following utility differences between express and regular lanes:

\[
U_{i}^{BR} = \theta_{i}^{BR} + \beta_{i}^{BR} X_{i}^{BR} + v_{i}^{BR} + \eta_{i}^{BR}
\]  
(8)

\[
U_{i}^{BS} = \theta_{i}^{BS} + \beta_{i}^{BS} X_{i}^{BS} + \rho v_{i}^{BR} + \eta_{i}^{BS}
\]  
(9)

\[
U_{i}^{C} = \theta_{i}^{C} + \beta_{i}^{C} X_{i}^{C} + \eta_{i}^{C}
\]  
(10)

with parameters $\theta_{i}$ and $\beta_{i}$ specified as in (3) and (4) except that the error terms $\xi_{i}^{BR}$ and $\xi_{i}^{C}$ are set to zero because, with only one observation per individual, they are redundant given the presence of $\eta_{i}^{BR}$ and $\eta_{i}^{C}$. Index $i$ in (8)-(10) runs through all individuals, with the appropriate equation(s) aligned with the sample that an individual is drawn from. We also assume that $\eta_{i}^{BR}$, $\eta_{i}^{BS}$, and $\eta_{i}^{C}$ have independent logistic distributions, which yields the familiar logit formula for the choice probability conditional on other random parameters.\(^8\) Our treatment of heterogeneity is therefore an example of a mixed logit model (Brownstone and Train (1999), McFadden and Train (2000)).\(^9\)

\(^8\) Equivalently, we assume each $\eta_{i}$ is the difference between two random variates with independent extreme-value (double-exponential) distributions.

\(^9\) One could also represent heterogeneity using a discrete rather than a continuous distribution across people; but as noted by Allenby and Rossi (1999), it is difficult in practice to estimate the number of parameters needed to make a discrete distribution adequately represent the actual distribution.
As is usual in combining RP and SP data sets, we allow the variances of \( \eta^{BR}_i \) and \( \eta^{BS}_i \) to differ, indicating that there may be different sources for random preferences over revealed and stated choices. We also let \( \eta^C_i \) have its own distinct variance because the data sets have somewhat different questionnaire formats. All this is accomplished by normalizing the variance of \( \eta^{BR}_i \) (to \( \pi^2/3 \) as in the binary logit model) and estimating the ratios

\[
\mu^{RS} \equiv \frac{\sigma^{BR}}{\sigma^{BS}}
\]

\[
\mu^C \equiv \frac{\sigma^{BR}}{\sigma^C}
\]

where each \( \sigma \) is the standard deviation of the corresponding \( \eta_i \) or \( \eta_{it} \).

Our specification allows considerable generality in how choices are determined in the three samples (BR, BS, C) relative to each other. Of course, one can improve statistical efficiency by combining the samples only if the model imposes some constraints. We therefore assume that some coefficients are identical in two or more of the samples. This enables the SP responses to help identify some key heterogeneity parameters, whose effects would otherwise be obscured by multicollinearity in the RP-only data; yet we still estimate different mean coefficients on travel variables in the RP and SP samples as a protection against contamination by SP survey bias. We calculate VOT and VOR from the RP and SP parameters separately, but we rely primarily on the RP values for policy analysis.

We also assume that commuters who travel at different times of day have a common set of parameters. This approach is necessary if we wish the RP portion of our sample to help identify the coefficients of travel variables, because in our data everyone traveling at a given time of day faces the same observable travel conditions. Fortunately, we can use the SP sample to test
the validity of this identifying assumption; we can also relax it somewhat by allowing each time
of day to have its own fixed effect on lane choice. We describe these experiments in Section 5.

The parameters of the model are estimated by maximizing the log of the Simulated Like-
lihood function:

$$SL(\psi_1, \psi_2) = \prod_i \frac{1}{R} \sum_{r=1}^{R} \prod_{k,t} P(y_{it}^k | \psi_1, \Theta_i^r)$$

(13)

where $\psi_1$ is the vector of all non-random parameters in the model; $\psi_2$ is the vector of parameters
describing the distribution of all random components other than $\eta_i^k$; $\Theta_i^r$ is a random draw from
the latter distribution; and $P(y_{it}^k | \psi_1, \Theta_i^r)$ is the individual’s choice probability for the observed
choice at instance $t$ in data subsample $k$, taking a binary logit form based on (8)-(10).10 We esti-
rate the model using $R=4,500$ random draws for each individual.11

4. Data and Specification

Table 1 summarizes commuters’ behavior and socioeconomic characteristics from the
Brookings RP and SP and the Cal Poly RP samples. Values for the Brookings data are broadly con-
sistent with population summary statistics, indicating it is a representative sample.12 Its median

---

10 The likelihood function itself is written out in an appendix available from the authors. It contains an integral over
the random parameter vector $\Theta_i$, which is numerically approximated by the Monte Carlo draws in (13). Lee (1992)
and Hajivassilio and Ruud (1994) show that under regularity conditions, the parameter estimates obtained by maxi-
mizing the simulated likelihood function are consistent when the number of replications rises at any rate with the
sample size and are asymptotically normal and equivalent to maximum likelihood estimates when the number of
replications rises faster than the square root of the sample size.

11 We found that the coefficients and standard errors both stabilized with $R$ somewhere between 1500 and 4500. We
also tried generating draws using randomized Halton sequences (Train 2003, pp. 234-235) with virtually no changes
in the results.

12 The distributions of the RP sample’s commuting times and route share are close to the ones in the Cal Poly data
and in other survey data collected by University of California at Irvine in 1998 (Lam and Small (2001)). The socio-
economic data are mostly consistent with Census information, and diverge where appropriate: our median income
(approximately $46,250) is higher than the average incomes in the two counties where our respondents lived
($36,189 and $39,729 in 1995, as estimated by the Population Research Unit of the California Department of Fi-
nance), presumably because the Brookings sample contains only employed people.
household income (assigning midpoints to the income intervals) is $46,250; we estimate the average wage rate to be about $23 per hour.\textsuperscript{13}

In the Brookings RP sample, which contains information for multiple days, choices do not vary much from day to day: 87 percent of respondents made the same choice every day during the survey week, and nearly all the others varied on only one day. Nearly half of the Brookings RP respondents do not have a transponder and thus have chosen in advance not to use the express lanes on a given day.\textsuperscript{14}

The Cal Poly sample is partly choice-based, some of it being obtained from license-plate observations on SR91. However, its express-lane share is so similar to the Brookings sample, which is random, that correcting for choice-based sampling makes virtually no difference to the estimation results. Time-of-day patterns in the Cal Poly data are also similar to those in the Brookings RP sample, as are most other observables including age and sex. The Cal Poly sample does have higher household incomes and shorter trip distances, evidently drawing from a narrower geographical area where people occupy closer-in and presumably higher-priced housing; these differences should not bias our estimates, however, because commuters’ choices are conditioned on income and trip distance.

\textit{Dependent Variable}

\textsuperscript{13} Data from the US Bureau of Labor Statistics (BLS) for the year 2000 record the mean hourly wage rate by occupation for residents of Riverside and San Bernardino Counties. We combine the BLS occupational categories into six groups that match our survey question about occupation, then assign to each person in our sample the average BLS wage rate for that person’s occupational group. We then add 10 percent to reflect the higher wages likely to be attracting these people to jobs that are relatively far away.

\textsuperscript{14} Among the 41 Brookings RP respondents who have a transponder, 11 made different choices on different days; this amounts to 27% of those with a transponder but only 13% from the entire Brookings RP sample. The latter statistic is relevant for judging constancy of choice because we model lane choice unconditionally on transponder (hence not getting a transponder is a natural concomitant of a persistent decision to take the free lanes). Six of those eleven respondents who varied their choice made trips on all five weekdays; four of the six chose the free lanes on all but one day, and one chose the express lanes on all but one day, leaving only one who made a 3:2 split.
Because the Brookings RP sample contains observations on more than one day for a
given individual, we can take each day as a separate observation; we call this a trip-based model.
Alternatively, we can take each person as a separate observation and define a dependent variable
based on the frequency of using toll lanes; we call this a person-based model.

The person-based model has certain advantages. First, it enables us to model the trav-
eler’s decision to get a transponder as an implicit part of lane choice rather than as a separate de-
cision, thus simplifying the analysis. Second, its simple error structure makes it easy to combine
with other data. These advantages may be weighed against only a small information gain from
the trip-based model because, as noted, few travelers changed behavior from day to day. Because
we model choice unconditionally on acquiring a transponder, it does not matter that a somewhat
larger proportion of those with a transponder (albeit only 27%) do change lanes from day to day.
Furthermore, preliminary estimations indicated that person-based and trip-based models yielded
similar results. We therefore focus here on the findings from person-based models. We explored
alternative ways of specifying the person-based dependent variable, and settled on a binary out-
come defined as one if the motorist used the express lanes for half or more of reported commut-
ing trips, zero otherwise.15

Measurement of RP Travel Time Variables

---

15 Under this definition, 24% of the Brookings RP sample chose the express lanes. We omitted the few respondents
who have a transponder but traveled two days or less, because defining a frequency for them involves too much er-
ror. Our use of a binary dependent variable can be thought of as a special case of an ordered logit model that divides
the possible [0,1] interval (for fraction of trips made on the express lanes) into two or more sub-intervals. We ex-
plored several such ordered models, including some with three or four choices; based on Vuong’s (1989) test for
non-nested models, we could not reject any of the specifications we tried in favor of any other one, and all gave
similar results for the parameters of interest. In those few cases where an independent variable varies across days, it
is defined as the average value over the days reported. An exception is the “flexible arrival time” dummy defined in
the next subsection; it is set to one if the respondent indicated a flexible arrival for half or more of the reported days,
zero otherwise. (In fact, only five people reported any daily variation in this variable.)
This analysis seeks to account for the effect of average travel time and travel-time reliability on behavior. Reliability is important to commuters because they may be less productive or suffer a loss in pay if they arrive later than planned. Under such circumstances, expected travel cost rises with uncertainty; and if the cost of being late is greater than that of being early, as indicated by empirical results of Small (1982), then expected costs are especially sensitive to the right tail of the distribution of travel time.

We estimated travel time and reliability based on actual field measurements on SR91 taken at different times during the six-hour morning period covered by our data. Students from University of California at Irvine drove repeatedly on the free lanes, clocking the travel time between prescribed points. These measurements were taken on eleven different days. Ten coincided with the days covered by the second and third waves of the Brookings survey; the eleventh day was two months before the first wave of the Brookings survey and one month before the Cal Poly survey.

We posit that for any given time of day, there is a distribution of travel times across all weekdays of the year that is known to travelers based on their experience. We assume that each travel time we observe is drawn randomly from that distribution. By asserting that motorists care about trip time and reliability when making travel decisions, we are saying that they consider the central tendency and the dispersion of that distribution; accordingly, we need measures of both as independent variables in the model. Plausible measures of central tendency include the mean and the median; we find the model fits slightly better (in terms of log-likelihood achieved) using the median. Possible measures of dispersion include the standard deviation and the inter-quartile

16 It is reasonable for several reasons to assume that motorists’ lane choices are based mainly on their knowledge of the distribution of travel times across days, not on the travel time encountered that day. Previous survey results described by Parkany (1999) suggest that whatever information travelers on this road have about conditions on a given day is mostly acquired en route through radio reports, and thus has limited value to them because it cannot affect their departure time. In addition, there is no sign displaying traffic information, and our field observations suggested that the amount of congestion encountered before the entrance to the express lanes could not be used to accurately predict travel delays along the full 10-mile segment.
difference. However, as noted, travelers are most concerned with the occasional significant delay that causes them to arrive late, so they are likely to pay particular attention to the upper tail of the distribution of travel times. Accordingly, we investigate the upper percentiles.

We use non-parametric smoothing techniques to estimate the distribution of travel-time savings from taking the express lanes, by time of day.\textsuperscript{17} A discussion of the methodology is contained in appendix B, and some results are shown in Figures 1 and 2. Figure 1a shows the raw field observations of travel-time savings. The non-parametric estimates of mean, median, and 80\textsuperscript{th} percentile are superimposed. Median time savings reach a peak of 5.6 minutes around 7:15 a.m. As indicated by the 90\% confidence band for the median in figure 1b (lines labeled “CI-UP” and “CI-LO”), the travel time savings are measured quite precisely.

Figure 2a shows the same raw observations after subtracting our non-parametric estimate of median time savings by time of day. An interesting pattern emerges. Up to 7:30 a.m., the scatter of points is reasonably symmetric around zero with the exception of three data points. But after 7:30 the scatter becomes highly asymmetric, with dispersion in the positive range (the upper half of the figure) apparently continuing to increase until well after 8:00 a.m. while dispersion in the negative range decreases. This feature is reflected in the three measures of dispersion, or unreliability, that are also shown in the figure: the standard deviation and the 80\textsuperscript{th}-50\textsuperscript{th} and 90\textsuperscript{th}-50\textsuperscript{th} percentile differences. The standard deviation peaks at roughly 7:45 a.m., the other two considerably later. The reason for these differences is that traffic in the later part of the peak is affected by incidents occurring either then or earlier. This mostly affects the upper tails of the distribution of travel-time savings and so is most apparent in the percentile differences. The

\textsuperscript{17} We never observed any congestion on the express lanes. Thus, to simplify the analysis, we assume that travel time on them is equal to the travel time we observed on the free lanes at 4:00 a.m., when there was no congestion: namely, 8 minutes, corresponding to a speed of 75 miles per hour.
standard deviation, by contrast, is higher early in the rush hour because of days with little con-
gestion—showing up as negative points in Figure 2a. Such dispersion is probably less relevant to
carpoolers than dispersion in the upper tails. In addition, the percentile differences are consid-
erably less correlated with median travel time than is the standard deviation. Thus, we prefer the
percentile differences as reliability measures. In our estimations, we obtained the best statistical
fits using the 80th-50th percentile difference.¹⁸ Note that the confidence bands in figure 2b sug-
gest that in all likelihood unreliability continues to rise until after 8:00 a.m., generating a pattern
that is different from that of median travel time.

Other Variables

The express-lane toll for a given trip is the published toll for the relevant time of day, dis-
counted by 50 percent if the trip was in a carpool of three or more.¹⁹ Other potentially important
influences on lane choice include trip distance, annual per capita household income, age, sex,
household size, and a dummy variable that indicates whether the commuter had a flexible arrival
time.²⁰ We also explored a number of other variables, such as occupation, education, vehicle oc-
cupancy, and size of workplace, but they had little explanatory power and did not influence the
other coefficients so they are omitted here.

¹⁸ The 90th-50th percentile difference fit almost as well as the 80th-50th difference (in terms of log-likelihood) and
resulted in similar coefficient estimates. The 75th-50th percentile difference (an additional measure we tried) and the
standard deviation fit noticeably less well and gave statistically insignificant results for the reliability measure.

¹⁹ The relevant toll is determined by the time of day that the commuter reported passing the sign that indicates the
toll level. Tolls on westbound traffic during the morning commute hours covered in this study ranged from $1.65 to
$3.30; they have been raised subsequent to our data collection. We asked respondents their vehicle occupancy, and
those who did not report it are assumed not to have carpooled. To guard against systematic bias from this assump-
tion, we specified a dummy variable identifying these respondents, but it had no explanatory power so it is not in-
cluded in the models reported here. Due to the uneven quality of answers about carpooling, and lack of knowledge
of ages or characteristics of passengers, we did not attempt to prorate the toll among vehicle occupants.

²⁰ The question, identical in both surveys, was: “Could you arrive late at work on that day without it having an im-
 pact on your job?” Similar measures of work-hour flexibility have been found to have important implications for
consumer choice of time of day of travel (Small (1982)). Including this measure here may control somewhat for the
possible endogeneity of the time-of-day choice.
Most SP variables correspond exactly to the RP variables. An exception is the measure of unreliability, because we did not think survey respondents would understand statements about percentiles of a probability distribution. Instead, we specified in our SP scenarios the probability of being delayed 10 minutes or more.21

In our base model, we specify random coefficients for the time and unreliability variables, as well as random constants. We allow the RP and SP values of time and of reliability to differ, but we combine the power of both data sources to estimate the random variation in those values. Specifically, for the RP and SP travel-time coefficients, we require the standard deviation of the random component to be identical. For the RP and SP reliability coefficients, because the measure of unreliability differs, we require the ratio of the standard deviation to the mean of the random coefficient to be identical.22

5. Estimation Results

Estimation results are presented in table 2. Beginning with the stochastic part of the model, the scale and correlation parameters describing the error structure (listed under “Other Parameters” in the table) are estimated quite precisely. The parameter \( \rho \) indicates that SP and RP responses from a single respondent are strongly correlated. The standard deviations of the random parameters (listed under “Pooled Variables”) are also estimated with quite good precision, and amount to roughly 25% and 100% of the corresponding mean coefficient.23

21 The probability was always stated for the trip as a whole. It was given as 0.05 for trips using 91X, and as 0.05, 0.1, or 0.2 for trips using 91F. The actual statement is: “Frequency of unexpected delays of 10 minutes or more: 1 day in X” where X=20, 10, or 5.

22 We also estimated a model that included random coefficients on cost as well as on travel time and reliability, with cost coefficients constrained the same way as the travel-time coefficients. However, the estimated standard deviation of the cost coefficient was somewhat unstable with respect to differences in the number and types of draws for the simulated likelihood, so we concluded that such a model is too rich for our data set.

23 The RP coefficient of travel time depends on distance; its median value in the sample is -0.69.
Most of the parameter estimates in the systematic part of the model are statistically significant and have the expected signs. Both the RP and SP coefficients indicate that commuters on average are deterred from the express lanes by a higher toll and are deterred from the free lanes by longer median travel times and greater unreliability. (Despite the interaction terms, this remains true throughout the full range of distance in our data.)

Observed heterogeneity is indicated by preferences that vary with income and trip distance. Consistent with expectations, motorists with higher incomes in the RP sample are less responsive to the toll. But, surprisingly, income is statistically insignificant in the SP sample, whether interacted with the toll (as shown) or entered as a lane-choice shift variable (as tried but not shown). For the RP variables, the effect of distance on the time coefficient is captured well by a cubic form with no intercept (i.e., median travel time is not entered by itself). The dependence of the value of time on distance is characterized by an inverted U, initially rising but then falling for trips greater than 45 miles. We conjecture, along with Calfee and Winston (1998), that this pattern results from two opposing forces: the increasing scarcity of leisure time as commuting becomes longer, and the self-selection of people with lower values of time into farther-out residences. In the SP sample, we allow the coefficient on travel time to differ between people with long or short actual commutes (who received different versions of the SP survey, as explained in Appendix A), but the difference is negligible.

We also find that women, middle-aged motorists, and motorists in smaller households are more likely to choose the toll lanes. Others have similarly found that women are more likely to use toll lanes (Parkany (1999), Lam and Small (2001), Yan et al. (2002)). To better understand why, we tried interacting gender, age, and household size under the hypothesis that working mothers, in particular, prefer the toll lanes due to tighter schedules; but we could not find a
measurable effect either on lane preference or on the disutility from unreliability or travel time. Surprisingly, we do not find a significant effect of having a flexible arrival time, possibly because this variable captures other job characteristics correlated with arrival flexibility. (We tried interacting it with time or reliability, but it still was statistically insignificant.)

We explored several specifications that differed in the extent to which they allowed coefficients to vary across RP and SP subsamples. In our preferred specification, we allow the mean values of their coefficients (including interactions with income and distance) to differ across RP and SP subsamples because one would expect people to react differently to cost, time, and reliability in a hypothetical than in an actual context; but we assume for simplicity that the individual characteristics affecting alternative-specific preferences (namely sex, age, flexible arrival, and household size) have the same effects across data sets. As described later, the constraint easily passes a likelihood ratio test and makes little difference to the estimated values of time and reliability.

As a methodological point, there was clearly a payoff from combining the RP and SP data because we were unable to obtain estimates of unobserved heterogeneity when we estimated the model using RP data alone. As for the rest of the specification, it is reassuring that if the model of table 2 is re-estimated without unobserved heterogeneity in parameters, its RP coefficients are nearly identical to the corresponding coefficients of the model estimated on RP-only data.

Motorists’ Preferences and Heterogeneity

We use the estimated parameters to calculate the distributions of motorists’ implied values of time (VOT) and reliability (VOR), making use of the Brookings RP and SP samples. (We use these samples because they are purely random and seem to better represent the characteristics
of the relevant population than the Cal Poly sample.) We compute the resulting median values for VOT and VOR and three types of heterogeneity—observed, unobserved, and total—across the population. We measure each type of heterogeneity by performing Monte Carlo draws from our estimated distribution of random parameters and enumerating across the sample; we report the interquartile range (i.e. the difference between 75\textsuperscript{th} and 25\textsuperscript{th} percentile values) since this measure is relatively robust to the high upper-tail values occasionally arising in ratios.

Results are shown in the first column of Table 3. The median value of time based on commuters’ revealed preferences is $21.46/hour; at 93 percent of the average wage, it is near the upper end of the range expected from previous work (Small (1992)). The median RP value of reliability is $19.56/hour.

What are the implications of these median values for observed travel conditions on SR91? In our data, median time savings at the height of rush hour are 5.6 minutes; thus, the average commuter would pay $2 to realize these savings. Unreliability peaks at 3 minutes; thus, the average commuter would pay $0.98 to avoid this possibility of unanticipated delay. Time savings therefore account for two-thirds and improved reliability accounts for one-third of the attraction of the express lanes. The median traveler would be willing to pay about $3 for the express lanes; the actual peak toll of $3.30 should therefore attract somewhat less than half of the total peak traffic to the express lanes—which, in fact, it does. This calculation does not include the alternative-specific constant, which may reflect other advantages of the express lanes such as safety or better emergency response, or disadvantages such as needing to obtain a transponder.

Our estimates of total heterogeneity are roughly 50\% to 100\% of the median value for VOT and greater than the median for VOR, indicating that commuters exhibit a wide distribution of preferences. It is interesting that there is considerable heterogeneity from unobserved sources,
verifying the importance of using random parameters to capture motorists’ taste variation. To be sure, unobserved heterogeneity reflects the limitations of empirical work and presumably could be reduced if it were possible to measure additional sources of individuals’ preferences.

The second column of the table reports 90 percent confidence intervals for all the computations, based on Monte Carlo draws from the estimated sampling distribution of the unknown parameters of the model. We find for all estimates, including the heterogeneity measures, that the confidence intervals exclude zero.

The implied SP values of time are much smaller on average than the RP values. This finding may reflect a tendency of travelers to overstate the travel time they lose or would lose in congestion. For example, suppose a motorist is in the habit of paying $2.00 to save 5 minutes, but perceives the savings as 10 minutes. That motorist may then answer SP questions as if he or she would pay $2.00 to save 10 minutes—yielding an SP value of time that is only half the value used in actual decisions. As for reliability, the median SP value of $5.40 per incident means that the median motorist in our sample would pay $0.54 per trip to reduce the frequency of 10-minute delays from 0.2 to 0.1.

*Sensitivity to Identifying Assumption and Alternative Specifications*

As noted, the parameters of our model are identified by combining people who travel at different times of day into a single sample. The assumption underlying this strategy is that any unobserved influences on lane choice do not vary systematically by time of day; if they did, they would be correlated with cost, time, and reliability and therefore bias those coefficients. The validity of the assumption depends partly on how well our observed variables capture taste variation across times of day. Fortunately, we have many variables that play this role including in-

---

24 Sullivan et al. (2000, p. xxiii) and Supernak et al. (1999, p. 31) provide evidence of this tendency by surveying travelers participating in two California road-pricing experiments, including the one used in this study.
come, trip distance, trip purpose, flexibility of arrival time, sex, age, household size, occupation, marital status, and education. For example, a motorist’s sex is likely to be an important source of taste variation; we know it is correlated with travel conditions because females constitute only 15% of those commuters traveling during the interval 4-5 a.m., but 39% of the 7-8 a.m. group. Another potential source of taste variation that varies by time of day is trip purpose; the proportion of respondents whose trips are work trips varies from 100% at the earliest time to 58% at the latest time. (All Brookings respondents, but not all Cal Poly respondents, are commuters.)

We provide a more formal test by taking advantage of the fact that 55 members of the SP sample, providing 433 observations, told us the time of day at which their actual trip normally took place. In this sample, travel time and reliability are uncorrelated with the time of day of travel as part of the SP survey design, so we can include time-of-day dummies without much loss of precision in other parameter estimates. If time of day is correlated with unmeasured influences on lane choice, we would expect those influences to be captured in such a model; by comparing the results with and without the dummies we can see whether VOT and VOR are affected. We carried out this comparison using five time-of-day dummies: one for each hour of the morning period other than the base hour of 7-8 a.m. We found that adding time-of-day dummies decreases the median SP values of time and reliability less than 10 percent and increases their unobserved heterogeneity (as measured by the inter-quartile range) less than 8 percent. Thus it appears that any unobserved influences on lane choice that vary by time of day do not have much effect on the results of interest, at least in SP data.

We performed a further check by re-estimating the joint RP/SP model with time-of-day
dummy variables.\textsuperscript{25} The SP results just described had suggested that the five hourly time-of-day dummies (with the omitted period 7-8 a.m., the busiest time of day) were not statistically significantly different from each other; thus, we specified a single 7-8 a.m. dummy. Its estimated coefficient is –1.64 with standard error 1.29. The first two columns of table 4 show the effects on our VOT and VOR estimates.\textsuperscript{26} Including the time-of-day dummy increases estimated median RP values of time and reliability and also increases slightly the unobserved heterogeneity in the RP values, although none of these changes move the estimates outside the original 90-percent confidence bounds. Therefore, our main conclusions—that values of time and reliability are high and contain considerable unobserved heterogeneity—are if anything strengthened. We view this experiment with caution, however, because it ignores relevant information. Thus, the base model has greater precision than the model that includes a time-of-day dummy, as indicated by the smaller 90% confidence intervals for most of the quantities shown. For this reason and because the base model includes variables that vary strongly with the time of travel and that capture essential aspects of preference variation, we prefer it for policy analysis.

We also estimated models with both fewer and more constraints across the RP and SP coefficients. For each, we computed the distributions of the values of time and reliability and tabulated the median, inter-quartile differences, and so forth, as in Table 3. The results are virtually indistinguishable from the base model. Furthermore, we found using likelihood-ratio tests that we could not reject our preferred specification against less constrained models (i.e., with additional distinct RP and SP coefficients). In fact, we found that it would be statistically justifiable

\textsuperscript{25} In this case we retained all the SP data. For those not answering the time-of-day question (26 people), we set the dummy variable to zero and included a distinct alternative-specific constant and a distinct error variance to avoid making any implicit assumption about their time of day of travel.

\textsuperscript{26} The parameter estimates that underlie all the VOT and VOR calculations in this section are in an appendix available from the authors upon request.
to impose a similar constraint on the cost coefficients, but we preferred not to do so because this coefficient is critical to our VOT and VOR calculations.

Finally, we estimated a model explaining the simultaneous choice of vehicle occupancy, transponder acquisition, and lane to explore the sensitivity of our findings to possible simultaneity bias. We created nine alternatives from the combinations of vehicle occupancy (1, 2, or greater); transponder acquisition (yes or no); and lanes (free or express). Alternatives were not created that involved the use of the express lanes without a transponder because such alternatives are illegal and therefore not available. Conditional on the random coefficients, choice among the nine alternatives is multinomial logit. The implications of the parameter estimates for the distribution of the values of time and reliability are summarized in the last column of Table 4. The values are quite consistent with those from the base model and they support our earlier arguments for assuming that the time of day of travel and vehicle occupancy are exogenous and for treating transponder choice as implicitly determined with lane choice.

6. Implications for Road Pricing Policy

Does preference heterogeneity create a strong case for differentiated services? We develop a simulation model using our econometric results to examine several pricing policies for a situation closely resembling the SR91 road-pricing experiment. Two 10-mile roadways, Express and

---

27 Thus we can observe all of the respondents’ choices regardless of whether they have a transponder. The resulting model has an unbalanced nesting structure where more alternatives involve a transponder than do not involve a transponder.

28 We used a random error structure that provides an analog of nested logit, in which the three alternatives involving express lanes can have a closer degree of substitutability among themselves than do other alternatives. This is one of several advantages of mixed logit described by Brownstone and Train (1999). In this extended model, because the alternatives are more disaggregated than in the lane-choice-only model, the sample shares of some alternatives deviated from the shares in our random subsamples and we could therefore no longer ignore the choice-based sampling schemes of our Cal Poly subsamples; we accounted for this using the weighted exogenous sample maximum likelihood estimator (WESMLE) of Manski and Lerman (1977).
Regular, are assumed to connect the same origin and destination and to have the same free-flow travel time. We model a four-hour peak period and find equilibria by iterating between the supply and demand sides of the model.

The supply side is a standard static congestion model in which travel delays are proportional to the fourth power of the volume-capacity ratio, as for example in Vickrey (1963). Following the well-known formula of US Bureau of Public Roads (1964), we assume that travel delay is 15% of the free-flow travel time when the volume-capacity ratio is equal to one. Capacity is taken to be 4,000 vehicles per hour for the Express roadway and 8,000 for the Regular roadway. We also assume that unreliability, as measured by the difference between 80th and 50th percentile travel times, is a constant fraction 0.3785 of travel delay—consistent with the fraction observed on the unpriced lanes in our SR91 data averaged over the four-hour peak period (5-9 a.m.).

The demand side of the model consists of a sample enumeration procedure using the parameters estimated for RP choices in Table 2. The enumeration uses the Brookings RP sample because it is random and representative of the population, as discussed in Section 4. Sample enumeration consists of calculating the choice probability for each of the 84 individuals in that sample using the observed characteristics of that individual and Monte Carlo simulation of the integral over random parameters. Each subsample member is assumed to represent \( N/84 \) potential travelers whose values of time, values of reliability, and choices are all distributed in the same way as those calculated for that member of the subsample. The calibration of \( N \) is described below.

Before enumeration, however, we augment our estimated demand model to allow for a non-zero price-elasticity of demand for automobile travel on the corridor. We cannot measure this elasticity from our data because we surveyed only people who travel on SR91. We therefore assume that our demand model is one branch of a nested-logit structure in which the other branch
consists of an “outside” alternative (denoted \(j=-1\)) indicating a choice not to travel by automobile on this corridor. The unconditional probability of individual \(i\) choosing roadway \(j\) (\(j=0\) for Regular, \(1\) for Express) is given by:

\[
P_{ij} = \int P_{ij} (j \geq 0; \beta) \cdot P_{ij} (j \geq 0; \beta) \cdot f_i(\beta) d\beta
\]

\[
= \int \frac{\exp[I_j (j \theta_{i_0} + \beta X_{ij})]}{\exp(I_i)} \cdot \frac{\exp(\lambda I_i)}{\exp(\theta_{i_0}) + \exp(\lambda I_i)} \cdot f_i(\beta) d\beta
\]

(14)

where

\[
I_i = \ln \sum_{j=0}^{J} \exp[I_j (j \theta_{i_0} + \beta X_{ij})]
\]

(15)

is the “inclusive value” for automobile travel on the corridor; \(\lambda \in (0,1]\) is a scale parameter for the nest \(\{j=0,1\}\); \(X_{ij}\) is the data vector such that \(X_{i1} - X_{i0} = X_{i}^{BR}\) of equation (8); \(\theta_{i_0}\) is the utility of the outside alternative; \(\theta_{i_0} = \bar{\theta} + \phi W_i\) (it is non-random in the BR sample as explained earlier); \(f_i(\cdot)\) is a normal probability density function with mean \(\bar{\beta} + \gamma Z_i\) as defined by equations (4)-(5); and \(J=1\) when two distinct roads are offered. In policy scenarios where the two sets of lanes have zero or identical tolls, we collapse them into one alternative on the assumption that the alternative-specific preference for the express roadway estimated in our econometric model, \(\bar{\theta} + \phi W_i\), reflects special characteristics of express lanes that disappear when the lanes are undifferentiated; so in this case \(J=0\) in (15), and (14) reduces to a binary logit model for road travel with scale \(\lambda\).

Consistent with previous analyses of the choice between two competing roadways, we assume that the substitutability between them is much greater than the substitution between either road and the outside option. Hence, the scale parameter \(\lambda\) is made as small as possible without causing numerical problems; we use \(\lambda=0.25\) and perform sensitivity analysis using a smaller value.
We then choose the two unknown parameters $\theta_{-1}$ and $N$ to produce a realistic price-elasticity of demand for road travel and a significant yet plausible amount of congestion when the roads are not tolled. Specifically, we choose parameters to produce a demand elasticity for road travel on the corridor of -0.21 and travel delays of 9.37 minutes under a no-toll policy. These values are conservative in the sense that a higher elasticity or greater initial congestion, either of which is plausible, would increase the potential benefits from pricing policies. We perform sensitivity analysis using an alternative elasticity value.

Based on this model, we can calculate tolls, travel times, traffic volumes, changes in consumer surplus, and the change in social welfare under several alternative pricing policies. Our base-case policy has no toll on either roadway. The change in consumer surplus relative to this base case is determined by the log-sum rule for nested logit (Choi and Moon (1997)):

$$\triangle CS_i = \frac{1}{\mu_i} \Delta \ln[\exp(\theta_{-1}) + \exp(\lambda I_i)]$$

where $\Delta$ indicates the difference between a given scenario and the base scenario, $I_i$ is given by

---

29 The demand elasticity is with respect to the “full price,” defined as the money price plus the perceived value of travel time and unreliability encountered. We obtained the values of the elasticity and the travel delays through the following procedure. First, we chose $\theta_{-1}$ and $N$ so that our model replicates conditions observed in summer 1999 as reported by Yan, Small, and Sullivan (2002): namely, peak travel delay (averaged over four hours) of 3.4 minutes in the free lanes and a money-price elasticity of -0.58 for travel in the Express Lanes, with the Express toll set to maximize revenue (presumed to be the policy of the private consortium that operated the Express Lanes at that time). This led to the implied full-price elasticity of -0.21 under no toll. To depict greater congestion (which in fact has occurred in the subsequent years), we then increased $N$ while adjusting $\theta_{-1}$ to maintain this same full-price elasticity, until travel delay doubled in the revenue-maximizing scenario; those parameters produced the no-toll travel delay of 9.37 min. shown in Table 5. We held the full-price elasticity constant because we believe it is a more fundamental parameter than the money-price elasticity of the Express Lanes.

30 Estimated elasticities of automobile travel with respect to fuel price tend to cluster around -0.25 (Luk and Hepburn (1993), Parry and Small (2003)); because fuel price is only one component of full cost and because reducing total automobile travel is only one way to reduce travel on SR91, we would expect a full-price elasticity for travel on SR91 to be larger in magnitude than the fuel price elasticity.
equation (15), and $\mu_i$ is the individual’s marginal utility of income. The change in social welfare is the sum of expected changes in all individuals’ consumer surplus, plus toll revenues.

Results are presented in Table 5. The tolls are determined using mean values of the estimated parameters ($\psi_1, \psi_2$). All other results are conditional on those tolls and are computed for each of 100 Monte Carlo draws from the sampling distribution of estimated ($\psi_1, \psi_2$); the numbers shown in the table are the median values from the resulting empirical distributions and, for the welfare change, the 5th and 95th percentile values (last row of the table).

The policies shown in the second and third columns set a price on the Express roadway that maximizes social welfare subject to the constraint that the Regular lanes have a zero toll. This “second-best” policy enables road pricing to be politically feasible, but sacrifices efficiency by not pricing all lanes. It improves welfare by a modest $0.35$ per base-case vehicle trip (column 2); whereas if users were identical that gain would be reduced by more than half (column 3). As expected, a pricing policy that enhances product differentiation yields greater benefits when consumers have heterogeneous preferences.

As a benchmark, we also calculated a first-best toll allowing for price differentiation. But we found that it contained very little differentiation. We suspect that many of the lowest value-of-time travelers choose the outside option when they face high tolls, so heterogeneity among the people actually traveling is reduced (cf. Verhoef and Small (2004)). We therefore model this policy as a uniform toll so that we do not attribute to it any product-differentiation advantages. This first-best

---

31 We express income in units of dollars per potential weekday trip, so that by Roy’s identity, $\mu_i$ is minus the coefficient of trip cost; see Small (1992), eqn. (2.22) and (2.35). The value of $\mu_i$ is thus computed using the third, fourth, and fifth coefficients shown in Table 2 and the values of the income dummies for that individual.

32 The bottom row of Table 5 shows that every scenario computed produces a welfare gain relative to the base case that is statistically significant because all the 5th percentile values shown are positive. Statistical tests in this section are performed by computing the relevant difference for each of the 100 Monte Carlo draws from the sampling distribution of estimated parameters.
toll achieves a much greater welfare gain than second-best pricing. But by imposing high direct costs on motorists, especially low and medium-income users, and by producing large disparities among travelers, it will most likely face resistance from policymakers.

Is there a politically feasible compromise between the second-best and first-best policies? Suppose we wish to achieve a greater welfare gain than the second-best policy, which has attained some degree of public acceptance; but we do not want to exacerbate distributional concerns by magnifying the largest consumer surplus losses. We find that accounting for heterogeneity can help. We propose a “limited differentiated toll” policy that charges tolls on both roadways, maximizing social welfare subject to two constraints imposed for political feasibility. First, the higher toll must be on the smaller-capacity “express” lanes. Second, the 25th percentile consumer surplus change (i.e., the 75th percentile consumer surplus loss) must not be any larger (within $0.01 tolerance) than in the second-best policy. This latter constraint ensures that the 25 percent of people with the lowest values of time and reliability are made no worse off than under second-best pricing.

The result, shown in column 5 of the table, is a lower but much more sharply differentiated toll than in the first-best case. From an efficiency perspective, it achieves a welfare gain nearly one-third greater than that of second-best pricing. From a distributional perspective, the limited differentiated toll treats users more evenly than other policies: compared to second-best, the inter-quartile variation in consumer surplus loss is reduced from $0.44 to $0.17 per trip, and the variation across our income groups is reduced from $0.94 to $0.20.\textsuperscript{33}

\textsuperscript{33} Based on the empirical distribution of these differences, obtained as part of computing confidence intervals, there is only a 10 percent chance that if the scenarios were designed as shown, with tolls based on calculations using our mean parameter estimates, the limited-differentiated toll would provide less welfare gain than the second-best toll. And this is a conservative test because in reality, tolls would probably be set using more information and therefore more precisely estimated parameters. Using the same test, there is a 15 percent probability that the inter-quartile variation would not be reduced, and a 6 percent probability that the income-group variation would not be reduced.
When we compare the limited differentiated toll with a limited uniform toll, which generates the same welfare gain, it is apparent that catering to heterogeneity is an important part of what enables the limited differentiated toll to soften the direct impacts of road pricing. As shown in the last column of the table, the uniform toll is much less effective than the differentiated toll at limiting the losses in consumer surplus.

Our qualitative findings hold up well under sensitivity analysis in which we repeated the comparisons using alternative sets of assumptions. We stress that our policy exercises are based on an experiment in which pricing is implemented on only a ten-mile segment of one highway, whereas congestion affects a much broader region. If the distributional advantages of differentiated pricing enabled it to be broadly adopted, its welfare gains would be greatly magnified.

7. Conclusion

Road pricing has been beloved by economists and opaque to policymakers for decades. Calfee, Winston, and Stempski (2001) rationalize this state of affairs by arguing that, in fact, few long-distance automobile commuters are willing to pay much to save travel time because those with a high value of time have self-selected into expensive housing located close to their workplace.

Our results confirm that those with very long commutes have substantially lower values of time. But we have also found rather high average values of time and reliability in our sample.

---

34 Three are especially worth mentioning. First, we reduced \( \lambda \) from 0.25 to 0.10, which hardly affected our findings but did produce some numerical instability. Second, we eliminated travelers’ loyalties for one roadway or the other, to reflect the results of applying our policies to otherwise identical roadways; we did this by setting \( \phi = 0 \) in equation (3) and adjusting \( \vec{\theta} \) to produce equal travel times under equal tolls, resulting in somewhat smaller welfare gains overall but showing a similar pattern across policies. (In this calculation we did not collapse the highway alternatives into one for cases of uniform tolls, as we did previously, because there would be no pure preference for having more highway options separate from their measured travel characteristics.) In the third alternative calculation, starting with this last model, we increased the full-price elasticity of travel from -0.21 (which we argued was conservative) to -0.33, a value used by Verhoef et al. (1996) and Small and Yan (2001); as expected, increasing the demand elasticity resulted in larger welfare gains for all the differential toll policies, especially the “limited differentiated toll.” All these results are contained in an appendix available from the authors.
of Southern California motorists. More importantly, we have found other sources of significant heterogeneity in highway travel preferences. One possible explanation of these findings is that in very expensive and congested metropolitan areas, many consumers who place a high value on travel characteristics are unable to find affordable housing to eliminate long commutes. In such a situation, we contend that pricing policies that cater to varying preferences enhance their chance of public acceptance.

Recent “value pricing” experiments have begun to explore this possibility by offering motorists the option to pay for congestion-free travel. But they leave part of the roadway un-priced, which severely compromises efficiency. We have demonstrated that pricing policies taking preference heterogeneity explicitly into account improve on the experiments by offering substantial efficiency gains with potentially more benign distributional consequences. By limiting the adverse net impact of tolls and time savings on consumer surplus, and reducing its variation across the affected population, differentiated pricing enhances the political viability of road pricing and thus offers policymakers a long-awaited opportunity to address the stalemates that impede transportation policy in congested cities.
Figure 1. Time Saving
Figure 2. Dispersion of Time Saving

2a, Standard Deviation, 80th-50th Percentile, and 90th-50th Percentile of Time Saving

2b, 80th-50th Percentile of Time Saving with 90% Confidence Interval
Table 1. Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Cal Poly-RP</th>
<th>Brookings-RP</th>
<th>Brookings-SP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Route Share:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>91X</td>
<td>0.26</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>91F</td>
<td>0.74</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>One-Week Trip Pattern:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never Use 91X</td>
<td></td>
<td></td>
<td>0.68</td>
</tr>
<tr>
<td>Sometimes Use 91X</td>
<td></td>
<td></td>
<td>0.13</td>
</tr>
<tr>
<td>Always Use 91X</td>
<td></td>
<td></td>
<td>0.19</td>
</tr>
<tr>
<td>Percent of Trips by Time Period:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4:00am-5:00am</td>
<td>0.11</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>5:00am-6:00am</td>
<td>0.22</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>6:00am-7:00am</td>
<td>0.23</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>7:00am-8:00am</td>
<td>0.20</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>8:00am-9:00am</td>
<td>0.14</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>9:00am-10:00am</td>
<td>0.10</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Age of Respondents:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;30</td>
<td>0.11</td>
<td>0.12</td>
<td>0.10</td>
</tr>
<tr>
<td>30-50</td>
<td>0.62</td>
<td>0.62</td>
<td>0.64</td>
</tr>
<tr>
<td>&gt;50</td>
<td>0.27</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>Sex of Respondents:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.68</td>
<td>0.63</td>
<td>0.63</td>
</tr>
<tr>
<td>Female</td>
<td>0.32</td>
<td>0.37</td>
<td>0.37</td>
</tr>
<tr>
<td>Household Income ($):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;40,000</td>
<td>0.14</td>
<td>0.23</td>
<td>0.24</td>
</tr>
<tr>
<td>40,000-60,000</td>
<td>0.24</td>
<td>0.60</td>
<td>0.59</td>
</tr>
<tr>
<td>60,000-100,000</td>
<td>0.40</td>
<td>0.15</td>
<td>0.13</td>
</tr>
<tr>
<td>&gt;100,000</td>
<td>0.22</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>Flexible Arrival Time:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.40</td>
<td>0.55</td>
<td>0.50</td>
</tr>
<tr>
<td>No</td>
<td>0.60</td>
<td>0.45</td>
<td>0.50</td>
</tr>
<tr>
<td>Trip Distance (Miles):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>34.23</td>
<td>44.76</td>
<td>42.56</td>
</tr>
<tr>
<td>Number of People in Household:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>3.53</td>
<td>2.91</td>
<td>3.44</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.51</td>
<td>1.63</td>
<td>1.55</td>
</tr>
<tr>
<td>Number of Respondents</td>
<td>438</td>
<td>84</td>
<td>81</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>438</td>
<td>377</td>
<td>633</td>
</tr>
</tbody>
</table>
### Table 2. Parameter Estimates: Joint RP/SP Model

**Dependent Variable:**
1 if chose toll lanes, 0 otherwise

**Independent Variable**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Coefficient (standard error)a</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RP Variables</strong></td>
<td></td>
</tr>
<tr>
<td>Constant: Brookings sub-sample ($\theta^{RR}$)</td>
<td>0.1489 (0.8931)</td>
</tr>
<tr>
<td>Constant: Cal Poly sub-sample ($\theta^{C}$)</td>
<td>-1.6349 (1.1040)</td>
</tr>
<tr>
<td>Cost ($)b,c</td>
<td>-1.8705 (0.5812)</td>
</tr>
<tr>
<td>Cost x dummy for medium household inc. ($60-100K$)</td>
<td>0.5438 (0.2549)</td>
</tr>
<tr>
<td>Cost x dummy for high household income (&gt; $100,000)</td>
<td>1.1992 (0.3849)</td>
</tr>
<tr>
<td>Median travel time (min.) x trip distance (units: 10 mi.)b</td>
<td>-0.4088 (0.1536)</td>
</tr>
<tr>
<td>Median travel time x (trip distance squared)</td>
<td>0.0695 (0.0276)</td>
</tr>
<tr>
<td>Median travel time x (trip distance cubed)</td>
<td>-0.0029 (0.0012)</td>
</tr>
<tr>
<td>Unreliability of travel time (minutes)b,d</td>
<td>-0.5778 (0.2435)</td>
</tr>
<tr>
<td><strong>SP Variables</strong></td>
<td></td>
</tr>
<tr>
<td>Constant ($\theta^{BS}$)</td>
<td>-1.6107 (0.8943)</td>
</tr>
<tr>
<td>Standard deviation of constantc ($\sigma_\xi$)</td>
<td>0.4800 (0.6305)</td>
</tr>
<tr>
<td>Costb,c</td>
<td>-1.0008 (0.2849)</td>
</tr>
<tr>
<td>Cost x dummy for high household income (&gt; $100,000)</td>
<td>0.2842 (0.9714)</td>
</tr>
<tr>
<td>Cost x dummy for medium household inc. ($60-100K$)</td>
<td>-0.2317 (0.5407)</td>
</tr>
<tr>
<td>Travel time (min.) × long-commute dummy (&gt; 45 min.)b</td>
<td>-0.1965 (0.0522)</td>
</tr>
<tr>
<td>Travel time × (1 − long-commute dummy)</td>
<td>-0.2146 (0.0618)</td>
</tr>
<tr>
<td>Unreliability of travel time (probability)b,f</td>
<td>-5.6292 (2.3819)</td>
</tr>
<tr>
<td><strong>Pooled Variables</strong></td>
<td></td>
</tr>
<tr>
<td>Female dummy</td>
<td>1.3267 (0.6292)</td>
</tr>
<tr>
<td>Age 30-50 dummy</td>
<td>1.2362 (0.5121)</td>
</tr>
<tr>
<td>Flexible arrival-time dummy</td>
<td>0.5903 (0.6994)</td>
</tr>
<tr>
<td>Household size (number of people)</td>
<td>-0.5497 (0.2248)</td>
</tr>
</tbody>
</table>
Standard dev. of coeff’s of travel time (derived from $\Omega$)                0.1658 (0.0457)
Ratio of std. dev. to the mean for coefficients of unreliability (derived from $\Omega$) 1.0560 (0.2754)

Other Parameters

Scale parameter: Cal Poly sample ($\mu^C$)               0.4118 (0.1688)
Scale parameter: SP sample ($\mu^{BS}$)                  1.3368 (0.3741)
Correlation parameter – RP and SP ($\rho$)               3.2882 (0.8320)

Summary Statistics

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>1155</td>
</tr>
<tr>
<td>Number of persons</td>
<td>548</td>
</tr>
<tr>
<td>Number of replications ($R$)</td>
<td>4,500</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-501.57</td>
</tr>
<tr>
<td>Pseudo R²</td>
<td>0.3704</td>
</tr>
</tbody>
</table>

---

$^a$ Standard errors reported are the “sandwich” estimate of standard errors from Lee (1995). That is, each is the square root of the corresponding diagonal element in the matrix $\hat{V} = (\hat{H})^{-1} \hat{P}(\hat{H})^{-1}$, where $H$ is the Hessian of the simulated log-likelihood function and $P$ is the outer product of its gradient vector (both calculated numerically). This estimate accounts for the simulation error in the likelihood function.

$^b$ All cost, travel-time, and unreliability variables are entered as the difference between values for toll and free lanes. In the RP data, the cost for free lanes is zero, travel time for toll lanes is 8 minutes, and unreliability for toll lanes is zero. In the SP data, cost, travel time, and unreliability are specified in the questions.

$^c$ Value of “cost” for the toll lanes is the posted toll for a solo driver (for RP data) or the listed toll in the survey question (for SP), less 50% discount if car occupancy is 3 or more. For SP, car occupancy is determined from a question asking whether the respondent answered as a solo driver or as part of a carpool, and if the latter what size carpool.

$^d$ Value of “unreliability” for the free lanes in the RP data is the difference between the 80$^{th}$ and 50$^{th}$ percentile travel times (see text).

$^e$ The estimation of a standard deviation of the constant $\theta_{it}^{BS}$, separate from the standard deviation of the overall random term $\eta_{it}^{BS}$, is made possible by the multiple observations for a given individual in the SP data sample (see equation (9)). Hence there is no comparable parameter for the RP samples, where $\sigma_\delta$ would be redundant with $\sigma_\eta$, and so is assumed to be zero (see equations (8) and (10)).

$^f$ Value of “unreliability” for either set of lanes in the SP data is the probability of unexpected delays of 10 minutes or more, as given in the survey question and applying to the entire trip.

$^g$ Scale parameters are defined in equations (11)-(12). A value less than one means there is more unexplained dispersion in this portion of the data than in the Brookings RP data.
### Table 3. Values of Time and Reliability from Joint RP/SP Model

<table>
<thead>
<tr>
<th></th>
<th>Median Estimate(^a)</th>
<th>90% Confidence Interval(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>[5%-ile, 95%-ile]</td>
</tr>
<tr>
<td><strong>Value of time ($/hour)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median in sample</td>
<td>21.46</td>
<td>[11.47, 29.32]</td>
</tr>
<tr>
<td>Observed heterogeneity</td>
<td>4.04</td>
<td>[2.60, 8.34]</td>
</tr>
<tr>
<td>Unobserved heterogeneity</td>
<td>7.12</td>
<td>[3.15, 16.87]</td>
</tr>
<tr>
<td>Total heterogeneity in sample</td>
<td>10.47</td>
<td>[5.82, 24.11]</td>
</tr>
<tr>
<td><strong>Value of reliability ($/hour)</strong>(^c)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median in sample</td>
<td>19.56</td>
<td>[6.26, 42.80]</td>
</tr>
<tr>
<td>Observed heterogeneity(^d)</td>
<td>0</td>
<td>n.a.</td>
</tr>
<tr>
<td>Total heterogeneity in sample</td>
<td>26.49</td>
<td>[8.60, 60.40]</td>
</tr>
<tr>
<td><strong>Value of time ($/hour)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median in sample</td>
<td>11.92</td>
<td>[7.09, 21.06]</td>
</tr>
<tr>
<td>Observed heterogeneity</td>
<td>2.60</td>
<td>[0.24, 8.86]</td>
</tr>
<tr>
<td>Unobserved heterogeneity</td>
<td>12.32</td>
<td>[6.90, 23.30]</td>
</tr>
<tr>
<td>Total heterogeneity in sample</td>
<td>13.31</td>
<td>[7.41, 23.88]</td>
</tr>
<tr>
<td><strong>Value of reliability ($/incident)</strong>(^c)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median in sample</td>
<td>5.40</td>
<td>[3.26, 10.12]</td>
</tr>
<tr>
<td>Observed heterogeneity(^d)</td>
<td>0</td>
<td>n.a.</td>
</tr>
<tr>
<td>Total heterogeneity in sample</td>
<td>7.95</td>
<td>[4.65, 14.38]</td>
</tr>
</tbody>
</table>

N.a.: not applicable

\(^a\) All results shown in the first column are expectations over the sampling distribution of the underlying unknown parameters \((\psi_1, \psi_2)\), obtained by taking Monte Carlo draws \((\psi_1, \psi_2)^r, r=1, ..., R\) from the asymptotic distribution of \((\psi_1, \psi_2)^r\). For each such draw, a random coefficient vector \(\beta^r\) is then drawn from the normal distribution defined by (4)-(5), where \(\Omega\) is part of the underlying parameter vector \((\psi_1, \psi_2)^r\). This is done for each of the 84 individuals \(i\) in the Brookings RP sample. \(VOT_i^r\) is computed from this coefficient vector, and its distribution across the 84 individuals yields the \(r\)-th estimate of the percentile values \(P_{25}^r, P_{50}^r, P_{75}^r\). The same is done for \(VOR_i^r\). Averaging over the \(R\) estimates yields the value shown for “Median in sample,” \(\Sigma P_{50}^r / R\), and “Total heterogeneity in sample,” \(\Sigma_r (P_{75}^r - P_{25}^r) / R\). The same procedure produces the values shown as “Observed heterogeneity” and “Unobserved heterogeneity” except the draw of coefficient vector \(\beta^r\) is from a restricted distribution: a degenerate distribution fixed at \(\bar{\beta} + \gamma Z_i\) in the case of observed heterogeneity, with \(\bar{\beta}\) and \(\gamma\) part of the draw \((\psi_1, \psi_2)^r\); and a normal distribution centered at \(\bar{\beta} + \gamma Z_i\) instead of \(\bar{\beta} + \gamma Z_i\) in the case of unobserved heterogeneity, where \(Z = \Sigma Z_i / 84\).

\(^b\) The confidence interval for each quantity is simply the 5th and 95th percentile points of the distribution of that quantity across the \(R\) Monte Carlo estimates of it.

\(^c\) The measures of unreliability differ between RP and SP; see notes d and f of Table 2.

\(^d\) The observed heterogeneity of VOR, as defined here, is zero despite the dependence of VOR on an observable quantity (income) in our specification. This is because the 25th through 75th percentile values of VOR all come from a single income category (namely, the lowest).
Table 4. Comparison of Alternate Specifications of the Joint RP/SP Model

<table>
<thead>
<tr>
<th></th>
<th>Base Model (from Table 2)</th>
<th>Model with dummy for travel in peak hour(a)</th>
<th>Model with simultaneous choice of occupancy, transponder, and lane(b)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RP Estimates</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Value of time ($/hour)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Value of reliability ($/incident)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median in sample</td>
<td>19.56 [6.26, 42.80]</td>
<td>24.31 [8.54, 42.95]</td>
<td>21.30 [5.27, 41.13]</td>
</tr>
<tr>
<td>Unobserved heterogeneity</td>
<td>26.49 [8.60, 60.40]</td>
<td>29.76 [10.01, 60.00]</td>
<td>30.64 [9.84, 61.05]</td>
</tr>
<tr>
<td><strong>SP Estimates</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Value of time ($/hour)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Value of reliability ($/incident)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median in sample</td>
<td>5.40 [3.26, 10.12]</td>
<td>5.54 [3.04, 13.50]</td>
<td>5.03 [3.50, 11.61]</td>
</tr>
</tbody>
</table>

See notes to Table 2. Estimated on full sample.

\(a\) This model contains a dummy variable equal to one for travel during time 7-8 a.m. and zero for travel at other times. It is also set to zero for the 26 individuals in the SP sample for whom the information is lacking; for those individuals, a separate constant and separate error-component variance are specified. The SP values shown are for the sample of 55 SP individuals for whom we have the time-of-day information.

\(b\) Nine-alternative model, with probabilities joint logit conditional on random parameters. Alternatives include the nine legal combinations of transponder (yes, no), car occupancy (1, 2, 3 or more), and lanes (free, express), given that it is illegal to use the express lanes without a transponder.
### Table 5. Results of Policy Simulations

<table>
<thead>
<tr>
<th>PRICING REGIMEa</th>
<th>Base case (no toll)</th>
<th>Second-best</th>
<th>Second-best: no heterogeneity</th>
<th>First-best</th>
<th>Limited differentiated toll</th>
<th>Limited uniform toll</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toll:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Express lanes</td>
<td>0</td>
<td>$3.56</td>
<td>$2.34</td>
<td>$6.93</td>
<td>$2.02</td>
<td>$2.09</td>
</tr>
<tr>
<td>Regular lanes</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$6.93</td>
<td>$0.94</td>
<td>$2.09</td>
</tr>
<tr>
<td>Travel delay (minutes):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Express lanes</td>
<td>9.37</td>
<td>5.4</td>
<td>6.4</td>
<td>3.7</td>
<td>8.1</td>
<td>8.3</td>
</tr>
<tr>
<td>Regular lanes</td>
<td>9.37</td>
<td>11.3</td>
<td>10.5</td>
<td>3.7</td>
<td>9.4</td>
<td>8.3</td>
</tr>
<tr>
<td>Consumer surplus change per base-case vehicle:b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distribution in population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75%-ile</td>
<td>0</td>
<td>-$0.50</td>
<td>-$0.55</td>
<td>-$2.99</td>
<td>-$0.78</td>
<td>-$1.43</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>-$0.76</td>
<td>-$0.55</td>
<td>-$4.02</td>
<td>-$0.94</td>
<td>-$1.59</td>
</tr>
<tr>
<td>25%-ile</td>
<td>0</td>
<td>-$0.94</td>
<td>-$0.55</td>
<td>-$4.85</td>
<td>-$0.95</td>
<td>-$1.72</td>
</tr>
<tr>
<td>Distribution by income groupc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High income (&gt;=$100K)</td>
<td>0</td>
<td>+$0.17</td>
<td>n.a.</td>
<td>-$0.78</td>
<td>-$0.73</td>
<td>-$0.86</td>
</tr>
<tr>
<td>Medium income ($60-100K)</td>
<td>0</td>
<td>-$0.74</td>
<td>n.a.</td>
<td>-$3.21</td>
<td>-$0.88</td>
<td>-$1.38</td>
</tr>
<tr>
<td>Low income (&lt;$60K)</td>
<td>0</td>
<td>-$0.77</td>
<td>n.a.</td>
<td>-$4.19</td>
<td>-$0.93</td>
<td>-$1.60</td>
</tr>
<tr>
<td>Social welfare change per base-case vehicleb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>$0.35</td>
<td>$0.16</td>
<td>$1.54</td>
<td>$0.46</td>
<td>$0.46</td>
<td></td>
</tr>
<tr>
<td>[90% confidence interval]d</td>
<td>[0.27, 0.65]</td>
<td>[0.09, 0.49]</td>
<td>[0.66, 2.01]</td>
<td>[0.35, 0.81]</td>
<td>[0.39, 0.62]</td>
<td></td>
</tr>
</tbody>
</table>

---

a “Second-best” maximize social welfare subject to regular lanes being free. “First-best” is calculated as a uniform toll because the first-best differentiated tolls were nearly identical. “Limited differentiated toll” maximizes social welfare subject to the lower toll being on the higher-capacity roadway (“regular lanes”) and the 25th percentile consumer surplus change being the same (within $0.01) as with the Second-best toll. “Limited uniform toll” is a uniform toll providing the same total welfare gain as “Limited differentiated toll”. Tolls are calculated within a tolerance of 10 cents.

b Consumer surplus and social welfare are measured relative to the no-toll scenario and are divided by the number of users in the no-toll scenario in order to put them in per-vehicle terms, including all people who would use the corridor under the no-toll policy. Social welfare is the sum of consumer surplus plus revenue.

c For these results, we combined the Brookings and Cal Poly respondents to obtain 101, 186, and 235 people in the high, medium, and low income groups, respectively. High income respondents were a small proportion of the Brookings sample but over-sampled in the Cal Poly data.

d The sampling distribution of social welfare is calculated by first drawing the estimated parameters from their asymptotic joint distribution and then recalibrating $N$ and $\theta_{-1}$ to keep the same travel time and full-price elasticity of demand as under the no-toll scenario. Conditional on this draw and the calibrated parameters, we calculated social welfare for each pricing scheme; tolls defining the scheme were calculated based on mean estimates of parameters. We repeated this procedure 100 times and used the empirical distribution to construct the 90% confidence interval for social welfare.
Appendix A. Stated Preference Survey Questionnaire

Eight hypothetical commuting scenarios were constructed for respondents who travel on SR91. Respondents who indicated that their actual commute was less (more) than 45 minutes were given scenarios that involved trips ranging from 20-40 (50-70) minutes. An illustrative scenario follows:

<table>
<thead>
<tr>
<th>Scenario 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Free Lanes</strong></td>
</tr>
<tr>
<td>Usual Travel Time: 25 minutes</td>
</tr>
<tr>
<td>Toll: None</td>
</tr>
<tr>
<td>Frequency of Unexpected Delays of 10 minutes or more: 1 day in 5</td>
</tr>
<tr>
<td><strong>Your Choice (check one):</strong></td>
</tr>
<tr>
<td>Free Lanes</td>
</tr>
</tbody>
</table>

Appendix B. Construction of RP Variables on Travel Time Savings and Reliability

Travel times on the free lanes (91F) were collected on 11 days: first by the California Department of Transportation on October 28, 1999 (six weeks before the first wave of our survey), and then by us on July 10-14 and Sept. 18-22, 2000 (which are the time periods covered by two later waves of our survey).

Data were collected from 4:00 am to 10:00 am on each day, for a total of 210 observations $y_i$ of the travel-time savings from using the express lanes at times of day denoted by $x_i$, $i=1,...,210$. Our objective is to estimate the mean and quantiles of the distribution (across days) of travel time $y$ conditional on time of day $x$. To do so, we use non-parametric methods of the class of locally weighted regressions: specifically, the form known as local linear fit. For each value of $x$ on a pre-chosen grid, it estimates a linear function $y_i = a + b(x_i-x) + \varepsilon_i$ in the region $[x-h, x+h]$, where $h$ is a bandwidth chosen by the investigator. It does so by minimizing a loss function $g(\cdot)$ of the deviations between observed and predicted $y$.

Denote the $p$-th quantile value of $y$, given $x$, by $q_p(x)$. Following Koenker and Bassett (1978), we estimate it with the local linear quantile regression:

$$\hat{q}_p(x) = \arg\min_a \sum_{i=1}^n g_p[y_i - a - b(x_i-x)] \cdot K[(x_i-x)/h]$$  \hspace{1cm} (B1)

where $n$ is the total number of observations and $g_p[\cdot]$ is the following loss function, which is asymmetric except when $p=0.5$: 38
\[ g_p(t) = \left[ |t| + (2p-1)t \right]/2. \]  \hspace{1cm} (B2)

in which case equation (B1) defines. Yu and Jones (1998) show that the estimated percentile values converge in probability to the actual percentile values as the number of observations \( n \) grows larger, provided the bandwidth \( h \) is allowed to shrink to zero in such a way that \( nh \to \infty \). In the case of the median \((p=0.5)\), this is a least-absolute-deviation loss function, and therefore the estimator can be thought of as a non-parametric least-absolute-deviation estimator.

Similarly, denoting the mean of \( y \) given \( x \) by \( m(x) \), its estimate is given by (B1) but with subscript \( p \) replaced by \( m \) and with loss function \( g_m(t) = t^2 \).

The choice of kernel function has no significant effect on our results. We use the biweight kernel function:

\[
K(u) = \begin{cases} 
\frac{15}{16} (1-u^2)^2, & |u| \leq 1 \\
0, & |u| > 1.
\end{cases} \hspace{1cm} (B3)
\]

The choice of bandwidth, however, is important. We first tried the bandwidth proposed by Silverman (1985):

\[
h = 0.9n^{-0.5} \min\{stdx, (iqdx/1.34)\} \hspace{1cm} (B4)
\]

where \( stdx \) and \( iqdx \) are the standard deviation and interquartile difference of the empirical distribution of \( x \). This bandwidth turns out to be about 0.5 hour for our data. However, there is rather extreme variation in our data at particular times of day, especially around 6:00 a.m., due to accidents that occurred on two days around that time. While these accidents are part of the genuine history and we want to include their effects, they produce an unlikely time pattern for reliability when used with the bandwidth defined by equation (B4) — namely, one with a sharp but narrow peak in the higher percentiles around 5:30 a.m., followed by the expected broader peak centered near 7:30 a.m. We therefore increased the bandwidth to 0.8 hour in order to smooth out this first peak.

The standard deviation shown in figure 2a of the text is the square root of the estimated variance of time saving, obtained by a similar nonparametric regression of the squared residuals \( [y_i - \hat{m}(x_i)]^2 \) on time of day.

The point-wise confidence intervals of the median time savings (shown in figure 1b) and the 80th-50th percentile of time savings (shown in figure 2b), which are used in estimation, are constructed using the paired bootstrap (Hardle (1990), Buchinsky (1998)). We randomly sample pairs \( (y_i, x_i) \) with replacement to form the bootstrap sample with the same size as the original data, and compute the local linear quantile estimator for both the median and the 80th percentile. The procedure is repeated 100 times. The empirical distributions of the median time savings and unreliability (80th - 50th percentile) are used to construct the upper and lower bounds of the 90% confidence intervals for the two estimates.
References


Sullivan, Edward, with Kari Blakely, James Daly, Joseph Gilpin, Kimberley Mastako, Kenneth Small, and Jia Yan (2000), *Continuation Study to Evaluate the Impacts of the SR 91 Value-Priced Express Lanes: Final Report*. Dept. of Civil and Environmental Engineering, California
Polytechnic State University at San Luis Obispo, December


