The Urban Transport Institute

TUTI Report 7-2001

A Simulation Study of the Estimation of Individual Specific Values of Time using an Adaptive Stated Preference Survey

A paper submitted for presentation at the
81st Annual Meeting of the Transportation Research Board

A.J. Richardson

July 2001

Abstract

Stated Preference surveys provide a good opportunity for obtaining information for the construction of choice models, but can run the risk of asking questions of respondents which are too complex to answer meaningfully, especially when many attributes and alternatives are considered in each question. An alternative technique is the use of Adaptive Conjoint Analysis or Adaptive Stated Preference (ASP) surveys, wherein simpler questions are used, each game is dependent on the answers given to previous games, and parameters are estimated at the level of the individual. This paper describes a simulation study to test the ability of ASP to produce unbiased estimates of the mean and standard deviation of the Value of Time (VOT). It is found that this is possible under certain conditions, namely that a 7-point rating scale is used to obtain responses, each respondent plays three games for each parameter to be estimated, the starting estimate of VOT used should be higher than the expected final mean value of VOT, and a logit model sensitivity parameter of c=5 is used in the binary logit model used to convert probabilities of choice into utility differences. The results of this simulation give reassurance that, in addition to being easier for respondents, the ASP survey method also produces unbiased estimates of the distribution of VOT (or any other parameter under consideration). Given the extreme flexibility of the disaggregate results produced by the ASP survey method, the demonstration of unbiased estimates is an important confirmation of the usefulness of this method.

Introduction

The rapid development of choice modelling procedures, especially using Stated Preference data, over the past twenty years, has created an opportunity for an imbalance between the models and the data used for construction of the models. Richardson and Ampt (1) have examined this imbalance and stressed the need for more attention to be paid to the needs of the respondent when collecting data for choice modelling. Drawing on the classic ideas of Alonso (2) on the sources of error in modelling studies, that paper highlighted the fact that the sophistication of the models and the data need to be commensurate. The paper gave some contrasting examples of studies in which sophisticated analysis has not been supported by equally sophisticated data collection, compared to studies where the emphasis has been placed on respondent-friendly designs without compromising the analytical outputs of the study.

The contrasting study used in that paper was the Singapore Value of Time Study, conducted for the Singapore Land Transport Authority. In this study, relatively simple methods of analysis are used, but with more attention paid to the needs of the respondents.

The Singapore VOT study had as its main objective the measurement of relative valuations of importance of various attributes to travellers. It did not seek to build a predictive model of choice or usage, but merely to measure the relative value of parameters to users (in this case, the value of travel time differences relative to cost differences). For users of public modes of transport, it was desired to obtain values of time for in-vehicle time, waiting time, walking time and transfers between vehicles. For users of private modes, it was desired to obtain a value of time for door-to-door travel time. It was also desired to obtain perceived values of various costs involved for private modes, such as fuel, parking, Electronic Road Pricing (ERP) and the Certificate of Entitlement (CoE) which must be paid in order to initially obtain a vehicle in Singapore.

Given the nature of the study, a stated preference survey was the logical candidate. However, a prime concern was to design a Stated Preference (SP) survey which would be respondent-friendly, and which would not overtax the cognitive capabilities of the respondents. Given the manner in which the results were to be subsequently analysed, involving the estimation of Values of Time for many different market segments, it was also desirable if the results could be obtain at a highly disaggregated level, i.e. at the level of the individual. Conventional SP surveys do not adequately fulfil either of these requirements. For this reason, the chosen method was a variation on a technique used widely in market research called Adaptive Conjoint Analysis (ACA). To distinguish the specific method used in the Singapore Value of Time Study, we refer to it as Adaptive Stated Preference (ASP).

Adaptive Stated Preference and Parameter Estimation

The ideas of ASP and the use of simplified questions are not entirely new to transport. Polak and Jones (3) describe adaptive SP surveys as “very appealing intuitively” and state that they “have become popular among practitioners, not least because they tend to reduce the amount of questioning that needs to be carried out and because respondents typically report adaptive exercises to be easier and more interesting to complete”. Wang, Oppewal and Timmermans (4) also report on the use of a simplified SP design which, while not being adaptive as in the Singapore survey, varies only two attributes at a time in an attempt to avoid the “information overload problem typical of the conventional fractional factorial design strategy”.

While recognising the benefits of ASP to the respondents, Polak and Jones (3) also note that there are some potential disadvantages. First, they note that ASP will result in non-orthogonal data, which reduces a traditional advantage of SP data (although Watson et al., (5) note that non-orthogonality can be a positive advantage for model estimation). Second, they note that poorly designed ASP surveys can “run out of control”, especially in the case of respondents with extreme preferences. In the Singapore ASP survey, this was accounted for by bounding the values of time estimated in each game so that the next game would present reasonable attribute values to the respondent (this restriction was removed in the later analysis of the data obtained from the ASP survey). Third, they note the reservations expressed by Bradley and Daly (6) concerning correlations being introduced between the levels of the design variables and the unmeasured components of utility that are not accounted for in model estimation. This presumably is also the basis for Wardman's (7) description of ASP designs as being "the now generally discredited adaptive approach", where they refer to Bradley and Daly's paper (6).

This apparently fatal flaw of adaptive designs, however, is very much dependent on the method of analysis used with the data. Bradley and Daly (6) analyse the data using relatively conventional choice modelling methods, estimated with the ALOGIT package. The non-orthogonality and correlations introduced by adaptive designs do indeed introduce bias and variability when performing such model estimations. This problem was avoided in the Singapore study by not estimating models for groups, but rather estimating individual specific parameters. Indeed, Bradley and Daly (6) had recognised this as an alternative approach, but appear to have discarded the method on the basis of some earlier work by Morikawa (8) in which he suggested that the bias and variability in individual specific estimates was much higher than that obtained from a conventional modelling process with group data. However, this conclusion by Morikawa appears to have been based on an incorrect procedure whereby the two methods were not compared on a like-for-like basis. Such an observation is supported by Kroes and Cirillo (9), who conclude that grouped models are not necessarily superior to individual specific models.

The Singapore ASP survey therefore estimated individual specific values of time (and other parameters). In doing so, it acknowledged the warning given by Fowkes (10) and Fowkes and Wardman (11) that there is a problem with estimating Values of Time in the presence of taste variation, since conventional model estimation estimates a ratio of averages (the average time coefficient divided by the average cost coefficient) when what is really needed is the average of the ratios of time coefficient and cost coefficient for each individual.

However, given the previously-expressed reservations about ASP designs, the bias and variability characteristics of individual specific parameters using ASP were tested extensively before the conduct of the Singapore survey using simulation techniques. The results of these tests are the subject of this paper.

The Nature of Adaptive Stated Preference Surveys

Adaptive Stated Preference surveys differ from conventional Stated Preference surveys in four major ways:

- The options presented to the respondent in an ASP game depend on the responses given by the respondent to previous games
- The individual ASP games have fewer options and fewer attributes presented to the respondent than in conventional SP games

- The respondent is often presented with more games in an ASP survey than in a conventional SP survey
- At the end of an ASP survey, it is possible to obtain estimates of the parameters of interest (such as Value of Time) for that individual. Conventional SP surveys require that data from many respondents be aggregated and the parameters estimated as the result of a model-fitting process.

ASP games start with an initial estimate of the parameters of interest and then present the respondent with ASP games that progressively refine the estimates of the parameters of interest. The ASP games are designed to deliberately present the respondent with games which force them to trade-off increases in one attribute against reductions in another attribute.

Each respondent answered 18 ASP questions in the Singapore study. However, each of the questions was fairly simple and easily understood, being only a trade-off between two attributes across two options. A typical screen from the computerised interview is shown in Figure 1, for a trade-off of total travel time against fare for a public transport trip.

After the respondent answered such a question, the survey program updated the estimate of this respondent's value of time (using a simple binary logit formulation) and generated the attribute levels for the next question in an attempt to refine the estimate.

Even though the ASP questions are easy for the respondent to complete, they generate highly detailed and disaggregate data on Values of Time (at the level of the individual respondent). These Values of Time are the same as any variable in a travel survey data set, and can be expanded to population totals using Census data and other external data sets. They can then be analysed using readily available computer programs, such as Excel or any statistical software.

Simulation Testing of ASP Parameter Estimates

While the type of disaggregate results obtainable from the ASP survey look very useful and informative, a basic question is whether the estimated parameters are indeed reflecting the underlying Values of Time of the population being surveyed. To test the ability of the ASP design to accurately estimate Values of Time for a population, this paper describes a simulation study which investigated the ability of the ASP design to reproduce the parameters of an underlying distribution of the Value of Time used to create a simulated population.

The structure of the simulation

The simulation model consisted of simulating a population of 1000 respondents, each with a value of time (VOT) sampled from a log-normal distribution of values of time. A log-normal distribution means that all VOT are positive. This is not a necessary condition, but is useful as a starting point for the simulation. Further simulations can relax this assumption by using a shifted log-normal distribution, a normal distribution or any other distribution which can produce positive and negative VOT. Each simulated respondent then played up to 10 ASP games of the type used in the Singapore study. At the start of the first game, a VOT is assumed as a starting point. Using the time and cost recorded for the sampled trip as a base (Option A), a variation in travel time is assumed (say 5 minutes longer) and a cost is calculated to generate an Option B so as to give a trade-off game which will produce an expected outcome in terms of the score given by the respondent on a rating scale (from strongly preferring Option A through to strongly preferring Option B). The expected score is varied on either side of the mid-point score. However, because the respondent does not necessarily have the same VOT as the initially assumed VOT, they will consider the game and give a score that is perhaps different from that expected in the design of the game. The score they give can then be used to update the estimate of the VOT for that respondent. This new VOT is then used in the design of the second game, which then gives rise to a new estimate of the VOT for that respondent. The process is repeated for as many games as the respondent plays.

Major assumptions in the simulation

There are several assumptions in the simulation model that can be varied to test the sensitivity of the ASP survey technique to various assumptions made in the games. The major assumptions include:

Shape, mean and standard deviation of the distribution of Value of Time

The Value of Time is assumed to be log-normally distributed, as shown in Figure 2. The mean and standard deviation of the distribution can be varied at will. In addition, the functional form of the distribution can also be changed as required.

Characteristics of the selected trip

The characteristics of the base trip used as Option A in the first game are taken as the characteristics of the respondent’s sampled trip in the real ASP survey. In the simulation, a typical public transport trip of 35 minutes with a one-way fare of $1.50 is used. Since the simulation assumes, as with nearly all choice models, that the VOT is independent of the length of the trip, then it is unnecessary to vary the characteristics of the trip. In the real ASP surveys, it can be investigated empirically whether the VOT really is independent of the length of the trip.

Size of the change in travel time

In the simulation, a travel time change of 5 minutes either way from the base trip travel time is used to generate the trade-off games. Since the simulation assumes, as with nearly all choice models, that the VOT is independent of the size or direction of the change in travel time, then it is unnecessary to vary the size or direction of the change in travel time. In the real ASP surveys, it can be investigated empirically whether the VOT really is independent of the size and direction of the change in travel time, by using different size changes in travel time in both directions.

The number of points on the rating scale

The ASP survey currently uses a 7-point rating scale to obtain the degree of preference of the respondent for one option over the other. However, the simulation allows for the testing of a 5-point scale and a 9-point scale. More or less points on the scale are seen to offer less resolution or to be more complex for respondents, respectively.

The conversion of rating scale scores to probabilities of choice

Respondents indicate their degree of preference for option by selecting a score on the rating scale described above. In order to use this score in the construction of a simple logit model of choice, the rating scale score needs to be converted into a measure of probability of choosing either of the options. It is assumed that a higher score indicates a lower probability of choosing Option A and a higher probability of choosing Option B. The correspondence between rating scale score and probability of choosing Option A is assumed to be linear. However, three different conversions are defined with different degrees of sensitivity.

The starting point for the VOT estimation

The ASP games must start with an initial estimate of VOT in order to construct the first ASP game. Since every respondent has a potentially different VOT, there is no unique VOT that will be a good starting point for all respondents. Therefore, the simulation model allows the starting VOT to be varied to observe the effect of the starting point on the accuracy of the final VOT and the speed of convergence to that final VOT.

The number of games

Each respondent plays 10 ASP games, with each game using an updated estimate of VOT to set the attribute levels for that game. The number of games needed for convergence can be observed in the results.

The logit model sensitivity parameter

In converting the utility difference calculated for each game into a measure of probability of choice, a simple binary logit model is assumed, where:

$$P_A^i = \frac{e^{cU_A^i}}{e^{cU_A^i} + e^{cU_B^i}}$$

and P_A^i = the probability of individual i choosing option A

U_A^i = the utility of option A to individual i

c = logit model sensitivity parameter

The logit model sensitivity parameter measures the sensitivity of probability of choice to changes in the utility difference between the two options. Higher values of c produce greater changes in probability for a given change in utility of an option. In the simulation, the sensitivity coefficient can be changed to see what effect it might have on the final VOT estimated for each individual.

Respondent error in marking the rating scale

In addition to the variability of VOT between respondents, there is also a variability in response within any one respondent. When presented with the same choice on many independent occasions, there is no guarantee that the respondent will always give the same score on the rating scale indicating their degree of preference for the options. This is modelled in the simulation by adding a random error term to the score which the respondent would give according to deterministic application of the above decision rules. This random component mimics the random errors which respondents would make in reality when choosing between options, and tests the robustness of the survey method to random errors in scoring.

Results of the Simulation Testing

The simulation model was run with a wide range of settings to check the sensitivity of the results to different levels of the inputs. The major results are summarised below.

The number of games

The change in estimated mean VOT with an increasing number of games played is shown in Figure 3. The mean of the distribution of VOT in the population is 12.9¢/minute, so it can be seen that the ASP estimation procedure does a good job in replicating that value. Importantly, it also reaches equilibrium fairly quickly, with good estimates being obtained after only three games, independent of the starting point (ranging between 4¢/minute and 24¢/minute). This finding confirms the general rule given by Sawtooth Software (12) that three games are needed for each parameter to be estimated in an ASP survey.

In addition to estimating the average VOT, the ASP survey also estimates the distribution of VOT, as summarised by the Standard Deviation of the VOT. The change in estimated Standard Deviation of VOT with an increasing number of games played is shown in Figure 4. The Standard Deviation of the distribution of VOT in the population is 5.4¢/minute, so it can be seen that the ASP estimation procedure does a good job in replicating that value, after starting with an initial estimate of the Standard Deviation of zero (i.e. the same starting value is used for all respondents). Importantly, it also reaches equilibrium fairly quickly, with good estimates being obtained after only three games provided that the starting estimate of VOT is on the high side of the final VOT. For a starting estimate of VOT below the final estimate (e.g. the starting estimate of 4¢/minute in Figure 4), the convergence process is much slower.

The number of points on the rating scale and the conversion of rating scale scores to probabilities of choice

The simulation allowed the use of 5, 7 and 9 point rating scales, and allowed three variations for converting each rating scale into probabilities of choice. The results of the testing of these nine combinations are shown in Figure 5. The errors in estimation of the mean VOT and the standard deviation of VOT are shown for each of the rating scale combinations. It can be seen that the 5-point scale has higher error levels, especially for method 3 of converting rating scores to probabilities (i.e. when the probabilities are more extreme – closer to 100% or 0% - for each rating scale score). The 7-point and 9-point scales both give acceptable results in terms of errors in estimation, with the 7-point scale having slightly lower errors for each of the scale conversion options. Since the 7-point scale is also slightly easier for respondents to use, the 7-point scale was adopted for use in the ASP survey design.

The logit model sensitivity parameter

Using a 7-point scale, with a starting VOT of 24¢/minute and a final VOT of 12.9¢/minute (SD=5.4¢/minute), the logit model sensitivity parameter was allowed to vary between 1 and 11. The effect of this variation is shown in Figure 6 in terms of the error in estimating the mean and standard deviation of the VOT. It can be seen that within a range of c=3 to c=5, the mean and standard deviation are accurately estimated. Below this range, both are over-estimated slightly, while above this range the mean is over-estimated considerably while the standard deviation is under-estimated considerably.

To test the robustness of this result, the relativity between starting and final VOT was changed by using new starting values of VOT and new mean values of VOT. On the basis of these tests, it is concluded that the mean and standard deviation of VOT are generally well estimated within a range of c=3 to c=6.

The starting point for the VOT estimation

A necessary feature of the ASP survey process is that a starting estimate of the VOT must be made in order for the first game to be designed. This starting estimate must be within a suitable range in order to produce a reasonable first game. Unfortunately, because the VOT is variable across the population, there is no single best starting value to begin the games.

To test the effect of the starting value of VOT, it was systematically varied from 2¢/minute up to 30¢/minute. The errors in estimation of the mean VOT and the standard deviation of VOT are shown in Figure 7, for a final mean VOT of 12.9¢/minute (with SD=5.4¢/minute). It can be seen that the mean and the standard deviation are both well estimated for starting VOT in the range of 18¢/minute to 24¢/minute.

To test the effect of the starting value of VOT, with different final values of VOT, the analysis is repeated with different final values of VOT. The results for final mean VOT=5.4¢/minute (with SD=2.3¢/minute) are shown in Figure 8. It can be seen that, for the final mean VOT=5.4¢/minute, the mean and the standard deviation are both well estimated for starting VOT in the range of 2¢/minute to 20¢/minute.

The general conclusion to emerge from this analysis is that for a wide range of final mean VOT (5.4ç/minute up to 12.9¢/minute), good estimates of the mean and standard deviation of VOT can be obtained by using a starting value of VOT which is greater than the final mean VOT. The starting value can be two to three times the final mean without loss of accuracy in the estimation of the final VOT.

Respondent error in assessment of utility

In all the above simulations, it has been assumed that respondents give their scores on the rating scale with absolutely no error. That is, they examine the options presented, apply their own VOT to the trade-off, estimate the difference in utility, convert this to a position on the rating scale and then select the nearest score on that scale. Naturally, most respondents don’t go through this process with mathematical rigour, and all the above processes are occurring in the sub-conscious mind. Therefore, if presented with the same game on several independent occasions, it is unlikely that the respondent would give exactly the same response. We would expect approximately the same answer, but not exactly the same answer. To approximate this mental process, the simulation model allows for the score on the rating scale to be randomised somewhat, by adding or subtracting a random amount from the respondent’s assessment of their position on the rating scale for each game before selecting the nearest score. There is some question as to how much randomness should be attached to the respondent's score, and what the shape of the distribution of error should be. A conventional logit model would assume that the error is Weibull distributed. In it's current form, the ASP simulation model selects a random error from a uniform distribution of errors within a given range. The effect of this error on the estimation of mean and standard deviation of VOT is virtually nil. Across a wide range of errors (up to three points on the 7-point rating scale), the maximum error in estimation of the mean or standard deviation is less than one-tenth of one cent. The only effect of allowing an error in marking the scale is to introduce a degree of scatter into the correspondence between estimated and actual VOT, as shown in Figure 9 for an error of +/-2.

Conclusion

This paper has described a form of stated preference survey entitled Adaptive Stated Preference (ASP). The distinguishing features of ASP surveys are that they use more, but simpler, questions for the respondent, they adapt the games presented to the respondent on the basis of answers given to previous games, and they estimate parameters for each individual rather than for the population or for segments of the population.

The paper has described the results of a simulation study to test the ability of the ASP survey to produce unbiased estimates of the mean and distribution of the Value of Time. As a result of the simulation study, the following recommendations are made with respect to the design of ASP surveys:

- The ASP survey is able to reproduce unbiased estimates of the underlying distribution of VOT for a population;
- A 7-point rating scale, with a linear conversion to probability of usage and a score of 1 equal to 12.5% probability, should be used;
- The starting VOT used should be higher than the expected final mean value of VOT. A starting VOT twice as large as the final VOT appears to work well;
- Each respondent needs to play three games for each parameter to be estimated;
- A logit model sensitivity parameter of c=5 appears to produce unbiased estimates of the mean and standard deviation of VOT;
- The estimates of mean and standard deviation of VOT are unaffected by random errors made by respondents in marking the rating scale.

The results of this simulation give reassurance that, in addition to being easier for respondents, the ASP survey method also produces unbiased estimates of the distribution of

VOT (or any other parameter under consideration). Given the extreme flexibility of the disaggregate results produced by the ASP survey method, the demonstration of unbiased estimates is an important confirmation of the usefulness of this method.

References

Richardson, A.J. and Ampt, E.S. (2001). "Never mind the Data – Feel the Model". *Keynote Paper for an International Conference on Transport Survey Quality*, Kruger Park, South Africa

Alonso W (1968). "The Quality of Data and Choice and Design of Predictive Model". *Special Report 97*, Highway Research Board, pp. 178-192.

Polak, J. and Jones, P. (1997). "Using Stated-Preference Methods to Examine Traveller Preferences and Responses", In P.Stopher and M. Lee-Gosselin (Eds.) *Understanding Travel Behaviour in an Era of Change*, Elsevier Science: UK., pp 177-207.

Wang, D., Oppewal, H. and Timmermans, H. (2000). "Pairwise Conjoint Analysis of Activity Engagement Choice". *Environment and Planning A*, Vol 32, pp 805-816.

Watson, S.M., Toner, J.P., Fowkes, T. and Wardman, M. (2000). "Efficiency Properties of Orthogonal Stated Preference Designs", In J de Dios Ortuzar (Ed.) *Stated Preference Modelling Techniques*, PTRC Education and Research Services Ltd: London, pg 91-101.

Bradley, M. and Daly, A.J. (2000). "New Analysis Issues in Stated Preference Research", In J de Dios Ortuzar (Ed.) *Stated Preference Modelling Techniques*, PTRC Education and Research Services Ltd: London, pg 121-136.

Wardman, M. (2001). "A Review of British Evidence on Time and Service Quality Valuations", *Transportation Research E*, Vol 37, pp 107-128.

Morikawa, T. (1989). *Incorporating Stated Preference Data in Travel Demand Analysis (Chapter 8: Simulation Study of Models with Taste Heterogeneity)*. PhD Dissertation, Department of Civil Engineering, MIT

Kroes, E. and Cirillo, C. (2000). "Individual versus Grouped SP Models – Some Experiments". *European Transport Conference*, Cambridge, UK.

Fowkes, T. (2000). "Recent Developments in Stated Preference Techniques in Transport Research", In J de Dios Ortuzar (Ed.) *Stated Preference Modelling Techniques*, PTRC Education and Research Services Ltd: London, pg 37-52.

Fowkes, T. and Wardman, M. (1998). "The Design of Stated Preference Choice Experiments with Particular Regard to Inter-personal Taste Variations", *Journal of Transport Economics and Policy*, Vol. 22, pp 27-44.

Sawtooth Software Inc. (1993). *Adaptive Conjoint Analysis – Version 4*. Sawtooth Software, Washington State, USA.

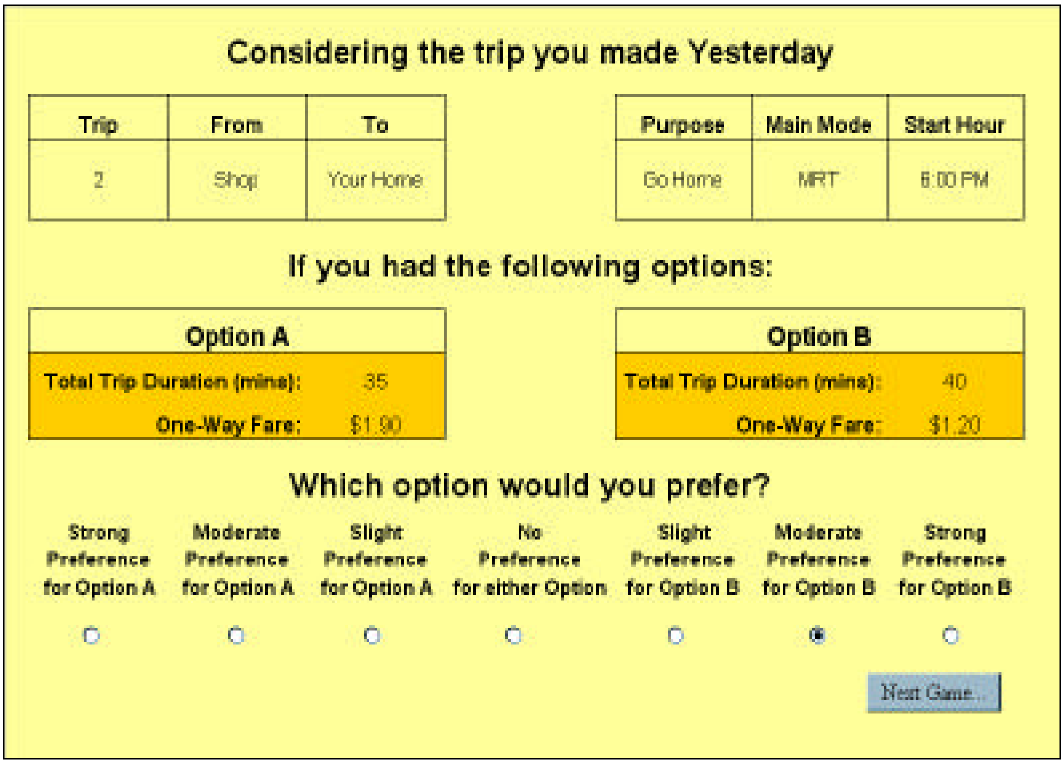

FIGURE 1 A Typical Question from the Adaptive Stated Preference Survey.

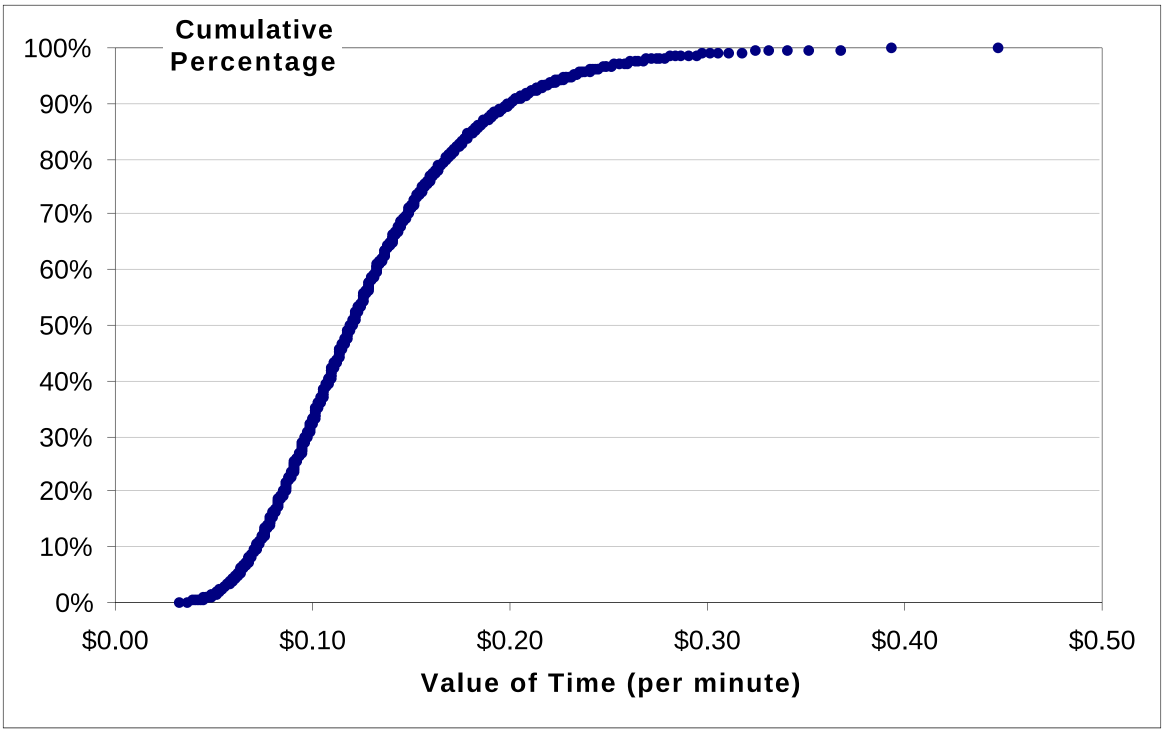

FIGURE 2 The Distribution of the Value of Time in the Population.

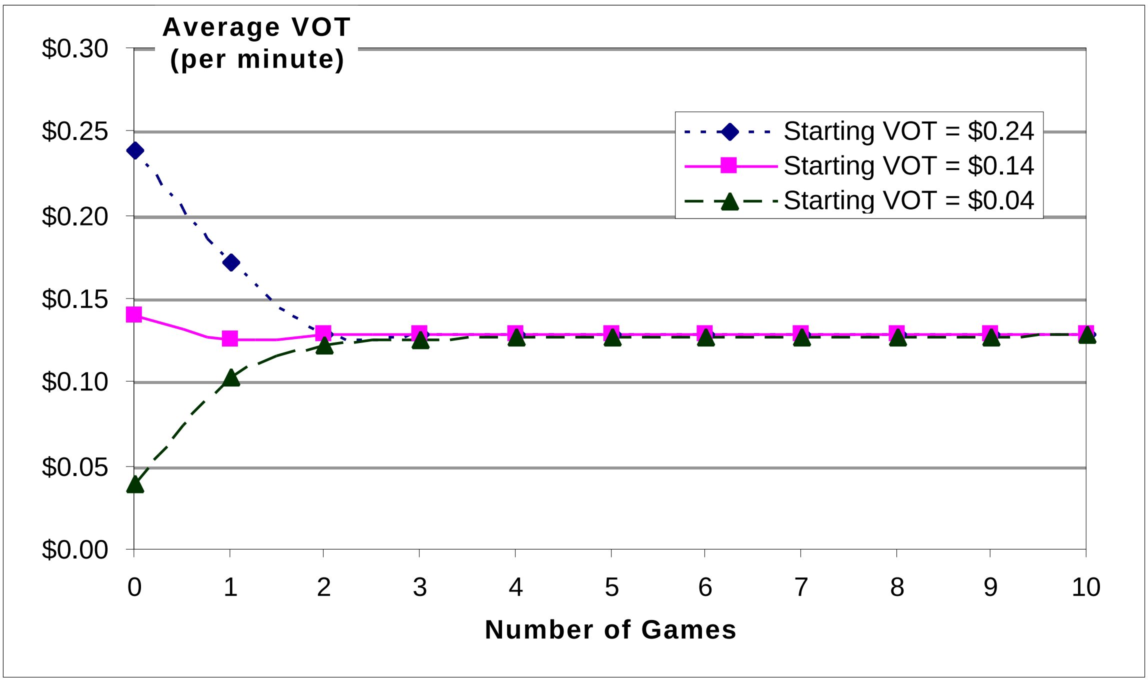

FIGURE 3 Average VOT as a Function of Number of Games Played.

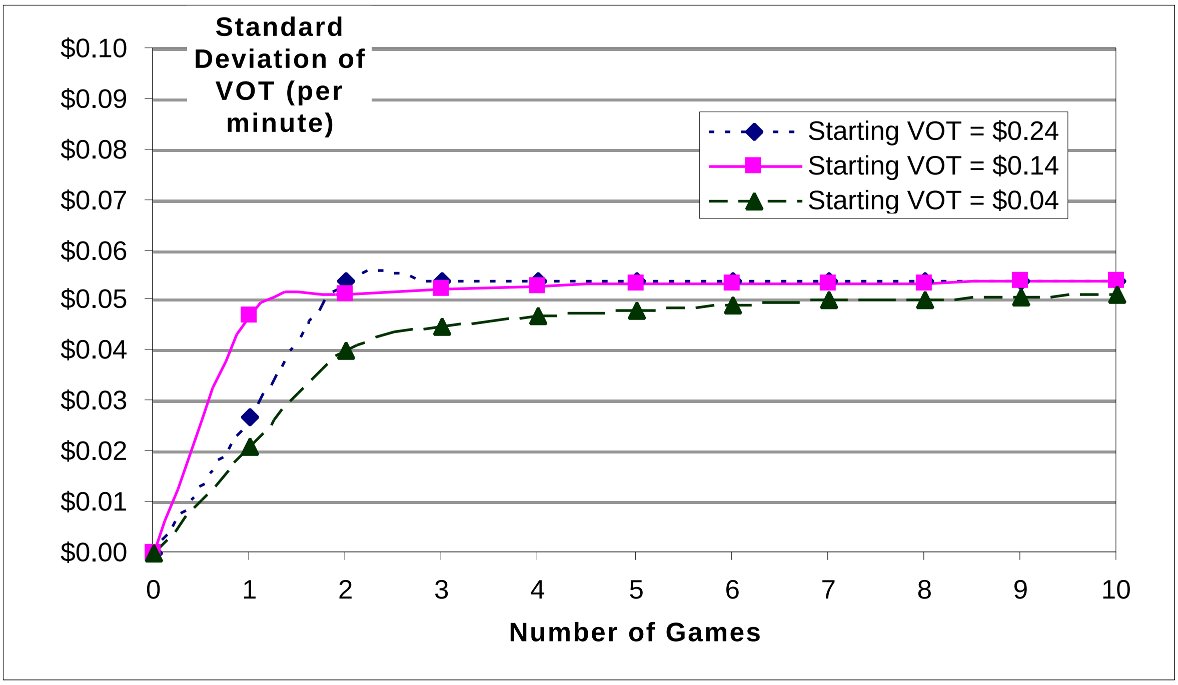

FIGURE 4 Standard Deviation of VOT as a Function of Number of Games Played.

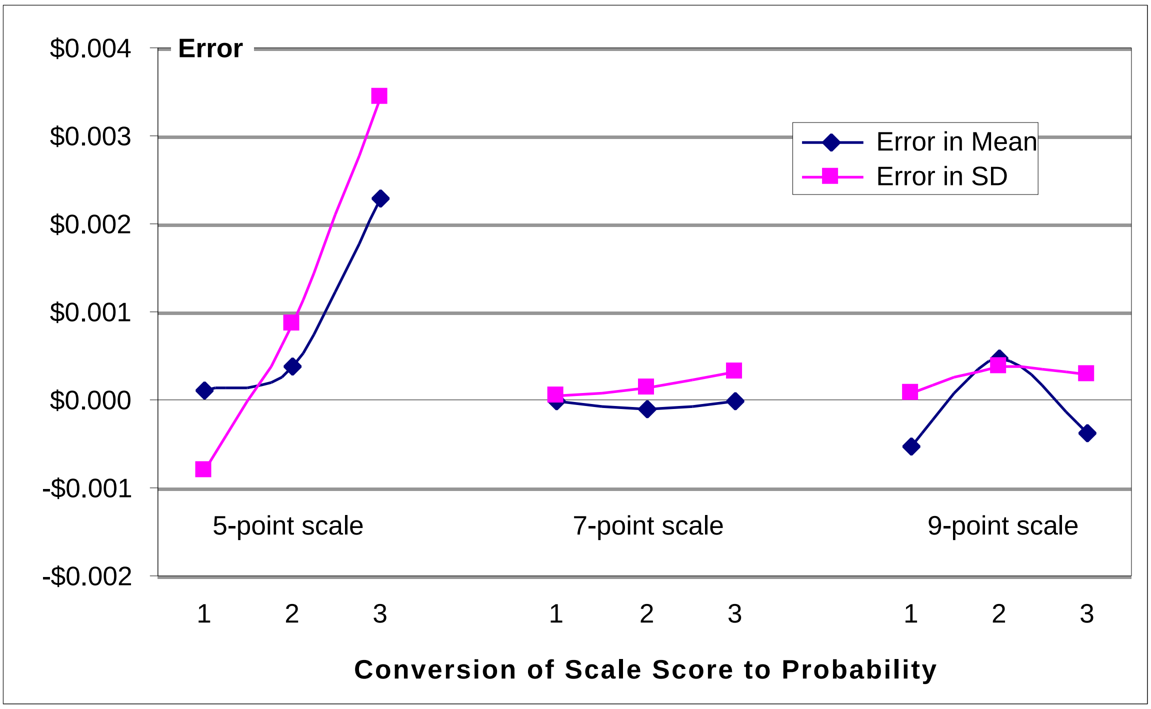

FIGURE 5 Errors in Estimation as a Function of Type of Rating Scale Used.

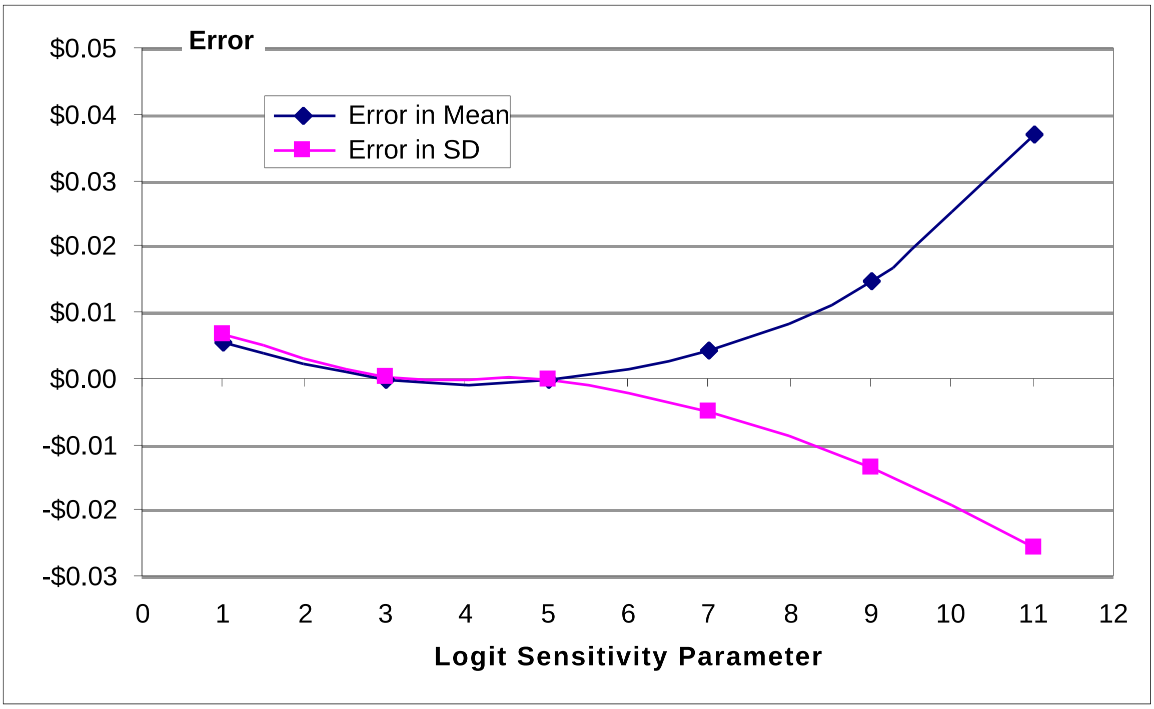

FIGURE 6 Errors in Estimation as a Function of Logit Sensitivity Parameter (starting VOT=24¢/minute).

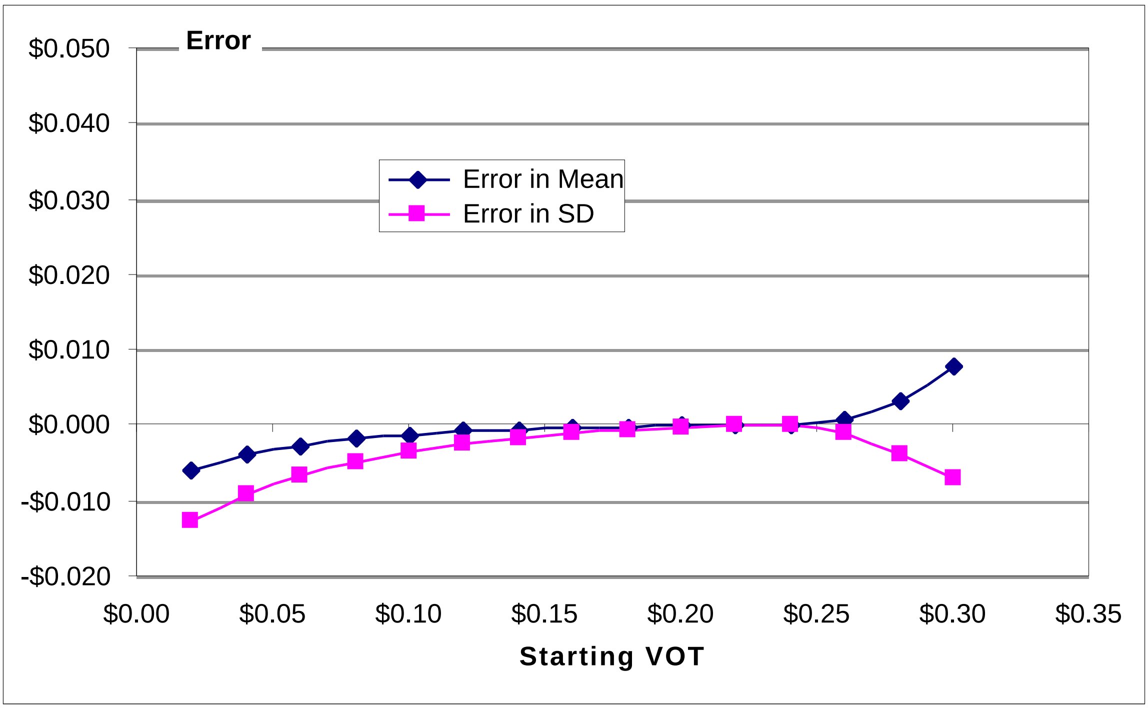

FIGURE 7 Errors in Estimation as a Function of Starting estimate of VOT (final mean VOT=12.9¢/minute, final SD = 5.4¢/minute).

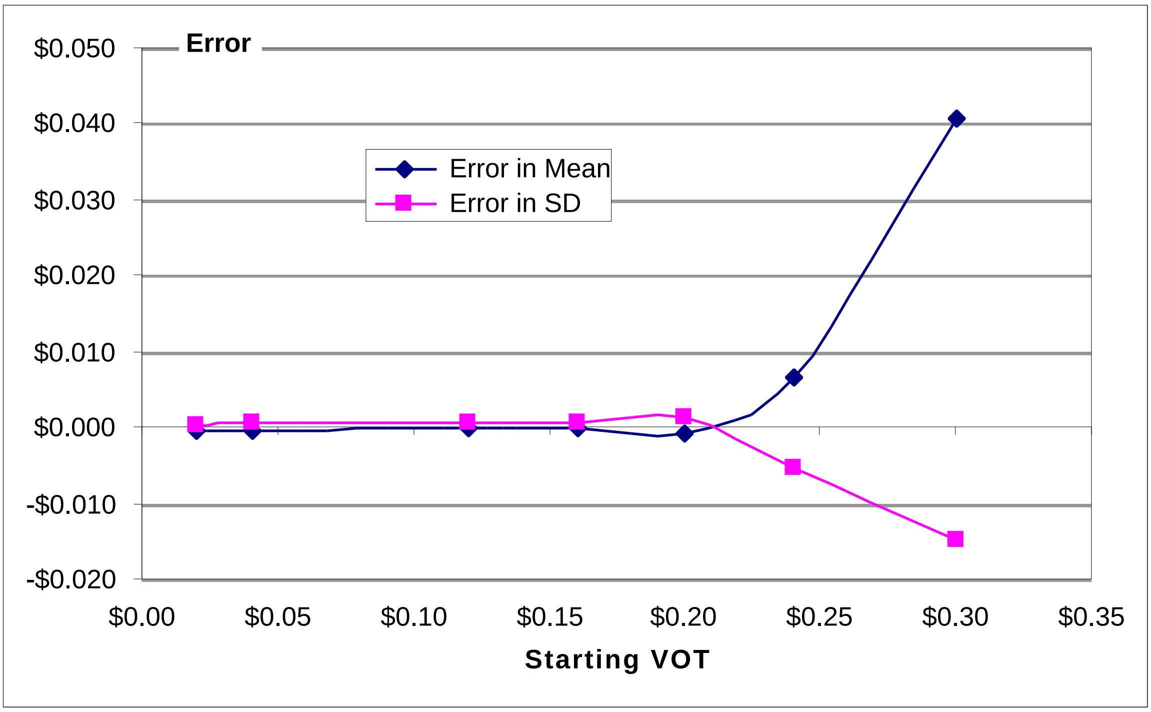

FIGURE 8 Errors in Estimation as a Function of Starting estimate of VOT (final mean VOT=5.4¢/minute, final SD = 2.3¢/minute).

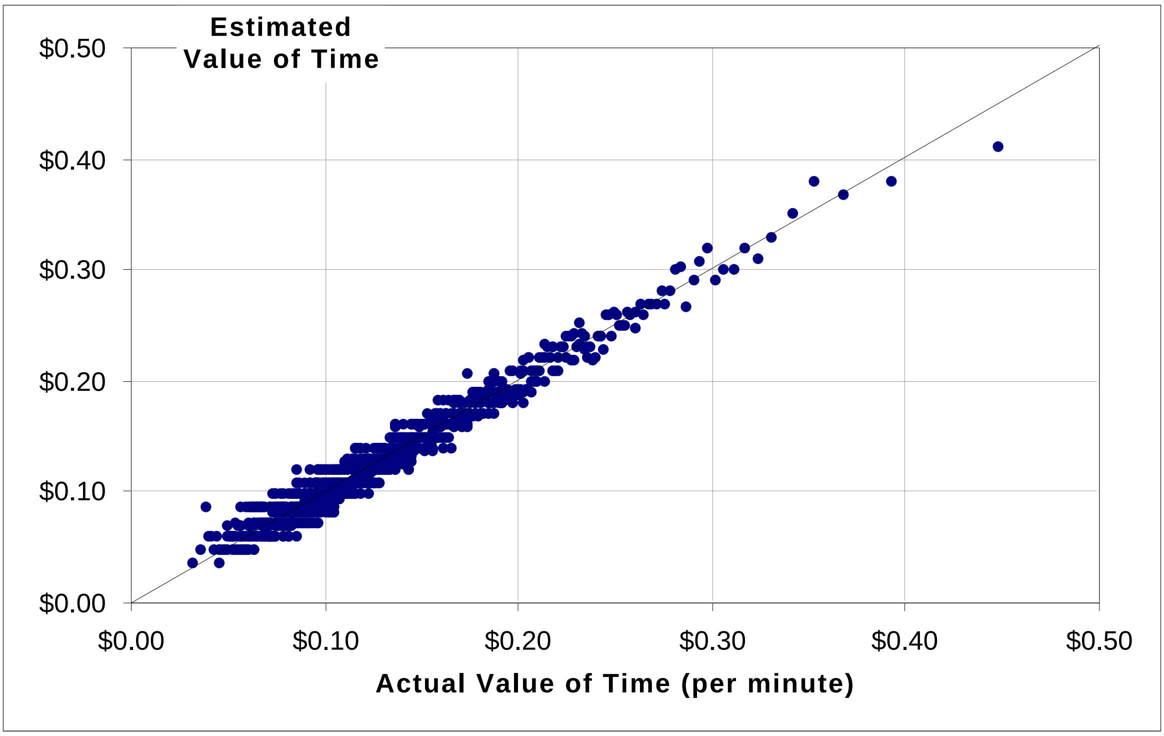

FIGURE 9 Comparison of Estimated and Actual Values of VOT (error = +/-2).