UTS CRICOS 00099F

Contact:
Professor John Rose
BIDA, UTS Business School
John.Rose-1@uts.edu.au
02 9514 5994

Business Intelligence & Data Analytics (BIDA) Research Centre

Water Pricing Submission Review: Response II

Final report: 12 May 2021

Table of Contents

Table of Contents

1. Response
2. References
3. The Research Team
4. Further Information

Project Name	Melbourne Water Pricing Submission Review
Commissioning Agency	Essential Services Commission, Victoria Government

1. Response

The study undertaken by Melbourne Water's consultant, Newgate Research, was thoroughly undertaken and well thought out. Nevertheless, this does not mitigate concerns related to the specific technique employed, and the validity of the results obtained from the study. The major issue is not that the adopted technique, SIMALTO, is likely to result in biased outcomes, but the fact that there is no known research into the technique that one can draw on to understand how the outputs may be impacted upon from a range of possible sources. This differs to other stated preference techniques that have been extensively studied in both academic and real-world settings.

This precise concern is highlighted by Melbourne Water's false dichotomy between real world and theoretical models with respect to stated preference methods. It is the position of the author of this document that all stated preference techniques have their origin in academia, and are subsequently used after the fact in real world settings. Indeed, stated preference techniques can be categorised in a number of different ways. One pertinent distinction between alternate stated preference techniques is between those that have a theoretical basis of behaviour, and those that do not. For example, discrete choice experiments (DCEs), sometime referred to as choice-based conjoint (CBC) or simply choice experiments (CEs), and contingent valuation (CV) methods have a strong theoretical basis describing how respondents answer such survey questions. Traditional conjoint (TC) and SIMALTO on the other hand do not (with respect to traditional conjoint, see Louviere et al. 2010), which discusses this precise issue). Note however, that like TC and SIMALTO, DCEs and CV methods originated in academic settings, and like TC and SIMALTO, the original work introducing both DCEs and CV survey approaches did not provide a theoretical representation of behaviour. Indeed, the theoretical underpinnings of the DCEs and CV survey methods were derived later. This has yet to occur for both TC and SIMALTO approaches.

To make clear why this is important, consider that each stated preference method is attempting to determine how different attributes or variables impact utility for some good or service. The impact is measured in units known as utilities (in the case of CV, the outputs are willingness to pay (WTP) outcomes, which is a monetarisation of utility). Let U represent the overall utility for a good or service, **x** a matrix of attributes, and **B** a vector of weights that reflect how much a given x impacts on U. Assuming a linear utility function, this can be written as

$$U = \mathbf{Bx}. \qquad (1)$$

The objective is typically to determine **B**.

In models of discrete choice (also used for CV) and regression (which is used in TC for example), there also exists an error term within the equation, such that Equation (1) becomes

$$U = \mathbf{Bx} + \boldsymbol{E}, \quad (2)$$

where ***E*** is a stochastic term, random over respondents completing the survey. Given that ***E*** is random, it stands to reason that U is also randomly distributed over the population. The ***E*** in DCE analysis and CV modelling have been given a theoretical explanation that explains behaviour in a population. In TC, the ***E*** is simply present as a result of the modelling process and has no behavioural interpretation. In estimating the models using a DCE, CV or TC survey approach, standard errors are estimated for each term within the **B** vector, which under a classical statistical interpretation determines how robust each estimated weight is within the analysis. In terms of interpretability, this allows one to undertake tests of statistical significance.

SIMALTO however differs to the above, in that the values contained the vector **B** are not estimated, but determined by the analyst, or in this case, Melbourne Water's consultant, Newgate Research. Note that the author does not dispute that extensive consultation was undertaken to determine these values, however it remains a fact that the Newgate Research developed the values contained within the vector **B**. Further, in the simulation, the weights are transformed or normalised using an unknown approach which has not been described. A second point of differentiation between SIMALTO and other stated preference techniques is that with SIMALTO, respondents choose values from within the matrix **x**, whereas in the other approaches, the values of **x** are given to respondents to form alternatives, and respondents are asked to provide feedback about the alternatives shown (either by ranking multiple alternatives, selecting an alternative out of a set of alternatives, or indicating how much they are willing to pay for the alternative). The above description highlights several concerns about the SIMALTO approach.

1. That the weights (**B**) are not modelled, but inputs determined by the analyst;
2. That the model assumes a utility function commiserate with Equation (1) in which there exists no error term;
3. The choice of attribute levels, from a theoretical and behavioural perspective would suggest that in addition to an overall error term, there should exist an error term around each **x** (the error is associated with each choice), such that one would expect utility to be $U = \mathbf{B}(\mathbf{x}+\mathbf{e}) + \boldsymbol{E}$; and
4. The lack of estimation, whilst beneficial to the analyst, means that there no standard errors estimated for the elements contained within **B**.

Points (1) and (4), above means that one must accept without comment, the values assigned by the analyst to each element within the vector **B**. Of course, one could always change the values of the weights post data collection to determine their impact on U, however this does not appear to have been done here. Points (2) and (3) means that there is no behavioural interpretation of the results, and give rise to concerns that by ignoring possible sources of error, SIMALTO is open to different sources of bias such as resulting from, for example endogeneity. If one considers for example the true behavioural model describing utility for Melbourne Water's services to be $U = \mathbf{B}(\mathbf{x}+\mathbf{e}) + \boldsymbol{E}$, and respondents are directly selecting the values of the attributes, x, then given the budget constraint implicit within SIMALTO, the error terms between any two attributes would be expected to be correlated, such that $r(e_k, e_l) \neq 0$ (one would also expect correlations between the values contained within **e** and **E**). Behaviourally, this would have an impact on the values of B_k and B_l, however given points (1) and (4), this is completely ignored within the SIMALTO system of analysis. With respect to point (4), a typical econometric modelling exercise involves estimating the weights in vector **B**, alongside standard errors that can be used to undertake hypothesis testing about these weights. Based on the estimated standard errors, hypothesis testing allows the analyst to test whether the weights are statistically significantly, meaning that they are different to zero or not. For example, $B_k \neq 0$ and $B_l \neq 0$. Given that there is no model, and the values of B_k and B_l are inputs assumed by the analyst, it is not possible to test whether the values chosen by the analyst represents the true decision calculus used by respondents in answering the SIMALTO questions, including whether the values that best reflect respondents choice behavior, including whether their true value should be zero or not.

A further categorisation that is useful to make between different stated preference techniques is between those techniques that have been validated against real world market settings and those that have not. To externally validate the outputs of a technique, one must undertake a stated preference survey using that technique and compare the outputs to models obtained using real world data. Within this categorisation, DCEs (CBCs), CVs, and to a lesser extent, TC survey approaches have been tested. Figure 1 presents a list of papers examining the issue of hypothetical bias and their findings with respect to DCEs (adopted from Haghani et al. 2021). On the other hand, SIMALTO appears not to have been validated as of yet, with no study examining to what extent, if any, SIMALTO will induce hypothetical bias. The above represents the primary, but not only, concerns of the author of this document with respect to the use of SIMALTO. That is, there is no documented evidence that allows one to understand the degree to which, if any, the approach suffers from hypothetical bias issues, nor whether mitigation techniques tested on other stated preference approaches, work with SIMALTO (assuming it does suffer from hypothetical bias concerns).

Questions have also arisen around whether SIMALTO represents a black box in respect to how the data is analysed. The issue originally arose given the original submission documents submitted on behalf of Melbourne Water by its consultant, Newgate Research, provided only mention of how the data was analysed using *a "bespoke mathematical model derived from approaches similar to neural network designs"*, alongside some form of simulation. Melbourne Water's consultant, Newgate Research, also provided a spreadsheet with their analysis, however the version provided to this author at the time did not allow one to view the calculations being undertaken, just the inputs and outputs. A version of the simulation that reveals the calculations was made available on the 27th May 2021, five months after the original review was complete. It was unclear at the time of the original review how the analysis using SIMALTO relates to a neural network model, and now that the simulation approach underlying the analysis has been revealed, such a comparison can be confirmed to be unwarranted. Indeed, it is worth noting that prior to the new simulation spreadsheet being provided, it was only the answers provided by Melbourne Water's consultant Newgate Research that allowed insights into the analysis performed. Key to this was their answer that SIMALTO utilises a self-explicative approach. From this, one can deduce how the analysis was undertaken. Some aspects of the analysis however remain unknown. Based on the discussion above, we know that the SIMALTO model assumes that $U = \mathbf{Bx}$, where $\mathbf{B}$ are weights provided by the analyst, $\mathbf{x}$ are attribute levels selected by respondents, and U is utility which is a function of both $\mathbf{B}$ and $\mathbf{x}$.

The answers given prior to the simulation process being provided however did not make clear how utility is related to the WTP outcomes, and hence how the values reported were derived. From the simulation spreadsheet, the process can now fully be understood. In the survey, the SIMALTO questions require respondents to allocate hypothetical points amongst the attribute options. After each SIMALTO round, respondents were then asked how much they were willing to pay in the form of a CV question for the alternative they constructed by allocating the points, relative to the current status quo. Without any additional information of SIMALTO or the simulation process, it was assumed that these CV questions were used to convert the preferences obtained from SIMALTO into WTP values. Given the newly provided simulation worksheet, this is now known not to be the case.

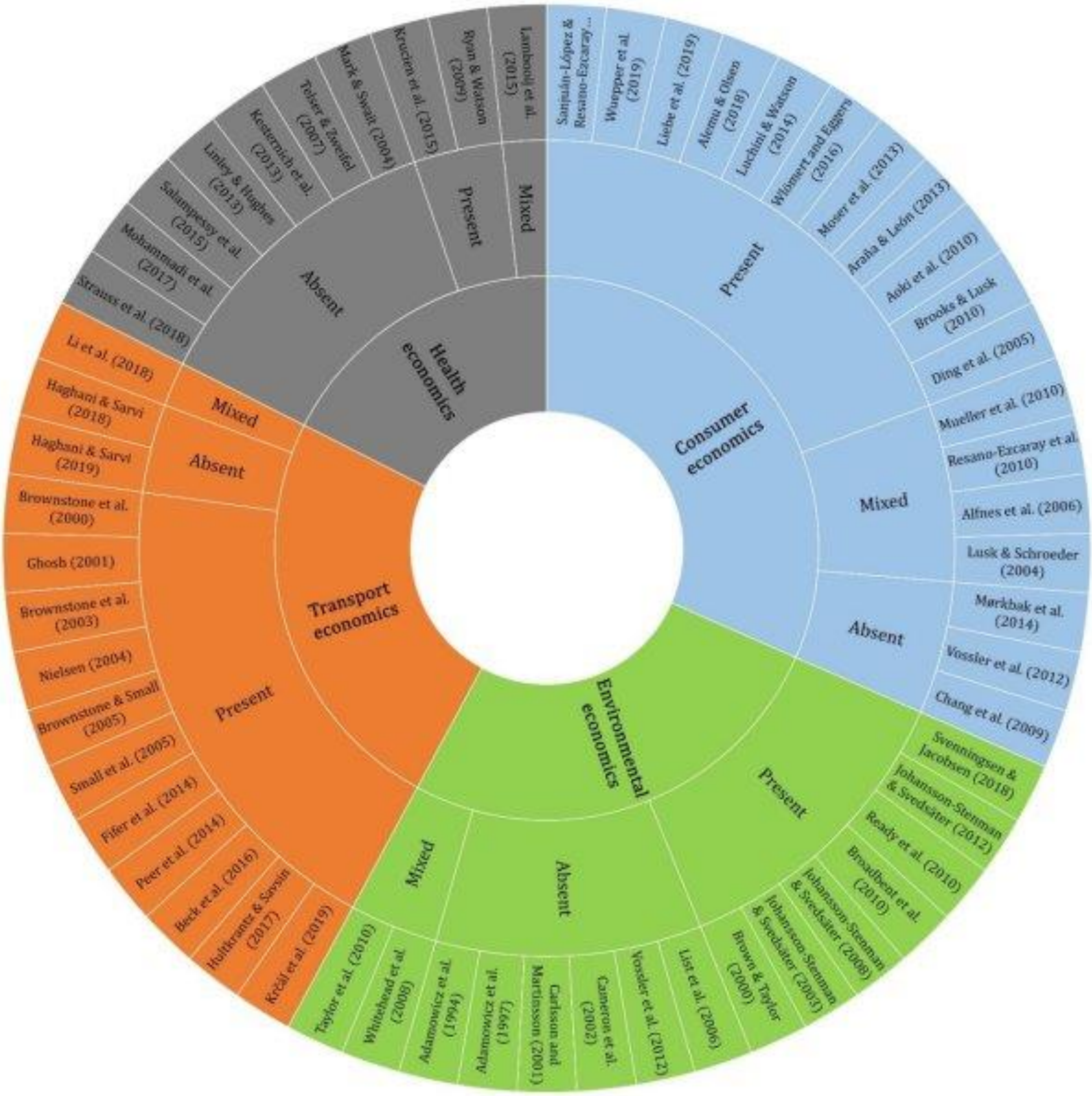

Figure 1: Empirical studies of HB in CE from environmental, consumer, health and transport economic disciplines (adopted from Haghani et al. 2021)

To derive the WTP values reported, the CV questions are ignored completely. The simulation establishes two alternatives, one for the status quo, and the second representing a move away from the status quo, represented by changes to the various attribute levels. Fixing the normalised or transformed values of parameter weights contained in B, the simulation then cycles through various attribute level combinations that form the second alternative, and computes the choice probabilities using the Bradley-Terry-Luce (BTL) probability function. By systematically changing the attribute levels, the preference share for each attribute is then determined. Note that the outputs of this process are preference shares, not WTP outputs.

To derive the WTP outcomes, dollar values associated with each attribute level are first assigned. For example, the four levels associated with the attribute stormwater quality are given dollar values of -\$0.59, \$1.27, \$5.63 and \$11.52 respectively. Next, for each respondent the optimal level obtained for that respondent is used to determine which price to assign in the overall WTP calculation. For example, if for a respondent, the second level is determined to be optimal in respect to the preference simulation, then the value $1.27 is assumed. This process is done for all attributes, with the WTP values obtained summed across all attributes and attribute levels. The WTP values reported are these sums.

A number of issues arise from this approach. Firstly, respondents were asked to trade hypothetical points for different attribute levels rather than dollar values. For example, in the SIMALTO survey, the four attribute levels for the stormwater quality attribute were 0, 9, 31 and 61 respectively. Not only were hypothetical points used rather than dollar values in the trade-offs made, respondents were explicitly told in the survey "The aim of this second round is to see what other options you would choose if you had a larger budget. **This doesn't mean the actual Charge is going to increase or decrease**; it's about helping us to understand what services customers value. This is part of why we've used points rather than dollars here." The bold statement above is taken directly from the survey and not added here for emphasis. The last sentence of this quote directly states that respondents were asked to trade hypothetical points rather than dollar values, and other than being told the current water charge is as part of the CV task completed after each SIMALTO round, it would appear that respondents were never told the actual costs in dollar values for each attribute they were being asked to trade-off during the survey.

WTP represents a monetarised form of a Marginal Rate of Substitution (MRS), which is defined as the value a person is willing to pay (or accept) for a change in an attribute that results in a zero change in utility. For example, if one were to change an attribute level from the status quo, how much would respondents be willing to pay for that change such that they are indifferent between the change, and the status quo alternative. This requires that a trade-off be made by the respondents between the attribute levels and a cost measured in dollar values. In the current survey, respondents made no such trade-off, and worse yet appear not to actually ever have been informed of the dollar values of the trade-offs, or if so informed, effectively told to ignore them. Indeed, the simulation process optimises choices based on a trade-off between non-monetarised points, and only after the fact assumes that the costs associated with each choice were fully known and accounted for by the respondent when making those choices. Simply put, such an assumption is completely unwarranted and even if one were to accept SIMALTO at face value, the WTP outputs obtained from the simulation model do not reflect a true WTP outcome made by the respondents.

A second issue arises in that the CV questions in the survey are not used in the calculation. No reason has been provided as to why the only actual questions that do ask respondents to actually make a monetary trade-off have not been used in the analysis, and not even reported. The question arises as to what is the relationship between the simulation WTP outputs reported and the actual WTP outputs provided by respondents in the survey? That is, SIMALTO as a method was designed to capture preferences, in the form of utilities and unlike other stated preference techniques, SIMALTO is not designed to compute WTP outcomes. In models of discrete choice, WTP outcomes are typically not directly estimated (it is possible to estimate them directly if the model is estimated in WTP-space as opposed to preference space), but are computed after the fact from the modelled outputs (WTP are simply derivatives of the parameter estimates obtained from the model). In the survey, the additional CV questions could be used as a dependent variable with the independent variables being the attributes and levels obtained from SIMALTO. The entire process is described in Equation (3).

$$\mathbf{Bx} = U \rightarrow WTP. \quad (3)$$

This was not done. The question then becomes, if such a model were estimated, would the WTP outputs for the respondent reported optimal alternative (the one they provide during each SIMALTO round) correspond to the one that is obtained from the simulation task that was used to compute the WTP outputs.

We note that in several replies, that it has been mentioned that SIMALTO was used in previous submissions. Examination of the previous submissions reveals that whilst this is true, the actual name SIMALTO was never previously used. Further, previous submission documents are vague in terms of how WTP values were obtained with references to work undertaken by LaTrobe University to compute the WTP component of the submission. It is unclear whether SIMALTO was used in these submission to also compute WTP. In the current submission, SIMALTO has been directly linked to the WTP component of the study, and the manner that it has been done is at odds with the task that respondents were asked to complete when undertaking the survey.

In addition to the above comments, we note that both Melbourne Water and their consultant Newgate Research make the assumption that existing approaches such as DCEs cannot be used to model complex choices. Whilst there is some evidence of this, and the author of this document is somewhat sympathetic to this argument, it is worthwhile noting that this issue is one that is still being debated within the literature. For example, Hensher (2006) argues that relevance is more important than trying to limit cognitive burden in discrete choice experiments (DCEs). Hensher argues that real life choices are complex and might involve lots more alternatives than given in a DCE task. DCEs have been applied to a wide range of complex choice questions. Further still, there exists a significant literature on making DCE tasks more realistic to respondents, via the use of pictures and videos for example, as well as developments around in experimental design techniques that are designed to reduce cognitive burden such as with the use of partial profile designs. The point being, that one cannot simply rule out the use of DCEs or CV in this instance.

2. References

Haghani, M., Bliemer, M.C.J., Rose, J.M., Oppewal, H. and Lancsar, E. (2021) Hypothetical bias in stated choice experiments: Part II. Macro-scale analysis of literature and effectiveness of bias mitigation methods, arXiv preprint arXiv:2102.02945.

Hensher, D.A. (2006) How do respondents process stated choice experiments? Attribute consideration under varying information load, Journal of Applied Econometrics, 21, 861–878.

Louviere, J.J., Flynn, T.N. and Carson, R.T. (2010) Discrete choice experiments are not conjoint analysis, Journal of Choice Modelling, 3(3), pp 57-72.

The Research Team

Dr. John Rose is Professor and founding Director of the Business Intelligence & Data Analytics (BIDA) Research Centre in the UTS Business School. John's research interest lies primarily in the area of modelling of behavioural agents in decision making, with a particular focus on understanding infrastructure planning. This includes large scale infrastructure projects of national interest, such as projects within the transport, telecommunications and energy sectors. He also researches extensively in Health and Environmental Economics. He is in the process of writing a book on generating efficient stated choice experimental designs (with Mike Bliemer). He is also actively engaged in contract research, working in the areas of toll road evaluation and modelling.

In his core area within the transportation sector, Professor Rose has worked on a number of major projects, including:

- Projects for the NSW Government to determine the demand for heavy rail for the Northwest sector of Sydney;
- Patronage forecasts for the proposed Sydney City Metro;
- The generation of values of statistical life for car drivers and passengers as well as pedestrians in NSW, which are used as an engineering guide to determine safety parameters in the construction of new road projects; and
- Numerous toll road patronage forecasts both in Australia and New Zealand.

All of these studies have implications on the engineering scope of work; as such studies are required to determine the minimum level of service requirements needed to result in sustainable transport related infrastructure projects.

Since graduating with a Ph.D in 2005, John has been involved in winning Category 1 grants worth over $3.4 million including four ARC Discovery grants. Funded projects include:

- Public health - examining community preferences for organ donations ($240,000);
- Transport - examining travel time reliability for public transport ($215,000);
- Applied economics - examining the concept of utility separability ($360,000); and
- The external validity of discrete choice experiments ($460,000).

Other grant areas John received vary in topic area, from transport to environmental engineering logistics. In addition, John has received $6 million in industry-based contract research since 2005.

4. Further Information

For further information, please contact:

Professor John Rose
Director/BIDA
Centre for Business Intelligence and Data Analytics (BIDA)
University of Technology Sydney (UTS), Business School
ABN: 77 257 686 961
Dr. Chau Chak Wing, Building 8
225 George Street, Level 10
Broadway, NSW 2007
E: John.Rose-1@uts.edu.au
T: +61 2 9514 5994