

Presented at the International Conference on Survey Nonresponse, October 28-31, 1999 Portland, OR. To appear in *Survey Nonresponse*, New York: Wiley.

Modeling non-ignorable attrition and measurement error in panel surveys: an application to travel demand modeling

David Brownstone (dbrownst@uci.edu) and Thomas F. Golob
University of California, Irvine

Camilla Kazimi
San Diego State University

1 Introduction

Modern panel surveys frequently suffer from high and likely non-ignorable attrition, and transportation surveys suffer from poor travel time estimates. This paper examines new methods for adjusting forecasts and model estimates to account for these problems. The methods we describe are illustrated using a new panel survey of 1500 commuters in San Diego, California. These data are being collected to evaluate a federally-funded “Congestion Pricing” experiment investigating the impacts of allowing solo drivers to pay to use freeway carpool lanes. The panel survey, begun in Fall 1997, collects data on travel behavior and attitudes at six-month intervals through telephone interviews. The panel sample is refreshed with new respondents at each wave to counteract the attrition between waves. Both the original and refreshment samples are stratified on commuters’ mode choice (solo drive in free lanes, pay to solo drive in the carpool lanes, or carpool for free in carpool lanes) to insure sufficient sample size for estimating our models.

We illustrate this methodology using a standard conditional logit model of commuters’ mode choice (solo drive in free lanes, pay to solo drive in the carpool lanes, or carpool for free in carpool lanes). The basic model is documented in Kazimi *et al.* (2000), and it is summarized in Sections 2 and 6 of this chapter. Our model is calibrated from the third wave of the panel study which was collected in Fall, 1998. We use data from the second wave to estimate an attrition model and then use this model to predict attrition probabilities as described in Section 5. We expect non-ignorable attrition because commuters who use the carpool lanes are more interested in the survey questions. It turns out that attrition does not significantly bias key parameter estimates even though there is some indication that it is non-ignorable.

We also have potentially non-ignorable measurement error in the time saved by using the carpool lane. Objective measurements of time savings are available from two types of data on speeds. First, floating car observations were obtained by driving cars down the corridor at frequent intervals and recording the actual travel times. During wave 3 of the panel survey, these floating car measurements were carried out for 5 days, but the panel survey data collection involved reported travel behavior over two months. Second, point speeds derived from magnetic loop detectors placed along the corridor for general traffic counting purposes were available during the entire data collection period, but these data are subject to significant errors as described in Section 4.

We have built a model that predicts the floating car data from the loop detector data. This model fits well (R-squared of .9), and we use it to predict the actual time savings faced by each survey respondent as a function of the date and time they entered the corridor. We use multiple imputations to account for the component of error in our estimates and predictions from this imputation model. Correcting for measurement error leads to significant differences in key model estimates as reported in Section 6.

We view measurement error as formally equivalent to nonresponse for the true key time savings variable. Since we have external data, we can model the measurement error process and use multiple imputations (which was originally devised to handle non-response) to correct for the problems caused by

measurement error. In our nonlinear model, measurement error in any independent variable causes all of the parameter estimates to be inconsistent. Even if measurement error doesn't bias some mean estimates in simpler models, it almost always biases inferences unless some corrections are made. In our application, the measurement error occurs in engineering data, but measurement errors are endemic in survey samples. (see Biemer, *et al.*, 1991, Groves, 1989, and Lessler and Kalsbeek, 1992). Fuller (1987) and Carroll, *et al.* (1995) review methods for modeling measurement error and correcting its effect on inference.

This chapter examines the impact of a number of common survey problems (non-random sampling, panel attrition, and measurement error) on estimates from a nonlinear model. Although only one of these problems (measurement error) led to significant biases, there is no way to know without carefully analyzing the impacts of all possible problems.

2 The San Diego Congestion Pricing Project

The pricing demonstration project (referred to as FasTrak) allows solo drivers to pay to use an eight-mile stretch of reversible high occupancy vehicle (HOV) lanes along Interstate Route 15 (I-15). The combination of free HOV use and priced solo driver use is generally referred to as high occupancy toll (HOT) lanes. The HOT Lanes are about eight miles long and are operated in the southbound (inbound to San Diego) direction for four hours in the morning and in the northbound (outbound) direction for four hours in the afternoon and evenings. The per-trip fee for solo drivers is posted on changeable message signs upstream from the entrance to the lanes, and may be adjusted every six minutes to maintain free-flowing traffic conditions in the HOT lanes. Solo drivers who subscribe to the FasTrak program are issued windshield-mounted transponders used for automatic vehicle identification. Each time they use the lanes, their accounts are automatically debited the per-trip fee. This is a dynamic form of voluntary congestion pricing, where solo drivers can choose to pay to reduce their travel time, and the payment is related to the level of congestion.

2.1 The Panel Survey

The panel survey consists of three samples of approximately equal size: 1) FasTrak program subscribers and former subscribers, 2) other I-15 users, and 3) a control group of users of another freeway corridor (I-8) in the San Diego Area. The analysis in this paper excludes the I-8 control group. The first wave of the panel was conducted prior to per-trip pricing. The second wave of the panel was conducted in spring 1998, during the first few months of dynamic pricing. For the purposes of this analysis, we focus primarily on program subscribers and other I-15 users in the third wave of panel data, collected during the fall of 1998 (October through November). During this time period, dynamic per-trip congestion pricing was well established.

FasTrak subscribers were picked at random from a list maintained by the billing agency, and the remaining respondents were recruited using random digit dialing (RDD) of residential areas along the respective corridors. In the initial wave of the panel, a partial quota sampling procedure was used to increase the number of carpoolers in non-subscriber parts of the sample. Panel attrition is about 33% per wave, and the sample is refreshed at each wave with a new random sample of FasTrak subscribers as well as I-15 and I-8 commuters recruited using RDD sampling. The partial quota sampling procedure implies that the resulting sample is choice-based and weights are needed to represent the population of regular I-15 corridor users. We estimated sampling weights from traffic counts carried out during the survey period.

Survey respondents were queried for detailed information about their most recent inbound trip along I-15 if that trip was made during the hours of operation of the HOT facility and covered the portion of I-15 corresponding to the facility. By design, trip lengths must be at least eight miles long (the length of the facility). There were 699 I-15 respondents with full information on morning peak-period inbound

trips, divided into three modes: 1) 304 solo drivers in the main lanes, 2) 279 solo drivers using FasTrak transponders to travel in the HOT facility, and 3) 116 carpoolers who also travel the HOT facility for free.

2.2 Dynamic Per-Trip Tolls

Solo drivers face tolls that are a function of arrival time at the HOT facility. The level of congestion in the HOT facility determines the toll (i.e. tolls increase to avoid exceeding preset capacity constraints). While program subscribers are provided with a profile of maximum tolls that vary by time-of-day, actual tolls may be less than the maximum tolls depending upon usage of the facility.

In October and November 1998 (excluding Thursday and Friday of Thanksgiving weekend), the actual maximum toll by time of day is flat at \$0.50 before 6:30, rising in an approximately linear fashion to \$4.00 over the 6:30 to 7:30 period. It stays at \$4.00 in the 7:30 to 8:30 period, then falls back down to \$1.00 by about 8:45 and \$0.75 by about 9:30. The average actual toll paid by the survey respondents who chose FasTrak varies by time of day in a similar manner from \$0.50 to a maximum of approximately \$3.50 in 7:45 to 8:00 period. Average tolls are remarkably similar across the days of the week. (Kazimi *et. al.*, 2000).

Based on the estimated arrival time at the HOT lanes, each survey respondent is assigned a toll price for that specific arrival time and date of travel. For respondents who choose to drive alone in the HOT lanes, this represents actual price paid. For solo drivers in the regular lanes and those who carpool, this represents the price they would have paid had they chosen to use FasTrak.

Arrival time at the HOT lanes is determined using a combination of information from the panel survey and speed estimates for the upstream portion of I-15. The panel survey queried respondents for onramp used in the morning commute and arrival time at that onramp. Travel time from the onramp to the beginning of the HOT lanes is estimated using time-of-day point speeds calculated from California Department of Transportation (CALTRANS) loop detectors embedded in the roadway. These loop detector data are computed every six minutes. Point speeds at loop detector locations are converted into speeds on intervening roadway segments using an algorithm that assumes that the point speed at the beginning of the segment applies to the first half of the segment and the point speed at the end applies to the second half of the segment (van Grol, 1997). Since loop detectors are placed near onramps, the freeway is effectively broken into segments traveling from onramp to onramp.

2.3 Time Savings From HOT Lane Use

For mode choice modeling, we must determine possible time saving from travel on the HOT lanes for all respondents regardless of mode choice. Time saving is defined as the difference in travel time on the HOT lanes and travel time on the parallel main lanes. Both are a function of when commuters arrive at the facility, speeds along the HOT lanes, and speeds in the main lanes. Speed on the HOT facility is assumed to be 70 miles per hour based on several days of floating car experiments. Speeds on the main lanes are estimated every six minutes during the entire survey period using the loop detector data. These speeds were also estimated by driving along the roadway every fifteen minutes for one week in the middle of the survey period (referred to as floating car measurements). Section 6 shows results using the loop detector speeds and using a combination of loop detector speeds and floating car speeds.

The median time saving, based solely on loop detector speed measurements by time of arrival at the HOT facility, peaks at about seven minutes at the same time period (7:30-8:00 AM) that average tolls peak at four dollars. Considerable variation occurs within each half-hour time period as indicated by the divergence between median, 90th percentile, and 10th percentile time savings. Ten percent of the time, peak time saving exceeds twelve minutes. Details are provided in Kazimi, *et al.*, (2000).

Those entering I-15 at one particular onramp (the Ted Williams Parkway onramp at the north end of the HOT Lanes) may also benefit from a special dedicated entrance to the HOT facility that avoids a congested main-lane onramp with a ramp-meter traffic signal. We estimated this additional time savings for each time interval from floating car observation of queuing times, and added it to the estimated time

savings from use of the HOT lanes for those respondents entering I-15 at this location (approximately 36 percent of the sample). These additional time savings ranged up to five minutes (Kazimi, *et al.*, 2000).

3 Mode Choice and Value of Time

The key ingredient in evaluating projects designed to reduce travel time is commuters' willingness to pay for these reductions. If commuters value time saved from congestion reduction highly, then it may be worthwhile to make costly investments in new transportation infrastructure. This section reviews the model structure and estimation methods that transportation economists use to estimate value of time (VOT) from reducing travel delays.

3.1 Conditional Logit Mode Choice Models

Suppose that respondent n faces a choice of three modes for travel to work indexed by j . In this paper the modes are drive alone, pay to drive alone in the HOT lanes (FasTrak), or carpool in the HOT lanes. In most previous studies the modes are automobile, bus, or subway. The Conditional Logit model assumes that the probability that respondent n takes mode j conditional on observed variables x_{jn} is given by:

$$P_{jn} = \frac{\exp(\mathbf{q}x_{jn})}{\sum_{i=1}^3 \exp(\mathbf{q}x_{in})} \quad (1)$$

The value of time saved (VOT) is given by the increase in cost required to keep P_{jn} constant after a small decrease in travel time. If time and cost only enter as linear terms in x , then the VOT is just given by $\mathbf{q}_{time}/\mathbf{q}_{cost}$.

Small (1992) and Wardman (1998) provide comprehensive reviews of VOT studies, and Gonzalez (1997) provides a review of the theory of consumer choice and its connection to value of time and mode choice modeling. Based on his review, Small (1992) suggests that 50 percent of gross wage rate is a reasonable value of time estimate. On the higher end of previous studies, Cambridge Systematics (1977) estimate that VOT for commuters in Los Angeles is 72 per cent of gross hourly wage. These previous studies are based upon mode choice models that consider differences between transit and automobile travel, and to the extent that differences between transit and private automobiles are not captured, the results will be biased. In more recent work, Calfee and Winston (1998) attempt to avoid this problem by using stated preference data that only considers the tradeoff between travel by automobile in slower, free lanes and travel by automobile in faster, priced lanes. Their results indicate that commuters have a lower VOT than previously estimated (roughly \$3.50 to \$5.00 per hour or 15 to 25 percent of hourly wage). Calfee and Winston rely upon stated preference data because they lack revealed preference data for the choices involved with congestion pricing. Our results are not subject to the same potential biases associated with stated preference data as we use revealed preference data.

Given a random sample of N commuters, the model in equation (1) is typically estimated by maximizing the likelihood function

$$L = \sum_{n=1}^N \sum_{i=1}^3 D_{in} \log(P_{in}), \quad (2)$$

where $D_{in}=1$ if respondent n chooses mode i and zero otherwise. This likelihood function is globally concave and therefore easy to maximize using standard algorithms. See Train (1986) for more information about this model and its application to transportation problems.

3.2 Choice-base Sampling

It is very common for one mode to have a very low market share, which makes collecting a random sample with a reasonable sample size for each mode very expensive. For example, in the I-15 corridor

the FasTrak users account for only 3.5 percent of the inbound peak period trips. To reduce data collection costs most transportation surveys stratify on mode choice, which results in a non-ignorable sampling scheme.

Maximizing a random-sample likelihood function as in equation (2) with a choice-based sample will generally yield inconsistent parameter estimates. McFadden (see proof in Manski and Lerman, 1977) shows that for the conditional logit model with a full set of mode-specific constants only the parameters associated with these mode-specific constants are inconsistent. Scott and Wild (1986) provide similar results and give links to case-control sampling schemes. These results imply that we can use unweighted maximum likelihood for our conditional logit model. However, it is useful to consider alternative estimators that are consistent for more general choice models such as the Nested Logit Model (see Train, 1986).

A relatively simple estimator which yields consistent estimates under choice-based sampling was developed by Manski and Lerman (1977). Their Weighted Exogenous Sample Maximum Likelihood Estimator (WESMLE) is the maximand of the weighted likelihood function:

$$\sum_n \mathbf{w}_n L_n(\mathbf{q}, x_n), \quad (3)$$

where L_n is the log likelihood function for the n^{th} observation and the sampling weight, \mathbf{w}_n , is the inverse of the probability that the n^{th} observation (individual) would be chosen from a completely random sample of the population. This estimator is also known as the “pseudo maximum likelihood estimator in the survey sampling literature (Skinner, 1989). If the sampling scheme were completely random, then all of the sampling weights would be equal and the WESMLE would simply be the usual maximum likelihood estimator. The WESMLE is inefficient, but Imbens (1992) gives an efficient method of moments estimator for choice-based samples.

4 Measurement Model

The loop detector data described in Section 2.3 can give inaccurate estimates of the actual time savings commuters get from taking the HOT lanes. Depending on the traffic flows between the loop detectors (which are miles apart on the I-15 corridor), actual speeds can be either over or under-predicted. Since these measurement errors will generally be larger when the road is congested, the measurement errors in time savings are likely to be larger for FasTrak and carpool lane users. Since time saved using the HOT lanes is a key independent variable in the choice models in Section 6, this measurement error will bias key parameter estimates.

We use the five days during the survey period where we have both floating car and loop detector data available to fit a model which we use to predict floating car travel time for the other seven weeks of the survey period. These predicted floating car data are then used to fit mode choice models in Section 6. This approach assumes that the floating car data are correct, and we will use multiple imputations to correct for the measurement error caused by imperfect predictions.

The floating car data are collected at 15 minute intervals while the loop detector data are at 6 minute intervals. To make these data compatible, we interpolated the floating car data into 6 minute intervals. The floating car estimates over the morning commutes from October 26 through October 30, 1998 are generally more than twice as large as the loop detector time savings. The median floating car time savings is 8.5 minutes, while the median loop detector time savings is 2.2 minutes. Obviously the loop detector estimates are badly biased for this corridor.

Table 1 shows the best fitting linear regression model for predicting floating car HOT lane time savings. To avoid unreasonable predictions we first transform both time savings measures to keep them bounded between zero and 35 minutes, which is the maximum observed loop detector time savings. The exact transformation for both time savings variables is given by the following transformed logit:

$$\log\left(\left(\frac{t}{35}\right)/\left(1-\frac{t}{35}\right)\right) \quad (4)$$

We tried a number of different specifications including higher order terms in loop detector time savings and toll variables, but none of them significantly improved the fit of the model. Since the purpose of this model is accurate prediction, we are looking for the most parsimonious model with the best fit. Although including the main effect is traditional when including interactions in a regression model, including the logit of loop detector time savings results in a coefficient of .06 with a standard error of .22. Since no other coefficients were changed, we deleted the main effect to avoid inflating the variance of the model's predictions. We also experimented with lagged values, but the cubic polynomial in time effectively removes the autocorrelation in the time savings measures (residual first-order autocorrelation is .08).

Although the variables involving the tolls are not individually significant, they are jointly significantly different from zero at the one percent level. If they are excluded from the model, then the R^2 drops slightly to .89. However, excluding the loop detector data reduces the R^2 to .82 and increases the MSE of the residuals to .46.

There are two general approaches for estimating a behavioral model with measurement error in the explanatory variables: joint maximum likelihood of the behavioral and measurement models, or Rubin's multiple imputation approach. Joint maximum likelihood would be very difficult for the model in Section 6 since the actual explanatory variables are complicated non-differentiable transformations of the variable explained by the measurement model in Table 1. We will therefore implement the multiple imputation approach as given in Rubin (1987 and 1996). Brownstone (1998) gives more detail using the same notation as this section. Rubin developed his methodology for missing data, and in our application floating car time savings are missing for approximately 80 percent of our respondents.

Suppose we are interested in estimating an unknown parameter vector θ . If no data are missing, then we would use the estimator $\tilde{\theta}$ and its associated covariance estimator $\tilde{\Omega}$. If we have a model for predicting the missing values conditional on all observed data, then we can use this model to make independent simulated draws for the missing data. If m independent sets of missing data are drawn and m corresponding parameter and covariance estimators, $\tilde{\theta}_j$ and $\tilde{\Omega}_j$, are computed for each of these imputed data sets, then Rubin's Multiple imputation estimators are given by :

$$\hat{\theta} = \sum_{j=1}^m \tilde{\theta}_j / m \quad (5)$$

$$\hat{\Sigma} = U + (1 + m^{-1})B, \text{ where} \quad (6)$$

$$B = \sum_{j=1}^m (\tilde{\theta}_j - \hat{\theta})(\tilde{\theta}_j - \hat{\theta})' / (m - 1) \quad (7)$$

$$U = \sum_{j=1}^m \tilde{\Omega}_j / m. \quad (8)$$

Note that B is an estimate of the covariance among the m parameter estimates for each independent simulated draw for the missing data, and U is an estimate of the covariance of the estimated parameters given a particular draw. B can also be interpreted as a measure of the covariance caused by the nonresponse (or measurement error) process.

Table 1. Imputation Model for Floating Car HOT Lane Time Savings

Dependent Variable: Logit of Floating Car Time Savings	R ² = 0.90 Root MSE = 0.36		
Independent Variables:	Coef.	Std. Err.	t-Stat.
Logit of Loop Detector Time Savings × Minutes Past 5:00 A.M.	0.0029	0.00031	9.3
Minutes Past 5:00 A.M.	0.222	0.0149	14.8
(Minutes Past 5:00 A.M.) ²	-0.00138	0.000121	-11.4
(Minutes Past 5:00 A.M.) ³	2.73E-06	2.91E-07	9.38
Toll	-0.229	0.188	-1.22
Toll × Minutes Past 5:00 A.M.	0.00222	0.00126	1.77
Constant	-11.4	0.52	-22.1

Rubin (1987) shows that for a fixed number of draws, $m \geq 2$, $\hat{\theta}$ is a consistent estimator for θ and $\hat{\Sigma}$ is a consistent estimator of the covariance of $\hat{\theta}$. Of course B will be better estimated if the number of draws is large, and the factor $(1 + m^{-1})$ in equation (6) compensates for the effects of small m . Rubin (1987) shows that as m gets large, then the Wald test statistic for the null hypothesis that $\theta = \theta^0$,

$$(\theta - \theta^0)' \hat{\Sigma}^{-1} (\theta - \theta^0), \quad (9)$$

is asymptotically distributed according to an F distribution with K (the number of elements in θ) and n degrees of freedom. The value of n is given by:

$$\begin{aligned} n &= (m - 1)(1 + r_m^{-1})^2 \text{ and} \\ r_m &= (1 + m^{-1}) \text{Trace}(BU^{-1})/K \end{aligned} \quad (10)$$

This suggests increasing m until n is large enough (e.g. 100) so that the standard asymptotic Chi-squared distribution of Wald test statistics applies. We used this stopping rule and found that the models in Section 6.2 required $m=20$ multiple imputations. Although this is more than the 4-5 multiple imputations used in most applications, recall that the proportion of missing floating car data is 80 percent in our application. Meng and Rubin (1992) show how to perform likelihood ratio tests with multiply-imputed data. Their procedures are useful in high-dimensional problems where it may be impractical to compute and store the complete covariance matrices required for the Wald test statistic (equation 9).

To draw one set of imputed values for the missing floating car data, first draw one set of slope and residual variance parameters from the asymptotic distribution of the linear regression estimators from Table 1. The slope parameters are drawn from the joint normal distribution centered at the parameter estimates with covariance given by the usual least squares formula $(s^2(X'X)^{-1})$. The residual variance, \mathbf{S}_*^2 , is drawn by dividing the residual sum of squares by a draw from an independent \mathbf{C}_d^2 distribution, where d is the residual degrees of freedom. An imputed residual vector is then drawn from independent normal distributions with mean zero and variance equal to \mathbf{S}_*^2 . The imputed values are then computed by adding this imputed residual to the predicted value from the regression using the imputed slope parameters. Additional sets of imputed values are drawn the same way beginning with independent draws of the slope and residual variance parameters. Observations where floating car data are observed are fixed at these observed values across all imputations. This imputation method, which Schenker and Welsh (1988) call the ‘‘full normal imputation’’ procedure, is equivalent to drawing from the Bayesian predictive posterior distribution when the dependent variable and the regressors follow a joint normal distribution with standard uninformative priors.

For each imputed value we add the mean time savings for those respondents entering the I-15 at Ted Williams Parkway. The medians and 90th percentiles across each month are computed for each 6-

minute time interval. These medians and the difference between the 90th percentiles and the medians are then used to estimate the parameters of the choice model in Section 6.2. The multiple imputation procedure described here has been implemented in STATA, and it could be programmed in most modern statistical packages.

5 Attrition Model

The 39% attrition rate between Waves 2 and 3 of our panel is not unusual for transportation panel surveys (Raimond and Hensher, 1997). The high attrition might be due to the required detailed questions about the commute trip which respondents find difficult to answer and/or intrusive. Although new respondents (the refreshment sample) are recruited each wave to maintain sample size, it is crucial to account for attrition when analyzing these data. Once the data are collected there is nothing to be done about the loss of efficiency due to the decreased sample size, but there are flexible modeling techniques to identify and correct for non-ignorable attrition.

The simplest approach is to compare the panel sample with the refreshment sample. There do not appear to be striking differences in the distribution of key variables across these samples, but the panel sample exhibits slightly higher income and longer commute distance. Since the samples are approximately equal size, it is also possible to fit the choice model in Section 6.1 separately for each sample. The hypothesis that attrition is ignorable is then equivalent to the hypothesis that the coefficients of the choice model are equal across the samples. A standard likelihood ratio test shows that this hypothesis cannot be rejected at any reasonable significance level for these data.

If there is no reasonable size refreshment sample, or if the data are used for dynamic analysis, then the attrition process can be modeled using the initial wave of the panel. The results from fitting a binomial logit attrition model show that the only significant predictors of attrition are refusal to disclose income, distance, and proportion of FasTrak use during the previous week. Commute distance enters as a quadratic term that has a maximum negative effect on attrition at 42 miles. This implies that for the relevant range of the data longer distance commuters are less likely to attrite. Proportion of FasTrak use is an endogenous variable in our choice models, so its significance in the attrition model implies that the attrition process is non-ignorable. The higher attrition of FasTrak users might be related to the substantial number of additional survey questions administered to this group.

Unless there are significant interactions between the dependent variable and other independent variables, the attrition process described above is just another form of choice-based sampling. Therefore unweighted maximum likelihood estimates of the conditional logit model will be consistent except for the alternative-specific constants. In our application, there are no significant interactions in the attrition model, so we will base our estimates in Section 6 on unweighted estimates. If there are significant interactions, then Brownstone (1998) and Brownstone and Chu (1997) show that the WESMLE estimator can be used with multiply imputed weights from the attrition model to get consistent inference.

6 Choice Model Results

Sections 6.1 and 6.2 compare mode choice model estimates using uncorrected loop detector data and correcting for measurement error. We use a model derived from the specification in Kazimi *et al.* (2000). The main difference in the specifications is that here we include a variable identifying sample respondents who do not pay their own tolls. Any teenager knows that if someone else is paying (here, typically the employer), then they will be less sensitive to the price.

In addition to the parameter estimates, we also report value of time (VOT) estimates for the models in Sections 6.1 and 6.2. Since toll enters the specification both linearly and interacted with variability (the difference between the 90th percentile and the median of time saved by taking the HOT lane over the month), the VOT in dollars per hour saved is given by:

$$\frac{(60 \times \mathbf{q}_{timesavings})}{(\mathbf{q}_{toll} + \mathbf{q}_{toll} * \mathbf{Variability} \times \mathbf{Variability})} \cdot \quad (11)$$

Since VOT varies across respondents, we give the distribution across respondents weighted by the choice-base sampling weights to match the population of morning commuters. We also give this VOT evaluated at the weighted mean of Variability. This latter quantity is useful for comparison with other studies that typically do not report the variable in equation (11). Our definition of variability is based on the notion that commuters are much more concerned about unexpected delays than about unexpected speedy trips.

6.1 Loop Detector Time Savings

The left panel of Table 2 gives parameter estimates for the mode choice model using loop detector time savings. High-income, home-owning, middle-aged females with a graduate degree are the most likely group to pay for FasTrak. Large households with more workers than cars are most likely to carpool. Both carpoolers and FasTrak users have similar positive coefficients for time savings, but the reduction in Variability from HOT lane use is not significant. However, if Variability is removed from the model then the toll coefficient drops and becomes insignificant. Relative to solo driving, commute trip drivers are more likely to choose FasTrak and non-commute trip drivers are more likely to carpool.

The middle column of Table 3 gives various VOT estimates (computed from equation 11) from the model using loop detector data. Note that the distribution is skewed and there is substantial variance across the population. The median values are much higher than Calfee and Winston's (1998) estimates, and they are on the high end of the estimates reviewed in Small (1992). These medians are similar to equation (11) evaluated at the weighted sample mean variability (labeled "VOT at Mean Variability"). This is the number typically presented in studies where VOT varies according to observed variables. Since this is just a scalar, it is straightforward to estimate the standard error of this estimate (caused by parameter estimation error) using the delta method. Although this estimate is significantly different from zero, the standard error is large enough to include almost all previous estimates. Calfee and Winston do not report standard errors for their VOT estimate of \$5.00, but the \$26/hour estimate in Table 3 is more than two standard errors away from their point estimate.

6.2 Predicted Floating Car Time Savings

The right panel of Table 2 gives the results of estimating the choice model using the predicted floating car data and multiple imputation algorithm described in Section 4. The coefficient estimates are roughly similar to the uncorrected loop detector estimates, but the key coefficients of toll and time savings for commuters are reduced in magnitude and significance. Overall the standard errors are considerably larger than the uncorrected loop detector estimates. This is due to the component of error caused by the error in the prediction model used to generate the predictions.

Since the floating car time savings are generally larger than the corresponding loop detector measures, we would expect that the value of time estimates would drop relative to the uncorrected loop detector estimates. The third column of Table 3 confirms this and shows that the VOT estimates have dropped \$5 - \$7. While this change is quite significant from a policy perspective, it is not statistically significant given the large standard errors of these measures.

If the error in the prediction model is ignored and only one set of imputed floating car time savings is used, then the standard errors are downward biased by over 50 percent for this model. Even though the prediction model fits very well, the prediction error is still an important component of the total estimation error.

7 Conclusion

This paper reviews techniques for handling attrition, choice-based sampling, and measurement error in panel surveys. Although we concentrate on commuter surveys and value of time measurement, the techniques are general and can be applied in other settings. It turns out that only measurement error is a

serious problem in our application, although there is no way to know this without first carefully modeling the attrition and sampling process.

Section 4 shows that measurement error in travel time is a serious problem for mode-choice models. The relatively cheap measures, loop detectors and respondents' perceptions of time savings, are both badly biased. When we collect additional data on all respondents' perceptions, then we can add these perceptions to our imputation models. In any case the multiple imputations approach used here to integrate the measurement error and choice models is a good general tool for these sorts of problems. Ignoring the component of error in the choice model parameters caused by the prediction model leads to serious underestimates of the precision of the choice model parameters.

Table 2. Conditional Logit Mode Choice Model Estimates

Number of obs. = 699	Loop Detector Data			Corrected Data		
	Pseudo R ² = 0.21 Log likelihood = -606.56			Pseudo R ² = 0.20 Log likelihood = -611.27		
FasTrak choice	Coef.	Std. Err.	t-Stat.	Coef.	Std. Err.	t-Stat.
Constant	-5.978	1.994	-3.00	-7.179	3.342	-2.15
Income ≥ \$100K + Refused to answer*	0.855	0.183	4.68	0.830	0.271	3.06
Income < \$40K*	-0.621	0.505	-1.23	-0.591	0.536	-1.10
Female*	0.730	0.183	3.98	0.704	0.251	2.81
Age between 35 & 45*	0.423	0.179	2.36	0.445	0.210	2.12
Has Graduate Degree*	0.741	0.195	3.80	0.747	0.266	2.81
Household owns home*	0.754	0.293	2.57	0.812	0.355	2.29
Distance (miles)	0.019	0.010	1.86	0.015	0.011	1.39
Toll paid by someone else*	1.747	0.454	3.85	1.816	0.633	2.87
Toll (\$/trip)	-0.787	0.220	-3.58	-0.600	0.387	-1.55
Median total time savings for commuters	0.182	0.047	3.87	0.074	0.037	2.04
Median total time savings for non-commuters	0.417	0.216	1.93	0.297	0.200	1.49
Toll × Variability	0.135	0.035	3.83	0.090	0.053	1.69
Commute trip*	3.395	1.939	1.75	4.495	3.004	1.50
Carpool Choice						
Constant	-2.265	1.006	-2.25	-2.139	1.145	-1.87
Workers per vehicle	1.005	0.366	2.74	0.982	0.435	2.26
Distance (miles)	0.102	0.056	1.82	0.099	0.060	1.64
Distance squared	-0.001	0.001	-1.27	-0.001	0.001	-1.23
Single worker household*	-0.973	0.350	-2.78	-1.005	0.426	-2.36
Two worker household*	-0.522	0.289	-1.81	-0.548	0.318	-1.72
Commute trip*	-1.762	0.414	-4.25	-1.747	0.588	-2.97
Median total time savings	0.144	0.045	3.19	0.056	0.033	1.71
Carpool ramp bypass*	0.556	0.278	2.00	0.634	0.315	2.01
Variability of solo drive time	0.098	0.076	1.29	0.039	0.076	0.51

* These are dummy variables defined to equal one if the condition is true and zero otherwise.

Table 3. Value of Time Saved Estimates

Value of Time (VOT) (\$/hour)	Loop Detector	Corrected
90 th Percentile	73.63	72.12
50 th Percentile	23.37	18.71
10 th Percentile	14.43	-20.72
Mean	32.64	25.63
Std. Dev	94.29	74.75
VOT at Mean Variability	25.96	18.63
Std. Dev. Of VOT at Mean Variability	7.70	13.88

The substantive conclusions from the models in Section 6 are largely negative. We cannot estimate value of travel time reduction accurately enough to resolve current controversies. In particular, the confidence bands from our estimates cover most existing estimates, and the differences between these estimates are important for planning new transportation infrastructure investments. Additional work is required to combine perceived time savings, loop detector time savings, and floating car time savings using data from more recent waves of the I-15 panel. Stated preference questions have also been added to the survey so that we can jointly model responses to hypothetical and real situations. Hopefully the enhanced models will shed more light on the problem of evaluating time savings.

8 Acknowledgements

John Eltinge, Rod Little, Don Rubin, and Doug Wissoker provided many useful comments on an earlier draft. We would like to acknowledge financial support from the U.S. Department of Transportation and the California Department of Transportation through the University of California Transportation Center. Arindam Ghosh provided excellent research assistance, and Dirk van Amelsfort calculated the loop detector time savings. Additional thanks go to Jackie Golob of Jacqueline Golob Associates, Kim Kawada of San Diego Association of Governments (SANDAG), and Kathy Happersett of the Social Science Research Laboratory of San Diego State University. None of these people or agencies is responsible for any errors or omissions.

9 References

- Biemer, P.P., Groves, R.M., Lyberg, L., Mathiowetz, N.A. and Sudman, S. (1991). *Measurement Errors in Surveys*. New York:Wiley.
- Brownstone, D. (1998). Multiple Imputation Methodology for Missing Data, Non-Random Response, and Panel Attrition. In *Theoretical Foundations of Travel Choice Modeling*, eds. T. Gärling, T. Laitila and K. Westin, Amsterdam: Elsevier, 421-450.
- Brownstone, D. and Chu, X. (1997). Multiply-imputed sampling weights for consistent inference with panel attrition. In *Panels for Transportation Planning*, eds. T.F. Golob, R. Kitamura and L. Long, Boston: Kluwer Academic Publishers, 261-273.
- Calfee, J. and Winston, C. (1998). The Value of Automobile Travel Time: Implications for Congestion Policy. *Journal of Public Economics*, 69, 83-102.
- Cambridge Systematics, Inc. (1977). *The Development of a Disaggregate Behavioral Work Mode Choice Model*. Prepared for California Department of Transportation and Southern California Association of Governments, Cambridge, MA.
- Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995). *Measurement Error in Nonlinear Models*. London:Chapman and Hall.

- Fuller, W.A. (1987). *Measurement Error Models*. New York:Wiley.
- Gonzalez, R. M. (1997). The Value of Time: A Theoretical Review. *Transport Reviews*, 17, 245-266.
- Groves, R.M. (1989). *Survey Errors and Survey Costs*. New York:Wiley.
- Imbens, G. (1992). An Efficient Method of Moments Estimator for Discrete Choice Models With Choice-Based Sampling. *Econometrica*, 60, 1187-1214.
- Kazimi, C., Brownstone, D., Ghosh, A., Golob, T. F., and van Amelsfort, D. (2000). Willingness-to-Pay to Reduce Commute Time and Its Variance: Evidence from the San Diego I-15 Congestion Pricing Project. Presented at Annual Meeting of Transportation Research Board, National Research Council, January 9-13, Washington, DC.
- Lessler, J.T. and Kalsbeek, W.D. (1992). *Nonsampling Error in Surveys*. New York:Wiley.
- Manski, C.F. and Lerman, S. (1977). The estimation of choice probabilities from choice-based samples. *Econometrica*, 45, 1977-1988.
- Meng, X-l. and Rubin, D.B. (1992). Performing Likelihood-Ratio Tests with Multiply-Imputed Data Sets. *Biometrika*, 79, 103-111.
- Raimond, T. and Hensher, D.A. (1997). A Review Of Empirical Studies And Applications. In *Panels for Transportation Planning*, eds. T.F. Golob, R. Kitamura and L. Long, 15-72. Boston: Kluwer Academic Publishers.
- Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley.
- Rubin, D.B. (1996). Multiple Imputation After 18+ Years. *Journal of the American Statistical Association*, 91, 473-489.
- Schenker, N. and Welsh, A.H. (1988). Asymptotic Results for Multiple Imputation. *Annals of Statistics*, 16, 1550-1566.
- Scott, A. J. and C. J. Wild (1986). Fitting Logistic Models under Case-control or Choice Based Sampling. *Journal of the Royal Statistical Society B*, 48(2), 170-182.
- Skinner, C. J. (1989). Domain Means, Regression and Multivariate Analysis. In *Analysis of Complex Surveys*, eds. C. J. Skinner, D. Holt, and T. M. F. Smith, 59-88. New York: John Wiley and Sons.
- Small, K. (1992). *Urban Transportation Economics*. Switzerland: Harwood Academic Publishers.
- Train, K. (1986). *Qualitative Choice Analysis: Theory, Econometrics, And An Application To Automobile Demand*. Cambridge, MA:MIT Press.
- Van Grol, H.J.M. (1997). Evaluating the Use Of Induction Loops For Travel Time Estimation. Presented at 8th IFAC/IFIP/IFORS Symposium on Transportation Systems, Chania, Greece, June 16-18, 1997.
- Wardman, M. (1998). The Value Of Travel Time: A Review Of British Evidence. *Journal of Transportation Economics and Policy*, 32, 285-316.