survey engine
decision modelling technology

Valuing Victoria's Heritage

Methodology

Version
30.07.2017 Version 3
Authors
Ben White
Mikołaj Czajkow ski

Abstract

This document outlines the proposed methodology for the project 'Valuing Victoria's Heritage – the value and benefits of Victoria's historic heritage assets', commissioned by the Victorian State Government Department of Environment, Land, Water and Planning (DELWP).

Two attached documents complement this methodology: a) Value of Heritage Literature Review conducted by SGS and SurveyEngine and b)the Heritage Attribute Matrix which defines the proposed classes, attributes and their levels for use in the Discrete Choice Experiment.

The scope of this document is limited to valuing assets currently on the Victorian Heritage Register or likely to be included in the future. Valuation of pre-settlement Aboriginal heritage is explicitly not included.

The methodology proposed is of an academic grade. The analysis and interpretation of the final results are expected to also be of the same level. Efforts have been made to describe the method to be comprehensible by a lay-person. However, it is acknowledged that there is a practical limit to a full understanding by the intended audience of this research.

To overcome this limit and provide confidence in the research, it is recommended that this methodology and results undergo an academic peer review process for publication in leading international scientific journals with relevant acknowledgement of the contributors. A successful publication would increase the gravity of the Heritage valuations and citation value for policy makers, planners and other users of the research.

1 CONTENTS

1 Summary of the Methodological Approach

This study aims to develop estimates of Willingness-to-Pay (WTP) for heritage assets in Victoria, such that

(1) they are updated with respect to the existing estimates (Allen Consulting Group, 2005),
(2) they can be used for past and potential future estimates for public policy, including cost benefit analysis of individual development decisions, and
(3) they are derived in accordance with the state-of-the-art methodology that assures their validity.

With respect to (1), this approach involves not only developing a new study that builds on the methodological approach selected to satisfy the project requirements, but also replicates the 2005 ACG study, to investigate the extent of the change in welfare estimates that could have taken place in the the last decade as a result of changes in public preferences and affluence, achievements of past conservation policies, and the availability of substitutes.

Regarding (2), a review of the project requirements, existing materials and literature review was done collectively by the SurveyEngine team. The point of departure for the new study was provided by real and hypothetical cases of the usage of the valuation results. The valuation instrument developed in consultation with SGS Economics and Planning to understand relevant inputs for valuation.

Key to the approach is acknowledgement that different types of heritage assets have different properties, threats, protection types and development options. Separating the heritage assets by type allows departure from the 'one size fits all' problems with the Allen Consulting Group (2005) study results, particularly when comparing heritage objects to heritage buildings and sites. Separating the heritage assets by types means only relevant attributes need be tested. This means that there are less constraints on the attributes selection as they don't need to be generally applicable to every type of heritage asset. Furthermore the choice tasks would be more meaningful and credible for respondents and the results more useful for users of the final results.

Another key change with respect to the Allen Consulting Group (2005) study was valuing protection of individual heritage assets, rather than evaluating policies that simultaneously target thousands of them. This is because incremental valuation, for example reflecting particular protection of an additional heritage asset, of a given type and set of characteristics, is a more adequate approach for providing support to most policy decisions (e.g., extending protection to an additional asset, or allowing for a specific development of a building that could have some cultural heritage value). We argue that such a bottom-up approach is more appropriate than the top-down approach, in which conservation as a whole is being valued and used to infer values resulting from marginal changes in the portfolio of assets.

With respect to methodology used by our approach, it is clear that studies that have invested considerable time and effort into understanding what people believe, and satisfies certain design conditions (e.g., presenting a credible choice scenario with a well-defined good and a coercive payment mechanism, highlighting survey consequentiality) generally appear to produce results that are well-behaved, which is not necessarily the case otherwise (Carson and Czajkowski, 2015). We note that Stated Preference (SP) valuation is constantly being developed and an accumulating body of evidence leads to establishing new recommendations for the state-of-the-art (see e.g., Johnston et al., 2017). Following these recommendations allows for SP estimates to be

more robust and is also crucial for their validity and usability in public policy or judicial processes (Cameron, Cragg and McFadden, 2013). Reviewing all guidelines and requirements for the development and administration of SP studies necessary for the validity of resulting WTP estimates is too extensive a task to include it here. Instead, in what follows we discuss the four main approaches used to evaluate the validity of SP methods that we incorporate in the design and the analysis stages of our study. They are convergent, construct, content, and criterion validity, also known as CCCC framework (Bishop and Boyle, 2017).

The main points:

- Identical replication of the 2005 ACG as a separate standalone research project
- A new 2017 study will be conducted separately and differs in many respects
- The 2017 study will value separate heritage classes and individual Heritage assets
- Heritage classes will be differentiated by Heritage Sites, Cultural Landscape, Historical Sites and Heritage objects
- The CCCC framework will be used to establish research validity

Convergent validity

Convergent validity verifies the correspondence between WTP estimates derived from a stated preference study with some other measure of the same theoretical construct, but obtained by different studies or methods, including by indirect valuation methods. Therefore, these tests usually compare value estimates from a stated preference study with their counterparts from other stated or revealed preference approaches such as the contingent valuation, discrete choice experiments, hedonic pricing or the travel cost method (Zawojska and Czajkowski, forthcoming).

Due to the pioneering character of this study there are no estimates that would provide exact reference for convergent validity test of our results. Some insights can be gained by comparing our results with those of the earlier study aimed at valuing conservation of the heritage places in Victoria (Allen Consulting Group, 2005) as well as with the other valuation studies of cultural goods (Value of Heritage Literature Review, 2017). We note, however, that due to differences in the definitions of the valued goods, extent of the change, location, target population, methodology and other aspects the results are expected to vary between these studies and the comparisons can only give overall impression if the range of the WTP estimates are reasonable.

Results will be compared with other similar studies to verify results are reasonable.

Construct validity

The tests of construct validity assess stated preference methods' validity by verifying the consistency of stated WTP values with predictions derived from the consumer demand theory, such as sensitivity to price changes, income levels, and other economic variables, which can confirm that responses to stated preference surveys are not random.

Construct validity requires well-defined theory as a reference point for the comparison of theoretical predictions and stated values - this is somewhat problematic because the neoclassical demand theory provides the basis for the comparisons in some cases only. In addition, even in real markets, consumers are observed not necessarily to behave in line with this theory (Poe, 2016).

Given that this theory does not appear to capture all aspects central to consumers' choices, it should be complemented by other concepts such as those provided by behavioral economics. However, because of the constant development of behavioral economics, the discrepancy between theoretical predictions and SP values may demonstrate not the SP methods' lack of validity, but rather the incompleteness of the theory. Although construct validity constitutes an important component of validity, it is questionable whether the current state of the economic theory development enables conducting a proper test of SP methods' construct validity, because the rational choice theory does not provide a sufficient reference point for testing.

We address construct validity of our approach by evaluating if the results are in line with the current state of the economic theory, wherever the theory offers clear predictions that are also satisfied for consumers' behavior in markets. Specifically, by accounting for various sources of respondents' observed preference heterogeneity we are able to test if individuals' responses are sensitive to the cost and other attribute levels presented (e.g., distance, heritage rating), and test if their WTP is sensitive to their income and the availability of substitutes.

Results will be compared against economic theory to verify validity. We expect:

- a decrease in preference for increased costs such as Tax Levy,
- a decrease in preference for increased distance from the asset and
- a reduced WTP for respondents from lower socio-demographic segments

Content validity

Content validity focuses on whether a survey applies state-of-the-art recommendations of best design practices. The interpretation often relies on the evaluator's experience and subjective opinion, however, some general recommendations follow from a number of seminal works developed for contingent valuation (e.g., Arrow et al., 1993; Mitchell and Carson, 1989; Bateman et al., 2004; Champ, Boyle and Brown, 2004), recent state-of-the-art methodology summaries (Johnston et al., 2017) and more general recommendations regarding the construction of surveys (Dillman, Smyth and Christian, 2008).

This criterion is satisfied by following the state-of-the-art recommendations. Specifically:

- A clear description the baseline conditions, the mechanism of change, and the changes to be valued and elicit evidence that these pieces of information are understood, accepted, and viewed as credible by respondents.
- A qualitative (elicit open comments at the development stage, confirm understanding and credibility) and quantitative (a dedicated pilot study) pretest of the instrument and review by domain experts
- Responses are elicited to support debriefing questions.

- The change being valued is described (the information content of valuation scenarios) in a way that is in line with how respondents tend to perceive the good.
- An experimental design makes use of information from prior empirical research and is pretested considering both statistical efficiency and respondents' cognitive abilities and attention budgets The design employs constraints on implausible attribute levels and combinations, is robust to alternative model specifications and considers the levels chosen for each attribute to influence design properties.
- The hypothetical situation setting review will be in accordance with scientific ethics (the survey design procedures avoid deception).
- A sample is drawn from a known frame that is consistent with the population for which values are to be estimated and respondents are randomly selected from the sample frame using an explicit sampling procedure.
- Willingness-to-Pay will be used as the most appropriate welfare measure from a conceptual perspective in our case.
- An incentive-compatible response format will be used.
- Valuation scenarios and valuation questions will be designed to enhance incentive compatibility and to encourage truthful responses.
- A payment vehicle is selected such that it is realistic, credible, familiar and binding for all respondents to as great an extent as possible and to ensure that payments are viewed as fixed and non-malleable.
- The questionnaire will include auxiliary questions to enhance the validity of the study and to evaluate the validity of responses to the value elicitation questions.
- Utility-theoretic, behavioral, statistical, and other assumptions underlying model selection and specification for data analysis are made explicit.
- Analysis of the data will allow for both observed and unobserved preference heterogeneity and will consider the relevance of this heterogeneity for the use of study results to support decision making.
- Data analysis will include both (a) the simplest, most parsimonious specifications and (b) more-complex models that impose additional investigator assumptions on the structure of responses.
- Undesirable response anomalies will be analysed for potential influence and data analysis will investigate these anomalies to determine whether they significantly affect responses (for example, we will consider whether protest or outlier responses are influential).
- Reported welfare estimates will, at a minimum, include estimates of central tendency and dispersion. Methods used to calculate welfare measures will be transparent and will ensure that estimates are theoretically and statistically well defined.
- Analysis of the data will include a set of core internal validity assessments.
- Full documentation of study design, implementation, analyses, and results.
- With the client's consent, the study would undergo a peer review process and will be undertaken and results published in high quality scientific journals. This will strengthen the value of the results in policy making.

- Formal scientific and technical measures will be taken to ensure the validity of the scientific validity of the study.
- The all-encompassing measure to ensure validity of this method and results is submission of the final report for peer academic review and publication in a scientific journal.

Criterion validity

Criterion validity tests investigate consistency of preferences stated in a survey with another measure which is thought to truly express preferences, or at least to be a good proxy of true preferences. This measure is a so-called criterion and provides a reference point for comparison. Such criteria are usually derived from real (field) or simulated (laboratory) market data. Unlike convergent validity tests, this approach does not utilise revealed-preference-based estimates as a benchmark, but typically elicits consumers' preferences for the same or very similar good in both hypothetical and actual payment settings. Typically, a reference point for validity verification is provided by real-payment-based estimates. However, what has been shown consistently is that a criterion standard is rare if not non-existent, because actual behavior may also suffer from various systematic errors (e.g., resulting from behavioural effects).

As a result, almost all comparisons of stated preference estimates to some other type of measurement should be treated as tests of convergent validity (Mitchell and Carson, 1989).

Criterion validity tests remain experimental in their nature, as they are only possible when equivalent estimates exist - such as when stated and actual decisions are compared in a lab experiment, or when there is an opportunity to compare stated responses and actual voting. While tests of content and construct validity may be conducted using data from a single study, criterion validity tests require data from two or more parallel studies or data sources. Hence, they are not feasible as part of most analyses (if criterion values were available, there would be no need for stated preference analyses to inform decisions). In addition, many researchers believe that almost all comparisons of stated preference estimates to some other type of measurement should be treated as tests of convergent validity (Mitchell and Carson, 1989; Bateman et al. 2002; Carson et al. 2014; Ryan et al. 2016). For these reasons, state-of-the-art recommendations do not include criterion validity among the suggested lines of validity investigations (Johnston et al., 2017). In our case, it is not possible to infer about criterion validity of our approach due to the lack of equivalent valuations using real or simulated data, nevertheless real-world market data will be sought to validate that the results are at least reasonable.

Independent real-world market data for heritage valuation will be sought for comparison to validate the results are reasonable.

2 Experiment Instrument

In each contingent valuation survey, respondents are asked to imagine a situation in which they choose one of the available alternatives regarding the good that is being valued. In alignment with the Lancasterian perspective of utility (Lancaster, 1966), every good is described in terms of a collection of its characteristics (attributes).

The selected combinations of levels of these attributes include the alternatives that are presented to respondents, who will be asked to choose the alternative that they consider the best (the most preferred). The choices observed in these hypothetical situations make it possible to apply statistical methods to estimate the utility function parameters that are related to the specific attributes of a good and to formally model consumer preferences.

Therefore, it is possible to evaluate changes in consumer welfare in the case of implementing of a particular scenario (e.g., providing a new public good) and predict consumers' behavior that is related to new goods or alternatives. In addition, identifying the marginal rates of substitution between particular characteristics of a good (including the pecuniary attribute, e.g., the cost of provision) makes it possible to identify respondents' WTP for non-market goods and their characteristics.

2.1 Defining the hypothetical situation

The context sets up a credible situation that a respondent can understand in order to make a choice. The context should be closely related to what is being attempted to model and measure – in this case WTP for various heritage attributes.

The state-of-the-art recommendations underline that the hypothetical situation should be constructed in such a way that respondents' choices are incentive compatible. We discuss satisfying this criteria below.

Incentive compatibility

Incentive compatibility is a concept that implies that the optimal strategy for a respondent is to answer truthfully by revealing their actual preferences. Based on a review of literature (e.g., Carson and Groves, 2007) one can summarise that for the stated preference study to be incentive compatible, i.e., to reveal respondents' true preferences, the necessary (albeit not necessarily sufficient) conditions are:

1. respondents should correctly understand and answer the question being asked including the requirement that the good(s) being valued, including the different attribute levels and cost, are seen as plausible; (Carson and Hanemann, 2005);
2. respondents need to see the survey as consequential, i.e. their responses should be viewed as potentially influencing the supply of a public good, and agents must care about these outcomes (Vossler, Doyon, and Rondeau, 2012; Vossler and Watson, 2013);
3. the payment has to be coercive, i.e. the payment vehicle must be able to impose costs on all agents if the government undertakes the project (Carson and Louviere 2011);
4. following from the Gibbard-Satterthwaite theorem, the message space of a choice question cannot be larger than binary without restricting the space of allowable preference functions, i.e. a binary choice is the only elicitation format that has a potential to be incentive compatible;

5. the survey should be seen as a 'take-it-or-leave-it offer', so that agents do not see their decisions as influencing any other offers that may be made (Carson, Groves, and List 2014).

Some of the above conditions (correct understanding, take-it-or-leave-it character) can only be evaluated via researcher's experience and careful qualitative testing and refinement of the survey instrument. Others (consequentiality, coercive payment mechanism) are possibly easier to satisfy, although they too require making sure that what is written in a survey script, and what is read and understood by respondents is the same thing. Finally, some authors impose rather stringent conditions on acceptable elicitation formats (Carson and Louviere 2011). While certain elicitation formats should probably be avoided altogether, in the light of the bias vs. efficiency trade-off, it remains an empirical question to what extent moving away from these incentive compatibility requirements actually biases results.

The validity of Stated Preference (SP) methods has been thoroughly investigated, particularly because the empirical evidence is often contradictory. Some studies report significant differences between stated and true preferences, whereas others find no significant difference. Recently, Zawojska and Czajkowski (2015) have critically re-evaluated this evidence. By reviewing the four main types of validity tests – content, construct, convergent, and criterion validity – they argue that comparing SP-based estimates with corresponding criterion measures is the most adequate approach to verify how well SP-based estimates reflect true preferences. By classifying the empirical evidence with respect to whether it (1) deals with private or public goods, (2) uses a coercive or voluntary payment mechanism, (3) can be perceived by respondents as consequential, and (4) uses a single binary choice format, they identified studies that provide meaningful results in terms of providing conditions in which rational respondents can be expected to answer in line with their true preferences. The results of such studies consistently point to the validity of stated preferences under such conditions. When the available evidence is limited only to studies that satisfy the requirements listed above, the evidence becomes univocal – hypothetical bias can be avoided. This conclusion is very encouraging for SP methods, although it obviously comes with many requirements for the design and administration of future SP studies (Hanley and Czajkowski, 2017).

Choice Context

In our case, we propose that the goals of the study can be credibly achieved by setting the survey context as being advisory for the authorities responsible for conservation of heritage sites. Respondents to the survey will be informed that the results may have a direct impact on the types and levels of heritage protection in the state of Victoria, including the personal cost to them. As a result, their personal responses will have an impact on the public policy that directly affects them.

In order to calculate WTP from Choice Models, it is necessary for at least one of the attributes be a monetary one. This allows for calculating marginal rate of substitution of the changes in any non-monetary attributes to the monetary variable - the trade-off respondents are willing to make in terms of money for policy improvements. This is how their WTP is inferred.

It s proposed to a use one-time tax levy as payment vehicle - as was used in the Allen Consulting Group (2005) study.

An artifact of standard implementations of DCE's is that respondents are asked to evaluate more than one choice scenario, typically up to 8 scenarios for each choice experiment. This is done principally for efficiency of data collection and to maximise the amount of data collected for each respondent.

Because of this it is important that the context of the tax levy is clear for a consistent comprehension by the respondents and interpretation during analysis.

In our case, respondents will be asked to treat each choice task independently. Each choice task would be presented with a choice between a (systematically generated) protection policy (and associated tax levy) and would answer in each case whether they would vote for or against that policy As a result, the tax levy is:

- independent from one scenario to the next (specifically non cumulative) and
- the choice to support a protection measure (at a cost) is akin to a 'vote' and so it does not automatically imply implementing conservation policy and the bearing associated cost, that would limit one's budget in the following choice tasks.

Estimated WTP results should be interpreted accordingly, namely that they reflect a respondent's maximum willingness to pay for the first heritage protection extension of a particular kind. WTP for subsequent extensions is likely to be smaller due to income effects (reduced budget) and preference changes (that might depend on how much is already protected).

In summary, a respondent is told that the survey results will be used to inform the decisions of the authorities responsible for the conservation of heritage assets in Victoria. His or her responses will help decide if it is worth to implement additional protection measures such as to extend protection to an additional heritage building or asset of particular characteristics at a the cost presented in the survey. Further that this cost of protection is eventually covered from his or her taxes.

Such an instrument is consequential, as the respondent knows his or her responses will be taken into account by policy makers and influence policy decisions regarding conservation of Victorian heritage and potential new tax levies. At the same time, the payment mechanism (one time tax levy) is coercive - if the policy goes through respondents may be requested to pay increased taxes. Awareness of one's responses increasing or decreasing the probability of making a particular policy decision makes the response situation incentive compatible.

An example of how a heritage site valuation context may be presented is given below:

Example Site Valuation Task

On the next screen we are going to show you a heritage site that is under review for changes to its conservation status under a local planning scheme.

We would like you to imagine that you are asked to be involved in the decision to protect the site.

Your task is simple, carefully review the site, the proposed conservation amendments and the additional taxation cost and choose whether you would support the conservation measures.

A decision to not support the conservation measures means that the site would have no special status as a Heritage site, and that it may be altered or possibly demolished within the normally applicable building regulations

For physical Heritage objects, the valuation task would be similar:

Example Object Valuation Task

On the next screens we are going to show you a heritage object that is under review for changes to its conservation status.

We would like you to imagine that you are asked to be involved in the decision to control how this object may be protected from custodial ownership or alteration.

Your task is simple, carefully review the object, the proposed conservation amendments and the additional taxation cost and choose whether you would support the conservation measures.

A decision to not support the conservation measures means that the object would have no special Heritage status, and that it may be altered, transferred or sold without conservation controls.

The instrument would have separate sections for each of the heritage classes, with one section being presented after each other.

2.2 Heritage Classes

Heritage Classes were selected by review of the register and grouping by type of asset such that the majority of attributes were meaningful to compare within a group but not across a group.

Heritage Site:	Typically a single building
Cultural Landscape:	An area defined by features evoking a common theme
Historical Site:	Typically places commemorating a specific event or events from history
Heritage Object:	A relocatable object with potential heritage value such as a flag or portrait.

2.3 Attributes and Levels

A matrix of attributes, and their possible levels, used in the experiment were obtained from various sources

- the attributes and levels from the ACG 2005 report,
- examination of the Victorian Heritage Register,
- examination of economic valuation use-cases with SGS,
- discussion and input from with the project stakeholders and
- the selected payment vehicles from the ACG 2005 report and similar WTP evaluations

The process for developing the attributes and their levels was performed using the following steps::

1. Using the sources above, attributes and levels were created per heritage class. This allowed focus on the attributes relevant to each class. Attributes were required to be independent of each other and their levels should span all conceivable values of interest.
2. A second stage looked at generalising attributes discovered in one class for generalisation across all classes.
3. Heritage assets were constructed randomly from the attribute matrix to verify that they were comprehensible and credible.
4. A final verification stage involved repeated selection of assets from a large sample of the heritage register and examining whether the object could be represented uniquely by the attribute and level structure.

The result of this process was documented in the Heritage Attribute Matrix Spreadsheet attachment.

2.4 Presentation Format

Valuation scenarios were drawn randomly from the attribute matrix as sample content, yielding examples as below:

Attributes	**Heritage Site**
Age	Interwar Period 1919-1945
Type	Industrial Warehouse
Condition	Poor condition
Heritage Rating	National Significance
Distance from You	3 to 5 km
Protection Type	no protection from demolition
Built Environment Context	Sympathetic External Development
Visitation Management	yes
Traffic Management	none
Noise Management	none
Public Access	Public Access - free
Protection Tax Levy (per year)	$150

Reviewing the register and the typical presentation format therein, the above scenario was reorganised into a more familiar and readable format as below and a typical image of the asset provided.

Industrial Warehouse
Period: 1900-1945 In Poor Condition
Of National Significance
Approximately 4 km from you
Currently free Public Access

Proposed Protection Control
No protection from demolition
Sympathetic development only within area
Control of visitation
No control of traffic
No control of noise
Tax Levy for this Protection: $150

Would you vote for or against extending protection to such a building, considering the associated cost to your household? ☐Yes ☐No

2.5 Experiment Design

There are many ways in which the combinations of the levels of attributes can be combined to create alternatives for each choice situation. Because the number of attributes (and their levels) is too large to take into account all the possible combinations in a study, a so-called fractional design is applied. This includes only selected combinations of attributes that reduce the number of unique combination while preserving information properties allowing the independent effects of each attribute to be estimated.

The design will follow two concurrent approaches to experimental design using a split sample treatment. The first one will follow a standard so-called orthogonal optimal in a difference fractional factorial design. It aims at guaranteeing orthogonality of attribute levels while keeping balance concerning the incidence (Street, Burgess and Louviere, 2005; Street and Burgess, 2007).

The second approach will follow a different suggestion from contemporary literature, which shows that it is possible to construct non-orthogonal research designs, which make it possible to obtain more information from each consumer's choice (Sándor and Wedel, 2001). In the case of these so-called efficient designs, instead of preparing orthogonal sets of levels of attributes for each choice situation presented to respondents, sets are generated in such a way as to minimise the determinant of the asymptotic variance-covariance matrix of parameters (the so-called D-error), with the assumption of certain initial estimates (priors) concerning the parameters of the utility function of a respondent a priori (Huber and Zwerina, 1996; Scarpa and Rose, 2008).

Because the parameters in the utility function are usually other than zero, orthogonal designs are not efficient. Researchers usually have some expectations concerning the values of parameters (or at least their characters), for example a decreasing preference for price. This allows researchers to generate designs of choice situations which reveal more information and therefore improve the statistical features of the final model or make it possible to decrease the sample necessary to estimate the model at a given level of significance. This means either less sample is required, more accurate estimates can be produced for the same sample or some optimal balance of both. A common practice is to carry out preliminary research making it possible to obtain better estimates of the parameters for the main study.

Finally, the state-of-the-art in designing choice situations is the application of Bayesian efficient research designs, which take into account the uncertainty related to the initial estimates of parameters through allowing these estimates to have the form of random variables with certain probability distributions (Sándor and Wedel, 2001). In this case, evaluating the value of the determinant of the asymptotic variance-covariance matrix of parameters requires the integration based on simulation, since it is not possible to determine it analytically. Still, the added value of this approach involves taking into account the uncertainty related to initial estimates through the application of parameter distributions for the most probable values.

This method of generating research designs for research on conditional choices makes it possible to decrease the uncertainty associated with parameter estimates, ceteris paribus. As a result, the application of Bayesian initial estimations allows better estimates to be obtained.

The experimental design will randomise the order of sets of questions for each heritage type. Instead of using fixed blocks we select choice tasks randomly, ensuring that each choice task is used an almost equal number of times (Czajkowski and Budzinski, 2016). The Bayesian efficient design will be updated after the pilot and twice throughout the data collection to account for the information already available (i.e., to take better priors into account). The Bayesian priors will be normally distributed, with means given by the MNL model estimated on the available data, and standard deviations equal to 0.25 to 0.5 of the means (with an absolute minimum for means which were relatively close to 0), representing our certainty level in these estimates. The experimental design will be optimized for d-efficiency of the MNL model (Bliemer and Rose, 2010).

3 Additional Data Collected

In addition to the choice data collected for modelling, the research data-set will be augmented by additional data types.

3.1 Survey Meta Data

This includes starting time and duration of time spent on each page

- start-time
- end time
- duration per page

3.2 Screening Data

Data required to screen out ineligible respondents such as ages under 18 or place of residence outside of Victoria. It is also proposed that respondents participation be voluntary. This may also include data to detect panel fraud including non-Australian points of origin through use of VPNs and TOR exit nodes and automated data entry agents such as web-bots.

3.3 Demographics Data

Data on individuals will be collected to allow representative sampling. This includes

- Gender
- Age
- Postcode
- Education
- Citizenship
- Income

3.4 Usage and Attitudinal Data

Replication some of the ACG usage and attitudinal questions (UA) including

- Attitudes to levels of current heritage protection
- Attitudes to cost of heritage protection
- Heritage statement agreement
- Ranking of Heritage spending
- Involvement in heritage activities, groups, governance and tourism

3.5 Open-ended comments

Free text input on aspects of the survey including

- Difficulty and comprehension
- Technical issues
- Un-prompted views on heritage and conservation

This data will be used variously and interchangeably for ensuring only valid respondents, sample representativeness but also for the opportunity to gain additional insights into usage, visitation and attitudes to heritage for supporting analysis of WTP.

4 Sampling and Fieldwork

Fieldwork will proceed in 2 phases:

4.1 Pilot Study

A pilot of approximately 200 Victorians will be conducted using the experiment instrument to allow direct review of the data and respondent feedback. This risk management measure allows the opportunity for review of initial models, feedback and to make for minor corrections before committing to the full data collection.

This short run study will also be used to evaluate dominance effects and will allow review of the potential dominance and correction of the selected payment vehicle and distance attributes ranges.

Optimisation of the experimental design may also be be performed from analysis of the pilot study data.

4.2 Main Study

A main data collection phase of 1,200 Victorian residents of 18 years or older will be conducted using a professional Market Research panel.

4.3 Sample Frame & Representativeness

Sample quotas will be established to ensure the sample proportions of

- Age,
- Gender and
- Regional Location

match those of the most recent Census

4.4 Experiment Sampling

The order of the heritage classes will be randomised by respondent to avoid any possible ordering biases.

Experiment treatments for each class will be drawn randomly, without replacement, for all experiments. This will eliminate ordering bias while also ensuring an even allocation of experiment treatments.

4.5 Privacy and Data

Respondent privacy and data management will be as per the standard SurveyEngine usage and privacy terms at

http://surveyengine.com/terms

http://surveyengine.com/privacy

5 Analysis

Respondents' utility function parameters will be modelled using the stated choices they made in the discrete choice experiment component of the survey. We utilised the mixed logit model (MXL, McFadden and Train, 2000; Hensher and Greene, 2003) which allows for incorporation of unobserved preference and scale heterogeneity (Hess and Train, 2017).

Formally, the discrete choice data is modelled using the random utility theory (McFadden 1974). This assumes that the utility an individual receives from an alternative he chooses depends on observed characteristics (attributes) and unobserved idiosyncrasies, which is represented by a stochastic component. Individual i's utility from choosing alternative j in situation t can be expressed as:

$$V_{ijt} = \mathbf{x}_{ijt}\boldsymbol{\beta}_i + e_{ijt}.$$

The utility expression is separable in the observed choice attributes $\mathbf{x}_{ijt}$ and e_{ijt} being the stochastic component allowing for unobservable factors that affect individuals' choices. The parameters $\boldsymbol{\beta}_i$ represent individual-specific taste parameters associated with marginal utilities of the choice attributes, allowing for heterogeneous preferences among the respondents. The multivariate (parametric) distribution of these parameters in the sample is $\boldsymbol{\beta}_i \sim f(\mathbf{b}, \boldsymbol{\Sigma})$ where $\mathbf{b}$ is a vector of sample means and $\boldsymbol{\Sigma}$ is a variance-covariance matrix. A convenient way of accounting for preference differences associated with accessing information is $\boldsymbol{\beta}_i \sim f(\mathbf{b} + z_i\boldsymbol{\delta}, \boldsymbol{\Sigma})$, where z is a binary indicator for accessing information and $\boldsymbol{\delta}$ is a vector of its estimated attribute-specific effects.[1]

The stochastic component of the utility function (e_{ijt}) has an unknown, possibly heteroskedastic variance $\left(\text{var}\left(e_{ijt}\right) = s_i^2\right)$. The model is usually identified by normalizing this variance, making the error term $\varepsilon_{ijt} = e_{ijt} \cdot \frac{\pi}{\sqrt{6}s_i}$ identically and independently, extreme value type 1 distributed with a constant variance $\text{var}\left(\varepsilon_{ijt}\right) = \pi^2/6$, leading to the following specification:

$$U_{ijt} = \sigma_i \mathbf{x}_{ijt}\boldsymbol{\beta}_i + \varepsilon_{ijt},$$

where $\sigma_i = \pi/\sqrt{6}s_i$ is the 'scale' parameter. Due to the ordinal nature of utility, this specification still represents the same preferences for individual i. Note that since the scale and preference parameters enter the model as

[1] The specific distributions must be assumed by the modeller; it is typically done based on model fit.

multiplication they are not separately identifiable. This does not restrict applicability of the model, because utility function parameters do not have absolute scale and can only be interpreted in relation to 0 and each other.

Finally, given that we are interested in marginal rates of substitution with respect to the monetary attribute p it is convenient to introduce the following modification, which is equivalent to using a money-metric utility function (also called estimating the parameters in WTP space; Train and Weeks, 2005):

$$U_{njt} = \alpha\left(p_{njt} + \boldsymbol{Y}_{njt}\boldsymbol{b}\right) + e_{njt} = \alpha\left(p_{njt} + \boldsymbol{Y}_{njt}\ \right) + e_{njt}\ .$$

In this specification, the vector of parameters $\boldsymbol{\beta} = \boldsymbol{b}/\alpha$ can be directly interpreted as a vector of implicit prices (marginal WTPs) for the non-monetary attributes $\boldsymbol{Y}_{njt}$, facilitating interpretation of the results.

The model is estimated using maximum likelihood techniques. An individual will choose alternative j if $U_{ijt} > U_{ikt}$, for all $k \neq j$, and the probability that alternative j is chosen from a set of J alternatives is given by:

$$P(j|J) = \frac{\exp\left(\sigma_i \boldsymbol{\beta}_{ijt\ \ i}\right)}{\sum_{k=1}^{J} \exp\left(\sigma_i \boldsymbol{\beta}_{ikt\ \ i}\right)}\ .$$

There exists no closed form expression of , but it can be simulated by averaging over D draws from the assumed distributions (Revelt and Train 1998). As a result, the simulated log-likelihood function becomes:

$$\log L = \sum_{i=1}^{N} \log \frac{1}{D} \sum_{d=1}^{D} \prod_{t=1}^{T_i} \sum_{k=1}^{J} y_{ikt} \frac{\exp\left(\sigma_i \boldsymbol{\beta}_{ijt\ \ i}\right)}{\sum_{k=1}^{J} \exp\left(\sigma_i \boldsymbol{\beta}_{ikt\ \ i}\right)}\ .$$

where y_{ikt} is a dummy taking the value 1 if alternative k is chosen in choice situation t, and zero otherwise. Maximising the log-likelihood function in gives estimates for the parameters. In the modelling, the cost variable was continuous and other attributes were dummy-coded. The parameters of alternative specific constants (ASC) and all other attributes, including the cost, will be modelled as random.

Econometric models estimated using maximum simulated likelihood are known to be relatively sensitive to starting values, optimisation techniques and selection of convergence criteria. Our model is no exception in this respect and to make sure we reached the global maximum in optimization, we will use different optimisation algorithms, derive gradients analytically and use multiple starting points. In addition, since using longer low-discrepancy sequences (as opposed to shorter sequences or using pseudo-random draws) is found to facilitate reaching the global optimum or revealing identification problems (Chiou and Walker, 2007; Czajkowski and Budziński, 2017) in simulation of the log-likelihood function, we will use 10,000 scrambled Sobol draws.

6 References

Allen Consulting Group, 2005. Valuing the Priceless: The Value of Heritage Protection in Australia, Research Report 2. Heritage Chairs and Officials of Australia and New Zealand, Sydney.

Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R., and Schuman, H., 1993. Report of the NOAA Panel on Contingent Valuation. Federal Register, 58:4601-4614.

Bateman, I., Munro, A., Rhodes, B., Starmer, C., and Sugden, R., 1997. A Test of the Theory of Reference-Dependent Preferences. The Quarterly Journal of Economics, 112(2):479-505.

Bateman, I. J., Carson, R. T., Day, B., Haneman, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Ozdemiroglu, E., Pearce, D. W., Sugden, R., and Swanson, J., 2002. Economic Valuation with Stated Preference Techniques: A Manual. Edward Elgar, Northampton, Massachusetts.

Bliemer, M. C. J., and Rose, J. M., 2010. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. Transportation Research Part B: Methodological, 44(6):720-734.

Bishop, R. C., and Boyle, K. J., 2017. Reliability and validity in nonmarket valuation. In: A primer on nonmarket valuation, P. A. Champ, K. J. Boyle, and T. C. Brown, eds., Springer, Amsterdam.

Cameron, L., Cragg, M., and McFadden, D., 2013. The Role Of Conjoint Surveys In Reasonable Royalty Cases. Law360, October 16, 2013

Carson, R. T., 1997. Contingent Valuation and Tests of Insensitivity to Scope. In: Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods, R. Kopp, W. Pommerhene, and N. Schwartz, eds., Kluwer, Amsterdam.

Carson, R., and Groves, T., 2007. Incentive and informational properties of preference questions. Environmental and Resource Economics, 37(1):181-210.

Carson, R. T., Groves, T., and List, J. A., 2014. Consequentiality: A Theoretical and Experimental Exploration of a Single Binary Choice. Journal of the Association of Environmental and Resource Economists, 1(1/2):171-207.

Carson, R. T., and W. M. Hanemann. 2005. "Contingent Valuation." In Handbook of Environmental Economics, edited by K. G. Mäler and J. R. Vincent. Amsterdam: Elsevier.

Carson, R., and J. Louviere. 2011. "A Common Nomenclature for Stated Preference Elicitation Approaches." Environmental and Resource Economics no. 49 (4):539-559.

Champ, P. A., Boyle, K. J., and Brown, T. C., 2017. A Primer on Nonmarket Valuation. Springer, Amsterdam.

Chiou, L., and Walker, J. L., 2007. Masking identification of discrete choice models under simulation methods. Journal of Econometrics, 141(2):683-703.

Czajkowski, M., and Budziński, W., 2017. Simulation error in maximum likelihood estimation of discrete choice models. Paper presented at the 6'th International Choice Modelling Conference, Cape Town.

Czajkowski, M., and Budziński, W., 2016. Choice task blocking and design efficiency. Paper presented at the 5'th Workshop on Discrete Choice Modelling, Warsaw, available from http://czaj.org/pub/presentations/Czajkowski_2016-10-06b.pdf.

Czajkowski, M., and Hanley, N., 2009. Using Labels to Investigate Scope Effects in Stated Preference Methods. Environmental and Resource Economics, 44(4):521–535.

Czajkowski, M., Hanley, N., and LaRiviere, J., 2014. The Effects of Experience on Preferences: Theory and Empirics for Environmental Public Goods. American Journal of Agricultural Economics, 97(1):333-351.

Diamond, P., 1996. Testing the Internal Consistency of Contingent Valuation Surveys. Journal of Environmental Economics and Management, 30(3):337-347.

Diamond, P. A., and Hausman, J. A., 1994. Contingent Valuation: Is Some Number better than No Number? The Journal of Economic Perspectives, 8(4):45-64.

Dillman, D. A., Smyth, J. D., and Christian, L. M., 2008. Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method. 3 Ed., Wiley.

Hanley, N., and Czajkowski, M., 2017. Stated Preference valuation methods: an evolving tool for understanding choices and informing policy. University of Warsaw, Department of Economics Working Paper 1(230).

Heberlein, T. A., Wilson, M. A., Bishop, R. C., and Schaeffer, N. C., 2005. Rethinking the scope test as a criterion for validity in contingent valuation. Journal of Environmental Economics and Management, 50(1):1-22.

Hensher, D. and Greene, W. 2003. The mixed logit model: the state of practice. Transportation 30(2):133-176.

Hess, S., and Train, K., 2017. Correlation and scale in mixed logit models. Journal of Choice Modelling, 23:1-8.

Huber, J., and Zwerina, K., 1996. The Importance of Utility Balance in Efficient Choice Designs. Journal of Marketing Research, 33(3):307-317.

Lancaster, K., 1966. A New Approach to Consumer Theory. Journal of Political Economy, 74(2):132-157.

McFadden, D., 1974. Conditional Logit Analysis of Qualititative Choice Behaviour. In: Frontiers in Econometrics, P. Zarembka, ed., Academic Press, New York, NY, 105-142.

McFadden, D., and Train, K., 2000. Mixed MNL Models for Discrete Response. Journal of Applied Econometrics, 15(5):447-470.

McFadden, D., and Train, K., 2017. Contingent Valuation of Environmental Goods. A Comprehensive Critique. Edward Elgar Publishing, Northampton, MA.

Mitchell, R. C., and Carson, R. T., 1989. Using Surveys to Value Public Goods: The Contingent Valuation Methods. Resources for the Future, Washington, DC.

Johnston, R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., Hanemann, W. M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R., and Vossler, C. A., 2017. Contemporary Guidance for Stated Preference Studies. Journal of the Association of Environmental and Resource Economists, 4(2):319-405.

LaRiviere, J., Czajkowski, M., Hanley, N., Aanesen, M., Falk-Petersen, J., and Tinch, D., 2014. The value of familiarity: Effects of knowledge and objective signals on willingness to pay for a public good. Journal of Environmental Economics and Management, 68(2):376–389.

Poe, G. L., 2016. Behavioral Anomalies in Contingent Values and Actual Choices. Agricultural and Resource Economics Review, 45(2):246-269.

Rollins, K., and Lyke, A., 1998. The Case for Diminishing Marginal Existence Values. Journal of Environmental Economics and Management, 36(3):324-344.

Sándor, Z., and Wedel, M., 2001. Designing conjoint choice experiments using managers' prior beliefs. Journal of Marketing Research, 38(4):430-444.

Scarpa, R., and Rose, J. M., 2008. Design Efficiency for Non-Market Valuation with Choice Modelling: How to Measure it, What to Report and Why. Australian Journal of Agricultural and Resource Economics, 52(3):253-282.

Street, D. J., and Burgess, L., 2007. The Construction of Optimal Stated Choice Experiments: Theory and Methods. Wiley-Interscience, Hoboken, NJ.

Street, D. J., Burgess, L., and Louviere, J. J., 2005. Quick and easy choice sets: Constructing optimal and nearly optimal stated choice experiments. International Journal of Research in Marketing, 22(4):459–470.

Train, K. E., and Weeks, M., 2005. Discrete Choice Models in Preference Space and Willingness-to-pay Space. In: Applications of Simulation Methods in Environmental and Resource Economics, R. Scarpa and A. Alberini, eds., Springer, Dordrecht, 1-16.

Vossler, C. A., Doyon, M., and Rondeau, D., 2012. Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments. American Economic Journal: Microeconomics, 4(4):145-171.

Vossler, C. A., and Watson, S. B., 2013. Understanding the Consequences of Consequentiality: Testing the Validity of Stated Preferences in the Field. Journal of Economic Behavior and Organization, 86:137-147.

Zawojska, E., and Czajkowski, M., forthcoming. Re-examining empirical evidence on stated preferences: Importance of incentive compatibility. Journal of Environmental Economics and Policy.