European Review of Agricultural Economics Advance Access published July 7, 2009

European Review of Agricultural Economics pp. 1–24
doi:10.1093/erae/jbp012

Modelling attribute non-attendance in choice experiments for rural landscape valuation

Riccardo Scarpa
University of Waikato, Hamilton, New Zealand

Timothy J. Gilbride
University of Notre Dame, Indiana, USA

Danny Campbell
Queen's University Belfast, Northern Ireland

David A. Hensher
University of Sydney, Australia

Received November 2008; final version revised February 2009

Abstract

Non-market effects of agriculture are often estimated using discrete choice models from stated preference surveys. In this context we propose two ways of modelling attribute non-attendance. The first involves constraining coefficients to zero in a latent class framework, whereas the second is based on stochastic attribute selection and grounded in Bayesian estimation. Their implications are explored in the context of a stated preference survey designed to value landscapes in Ireland. Taking account of attribute non-attendance with these data improves fit and tends to involve two attributes one of which is likely to be cost, thereby leading to substantive changes in derived welfare estimates.

Keywords: attribute non-attendance, discrete choice modelling, stated preference, latent class models, stochastic attribute selection models

JEL classification: C25, Q26, Q51

Review coordinated by Alison Burrell.

1. Introduction

Non-market benefits from agricultural policies are becoming increasingly important for multifunctional policy evaluation. Stated preference studies are well-suited to explore this issue (Pruckner, 1994; Cicia and Scarpa, 2000; Randall, 2002), especially in their multi-attribute format (Scarpa *et al*., 2007; Campbell *et al*., 2008).. More generally, choice modelling from stated preferences has generated a great deal of interest in agricultural economics research (see, for example, Rigby and Burton, 2005, 2006; Hanley *et al*., 2006; Carlsson *et al*., 2007). In this framework, respondents are typically asked to choose their preferred alternative among several hypothetical alternatives in a sequence of experimentally designed choice tasks. A basic assumption, which gives rise to the continuity axiom, is that of unlimited substitutability between the attributes used to describe the alternatives in each of the choice tasks in the sequence. This implies that respondents make trade-offs between all attributes describing each of the alternatives, and are expected to choose their most preferred alternative in a choice set. Thus, the continuity axiom rules out situations where respondents focus solely on a subset of attributes, ignoring all other differences between the alternatives (see Hensher, 2006). Ignoring attributes in the choice task implies non-compensatory behaviour, because no matter how much the level of a given attribute is improved, the improvement will fail to compensate for worsening in the levels of other attributes if the attribute itself is ignored by the respondent (Lockwood, 1996; Spash, 2000; Sælensminde, 2002; Rekola, 2003). Therefore, respondents using such discontinuous preference orderings pose a problem for neoclassical analysis as they cannot be represented by a conventional utility function (Lancsar and Louviere, 2006).

Without continuity, there is no trade-off between two different attributes (Gowdy and Mayumi, 2001; McIntosh and Ryan, 2002; Rosenberger *et al.*, 2003). This is a key issue when computing the marginal rate of substitution between attributes. While the marginal rate of substitution can be calculated from the estimated parameters at the level of the sampled population, it is not computable for individual respondents who do not make trade-offs between some attributes. Crucially, for non-market valuation, no relative implicit price can be computed for these respondents. From a statistical perspective, pooling observations where some respondents attend to all attributes while others attend to only a subset will lead to erroneous and biased estimates.

While substantial literature exists in the fields of transport (e.g. Swait, 2001; Hensher *et al.*, 2005; Hensher, 2006, 2008, 2010), marketing (e.g. Swait and Adamowicz, 2001a, 2001b; Fasolo *et al.*, 2007; Islam *et al.*, 2007) and health (e.g. McIntosh and Ryan, 2002; Lancsar and Louviere, 2006), indicating that not all respondents attend to all attributes, with few exceptions (see, for example, Rigby and Burton, 2006; Campbell *et al.*, 2008; Carlsson *et al.*, 2008) the issue remains relatively unexplored in the literature on the nonmarket valuation of agricultural public goods. Nevertheless, it seems obvious that the issue of attribute non-attendance (henceforth AN-A) is likely to have serious consequences on the derivation of welfare estimates, especially when the object of neglect is the monetary attribute, such as the cost of an alternative.

The detection and statistical handling of AN-A raises a few technical issues for the practice of discrete choice modelling. A relatively unexplored issue regards how to incorporate this phenomenon in statistical models when data on self-reported AN-A are not available. Many past studies do not have such information, but given the consequences highlighted in the few studies conducted so far, researchers involved in previous studies might want to explore whether the conclusions drawn from previous data analysis are robust to models accounting for AN-A. Our main contribution is to explore some intuitive ways of addressing this issue, building on basic models that are commonly employed by practitioners. In particular, the present paper contributes to this literature by proposing two different panel mixed logit models to account for repeated attribute exclusion in the evaluation of proposed alternatives by a given respondent. What is intended here is that the identification of AN-A behaviour is achieved by analysing the observed response pattern using a statistical model with degenerate distributions of taste intensities at zero, which implies non-attendance. In contrast, past research (Campbell *et al.*, 2008; Carlsson *et al.*, 2008; Hensher, 2008) have asked respondents which attributes they paid attention to or were important. Our approach here has the advantage that it can be applied in the absence of self-reported information on attendance.

The first of the two panel mixed logit models we propose is developed from the frequentist framework and is based on a latent class structure (finite mixing panel), whereas the second is based on the Bayesian framework with a variable selection structure embedded in a continuous mixing panel model. The latent class approach is relatively easy to implement with commercially available software, and offers a quick check on attribute non-attendance. The Bayesian approach offers a more flexible distribution of heterogeneity but at the expense of more complicated programming. Once the data set has been collected, it may be erroneous to estimate specifications that assume all attributes have been attended to and traded-off among each other when in fact the decision heuristics adopted by respondents include systematic AN-A.[1] At the very least, our approach provides a means for testing the sensitivity of model performance and, importantly according to our results, the implied willingness to pay (WTP) estimates derived from such a crucial implicit assumption in the absence of supplementary data. Non-attendance to a single attribute in choice experiments may be a substantial concern. For example, using supplementary self-stated measures of non-attendance in a discrete choice study, Hensher (2008) found that 38 per cent of respondents did not attend to at least

1 AN-A is not the only processing rule used by individuals, but it may be used here as a simplified heuristic for other choice behaviours. Other rules include thresholds (Swait, 2001) on attributes relative to prior attribute experience (Greene and Hensher, 2008; Hensher, 2008), and attribute aggregation when the units are common (see Hensher, 2006).

one attribute, while non-attendance to any one attribute ranged from 6 to 21 per cent.

To illustrate the methods, we apply them to a data set already used in the literature in Scarpa *et al.* (2007). Results from both approaches support the notion that a sizeable proportion of respondents adopt some simplified, albeit relevant, choice heuristic involving non-attendance to some attributes. The effects of ignoring this issue vary according to the model used, but they are sizeable in all instances examined here.

The rest of the paper is organised as follows. Section 2 describes the features of two estimation methods, the equality-constrained latent class and the Bayesian attribute selection models. The empirical context is then presented in Section 3, followed by model estimation results and implications for WTP of AN-A in Section 4. Key findings and directions for ongoing research are given in the concluding section.

2. Method

We propose two estimation methods to account for respondent non-attendance of each attribute included in the indirect utility of discrete choice models. The first is grounded in the frequentist estimation approach based on maximising a sample log-likelihood function under the constraint that some of the latent classes of AN-A take up a specific indirect utility structure via adequately posed equality constraints. The second method is grounded in Bayesian estimation and relies on Monte Carlo Markov Chain methods to achieve adequate stationarity in the posterior distributions of the model parameters.

Both methods are based on random utility theory and hypothesise the existence of a utility function whose arguments are the levels of the attributes used (in our empirical application, these attributes are alternative rural landscape outcomes).

Respondent n evaluates alternative k in the available choice set according to the linear additive utility function given by

$$U_{nk} = \sum_{h=1}^{H} \beta_{nh} x_{kh} + \beta_n (m_n - t_k) + e_k, \tag{1}$$

where e_k is the classic i.i.d. Gumbel-distributed error term with scale λ, m_n is income, t_k is the cost of alternative k and x_{kh} denotes the level of the h attribute in alternative k, while the attendant utility coefficients β_{nh} and the cost coefficient $\beta_{n\$}$ are to be estimated from data. The probability of selecting alternative i in this context is logit:

$$\Pr(i) = \frac{\exp\left(\lambda \sum_{h=1}^{H} \beta_{nh} x_{ih} + \beta_{n\$}(m_n - t_i)\right)}{\sum_k \exp\left(\lambda \sum_{h=1}^{H} \beta_{nh} x_{kh} + \beta_{n\$}(m_n - t_k)\right)}, \quad \lambda > 0 \tag{2}$$

where λ remains unidentified in estimation and is set equal to 1, as is common in the literature. This model represents the conventional approach underlying most non-market valuation studies, which involves maximising the sample log-likelihood of the panel of observed choices given by

$$\ln L = \sum_{n=1}^{N} \ln\left[\prod_{t=1}^{T_n} \left(\Pr(i_t)_n\right)^{y_t}\right], \tag{3}$$

and then using the estimated value of β to derive welfare estimates. The expression inside square brackets is the joint probability of the panel of choices made by each respondent, while y_t is the indicator of choice.

2.1. The equality-constrained latent class approach

Our first proposed model is the equality-constrained latent class (ECLC) model, in which AN-A is operationalised by allowing some respondents to belong (with a probability to be estimated) to latent classes with zero utility weights for selected attributes, while non-zero attributes are assumed to take the same value across classes. Suppose we want to investigate the sensitivity of our choice data to the assumption that some respondents, n, in our sample systematically ignored one or more attributes in the sequence of observed discrete choices. Where this is the case, then the statistical fit of the model should increase if we set the relevant utility coefficients of this respondent to zero, thereby implying 'non-attendance' and hence discontinuous preferences for this individual. One way to approach the issue is to assume a split in behaviour across respondents. Some respondents might have traded-off all attributes, as is commonly assumed. We call these *fully compensatory or conventional* respondents, and denote them by *TA* for total attendance. We call respondents who ignore all attributes and give responses based on chance alone *totally non-compensatory* respondents, and denote them by *TNA* for total non-attendance. The remaining respondents might systematically ignore *one* attribute only in their entire sequence of preferred choices, thereby displaying *partial non-compensatory* preferences, a behaviour we denote by *PNA*1, for partial non-attendance of one attribute. Others might systematically ignore *two* of these attributes, displaying *PNA*2 behaviour, and so on. Unlike, for example, Swait (2001), we do not establish a mapping of process rules to specific individuals, but we impose some distributional assumptions within a homogeneous group of individuals or class (e.g. latent class), and infer membership probability up to groups exercising non-attendance to specific attributes.

PNA respondents might just be seeking to reduce their cognitive effort in rationalising a decision rule. Hensher (2006, 2008) argued that individuals appear to adopt a range of 'coping' or editing strategies in hypothetical choice settings that are consistent with how they normally process information in real markets. Choice experiments have varying amounts of information to process. However, aligning 'choice complexity' with the amount of information to process is potentially misleading as *relevancy* is what matters (Hensher, 2006). The heuristics individuals use to evaluate choices is what needs to be captured through frameworks that can empirically identify rules adopted by individuals. This can be done by a mix of supplementary questions as well as particular probabilistic conditions imposed on the model specification.[2]

Postulating the existence of these groups is equivalent to identifying separate *classes* of choice behaviour among respondents. Hence, our first model is cast in the frame of latent class models,[3] albeit with a common equality constraint on the parameters that are allowed to be non-zero, so that when attributes are attended to, their utility weights take the same value across all classes as is commonly assumed in multinomial logit models. This can be relaxed if the objective is to identify preference heterogeneity across classes, but this is not our emphasis here. Our first objective is to provide a model that allows for the effect of attribute non-attendance. The impact of taste heterogeneity is explored in our subsequent model based on Bayesian attribute selection.

In the ECLC, the *TNA* class has all parameters simultaneously constrained to be equal to zero. At the opposite end of the spectrum, for the *TA* class the coefficients of all attributes are freely estimable, but constrained to take the same values across classes. The exact way of identifying the frequency of *PNA*1, *PNA*2, etc. behaviour depends on the coding chosen for the choice attributes. If all k non-monetary attributes are numerically coded, then $k+1$ classes are necessary to implement all forms of *PNA*1 behaviour. If the coding of non-monetary attributes is either dummy- or effect-coded with J_k dummy variable coefficients identifiable for each attribute, then the classes will still be $k+1$, but each class will involve J_k coefficients constrained to be zero. Finally, with the class *PNAz*, where z stands for $2, 3, \ldots, k$, it is necessary to count all the combinations $\binom{k+1}{z}$ in which the attributes can take zero coefficient values. For example, in our empirical study (see below) of $k+1=5$ attributes, the number of different classes needed to cover all the combinations for two given attributes to be simultaneously set to zero is $\binom{5}{2}=10$. Each class will have a number of coefficients set to zero, depending on the variable coding adopted. In our case, we have two identifiable dummy variable coefficients (e.g. three-level attributes) for all k non-monetary attributes, plus the single coefficient for cost (coded numerically). We explain below the implementation of this model given the typical data structure of a choice experiment.

Note that no assumption is required as to whether the real utility function of those respondents who appear to ignore some attribute is the same or different from those of respondents attending to those attributes. One simply takes the sequence of choices made by attribute non-attendants as observationally

2 Hensher and Layton (2008) developed a non-linear utility specification that allows the data to reveal a probability distribution of non-attendance and attribute accumulation across a sample.

3 The latent class approach allows membership of a class only up to a probability, with the full probability per respondent allocated across all classes.

equivalent to a utility function that assigns zero values to the attributes subject to non-attendance. In the absence of further information, which in other studies might well be derived by debriefing questions, one cannot distinguish between the case where a zero is the outcome of – for example – a simplifying heuristic, and where it is instead a true manifestation of individual preferences.

2.2. Modelling strategy for the ECLC approach

We present three-panel latent class models to identify the membership probabilities corresponding to various forms of AN-A, and contrast the results with the basic multinomial logit model (Model 1).

Conditional on belonging to a given class, the probability of observing a set of T choices $\mathbf{y}_n = \{y_1, y_2, \ldots, y_T\}$ by respondent n is a product of logits:

$$P_n(\mathbf{y}_n|\boldsymbol{\beta}_c) = P_t(i_1, \ldots, i_T|\boldsymbol{\beta}_c) = \prod_{t=1}^{T} \frac{\exp(\mathbf{x}_{it}\boldsymbol{\beta}_c)}{\sum_k \exp(\mathbf{x}_{kt}\boldsymbol{\beta}_c)}. \tag{4}$$

The unconditional probability of the panel of choices of respondent n is obtained, using the law of total probability to obtain the unconditional probability of choice, by adding up the conditional probabilities over the finite set of c logit membership probabilities of the various attribute attendance classes and scale class. This implies that

$$P_n(\mathbf{y}_n) = \sum_c P(c)P_n(i_t|\boldsymbol{\beta}_c) = \sum_c \frac{\exp(\alpha_h)}{\sum_c \exp(\alpha_c)} \prod_{t=1}^{T} \frac{\exp(\mathbf{x}_{it}\boldsymbol{\beta}_c)}{\sum\limits_k \exp(\mathbf{x}_{kt}\boldsymbol{\beta}_c)}, \tag{5}$$

where α denotes class-specific constants identified by ensuring they sum to zero.

In Model 2 we identify the frequency of *PNA*1 and hence have seven classes:

1. Class 1 (total attendance, *TA*) comprises respondents simultaneously considering *all* attributes. The values of these non-zero coefficients are constrained to be the same across all classes, so that only attendance and non-attendance, but not taste heterogeneity across respondents, are captured.
2. Class 2 (total non-attendance, *TNA*) comprises those who simultaneously disregard all attributes. In this class, all coefficients are constrained to be zero.
3. Classes 3–7 (partial non-attendance ignoring one attribute ignored or *PNA*1) represent the various classes in which only one attribute is ignored in each class. So in Class 3 the two coefficients referring to the two identifiable levels of the first attribute are set to zero, while the other take the same values as in the other classes; in Class 4 a similar

restriction is placed on the two identifiable coefficients for the second attribute, etc.

Model 3 serves another purpose. Apart from modelling simultaneously the class membership probabilities of those who do not attend to any attribute (*TNA*), attend to them all (*TA*) and ignore one only (*PNA*1), we are specifically interested in non-attendance of the monetary attribute in combination with at least one non-monetary one (*PNA*2). Classes 8–11 of Model 3 represent these groups, whereas Classes 1–7 are as defined in Model 2.

Finally, Model 4 covers the case in which all the different combinations of simultaneous neglect of two attributes are observed (Classes 3–12). This model systematically explores the notion that two out of the five attributes are neglected. Table 1 illustrates the class structure and different equality constraints used in Models 1–4. Our approach can be immediately extended to identify classes involving the simultaneous non-attendance to more than two attributes. However, this makes the number of classes proliferate very quickly. For example, a model that explores neglect for all, none, two-at-a-time and three-at-a-time attributes would involve, in our case, 27 classes.[4] Models with such structure were explored and discarded as they did not significantly improve the fit to the observed data. All estimation was conducted using Latent Gold Choice v. 4.0 (Vermunt and Magidson, 2005).

2.3. The Bayesian attribute selection approach

A shortcoming of the previous approach is to ignore taste heterogeneity among non-zero taste intensities. Although taste heterogeneity could be built into such a framework, it tends to produce over-parameterised models and increases sample size requirements for identification. A more parsimonious alternative accounting for taste heterogeneity is one that allows for a continuous distribution of heterogeneity, as well as non-attendance. Representing heterogeneity via a continuous distribution in discrete choice models has been used extensively in social science research, including transportation, marketing, economics, land use and others. These models have been estimated via both classical and Bayesian methods.[5]

Attribute (or variable) selection has been studied from the Bayesian perspective by several researchers. For aggregate level data, a Gibbs sampling approach was proposed by George and McCulloch (1993, 1997); alternatives

4 Calculated as

$$1 + 1 + 5 + \binom{4+1}{2} + \binom{4+1}{3} = 7 + 10 + 10 = 27.$$

.

5 See Train (2003) for a general introduction, Greene and Hensher (2008) for generalised ordered choice models and Rossi *et al.* (2005) for a detailed treatment of Bayesian methods.

Table 1. ECLC models and class structure

Model	Class	Mountain Land A lot	Mountain Land Some	Stonewalls A lot	Stonewalls Some	Farmyard Tidiness A lot	Farmyard Tidiness Some	Cultural Heritage A lot	Cultural Heritage Some	Cost	
Model 1 (Model 2, Model 3)	Class 1	β_{11}	β_{12}	β_{21}	β_{22}	β_{31}	β_{32}	β_{41}	β_{42}	$\beta_{\$}$	*TA*
Model 1 (Model 2, Model 3)	Class 2	0	0	0	0	0	0	0	0	0	*TNA*
Model 2 (Model 3)	Class 3	0	0	β_{21}	β_{22}	β_{31}	β_{32}	β_{41}	β_{42}	$\beta_{\$}$	*PNA1*
Model 2 (Model 3)	Class 4	β_{11}	β_{12}	0	0	β_{31}	β_{32}	β_{41}	β_{42}	$\beta_{\$}$	*PNA1*
Model 2 (Model 3)	Class 5	β_{11}	β_{12}	β_{21}	β_{22}	0	0	β_{41}	β_{42}	$\beta_{\$}$	*PNA1*
Model 2 (Model 3)	Class 6	β_{11}	β_{12}	β_{21}	β_{22}	β_{31}	β_{32}	0	0	$\beta_{\$}$	*PNA1*
Model 2 (Model 3)	Class 7	β_{11}	β_{12}	β_{21}	β_{22}	β_{31}	β_{32}	β_{41}	β_{42}	0	*PNA1*
Model 3	Class 8	0	0	β_{21}	β_{22}	β_{31}	β_{32}	β_{41}	β_{42}	0	*PNA2*
Model 3	Class 9	β_{11}	β_{12}	0	0	β_{31}	β_{32}	β_{41}	β_{42}	0	*PNA2*
Model 3	Class 10	β_{11}	β_{12}	β_{21}	β_{22}	0	0	β_{41}	β_{42}	0	*PNA2*
Model 3	Class 11	β_{11}	β_{12}	β_{21}	β_{22}	β_{31}	β_{32}	0	0	0	*PNA2*

Model	Class	Mountain Land A lot	Mountain Land Some	Stonewalls A lot	Stonewalls Some	Farmyard Tidiness A lot	Farmyard Tidiness Some	Cultural Heritage A lot	Cultural Heritage Some	Cost	
Model 4	Class 1	β_{11}	β_{12}	β_{21}	β_{22}	β_{31}	β_{32}	β_{41}	β_{42}	$\beta_{\$}$	*TA*
Model 4	Class 2	0	0	0	0	0	0	0	0	0	*TNA*
Model 4	Class 3	0	0	0	0	β_{31}	β_{32}	β_{41}	β_{42}	$\beta_{\$}$	*PNA2*
Model 4	Class 4	0	0	β_{21}	β_{22}	0	0	β_{41}	β_{42}	$\beta_{\$}$	*PNA2*
Model 4	Class 5	0	0	β_{21}	β_{22}	β_{31}	β_{32}	0	0	$\beta_{\$}$	*PNA2*
Model 4	Class 6	β_{11}	β_{12}	0	0	0	0	β_{41}	β_{42}	$\beta_{\$}$	*PNA2*
Model 4	Class 7	β_{11}	β_{12}	0	0	β_{31}	β_{32}	0	0	$\beta_{\$}$	*PNA2*
Model 4	Class 8	β_{11}	β_{12}	β_{21}	β_{22}	0	0	0	0	$\beta_{\$}$	*PNA2*
Model 4	Class 9	0	0	β_{21}	β_{22}	β_{31}	β_{32}	β_{41}	β_{42}	0	*PNA2*
Model 4	Class 10	β_{11}	β_{12}	0	0	β_{31}	β_{32}	β_{41}	β_{42}	0	*PNA2*
Model 4	Class 11	β_{11}	β_{12}	β_{21}	β_{22}	0	0	β_{41}	β_{42}	0	*PNA2*
Model 4	Class 12	β_{11}	β_{12}	β_{21}	β_{22}	β_{31}	β_{32}	0	0	0	*PNA2*

to this algorithm have been suggested by Raftery *et al.* (1997) and Geweke (1996). A model applicable to aggregate level logistic regression was proposed by Chen *et al.* (1999). Gilbride *et al.* (2006) reviewed these methods and proposed an algorithm for heterogeneous attribute selection that is analogous to the approach suggested by George and McCulloch (1993). Their method is appropriate for discrete choice data when, as in the case at hand, attribute selection is conducted at the individual rather than the sample level. In this study, we apply the method by Gilbride *et al.* (2006) to the problem of respondent attribute selection to determine its effect on WTP estimates. In a Bayesian model, the likelihood specified above is augmented with the distribution of heterogeneity for the β_n, e.g. $\beta_n \sim N(\bar{\beta}, V_\beta)$ where $N(\cdot,\cdot)$ is the multivariate normal distribution (but other analytical distributions can be used). The distribution of heterogeneity serves as the prior for β_n and the hierarchical model is completed by specifying prior distributions for $\bar{\beta}$ and V_β. Markov Chain Monte Carlo (MCMC) methods are then used to draw elements from $\pi(\{\beta_n\}, \bar{\beta}, V_\beta | Data)$, i.e. the posterior distribution of the individual β_n's and parameters describing the distribution of heterogeneity, given the data. Conditional mixed logit has an equivalent specification via classical inference methods.

Respondent-level attribute selection could be introduced via $\tilde{\tau}_n$, a vector of the same length as β_n. Using the notation above, β_n is of length $H+1$ where the last element represents the cost coefficient. If element h of $\tilde{\tau}_n$ equals 1, then that element is attended to, and used by respondent n in his evaluation of choice alternatives. Here, the coefficient set indexed by h includes the cost coefficient. Otherwise, we set $\tilde{\tau}_n = 0$. Modifying our notation slightly, the utility function in equation (1) becomes

$$U_{nk} = \sum_{h=1}^{H} \tilde{\tau}_{nh} \tilde{\beta}_{nh} x_{kh} + \varepsilon_k \quad \text{where } \tilde{\tau}_{nh} \sim \text{binomial } (\tilde{\theta}_h). \tag{6}$$

While conceptually appealing, the parameters in this representation of utility are hard to estimate using standard hierarchical models. For instance, if $\tilde{\tau}_{nh} = 0$, then $\tilde{\beta}_{hn}$ is unidentified. We address this issue as follows: let $\tau_{nh} = 1$ if attribute h is attended to by the person n. Otherwise, we set $\tau_{nh} = c$, where c is some small constant indicating that this variable or attribute is not used by this respondent when making choices. Let $C_{\tau n} = \text{diag}(\tau_n)$, an $(H+1) \times (H+1)$ square matrix with the vector τ_n along the diagonal. The distribution of heterogeneity is now $\beta_n \sim N(C_{\tau n}\bar{\beta}, C_{\tau n} V_\beta C_{\tau n})$ and utility is again represented by equation (1).

Several comments on this approach are in order. First, when $\tau_{nh} = c$, β_{nh} is not exactly equal to zero, but drawn from a distribution where the hth element's mean is very close to zero with a very small variance. In the application below, c is set equal to 0.01. For all practical purposes, β_{nh} is indistinguishable from zero but this formulation permits the use of standard methods for updating parameters and drawing from posteriors. Specifically, let $\beta_n^* = C_{\tau n}^{-1}\beta_n$, then $\beta_n^* \sim N(\bar{\beta}, V_\beta)$. Heterogeneity is introduced by letting the probability that $\tau_{nh} = 1$ be θ_h, and hence $\tau_{nh} = \mathrm{c}$ is $1 - \theta_h$. Therefore, θ_h is a measure of the aggregate level probability that attribute h is being used in the choice task by respondents. An appropriate prior is used for each θ_h, and the full posterior is given by $\pi(\{\beta_n\}, \{\tau_n\}, \theta, \bar{\beta}, V_\beta|Data)$. Gilbride *et al.* (2006) provide complete details on priors and the estimation algorithm.

2.4. Modelling strategy for the stochastic attribute selection approach

We call the Bayesian approach to attribute selection Stochastic Attribute Selection (STAS). George and McCulloch (1993) named their original model Stochastic Search Variable Selection for the case of linear regression where each attribute (variable) can be included or excluded in the model. In discrete choice models, 'dummy level' coding is frequently used to represent the discrete levels of particular attributes. Similar to the ECLC approach, we conceptualise 'variable selection' at the attribute level: an individual either attends to all the levels of an attribute or ignores that attribute altogether. For example, if τ_{n1} and τ_{n2} map to different levels of the same attribute, then we force τ_{n1} to be equal to τ_{n2} and use $\theta_{\{1,2\}}$. From a purely technical standpoint, this assumption is not necessary, but it is consistent with behavioural theories of respondents attending to only a subset of *attributes*, as opposed to *levels* in an attribute.

There are several differences between the ECLC and STAS approaches. First is the basic difference between finite mixture models and continuous models for representing heterogeneity. Researchers in many instances have found that the continuous distribution of heterogeneity fits the data better (in terms of overall goodness of fit and model parsimony) and provides a richer description of heterogeneity between respondents (see, for instance, the review by Allenby and Rossi, 1999). However, because of their conceptual appeal, easier interpretability and the availability of robust commercial software, many researchers prefer finite mixture models. In the ECLC approach, the researcher must specify *a priori* the combinations of attributes or β_{nh}s that are set equal to zero, and this can lead to a relatively large number of latent classes to estimate. In the standard STAS approach, any combination of β_{nh}s corresponding to different attributes may be set equal to zero as determined by the data. An analysis of the posterior distribution of τ_ns would be necessary to determine combinations of attributes that are frequently excluded from consideration. Our purpose is not to advocate one approach over the other, but to illustrate the gains from explicitly modelling attribute attendance in terms of model fit and substantive conclusions using each method.

3. Data

We utilise the survey data collected and described by Scarpa *et al.* (2007) on the general public's attitudes and preferences regarding rural environmental landscape improvements in the Republic of Ireland. The four landscape improvements focused on in this paper are: the protection of *Mountain Land* from overstocking; enhancement of the visual aspect of *Stonewalls*; maintenance of *Farmyard Tidiness*; and safeguarding of *Cultural Heritage*. Three levels were used to depict each of these landscape attributes according to the effort made to conserve or enhance the attributes. To minimise respondent confusion, the levels for each landscape attribute were labelled: 'A lot of Action', 'Some Action' and 'No Action'–representing a high level of improvement, an intermediate level of improvement and the unimproved or status quo condition, respectively. As valuation of landscape components is subjective, and verbal descriptions can be interpreted differently on the basis of individual experience, each level of improvement was visualised by digitally manipulating a 'control' photograph to depict either more or less of the attribute in question. This method was used so that changes in the attribute levels could be easily identified while holding other features of the landscape constant.

Different stocking densities in an upland area reflecting overgrazing and soil erosion were used to depict the *Mountain Land* attribute. The *Stonewalls* attribute illustrated the consequence that their condition and their removal have on the appearance of the countryside. Similarly, the *Farmyard Tidiness* attribute portrayed a farmyard in different states of tidiness and the *Cultural Heritage* attribute showed the impact of different management practices on old farm buildings and historical features. All images and accompanying texts were tested in focus group discussions and a pilot study, to ensure a satisfactory understanding and scenario acceptance by respondents. The *Cost* attribute was specified as the value in Euros that the respondent would personally have to pay per year through their income tax and value added tax contributions. An example of a choice task can be found in Campbell *et al.* (2009).

The discrete choice experiment consisted of a panel of at least five repeated choice sets. Each choice set consisted of two experimentally designed alternatives (labelled 'Option A' and 'Option B') and a status quo alternative (labelled 'No Action') which portrayed all the landscape attributes at the No Action level with zero cost to the respondent. The study employed a sequential experimental design with a Bayesian information structure to maximise the D_b-optimal criterion, which is outlined in Sándor and Wedel (2001). Starting from a conventional main-effects fractional factorial in the first phase, a Bayesian design was employed in the second wave of sampling. The design for the final phase incorporated information from the first and second phases in a similar fashion. For further information and an evaluation of the efficiency of the sequential experimental design approach used in this study, the interested reader is directed to Scarpa *et al.* (2007) and Ferrini and Scarpa (2007).

The study adopted a stratified random sample to reflect the geographic distribution of the Irish adult population, the approximate rural/urban spilt, the approximate socio-economic status of the regional population and the approximate gender and age profile of the population. The survey was administered to a random sample of 766 respondents drawn from the Irish adult population. Of these, 600 respondents agreed to participate. The overall response rate of 78 per cent is in line with similar studies in Ireland.

Respondents who consistently choose the status quo (across all choice sets) do not appear to be willing to make trade-offs among the attributes and their inclusion might consequently distort our findings. For this reason, we chose to exclude 47 respondents from the analysis. Thus, a final sample of 553 respondents, with each respondent tackling five to six choice tasks, resulted in 3,127 observations for model estimation.

This sample consisted of approximately equal numbers of male and female respondents. In line with the breakdown of the Irish population, the majority of respondents were younger than 45 years of age (58 per cent), resided in areas classified as non-rural (58 per cent) and worked, on either a full-time or part-time basis (61 per cent). Of the 65 per cent of respondents who disclosed their income, the average annual gross income was approximately €26,000.

4. Results

4.1. Latent class with zero-constrained attributes

Table 2 reports the estimates of latent class panel models (Model 2, 3 and 4) and contrasts them with the basic multinomial logit model (Model 1) where allowance is made neither for non-attendance to single attributes nor for the panel nature of the data.[6] Model 2 allows for seven classes. The first two include, respectively, respondents who attended to all attributes (the *TA* group, 1.41 per cent) and to no attribute (the *TNA* group, 5.93 per cent). Classes 3 to 7 allow for one attribute to be ignored and its coefficient constrained to zero in each class, while other attributes have identical coefficient values to those of Class 1 and represent the *PNA*1 group of respondents. The total non-attendance rate for a given attribute is the sum of Class 2 and of the class corresponding to that attribute. According to this model, *Farmyard Tidiness* is the least attended to of all attributes, with respondents having, on average, a 0.502 probability of ignoring only this attribute and belonging to Class 5, plus on average a 0.0593 probability of ignoring this and all other attributes (Class 2), giving a total of 0.5613. This implies that respondents have on average only a 0.4387 probability of attending to this attribute. According to this model, the most attended to attribute is *Mountain Land*, closely followed by *Cost* with average probabilities of 0.9391 and 0.9295, respectively, while *Stonewalls* and *Cultural Heritage* have a probability of attendance of around 0.7083 and 0.7674, respectively.

However, Model 2 only allows for two forms of non-attendance: (a) either respondents do not attend to any (*TNA*) or (b) they ignore only one attribute (*PNA*1). This model produces unsubstantive WTP estimates due to an

6 Cross-section MNL models were estimated that allowed for non-attendance and are available on request from the first author.

Table 2. Equality Constrained Latent Class model estimates

	Model 1: MNL		Model 2: 7 classes			Model 3: 11 classes			Model 4: 12 classes		
	$\hat{\beta}$ (z-value)	$\hat{WTP}$ (st. err.)	$\hat{\beta}$ (z-value)	$\hat{WTP}$ (st. err.)	$\hat{Pr}(\beta \neq 0)$ (z-value)	$\hat{\beta}$ (z-value)	$\hat{WTP}$ (st. err.)	$\hat{Pr}(\beta \neq 0)$ (z-value)	$\hat{\beta}$ (z-value)	$\hat{WTP}$ (st. err.)	$\hat{Pr}(\beta \neq 0)$ (z-value)
Mountain land (A lot)	0.9829	702.07	1.2656	791.00		1.3066	20.91		1.3210	21.14	
	(15.5)	(514.82)	(15.5)	(625.02)	0.9331	(15.9)	(3.71)	0.9430	(15.9)	(5.86)	0.9305
Mountain land (Some)	0.7132	509.43	0.9400	587.50	(55.9)	0.9593	15.35	(68.4)	0.9662	15.46	(44.9)
	(11.4)	(371.40)	(11.9)	(462.07)		(12.1)	(2.77)		(12.2)	(4.33)	
Stonewalls (A lot)	1.0351	739.36	1.7961	1122.56		1.8076	28.92		1.8011	28.82	
	(15.3)	(542.95)	(15.7)	(887.21)	0.7083	(15.6)	(5.56)	0.7235	(15.6)	(8.17)	0.7314
Stonewalls (Some)	0.8900	635.71	1.5095	943.44	(16.2)	1.4960	23.94	(18.5)	1.4918	23.87	(19.1)
	(11.3)	(465.47)	(10.8)	(744.43)		(10.3)	(5.00)		(10.2)	(7.11)	
Farmyard tidiness (A lot)	0.8565	611.79	2.4332	1520.75		2.6575	42.52		2.6630	42.61	
	(12.3)	(453.88)	(8.2)	(1227.81)	0.4387	(8.1)	(6.24)	0.4447	(7.9)	(9.83)	0.4442
Farmyard tidiness (Some)	0.7341	524.36	1.9284	1205.25	(8.9)	2.0405	32.65	(9.3)	2.0445	32.71	(9.0)
	(11.1)	(385.79)	(9.5)	(959.09)		(9.1)	(4.88)		(9.0)	(7.70)	
Cultural heritage (A lot)	0.8113	579.50	1.3217	826.06		1.3413	21.46		1.3386	21.42	
	(13.1)	(426.72)	(12.2)	(654.83)	0.7674	(12.4)	(4.35)	0.7968	(12.0)	(6.40)	0.7989
Cultural heritage (Some)	0.7295	521.07	1.1725	732.81	(15.7)	1.1937	19.10	(18.8)	1.1943	19.11	(18.6)
	(11.3)	(383.77)	(10.8)	(583.26)		(11.3)	(3.89)		(11.0)	(5.71)	
Cost	−0.0014		−0.0016		0.9295	−0.0651		0.0865	−0.0625		0.0915
	(−1.3)		(−1.2)		(50.4)	(−5.6)		(3.8)	(−3.8)		(3.6)
LL	−2,356.17		−2,225.85			−2,204.58			−2,203.88		
BIC(LL)	4,769.18		4,546.43			4,529.14			4,534.06		
AIC(LL)	4,730.35		4,481.70			4,447.15			4,447.76		
AIC3(LL)	4,739.35		4,496.70			4,466.15			4,467.76		

Note: Standard errors in brackets, approximated via delta method for WTP.

insignificant *Cost* coefficient. In all likelihood, decision heuristics implying non-attendance are more articulated than the two extreme forms (a) and (b) and it makes sense to also allow for respondents to ignore a pair of attributes simultaneously (*PNA*2). In applications with many attributes, this may substantially increase the number of classes, depending on how many attributes are used to describe alternatives. One way to reduce the overall number of classes is to extend non-attendance to paired attributes, but limiting the latter to pairs including *Cost*, which in a non-market valuation study is of major interest.[7]

Model 3 imposes this assumption, and involves the estimation of 11 classes. The first two are identical to those of Model 2, representing *TA* and *TNA* respondents. Classes 8 to 11 are new and allow for the four combinations of non-monetary attributes and *Cost*. Together they are found to account for a membership probability on average of 0.8527 while Classes 3 to 7 are the single attribute non-attendance classes *PNA*1 and account for only 0.035. The estimated membership probabilities suggest that, when non-attendance to pairs of attributes is considered, the *Cost* attribute is likely to be one of them.

We note that in this case the probability of non-attendance of each attribute is made up of three components:

- total non-attendance (Class 2 with *TNA* respondents),
- pair-wise non-attendance (the relevant class in the Groups 8 to 11, or *PNA*2 respondents) and
- single-AN-A (the relevant class in the Group 3 to 7, or *PNA*1 respondents).

According to this model, the attribute with lowest probability of attendance (0.0865) is the *Cost* attribute. This is remarkably low, and much lower than what was found in the restricted Model 2. The most attended to attribute in Model 3 remains *Mountain Land*, with a probability of 0.943 and the least attended to non-monetary attribute is again *Farmyard Tidiness*. The class attending to all attributes (Class 1) has a membership probability on average of 0.0566, four times higher than in Model 2, while the class ignoring all attributes has a similar membership probability to that of Model 2, 0.0512. We note that the proportion of each AN-A combined with *Cost* is much higher for each attribute than the proportion of non-attendance for a single attribute. For example, *Stonewalls* have a probability of being neglected jointly with *Cost* of 0.1995 but only 0.0257 ignore only *Stonewalls*. Similarly, *Farmyard Tidiness* has a probability of being ignored jointly with *Cost* of 0.5033, but of only 0.008 on its own. Model 3 fits the data much better than Models 1 and 2, with a significant increase in fit according to both the maximised log-likelihood value and to statistical criteria accounting for parameter proliferation (the additional constants for the membership probability equations),

7 In this case, this restriction does not appear to be binding since models allowing for simultaneous non-attendance to three attributes show that following this strategy has an average probability of less than 0.03.

such as the BIC, AIC and AIC3. Finally, the individual parameter estimates are much sharper as denoted by the much higher *z*-values, especially the estimate for the *Cost* coefficient whose efficiency has repercussion in the estimation of marginal WTPs.

Such dramatic increase in the fit and efficiency of the model combined with the much larger fraction of the sample included in the four classes of pair-wise non-attendance, are consistent with the notion that when respondents adopt non-attendance as a decision-making heuristic for attribute processing, such a process is likely to involve pairs of attributes, rather than only one. For this reason, Model 4 includes all possible combinations of classes with two attributes unattended to, introducing six combinations of pairs of non-monetary attributes not attended to. Such a model, however, produces results that are not dissimilar to those from Model 3, in terms of both fit and membership probability.

Turning to the implied welfare measures, the point estimates for marginal WTP implied by the basic multinomial logit model (Model 1) and by the model accounting only for total non-attendance of all attributes and of single attributes (Model 2) are very large, while their precision is low, suggesting a lack of face-validity for the purpose of nonmarket valuation.[8] The fact that the WTP estimates from the MNL model are large is unsurprising as this model forces those who have ignored price to be fitted a non-zero coefficient whose estimate is inevitably much closer to zero than it would be if non-attendance (i.e. zero value to the *Cost* coefficient) were allowed for separately from attendance. The outcome is a *Cost* coefficient that induces a much higher WTP value, albeit more imprecisely estimated and with a relatively poor overall model fit. Relaxing only partially this restrictive assumption that all attributes are attended to by all respondents, as is done in Model 2, although significantly improving overall model fit still does not allow sufficient flexibility to accommodate the different ways in which the C*ost* attribute is neglected. A salient feature of our results is that when non-attendance is practised by respondents as an attribute processing strategy it seems frequently to involve *pairs* of attributes (i.e. *PNA*2 respondent types), with one element of the pair often being the *Cost* attribute.

If most people chose alternatives as if cost was of little or no consequence, they behaved as if they had a marginal utility of income of zero or close to zero. This, of course, is probably not a realistic representation of their preferences with respect to money. We can only speculate as to why this choice behaviour was so frequently shown. It might be an outcome of specific decision heuristics compounded with the fact that the hypothetical context gives insufficient penalties for incorrect choices. Perhaps people were distracted by the rural landscape images they were asked to evaluate and decided to focus on those and neglect cost, or perhaps respondents did not have sufficient

8 We collected supplementary data on stated non-attendance and estimated models accounting for this but the results were very similar, adding confidence to the evidence reported herein. These findings are available from the first author.

familiarity with choices that trade off rural landscape types against money and so could not rationalise the choice task. How to identify the true causes of this behaviour and how to optimally select the ranges of attribute levels that are of true relevance for each respondent remain the topic of further research.

We note though that there is an increasing body of evidence consistent with the notion that cost non-attendance is quite common. For example, Campbell (2008) found that 70 per cent of respondents seem to have ignored cost in a study of endangered fish species conservation; Scarpa *et al.* (2009) found that, depending on the set of responses used, the range of non-attendance varies from 40 to 80 per cent in rank-ordered data of Alpine grazing areas; Gilbride *et al.* (2006), in a completely different context, found non-attendance to cost equal to 57 per cent; Hensher (2008) found that non-attendance varies between 5 and 30 per cent, while Puckett and Hensher (2008) reported it to be up to 5 per cent.[9]

In this context, can we infer that the 90 per cent of respondents who did not attend to the cost attribute have the same marginal utility of money as the minority who attended to the cost coefficient? Under the ECLC and STAS model assumptions, such an estimate represents the best estimate of marginal utility of income because it is conditional on attending to cost, and corroborated by the fact that the WTP estimates seem of appropriate magnitude. In contrast, using the standard MNL model, and hence ignoring non-attendance by assuming that every respondent attends to all attributes, gives unreasonably high WTP estimates. The fact that both the ECLC and STAS models produce estimated marginal WTP values for all attributes that are much more reasonable than those derived from the MNL model is encouraging as it represents a better alternative than ignoring the issue and relying on much higher WTP estimates, which for many would invalidate the whole stated preference approach to valuation of landscape attributes.

It is unfortunate that in the study reported here so few people attended to cost. We can only report these results and signal the issue as a research topic to be addressed by future research. This clearly remains a critical assumption. A recently introduced alternative approach, which makes marginal WTP estimates somewhat less dependent on the estimates of marginal utility of income, employs utility specifications in the WTP space (see Train and Weeks, 2005; Scarpa *et al.*, 2008; Balcombe *et al.*, 2009). The use of 'cheap talk' scripts during interviews could also be a possible solution.

4.2. STAS results

The Bayesian estimates for the two continuous mixing taste parameter models are reported in Table 3. The first model is the Bayesian equivalent to a panel random parameter logit, whereas the second model is an STAS model, where

9 In these last two studies, there were two cost attributes and hence percentages vary depending on the cost attribute. For example, running cost and toll or fuel cost and other variable user charges were used.

Table 3. Stochastic attribute selection estimates

	Model B1 — MNL		Model B2 — STAS		
	$\bar{\beta}$ (post. std. dev.)	$\overline{WTP}$ (post. std. dev.)	$\bar{\beta}$ (post. std. dev.)	$\overline{WTP}$ (post. std. dev.)	$\bar{\theta}$ (post. std. dev.)
Mountain land (A lot)	1.893	309.64	3.108	40.80	
	(0.13)	(272.8)	(0.40)	(7.12)	0.750
Mountain land (Some)	1.239	202.60	1.978	25.95	(0.08)
	(0.14)	(185.32)	(0.28)	(4.72)	
Stonewalls (A lot)	1.907	317.46	2.647	34.78	
	(0.16)	(320.92)	(0.28)	(5.64)	0.864
Stonewalls (Some)	1.496	248.48	1.971	25.91	(0.07)
	(0.17)	(250.91)	(0.28)	(4.90)	
Farmyard tidiness (A lot)	1.525	251.49	3.202	42.19	
	(0.17)	(238.67)	(0.68)	(11.06)	0.632
Farmyard tidiness (Some)	1.188	195.43	2.379	31.35	(0.09)
	(0.15)	(182.31)	(0.49)	(8.10)	
Cultural heritage (A lot)	1.596	262.94	2.197	28.80	
	(0.14)	(237.52)	(0.26)	(4.50)	0.896
Cultural heritage (Some)	1.322	218.20	1.807	23.69	(0.07)
	(0.14)	(198.44)	(0.23)	(3.95)	
Cost	−0.728		−7.712		0.168
	(0.23)		(0.82)		(0.04)
LMD	−1,310.79		−1,201.61		

the parameter $\bar{\theta}$ is interpretable as a probability of the attribute being attended to at the respondent level. $\bar{\beta}$ is the vector of normal means (additional details on model parameters are available from the authors). In the Bayesian analysis, the *Cost* variable was divided by 100 so that the estimated βs would be of similar scale and a common (proper, but non-informative) prior distribution could be specified. All MCMC chains were run for 50,000 iterations and a sample of every tenth from the last 10,000 iterations was used to describe the posterior moments of the distributions.

The STAS model accounts for taste variation across individuals beyond that caused by non-attendance alone, and it is hence a more realistic model than both MNL and ECLC. The values of the log of the marginal densities (LMD) suggest that the STAS model outperforms the conventional specification implying total attribute attendance. The LMD is a Bayesian measure of model fit with an implicit penalty for the number of parameters (see Newton and Raftery, 1994). Similar to the latent class analysis, the implied estimates of marginal WTP are substantially smaller in the STAS model, and possibly would be considered more realistic by many analysts. WTP estimates and posterior standard deviations are calculated using the posterior distribution of $\bar{\beta}$. Alternative calculations were also conducted using the individual level posteriors β_n. For the STAS model, individual level estimates of

WTP are calculated as $-\beta_{nh}/\beta_{n\$}$ only if $\tau_{nh}=1$ and $\tau_{h\$}=1$. This method produced comparable mean values for WTP, but posterior standard deviations were an order of magnitude higher. This accurately reflects the uncertainty in obtaining individual level estimates with only five or six choice occasions and the inherent instability of taking the ratio of two normally distributed variables. The lower WTP values of Tables 1 and 2 are consistent with what is reported by Hensher (2008) and Campbell *et al.* (2008) when AN-A is accounted for.

Comparing the results of the probability of non-attendance derived from the STAS results to those obtained by Models 3 and 4 in Table 2, we note that the range of variation of attendance probabilities is smaller, with a minimum for the *Cost* attribute of 0.168 and a maximum for *Cultural Heritage* of 0.896, with *Stonewalls* taking the second place with 0.864 followed by *Mountain Land* (which was instead found to be the most attended to by the ECLC approach) with 0.750 and *Farmyard Tidiness* with 0.632. While the ranking in attendance frequencies are somewhat different between the two approaches, both results suggest an important role for non-attendance of the *Cost* attribute and a relatively low attendance to *Farmyard Tidiness*.

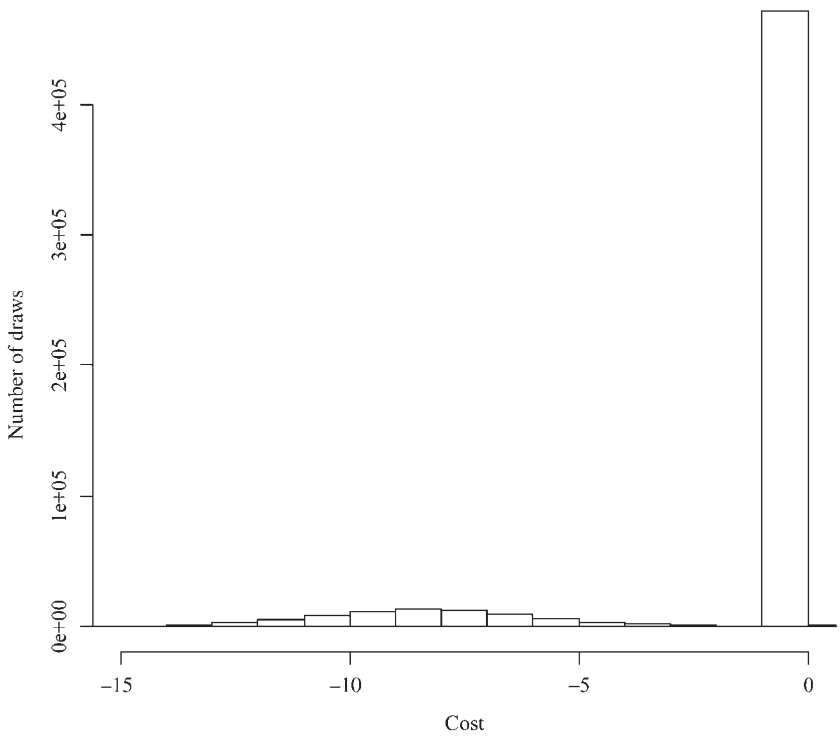

Figure 1. Histogram of 1,000 draws from the posterior distribution for all 553 respondents from model B2.

The Bayesian model also indicates a high probability of non-attendance for more than one attribute. Similar to the ECLC Model 3, there is a 0.067 probability of attending to all attributes. There is a 0.372 probability of non-attendance to only one attribute and a 0.561 probability of non-attendance to two or more attributes. The results also suggest that *Cost* plays a prominent role in the pattern of non-attendance: there is a 0.301 probability of non-attendance to cost alone, but a 0.532 probability of non-attendance to cost in conjunction with one or more other attributes. Figure 1 illustrates the distribution of the cost coefficient and shows the spike in density in the proximity of zero.

5. Conclusion

While multi-attribute stated preference methods have been largely applied under the assumption that respondents process *all* attributes used to illustrate choice alternatives, there is a growing concern and evidence that the real decision heuristics employed by respondents often involve selective attribute attendance. To date, little exploration of the effects of such phenomenon has been reported in the public goods non-market valuation literature, yet it appears to be crucial. When respondents systematically neglect the cost attribute, the effect of modelling such neglect on final welfare estimates can be of particular relevance. This paper introduces two ways of modelling probabilities of non-attendance to single attributes at the respondent level, and uses a well-known data set to illustrate the advantages of such sensitivity analysis in the context of non-market valuation. Both approaches produce concordant results in suggesting that: (a) AN-A is frequent in these data, (b) its treatment and identification are relevant for estimation outcomes as they significantly improve goodness-of-fit as well as efficiency of coefficient estimates and (c) strongly affect the estimation of non-market values. In particular, WTP estimates are much lower and of a magnitude that many analysts would probably find more realistic when such heuristics are explicitly addressed in modelling of discrete choice, as also evidenced by studies using self-stated AN-A (Hensher, 2008; Campbell *et al.*, 2008).

Both modelling approaches indicate that, in our sample, the probability of respondents acting according to the conventional assumption of considering (and valuing) all attributes proposed by the researcher is less than 0.1. Most respondents seem to have ignored at least two of the five attributes and the *Cost* attribute was often one element of the pair. Only a minority ignored only one attribute. Importantly for non-market valuation, the money coefficient appears to have been ignored by 80 to 90 per cent of respondents, depending on which model results one considers. Apparently for many of these respondents, an annual tax increase of €15–€80 was too small to influence their decision-making. The resulting differences in estimated marginal WTP are found to be of one order of magnitude smaller than the excessively large estimates implied by a naïve MNL model. While this might be partly due to the nature of the study, in which respondents engaged in comparative image evaluation of alternative landscape descriptions, such a high fraction of non-attendants to the cost attribute should by itself justify the routine application of this kind of statistical investigations.

The model that allows for taste heterogeneity across respondents with STAS suggests a slightly different frequency of attendance by attribute, but confirms that *Farmyard Tidiness* and *Cost* were the attributes least attended to, with significant repercussions in the implied welfare estimates. We deliberately chose a data set that shows a conspicuous discrepancy between WTP estimates in the presence and absence of non-attendance. Applications of the latent class approach to other data sets are available from the authors and also showed both significant model improvement and similar sensitivity of the implied WTP estimates to addressing the issue of attribute attendance. Hensher (2008) and Carlsson *et al.* (2008) found high incidence of non-attendance based on actual evidence from supplementary questions, although not as high as herein.

Analysts engaged in non-market valuation of benefits from discrete choice multi-attribute stated preference data should benefit from using such approaches to evaluate the robustness of their WTP estimates to violations of the common assumption of complete attribute attendance. Indeed, this should become a recommended course of action in practice. We note that both approaches proposed here are applicable to data sets without supplementary questions, and can hence be employed to investigate the sensitivity of estimates of older studies. Extending this methodology to 'standard' mixed logit models would be a useful area for further research and for development of commercial software.

Further research on methods to evaluate the robustness of welfare estimates from stated preference data to attribute attendance, and more generally to attribute processing strategies, could benefit from approaches based on utility specifications on the WTP space. Bayesian estimators for such models have been proposed by Train and Weeks (2005) and Sonnier *et al.* (2007) and may be a useful extension of the STAS model proposed here, whereas Scarpa *et al.* (2008) compared results from Bayesian and simulated maximum likelihood estimates. More research is required to identify appropriate supplementary questions that give an opportunity to test choice process and outcome models based on information that is individual specific, in contrast to reliance on analytical assumptions within a model, no matter how appealing they might be. Cognitive psychologists would support embedding supplementary questions as illustrated by recent research by Hensher (2008). The use of 'cheap talk scripts' (Cummings and Taylor, 1999) could also be explored.

References

Allenby, G. M. and Rossi, P. E. (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics* 89: 57–78.

Balcombe, K., Fraser, I. and Chalak, A. (2009). Model selection in the Bayesian mixed logit. *Journal of Environmental Economics and Management*, 57: 226–237

Campbell, D. 2008. Identification and analysis of discontinuous preferences in discrete choice experiments. Paper presented to the European Association of Environmental and Resource Economists Annual Conference Gothenburg, Sweden, 25–28 June 2008.

Campbell, D., Hutchinson, W. G. and Scarpa, R. (2008). Incorporating discontinuous preferences into the analysis of discrete choice experiments. *Environmental and Resource Economics* 41: 401–417.

Campbell, D., Hutchinson, W. G. and Scarpa, R. (2009). Using choice experiments to explore the spatial distribution of willingness to pay for rural landscape improvements. *Environment and Planning A*, 41: 97–111.

Carlsson, F., Frykblom, P. and Lagerkvist, C. J. (2007). Consumer willingness to pay for farm animal welfare: mobile abattoirs versus transportation to slaughter. *European Review of Agricultural Economics* 34: 321–344.

Carlsson, F., Kataria, M. and Lampi, E. (2008). Ignoring attributes in choice experiments. Working paper, School of Business, Economics and Law, Göteborg University.

Chen, M., Ibrahim, J. G. and Yiannoutsos, C. (1999). Prior elicitation, variable selection and Bayesian computation for logistic regression models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 61: 223–242.

Cicia, G. and Scarpa, R. (2000). Willingness to Pay for Rural Landscape Preservation: A Case Study in Mediterranean Agriculture, FEEM Working Paper No. 59.2000.

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: 649–665.

Fasolo, B., McClelland, G. H. and Todd, P. M. (2007). Escaping the tyranny of choice: when fewer attributes make choice easier. *Marketing Theory and Decision* 7: 13–26.

Ferrini, S. and Scarpa, R. (2007). Designs with a-priori information for nonmarket valuation with choice-experiments: a Monte Carlo study. *Journal of Environmental Economics and Management* 53: 342–363.

George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs Sampling. *Journal of the American Statistical Association* 88: 881–889.

George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica* 7: 339–373.

Geweke, J. (1996). Variable selection and model comparison in regression. In: J. M., Bernardo, J. O., Berger, A. P., Dawid and A. F. M., Smith, (eds), *Bayesian Statistics 5*. Oxford: Oxford Press, 609–620.

Gilbride, T. J., Allenby, G. M. and Brazell, J. D. (2006). Models for heterogeneous variable selection. *Journal of Marketing Research* 43: 420–430.

Gowdy, J. M. and Mayumi, K. (2001). Reformulating the foundations of consumer choice theory and environmental valuation. *Ecological Economics* 39: 223–237.

Greene, W. H. and Hensher, D. A. (2008). Ordered choice, heterogeneity attribute processing. Working paper, Institute of Transport and Logistics Studies, The University of Sydney.

Hanley, N., Colombo, S., Tinch, D, Black, A. and Aftab, A. (2006). Estimating the benefits of water quality improvements under the Water Framework Directive: are benefits transferable? *European Review of Agricultural Economics* 33: 391–413.

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

Hensher, D. A. (2008). Joint estimation of process and outcome in choice experiments and implications for willingness to pay. *Journal of Transport Economics and Policy* 42: 297–322.

Hensher, D. A. (2010). Hypothetical bias, stated choice studies and willingness to pay. For special issue of *Transportation Research* B (forthcoming).

Hensher, D. A. and Layton, D. (2008). Attribute referencing, cognitive rationalisation and implications for willingness to pay. Working paper, Institute of Transport and Logistics Studies, The University of Sydney.

Hensher, D. A., Rose, J. M. and Greene, W. H. (2005). The implications on willingness to pay of respondents ignoring specific attributes. *Transportation* 32: 203–222.

Islam, T., Louviere, J. J. and Burke, P. F. (2007). Modeling the effects of including/excluding attributes in choice experiments on systematic and random components. *International Journal of Market Research* 24: 289–300.

Lancsar, E. and Louviere, J. J. (2006). Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Economics* 15: 797–811.

Lockwood, M. (1996). Non-compensatory preference structures in non-market valuation of natural area policy. *Australian Journal of Agricultural Economics* 40: 73–87.

McIntosh, E. and Ryan, M. (2002). Using discrete choice experiments to derive welfare estimates for the provision of elective surgery: implications of discontinuous preferences. *Journal of Economic Psychology* 23: 367–382.

Newton, M. A. and Raftery, A. E. (1994). Approximating Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society B* 56: 3–48.

Pruckner, G. J. (1994). Agricultural landscape cultivation in Austria: An application of the CVM, *European Review of Agricultural Economics* 22: 173–190.

Puckett, S. M. and Hensher, D. A. (2008) The role of attribute processing strategies in estimating the preferences of road freight stakeholders under variable road user charges, *Transportation Research E*, 44: 379–395.

Randall, A. (2002). Valuing the outputs of multifunctional agriculture. *European Review of Agriculture Economics* 29: 289–307.

Raftery, A. E., Madigan, D. and Hoeting, J. A. (1997). Bayesian model averaging for linear regression model. *Journal of the American Statistical Association* 92: 179–191.

Rekola, M. (2003). Lexicographic preferences in contingent valuation: a theoretical framework with illustrations. *Land Economics* 79: 277–291.

Rigby, D. and Burton, M. (2005). Preference heterogeneity and GM food in the UK. *European Review of Agricultural Economics* 32: 269–288.

Rigby, D. and Burton, M. (2006). Modeling disinterest and dislike: a bounded Bayesian mixed logit model of the UK market for GM food. *Environmental and Resource Economics* 33: 485–509

Rosenberger, R. S., Peterson, G. L., Clarke, A. and Brown, T. C. (2003). Measuring dispositions for lexicographic preferences of environmental goods: integrating economics, psychology and ethics. *Ecological Economics* 44: 63–76.

Rossi, P. E., Allenby, G. M. and McCulloch, R. (2005). *Bayesian Statistics and Marketing*. West Sussex: John Wiley & Sons Ltd.

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

Sælensminde, K. (2002). The impact of choice inconsistencies in stated choice studies. *Environmental and Resource Economics* 23: 403–420.

Scarpa, R., Campbell, D. and Hutchinson, W. G. (2007). Benefits estimates for landscape improvements: sequential Bayesian design and respondents' rationality in a choice experiment. *Land Economics* 83: 617–634.

Scarpa, R., Thiene, M. and Train, K. E. (2008). Utility in WTP space: a tool to address confounding random scale effects in destination choice to the Alps. *American Journal of Agricultural Economics* 90: 994–1010.

Scarpa, R., Notaro, S. and Raffelli, R. (2009). *Non-attendance in Best/Worst Rank Ordered Choice Data: an Explorative Investigation.* Department of Economics, University of Waikato, New Zealand.

Sonnier, G., Ainslie, A. and Otter, T. (2007). Heterogeneity distributions of willingness-to-pay in choice models. *Quantitative Marketing and Economics* 5: 313–331.

Spash, C. L. (2000). Ecosystems, contingent valuation and ethics: the case of wetland recreation. *Ecological Economics* 34: 195–215.

Swait, J. D. (2001). A non-compensatory choice model incorporating attribute cutoffs. *Transportation Research Part B* 35: 903–928.

Swait, J. D. and Adamowicz, W. L. (2001a). The influence of task complexity on consumer choice: a latent class model of decision strategy switching. *Journal of Consumer Research* 28: 135–148.

Swait, J. D. and Adamowicz, W. L. (2001b). Choice environment, market complexity, and consumer behavior: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organizational Behavior and Human Decision Processes* 49: 1–27.

Train, K. E. (2003). *Discrete Choice Methods with Simulation*. Cambridge: Press Syndicate of the University of Cambridge.

Train, K. E. and Weeks, M. (2005). Discrete choice models in preference space and willing-to-pay space. In: R., Scarpa and A., Alberini, (eds), *Applications of Simulation Methods in Environmental and Resource Economics*. Dordrecht: Springer, 1–16.

Vermunt, J. K. and Magidson, J. (2005). *Technical Guide for Latent GOLD 4.0: Basic and Advanced*. Belmont Massachusetts: Statistical Innovations Inc.

Corresponding author: Riccardo Scarpa, Department of Economics, Waikato Management School, University of Waikato, Hamilton, New Zealand. E-mail: rscarpa@waikato.ac.nz, Web: http://www.mngt.waikato.ac.nz/ric/