Please do not quote or circulate.

UNCOVERING THE DISTRIBUTION OF MOTORISTS' PREFERENCES

Kenneth A. Small
University of California at Irvine

Clifford Winston
Brookings Institution

Jia Yan*
University of California at Irvine

April 2002

***The authors are grateful to the Brookings Center on Urban and Metropolitan Policy and the University of California Transportation Center for financial support.**

Introduction

Urban and intercity highways provide the most ubiquitous means of transportation in the United States and serve the greatest number of travelers, yet the variety of motorists' preferences is barely recognized. Public management of roads has mainly offered a uniform class of service financed by the gasoline tax, resulting in growing congestion during ever-expanding peak periods. The standing recommendation of economists—that road pricing could spur motorists to make better use of highway capacity by spreading their travel throughout the day—has gone unheeded by policymakers presumably because the direct welfare impact of such policies on road users is generally negative.[1]

But this perspective is based on a policy where motorists face a single price. Analysts have rarely considered the advantages of differentiated prices that would cater to the differences in travelers' preferences. Indeed, experience with deregulation of transportation, telecommunications, energy, and other industries has taught us that firms have increased capacity utilization, developed niche markets, and benefited consumers by offering a variety of prices and services that respond to consumer desires (Winston (1998)). Could highway pricing that recognized the heterogeneity in motorists' preferences increase efficiency and curtail adverse distributional effects?

Recent pricing experiments in the Los Angeles, San Diego, and Houston areas give motorists the option 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' preferences in automobile-dominated environments where real money is at stake.[2]

---

[1] For homogeneous users, the full price (combined effect of toll and time savings) must be raised in order to reduce travel and thereby reduce congestion.

[2] The term "value pricing" originated as a marketing tool for the first of these experiments. Interestingly the term, which emphasizes precisely the differential-pricing aspect studied here, was found so efficacious that the US Congress substituted it for "congestion pricing" in the 1998 reauthorization of what was then called the "Congestion Pricing Demonstration Program." See *Federal Register*, 63 (192), October 5, 1998, pp. 53487-91.

At the same time, econometric advances are making it possible to identify the varied nature of consumer preferences. These advances include random-parameters models of discrete choice that account for unobserved heterogeneity, error-components models that control for the correlation among repeated choices by a given individual, methodologies that combine the advantages of data concerning consumers' actual and hypothetical choices, and non-parametric techniques that yield plausible characterizations of difficult-to-measure variables such as the reliability of travel time.

This paper measures the preferences of automobile commuters by applying these methodological advances to newly collected data concerning route choices in the Los Angeles-area pricing experiment. Based on their choice of whether to pay a toll to use express lanes, we find that commuters exhibit substantial variation in their valuations of travel time and of travel-time reliability. We then study how the efficiency and distributional effects of road pricing are affected when commuters' heterogeneity is taken into account. Compared with a uniform price, we find that differentiated road prices can significantly reduce the losses in consumer surplus and the distributional disparities between groups of motorists, while still producing sizable efficiency gains. Such prices enhance the political viability of road pricing because only a modest portion of the toll revenues would be necessary to compensate road users.

A Brief Methodological Overview

At first blush, our empirical question—how do motorists value travel time and reliability?—is hardly original or one that would require a sophisticated methodology. A conventional approach would be to estimate a model of a commuter's choice of whether to pay a toll to use a less congested express lane or use a free lane as a function of the toll, the travel times on the lanes, the reliability of travel time on the lanes, and the driver's socioeconomic characteristics.

This model, however, is likely to be flawed unless one accounts for commuters' unobserved heterogeneity, the strong correlation of travel cost and travel time, and the difficulty of obtaining an accurate measure of reliability. Recent advances in econometrics enable us to address these issues.

Unobserved Heterogeneity. Preference heterogeneity may be explained by observable characteristics and unobserved influences. The latter can be captured using models with random coefficients. We will use the mixed logit specification, which extends the random-utility model that underlies multinomial logit (Brownstone and Train (1999), McFadden and Train (2000)).[3] In our analysis, mixed logit divides the stochastic part of utility into three components: one that depends on independent variables that capture unobserved heterogeneity (deviation from mean tastes), another that handles a panel-type error structure arising from repeated route choices by a given commuter, and a third that has the double-exponential distribution standard for logit models. Choice probabilities are estimated using Monte-Carlo simulation to integrate the difficult parts of the error distribution. Conditional on the Monte-Carlo draws, the probabilities take the computationally convenient logit form.

Revealed and Stated Preferences. Most previous research attempting to determine the value of urban travel time has analyzed revealed preference (RP) data based on the choice between travel by car and public transit. This research has found that travelers' value of time varies with trip purpose, income, trip distance, and other observed variables (Small (1992)). Some recent studies have analyzed stated preferences (SP) that are elicited from individuals who are faced with hypothetical commuting situations (Calfee and Winston (1998), Hensher (2001)).

[3] Based on data generated by hypothetical questions, mixed logit has been used to estimate the value of time in analyses of long-distance commuting (Calfee, Winston, and Stempski (2001)), urban trucking (Kawamura (2000)), and residential and workplace location (Rouwendahl and Meijer (2001)).

Both RP and SP data have drawbacks. Use of RP data is often hindered by strong correlations among cost, travel time, and reliability, and by the difficulty of obtaining accurate values of these variables for all the alternatives faced by each individual. SP data cannot overcome the lingering doubt about whether the behavior exhibited in hypothetical situations applies to actual choices. Methodologies have been developed to combine both types of data, thereby taking advantage of the strengths of each (Ben-Akiva and Morikawa (1990), Hensher (1994)). The key insight is that some parameters or parameter combinations are likely to be identical in the choice functions generating RP and SP choices, whereas others are likely to be different. For example, the variance of the error term describing the choice process is likely to differ across data types, as is the ratio of the coefficients of travel time and cost. The latter difference arises because people commonly overstate the time delays they actually incur, and thus respond more to a given actual time saving than to a hypothetical time saving of the same amount.[4] By combining data sets, one can greatly improve the precision in estimating shared coefficients, while controlling for important differences.

Reliability. Travel-time reliability is a potentially critical influence on any mode or route choice, but it can be difficult to measure (Bates, Polak, Jones, and Cook (2001)). Based on data from actual driving conditions, we use non-parametric methods to develop plausible characterizations of reliability to include in our RP estimations. We also specify the reliability of trips in our hypothetical (SP) questions.

Empirical Setting

[4] Sullivan et al. (2000, p. xxiii) provide evidence of this from questions asked of travelers affected by two California road-pricing experiments, including the one used in this study.

The commuter route of interest is California State Route 91 (SR91). It connects rapidly growing residential areas in Riverside and San Bernardino Counties—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 preset schedule. Tolls during the times of our survey ranged from $0.75 to $3.30; carpools of three or more received a 50 percent discount.[5] In contrast to the regular lanes, the express lanes have no entrances or exits between their end points.

Survey Design. A market research firm, Allison-Fisher, Inc., mailed a survey custom-designed to our specifications to SR91 commuters who were members of two nationwide household panels, National Family Opinion and Market Facts. 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 roadway choice, time of commute, trip distance, vehicle occupancy, and whether they had a flexible work-arrival time. They also provided various socioeconomic characteristics.

The same people were also given an SP survey in which they were presented with eight hypothetical commuting scenarios describing the essential characteristics of express and regular lanes. For each scenario, they were given hypothetical tolls, travel times, and probabilities of delay on the two routes, and asked which they would choose. The values presented in the scenarios were

[5] For this reason the express lanes are known as High-Occupancy/Toll (HOT) lanes.

roughly aligned with a respondent's normal commute. An illustrative scenario is shown in Appendix A.[6]

Because the initial screener did not adequately identify travelers who had the option of using the express lanes, we had to survey three waves of potential respondents—in December 1999, July 2000, and September 2000—to assemble an adequate sample. Our final sample consists of 89 respondents providing 385 daily observations on actual behavior (RP) and 82 respondents providing 641 separate observations on hypothetical behavior (SP). Of these respondents, 56 answered both surveys.

Table 1 summarizes the responses. Values are broadly consistent with prior studies and with population summary statistics, indicating that we have a representative sample.[7] The median household income (assigning midpoints to the income intervals) is $46,250. We estimate the average wage rate to be about $23 per hour.[8]

About half of our RP respondents have committed to never choosing the express lanes in the short run by not getting a transponder. Inertia is generally a powerful force in our RP sample

---

[6] Respondents were asked to fill out the SP survey after they completed the RP survey.

[7] The distributions of the RP sample's commuting times and route share are close to the ones in 1998 survey data collected by University of California at Irvine (Lam and Small (2001)) and 1999 survey data collected by California Polytechnic State University at San Luis Obispo (Sullivan et. al. (2001)). Our socioeconomic data appear consistent with Census information, and diverge where appropriate. For example, our median income (approximately $46,250) is higher than the average income in the two counties where our respondents lived ($36,189 in Riverside County and $39,729 in San Bernardino County in 1995, as estimated by the Population Research Unit of the California Department of Finance). But this should be expected because our sample only includes people who are employed and commute to work by car. The median number of people per household (which can be expected to be stable across time) is 2.81 and 3.47 in our RP and SP subsamples respectively; these are not far from the 1990 Census figures of 2.85 for Riverside County and 3.15 for San Bernardino County.

[8] 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 the appropriate 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.

because 85 percent of the respondents make the same choice every day—two-thirds of them never take the express lanes and 19 percent of them always take the express lanes.

Construction of Variables. Obtaining reliable measures of travel conditions facing survey respondents is a challenging part of any travel demand analysis using RP data. Our case is no exception, and is made more difficult by the desire to include travel-time reliability. Our strategy is to use actual field measurements of travel times on SR91 taken at many different times of day over the six-hour morning period covered by our data. Measurements were taken on eleven days, ten of which coincided with the days covered by the second and third waves of our survey; the eleventh day was two months prior to the first wave.

We posit that for any given time of day, observed travel times are random draws from a distribution that travelers know from past experience. By asserting that motorists care about trip time and reliability, we mean that they consider both the central tendency and the dispersion of that distribution.[9] Plausible measures of central tendency include the mean and the median; we find the median fits slightly better (in terms of log-likelihood achieved by the model). Measures of dispersion include the standard deviation and the inter-quartile difference; however, given that motorists—especially commuters—are concerned with occasional significant delays, they are likely to pay particular attention to the upper tail of the distribution of travel times. We therefore investigate the upper percentiles of our travel time distributions.

[9] It seems 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 travel delays along the full 10-mile segment were not well predicted by the amount of congestion encountered prior to the entrance to the express lanes.

We use non-parametric smoothing techniques to estimate the distribution of travel-time savings from taking the express lanes, by time of day.[10] Details are presented in appendix B, and some results are shown in Figures 1 and 2. Figure 1 shows the raw field observations of travel-time savings. The non-parametric estimates of mean, median, and 80th percentile are superimposed. Median time savings reach a peak of 5.6 minutes around 7:15 a.m.

Figure 2 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 that time the scatter becomes highly asymmetric, with dispersion in the positive range (the upper half of the picture) continuing to increase until 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 80th-50th and 90th-50th percentile differences. The standard deviation peaks at roughly 7:45 a.m., the other two between 8:15 and 9:30. 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 standard deviation, by contrast, is higher early in the rush hour because of days with little congestion—showing up as negative points in Figure 2. Such dispersion is probably less relevant to travelers than dispersion in the upper tails, leading us to prefer the percentile differences as reliability measures. These measures are also considerably less correlated with

[10] We observed no congestion on the express lanes at any time. Thus, to simplify the problem, 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.

median travel time than is the standard deviation. In our estimations, we obtained the best statistical fits using the 80^{th}-50^{th} percentile difference.[11]

The express-lane toll for a given trip is from the published toll for the relevant time of day, discounted by 50 percent if the trip were in a carpool of three or more.[12] Other potentially important variables include trip distance, annual per capita household income, and a dummy variable that indicates whether the commuter had a flexible arrival time.[13] We tried a number of others including age, sex, and size of workplace, but they had little explanatory power and did not influence other coefficients.

Most variables in the SP model correspond exactly to variables in the RP model. 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.[14] In addition, SP respondents indicated whether they were answering the questions as solo drivers or as part of a carpool of a specified size, enabling us to determine vehicle occupancy for the SP choices.

Econometric Framework

[11] In our RP and joint RP/SP models, the 90^{th}-50^{th} percentile difference fit almost as well as the 80^{th}-50^{th} difference (in terms of log-likelihood) and resulted in similar coefficient estimates. The 75^{th}-50^{th} percentile difference, an additional measure, and the standard deviation fit noticeably less well and gave statistically insignificant results for the reliability measure.

[12] Specifically, the toll is for the time of day that the commuter reported passing the sign stating the applicable toll. Respondents who did not provide information on vehicle occupancy are assumed not to have carpooled. To guard against systematic bias from this, we specified a dummy variable identifying these respondents, but it had no explanatory power so it is not included in the models reported here.

[13] The question was: "Could you arrive late at work on that day without it having an impact on your job?"

[14] The probability was always stated for the trip as a whole. It was given as 0.05 for all trips using 91X, and either 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.

We assume that a motorist i, facing an actual or hypothetical choice between commuting lanes at time t, chooses the alternative j that maximizes a random utility function of the form:

$$U_{ijt} = \beta x_{ijt} + u_{ijt} \ , \tag{1}$$

where β is a coefficient vector, x_{ijt} is a vector of explanatory variables (encompassing travel conditions, user characteristics, and a constant), and u_{ijt} is a random error term. Assuming that measures of cost c, time t, and unreliability r are included among the travel variables, the values of travel time and (un)reliability are defined as:

$$VoT_i = \frac{\partial U_{ijt} / \partial t_{ijt}}{\partial U_{ijt} / \partial c_{ijt}}; \qquad VoR_i = \frac{\partial U_{ijt} / \partial r_{ijt}}{\partial U_{ijt} / \partial c_{ijt}} \ . \tag{2}$$

In general *VoT* and *VoR* could depend on all three indices, but in our models they depend only on the individual traveler. We will add superscripts *RP* and *SP* to indicate that separate utility functions may govern the responses to RP and SP questions.

Preference Heterogeneity and the Mixed Logit Formulation. Observed heterogeneity in *VoT* and *VoR* is captured by including among the variables some interactions between socioeconomic characteristics (like trip distance) and travel time or unreliability. Unobserved heterogeneity, arising for example because motorists have different personalities, is captured in the mixed logit model by letting the error term have two additive components: one depending on the variables x and the other purely random:

$$u_{ijt} = \eta_i x_{ijt} + e_{ijt} \ .$$

Thus the derivatives in (2) contain one or more components of η_i, making *VoT* and *VoR* stochastic.

Let y_{it} be the binary choice variable, defined as y_{it}=1 if the express lanes are chosen. The criterion for making this choice is then U_{i1t}-U_{i0t}>0, or:

$$Y_{it} \equiv \beta X_{it} + \eta_i X_{it} + \varepsilon_{it} > 0 \tag{3}$$

where $X_{it} \equiv x_{i1t} - x_{i0t}$ and $\varepsilon_{it} \equiv e_{i1t} - e_{i0t}$. This formulation clarifies that whenever a component of η has a non-zero variance, the corresponding variable X can be viewed as having a stochastic coefficient $\beta + \eta_i$. The values of e_{ijt} are assumed to be independently and identically distributed as extreme value for all i, j, t. Therefore ε_{it} has a logistic distribution, which yields the familiar logit formula for the choice probabilities conditional on the value of η_i.

The panel-type data that arises from individuals' repeated observations is handled by this formulation if a constant term is included among the variables having stochastic coefficients. Let the constant be denoted by $X_{1,it} \equiv 1$ and its coefficient by $\beta_1 + \eta_{1,i}$. Then the error term in (3) includes a term $\eta_{1,i}$ that is constant for each individual but varies randomly across individuals, exactly like a random-effects panel model. The standard deviation $\sigma_{\eta 1}$ of $\eta_{1,i}$ helps determine how closely the combined error terms in (3) are correlated across observations from the same individual. Let $v_{it} = \eta_i X_{it} + \varepsilon_{it}$ be that combined error term, we have for each i,j,t,s:

$$E(\nu_{it}) = 0 \tag{4a}$$

$$E(\nu_{it}\nu_{it}) = X_{it}^{'}\Omega X_{it} + \sigma_{\varepsilon}^2 \tag{4b}$$

$$E(\nu_{it}\nu_{is}) = X_{it}^{'}\Omega X_{is} \qquad \text{if } t \neq s \tag{4c}$$

$$E(\nu_{it}\nu_{js}) = 0 \qquad \text{if } i \neq j \tag{4d}$$

where Ω is the variance-covariance matrix of η_i. Hence the correlation between v_{it} and v_{is} for $t \neq s$ is the ratio of the right-hand sides of (4c) and (4b). As in a binary logit model, we can normalize σ_{ε}^2 to $\pi^2/3$ (which follows from e_{ijt} being independent extreme-value variates). To conserve on unknown parameters, we specify Ω to be diagonal.

The probability of motorist i choosing the express lanes at situation t, conditional on η_i, is

$$P(y_{it}=1 \mid \beta,\eta_i)=\frac{1}{1+\exp(-\beta X_{it}-\eta_i X_{it})} \ . \tag{5}$$

The unconditional joint probability of motorist i's choice sequence (y_{it}) over several choice situations t is (using a simplified notation):

$$P_i(\beta,\Omega)=\int_{\eta_i}\left\{\prod_t P(y_{it} \mid \beta, X_{it},\eta_i)\right\} f(\eta_i \mid \Omega) d\eta_i \ , \tag{6}$$

where $f(\bullet)$ represents the joint density function of the components of random vector η_i. The integration in (6) is performed using Monte-Carlo simulation methods. Given a trial value for Ω, we draw η_i^r from the assumed distribution $f(\bullet)$ and evaluate the probabilities conditional on η_i^r, repeating for r=1,...,R. The simulated estimate of P_i in (6) is therefore:

$$P_i^S(\beta,\Omega)=\frac{1}{R}\sum_{r=1}^{R}\left\{\prod_t P(y_{it} \mid \beta,\eta_i^r)\right\} \ .$$

We estimate β,Ω by maximizing the simulated log-likelihood function:

$$L^S(\beta,\Omega)=\sum_i \ln P_i^S(\beta,\Omega) \ .$$

Lee (1992) and Hajivassiliou and Ruud (1994) show that under regularity conditions, the parameter estimates are consistent and asymptotically normal and, when the number of replications rises faster than the square root of the number of observations, asymptotically equivalent to maximum likelihood estimates.

Combining RP and SP Observations. When RP and SP data are combined, we must allow for the possibility that the RP and SP utility functions have error terms with different variances, and that the RP and SP error terms corresponding to the same individual are correlated.

Following standard practice, we accommodate the first possibility by normalizing σ_ε^2 to $\pi^2/3$ for RP only, while estimating a new parameter $\mu \equiv \sigma_\varepsilon^{RP} / \sigma_\varepsilon^{SP}$. If μ<1 (as it turns out in our model), then there is more randomness in SP than in RP responses, not counting random parameters.

We account for the correlation between RP and SP responses by following Morikawa (1994) and force the random constants $\eta_{1,i}^{RP}$ and $\eta_{1,i}^{SP}$ for a given individual *i* to be identical up to a constant of proportionality, θ. This is equivalent to splitting the error terms ν^{RP} and ν^{SP} in (4) into three components:

$$\nu_{it}^{RP} = \eta_{-1,i}^{RP} X_{-1,i}^{RP} + \lambda_i + \varepsilon_{it}^{RP} \qquad (7)$$

$$\nu_{it}^{SP} = \eta_{-1,i}^{SP} X_{-1,i}^{SP} + \theta\lambda_i + \varepsilon_{it}^{SP} \qquad (8)$$

where subscript -1 denotes a vector after removing its first component, and where $\lambda_i = \eta_{1,i}^{RP}$. The components λ_i in (7) and $\theta\lambda_i$ in (8) represent individual effects and therefore account for correlation among the responses for a given individual, including (to the extent that θ is positive) correlation between an individual's RP and SP responses.

To see more clearly what is identified, it is helpful to rewrite (3), the criterion for choosing the express lanes, separately for RP and SP and in a form in which the last error term has the same variance, $\pi^2/3$, in both equations:

$$Y_{it}^{RP} \equiv (\beta_{-1}^{RP} + \eta_{-1,i}^{RP}) X_{-1,it}^{RP} + (\beta_1^{RP} + \lambda_i) + \varepsilon_{it}^{RP} > 0 \qquad (9)$$

$$\mu Y_{it}^{SP} \equiv \mu(\beta_{-1}^{SP} + \eta_{-1,i}^{SP}) X_{-1,it}^{SP} + \mu(\beta_1^{SP} + \theta\lambda_i) + \mu\varepsilon_{it}^{SP} > 0 \,. \qquad (10)$$

It is worth stressing that the joint model in equations (9) and (10) can improve statistical efficiency by using both RP and SP observations to estimate some coefficients that are assumed identical (up to scale factor μ) in the two choice processes. This enables us to estimate some key heterogeneity parameters whose effects would otherwise remain indeterminate in our RP data because of the small sample size and multicollinearity. At the same time, we can protect the key RP coefficient estimates against potential SP survey bias by allowing those coefficients most in danger of such contamination to take different values for the two data sets.

Estimation Results

Our primary objective is to estimate distributions of the value of time and unreliability based on parameters obtained from a combined RP and SP model. The final specification of this model is sufficiently complex that our findings will be easier to understand if we proceed in steps by first presenting estimates of the separate RP and SP models that form the joint model.

Revealed Preference Estimates. There are at least two ways to analyze our RP data. One is to estimate a binary logit model of lane choice on the entire sample of observations, including those from the same motorist on different days. We call this a *trip-based* model. The other is to convert each motorist's multiple observations into one and estimate a model whose dependent variable is the frequency of using toll lanes. We call this a *person-based* model.

The latter has certain advantages. First, it correctly assumes that a traveler's decision to get a transponder (a prerequisite for choosing the express lanes) is based on long-run tradeoffs, not a daily choice. Second, its simpler error structure makes it easier to combine with an SP model. Our data revealed that few travelers change behavior from one day to the next, so little information is gained from the extra observations contributing to trip-based models. Indeed, results from the two models were quite similar; thus, we 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 outcome defined as 1 if the motorist used the express lanes for half or more of reported commuting trips, 0 otherwise.[15] Independent variables are defined as the average value over the days reported.[16]

We were unable to satisfactorily identify the unobserved heterogeneity in an RP-only model because of the small sample size, but we did find substantial *observed* heterogeneity. Distance has a strong nonlinear effect on the time coefficient, captured well by a cubic form with no intercept (i.e., median travel time is not entered by itself). We also follow convention by interacting travel cost with income.

Most of the parameter estimates are statistically significant and have the expected signs (table 2). Commuters are deterred from the express lanes by a higher toll and from the free lanes by longer median travel times and greater unreliability. Commuters with flexible arrival times are more likely to use the express lanes; we speculate that this variable serves as a proxy for unmeasured job characteristics requiring punctuality on a given day even though a commuter is not normally constrained to arrive at a particular time.[17] We also investigated the influence of a motorist's sex on lane choice, but were unable to find a clear-cut effect.[18]

---

[15] We omitted the few respondents who have a transponder but traveled two days or less, because defining a frequency for them involves too much error. Our specification 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 *j* intervals. We explored several ways of dividing the [0,1] interval into two, three, or four regions. 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 main parameters of interest.

[16] An exception is the "flexible arrival time" dummy, which is set to one if the respondent indicated flexible arrival for half or more of the reported days. (In fact, only five people reported any daily variation in this variable.)

[17] Although theory suggests that this variable might affect value of reliability, experimentation revealed it does not fit as well when interacted with the time or reliability measure as it does in the form shown.

[18] This contrasts with Parkany (1999), Lam and Small (2001), and Yan, Small, and Sullivan (2001) who find that women are more likely to choose the SR91 express lanes, other things equal.

As noted, observed heterogeneity is captured by the interactions between toll and income and between median travel time and distance.[19] Consistent with expectations, motorists with higher incomes are less responsive to the toll. The effect of distance on the value of time is characterized by an inverted U, initially rising but then falling for trips greater than 28 miles (which applies to most of our sample). We conjecture that this pattern results from two opposing forces: the increasing scarcity of leisure time as commuting takes up a greater fraction of it, and the self-selection of people with low values of time to live in lower-cost housing further from their workplace (Calfee and Winston (1998)).

The implied VoT and VoR are summarized at the bottom of the table. The 90 percent confidence intervals, representing statistical uncertainty not heterogeneity, are computed by Monte-Carlo simulation based on the asymptotic joint distribution of the estimated parameters. There is substantial observed heterogeneity captured by the interactions of distance with median travel time and income with cost.[20] We will assess the magnitudes of VoT and VoR when we present the results of the combined RP/SP model.

Stated Preference Estimates. We can successfully estimate both observed and unobserved heterogeneity using SP data because we have an adequate sample size and the independent variables are, by design, not highly correlated. The dependent variable is the respondent's choice of whether to use the express lanes in a given scenario. Mixed-logit

[19] The estimated coefficients of the interaction terms are such that the net marginal effects of the toll and median travel time remain negative throughout the full range of our data.

[20] For example, the values of time for two given commuters, each having the median household income and average family size but with one-way commute distances of 30 and 60 miles are \$16.78 and \$7.67 per hour, respectively. Doubling the first commuter's income raises its VoT to \$18.86.

estimations were performed using 1,000 random draws for the simulations, assuming that the random parameters for time and unreliability have independent normal distributions.[21]

The results presented in table 3 indicate that key parameters are precisely estimated and have the expected signs. As before, respondents trade off tolls with travel time and unreliability, and are more likely to use a toll lane if they have a flexible arrival time. Consistent with the predominant effect of distance in the RP model, those with actual commutes over 45 minutes find travel time less onerous than those with shorter commutes.[22] The parameters indicating the standard deviations of random coefficients are comparable in magnitude to the corresponding means, implying considerable unobserved heterogeneity. Surprisingly, income is statistically insignificant, whether entered as a lane-choice shift variable or (as tried but not shown) interacted with the toll.[23]

Joint RP/SP Estimates. Separate RP and SP models suggest that motorists' values of time and unreliability vary in accordance with observed and unobserved influences. By combining them, we can obtain a more precise understanding of preference heterogeneity. Because we use a person-based RP model, there is only one value of t in equations (7) and (9) and so we cannot effectively identify the variance of λ_i. We therefore normalize it to one. We also economize on the number of parameters that describe heterogeneity by constraining the RP and SP cost

---

[21] This assumption leads to the possibility of a traveler having the "wrong" sign for these two coefficients. We tried log-normal and truncated normal distributions for the random coefficients, but were unable to get the results to converge—a problem noted by other researchers using these distributions such as Train (2001), although Bhat (2000) and Calfee, Winston, and Stempski (2001) were successful with the log-normal. We also tried treating the toll coefficient as random, but this created problems in calculating the ratios in (2) because the net marginal effect of the toll, which appears in the denominator of these ratios, is then zero or wrong-signed for some people.

[22] As discussed in Appendix A, these two groups received different versions of the survey, each designed to make the stated travel times approximately aligned with their actual commutes.

[23] We also estimated models with "inertia effects," in which the SP choice is conditioned on the actual choice of whether to obtain a transponder. Although this improves the goodness of fit considerably, it may introduce bias because the actual choice is not fully exogenous to the SP choice given the likely correlation of their error terms. Morikawa (1994, pp. 158-159) discusses this type of model.

coefficients to be equal (except for the scale factor μ) and by assuming that all the RP and SP time coefficients share a single additive random component δ_i.[24] Thus the marginal effect of travel time on utility is:

$$\frac{\partial U_{ijt}^{RP}}{\partial t_i} = \beta_2^{RP} dist + \beta_3^{RP} dist^2 + \beta_4^{RP} dist^3 + \delta_i \tag{11}$$

$$\frac{\partial U_{ijt}^{SP}}{\partial t_i} = \beta_2^{SP} Long + \beta_3^{SP}(1 - Long) + \delta_i \tag{12}$$

where *dist* refers to trip distance in the RP model, *Long* refers to the commute distance dummy in the SP model, the βs refer to the corresponding coefficients, and δ_i is assumed normal with mean zero and variance σ_δ^2 to be estimated.[25] This specification allows the RP and SP values of time to differ, but combines the power of the two data sets to estimate their random variation.

Because the measures of unreliability have different meanings in the RP and SP data, we cannot assume that their coefficients have the same random component. Instead we assume that the ratio of the standard deviation to the mean of the unreliability coefficient is the same across samples. Thus, using subscript *r* to denote the unreliability coefficient, we reparameterize the specification as follows:

$$\beta_r^{\,k} + \eta_{r,i}^k = \beta_r^k\left(1 + \frac{\eta_{r,i}^k}{\beta_r^k}\right) \equiv \beta_r^k(1+\omega_i), \qquad k{=}RP,\ SP \tag{13}$$

where ω_i is assumed normal with variance σ_ω^2 to be estimated.

---

[24] Using likelihood-ratio tests, we tested and did not reject the constraint that the cost coefficients are equal. The SP and RP cost variables in the joint model are defined the same way, namely commuters pay the full toll if occupancy is less than three and half the toll otherwise.

[25] Formally the assumption behind (11)-(12) is that a variable equal to median travel time, call it X_5, is included in the RP and SP models described by (3) with constraints $\beta_5^{RP} = \beta_5^{SP} = 0$ and $\eta_{5,i}^{SP} = \eta_{5,i}^{RP} \equiv \delta_i$; while the η's corresponding to parameters β_2, β_3, and β_4 are all zero.

We estimate the model by simulated maximum likelihood, using 2000 random draws.[26] Parameter estimates are presented in table 4. Most parameters are similar in magnitude to their values in the separate estimations, once error variances are equalized by multiplying SP coefficients by scale factor μ as in (10). But there is a payoff from joint estimation because some coefficients relevant to the RP choice are estimated with greater precision than before, notably those associated with cost and unreliability. The two parameters capturing unobserved heterogeneity in the coefficients of time and unreliability are also precisely estimated, as are the scale and correlation parameters describing the error structure. The scale parameter μ suggests that there is substantially more noise in the SP responses than in actual behavior (RP), while the parameter θ indicates that SP and RP responses from a single respondent are strongly correlated.

Motorists' Preferences and Heterogeneity. We use the estimated coefficients of the joint RP/SP model to calculate motorists' implied values of time and unreliability and indicate the extent of their heterogeneity (table 5). All quantities shown are significantly different from zero at a 5% confidence level (one-sided test), as can be seen from the last column. The median value of time based on commuters' revealed preferences is \$15.64/hour; at 68 percent of the average wage it is within the range expected from previous work (Small (1992)). Median time savings in our data peak at 5.6 minutes; thus, the average commuter would pay \$1.46 to realize these savings. Motorists place an even higher value, \$23.26/hour, on unreliability.[27] Unreliability peaks at 3 minutes; thus, the average commuter would pay \$1.16 to avoid this possibility of

[26] We found that the calculated standard errors were sensitive to the number of random draws up to 1500, but did not change when they were increased to 2000.

[27] Similar estimates of the value of time and unreliability were obtained by estimating a trip-based model that treats each respondent's lane choice as a separate observation and accounts for multiple observations of a single respondent by adopting a robust estimator.

unanticipated delay. Given these estimates, the actual peak toll of $3.30 would be expected to attract substantially less than half of the total peak traffic—which, in fact, it does.

We are also interested in how much motorists' preferences vary. We use the interquartile difference (the difference between 75th and 25th percentile values) as our heterogeneity measure because it is unaffected by high upper-tail values occasionally found in the calculations of observed heterogeneity. This measure exceeds the median values of time and reliability, indicating that commuters exhibit a wide distribution of preferences for speedy and reliable travel. It is interesting that most of the heterogeneity is from unobserved sources, two-thirds in the case of VoT and more than 90 percent in the case of VoR, verifying the importance of "taste variation" in motorists' behavior and our attempt to capture it.

The implied values of time and unreliability from stated preferences are lower on average than those based on revealed preferences. This finding may reflect the aforementioned tendency of travelers to overstate the travel time they lose or would lose in congestion. For example, if motorists are in the habit of paying $1.56 to save 10 minutes but they perceive that saving as 15 minutes, then they may answer SP questions as if they would pay $1.56 to save 15 minutes—yielding a downward-biased value of actual travel time. The SP value of unreliability may be similarly biased, but we have no point of comparison. The median value of $5.15 per incident means that the median motorist in our sample would pay $0.51 per trip to reduce the frequency of 10-minute delays from 0.2 to 0.1.

Implications for Road Pricing Policy

Does preference heterogeneity create a strong case for differentiated services? We use Small and Yan's (2001) simulation model to investigate the potential effects of differential road-pricing.

The model resembles the SR91 road-pricing experiment, in which two 10-mile roadways, Express and Regular, connect the same origin and destination and have the same free-flow travel-time. Users are of two types: high value of time (i=1) and low value of time (i=2). Each user chooses the best option while congestion on each road adjusts endogenously according to a standard fourth-power relationship with the volume-capacity ratio. Each user group has the same size when roads are free, but each also has a demand elasticity with respect to the full price. Because SR91 has twice as many regular lanes as express lanes, in equilibrium the express roadway contains only high-VoT users while the regular roadway contains both types of users. Equilibrium is therefore reached when the full cost to high-VoT users is the same on both roadways.

In Small and Yan's model, the value of time can also include a value of unreliability, with unreliability perfectly correlated with time. To apply the results obtained here, we specify the "full price" p_{ir} for a user of type i on roadway r to be $p_{ir} = \tau_r + \varphi_i T_r + \delta_i R_r$, where T is travel time and R is (un)reliability. We assume $R_r / T_r = \varsigma$ =0.3785, the ratio of our measured average values of unreliability and travel time over the 4-hour peak period (5am – 9am). Thus $p_{ir} = \tau_r + \alpha_i T_r$, where $\alpha_i = \varphi_i + \varsigma\delta_i$. For φ_i and δ_i we use the value of time and unreliability estimates in Table 5 based on RP behavior, taking the two user groups to be represented by the 75th and 25th percentiles, respectively.[28] This yields values of $\alpha_1 = \$40.28 / hr$, and $\alpha_2 = \$8.60 / hr$.

Letting p_i be the lower of the two prices facing user type i, the demand for trips by these users can be written as $N_i(p_i)$. We assume this function is linear, with parameters calibrated to reproduce real traffic conditions observed on SR91 in summer 1999. Thus each group's money-price elasticity is –0.58, as estimated by Yan, Small, and Sullivan (2001), and the time difference

[28] The third and sixth rows of the table show the difference between 75th and 25th percentiles. The percentiles themselves are: \$25.24 and \$6.04 for VoT, and \$39.74 and \$6.79 for VoR.

between the lanes is 6 minutes when the price on the express lanes maximizes profit subject to the regular lanes being free. This calculation yields a plausible express-lane toll of $4.02.

Based on these parameters, we calculate tolls, travel times, changes in consumer surplus, and social welfare under several alternative pricing policies. Our base case scenario has no toll on either roadway. The improvement in social welfare is the change from the base scenario in toll revenues and both groups' consumer surplus. Results are shown in table 6.

The first two policies set a price on the express lanes that maximizes social welfare subject to the current constraint that the regular lanes have zero toll. This "second-best" policy sacrifices efficiency by not pricing all lanes. Nonetheless, when there is user heterogeneity, welfare improves by a nontrivial $0.25 per vehicle, whereas if users are all alike there is virtually no welfare gain. Thus, in this instance the presence of heterogeneity increases the potential efficiency of road pricing.

Can accounting for heterogeneity also improve road pricing's distributional effects? A first-best differentiated toll, shown in the fourth column of the table, achieves a substantial welfare gain ($0.82 per vehicle). However, it imposes high direct costs on motorists, especially the low-VOT users. Because these costs are undoubtedly a major political barrier to implementing road pricing, we consider a "limited differentiated toll" that maximizes welfare subject to a limit of $0.67 per vehicle on the regular lanes. The result is a sharply differentiated toll that substantially reduces both groups' losses in consumer surplus and narrows the gap between them, while achieving a welfare gain of more than half that of first-best pricing. This desirable outcome depends on catering to heterogeneity as indicated by a "limited uniform toll" that generates the same efficiency gains but without differentiating the toll rate. As shown in the last column of the table, this policy harms the low-VOT group far more than the high-VOT group.

By reducing the adverse direct impact of combined tolls and time savings on consumer surplus, differentiated pricing enhances the political viability of road pricing because policymakers must apportion only a modest fraction of the toll revenues to fully compensate road users.

Conclusion

Road pricing has been beloved by economists and opaque to policymakers for decades. Calfee, Winston, and Stempski (2001) rationalized this state of affairs by arguing that few long-distance automobile commuters were willing to pay much to save travel time because high value-of-time commuters spend more on housing to live close to their workplace.

We have applied recent econometric advances to analyze the behavior of commuters in Southern California and found that those with longer commutes have lower values of time, which is consistent with residential selectivity. But we have also found great heterogeneity in motorists' preferences for speed and reliability suggesting that in expensive and congested metropolitan areas, such as Southern California, the San Francisco Bay Area, New York, and Washington, DC, consumers face significant constraints in trading off housing costs for commuting time. In these parts of the United States, there is an opportunity to design pricing policies with a greater chance of public acceptance by catering to varying preferences.

Recent "value pricing" experiments have made a start by offering motorists the opportunity to pay for congestion-free travel. But by leaving part of the roadway unpriced, efficiency is severely compromised. We have demonstrated that pricing policies taking preference heterogeneity explicitly into account can realize substantial efficiency gains and ameliorate distributional concerns. Differential pricing, embedded in both the design and marketing of these recent experiments, may thus be the key to addressing the stalemates that impede transportation policy in congested cities.

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:

Scenario 1

Free Lanes	**Express Lanes**
Usual Travel Time: 25 minutes	Usual Travel Time: 15 minutes
Toll: None	Toll: $3.75
Frequency of Unexpected Delays of 10 minutes or more: 1 day in 5	Frequency of Unexpected Delays of 10 minutes or more: 1 day in 20
Your Choice (check one):	
Free Lanes ☐	Toll Lanes ☐

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. In these methods, the range of the independent variables is divided arbitrarily into a grid, and a separate regression is estimated at each point of the grid. In our case, there is just one variable, x. For given x, the regression makes use only of observations with x_i near x, the importance of each being weighted in a manner that declines with $|x_i\text{-}x|$. The weights are based on a kernel function $K(\bullet)$, and how rapidly they decline is controlled by a *bandwidth* parameter h; typically only observations within one bandwidth of x get any positive weight.

The specific form of locally weighted regression we use is known as *local linear fit*. For each value of x, it estimates a linear function $y_i = a + b(x_i\text{-}x) + \varepsilon_i$ in the region $[x\text{-}h, x\text{+}h]$ by minimizing a loss function of the deviations between observed and predicted y. Denote the p-th quantile value of y, given x, by $q_p(x)$. Its estimator is then:

$$\hat{q}_p(x) = arg\ min_a \sum_{i=1}^{n} g_p\left[y_i - a - b(x_i - x)\right] \bullet K\left[(x_i - x)/h\right] \qquad \text{(A1)}$$

where $g_p(t)$ is the loss function. Similarly, denoting the mean of y given x by $m(x)$, its estimate is given by the same formula but with subscript p replaced by m.

In the case of mean travel-time savings, we use a simple squared-error loss function, $g_m(t) = t^2$, in which case equation (A1) becomes the *local linear least square regression*. In the case of percentiles of travel-time savings, including the median, we follow Koenker and Bassett's (1978) suggestion and use the following loss function, which is asymmetric except for the median (p=0.5):

$$g_p(t) = \{|t| + (2p-1)t\}/2 \qquad \text{(A2)}$$

With this loss function, equation (A1) defines the *local linear quantile regression* (Yu and Jones, 1997). It can be shown 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.

The choice of kernel function has no significant effect on our results. We use the *biweight* kernel function, which has the following form:

$$K(u) = \begin{cases} \frac{15}{16}\left(1-u^2\right)^2 & |u| \le 1 \\ 0 & |u| > 1. \end{cases} \qquad \text{(A3)}$$

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

$$h = 0.9n^{-0.5} \min\left\{\mathrm{std}(\mathrm{x}), \frac{d}{1.34}\right\} \qquad \text{(A4)}$$

where n is the size of the data set, "std" means standard deviation, and d is the difference between the 75th and 25th percentile 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 (A4) -- namely, one with a sharp but narrow peak in the higher percentiles around 5:30 a.m., followed by the expected broader peak centered around 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 2 of the text is the square root of the estimated variance of time saving, obtained by a similar nonparametric regression of the squared residuals $\left(y_i - \hat{m}(x)\right)^2$ on time of day.

References

Bates, J., J. Polak, P. Jones, and A. Cook (2001), “The Valuation of Reliability for Personal Travel,” *Transportation Research E: Logistics and Transportation Review*, 37, pp. 191-229.

Ben-Akiva, M., and T. Morikawa (1990), “Estimation of Travel Demand Models from Multiple Data Sources,” in Koshi, M., ed., *Transportation and Traffic Theory*. New York: Elsevier, pp. 461-478.

Bhat, Chandra R. (2000), “Incorporating Observed and Unobserved Heterogeneity in Urban Work Travel Mode Choice Modeling,” *Transportation Science*, 34, pp. 228-238.

Brownstone, D., and K. Train (1999), “Forecasting New Product Penetration with Flexible Substitution Patterns,” *Journal of Econometrics*, 89, pp. 109-129.

Calfee, J., and C. Winston (1998), “The Value of Automobile Travel Time: Implications for Congestion Policy,” *Journal of Public Economics*, 69, pp. 83-102.

Calfee, J., C.Winston, and R. Stempski (2001), “Econometric Issues in Estimating Consumer Preferences from Stated Preference Data: A Case Study of the Value of Automobile Travel Time,” *Review of Economics and Statistics*, 83, pp. 699-707.

Hajivassiliou, V., and P. Ruud (1994), “Classical Estimation Methods for LDV Models Using Simulation,” *Handbook in Econometrics*, IV, R. Engle and D. MacFadden eds., Elsevire Science B. V., New York.

Hensher, D.A. (1994), “Stated Preference Analysis of Travel Choices: The State of Practice,” *Transportation*, 21, pp. 107-133.

Hensher, D.A. (2001), “The Sensitivity of the Valuation of Travel Time Savings to the Specification of Unobserved Effects,” *Transportation Research E: Logistics and Transportation Review*, 37, pp. 129-142.

Kawamura, Kazuya (2000), “Perceived Value of Time for Truck Operators,” *Transportation Research Record* 1725, pp. 31-36.

Koenker, R., and G. S. Basset (1978), “Regression Quantiles,” *Econometrica*, 46, pp. 33-50.

Lam, Terence C., and Kenneth A. Small (2001), “The Value of Time and Reliability: Measurement from a Value Pricing Experiment,” *Transportation Research Part E: Logistics and Transportation Review*, 37, pp. 231-251.

Lee, L. (1992), “On Efficiency of Methods of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Response Models,” *Econometric Theory*, 8, pp. 518-552.

McFadden, D., and K. Train (2000), "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15, pp. 447-470.

Morikawa, T. (1994), "Correcting State Dependence and Serial Correlation in the RP/SP Combined Estimation Method," *Transportation*, 21, pp. 153-165.

Parkany, E. A. (1999), *Traveler Responses to New Choices: Toll vs. Free Alternatives in a Congested Corridor*, Ph.D. Dissertation, University of California at Irvine.

Rouwendahl, J., and E. Meijer (2001), "Preferences for Housing, Job, and Commuting: A Mixed Logit Analysis," *Journal of Regional Science*, 41, pp. 475-505.

Silverman, B. W. (1985), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Small, K. A. (1992), *Urban Transportation Economics*, Vol. 51 of *Fundamentals of Pure and Applied Economics Series*, Harwood Academic Publishers.

Small, K.A. and J. Yan (2001), "The Value of "Value Pricing" of Roads: Second-Best Pricing and Product Differentiation," *Journal of Urban Economics*, 49, pp. 310-336.

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 (http://ceenve.ceng.calpoly.edu/sullivan/SR91/).

Train, Kenneth (2001), "A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit," working paper, Dept. of Economics, Univ. of California at Berkeley, May (http://elsa.berkeley.edu/~train/papers.html).

Vuong, Q. (1989), "Likelihood Ratio Tests for Models Selection and Non-nested Hypotheses," *Econometrica*, 57, pp. 307-333.

Winston, C. (1998), "U.S. Industry Adjustment to Economic Deregulation," *Journal of Economic Perspectives*, 12, pp. 89-110.

Yan, J., K. Small, and E. Sullivan (2001), "Choice Models of Route, Occupancy, and Time-of-Day with Value Priced Tolls," *Transportation Research Record*, forthcoming.

Yu, K., and M. C. Jones (1997), "A Comparison of Local Constant and Local Linear Regression Quantile Estimation," *Computational Statistics and Data Analysis*, 25, pp. 159-166.

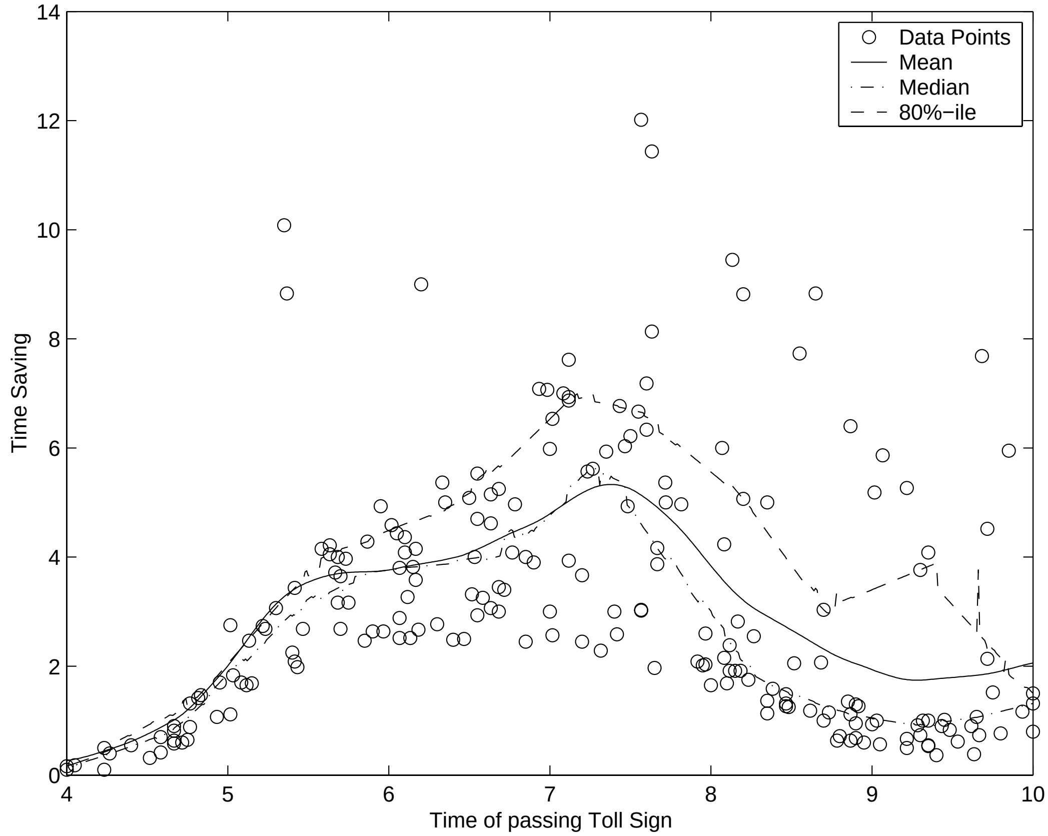

Figure 1. Time Saving

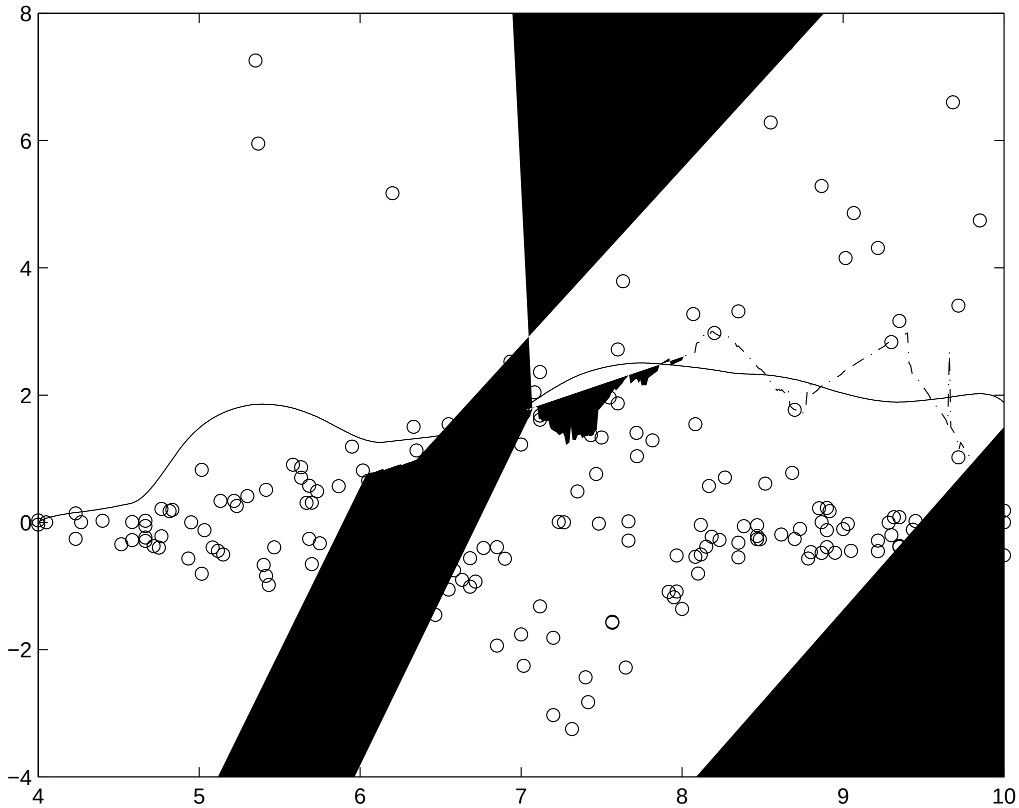
8
6
4
2
0
−2
−4
4
5
6
7
8
9
10

Table 1. Descriptive Statistics

	Value or Fraction of Sample	
	RP	SP
Route Share:		
91X	0.25	
91F	0.75	
One-Week Trip Pattern:		
Never Use 91X	0.66	
Sometimes Use 91X	0.15	
Always Use 91X	0.19	
Trip Distribution Across Time-of-Day:		
4:00am-5:00am	0.15	
5:00am-6:00am	0.12	
6:00am-7:00am	0.26	
7:00am-8:00am	0.20	
8:00am-9:00am	0.15	
9:00am-10:00am	0.12	
Age of Respondent:		
<35	0.25	0.26
35-60	0.69	0.68
>60	0.06	0.06
Gender of Respondent:		
Male	0.63	0.63
Female	0.37	0.37
Household Income:		
<20,000	0.13	0.10
20,000-40,000	0.12	0.14
40,000-70,000	0.71	0.70
>70,000	0.04	0.06
Flexible Arrival Time:		
Yes	0.55	0.50
No	0.45	0.50
Trip Distance (Miles):		
Mean	44.48	42.56
Standard Deviation	26.85	26.85
Number of People in Household:		
Mean	2.81	3.47
Standard Deviation	1.57	1.54
Number of Respondents	89	82
Number of Observations	385	641

Table 2. Binary Logit RP Parameter Estimates

Dependent Variable: 1 if respondent chose express lanes for >50% of reported work trips; 0 otherwise

Independent Variable	**Coefficient** (standard error)
Constant	-0.7094 (1.5780)
Posted toll ($) for a solo vehicle, divided by 2 if car occupancy is 3 or more	-2.5269 (1.0477)
Posted toll, as defined above, multiplied by per capita household income ($000)	0.0152 (0.0068)
Median travel time (minutes) multiplied by trip distance (in units of 10 miles)[a]	-0.4391 (0.1858)
Median travel time multiplied by trip distance squared[a]	0.0861 (0.0334)
Median travel time multiplied by trip distance cubed[a]	-0.0035 (0.0015)
Unreliability of travel time; given as the difference between the 80th percentile and median times[a]	-0.9368 (0.4745)
Flexible arrival time dummy (1 if traveler has a flexible arrival time; 0 otherwise)	1.200 (0.6731)
Summary Statistics	
Number of Observations	84
Log-likelihood	-37.72
Pseudo R^2	0.1819
Implied Tradeoffs	
Mean VoT in sample ($/hour)	17.38
[90% confidence interval]	[1.48, 34.25]
Mean VoR in sample ($/hour)	25.29
[90% confidence interval]	[5.64, 58.06]

[a] Travel time and unreliability are each calculated as the value in the express lanes minus that in the untolled lanes.

Table 3. Mixed Logit SP Parameter Estimates

Dependent Variable: 1 if respondent chose toll lanes; 0 otherwise

Independent Variable	**Coefficient** (standard error[a])
Constant (β_1)	-5.8872 (1.8155)
Standard deviation of constant ($\sigma_{\eta 1}$)	4.2647 (0.8695)
Listed toll in survey question ($), divided by 2 if car occupancy is 3 or more	-1.3142 (0.2964)
Travel time in survey question (minutes) x commute distance dummy (1 if respondent's actual commute time is greater than 45 minutes; 0 otherwise) [b]	-0.2386 (0.0551)
Travel time x (1 - commute dummy as defined above) [b]	-0.2855 (0.0631)
Standard Deviation of Coefficients for travel time[c]	0.1942 (0.0561)
Unreliability of travel time, given as the frequency of a delay of 10 minutes or more [b]	-6.9323 (1.7815)
Standard Deviation of Coefficient for unreliability of travel time	7.8915 (2.2432)
Number of people in vehicle	0.8096 (0.4374)
Per capita household income ($000)	0.0556 (0.0517)
Flexible arrival time dummy (1 if traveler has a flexible arrival time; 0 otherwise)	2.6439 (1.3264)
Summary Statistics	
Number of Observations	641
Number of Persons	82
Log-likelihood	-251.39
Pseudo R^2	0.4024

(table continued next page)

Implied Tradeoffs	
Mean VoT in sample ($/hour)	12.61
[90% confidence interval]	[8.54, 19.90]
Mean VoR in sample ($/incident of 10+ minute delay)	4.97
[90% confidence interval]	[2.61, 18.55]

[a]Standard errors are calculated by the square root of the corresponding diagonal element in the inverse of the negative Hessian of the simulated log-likelihood function.
[b]Travel time and unreliability are each calculated as the value in the express lanes minus that in the free lanes.
[c]The coefficients of the travel time variables are specified to have a single random error in common.

Table 4. Mixed Logit Joint RP/SP Parameter Estimates

Independent Variable	Coefficient (standard error[a])
Pooled Variables	
Toll ($), divided by 2 if car occupancy is 3 or more	-2.6023 (0.4752)
Toll, multiplied by per capita household income ($000)	0.0235 (0.0111)
Flexible arrival dummy (1 if traveler has a flexible arrival time; 0 otherwise)	1.9626 (0.9503)
Standard deviation of Coefficients of travel time[c]	0.3478 (0.0898)
Ratio of standard deviation to mean of Coefficients of unreliability[d]	1.0334 (0.3757)
RP Variables	
Constant	-0.2489 (1.1505)
Median travel time (minutes) multiplied by trip distance (in units of 10 miles)[b]	-0.5318 (0.1969)
Median travel time multiplied by trip distance squared[b]	0.1182 (0.0431)
Median travel time multiplied by trip distance cubed[b]	-0.0053 (0.0020)
Unreliability of travel time; given as the difference between the 80th percentile and median times[b]	-0.8355 (0.3380)

(table continued next page)

SP Variables	
Constant	-4.5777 (1.0723)
Travel time (minutes) x commute distance dummy (1 if respondent's commute time is greater than 45 minutes; 0 otherwise)	-0.3594 (0.0844)
Travel time (minutes) x (1 - commute dummy as defined above)	-0.4616 (0.1042)
Unreliability of travel time, given as frequency of a delay of 10 minutes or more[b]	-10.9984 (1.7075)
Other Parameters	
Scale parameter μ (ratio of RP to SP standard deviations of the iid error components)	0.6286 (0.1137)
Correlation parameter θ (ratio of SP to RP individual-specific error terms)[e]	6.5871 (1.0312)
Summary Statistics	
Number of Observations	725
Number of Persons	110
Log-likelihood	-285.49
Pseudo R^2	0.4319

[a]Standard errors are calculated by the square root of the corresponding diagonal element in the inverse of the negative Hessian of the simulated log-likelihood function.

[b]Travel time and unreliability are each calculated as the value in the express lanes minus that in the free lanes.

[c]The coefficients of the travel time variables are each specified to have a common additive random error: see equations (11)-(12).

[d]The coefficients of the unreliability variables are each specified to each have a common multiplicative random error: see equation (13).

[e]The part of the RP and SP error terms for a given individual that do not depend on independent variables (i.e. the second and third terms in (7) and (8)) have a correlation coefficient ranging from zero when $\theta=0$ to one as $\theta \to \infty$.

Table 5. Values of Time and Unreliability from Joint RP/SP Models

	Mean Estimate	90% Confidence Interval[a] [5%-ile, 95%-ile]
RP Estimates		
Value of time		
Median in sample	$15.64/hour	[$1.93, $27.53]
Unobserved heterogeneity[b]	$12.95/hour	[$7.25, $19.55]
Total heterogeneity in sample[b]	$19.20/hour	[$10.54, $30.86]
Value of unreliability		
Median in sample	$23.26/hour	[$7.63, $41.53]
Unobserved heterogeneity[b]	$30.49/hour	[$8.37, $58.22]
Total heterogeneity in sample[b]	$32.95/hour	[$9.31, $62.66]
SP Estimates		
Value of time		
Median in sample	$10.20/hour	[$6.42, $14.45]
Unobserved heterogeneity[b]	$12.74/hour	[$7.19, $19.04]
Total heterogeneity in sample[b]	$13.49/hour	[$7.80, $20.37]
Value of unreliability		
Median in sample	$5.15/incident	[$3.02, $8.12]
Unobserved heterogeneity in sample[b]	$6.65/incident	[$3.33, $10.24]
Total heterogeneity in sample[b]	$6.87/incident	[$3.46, $10.42]

[a] The confidence interval represents uncertainty due to statistical error, not heterogeneity. It is determined by Monte Carlo draws from the assumed statistical distributions of the parameter estimates. This method is more accurate than approximation formulas based on the standard errors of an correlation among coefficient estimates. The distributions of these ratios are skewed, so the standard deviation would give a misleading characterization of precision. A positive 5th percentile value means the quantity is significantly greater than zero according to a conventional one-sided hypothesis test at a 5 percent significance level.

[b] Heterogeneity is measured as the interquartile difference, i.e., the difference between the 75th and 25th percentile values, computed from Monte Carlo draws. For unobserved heterogeneity, these draws are from the estimated distribution of random parameters δ_i (for the value of time) or ω_i (for the value of unreliability in equations (11)-(13)). For total heterogeneity, the draws are from that distribution and from the relevant RP or SP sample.

Table 6. Simulation Results

PRICING REGIME[a]	Base Case: No toll	Second-best toll: Heterogeneity present	Second-best toll: Heterogeneity not present	First-Best Differentiated Toll	Limited Differentiated Toll	Limited Uniform Toll
Toll:						
Express lanes	0	$2.20	$0.39	$3.11	$2.05	$1.04
Regular lanes	0	0	0	$2.71	$0.67	$1.04
Travel time (minutes)						
Express lanes	14	11	11	10	11	12
Regular lanes	14	14	12	11	13	12
Consumer surplus:[b]						
High-VOT users	0	-$0.59		-$0.86	-$0.46	-$0.16
Low-VOT users	0	-$0.13		-$1.82	-$0.59	-$0.79
Homogeneous users	0		-$0.09			
Social Welfare[b]	0	$0.25	$0.02	$0.82	$0.48	$0.48

[a] Notes on pricing regimes: "Second-Best Toll (heterogeneity present)" and "Second-Best Toll (heterogeneity not present)" maximize social welfare subject to the price of the regular lanes constrained to zero, with and without heterogeneity, respectively. "First-Best Differentiated Toll" maximizes social welfare using differentiated tolls. "Limited Differentiated Toll" limits the toll in the regular lanes to $0.67 per vehicle. "Limited Uniform Toll" is the uniform toll providing the same total welfare gain as "Limited Differentiated Toll".

[b] Consumer surplus and social welfare are measured relative to the no-toll (NT) scenario and divided by the number of users in the no-toll scenario. Social welfare is equal to average of the two groups' consumer surplus plus average revenue.