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

Final report: 15 April 2021

Executive summary

This report provides a final review of a submission made by Melbourne Water on its pricing strategy for the period 2021 to 2026, including a reply provided to questions posed regarding the original submission. In providing this review, we have concentrated on the customer survey designed to capture willingness to pay outcomes for different level of service outcomes, as well as responses provided by Melbourne. In conducting this final review, we conclude that whilst the research design was competently undertaken, serious questions arise as to methodology employed and the estimated values obtained. We conclude that the initial Melbourne Water report was lacking in details that could be used to substantiate what analysis was undertaken. Indeed, some of the answers provided contradict the original report, whilst others fail to answer the question or provide misleading evidence. Nevertheless, it is possible to determine the specific methodology employed, and determine how the data was analysed. Based on the evidence provided, it is determined that the analysis conducted will result in known biases, however the direction of these biases cannot be determined. It is therefore the conclusion of this report that the methods applied are not fit for purpose of assessing the willingness to pay by Melbourne Water Customer for different level of service outcomes.

Table of Contents

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

1. Background and context

This report provides a review of the Melbourne Water Pricing Submission 2021, including subsequent responses to queries related to the methods employed. The primary focus of this report is on the Waterways and Drainage Charge: Willingness to Pay study undertaken by Melbourne Water's consultant Newgate Research. The research undertaken represents an extensive customer centric study of Melbourne Water customers that investigates their preferences for the provision of improved level of service. Using a marketing research tool, SImultaneous Multi-Attribute Level Trade-Off Analysis (SIMALTO), 1,354 respondents consisting of 1,069 residential (metropolitan), 135 residential (rural), and 150 non-residential (business) customers were surveyed. Extensive qualitative research was undertaken to inform the survey design, including eight focus groups, a series of cognitive interviews, and stakeholder workshops. Follow up surveys with 77 respondents occurred after the survey, suggesting an extremely well thought out and thorough research design was implemented.

Whilst the research design undertaken as part of the study is extensive and competently undertaken, we find significant deficiencies with the specific approach employed, that being SIMALTO. SIMALTO represents a form of stated preference experiment in which respondents are asked to make trade-offs between various attribute levels under a budget. SIMALTO suffers from a number of issues that give rise to the validity of the willingness to pay values obtained from its use. The methodological approach underlying SIMALTO relates to work undertaken in the 1970s through to the 90s after which it largely disappeared completely from the academic literature for a number of reasons. The approach was found to suffer from issues related to **measurement error**, **response bias, lack of realism, difficulties in interpretation, social acceptability biases**, **an inability to detect non-linearity effects,** and **double counting**. In addition to these documented deficiencies, the approach also is likely to induce further biases with the sequential nature of the questionnaire approach likely to give rise to endogeneity, a well-known source of bias in statistics. Other sources of potential bias likely to arise from using SIMALTO include simulation error, and hypothetical bias. Finally, we find that the technique has no underlying theoretical basis linking it to behaviour, meaning that it is not clear what economic interpretation should be given to the results presented. The issue of simulation error is of particular concern, as the Willingness to Pay (WTP) outcomes hinge on a simulation exercise, which requires the simulation of how respondents would react to billions of possible combinations, based on their responses to three questions. Statistically, the ability to do this is highly improbable, with the likely outcome being that the outputs represent nothing but simulation error.

In the next section of this report, we present a brief description of what we understand the SIMALTO process to be. In doing so, we attempt to place the technique within the stated preference framework so as to provide the reader with a better understanding as to what the identified issues identified are, and why we believe they exist. In Section 3, we discuss each of the identified issues related to the technique, providing possible solutions where possible. Each solution discussed has been identified and tested using other stated preference techniques. Nevertheless, in providing possible solutions to each concern raised, we note that they cannot be implemented for the current study given that the survey data has already been collected and the research completed. As such, we

do so, simply to highlight what improvements could be applied to the technique in future research. The final section provides some concluding comments

2. Background

SIMALTO as a research method is a form of stated preference (SP) survey task in which respondents are asked to make a series of choices based on different hypothetical scenarios. Nevertheless, SIMALTO differs from other more traditional SP formats, such as traditional conjoint, choice based conjoint (CBC), and contingent valuation (CV) type questions in a number of ways. Further, it should be noted that the version of SIMALTO implemented as part of the study incorporates both CV and CBC type questions. Traditional conjoint presents respondents with a large number of alternatives, usually constructed from a full factorial or fractional factorial, which they are asked to rank or rate in a single task. Developed in the 1960s, traditional conjoint has gradually fallen out of favour since the 1980s, being largely replaced by CBC questions. CBC survey approaches differ to traditional conjoint insofar as CBC presents respondents with subsets of alternatives from which they are asked to choose their most preferred alternative in multiple tasks. Derivatives of CBC include asking respondents to rank the subset of alternatives in each task rather than choose their single most preferred or having respondents select their most and least preferred alternatives in a process termed best-worst scaling. An alternative to traditional conjoint and CBC, used widely in environmental economics studies and research dealing with non-market goods, is CV. Typically, CV presents respondents with a single scenario and directly asks them how much they are willing to pay for the scenario. Different types of questions have been used, including allowing open-ended responses, asking respondents would they be willing to pay a specific amount, or being asked if they would be willing to pay a specific amount followed by an even larger amount if they answer yes, or a lesser amount if they answer no.

SIMALTO on the other hand presents respondents with a pre-specified budget in the way of points, which they are asked to allocate amongst a series of attributes. Each attribute has a number of levels, representing increasing improvements in the attributes level of service, and require a certain number of points in order to be selected. The survey extends SIMALTO beyond its original formulation by requiring respondents to indicate how much they would be willing to pay, selected from pre-selected values (a form of CV question), for the alternatives they construct given the points budget allocated. This additional willingness to pay question was not part of the original SIMALTO question format. After answering the willingness to pay question, the process is repeated twice more (including with the additional willingness to pay question) with the respondent a larger budget of points to allocate in each round. After completing three rounds, respondents are next presented with the choices they made in the three budget allocation tasks, as well an alternative representing the level of service currently exhibited by Melbourne Water. Presented with these three alternatives, respondents are asked a discrete choice question as to which alternative they would be choose given prices calculated based on their previous selections (a form of CBC question). This too represents a deviation from the original SIMALTO questionnaire format. Finally, respondents are asked to respond to an open-ended CV question.

Despite differences to traditional SP formats, the underlying data collection process used in SIMALTO is likely to suffer from the same deficiencies inherent in other SP methods. This is particularly the case given that the extended version of SIMALTO used within the study, incorporates not only SIMALTO, but makes use of two additional SP type questions (i.e., CBC and CV) which are used the willingness to pay outcomes of interest. As such, not only are the SIMALTO questions likely to result in biased outcomes, but the inclusion of additional SP questions will also provide additional sources of biases that have not been accounted for. It is therefore the conclusion of this report that the results provided within the Melbourne Water Pricing Submission 2021 document are highly likely to be compromised, suffering from several sources of bias, each of which may affect the outputs in different directions.

3. SIMALTO

SIMALTO as a technique has an extremely limited academic footprint, being confined to a conference paper that is no longer accessible, and a few references that do not describe what the approach is or actually does. A Wikipedia page is available (a non-academic source that can be written by any person and is not referred) which provides only scant detail. Hence, the only information available that can be used to decompose and understand the method is the Melbourne Water Pricing Submission 2021 report, the accompanying research report, as well as a document prepared by Melbourne Water in response to questions asked of it by the Essential Services Commission (ESC) about the study. Unfortunately, the original Melbourne Water Pricing Submission 2021 report and the accompanying research report provide enough details to determine precisely what SIMALTO is, and how it works, and actually confuse the matter considerably. The reply to one question posed to Melbourne Water however allows one to understand how the method works in theory, if not in practice.

The original research report suggested that SIMALTO data is analysed using a "bespoke mathematical model derived from approaches similar to neural network designs". When asked about the neural network model employed, Melbourne Water's consultant Newgate Research replied that the analysis does not make use of a neural network model, but rather uses a version of what is called **self-explicative conjoint**. It is reference to self-explicative conjoint that allows one to untangle SIMALTO.

Figure 2 shows a neural network estimated on Brexit data. The left-hand side variables (referred to as input nodes are linked to output nodes (the right-hand side of the figure) which reflects the variable being forecast. The middle nodes (i.e., anything between the input and output nodes) are referred to as hidden nodes, which constate constitute different components that the network is learning to recognize (e.g., in image analysis, each hidden node may represent a different facial feature such as a noise or a mouth). In non-image type data analysis, the hidden nodes reflect different latent variables, the interpretation of which, are difficult to describe. The numbers shown in the figure in black are the estimated weights, whereas the numbers shown in blue are referred to as bias weights (indicated by the nodes labelled with the number 1). The bias terms are numeric constants that allow the value at the indicated nodes to be shifted upward or downward, much like the intercept in a linear equation. Although not shown, Neural networks also have error terms, described by a Sum of Squared Errors (SSE), which as you might expect, is the sum of the squared predicted minus actual values.

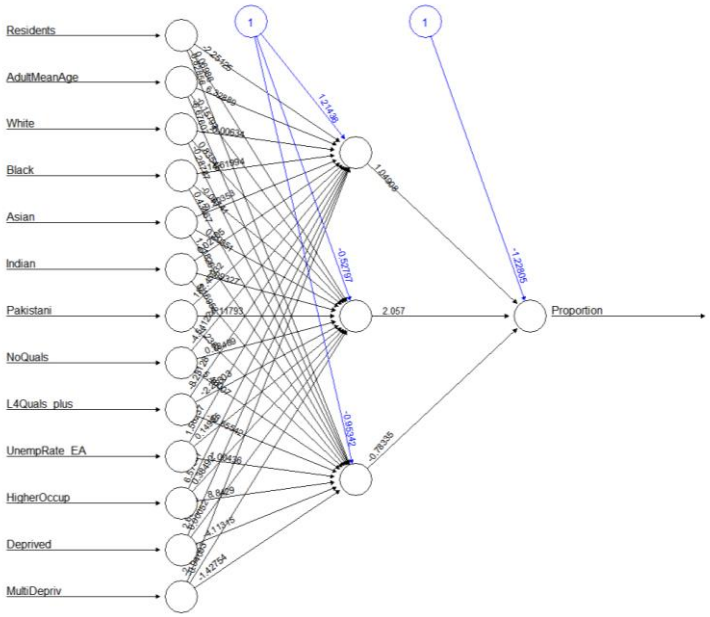

Figure 2: An example Neural Network

The analysis undertaken for SIMALTO bares no resemblance to a Neural Network, bespoke or otherwise. Indeed, self-explicative conjoint is simply a scoring technique that multiplies the answers provided to different questions together. There is no estimation process, no constant terms and no error.

Traditional conjoint approaches present respondents with several profiles, each of which describe some alternative. Respondents are then asked to either rank or rate each profile. The ranks or ratings provided by each respondent represent the dependent variable in the analysis, whilst the attributes of each various profiles are used as independent variables. Parameter weights using regression models (or similar) are then obtained reflecting the preferences of respondents. Depending on the study design, models can be estimated for each individual respondent, or a single model estimated on the aggregate data.

Self-explicative conjoint differs from traditional conjoint, both in terms of how questions are asked, and how the data is analysed. Whilst several forms of self-explicative conjoint were developed in the 1980s and 1990s, the main version involved presenting respondents with multiple attributes, each defined by different levels. Respondents were then asked to evaluate the desirability of the levels for each attribute (say on a 0-10 scale), holding all other attributes constant. Next, respondents were allocated points (usually 100 in practice) which were allocated across the attributes (and levels) to reflect relative importance. Preferences are then derived using a simple scoring mechanism which involves multiplying the importance weights with the attribute level desirability weights (see Green and Srinivasan 1990). In this respect, self-explicative conjoint is model free, with the outputs obtained via the above simple scoring mechanism (or similar). The above-described process matches the SIMALTO questionnaire utilised within the study, however it is unclear where the relative importance measures are provided by each respondent. As such, given the reply provided by Melbourne Water's consultant Newgate

Research, it is possible to determine that SIMALTO is a form of self-explicative conjoint, but not how the precise scoring is calculated.

It is worth noting that like traditional conjoint approaches, self-explicative conjoint as a data collection approach disappeared largely from the academic literature in the early 2000s. Studies conducted in the 1980s and 1990s exploring self-explicative conjoint as a preference elicitation approach found that it suffers from issues, including **measurement error**, **response bias, lack of realism, difficulties in interpretation, social acceptability biases**, **an inability to detect non-linearity effects,** and **double counting**. These issues are detailed below.

Measurement error: The method assumes respondents are machines and error free. Respondents are able to deterministically state their desirability and importance with exact precision. Being model free, respondents' answers to questions is assumed to be inviolate. If any errors do exist in responses, this will lead to endogeneity biases in the outputs produced. As mentioned by Louviere et al. (2010) (italics are added as the original quote utilises acronyms)

"there is no error theory associated with *conjoint measurement (CA)*, statistical or otherwise, which allow the theory to be represented as testable statistical models; hence learning that preference data do not correspond to simple conjoint models contributed to the development of statistical error theories in psychology." and

"The key difference in traditional CA and *Discrete choice experiments (DCE)*s lies in the critical role of error components. For CA, treatment of error components is an afterthought, whereas in DCEs, it is the starting point." and

"If one wants to make accurate predictions using s*tated preference (SP)* data in an *revealed preference* context where there are multiple choice options, one typically must rescale the SP estimates, yet CA has no mechanism to achieve this because it has no real error theory." And finally

"traditional CA is largely a curve fitting/scaling exercise where error components are largely ad hoc and lack clear interpretations."

Response bias: it is not clear whether this is an issue or not as the precise questions used to compute the parameter weights is never fully explained. The assumption is that different respondents use scales exactly the same. For example, using a 0-10 scale, do two different respondents understand and use the same importance rating for say the value 2? Again, this may or may not be an issue for this method, as the importance ratings, if used at all, are never explained.

Realism: traditional conjoint and CBC methods have been shown to represent more realistic choices for respondents to make than self-explicative conjoint approaches. Similarity to real choices has been shown to result in higher predictive validity (Green et al. 1981).

Interpretation: Traditional conjoint and CBC have been shown to be more likely to detect real importance weights, compared to self-explicative conjoint approaches which tend to ask respondents directly how important each attribute level is (Srivansan, 1988).

Social acceptability: when asked directly, self-explicative conjoint questions have been shown to result in respondents providing more socially acceptable responses rather than reveal their true preferences (Hensel-Borner and Sattler, 1999).

Sensitivity: Studies have found that self-explicative conjoint questions tend to produce less sensitivity than other SP approaches (Gedenk et al. 1999). Nitzsch and Weber (1993) report instances where self-explicative conjoint approaches result in total insensitivity to changes in levels, resulting in significant validity issues.

Non-linearity: Studies have found that self-explicative conjoint questions tend to produce linear relationships between attribute levels compared to other SP approaches (Green and Srinivasan, 1990).

Double counting: If there are redundancies in the attributes, self-explicative conjoint approaches tend to result in double counting. For example, if respondents see a relationship between flood mitigation and flood preparedness, the fact that both are questioned separately in self-explicative conjoint approaches (with respect to the desirability questions), rather than traded off, can result in double counting of preferences (Green and Srinivasan, 1990).

In addition to the above concerns raised by the literature, we would add a further concern related to the method itself. Unlike other SP approaches, traditional conjoint and self-explicative conjoint approaches have no theoretical links to actual behaviour.

Outside of environmental economics, traditional conjoint was the dominant SP approach used in both industry and academia in the 1970s and early 80s. As a technique, the approach suffered from a number of major criticisms that

saw its popularity wane, until today, it has disappeared completely from academic use, and is used only rarely in marketing research studies. One of the major criticisms of the approach was that it was borne primarily from an industry need and lacked any behavioural or theoretical basis. CBC on the other hand has a long history dating back to the 1920s and 30s (Thurstone 1927, 1931) in mathematical psychology, and has strong theoretical connection with economics theory, in particular random utility theory. Whereas Hull (1943) and Thurstone (1945) provided the mathematical foundations with respect to probabilistic choice theory, and Luce (1959) developed the theoretical foundations of discrete choice models, Marschak (1959) generalised these earlier models to allow for stochastic utility maximisation over multiple alternatives, referring for the first time to Random Utility Maximisation (RUM). It was this later paper, followed by the work of McFadden in the 1960s and 70s that firmly implanted the approach within the economic theory of consumer choice (McFadden, 1968, 1974).

CV was first proposed as a valuation method by Bowen (1943) and Ciriacy-Wantrup (1947) before being operationalised by Davis (1963). As with CBC, CV has a strong grounding in theory. Whereas the theoretical basis for CBC lies in micro-economic theory of preference revelation, the theoretical roots of CV lies more in welfare economics. The original papers on CV argued that the method provided a direct measure of the marginal rate of substitution curve of individuals attempting to place a value on public goods. Later, CV was tied to random utility theory, similar to CBC via the incorporation of a theoretical error structure associated with respondent's answers to CV question responses (see Hanneman 1984).

As with traditional conjoint, SIMALTO being based on self-explicative conjoint has no underlying theoretical basis for how or why respondents answer the questions asked of them. It is also important to note that neither cognitive interviews or follow up interviews are substitutes for a theoretical grounding. Such techniques may assist to identify whether people understand the questions being asked, but not confirm whether the questions asked conform to their actual behaviour. To demonstrate, consider the actual questions used within the SIMALTO questionnaire. The initial set of questions force respondents to undertake a constrained optimization process in which they are required to allocate points to a series of attributes and attribute levels. Behaviourally, such decision processes do exist in many real markets, such as buying computers from companies such as Dell, or when undertaking household budgeting. Over the past three decades, there is a growing literature used to describe such choice behaviour, linked to multiple discrete continuous demand models. SIMALTO does not make use of such literature however. This is clear in how the budget is presented, and the sequence of questions asked.

4. Beyond SIMALTO

Self-explicative conjoint approaches provide respondent specific scores for each attribute and attribute level used in the study. These scores are not willingness to pay estimates, and cannot be used to infer willingness to pay outcomes without the use of additional questions. The research questionnaire therefore has respondents answer willingness to pay questions, in the form of CV questions, after each SIMALTO/ self-explicative conjoint round. These CV questions are not part of the original self-explicative conjoint questionnaire approach.

Whilst, self-explicative conjoint is a form of stated preference question, the additional CV and CBC questions used by within the study are also stated preference questions. As such, the survey utilised for this study is also subject to the same issues associated with stated preference surveys in general. We detail these issues below.

4.1 Double counting?

Economists are quick to point out that unless dealt with directly, SP experiments whose main output of interest is the willingness to pay for some good or service are prone to double counting. Double counting can occur when respondents answering SP surveys consider not only their own preferences when making choices, but also those of others. Examining the instructions provided to respondents when completing the willingness to pay CV question, they are asked how much they would be willing to pay relative to the current charge (see Figure 1). In asking this question, respondents should be instructed to do so considering only their own preferences, and not those of other Melbourne Water customers. This is particularly pertinent given the sequence of questions, where respondents are forced to allocate points across a range of service improvements, before being asked how much they are willing to pay for those services. In selecting the original most desirable levels of services (prior to undertaking the CV questions), respondents may have to select services that they personally don't want or need, but given the need to spend the entire budget (plus or minus five points for round 1 (between 35 and 45 points), plus or minus three points in round 2 (between 77 and 83), and plus or minus three points in round 2 (between 137 and 143)), they may feel that other customers may benefit from the selections made, and hence provide a higher willingness to pay valuation than reflects their actual true value.

Now imagine that all of your selections from Round 1 have all been put in place (i.e. your whole budget of 80 points). How much would you be prepared to pay per year for these services? This could be below, the same as or higher than the current charge. *Pick one.*

$1 less per year	☐	I don't want to pay anything extra per year ☐	$1 more per year	☐
$2 less per year	☐		$2 more per year	☐
$5 less per year	☐		$5 more per year	☐
$10 less per year	☐		$10 more per year	☐
$15 less per year	☐		$15 more per year	☐
$20 less per year	☐		$20 more per year	☐
$25 less per year	☐		$25 more per year	☐

Figure 1: Contingent Valuation question

4.2 A black box

The absence of academic papers using SIMALTO, and the lack of detail provided in the original submission documents provided makes it extremely difficult to disentangle precisely what analysis was undertaken. Only in responses provided in a separate document to questions is it possible to begin to work out what analysis was employed. Even then however, questions remain. Whilst it is clear that SIMALTO is a form of self-explicative conjoint, the it is unclear where the relative importance measures are provided by each respondent that is necessary for such analysis. As such, it is impossible to determine how the precise scoring mechanism used is calculated.

4.3 There is no error term

It is important to note that SIMALTO represents a form of self-explicative conjoint, and hence does not use any statistical model to impute the preference weights of individual respondents. Traditional conjoint methods were discarded in the early 2000s due to the lake of theory associated with the error structure behind the model. Self-explicative conjoint methods, such as SIMALTO, apply simple scoring mechanisms that have no error term. This means that respondents are assumed to be error free when answering questions, having full information and acting purely rational, Further, respondents are assumed to make no mistakes and are perfectly consistent in their responses over the entire survey.

Beyond the SIMALTO self-explicative conjoint questions, the survey questionnaire has respondents answer several CV and CBC questions as part of their survey. As noted in Section 3, the theoretical roots underpinning CV lie in random utility theory, which incorporates a theoretical error structure associated with respondent's answers to CV question responses (see Hanneman 1984). In assuming no error structure in their modelling, the approach adopted removes the theoretical basis underlying the CV questions, the very questions needed to obtain willingness to pay estimates.

4.4 Endogeneity

In statistical models, endogeneity occurs when one or more independent variables are correlated with the error term of the model. Putting aside the fact that the analysis employed does not assume any error structure, it is highly unlikely that respondents answering a survey of this nature will be perfectly deterministic decision makers, making no errors, particularly given the large number of choices they are expected to make, not to mention that such choices are required to be made within the form of a constrained optimisation decision. The very structure of SIMALTO makes the likelihood of endogeneity highly probable, even if the researchers assume otherwise.

Assume that there is an error term within the utility structure underlying each respondents observed choices. Then for SIMALTO round *r*, respondent *n* will have utility

$$U_{nr} = V_{nr} + \varepsilon_{nr},$$

where V_{nr} is the utility obtained from SIMALTO, ε ε_{nr} is the error term, and U_{nr} is the actual unobserved utility held by respondent *n* in round *r* for the Melbourne Water profile they select.

After each round, each respondent is asked to indicate their willingness to pay for their selected profile. As with their selection of their preferred attribute levels, their selection of a willingness to pay value is also assumed to have an error term (which is what links this CV question to random utility theory). Let the utility associated with the CV willingness to pay task be

$$WTP_{nr} = G_{nr}\left(U_{nr}\right) + \zeta_{nr},$$

where $G_{nr}\left(U_{nr}\right)$ is the utility associated with the expressed willingness to pay, WTP_{nr}, and ζ_{nr} is an error term reflecting unobserved effects, and respondent errors related to their selected set of attribute levels in round *r*.

Expanding the willingness to pay equation, we obtain

$$WTP_{nr} = G_{nr}\left(V_{nr} + \varepsilon_{nr}\right) + \zeta_{nr},$$

such that any errors in the original choice enter into the utility function of the willingness to pay value selected by the respondent in subsequent questions. Indeed, the error terms ε_{nr} and ζ_{nr} will almost certainly have a non-zero correlation, meaning any mistake made by the respondent in answering the original questions, or uncertainty in their responses, will impact on their willingness to pay.

The issue of endogeneity extends beyond the rounds and willingness to pay questions. Indeed, respondents completing the survey are asked to choose between the alternatives they constructed during the three rounds, and the current Melbourne Water level of service. Again, any errors associated with the previous choices will propagate to this choice, in much the same way as they do to the willingness to pay questions.

Note that we have used subscript *r* in the above description. In the current study, survey respondents were asked to complete three rounds of the SIMALTO task. In CBC models, it is not uncommon to assume that respondent's preferences are correlated across choices. Indeed, research over the past 20 years has consistently shown that this is the case, particularly when dealing with SP type data.

Whilst SIMALTO by it's very nature assumes there are no error terms, the approach itself is highly likely to result in error. As such, it is possible, and indeed probable that SIMALTO data suffers from endogeneity across rounds, leading to another possible source of bias. Endogeneity of the type discussed above will result in biased utility estimates, which is completely being ignored. If respondents are not free from making errors, or if they make assumptions about things the attributes and levels beyond what the researcher provides, then endogeneity will be present within the data. Ignoring the issue will not make it disappear. Unfortunately, the data generation process of SIMALTO is likely to induce endogeneity, and without correction, any outputs from have an extremely high probability of being biased.

4.5 Validity and Hypothetical Bias

Do responses to hypothetical choice scenarios presented in surveys allow one to measure preferences and predicting choices in real-world settings? Are such experiments capable of producing unbiased and valid outputs that reflect real values, such as willingness to pay? Such questions relate to what is commonly referred to as the problem of hypothetical bias (HB), arguably the most fundamental issue regarding the legitimacy of SP surveys in providing evidence for policy making and cost-benefit analysis. A particular issue regarding SP experiments of which SIMALTO represents a specific case or subcategory, is the ability to derive unbiased measures such as willingness to pay or willingness to accept and the extent to which the values obtained for these measures in hypothetical settings correspond to their values in real-world settings. In the absence of an inherent behavioural realistic decision task, in hypothetical choice data, the use of SP surveys is somewhat fruitless, regardless of the methodological advancements in capturing econometric phenomena using sophisticated model structures or the improvements in statistical efficiency of choice surveys.

The validity of a study concerns the extent to which the study's results represent and apply to the true state of some observed phenomenon. It is often discussed in terms of internal validity and external validity. Internal validity refers to whether observed effects represent causal effects. SP surveys generally have high internal validity because the analyst can control for many aspects in the data collection including random allocation of participants to conditions. External validity (EV) refers to the generalisability of results beyond the study setting. As will be further defined later, HB is the bias in choice model estimates that results when data are collected in a hypothetical setting instead of in a more realistic setting. Hypothetical bias as such relates to EV. SP surveys have been shown to exhibit varying degrees of hypothetical bias. Where hypothetical bias has been found, it has resulted in biased willingness to pay estimates. In the majority of cases, the bias has been found to result in higher willingness to pay outcomes than have been found to exist in real markets. The extensive set of empirical evidence on HB in CV experiments has enabled economists to conduct meta-analytical studies, estimate the likely magnitude of HB and better understand what experimental factors and protocols aggravate HB (Carson et al., 1996; List and Gallet, 2001; Little and Berrens, 2004). While these meta-analyses adopted different methods of calibration and different sets of variables, their findings suggest that HB can affect the estimates of WTP or WTA by a factor between 1.2 (Schmidt and Bijmolt, 2019) and 3.13 (Little and Berrens, 2004). Although the degree to which SIMALTO suffers

from issues of hypothetical bias has never been formally tested, the approach falls under the SP framework and makes use of both CV and CBC type questions. As such, it is highly probable that respondents answering SIMALTO questions will also display hypothetical bias, and if SIMALTO does suffer from the issue, it is likely that the bias will be such that it will result in higher willingness to pay outcomes.

Whilst efforts were undertaken within the questionnaire to instruct respondents as to how to answer the questions asked, best practice was not followed in terms of how to mitigate HB. Numerous mitigation approaches have been proposed and tested within the SP literature, each of which has been shown to reduce, although not eliminate completely hypothetical bias concerns. These cheap talk scripts (Cummings and Taylor, 1999); certainty scales (Herriges et al., 2010); honesty priming (De-Magistris et al. 2013); induced truth telling and inferred valuation (Prelec 2004); solemn oath (Jacquemet et al. 2013); opt-out option or budget reminders (Ladenburg and Olsen 2014); time-to-think method (Whittington et al., 1992); RP-assisted estimations; referencing and pivot (contextually realistic) designs; and perceived consequentially scales or consequentiality scripts (Ding et al., 2005).

None of these measures were used in the current study, an omission which leaves question marks over whether or not the willingness to pay values reported should be taken at face value. If the evidence from other SP surveys is considered, and assuming that that such evidence extends to SIMALTO, then we argue that the willingness to pay outputs should not be taken at face value. Interestingly, although evidence suggests SP experiments tend to result in an upward bias in the willingness to pay outcomes, this is not always the case. As such, assuming such biases do exist, the directionality of the bias, whilst likely upward, cannot be determined with any certainty.

4.6 Simulation or simulation error

SIMALTO represents a form of self-explicative conjoint, the output of which are preference scores for each of the attribute levels examined within the study design. In order to link these preference scores to willingness to pay outcomes, the survey had respondents complete a CV task after each SIMALTO round. Next, the analysis undertaken involves the use of a simulation to determine for each individual respondent what their willingness to pay would be under different combinations of attribute levels. In effect, from three rounds of SIMALTO questions, the individual preference weights and their relationship to the CV data are used to simulate, ignoring the budgetary constraints, over 110,592,000 possible combinations of levels. Of course, this number will be reduced considerably given both a lower and upper constraint is imposed on the budgetary requirements (the constraints are somewhat flexible, for example, respondents actually have to spend between 35 and 45 points in the first round, not 40).

Given the limited data available for each respondent, and the number of different possible combinations, the resulting simulation is highly likely to represent nothing more than simulation noise. First, it is difficult to believe that there does not exist numerous local optima for such a problem. Indeed, the amount of data observations required to cover such a space in sufficient number in order to estimate possible outcomes would be prohibitively large, even if only simple relationships are observed within the data. Second, the small number of observations collected per respondent (3 rounds of SIMALTO) is extremely small from a data perspective. Assuming typical simulation procedures are undertaken, then mathematically, if the number of observations per respondent is fixed (as here), then the simulated willingness to pay value will not be consistent with the true value. If the number of observations observed for an individual is fixed, then consistency requires that the estimate of utility converges to the true value when sample size rises without bound. If sample size rises, but the choice observations per respondent is fixed, then the population distribution and its mean do not change. Raising the sample size improves the estimates and hence however it does not does ensure that the mean willingness to pay for the population is equal to the samples individual utility values. However, as the simulation is conducted at the single respondent level, it is clear that the results will be inconsistent.

5. Conclusion

Melbourne Water is to be commended for undertaking a thorough and well thought out research study designed to capture community preferences for changes to Melbourne Water operations. After examining the submission documents provided, no fault with the research design implemented for this project. That said however, we believe that the statistical approach adopted, that being SIMALTO, is problematic and will result in significant biases. Based on the above, it is the conclusion of this report that the method and results provided in the original Melbourne Water Pricing Submission are highly likely to be biased in unknown directions, and should be considered to be highly suspicious. The issues of the original report provided by this author about the method and approach remain, with the new concerns raised above also arising. To summarise the original concerns, the method has

1. No theoretical basis linked to behaviour, or for use in estimating willingness to pay;
2. Is likely to exhibit double counting resulting in higher WTP outputs;

3. Is a black box insofar as the methodology and analysis remains conjecture, albeit with more certainty than before;
4. Is likely to suffer from issues of endogeneity, that will bias the outputs in unknown directions;
5. Has no background (internal and external) validity;
6. Is likely to suffer from Hypothetical Bias issues; and
7. Is likely to produce results, which is simply simulation noise as opposed to representing respondent's actual WTP outcomes.

It is the therefore the view of the author that original document and responses provided by Melbourne Water to the questions provided do not induce any level of confidence as to the robustness of the process. Indeed, the responses provide even less confidence in the validity and robustness of original work than previously indicated.

6. References

Cummings, R.G. and Taylor, L.O. (1999) Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method, *American Economic Review*, 89(3), 649-665.

De-Magistris, T., Gracia, A. and Nayga Jr, R.M. (2013) On the use of honesty priming tasks to mitigate hypothetical bias in choice experiments, *American Journal of Agricultural Economics*, 95(5), 1136-1154.

Ding, M., Grewal, R. and Liechty, J. (2005) Incentive-aligned conjoint analysis, *Journal of Marketing Research*, 42(1), 67-82.

Gedenk, K. Hensel-Borner, S. and Sattler, H. (1999) Bandbreitensensitivitat von Verfeahren zur Prafernezmessung, working paper University of Jena.

Green, P.E. and Srinivasan, V. (1990) Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice, *Journal of Consumer Research*, 5, 103-123.

Green, P.E. Goldberg, S.M. and Montemayor, M. (1981) A Hybrid Utility Estimation Model of Conjoint Measurement, *Behavioral Science*, 17, 288-299.

Hensel-Borner, S. and Sattler, H. (1999) Validitat der Customized Computerized Conjoint Analysis (CCC), working paper University of Jena.

Herriges, J., Kling, C., Liu, C.-C. and Tobias, J. (2010) What are the consequences of consequentiality? *Journal of Environmental Economics and Management*, 59(1), 67-81.

Jacquemet, N., Joule, R.-V., Luchini, S. and Shogren, J.F. (2013) Preference elicitation under oath, *Journal of Environmental Economics and Management*, 65(1), 110-132.

Ladenburg, J. and Olsen, S.B. (2014) Augmenting short Cheap Talk scripts with a repeated Opt-Out Reminder in Choice Experiment surveys, *Resource and Energy Economics*, 37, 39-63.

Louviere, J.J., Flynn, T.N. and Carson, R.T. (2010) Discrete Choice Experiments Are Not Conjoint Analysis, *Journal of Choice Modelling*, 3(3), 57-72.

Nitzsh, R.V. and Weber, M. (1993) The effect of Attribute Ranges on Weights in Multiattribute Utility Measurements, *Management Science*, 29, 937-943.

Prelec, D. (2004) A Bayesian truth serum for subjective data, *Science*, 306(5695), 462-466.

Srivansan, V. (1988) A Conjunctive-Compensatory Approach to the Self-Explication of Mutilattribute Preferences, *Decision Sciences*, 19, 295-305.

Whittington, D., Smith, V.K., Okorafor, A., Okore, A., Liu, J.L. and McPhail, A. (1992) Giving respondents time to think in contingent valuation studies: a developing country application, *Journal of Environmental Economics and Management*, 22(3), 205-225.

7. 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.

8. 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