



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Transportation Research Part A 39 (2005) 279–293

TRANSPORTATION
RESEARCH
PART A

www.elsevier.com/locate/tra

Valuing time and reliability: assessing the evidence from road pricing demonstrations [☆]

David Brownstone ^{*}, Kenneth A. Small

Department of Economics, 3151 Social Science Plaza, University of California, Irvine, CA 92697-5100, USA

Abstract

This paper compares results from evaluations of two recent road pricing demonstrations in southern California. These projects provide particularly useful opportunities for measuring commuters' values of time and reliability. Unlike most revealed preference studies of value of time, the choice to pay to use the toll facilities in these demonstrations is relatively independent from other travel choices such as whether to use public transit. Unlike most stated preference studies, the scenarios presented in these surveys are real ones that travelers have faced or know about from media coverage. By combining revealed and stated preference data, some of the studies have obtained enough independent variation in variables to disentangle effects of cost, time, and reliability, while still grounding the results in real behavior. Both sets of studies find that the value of time saved on the morning commute is quite high (between \$20 and \$40 per hour) when based on revealed behavior, and less than half that amount when based on hypothetical behavior. When satisfactorily identified, reliability is also valued quite highly. There is substantial heterogeneity in these values across the population, but it is difficult to isolate its exact origins.

© 2004 Elsevier Ltd. All rights reserved.

[☆] An earlier version of this paper is being published in French as “L'évaluation du temps et de la fiabilité: Résultats issus des études de tarification routière”, in André de Palma and Emile Quinet, eds., *La Tarification des Transports, Pourquoi? Pour Qui?: Les défis d'aujourd'hui et de demain (Transport Pricing—Why? For Whom?: Challenges for Today and Tomorrow)*, Paris: Economica, 2004. Copyright permission is gratefully acknowledged.

^{*} Corresponding author. Tel.: +1 949 824 6231; fax: +1 949 824 2182.

E-mail addresses: dbrownst@uci.edu (D. Brownstone), ksmall@uci.edu (K.A. Small).

1. Introduction

Since 1994, a series of demonstration projects in the United States has provided some real-life experience with congestion pricing. One form, represented by two projects in southern California and one in the Houston area, combines pricing with priority for high-occupancy vehicles in the form of “High Occupancy/Toll” (HOT) lanes. In this scheme, a set of express lanes on an otherwise free and congested road offers high-quality service to people who are willing to pay a time-varying toll and/or who ride in carpools.

These projects provide an opportunity to study some behavioral parameters that are central to the evaluation of transportation projects. The most important is the “value of time” (VOT), i.e. the marginal rate of substitution of travel time for money in a travelers’ indirect utility function. Another is the “value of reliability” (VOR), which measures willingness to pay for reductions in the day-to-day variability of travel times facing a particular type of trip. In addition, the extent of heterogeneity in VOT and VOR across the population of travelers has been shown to sometimes significantly affect project evaluation (Small and Yan, 2001).

This paper reviews and compares results on VOT, VOR, and their heterogeneity from five different data sets taken from the two HOT-lane projects in southern California. These projects provide particularly useful data for a number of reasons. First, the choice to pay to use the toll facilities is relatively independent from other travel choices because very little transit service exists on these corridors. Second, one of the projects contains sources of independent variation of travel time, reliability, and cost that makes it possible to sort out their separate effects on travel choices. Third, local travelers’ familiarity with these projects, through direct experience or media coverage, makes it possible to collect “stated preference” (SP) data on hypothetical choices in a setting where travelers are likely to understand the context.

The ability to collect meaningful SP data presents an opportunity to examine reasons for the rather large differences seen in recent literature between estimated values of time from revealed preference (RP) studies and those from SP studies. It also opens the possibility of combining RP and SP data so as to take advantage of both the realism of the former and the controllability of the latter.

2. Project descriptions

The State Route 91 (SR91) facility in Orange County, California, is a privately-funded set of express lanes in the center of 10 miles of the very congested Riverside Freeway, linking job centers in Orange and Los Angeles counties to outlying residential areas. It uses electronic toll collection and has a complex but preset time-varying pricing structure, presumably designed to maximize revenue for the private toll operator. Prices vary hour by hour and follow different schedules on different days of the week, including weekends. There are separate schedules for inbound and outbound directions, but all our data are inbound. See Sullivan et al. (2000) and OCTA (2003) for more information. Between 1996 and 2000, one author of this paper (Small) and various colleagues collected five separate data sets on this corridor, mostly using mail surveys. Three of these data sets are included in the comparisons of this paper.

The Interstate 15 (I-15) facility is a publicly-funded project that allows solo drivers to pay to use reversible carpool lanes over an 8-mile congested segment of the I-15 freeway linking San Diego

employment centers with inland northern suburbs. The I-15 project also uses electronic toll collection, with carpools exempted. In this case, the price is varied in real time to prevent the express lanes from becoming congested, a procedure sometimes called “dynamic pricing”. The other author of this paper (Brownstone) and his colleagues have administered five waves of a telephone survey to a panel of travelers on this corridor; the third and fifth waves, collected in Fall 1998 and Fall 1999, are used in the results described here. See Golob and Golob (2001), Golob et al. (1998), and SANDAG (2003) for more information.

On both projects, the toll is announced on a message sign prior to the point where the driver has to choose which lane to enter. Travel time, by contrast, cannot be observed in advance as we explain below. We assume that drivers are familiar with the distribution of travel times across days, for any given time of day, and we measure value of time and reliability as preferences about this distribution.

The two projects differ in some ways that affect efforts to measure behavioral parameters. On SR91, the fixed toll schedule and the revenue needs of the operator result in considerable variation in the ratio of time savings to toll across different hours of the week.¹ Furthermore, the daily pattern of reliability is still different because the effects of incidents tend to be long-lasting and so unreliability peaks at a later time of day than does mean or median travel time. Another advantage of this corridor to the analyst is that carpools of three or more people can travel in the express lanes for half price, thereby providing additional variation in the price paid. On I-15, by contrast, the toll is varied in real time in 6-min intervals, and is targeted explicitly to maintain a constant speed in the express lanes. This makes the toll more closely correlated with the travel-time savings and the improved reliability realized in the express lanes.² Also, carpools can use the I-15 express lanes for free, so they do not face an interesting lane choice.

Dynamic pricing on I-15 introduces a new feature that is a nuisance for the analyst but also an opportunity to study a very interesting phenomenon. Travelers can use the observed toll to learn something about how bad congestion is in the unpriced lanes. For example, if on a particular day the observed toll is unusually high for that time of day, this fact informs the driver that congestion in the unpriced lanes ahead is unusually severe. Thus the I-15 toll plays two different roles, pricing and signaling, which must be separately accounted for. This is done by using the actual toll on the day of travel as the value for the cost variable, and by including an additional signaling variable equal to the difference between this actual toll and the median toll (across days at the same time of day). Our reported values of time on I-15 are calculated using only the toll coefficient itself, not the coefficient of this additional variable.

On SR91, by contrast, informal observations by our test drivers suggest that drivers do not have much information about congestion beyond the distribution across days. This is because

¹ At the times of data collection, this project was fully private based on a franchise agreement with the State of California. The operator had a free hand in setting tolls, subject only to an overall rate of return cap. The ratio of toll to time saving was much higher during off-peak than peak periods, presumably reflecting a policy resembling profit-maximizing congestion pricing (see Small, 1992, Section 4.6A). On January 3, 2003, the franchise was purchased by a public agency, the Orange County Transportation Authority. The purchase was financed with revenue bonds, whose debt payments create significant financial pressure to maintain tolls that are near profit-maximizing.

² Fortunately for us, there is one interchange where traffic entering the express lanes bypasses an entry-ramp queue, thereby creating additional variation across our sample in the ratio of time savings to toll.

congestion ahead cannot be predicted very well by conditions upstream of the toll lanes. Furthermore, an earlier study on SR91 by Parkany (1999) found that drivers made little use of information sources about congestion other than the radio, which provided only crude indications of travel times to be expected on the specific stretch of highway for which the lane choice applies. Thus we believe the travel-time distribution, not its realization on the day in question, is the most important travel variable affecting traveler choice on SR91.

3. Empirical models

All the models we discuss assume that a traveler i faces an actual or hypothetical choice at time t among alternatives j . The alternatives include commuting lanes (toll or free) and possibly other travel features like carpooling, time of day, or acquisition of an electronic transponder. Using notation adapted from Small et al. (2002), the traveler chooses the option that maximizes a random utility function

$$U_{ij} \equiv \theta_{ij} + \beta_i X_{ij} + \varepsilon_{ij}. \quad (1)$$

Variables included in X_{ij} measure the toll C_{ij} , travel-time T_{ij} , and (un)reliability R_{ij} . The values of travel time and reliability are defined as

$$\text{VOT}_i = \frac{\partial U_{ij} / \partial T_{ij}}{\partial U_{ij} / \partial C_{ij}}; \quad \text{VOR}_i = \frac{\partial U_{ij} / \partial R_{ij}}{\partial U_{ij} / \partial C_{ij}}. \quad (2)$$

In general, the derivatives in (2) allow VOT and VOR to depend not only on the individual traveler i but also on the alternative j or the time t that a choice is made; but all the specifications we consider restrict such variation to dependence on i , as indicated by the subscripts on VOT and VOR. However, we do allow them to depend on whether a given individual is answering an RP or an SP question. The properties of the random term ε_{ij} depend on the situation: in some cases correlation is assumed across certain alternatives j (for example, those alternatives requiring a transponder) and/or times t (for example, separate hypothetical questions on a single questionnaire), while in others different variances are assumed for subsamples collected under different conditions (as recommended by Louviere and Hensher, 2001).

Some of the models distinguish between two types of heterogeneity in preferences across individuals: observed and unobserved. This is accomplished by specifying the estimable parameters, θ_{ij} and β_i , as functions of both observed characteristics of the individuals and unobserved random terms

$$\theta_{ij} = \bar{\theta} + \phi W_i + \xi_{ij}, \quad (3)$$

$$\beta_i = \bar{\beta} + \gamma Z_i + \zeta_i. \quad (4)$$

Observed heterogeneity is captured by variables X_{ij} , W_i , and Z_i , while unobserved heterogeneity is captured by the random terms ξ_{ij} and ζ_i . That part of heterogeneity represented by W_i or ξ_{ij} concerns absolute preferences for alternatives, while that represented by Z_i or ζ_i concerns preferences about tradeoffs and is therefore what is relevant to measurement of VOT and VOR. Heterogeneity represented by X_{ij} may be of either type, depending on whether X_{ijt} affects ratios of marginal

utilities such as appear in (2). An example of this distinction is the dependence of VOT on income, a topic considered extensively in the literature.³ In some models, income is included in Z and/or in X in such a way as to affect VOT; in others, it is included in W and/or in X in such a way as to affect only absolute preferences for particular choices.

4. Empirical results

In this section we consider and attempt to answer a number of empirical questions that arise from studies of value of time (VOT) and value of reliability (VOR). Table 1 provides selected results. Since VOT varies across individuals, as described in the previous section, Table 1 reports medians and interquartile ranges across the samples after weighting to reverse the effects of choice-based sampling. We have given confidence bands only for a few of the estimates in Table 1. In some cases they are not readily available, and they are similar within the two corridors. The SR91 results are more precisely estimated because there is more independent variation between tolls and time savings.

4.1. How big is the value of time?

While there has always been variation among empirical estimates, a decade ago there seemed enough consistency for Small (1992) to suggest that

[A] reasonable average value of time for journey to work is 50 percent of the gross wage rate... [I]t varies among different industrialized cities from perhaps 20 to 100 percent of the gross wage rate, and among population subgroups by even more. (p. 44)

Recently, however, Calfee and Winston (1998) have argued, based on results from a nationwide SP survey, that the average value of time is much lower for long-distance automobile commuters. They suggest this is due to self-selection of low-VOT individuals into residential locations requiring long automobile commutes.

Results from the SR91 and I-15 corridors based on RP lane-choice data almost always obtain median values of time of \$20 to \$40 per hour, at least five times as large as the results of Calfee et al. (2001). These values are typically 50–90% of the average wage rate in the sample, consistent with Small's earlier summary.⁴ It may be that high prices in the Los Angeles area have skewed the residential sorting process described by Calfee and Winston so that even long-distance commuters there have higher values of time than in other regions. But there are also other potential explanations for the differences, as discussed later.

Furthermore, there is consistency in the results for these two corridors when differences in the samples are accounted for. Row 10 of Table 1 shows the results of re-estimating the I-15 results with weights applied to the sample so that it reproduces the joint distribution of income and distance observed in the SR91 sample. This lowers the estimated median VOT to a value quite similar to that found on SR91.

³ See, for example, MVA Consultancy et al. (1987) or Wardman (2001).

⁴ See, for example, Lam and Small (2001, p. 238, 240, 250), Small et al. (2002, Section 5).

Table 1
Comparison of selected model results (with selected 90% confidence bands in parentheses)

		Estimation		Coefficients used for computing VOT, VOR	VOT (\$/hour)			VOR (\$/hour) ^a (median)	
		Data source used	Error components included		Median	Heterogeneity (inter-quartile range)		Male	Female
						Observed	Unobserved		
<i>SR91</i>									
1	Lam–Small: route only ^b	RP	None	RP	24	8	NA	\$12/hr	\$30/hr
2	Lam–Small: route, mode, transponder ^c	RP	None	RP	23	NR	NA	\$15/hr	\$32/hr
3	SWY ^d	RP	None	RP	25	13	NR	NA	
4	SWY ^{e,f}	RP/SP	Full ^g	RP	20* (15 – 26)	4*	11*	\$20/hr ^h (8 – 31)	
5	SWY ^{e,i}	RP/SP	Full ^g	SP	9*	2*	13*	\$4/incident	
6	SWY ^j	RP/SP	Partial ^k	RP	21*	NR	NA	NA	
7	SWY ^j	RP/SP	None	RP	20*	NR	NA	NA	
<i>I-15</i>									
8	Brownstone et al. (wave 3) ^l	RP	None	RP	30	20	NA	NA	
9	Steimetz–Brownstone (wave 5) ^m	RP	None	RP	45, 30*	52	NA	NR	
10	Steimetz–Brownstone (wave 5) weighted to match SR91 sample ⁿ	RP	None	RP	22 (9 – 40)	26	NA	NR	
11	Ghosh: route, mode, transponder ^o	RP	4 coeff's	RP	24, 21*	19, 22*	30 ^p	NA	
12	Ghosh: route, mode, transponder ^o	RP	None	RP	27	23	NA	NA	
13	Ghosh: route, cond'l on transponder ^q	RP	None	RP	40	31	NA	NA	
14	Ghosh: route, cond'l on transponder ^q	SP	None	SP	16	3	NA	NA	
15	Ghosh: route, cond'l on no trnspndr ^q	SP	None	SP	13	3	NA	NA	
<i>Other</i>									
16	Calfee–Winston–Stempski ^r	SP	4 coeff's	SP	4	NR	1	NA	

NA: not applicable (variable not included in model). NR: not reported (variable included but resulting distribution not calculated).

* Calculated by drawing randomly from the asymptotic distribution of the parameter estimates, as well as from the distribution of error terms as in the other entries in this column.

^a Unreliability is measured as 90th minus 50th percentile of travel-time distribution except where otherwise noted.

^b Lam and Small (2001), Table 6, Model 1f.

^c Lam and Small (2001), Table 11, Model 4d. This model specifies 9 choice alternatives.

^d Calculations for this paper, from coefficients of Small et al., (2002), Table 2, “RP Only” model.

^e Model was estimated on combined RP and SP data, with separate coefficients for RP, SP observations. Observed heterogeneity is not reported in SWY, but was calculated for this paper. Numbers in parentheses describe the 90–percent confidence band for the estimate shown.

^f SWY, Table 3, “RP Estimates.”

^g Error components include constants, repeated SP observations with correlation between SP and RP observations for the same individual, and random coefficients of cost, travel time, and unreliability.

^h Unreliability was measured as the 80th minus 50th percentile of travel-time distribution.

ⁱ SWY, Table 3, “SP Estimates.”

^j Additional models estimated for this paper: same specification and data as in lines 4–5 except with fewer error components.

^k Includes an error component for repeated SP observations to allow them to be correlated, and allows for correlation between the SP and RP observations for the same individual.

^l Brownstone et al. (2003), VOT calculations described in text.

^m Steimetz and Brownstone (2005), Table 4. Uncertain travel times are multiply imputed based on prediction models using loop detectors and other variables.

ⁿ From same model as previous row, but VOT distribution recomputed for this paper after reweighting individuals in I-15 sample to have same joint distribution of income and distance as the SR91 sample. Numbers in parentheses describe the 90–percent confidence band for the estimate shown.

^o Ghosh (2001), Table 6, p. 78, weighted models, parts 1 and 3. Logit error structure with 5 alternatives: (1) Free lanes, solo driver, no transponder; (2) Free lanes, solo, with transponder; (3) Express lanes, solo, with transponder; (4) Express lanes, carpool, no transponder; (5) Express lanes, carpool, with transponder. Estimated on data from wave 5 of the I-15 panel study.

^p Approximated as 1.3 times the reported standard deviation of drawn values of time. The factor 1.3 is the difference between the 75th and 25th percentiles of the standard normal distribution.

^q Ghosh (2001), Table 16, p. 113. Estimated on data from wave 5 of the I-15 panel study.

^r Calfee et al. (2001), Table 5, Scenario 1. We have approximated the inter-quartile range IQR of the value of time from the reported estimates of the mean (\$3.71) and the 90th and 10th percentiles (\$5.06 and \$2.53), based on the fact that VOT is log–normally distributed in their specification.

4.2. How important is reliability?

Not all the studies have been able to disentangle reliability from other factors well enough to measure its effects. Measuring reliability requires substantial numbers of observations of speeds across days, for any given time of day. Most such measurements rely upon magnetic loop detectors embedded in the roadway, which measure the density of vehicles and the time between vehicles. Converting this information to speed requires assumptions about the distribution of vehicle lengths. In addition, loop detectors are notoriously subject to failures that result in mistaken readings, which must be recognized and deleted, as well as missing readings. Thus there are numerous uncertainties in the mathematical algorithms used to derive speed information from loop detector data, as well as maddening gaps in the data coverage. In most of the studies described here, loop detector data are used only as supplements to data obtained painstakingly from students driving on the relevant sections of roadway with stopwatches. Brownstone et al. (2001) document the large systematic errors in the loop detector data for the I-15 study, and they describe a multiple imputations methodology that accounts for these errors.

Theory suggests that the most common measure of unreliability, the standard deviation of travel time across days, will only imperfectly capture traveler preferences. Most theories explaining aversion to unreliable travel are based on costs of unexpected arrival times at work (Bates et al., 2001), which are greater for being late than for being early; but using the standard deviation to measure unreliability assumes that negative and positive deviations have symmetric effects. For this reason, the SR91 and I-15 studies rely instead on measures of the upper tail of the distribution of travel times, such as the difference between the 90th and 50th percentile travel times. Such a measure should be closely related to the chance of being substantially later than expected.⁵

Furthermore, reliability is closely intertwined with the signaling function of the real-time price on the I-15 express lanes, because travelers care more about the information content of the signal if travel time is unreliable. For this reason, some of the I-15 studies contain as a variable the interaction between actual toll and travel-time unreliability.⁶ This control seems to be necessary for obtaining sensible results on reliability, but unfortunately it causes the coefficient of unreliability to be measured only imprecisely.

Because of these difficulties, only the SR91 studies appear to have satisfactorily identified the coefficients of unreliability in RP data. The results are fairly sensitive to specification. They suggest that reliability as defined above is valued at roughly 95–140% as highly as median travel time, depending on the measure.⁷ Multiplying these values by the actual travel-time and reliability differences found during the peak period on SR91 in 1999, Small et al. (2002) (hereafter SWY) find that travel time accounts for about two-thirds, and reliability one-third, of the service quality differential between the free and express lanes.

⁵ Small et al. (2002) use the 80th – 50th percentile difference rather than the 90th – 50th difference, because the data are too sparse to accurately determine the 90th percentile.

⁶ For example Brownstone et al. (2003, Table 3).

⁷ For example, Lam and Small (2001) find $VOR/VOT = 1.39$ with unreliability measured as the 90th – 50th percentile difference (Table 3, Model 1c). Small et al. (2002) measure *median* VOR and *median* VOT in the population, finding they have a ratio of 0.97 (Table 3, RP estimates).

One of the studies finds a much higher VOR for women than men—roughly twice as high.⁸ A possible reason is that women have more child-care responsibilities, which reduce their scheduling flexibility. A higher VOR provides one possible explanation for the consistent finding across nearly all studies on these two corridors that, other things equal, women are more likely than men to choose the toll road.

4.3. Do stated preference estimates underestimate value of time?

Two of the studies, one on each facility, have collected data on revealed preferences (RP) and stated preferences (SP) from the same or overlapping subsets of respondents. They have then attempted to compare values of time from these two sources in a controlled manner. In both cases, the median SP estimates are about half the median RP estimates, a statistically significant difference. These median values are shown in [Table 1](#), rows 4–5 and 13–14.

In the case of SWY, the values shown are from a joint model in which the RP and SP observations were allowed to have separate coefficients on cost, travel time, and unreliability, as well as separate alternative-specific constants and variances. Similar distinctions were made between two different subsets of RP data. There are 522 distinct individuals with RP information and 81 distinct individuals with SP information (for a total of 633 SP observations, since each respondent was given up to eight SP scenarios). These two groups have 55 distinct individuals in common. As we see from [Table 1](#) (rows 4–5), the SP results imply a much lower median VOT than the RP results, and also a greater amount of heterogeneity as a fraction of the median value.

In the case of Ghosh (2001), two separate models are estimated on the RP and SP observations, which were obtained mostly from the same individuals in the same telephone interview (our [Table 1](#), rows 13–14). These are all individuals who had already obtained a transponder that enables them to use the express lane if they so choose, and so they are self-selected to have higher values of time. The sample sizes are 266 for RP and 306 for SP. These models account for a more complete choice set than most others, including as possible choices five combinations of route (i.e. express versus regular lanes), car occupancy, and transponder acquisition. Again, the median VOT from the SP responses is less than half that from RP. In this case the heterogeneity is relatively greater for RP than SP. Ghosh's results were also replicated using a joint mixed logit model that is very close to SWY's model; this mixed logit model (not shown in the table) yields median VOT from RP data three times the size of that from SP data, based on a sample of 456 commuters on I-15.

These results favor the hypothesis that earlier differences among RP and SP studies are caused by systematic differences in RP and SP methodologies. While we cannot say with any certainty which type of data is more trustworthy, two considerations argue in favor of the RP results. First, as noted by Louviere and Hensher (2001), values of time measured from SP data have been found to depend strongly on such elements of experimental design as range of variables presented in the survey. Second, it is clear that RP results correspond to what planners need to know in order to evaluate transportation projects. This is because the analyses that form the basis for project evaluation are in terms of actual travel times, similar to what is measured in RP variables. If people

⁸ Lam and Small (2001, Tables 8–11).

react differently to hypothetical scenarios than to real ones with the same conditions, it is the latter that normative studies must replicate in order to provide guidance about the benefits or costs of changes in the transportation environment.

Interestingly, the possibility that SP values understate the true values is the opposite of the concern over use of “contingent valuation” measures, another form of questions about hypothetical situations, in the literature on environmental economics (Carson, 2000). Of course the contexts are very different, but there are at least two possible explanations for understated SP results that are specific to congested highways.

The first is that people display time inconsistency in their actual behavior, but not in their hypothetical behavior. In their actual travel they may intend to choose the cheaper roadway, but then neglect to allow sufficient time and so be forced by a scheduling constraint to take the express lanes. It is entirely plausible that they would not account for such errors in implementing their own plans when answering SP questions. Thus, they make the higher-cost choice more often in real life than on hypothetical surveys.⁹

The second possibility is that the difference is caused by systematic misperception of travel times. In studies of both SR91 and I-15, people have been asked to report the travel-time savings they think they could realize by using the express lanes. Their responses are typically about twice the actual travel-time savings. A possible reason for this could be that they focus on total delays on a portion of their trip that is actually longer than the section where there is an express-lane option; they may mistakenly think that this larger delay would all be eliminated by using the express lanes. Another possible reason is impatience with heavy traffic that leads them to exaggerate how much time the delays are actually causing them. (The latter might be related to the finding of a few studies that congested travel time is valued at about twice as much per minute as uncongested travel time—see also Steimetz, 2004.) Whatever the reason, if people experiencing a 10-min time delay remember it as 20 min, then they probably react to a hypothetical question involving a 20-min delay in the same way that they react to a real situation involving a 10-min delay. This would cause their measured value of time in the hypothetical situation to be exactly half the value observed in real situations.

To investigate this issue further, Ghosh (2001, Section 5.2) uses perceived time savings to help explain route, mode, and transponder choice. In the best specification, perception error (defined as perceived minus actual time savings) is simply added as an explanatory variable to the model of row 12 in Table 1. He finds that commuters with larger positive perception errors are more likely to use the toll facility. However, the RP values of time are not changed by including this variable, suggesting that RP results may not be affected by perceptual problems. Whether or not SP results are so affected, as our hypothesis implies, remains to be studied.

4.4. Are estimated VOT and VOR sensitive to error structure?

Hensher (2001) compares values of time estimated from several alternative statistical models, using SP responses from a sample of people asked to consider intercity high-speed rail service for nonwork trips in Australia. He finds that the estimated median values of time are substantially

⁹ We are indebted to Steven Berry for suggesting this possibility and to Kay Axhausen for suggesting an empirical test using data on unforeseen scheduling change such as child-care emergencies.

larger when the model allows for more general correlation structures in the error terms and for more unobserved variability across the sample. He interprets this finding to mean that at least for these data, the more sophisticated models correctly measure values of time and the simplified models underestimate them.

We consider here whether the SR91 and I-15 studies provide similar evidence of bias. In three cases, otherwise identical models have been estimated with varying degrees of complexity in the error structure. Based on these, we are unable to replicate Hensher's finding. For one example, earlier results from the SP portion of the SR91 data, not reported in SWY, compared a binary logit model with no random parameters to one with a random constant and another with a random constant and two additional random parameters. The results are shown in Table 2. (These include some variables solely for comparison with other models not described here.) Accounting for random parameters produces great improvements in goodness of fit, and the estimated

Table 2
Comparison of SP Models with and without random parameters (SR91)

Independent variable	Parameters randomized		
	None	Constant	Constant, time, reliability
<i>Constant</i>			
Mean	-1.502 (0.858)	-2.452 (2.213)	-4.870 (3.497)
Std. dev.		3.368 (0.603)	4.364 (1.395)
<i>Cost (\$)</i>			
	-0.497 (0.118)	-1.060 (0.194)	-1.380 (0.344)
<i>Travel time (min)</i>			
Mean	-0.104 (0.018)	-0.218 (0.037)	-0.277 (0.066)
Std. dev.			0.200 (0.085)
<i>Unreliability^a</i>			
Mean	-2.496 (0.618)	-5.402 (1.188)	-6.690 (1.762)
Std. dev.			7.148 (2.122)
Car occupancy	0.438 (0.158)	0.955 (0.347)	1.455 (0.798)
Work-site size ^b	-0.390 (0.135)	-0.670 (0.234)	-1.180 (0.766)
Distance ^c	-0.345 (0.338)	-1.015 (0.898)	-0.536 (1.163)
Distance squared	0.047 (0.035)	0.119 (0.089)	0.080 (0.101)
Work-hour flexibility ^d	0.707 (0.385)	1.485 (0.959)	2.129 (1.724)
Education ^e	0.310 (0.421)	0.626 (1.064)	1.023 (2.068)
Log-likelihood	-326.26	-235.75	-227.20
# of observations	601	601	601
Implied mean VOT (\$/h)	12.55	12.35	12.08
Implied mean VOR (\$/incident)	5.02	5.09	4.85

Notes: Dependent variable: choice of toll road. Standard errors in parentheses. All travel variables describe a one-way trip.

^a Fraction of trips with unexpected delays of 10 min or more (SP question).

^b Number of workers at work site, in 1000 s.

^c One-way commute distance, in units of 10 miles.

^d Dummy variable equal 1 if work arrival time or work departure time are flexible.

^e Dummy variable equal 1 if have bachelor degree or higher.

standard deviations of the parameters are strongly statistically significant. Yet these results do not show substantial differences in either mean VOT or mean VOR.

The full version of the SWY model (Table 1, rows 4–5) was also re-estimated for this paper with simpler error structures (rows 6 and 7). Comparisons of the RP values of time (rows 4, 6, and 7) show no apparent trend in median VOT.

Ghosh (2001, Section 3.1) compares conditional logit, nested logit, heteroscedastic extreme value, and mixed logit models for a 5-alternative choice model including the choice of whether to carpool and whether to acquire a transponder. The heteroscedastic extreme value model showed the best improvement in overall fit, but the implied values of time show no clear pattern and are not statistically different from each other (Ghosh, Table 6). The most general model, mixed logit, yielded the lowest values of time, which contradicts Hensher's findings.

4.5. *Heterogeneity and variable specification*

Advances in computing power and simulation methods for estimating nonlinear models have heightened interest in the sources of heterogeneity in behavioral parameters such as VOT and VOR. Recent work has shown that such heterogeneity has important implications for design of road pricing schemes, making differentiated pricing such as that on HOT lanes much more attractive (Small and Yan, 2001; Small et al., 2002). Thus it seems natural to try to isolate the sources of heterogeneity.

As noted earlier, the observed heterogeneity in VOT or VOR depends on what variables are interacted with cost, travel time, or unreliability in the specification. In the SR91 and I-15 studies, the models contain interaction terms between cost and income, travel time and income, and/or travel time and powers of distance, the latter leading to an inverted U-shaped relationship between value of time and distance. In some models, unreliability is interacted with toll (I-15 only) and/or with various traits of individuals such as work-hour flexibility, gender, and income.

In many cases, the model fits as well or better if instead these traits are interacted with alternative dummies—e.g., they are allowed simply to shift the alternative-specific constant for choosing the express lanes. The two forms often produce similar estimates of most coefficients, including those determining median VOT and VOR. However, they produce drastically different predictions of the observed heterogeneity of VOT and VOR. If unobserved heterogeneity is also allowed by the specification, through random parameters, then it will also be different. Complicating the interpretation of results is the fact that unobserved heterogeneity typically affects the normalization of the error term and therefore may cause most coefficients to be scaled upward or downward.

Currently, this inability to distinguish clearly between different specifications of interaction terms is an important limiting factor for understanding the forms of heterogeneity in travel behavior. We suspect that the greatest future advances in this understanding will come not from more sophisticated specifications of error terms or random parameters, but rather from better measures of the individual traits that systematically affect the value travelers place on time and reliability.

Observed heterogeneity also complicates the estimation of summary statistics like median VOT. Two of the studies in Table 1, SWY (2002) and Steimetz and Brownstone (2005), include various interaction terms among toll, time savings, income, and commute distance. Their specifications imply that estimated VOT is a highly nonlinear function of observed data and parameter esti-

mates. Most of the results for median VOT reported in Table 1 are evaluated at the point estimates of the unknown parameters. A better estimate is the expected value of median VOT, where the expectation is taken over the sampling distribution of the parameter estimates. (This expected value is also asymptotically equal to the optimal Bayesian posterior estimate of median VOT.) This calculation, where performed, is marked * in Table 1. In the two cases where it is calculated both ways (rows 9, 11), the use of point estimates of parameters rather than their entire distributions appears to overstate the median VOT. Since the VOT functions are highly nonlinear, there is no reason to expect that this overstatement will apply to other data and models.

5. Conclusions

Perhaps the most satisfying conclusion from studies of the I-15 and SR91 toll facilities is that they yield very similar estimates of the value of time. When the I-15 sample is weighted to match the income and commute distance distribution in the SR91 sample, all studies find a roughly \$20 VOT for people with characteristics of SR91 commuters. This level of agreement is surprising since the corridors differ in their pricing schemes and the studies use different survey questionnaires and methods. This close agreement gives us confidence that our empirical findings are not just artifacts of a particular survey and model.

We find that commuters' values of time on these corridors, when based on RP data, are roughly consistent with Small's (1992) earlier survey, and at least twice as large as more recent SP results. We are able to replicate this kind of difference between SP and RP estimates of value of time within our samples; thus at least some of the difference is inherent in the methodologies, at least as currently practiced in these studies. Our results suggest that using SP data in computing benefits from travel-time savings will undervalue projects whose purpose is to reduce congested travel time. Explaining the large differences between SP and RP estimates is clearly an important topic for future research.

We are encouraged by the possibility of accurately measuring the value of improvements in the reliability of travel. Further advances in studying reliability should be a top research priority. They depend primarily upon collection of data showing more precisely the uncertainty that is experienced by the individual at the time of decision-making. Further theoretical progress identifying the information available and nature of the decision to be made at various points in time, along the lines of Lam (2000), would also be very helpful. We believe the studies reviewed here have made some progress by showing that reliability can be fruitfully modeled as a property of the upper tail of the travel-time distribution across days. But clearly the way travelers respond to such distributions, and how they acquire information about the random draw they are about to experience, is complex and important in understanding many policy issues facing transportation planners.

Acknowledgement

We thank Seiji Steimetz and Jia Yan for research assistance and help with interpretation, and Steven Berry, André de Palma, David Hensher, and Emile Quinet for helpful comments and suggestions on earlier drafts. Arindam Ghosh and Tom Golob also provided many useful insights

and comments on the I-15 work. This research has been supported by the US Department of Transportation and the California Department of Transportation, through the University of California Transportation Center. The views expressed here and any errors are the sole responsibility of the authors.

References

- Bates, J., Polak, J., Jones, P., Cook, A., 2001. The valuation of reliability for personal travel. *Transportation Research E* 37 (2–3), 191–229.
- Brownstone, D., Golob, T.F., Kazimi, C., 2001. Modeling non-ignorable attrition and measurement error in panel surveys: an application to travel demand modeling. In: Groves, R.M., Dillman, D., Eltinge, J.L., Little, R.J.A. (Eds.), *Survey Nonresponse*. Wiley, New York, pp. 373–388.
- Brownstone, D., Ghosh, A., Golob, T.F., Kazimi, C., Van Amelsfort, D., 2003. Drivers' willingness-to-pay to reduce travel time: evidence from the San Diego I-15 Congestion Pricing Project. *Transportation Research A* 37, 373–387.
- Calfee, J., Winston, C., 1998. The value of automobile travel time: implications for congestion policy. *Journal of Public Economics* 69, 83–102.
- Calfee, J., Winston, C., Stempski, R., 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, 699–707.
- Carson, R.T., 2000. Contingent valuation: a user's guide. *Environmental Science and Technology* 34, 1413–1418.
- Ghosh, A., 2001. Valuing time and reliability: commuters' mode choice from a real time congestion pricing experiment. Ph.D. Dissertation, Department of Economics, University of California at Irvine.
- Golob, J.M., Golob, T.F., 2001. Studying road pricing policy with panel data analysis: the San Diego I-15 HOT Lanes. In: Hensher, D.A. (Ed.), *Travel Behaviour Research: The Leading Edge*. Pergamon, Oxford, pp. 869–883.
- Golob, J.M., Supernak, J., Golob, T.F., Kawada, K., 1998. An evaluation of a high occupancy toll (HOT) lane demonstration project in San Diego. In: *Proceedings of the European Transport Conference: Seminar C: Policy, Planning and Sustainability*, vol. 2. PTRC Education and Research Services, London, pp. 255–269.
- Hensher, D.A., 2001. Measurement of the valuation of travel time savings. *Transportation Research E* 37 (2–3), 71–98.
- Lam, T.C., 2000. Route and scheduling choice under travel time uncertainty. *Transportation Research Record* 1725, 71–78.
- Lam, T.C., Small, K.A., 2001. The value of time and reliability: measurement from a value pricing experiment. *Transportation Research E* 37 (2–3), 231–251.
- Louviere, J.J., Hensher, D.A., 2001. Combining sources of preference data. In: Hensher, D.A. (Ed.), *Travel Behaviour Research: The Leading Edge*. Pergamon, Oxford, pp. 125–144.
- MVA Consultancy, Institute for Transport Studies of the University of Leeds, and Transport Studies Unit of the University of Oxford, 1987. The value of travel time savings: a report of research undertaken for the Department of Transport. *Policy Journals*, Newbury, UK.
- OCTA, 2003. 91 Express Lanes, Orange County Transportation Authority, California <http://www.91expresslanes.com/>, accessed 23 May 2003.
- 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.
- SANDAG, 2003. I-15 FasTrak Online, San Diego Association of Governments <http://argo.sandag.org/fastrak/>, accessed 23 May 2003.
- Small, K.A., 1992. *Urban Transportation Economics*, Fundamentals of Pure and Applied Economics Series, vol. 51. Harwood Academic Publishers, New York, Reprinted by Routledge, Abingdon, UK.
- Small, K.A., Yan, J., 2001. The value of “value pricing” of roads: second-best pricing and product differentiation. *Journal of Urban Economics* 49, 310–336.
- Small, K.A., Winston, C., Yan, J., 2002. Uncovering the distribution of motorists' preferences for travel time and reliability: implications for road pricing, working paper #546 University of California Transportation Center, August <http://www.uctc.net/papers/546.pdf>. *Econometrica*, shorter version forthcoming.

- Steimetz, S.S.C., 2004. Defensive driving and the external costs of accidents and travel delays. Working paper. Department of Economics, University of California at Irvine, April.
- Steimetz, S.S.C., Brownstone, D., 2005. Estimating commuters' 'Value of time' with noisy data: a multiple imputation approach. *Transportation Research B*, in press, doi:10.1016/j.trb.2004.11.001.
- Sullivan, E., Blakely, K., Daly, J., Gilpin, J., Mastako, K., Small, K., Yan, J., 2000. Continuation study to evaluate the impacts of the SR91 value-priced express lanes: final report. California Polytechnic State University at San Luis Obispo, December <http://ceenve.ceng.calpoly.edu/sullivan/SR91>.
- Wardman, M., 2001. A review of British evidence on time and service quality valuations. *Transportation Research E* 37 (2–3), 107–128.