**17**

**Stated Choice Experiments: Experimental Design and Data Analysis**

***Jeffrey J. LaMondia***

*Auburn University, Auburn, USA*

***Mark Wardman***

*University of Leeds, Leeds, UNITED KINGDOM*

1. **Introduction**

Over the past decade, stated choice experiments have become an important component of travel demand research, as they can provide more detailed travel behavior data as well as opportunities for making choice tasks more realistic. Consequently, the freedom associated with hypothetical scenarios and questioning requires researchers to be more conscious of their survey design, processing, and conclusions. As such, this workshop was focused on discussing the challenges of stated choice experiments as well as the techniques that are currently being used to address common issues. Specifically, participants were concerned with whether complexity, realism, or relevancy is more important. They highlighted the tradeoffs between having too many factors, making experimental designs intuitive, and keeping sample sizes reasonable. The typical format of experimental design is comprised of multiple choice tasks, which ultimately affect the econometric modeling of choices as well as the required survey sample size. Therefore, the topics detailed below include the inherent challenges of stated choice experiments, stated choice experiment processing, and addressing these challenges through adaptive experimental design and statistical advancements. The final emphasis of the workshop was on using the discussion topics to identify future research objectives.

1. **Challenges and Opportunities**

**2.1 Self-Forecasting**

One of the most common challenges associated with stated-choice experiments is individuals’ “self-forecasting”. This means that individuals’ approaches to a stated experiment are based on their own preferences, interpretation, and memory, which can vary greatly. Depending on how a question or scenario is presented, it can be interpreted many ways. This leads to concerns about the way stated choice experiments are designed and collected. At the most basic level, fewer observations or limited variability across sociodemographics can limit modeling applications or generalizability because a wide enough range of self-forecasts are not captured in the survey. Additionally, the content of surveys must also be carefully composed, as the quality and quantity of responses can depend on how relevant experiments are to individuals. As such, four main techniques for collecting this data were identified. First, it is important to target surveys to specific audiences with stated choice experiments rather than try to reach a large audience. This allows you to anticipate common types of self-forecasts and proactively address them. Second, surveys should be kept short and free from bias. Participants recognized that self-forecasts become more prevalent in longer surveys when respondents feel fatigued or in overly-directed surveys when they feel an agenda is being promoted. Third, it is better to use “demand effects” (i.e., real world scenarios based on respondent’s actual experiences or past behaviors) than ‘imagination experiments” (i.e., written scenarios that have little connection to each individual’s experience). For example, there is less room for interpretation if questions revolve around hypothetical changes to a specific commute trip relative to a generic leisure trip. In some instances, it is helpful to first gather GPS or other real-world data from the respondents to then use as a baseline in the stated experiment. Fourth, researchers can study heuristics, such as navigation, choose sets, and minimum regret, by using changing options during stated choice experiments.

Self-forecasting also leads to concerns about the impact that individuals’ perceptions, overt experiences and knowledge have on responses. Workshop participants emphasized that these factors can significantly affect responses; as a result, many researchers are wary of stated choice surveys. For example, reliability is based on managing the role perceptions play in completing the stated choice experiments. It’s not the actual attributes of alternatives that are important, but the way individuals’ perceive these attributes. Therefore, researchers need to condition results on individuals’ degrees of knowledge, more so than their sociodemographics. Specifically, these perceptions and experiences affect individuals’ reference points from which they gauge alternatives, and these reference points vary for everyone. Workshop participants were concerned that results from models that do not distinguish individuals’ reference points can be misinterpreted. Furthermore, we need to be conscious of the fact that as individuals complete stated choice experiments, they are being conditioned for each subsequent set of questions. Ultimately, it was concluded that the most robust method for using stated choice experiments is to calculate elasticities, as they provide the scale for model calibration.

**2.2 Experiment Processing**

In addition to having preconceived ideas that affect individuals’ responses, the way in which individuals process information during a choice experiment can also affect their responses. Researchers found that individuals are more engaged with experiments that they perceive as more relevant. As a result, relevancy is more important than complexity. More importantly, individuals’ tend to learn during an experiment, which can affect their future responses. They can read into experiments quite a bit, so it may be helpful to have random individuals help design choice experiments (prior to a pilot study). Additionally, participants acknowledged that including multiple questions on a topic (which can later be used in a factor analysis) need to be used to fully capture individuals’ responses.

This processing, however, can lead to variations in individuals’ thresholds (e.g., personal limits for alternative characteristics). A large variety in thresholds can exist between and within respondents, and it was recognized that ‘experienced’ people demonstrate more thresholds than others. In fact, participants agreed that the more thresholds an individual has (i.e., less information), the higher the quality of their response. Fatigue can also affect individuals’ stated choice responses. This most commonly takes the form of simplifying tasks as their get tired, resulting in reducing complex down to single factoring. Overall, the way individuals process information in experiments is similar to what they do in real markets, so it may also useful to collect data on the decision-making process as well as their final responses. However, workshop participants were concerned about whether we should actually ask supplementary questions about the experiment process; some felt that supplementary questions are unreliable since respondents may not even be aware of how they process information. Still, if researchers choose to ask, there are many ways to do it, from simple feedback to walking back through the experiment and asking individuals’ to comment on their thoughts. Ultimately, workshop participants agreed that we need to more explicitly integrate processing into stated choice experiments. It then follows that processing needs to be integrated into the econometric modeling. For example, a random mixed multinomial logit model would allow us to see if people who respond differently to supplementary questions will also respond differently to other questions.

**2.3 Adaptive Experimental Design**

A considerable amount of time was spent discussing the best procedures for adaptive experimental design as a means for addressing the processing issues outlined previously. There are two types of adaptive methods: (1) across individuals, in which choice tasks and parameters are updated after each individual has completed the experiment and (2) within individuals, in which choice tasks and parameters are updated as each individual completes their experiment. It was generally agreed that ‘across individuals’ adaptive methods are more proper to apply in large sample choice experiments, but ‘within individuals’ adaptive methods may be better for small, targeted experiments. The process for adapting experimental designs across individuals is also quite flexible. Researchers can either update hypothetical tasks/ parameters after each individual or after a specific number of individuals. In either case, previous responses can be used to narrow choice task values down to reach specific choices or behaviors quickly. A number of methods for completing this choice process can be used, including the adaptive conjoint method. Some participants were concerned that without proper planning, adaptive designs can lead to suboptimal results. While this may occur, insignificant results may also indicate that there simply isn’t enough variability at the current scale; if it is dropped before more variability is introduced then we lose information. Regardless, many participants were supportive fully embracing Bayesian statistical methods to improve the efficiency and effectiveness of adaptive experiments. In these situations, researchers have the option of conditioning on every phase/moment in time for each individual.

**2.4 Statistical Optimization**

Another important topic, tied to adaptive experimental design, was the use of statistics to optimize stated choice experimental design. While many types of stated choice experimental designs exist (that support specific sample sizes or model structures), researchers are now interested in developing designs that relax traditional assumptions or optimize over multiple model types. This advanced process reduces the required sample size as well as the number of questions in the experiment, saving time and reducing costs. The process for developing an optimized design is comprised of three main steps: First, one specifies the end utility function, which details what parameters are included, interaction effects, correlation, etc. If researchers skip this step, they are making many implicit assumptions. Then, one determines which model specification you expect to use. Finally, one develops the experimental design based on the model specification. In many cases, using Bayesian information or pre-information to predict estimates is recommended. It is also important to recognize that standard errors are just as important as the beta parameter estimates, as they describe the reliability and statistical meaning. As such, they should be emphasized in the model selection as well.

Workshop participants were especially interested in using S-estimates to calculate required sample sizes. In this case, larger samples sizes typically provide better beta and standard error estimates. As the number of respondents increase, we affect standard errors because we are dividing by N (i.e., the sample size). Therefore, we can use the covariance matrix to calculate required sample sizes. This doesn’t require simulation or advanced computing but one does need to assume a model type to calculate the variance/covariance matrix. Fortunately, one can simplify the process by using the second derivative of the log-likelihood functions, which are often presented in the current literature. Sample sizes can be further impacted by varying the number of levels and ranges of attributes. Large sample sizes are associated with narrow ranges and more levels. Better sample sizes are associated with wider ranges and fewer levels. However, it is important to not let this method run your stated choice experimental design. Rather, it is simply a tool to help in the development and there are many tradeoffs between the different modeling structures. All experimental designs will generate a result, but some will require larger sample sizes relative to those designs that make assumptions for a more efficient model.

Throughout the workshop, participants considered whether stated choice experiments were better than revealed choice experiments. In the end, they concluded that revealed choice data may be even more challenging with which to work. While revealed choice data is needed for proper forecasting, it is generally filled with a lot of noise. Since the largest concerns associated with stated experiments is how accurate responses as, experience shows that reality typically doesn’t match results of revealed choice models either. This is based on the fact that revealed choice experiments typically do not consider ‘real’ alternatives (i.e., transit is perceived as bad, so it is not really an option). Participants recognized the need to further study how stated choice and revealed choice experiment results can be combined, perhaps calibrating revealed choice data with stated choice estimations. Still, many argued to keep revealed and stated choice data separate to reduce addition error.

1. **Future Research Directions**

Finally, the workshop participants identified a number of future research topics within six broad categories. These included hypothesis bias, information processing, technology, individual-based design, next generation discrete choice models, and next generation response formats. First, research needs to consider hypothesis bias through the use of reference alternatives/ bases. Most people think incrementally and make decisions relative to some point or situation (either current or accumulated). As a result, further research needs to determine if we get ‘better’ estimates with SP calibrated with RP or with combined SP and RP estimation. There is often noise in RP data, so SP calibrated with RP may be more accurate. However, we then need to incorporate experience, information processing, etc. into the research. Second, overt experience (or contextual knowledge) matters, and we need to consider this ‘within’ as well as ‘between’ individuals in the information processing. Fatigue effects can also be significant, and its effect on stated choice experiments needs to be further explored. As fatigue sets in, simplification and single processing occur. Researchers also need to ask supplementary questions about processing. Additionally, it may be helpful to pursue mixed methods of qualitative analysis (such as interviews). Third, as new technology is introduced, research needs to explore how it and be applied to stated choice experiments. For example, how can we integrate GPS, Wi-Fi, and phone technology with data exchange? Technology suitability and transferability (i.e., individuals’ experience with different technologies) also need to be studied, and this is especially true in the developing countries. As immersive technologies become more prevalent, participants were interested if exposure to simulated hypothetical scenarios will lead to better responses, and what is the best way to present/experience these scenarios. Fourth, new individual-based designs need to be explored, including adapting across respondents (which improves efficiency, but can complicate the startup portion of the updating process) and adapting within respondents in real-time (which may cause respondents to fall into confusing choice processes). Fifth, given that people respond differently based on the design of experiment, should we have a range of next generation discrete choice model designs? How much change is due to the design or the way people are processing the design? To answer these questions, it will be important to develop non-compensatory models, dynamic modeling that incorporates anticipated changes in scenarios, and models that include respondent created/ guided data response forms that compare with real choices. Sixth, research needs to evaluate the reliability of how responses to stated choice experiments are given. For example, we need to improve methods for collecting psychometric parameters in various formats, especially in developing country contexts. We also need to identify shift thresholds and further study priority evaluation, including revisiting things such as adaptive SP. Designs for continuous dependent variables, such as expenditure, car use and time use need to be developed. Finally, when we have too many attributes that we can’t process all of them, what are the best methods for incorporating rankings of best/worst attributes?

**Acknowledgement**

The workshop chairs sincerely thank the resource paper authors and all of the workshop participants for their valuable insights and intellectually stimulating perspectives that greatly contributed to the success of the workshop and made this report possible.