



**UNIVERSITY OF
STIRLING**

DEPARTMENT OF
ECONOMICS

Modelling preference heterogeneity in stated choice data: an analysis for public goods generated by agriculture

Sergio Colombo

Nick Hanley

Jordan Louviere

Stirling Economics Discussion Paper 2008-28

December 2008

Online at <http://www.economics.stir.ac.uk>

**Modelling preference heterogeneity in stated choice data: an analysis for public
goods generated by agriculture.**

Sergio Colombo¹ Nick Hanley² and Jordan Louviere³

1. Department of Agricultural Economics, Andalusian Institute of Agrarian Research (IFAPA), Granada, Spain.

2. Economics Department, University of Stirling, Stirling FK9 4LA, Scotland.

3. Centre for the Study of Choice (CenSoC) and School of Marketing, University of Technology, Sydney

December 2008.

Abstract.

Stated choice models based on the random utility framework are becoming increasingly popular in the applied economics literature. The need to account for respondents' preference heterogeneity in such models has motivated researchers in agricultural, environmental, health and transport economics to apply random parameter logit and latent class models. In most of the published literature these models incorporate heterogeneity in preferences through the systematic component of utility. An alternative approach is to investigate heterogeneity through the random component of utility, and covariance heterogeneity models are one means of doing this. In this paper we compare these alternative ways of incorporating preference heterogeneity in stated choice models and evaluate how the selection of approach affects welfare estimates in a given empirical application. We find that a Latent Class approach fits our data best but all the models perform well in terms of out-of-sample predictions. Finally, we discuss what criteria a researcher can use to decide which approach is most appropriate for a given data set.

JEL codes: Q51, Q57, C52

Keywords: choice experiments, covariance heterogeneity model, agri-environmental policy, landscape values, latent class model, preference heterogeneity, random parameter logit model, error component models, welfare measures.

Corresponding author: Dr. Sergio Colombo, Agricultural Economics Department, IFAPA, 18080, apdo 2027, Granada, Spain. Telephone (+) 34 958 895267. E-mail: scolombo@ugr.es.

We thank the UK Department for the Environment, Food and Rural Affairs for funding data collection for this paper, and two referees for their extremely helpful comments on an earlier version of this paper.

1. Introduction

In the last 15 years, Choice Modelling using stated preference data has become an increasingly valuable tool in agricultural, transport, health and environmental economics (Louviere et al., 2000). This is largely due to the ability of this method to estimate marginal values for different attributes of various goods and services (both market and non-market); to estimate welfare effects of changes in these attributes; and to predict market shares. Choice Modelling is an empirical application of the Random Utility Model (Manski, 1977), in which it is assumed that individual i 's indirect utility function U can be represented by two, separable components:

$$U_{in} = V_{in} + \varepsilon_{in} \quad (1)$$

where U_{in} is latent utility, V_{in} is the systematic, or observable element of utility for individual n from choice alternative i , and ε_{in} is the random, or unobservable element of utility associated with option i and individual n . Discrete “choice alternatives” may be alternative travel-to-work modes, recreational site choices, or health care options. The original statistical “workhorse” for Choice Modelling was the multinomial (MNL) or conditional logit model (McFadden, 1973), which possesses many advantages in terms of closed-form solution and simplicity of interpretation and use¹. For example, most applications in environmental economics in the 1990s used the conditional logit model (Hanley et al., 2003).

However, the conditional logit model assumes preference homogeneity across respondents, such that only one fixed vector of parameter estimates is estimated for for the choice attributes. If one interprets the parameter associated with any attribute as its marginal utility (albeit confounded by a scale parameter), this implies that all respondents have the same tastes for that attribute. Socio-economic variables can be included as interactions with attributes or as interactions with alternative-specific constants, or different models can be estimated for different subsets of data (e.g. rural versus urban households, or higher versus lower income respondents), but these are relatively crude ways of representing preference heterogeneity. Such drawbacks have led to increasing

dissatisfaction with the conditional logit approach. Finding better ways to represent heterogeneity in choice modelling is important if researchers are to improve their understanding of the factors underlying consumer behaviour and willingness to pay, and how the benefits and costs of policies are distributed across recipients. Resource managers (for example, national park managers) also can benefit from knowing which groups of users derive relatively higher values from management decisions such as changes in site access or site quality.

Partly as a response to the perceived weaknesses of the Conditional Logit approach in this respect, the Random Parameters Logit (RPL) or “mixed logit” model has grown in popularity with discrete choice modellers (Train, 1998; McFadden and Train, 2000). In this approach (described in more detail below) the utility function for respondent n choosing over alternatives J is augmented with a vector of parameters that incorporate individual preference deviations with respect to the mean. Other models also have been developed to represent heterogeneity, principally the Latent Class (LC) model (Kamakura and Russell, 1989; Boxall and Adamowicz, 2002). LC models capture heterogeneity by assuming that the underlying distribution of tastes can be represented by a discrete distribution, with a small number of mass points that can be interpreted as different groups or segments of individuals. Preferences in each “latent” (that is, unobserved) class are assumed homogeneous; but preferences, and hence utility functions, can vary between segments (more details are again provided below).

Both the RPL and LC approaches incorporate heterogeneity in preferences through the systematic component of utility, V^j . An alternative is to include heterogeneity in terms of the random component of utility, ε . The Covariance Heterogeneity (Cov-Het) model allows the scale parameter (error variance) to be a function of choice attributes and respondents’ socioeconomic characteristics, by specifying scale (or equivalently, the inverse of error variance) to be a function of choice alternative attributes and/or individuals’ characteristics (Bhat, 1997). This adds useful information regarding the sources of sample heterogeneity.

¹ We discuss an alternative interpretation of the mixed logit approach later on, which focuses on heterogeneity in the random component of utility. For now, we focus on the currently-dominant way of applying RPL models.

As others have noted (eg, Carlsson et al 2003; Birol et al 2006), incorrect treatment of preference heterogeneity in stated choice data can lead to misleading estimation results. In the case of this particular paper, our interest lies in the consequences for welfare measures of how one chooses to capture preference heterogeneity, focussing on a comparison of random parameter, latent class and Cov Het approaches. Whilst several authors have compared random parameter and latent class approaches to choice data, we are unaware of other papers which contrast these approaches with the Cov Het model. We also consider the relative predictive ability of these three approaches. The empirical context of our comparison relates to the provision of a public good (landscape quality) in agricultural upland areas of England, where it is reasonable to expect a high degree of preference heterogeneity to exist. That is, it would be unusual not to find a considerable degree of variation in how people value landscape, even within a given geographic region, since work in landscape planning has shown that perceptions of landscape “quality” are fluid and vary greatly across individuals (Strumse, 1994; Van den Berg, 1998). Moreover, existing environmental economic valuation studies have shown preference heterogeneity to exist for landscapes, for example between residents and visitors, and according to education and income levels (Hanley et al, 1998; Willis et al, 1995). Better understanding of preferences for landscapes can help improve design of “agri-environmental” schemes that pay farmers to produce public environmental goods, such as landscape features related to the manner in which farming is carried out in upland areas (Hanley et al, 2007).

2. Alternative Approaches to modelling preference heterogeneity

2.1. The Random Parameters Model

The Random Parameters (RPL) or mixed logit model has grown rapidly in popularity with discrete choice modellers (Train, 1998; McFadden and Train, 2000), despite concerns about the distributional assumptions that researchers use in applications (Rigby and Burton, 2006). Several reasons underlie this growth: RPL avoids the Independence of Irrelevant Alternatives property of

the Conditional Logit model, allows for random taste variations, and can incorporate correlations in unobserved factors over choice alternatives. Furthermore, McFadden and Train (2000) have demonstrated that any random utility model can be approximated, to any degree of accuracy, by a mixed logit model with the appropriate choice of variables and mixing distribution. As we note below, this extends to heterogeneity in the random component of the utility function.

The typical formulation of the RPL model decomposes (1) into an unobserved, preference heterogeneity component and a deterministic component, the latter representing the utility of respondent n choosing alternative j in choice situation t as a function of that alternative's attributes, X :

$$U_{njt} = \beta X_{njt} + \eta_n X_{njt} + \varepsilon_{njt} \quad (2)$$

where X_{njt} is a vector of observed attributes for the good in question, β is the vector of coefficients associated with these attributes, η_n is a vector of k standard deviation parameters and ε_{njt} is an unobserved random term which is independent of the other terms in the equation, and independently and identically Gumbel distributed. Under this specification each person has her own vector of parameters, β_n , which deviates from the population mean β by the vector η_n . Preference heterogeneity is thus directly incorporated into the random parameters approach. A further advantage of the RPL approach is that one can allow for the unobserved portion of utility ($\eta_n X_{njt} + \varepsilon_{njt}$) being correlated across choices for each respondent². The probability of individual n observed sequence of choices $[y_1, y_2, \dots, y_T]$ is calculated by solving the integral³:

$$P_{n[y_1, y_2, \dots, y_T]} = \int \dots \int \prod_t \left[\frac{e^{X_{njt}\beta_n}}{\sum_{i=1}^J e^{X_{mit}\beta_n}} \right] f(\beta) d\beta \quad (3)$$

Where j is the alternative chosen in choice occasion t . Integral (3) has no analytical solution but can be approximated by simulation. However, to estimate the model one must make assumptions about how the β coefficients are distributed over the population, $f(\beta)$, take a set of R draws from $f(\beta)$ and then calculate the logit probability for each draw. Train (2003) shows that the

simulated probability \hat{P}_n (equation 4) is an unbiased estimator of P_n whose variance decreases as R increases. Note that in equation (4) the index nr on β indicates that the probability is calculated for each respondent using R different sets of β vectors.

$$\hat{P}_n = \frac{1}{R} \sum_{r=1}^R \left(\prod_t \left[\frac{e^{X_{njt}\beta_{nr}}}{\sum_{i=1}^J e^{X_{nit}\beta_{nr}}} \right] f(\beta) d\beta \right) \quad (4)$$

The RPL analyst also has to decide on the parameterization of the covariance matrix. In this paper we assume preference parameters to be independent, so that the r^{th} draw of β_{nk} is taken using a diagonal variance-covariance matrix⁴.

RPL approaches to discrete choice data began with applications by Boyd and Mellman (1980) and Cardell and Dunbar (1980). However, advances due to faster computer calculations and simulation techniques made the model accessible to a wider audience by incorporating estimation routines into standard statistics software. Recent applications of the model extend over the fields of transportation (Hensher, 2001; Amador et. al, 2005), consumer choice (Revelt and Train, 1998), recreation (Train, 1998; Hanley et al. 2002), health (Personn, 2002) and waste management (Layton 2000). Environmental applications have increased markedly, see for example Carlsson et al. 2003; Colombo et al. 2005; Hanley et al. 2006; and Birol et al. 2006.

2.2. The Latent Class Model

Preference heterogeneity is captured in Latent Class (LC) models by simultaneously assigning individuals into behavioural groups or latent segments whilst estimating a choice model. Within each “latent” (ie, unobserved) class, preferences are assumed homogeneous, but preferences, and so utility functions can vary between segments. Thus, LC models allow one to explain preference differences across individuals conditional on the probability of membership in a latent segment (grouping).

In the random utility model, if error terms are iid across individuals and classes with a type I extreme value distribution, the choice probability of the sequence of choices of individual n , who belongs to class s , is expressed as:

$$P_{n|s} = \prod_t \left[\frac{\exp(\beta_s X_{nit})}{\sum_{j=1}^J \exp(\beta_s X_{njt})} \right] \quad s = 1, \dots, S \quad (5)$$

where β_s is the parameter vector of class s associated with a vector of explanatory choice attributes X_{nit} . Additionally, one can allow the classification model to be a function of individual-specific covariates that underlie allocation of individuals to the s classes. The membership probability of class s is given by:

$$P_{ns} = \frac{\exp(\alpha_s Z_n)}{\sum_{s=1}^S \exp(\alpha_s Z_n)} \quad s = 1, \dots, S, \quad (6)$$

which is a multinomial logit process where individual-specific characteristics (Z_n) and not attributes (X_{nit}) underlie choice probabilities⁵.

Given the assumption of independence between the probabilities of equation (5) and (6)⁶ the joint probabilities that a randomly chosen individual n sequence of choices is given by:

$$P_n([y_1, y_2, \dots, y_T]) = \sum_{s=1}^S \left[\frac{\exp(\alpha_s Z_n)}{\sum_{s=1}^S \exp(\alpha_s Z_n)} \right] \left[\prod_t \frac{\exp(\beta_s X_{nit})}{\sum_{j=1}^J \exp(\beta_s X_{njt})} \right] \quad s = 1, \dots, S \quad (7)$$

where the first expression in brackets is the probability of observing the individual in class s , and the second term is the probability of the sequence of choice $[y_1, y_2, \dots, y_T]$ conditional on belonging to class s . Equation (7) encapsulates the LC approach to choice modelling.

Applications of LC models were first proposed by Kamakura and Russell (1989), and are widely used in marketing, as noted by McFadden (2000); however, their use in applied economics is relatively new. For example, Boxall and Adamowicz (2002) estimated a LC model to describe recreational choices of wilderness parks; Green and Hensher (2003) used a LC approach to model

choice of long distance travel mode; Provencher and Bishop (2004) applied a LC model to the choice of recreational fishing sites, Birol et al. (2006) used it to model preferences for wetland attributes, and Shen et al. (2006) for transport mode choice. Other recent LC applications include Morey et al (2006), Milon and Scrogin (2006) and Ruto et al (2008).

2.3 The Covariance Heterogeneity Model

The Cov Het model estimated in this study is a generalization of the nested logit model (Bhat, 1997) where the inclusive value parameter for branch j is specified as an exponential function of covariates:

$$\tau_{jn} = \tau_j^* \exp[\delta z_n] \quad (8)$$

where τ_j^* is the inclusive value (IV) parameter of the nested logit model, z_n is a set of individual characteristics and δ a vector of parameters to be estimated. Because the IV is a scaling parameter for a common random component in the alternatives within a choice branch, a Cov Het model can be used to explain the heteroskedastic error structure present in the data. Individual covariates and attributes of choice alternative can influence both deterministic and stochastic utility components; and error variances of the conditional choice model (i.e., conditional on the nesting structure used) can be allowed to vary across individuals. Following Louviere et al. (2000) the probability that individual n chooses alternative i given branch j , $P(ni|j)$ and the probability that the same individual chooses branch j , $P(jn)$, are given respectively by:

$$P(ni | j) = \frac{\exp(\beta' x_{ni|j})}{\sum_{q=1}^Q \exp(\beta' x_{nq|j})} \quad (9)$$

$$P(nj) = \frac{\exp(\alpha' y_j + \tau_{jn} IV_j)}{\sum_{j=1}^J \exp(\alpha' y_j + \tau_{jn} IV_j)} \quad (10)$$

where $x_{ni|j}$ are variables that vary within nests, Q is the total number of alternatives in the branch j , J is the total number of branches, y_j are variables that vary across nests, IV_j is the inclusive value of

nest j ($IV_j = \ln \sum_{q=1}^Q e^{V_{qj}}$), V_{qj} is the utility of alternative q which belongs to nest j , and β , α , τ , are parameters to be estimated.

2.4. Considerations regarding “scale”.

As Louviere and Eagle (2006) note, RPL models are “... long on statistical theory, but short on behavioural theory”. Although this assertion may seem paradoxical (a motive to use such approaches being to better describe respondents’ choices, i.e., behaviour), it is true that there is little in the applied literature which tests the assumptions commonly adopted when estimating RPL models. For example, most published articles that use RPL models assume constant error variances across individuals and alternatives. If this is not true, estimated parameters are biased because the analyst cannot estimate unconfounded distributions of preferences. This is because the distribution of taste parameters will be confounded with distributions of scales. Louviere et al. (2002) argued that each person has a scale factor that is perfectly correlated with their parameter vector; hence the distribution estimated in RPL choice models is indeterminate. These authors also show that scale (which is inversely proportional to error variance) can be determined by many factors (e.g. individual characteristics, factors that vary over conditions, contexts or situations, time varying factors and geographical-spatial factors) and conclude that it is very unlikely that scale is constant. The LC model also suffers from the confounding between scale and estimated parameters. In particular, in LC models two types of scale factors cannot be estimated along with parameters⁷. One is the scale across the segment membership function (equation (6)); the second is the scale for the s^{th} segment’s utility function (equation (5))⁸.

Recently, a number of empirical applications have shown that the constant scale assumption is indeed violated. Dellaert et. all. (1999) parameterised scale as function of price attribute level differences and absolute values, and showed the variance of the error component (the inverse of scale) increased as price level differences and absolute price levels increased. Swait and Adamowicz (2001) used a heteroskedastic multinomial logit model where scale varied over

respondents as function of a measure of choice task complexity. They found scale was a function of complexity faced by individuals in 8 out of 10 studies they reviewed. De Shazo and Fermo (2002) also used a heteroskedastic logit model to test if scale varied as function of choice set complexity, finding that the variance of the error terms increased for all five measures of choice set complexity considered. Finally, Magidson and Vermunt (2007) found that the equal scale parameter assumption used to justify traditional latent class modelling resulted in misclassifying 37% of cases.

However, there is very little discussion in the literature about effects of violating a constant scale assumption on the measurement of willingness to pay. De Shazo en Fermo (2002) noted that failure to control for heteroskedasticity may overestimate measures of welfare change as much as 33%. When compensating surplus measures are of interest, and where several attributes change at the same time, it may be that an overstated value for one attribute is compensated by a lower value for another attribute in the same policy alternative. If so, the difference in the compensating surplus estimates between models which parameterise the scale compared with models that do not may be statistically insignificant.

2.5 On comparing alternative modelling approaches

As noted above, there are at least three contending ways to model preference heterogeneity in stated choice data. What criteria should be adopted to compare these approaches? As both RPL and LC methods focus on the deterministic component of utility, one criterion is to ask whether preference heterogeneity is more likely to be better represented via the deterministic component, or instead by the random (stochastic) component via a Cov Het approach. However, we also noted earlier most published articles that use RPL and LC assumed constant error variances; so, if error variances are in fact related to observable factors and non-constant, this suggests a Cov Het approach. A researcher also can ask if preferences are more likely to be unique for a given good, or grouped, which suggests a way to compare RPL and LC. A researcher also simply can ask which model fits choice data better (more comments follow on this below). Finally, an important issue is

whether one's choice of RPL, LC or Cov Het routines actually matters in terms of welfare estimates.

We now describe the empirical data that we will use to examine differences in marginal willingness to pay and compensating surplus estimates to compare RPL, LC and Cov Het, as noted above.

3. Empirical data

The data were derived from a choice experiment study aimed at estimating the public good benefits resulting from conserving upland hill farming in the North West region of England (more details can be found in Hanley et al, 2007). Historically, support payments to farmers in such areas were provided throughout the European Union based on livestock levels; since 2003 these payments were replaced by area-based payments under revisions to the "Less Favoured Areas" scheme. However, recent changes in the Common Agricultural Policy have created a need to replace area-based payments with something else, since area-based payments violate the principle of decoupling. The main alternative for support payments to upland farmers being considered by the UK government is a scheme based on the provision of public goods, in terms of landscape features and wildlife habitats "produced" by hill farming. Landscape is a good that can be described with an attributes-based approach typical of stated choice modelling in a manner useful for public policy design.

Focus groups with members of the general public in North-West England were used to identify relevant upland landscape attributes for inclusion in the study. The final attribute list comprised heather moorland and bog, rough grassland, broadleaf and mixed woodland, field boundaries (stone walls and hedges), and "cultural heritage". Cultural heritage was defined to include the presence in the landscape of traditional farm buildings, keeping of traditional livestock breeds, and traditional farming practices like shepherding with sheep dogs. Attributes were

described to respondents in the survey in both words and pictures. A copy of the survey is available from the authors.

In any choice experiment, alternatives are presented to respondents described in terms of the attributes and the “levels” that these take: for example, for an attribute representing the conservation of wetlands in New South Wales, these levels might be “20% of wetlands are conserved”, “50% of wetlands are conserved” and “80% of wetlands are conserved”. In our case, the selection of attribute levels was difficult due to the need to make quantitative predictions of the impacts of future agricultural policy changes. These predictions were made by experts, based on a literature review of recent rates of changes in the attributes (Cumulus et al., 2005). A short description of the final list of attributes levels is given in Table 1. We decided to describe changes in quantitative attributes in percentage terms instead of absolute value terms because we wanted to be able to make comparisons across other upland regions of England to which the same discrete choice experiment was applied (see Colombo and Hanley (2008) for details of this benefits transfer test). Attribute levels were designed to span a wide spectrum of policy options for the reform of support mechanisms for farmers in Less Favoured Areas in the UK for the period 2007-2013 (DEFRA, 2006). For example, this meant that the landscape attribute “heather moorland and bog” was described in any choice alternative using either a 12% decline, a 2% decline or a 5% increase, according to the experimental design plan. A tax attribute was included to allow calculation of welfare measures: this was specified as “higher national or local taxes”, in terms of additional tax payments per household per year. The survey explained to respondents that higher taxes would be needed during the period 2007-2013 to pay for the policy changes outlined in the choice cards. The tax levels were £2, £5, £10, £17, £40 and £70, with a baseline of no increase.

A shifted fractional factorial main effects experimental design (Louviere et al., 2000) was used to vary attributes and levels, creating eighteen initial profiles that were used to generate additional choice options to form the choice cards, following the approach proposed by Street et al. (2005) (see also Street and Burgess, 2007, and Ferrini and Scarpa, 2007). The initial eighteen

profiles were blocked into three versions of six choice cards each containing three alternatives: option A, option B and a status quo, as shown in Figure 1. The status quo alternative represented a continuation of current hill farming support mechanisms that was converted into a loss of 2% of the heather moorland and bog area, a reduction of 10% in the rough grassland, a 3% increase in the woodland area, 100 meters of restoration for every kilometre of field boundaries (hedges and stone walls), and a rapid decline of cultural heritage. A zero additional tax (price) was associated with the status quo. Generic alternatives A and B contained variations in the attribute levels, but with a positive tax price, representing modification to current policy support. These tax prices varied between £2 and £70 per household per year, levels chosen based on focus group discussions and pilot tests.

A pilot survey was conducted during July 2005 with a sample of 50 respondents to test the coverage, wording, length and the design of the survey. The final survey was conducted by two market research firms in the summer and autumn of 2005. Three-hundred respondents were interviewed according to quotas for age, gender, socio-economic group and whether they lived in an urban or rural area. The survey mode was face-to-face, door-to-door personal interviews. Survey de-briefing questions and focus and pilot group responses convinced researchers that respondents could understand the choice tasks in the main survey: for example, 69% of the sample stated that the survey was “easy” to complete (Eftec, 2006)..

4. Methodology

Initial analysis showed that the sample was representative of the population in terms of age, gender, and income group, but slightly over-represented urban residents compared with rural respondents (further details are in Eftec, 2006). We now describe how each of the three modelling methods outlined in section 2 was applied.

4.1. Random Parameters Logit

Several possible distributions of the coefficients were considered (Rigby and Burton, 2006).

Because respondents may either like or dislike the landscape attributes in the choice experiment, we

assumed preferences for all random attributes followed a normal distribution. To test the effect of this assumption on parameter estimates we also specified a triangular distribution for the random parameters, and estimated this model. We found no significant differences in the normal distribution and triangular distribution models. Preferences towards the cost attribute were assumed homogeneous to facilitate interpretation of resulting welfare measures (Chen and Cosslett, 1998). Also, preferences of the “broad mixed woodland” attribute were considered homogeneous, as initial analysis showed that respondents had homogeneous tastes for this attribute. These assumptions implicitly constrain the scale parameter to be constant across respondents, as noted by Train and Weeks (2005)⁹.

4.2. Latent class

To use an LC model, analysts must determine the number of latent classes. There is no rigorous way to select the number of classes, and several ways have been used in the literature¹⁰. Here, we use the Akaike Information Criterion (AIC) and its corrected form based on sample size, the Consistent Akaike Information Criterion, CAIC. Both criteria penalise improvements in the log likelihood due to additional parameters included. The number of classes that minimise each of the measures suggests the preferred model. When different criteria indicate different preferred numbers of segments, Scarpa and Thiene (2005) note that selection “must also account for significance of parameter estimates and be tempered by the analyst’s own judgement on the meaning of the parameter signs”. In this study, both the AIC and CAIC were lowest for a three class model, so we retained this number of segments in model reporting..

4.3. Covariance Heterogeneity model

We use a two level nesting structure with the two environmental improvement alternatives (policy alternatives A and B) grouped in one branch and current conditions (status quo) in the other. This assumes that people first choose between the status quo and the alternatives given their

particular socio-economic characteristics. Later, at the second level of the nest, respondents are assumed to choose their preferred option (A or B) on the basis of landscape and cost attributes and a set of respondent characteristics. We added these covariates to capture observed heterogeneity in respondent preferences in the systematic component of utility. Heterogeneity in the random component of utility was parameterised by estimating the scale parameter as a function of a set of landscape attributes and individuals' characteristics.

4.4. Model comparison

Comparing different models is always challenging given the many domains of contrasts. A comparison between RPL, Cov Het and LC models cannot be carried out using conventional log likelihood ratio tests because the models are non-nested. Hence, we used the test proposed by Ben-Akiva and Swait (1986) for non-nested choice models. The test works as follows: suppose the two models we wish to compare use K_1 and K_2 variables to explain the same choices and assume that $K_1 \geq K_2$. Models either have different functional forms, or the two sets of variables differ by at least one element. The fitness measure for model j , $j=1,2$ is:

$$\rho_j^2 = 1 - \frac{L_j - K_j}{L(0)} \quad (11)$$

where L_j is the log likelihood at convergence for model j and $L(0)$ is the log likelihood assuming that choices are random. Ben-Akiva and Swait (1986) show that under the null hypothesis that model 2 (the more parsimonious specification) is the true model, the probability that the fitness measure of equation (11) for model 1 will be greater than that of model 2 is asymptotically bounded by the function:

$$\Pr(|\rho_2^2 - \rho_1^2| \geq Z) \leq \Phi(-\sqrt{-2ZL(0) + (K_1 - K_2)}) \quad (12)$$

where Z is the difference of the fitness measures between model 1 and model 2 and is assumed greater than 0, and Φ is the standard normal cumulative distribution function. Thus, equation (12) sets an upper bound for the probability that one incorrectly selects model 1 as the true model when

model 2 is the true model. Comparing absolute parameter estimates across models is not useful due to scale differences (Louviere et al., 2000). However, contrasting willingness to pay and compensating surplus measures is informative as it cancels the scale parameter, so we adopt this approach here.

4.5. Welfare analysis.

In choice experiments the coefficient of the cost attribute is interpreted as an estimate of the negative of the marginal utility of income. Model coefficients can therefore be used to provide welfare estimates for changes in attribute levels: the ratio of the coefficient of any attribute to the negative of the coefficient of the monetary attribute provides the “implicit prices” that represents willingness to pay (WTP) for a 1% or 1 unit increase in the quantity of the attribute in question if it is quantitative (eg area of heather moorland) or the WTP for a discrete change in the attribute (i.e., from “rapid decline” to “no change” in cultural heritage features, for instance) if it is qualitative. Welfare changes that relate to the outcome of a hypothesised policy option that changes several attribute levels simultaneously can be obtained by using the compensating surplus formula described by Hanemann (1984):

$$CV = -\frac{1}{\beta_m}(V^1 - V^0). \quad (13)$$

where β_m is the parameter estimate of cost, and V^0 and V^1 represent a representative agent’s utility before and after the change under consideration. In the case of the Cov Het model, we adapt this formula provided by Kling and Thomson (1996) for estimating compensating variation for a Nested logit model.

$$CV = -\frac{1}{\beta_m} \left\{ \ln \left[\sum_{m=1}^M \left(\sum_{j=1}^{J_m} e^{V_{mj}^1 / \alpha_m} \right)^{(\alpha_m)} \right] - \ln \left[\sum_{m=1}^M \left(\sum_{j=1}^{J_m} e^{V_{mj}^0 / \alpha_m} \right)^{(\alpha_m)} \right] \right\} \quad (14)$$

where M are the nodes and J the alternatives in each node and α_m is the inclusive value parameter of node M . Given the covariance heterogeneity structure, the inclusive value parameter is a function of respondent characteristics in this study. Compensating Surplus is arguably the most important measure from a policy perspective, as it describes how the results of alternative policy options affect social welfare.

5. Results

We first removed “protest bids” and surveys that were not fully completed from the 300 interviews, leaving 1,187 valid choice observations for estimating models. The percentage of protest bids was 26%. Coefficients estimates for RP, LC and Cov-Het models are in Table 2, whilst Table 3 describes the coding used in the tables. All models provide good fits to the data as measured by McFadden’s ρ^2 , and all are highly significant in explaining respondents’ choices.

Starting with the RPL model, the coefficients that are significant have signs as expected a priori. Increasing the area of “heather moorland and bog”, “rough grassland”, “broad and mixed woodlands” and improving “cultural heritage” gives positive utility to respondents. The tax coefficient is negative. Socio economic variables were introduced into the model as interactions with the constant, as the generic format of the choice task allows a convenient interpretation of the resulting coefficients¹¹. Respondents’ income was excluded from analysis due to high percentages of people who did not reveal their income, and because in analyses not shown here we found that income was not significant in driving choice. Younger and more educated respondents were more likely to choose support such initiatives. The sign of the “number of years respondents have been living in the region” is highly significant and indicates that people who have been living for longer in the region have a lower probability of contributing to landscape conservation. The significance of the landscape attributes’ standard deviation coefficients reveal that preferences for landscape are indeed heterogeneous. Taking into account this heterogeneity increases model fit a lot. In analyses not shown here, we estimated a conditional logit model that indicated that the sources of preference

heterogeneity in the data go well beyond inclusion of respondents' observable characteristics as interactions with the status quo choice, a result also observed by Revelt and Train (1998) and Scarpa et al. (2005).

Turning to the *LC* model, the three-class model specification allocated 30% of respondents to class one, 40% to class two, and 30% to class three. Note that the segment function parameters for segment three equal zero due to normalization in estimation. Thus, factors driving membership of the other two segments are described relative to segment three. We consider the utility function parameters first; none of them are significant at 95 confidence level for segment one, showing that respondents in this class are not interested in the proposed changes in landscape attributes (although two are significant at 90%). Class two utility parameters show positive preferences towards improving the area of "heather moorland and bog", "rough grassland" and the "cultural heritage" attributes. The sign of the "field boundary" attribute is contrary to expectations, since utility increase towards a diminishing of the length of field boundary. Importantly, the tax coefficient is negative, with a high absolute value relative to the coefficient for class three. This may be because members of class two are keen to preserve landscape features but are not willing to pay a lot of money for this. Class three utility parameters exhibit higher preferences for alternatives that offer increased levels for all the landscape attributes. For class membership functions (relative to class three), respondents in both class one and two are older and less educated.. The variable measuring whether people belong to an environmental organisation also influences class membership.

With regard to the Cov-Het model, starting from the branch choice variables, the sign and significance of socioeconomic variables reveal whether respondents are more likely to choose either options A or B than the status quo for each socioeconomic characteristic included. Interpretation of model estimates reveals, as observed in the RPL model, that younger and more educated people, who have not lived in the area long, and who are members of environmental organizations are more likely to be willing to pay for upland landscape conservation. Now considering interpretation of the attributes in the utility function, we again see that an increase in any of the landscape attribute

considered increases respondent utility. Inspection of the estimates of covariates in the inclusive value (IV) function reveals that individual characteristics affect the random component of utility. In turn, this indicates that error variances in the conditional choice model (conditional on one of the A or B alternatives being chosen) are systematically related to differences in respondents' socioeconomic characteristics. For example, conditional on respondents who chose the “do something” branch, a positive and significant “age” estimate implies that older people have higher scale parameters, (i.e., lower random component variances). The opposite is true for people who belong to any environmental or recreational organization, their scale parameters being higher.

Comparing the approaches

Comparison of RPL and LC models using the Ben-Akiva and Swait test gives a probability of $P \leq \Phi(-17.78) \cong 0$, indicating that the more parsimonious RPL is not the preferred model. The same test comparing LC and Cov-Het models gives a $P \leq \Phi(-20.58) \cong 0$. So, we can conclude that the LC model is superior to both RPL and Cov-Het models, implying that for these data 1) preference heterogeneity is better explained at a segment level than an individual level in terms of deterministic utility; 2) it is better to capture heterogeneity in the systematic utility component than the random component.

Comparing absolute parameter estimates across models is uninformative due to scale differences (Louviere, et al. 2000). So, we focus on comparing implicit prices and compensating surplus to test the effect of different approaches to modelling respondent heterogeneity in welfare measures. Table 4 gives attribute implicit prices (marginal willingness to pay values) for all models, along with the 95% confidence intervals estimated using the procedure proposed by Krinsky and Robb (1986). Respondents have positive WTP for increasing the area of all desirable landscape characteristics considered except for field boundaries. Interestingly, the models give different mean implicit price values. Comparing implicit prices of RPL to LC¹² (we consider the weighted sum of LC values in this comparison), all implicit prices differ at the 90% confidence level, except cultural

heritage at the highest level. Comparing implicit prices of LC and Cov-Het models gives similar results, with the exception that the implicit price of the field boundaries attribute no longer differs between them. Finally, only two implicit prices in the RPL model (“heather moorland and bog” and “broadleaf and mixed-woodland”) do not differ from implicit prices in the Cov-Het model. Thus, it seems clear that the way in which analysts treat heterogeneity can have an effect on marginal willingness to pay estimates.

Although implicit prices are useful to policy makers when defining priorities for policy interventions, they are not welfare measures that can be used in cost-benefit analyses of future policy scenarios as they do not give the willingness to pay of individuals for multiple changes in landscape attributes, nor do they give changes in the probability of a policy option being selected in the random utility model. To obtain compensating surplus figures requires the definition of the policy scenarios to be appraised. Table 5 shows the three policy scenarios used, taken from Cumulus et al. (2005) and the baseline to which these were compared, in terms of attribute levels¹³.

Table 6 shows compensating surplus estimates and 95% confidence intervals for each scenario for the three econometric models. The surplus estimates represent respondents’ average WTP to move from the state of the world given in the baseline to the state of the world given in Scenarios 1, 2 or 3. We observe a degree of similarity between the welfare estimates generated from the three different models and tight confidence intervals. Results of the Poe et al. (2005) test reveal that the welfare estimates of the RPL model do not differ significantly when compared to the LC or Cov-Het models at the 95% confidence level. This also holds when we compare welfare estimates of the LC model with those from the Cov-Het model. This suggests that 1) different ways of modelling preference heterogeneity do not seem to have big impacts on final welfare estimates, at least in this data set; and 2) whilst implicit prices differ across the models, the overall effect on WTP for different alternative policies can cancel out across attributes so that differences in WTP for alternative policies are not observed.

6. Discussion

Generally, it would be desirable to identify a way to model preference heterogeneity in choice experiments that is unambiguously preferred, or – more realistically- conditions under which analysts should prefer one approach to others. One way to do this is to think about the relative importance of systematic and random components of utility (as in equation 1) in the sample. If analysts have information that leads them to think that the sample heterogeneity is greater in the systematic component than the random component of $U(.)$, then a RPL (as implementd above)_or LC logit approach should be used. If they think that the a larger portion of heterogeneity is in the random component of $U(.)$, a COV-HET model should be used. So, analysts face the following possibilities:

		Random component	
		Low	High
Systematic component	Low	LL	LH
	High	HL	HH

In the first case (LL), there is a low heterogeneity in both utility components. Models such as conditional logit can be used to analyse the data as preference homogeneity is a reasonable working assumption. In case two (LH), heterogeneity is high in the random component and low in the systematic component, suggesting a Cov-Het model should be preferred. In case three (HL), heterogeneity is high in the systematic component and low in the random component, suggesting a RPL or LC model should be used.

That said, knowing in advance which utility component has more heterogeneity is difficult. Analysts might have some insight from focus groups, or from other studies in the literature, but it will typically not be possible to decide ex ante which quadrant in the above figure a particulare choice experiment data set will lie in. One way to analyse this positioning ex post may be to

consider the number of significant parameters exhibiting heterogeneity in different models, i.e. the standard deviations in RPL or covariates of IV parameters in Cov-Het models. One also can test in empirical applications how heterogeneity in systematic or random components affects choice probabilities. To do this, consider for example an RPL model and calculate the value of utility of an alternative that has the following attribute levels relative to the status quo: {HMB:+2; RG: +5; BMW:+3; FB:+50; CH: No change}. To account for the effect of heterogeneity in the systematic component we can measure the utility using the 10th and 90th percentiles of the distribution of each random coefficient¹⁴. In this example, the value of utility for an individual with 10th percentile parameters is -1.86 units¹⁵. The utility for an individual with 90th percentile parameters is +3.34 units. For a unitary scale parameter, the probability of choice of the example alternative is 0.13 for the 10th percentile individual and 0.96 for the 90th percentile individual. Thus, heterogeneity in the systematic component dramatically affects the probability of choice, and it is an important aspect of this dataset.

For the Cov-Het model the analyst can instead consider two “very different individuals” in the sample in term of socio-economic characteristics that affect the scale parameter, and use the model to determine how the different scale parameters affect choice probabilities for the same change. If heterogeneity in the systematic component has the greatest effect on the choice probability, then a “systematic” heterogeneity modelling approach should be preferred. If the choice probabilities are affected more by differences in scale, an approach that considers heterogeneity in scale should be preferred. For our data, we selected two individuals (A,B) who differ in their age, gender, membership of an environmental organisation, education level, and the time they have been living in the region¹⁶. For the choice alternative described above the choice probability of individual A is equal to 0.44 whilst for individual B it corresponds to 0.41, showing that the heterogeneity effects in the random utility component in the choice probability are much more “attenuated” than effects of heterogeneity in the systematic component. Because of that we consider

that an approach that models heterogeneity in the systematic component of utility should be preferred for this data.

To obtain additional insights about the different modelling approaches to preference heterogeneity we also compared in- and out of sample predictive performance of RPL, LC and Cov-Het models. This analysis was carried out by splitting the sample randomly into two halves and comparing the observed and the predicted choice probabilities for each alternative. The observed choice probabilities were calculated in each half of the sample by counting the number of times each alternative was chosen in each choice set, resulting in choice frequency counts, which were then normalised¹⁷ across the number of respondents. The experimental design contains 18 choice sets, and each choice set had three alternatives; hence, there are 54 (18*3) observed choice proportions that can be calculated from the experiment. These proportions are “model-free” in so far as they are whatever they are and do not depend on the model assumptions, and they represent the empirical estimates of the choice probabilities.

Predicted choice probabilities were calculated in the following way: a) re-estimate each model in the first half of the sample; b) use the resulting models to predict the expected choice probability for each alternative in the 18 choice set in the second half of the sample; c) repeat the process of steps a) and b) for the second half to predict the choice probabilities in the first half. We thus obtained 54 predicted probabilities in each half that were compared with the observed choice probabilities for each alternative and choice card. We then graphed the observed choice proportions against the predicted choice proportions for each of these split-half comparisons. This split-half predictive validity test again supports the superiority of the LC model for this dataset, both within- and cross-samples. We do not report these values or display the graphs in the interests of space, but they can be obtained from the authors on request.

Finally, it may be that there is high heterogeneity in both systematic and random utility components (HH above). In this case one solution is to estimate models allowing heterogeneity in both systematic and random utility components.. One way of modelling this kind of choice data is

to use a variant of the RPL model discussed above, namely an error component model (RPL-EC). This has been recently proposed in the environmental economics literature as an interesting new approach (Termansen et al., 2004; Scarpa et al, 2007). In the case of the choice experiment employed in this study, the error component approach can take the form:

$$U_{njt} = \beta X_{njt} + \eta_n X_{njt} + \mu_{jt} + \varepsilon_{njt} \quad j = \text{alternatives A or B} \quad (15)$$

$$U_{sq} = \beta X_{nsqt} + \eta_n X_{nsqt} + \varepsilon_{njt}$$

where μ_{jt} are additional error components distributed normally with zero mean and variance σ^2 , which allow for correlation patterns between the stochastic portions of the alternatives A and B. Thus, the EC-RPL model allows for heterogeneity in both systematic and random component of the utility¹⁸. Parameters estimates of this model are shown in table 7 along with the resulting welfare measures. All model coefficients are very significant and with the expected sign. Of particular interest is the high significance of the error component coefficient which shows the existence of correlation between the stochastic portions of utility A and B. This was expected, given the value of the inclusive value parameter in the CovHet model¹⁹. Also, most of the standard deviations of the random parameter distribution are still significant, revealing the presence of heterogeneity in the systematic component of utility. Although this model does not allow us to clearly disentangle the heterogeneity in the systematic or random component of utility, the analyst can easily determine which component is important by observing the significance of the standards deviation terms of the random parameter distribution and the standard deviation of the latent random effect. When both are highly significant, there exists heterogeneity in both systematic and random component of utility. However, if we compare this model to the LC model we still find that the LC model is superior, indicating that describing the heterogeneity in three discrete classes is the best approach in this dataset.

The comparison of implicit prices of the EC-RPL and the three models described so far reveals that the none of the IPs are different between the EC-RPL model and the CovHet model; whilst “mixed and broadleaved woodlands” and “field boundaries” are not statistically different

between the EC-RPL and the LC model at the 95% confidence level. “Heather moorland and bog”, “rough grassland”, and “mixed and broadleaved woodlands” model do not differ, at the same level of confidence, between the EC-RPL and the RPL model. This is a clear warning of the effect of modelling respondent heterogeneity if marginal WTP are the measures of interest..The comparison of the compensating surplus measures shows that they do not differ between the EC-RPL model and the CovHet model at the 90% confidence level, but do differ for scenarios 1 between the LC and EC-RPL model and for scenarios1 and 2 between the RPL and EC-RPL model. The tight confidence intervals of the EC-RPL model are the responsible for these differences.

7. Conclusions.

This paper compared three modelling approaches to incorporate and explain sources of preference heterogeneity in random utility choice models, using an example based on the value of public goods generated by upland farming. Understanding what underlies differences in values that people place on changes in public goods has been recognised as important for some time, especially in policy analysis (Randall, 1997). We wish to understand who benefits, and why some benefit more than others for a change in environmental quality. Land managers also can benefit from understanding preference heterogeneity: as for example, improvements in recreational site quality can be targeted at groups who value it most. As explained above, the three modelling approaches we considered differ in their basic approach to modelling preference heterogeneity, albeit within a common framework of the Random Utility model. RP and LC models consider respondent taste heterogeneity in systematic utility components and rely on the assumption of Gumbel distributed error terms whilst Cov-Het models analyse the heterogeneity in random utility components assuming that the error terms follow a Generalized Extreme Value distribution. Finally, the RPL model can be specified so that it takes into account the heterogeneity in systematic and random components of utility..Although several comparisons of RPL and LC models exist in the literature, we are unaware of peer reviewed publications that compare these with Cov-Het and EC-RPL

models This comparison is of interest as the way analysts deal with heterogeneity may impact on welfare measures used in cost-benefit analysis.

In this data set we considered, the LC model outperformed all other models on a number of criteria. However, the Cov-Het model results showed that the variance of the error component is systematically related to respondents' socio-economic characteristics and beliefs. This is not obviously consistent with preference modelling through LC models. So, an important issue is to determine under which conditions analysts should choose one approach over others. This will not be easy as each model has advantages and disadvantages, suggesting a need to develop criteria for choosing when one approach is better. RPL and LC models capture heterogeneity in the observable utility component in a different way: RPL models individual level heterogeneity, but requires assumptions of a distribution of taste parameters across the population; LC models are less flexible in structure, as the attribute and covariate parameters in each class are fixed, but they allow "clearer" descriptions of segment heterogeneity in the data. Yatchew and Griliches (1984) showed that in a logit model, loss of homoscedasticity leads to biased parameters estimates, with the bias increasing as the heteroskedasticity is itself a function of the independent variables in the utility function. Cov-Het models allow the analyst to explicitly test for this heterogeneity in the variance of the stochastic utility component. This, in turn, can depend on choice alternative attribute levels and/or individual socio-economic characteristics, and indeed in our data both influence error component variance.

If RPL and LC models assume constant scale across respondents, and a Cov Het model shows this assumption is not satisfied, should analysts rely on a Cov-Het model, or is there still some flexibility in choice of model specification? A possible solution would be to determine the extent of heterogeneity in each component for a specific data set, along the lines of the matrix described in the previous section. If heterogeneity in the random component is low relative to heterogeneity in the systematic component, one can use models that focus on heterogeneity in the latter component. A second solution is to estimate models that allow heterogeneity in both

components of utility. The specification of an error component in the RPL modelling framework is a way to do this. This model can in turn be specified to accommodate the correlation structure across alternatives or go further and model the heteroskedasticity of the error components as a function of respondent characteristics.

Importantly, the way that analysts treats preference heterogeneity in the random utility theoretical framework has an impact in the estimates of welfare measures, being the impact more important in the marginal WTP estimates than in the compensating surplus estimates, due to a sort of compensation between the overstated WTP of one attribute and the lower value for another attribute in the same policy alternative. More research is clearly needed into these more advanced choice models to develop guidelines for how to select the most suitable approach to deal with preference heterogeneity, in order to provide more reliable welfare measures for use in cost-benefit analysis and policy appraisal.

Figure 1: Example Choice Card







	Policy Option	Current policy	Policy Option A	Policy Option B
	Change in area of Heather Moorland and Bog	A loss of 1,560 hectares (-2%)	A gain of 1,560 hectares (+2%)	A loss of 1,560 ha (-2%)
	Change in area of Rough Grassland	A loss of 17,700 ha (-10%)	A loss of 3,500 ha (-2%)	A loss of 3,500 ha (-2%)
	Change in area of Mixed and Broadleaf Woodlands	A gain of 1,000 ha (+3%)	A gain of 5,500 ha (+20%)	A gain of 2,700 ha (+10%)
	Condition of field boundaries	For every 1 km, 100m is restored	For every 1 km, 200 m is restored	for every 1 km, 50 m is restored
	Change in farm building and traditional farm practices	Rapid decline	no change	Much better conservation
	Increase in tax payments by your household each year	£0	£20	£10

Table 1. Landscape Attributes and Levels used to describe choice alternatives in the study.

Attribute	Definition	Attribute levels
Heather moorland/bog	Heather dominated moorland. Bog habitat in wetter areas.	-12% -2% +5%
Rough grassland	Areas typically used for extensive sheep grazing.	-10% +5% +10%
Broadleaf and mixed wood-land	May consist of a mix of native species, or be dominated by one.	+3% +10% +20%
Field boundaries (for every 1km, X metres is restored)	Hedges, stone walls, ditches, banks and lines of trees, but not modern fences.	$X = 50$ $X = 100$ $X = 200$
Cultural heritage	Traditional farm buildings and farming practices such as shepherding with sheep dogs.	Rapid decline No change Much better conservation
Cost	Amount paid per household per year through higher tax payments.	£2, £5, £10, £17, £40, £70 (baseline £0 was not given as an option in other scenarios)

Note. The status quo levels of these attributes are shown in BOLD text.

Table 2. Model coefficients and standard errors.

Model		RPL		Latent Class		Covariance Het.		
Variables	Coef.	S.e.		Coef.	S.e.	Variables	Coef.	S.e.
<i>Attributes in the utility function</i>			<i>Attributes in Utility functions: Latent class1</i>			<i>Attributes in utility function</i>		
HMB	.054***	0.012	K	-3.18	2.264			
RG	.047***	0.012	HMB	.517*	0.269	HMB	.041***	0.009
FB	.000	0.001	RG	-.181*	0.103	RG	.057***	0.008
CH1	.397**	0.192	BMW	0.208	0.136	BMW	.031***	0.008
CH2	.668***	0.230	FB	0.001	0.010	FB	.002**	0.001
K	-0.475	0.472	CH1	-1.47	2.459	CH1	.590***	0.134
BMW	.050***	0.010	CH2	-0.289	2.268	CH2	.868***	0.162
TAX	-.071***	0.006	TAX	-0.220	0.141	TAX	-.056***	0.006
AGE	-.834***	0.199						
SEX	-0.215	0.245	<i>Latent class2</i>			<i>Attributes of Branch Choice Equation</i>		
LIVING	-.024***	0.008	K	.622*	0.326	K	-0.007	0.311
ASSO	0.306	0.338	HMB	.031**	0.016	AGE	-.950***	0.169
EDU	.614***	0.092	RG	.096***	0.016	SEX	-.394**	0.199
			BMW	0.0007	0.016	LIVE	-.017**	0.007
<i>Standard deviations</i>			FB	-.005***	0.002	ASSO	.902***	0.270
Sd HMB	.071***	0.019	CH1	0.433**	0.175	EDU	.446***	0.070
Sd RG	.098***	0.011	CH2	.987***	0.249			
Sd FB	.008***	0.002	TAX	-.333***	0.024			
Sd CH1	.980***	0.227						
Sd CH2	1.39***	0.243	<i>Latent class3</i>			<i>Inclusive value parameters</i>		
			K	1.237***	0.183	NO Change	1.000	0.000
			HMB	.040***	0.007	CHANGE	.280***	0.082
			RG	.032***	0.007			
			BMW	.036***	0.008	<i>Covariates in the IV parameter</i>		
			FB	.004***	0.001	AGE	.447***	0.098
			CH1	.405***	0.120	SEX	.291**	0.131
			CH2	.642***	0.131	LIVE	0.002	0.004
			TAX	-.036***	0.004	ASSO	-.615***	0.219
						EDU	0.056	0.049
Segment Function 1								
			K	0.242	0.893			
			AGE	.944**	0.385			
			SEX	-0.061	0.473			
			LIVE	0.020	0.016			
			ASSO	-.969***	0.227			
			EDU	-0.496	0.663			
Segment Function 2								
			K	-0.171	0.755			
			AGE	.689**	0.322			
			SEX	0.357	0.411			
			LIVE	0.018	0.014			
			ASSO	-.441***	0.157			
			EDU	-0.724	0.598			
L(0)	-1305.1			-1305.1			-1305.1	
LL	-871.4			-730.1			-926.0	
LR	867.6***			1150.1***			539.4***	
Rho ² (%)	.33			.44			.29	

Asterisks denote significance level (***) = 1%; ** = 5 %; * = 10 %).

Table 3. Explanation of variable abbreviations and coding.

Const	constant term (= 0 for the current policy, = 1 for alternatives A or B)
HMB	percentage change in area of heather moorland and bog
RG	percentage change in area of rough grassland
BMW	percentage change in area of broadleaf and mixed woodland
FB	change in the length of field boundaries (in metres restored)
CH1	change in cultural heritage from “rapid decline” to “no change” (1 = yes, 0 = no)
CH2	change in cultural heritage from “rapid decline” to “much better conservation” (1 = yes, 0 = no)
TAX	Additional tax payment per year
AGE	respondent’s age (1=18-34; 2=35-54;3=55-70)
SEX	respondent’s gender (1 = male, 0 = female)
LIVE	number of years respondents have been living in the region
ASSO	whether respondent belongs to an environmental organization (1 = yes, 0 = no)
EDU	respondent’s education level 1= pre-A level secondary, 6= higher degree)

Table 4. Attribute marginal willingness to pay (implicit prices) and 95% confidence intervals (all figures are in pounds sterling per household per year).

Model	Random Parameters	Latent Class				Cov Het.
Attribute:		Class1	Class2	Class3	Sum^a	
HMB	0.76 (0.54 1.01)	-	0.09 (0.00 0.19)	1.09 (0.30 1.52)	0.37 (0.21 0.53)	0.73 (0.47 1.02)
RG	0.66 (0.38 0.96)	-	0.29 (0.20 0.37)	0.87 (0.54 1.22)	0.37 (0.24 0.51)	1.02 (0.80 1.29)
BMW	0.70 (0.42 1.01)	-	.000 (-0.09 0.10)	0.97 (0.55 1.45)	0.30 (0.13 0.48)	0.57 (0.27 0.85)
FB	0.00 (-0.02 0.02)	-	-0.01 (-0.02 0.00)	0.11 (0.07 0.15)	0.03 (0.01 0.05)	0.04 (0.00 0.06)
CH1	5.58 (2.82 8.46)	-	2.86 (1.14 4.33)	11.03 (4.76 18.20)	4.40 (1.86 7.11)	10.60 (6.20 16.11)
CH2	9.39 (5.54 13.88)	-	2.96 (1.53 4.54)	17.50 (11.26 23.37)	6.40 (3.98 8.76)	15.60 (11.07 20.87)

^a: this is the weighted WTP for the attributes estimated by considering the class probabilities

HMB: heather moorland and bog

RG: rough grassland

BMW: broadleaved and mixed woodland

FB: field boundaries

CH: cultural heritage

Table 5: Attribute levels for the baseline and the three policy scenarios used in the calculation of compensating surplus

Upland attribute	Baseline	Scenario 1	Scenario 2	Scenario 3
Heather moorland and bog	+1%	+3%	+5%	-2%
Rough grassland	+1%	-1%	-3%	+3%
Mixed and broadleaved woodland	+3%	+4%	+6%	+5%
Field boundaries	+5%	+6%	+10%	+2%
Cultural heritage	Rapid Decline	No change	No change	Rapid Decline

Note: Modified from table 1 of Cumulus et al. (2005).

Table 6. Compensating surplus for three future policy scenarios (all figures are in pounds sterling per household per year).

	Random Parameter	Latent Class			Cov Het.
		Class2	Class3	Sum^a	
Scenario 1: Agri- Environment	6.47 (3.31 9.62)	2.33 (0.68 3.89)	13.55 (7.11 21.12)	4.97 (2.41 7.83)	7.05 (3.78 10.62)
Scenario 2: environment only	8.09 (4.48 11.84)	1.39 (-0.20 2.95)	20.36 (13.20 28.63)	6.68 (3.93 9.76)	8.49 (4.91 12.52)
Scenario 3: no support	0.44 (-0.69 1.60)	0.71 (0.35 1.10)	-2.91 (-5.12 -0.83)	-0.62 (-1.42 0.15)	-0.06 (-0.83 0.74)

^a: this is the weighted WTP for the scenarios estimated by considering the class probabilities.

Table 7: Error component random parameter model and welfare measures

Model		EC-RPL	
Variables	Coef.	S.e.	
<i>Attributes in the utility function</i>		<i>Implicit prices</i>	
		<i>CS</i>	
HMB	.063***	0.013	HMB 0.71 (0.53 0.90) Scenario1 8.87 (7.77 10.19)
RG	.074***	0.010	RG 0.84 (0.72 0.98) Scenario2 10.75 (9.03 12.85)
FB	.003*	0.001	BMW 0.48 (0.22 0.77) Scenario3 -0.35 (-1.21 0.47)
CH1	.737***	0.181	FB 0.03 (0.02 0.04)
CH2	1.209***	0.205	CH1 8.29 (7.41 9.43)
K	-1.480*	0.857	CH2 13.6 (11.55 16.12)
BMW	.042***	0.012	
TAX	.088***	0.006	
AGE	.979**	0.399	
SEX	0.048	0.470	
LIVING	0.025	0.017	
ASSO	0.740	0.770	
EDU	1.034***	0.207	
<i>Standard deviations of parameter distribution</i>			
Sd HMB	.073***	0.019	
Sd RG	.035**	0.017	
Sd FB	.006**	0.002	
Sd CH1	0.029	0.912	
Sd CH2	.769***	0.281	
<i>Standard deviation of the latent random effect</i>			
σ	2.636***	0.277	

Asterisks denote significance level (***) = 1%; ** = 5%; * = 10 %).

ENDNOTES

¹ The Multi-Nomial Logit model assumes that it is variations in individual-specific characteristics that affect choice probabilities; the Conditional Logit model assumes that it is variation in the choice alternative-specific characteristics that affect choice probabilities. These terms are often confused, however, in the literature.

² The conditional logit model treats repeated respondents' choices as independent observations. This assumption is difficult to defend since individual choices may be affected by their previous choices within a series of choice tasks, such as respondents are required to complete within a Choice Experiment.

³ This specification assumes that the person's taste, as represented by β_n , are the same for all choice situations.

⁴ For a discussion and comparison of the implications of alternative assumptions, see Hanley et al (2006).

⁵ The S^{th} parameter vector is normalised to zero to enable identification of the model.

⁶ As Boxall and Adamowicz (2002) pointed out, the membership function determining the structure of the latent class is not a behavioural relation, but a statistical classification process (which maximises the likelihood function). Thus, one can ignore the correlation between the error term in the utility function and the error in the classification function.

⁷ These scale factors could be identified assuming that the segment specific utility parameters are the same. However, as point out by Boxall and Adamowicz (2002) the assumption of parameter equality across segments is contrary to the spirit of latent class model since a researcher would not want to impose utility parameter equality.

⁸ In equation 5 and 6 we did not specify the scale parameters since they are assumed to be equal to 1 to allow model estimation.

⁹ Avoiding this assumption would require to allow all the parameters to vary randomly. As Ruud (1996) points out, a model with all coefficients, including the price, specified as random can be practically unidentified empirically. In this specific data set, we experimented with allowing all the coefficients to vary randomly with the result that the model did not converge.

¹⁰ Typically the selection is done by following the information criteria statistic developed by Hurvich and Tsai (1989). This criteria is specified as $-2\ln L + J\delta$ where $\ln L$ is the log-likelihood of the model at convergence, J is the number of estimated parameters in the model and δ is a penalty constant. A very common criteria used is the Akaike Information Criteria (AIC) where $\delta=3$. A recent simulation study by Andrews and Currim (2003) concluded that the AIC was the best criterion for LC model.

¹¹ When new policy designs are investigated it is of interest to know which respondent characteristics increase the probability of agreeing with the "policy-on" options and which the probability of the "policy-off" option.

¹² All the welfare comparisons are undertaken using the procedure suggested by Poe et al. (2005). This test is the most commonly used in the choice experiment literature and consists of calculating the differences between two random distributions generated from the asymptotic distribution of the parameters by the Krinsky and Robb procedure (Krinsky and Robb, 1986). A one-sided approximate significance level is calculated by the proportion of negative values in the distribution of differences, depending on which mean is thought to be greater.

¹³ Scenarios 0, 1 and 2 assume that there is continued and adequate funding to support varying degrees of upland farming in the SDA, and these policy options therefore describe differing emphasis (in direction and degree) of available funding. Option 3 (Abandonment-intensification) is a 'no subsidy' scenario, assuming that support for upland farming has been withdrawn.

¹⁴ The 10th percentiles of parameter values are: HMB: -0.04; RG: -0.08; FB: -0.011; CH1: -0.8654; CH2: -1.11. The 90th percentiles of parameter values are: HMB: 0.15; RG: 0.17; FB: 0.011; CH1: 1.65; CH2: 2.45.

¹⁵ We are considering two respondents with the same observable socio-economic characteristics, so that the utility difference is only due to attribute taste heterogeneity.

¹⁶ Individual 1 is a male between 55 and 70 years old who does not belong to any environmental organization, has a low education level and has been living for 40 year in the region. Individual 2 is a young female (between 18 and 35 years old) who is a member of an environmental organization, holds a higher degree education level and resides in the region from 10 years.

¹⁷ When accounting for the observe choice proportions, it is necessary to take into account that the choice cards have been answered a different number of times, depending on how many respondents faced each choice cards. After counting the number of times each alternative in each choice card had been chosen, we divided it (normalized) by the number of respondents that faced each alternative in each choice card.

¹⁸ In the model specified in this study we employ a simple specification of the random component in which the error component is uniquely assigned to the alternatives not in the status quo. As pointed out by Scarpa et al. (2007), additional information can be obtained by disaggregating the error component into socio-economic determinants of error. We estimated an error component random parameter logit model where we explained the heterogeneity in the variance of the error component. This model did not converge

¹⁹ The inclusive value parameter can take values from 0 to 1 which can be interpreted as a measure of correlation in unobserved factors within each nest. Values close to one reveal the absence of correlation; values close to 0 reveal a very high degree of correlation.

References

- Amador, F. J., Gonzales R., and Ortuzar J. (2005). Preference heterogeneity and willingness to pay for travel time savings. *Transportation* 32: 627-647.
- Andrews, R. and I. Currim. (2003). Retention of Latent Segments in Regression-Based Marketing Models. *International Journal of Research in Marketing* 20:315-321.
- Ben-Akiva, M. E. and Swait, J. D, (1986). The akaike likelihood ratio index. *Transportation Science* 20: 133-136.
- Bhat, C.R, (1997). Covariance heterogeneity in nested logit models: econometric structure and application to intercity travel. *Transportation Research-B* 31: 11-21.
- Biol, E., Karousakis, K. and Koundouri, P, (2006). Using choice experiment to account for preference heterogeneity in wetland attributes: the case of Cheimaditita wetland in Greece, *Ecological Economics* 60: 145.156.
- Boxall, P.C., Adamowicz, W, (2002). Understanding Heterogeneous Preferences in Random Utility Models: A Latent Class Approach. *Environmental and Resource Economics* 23: 421-446.
- Boyd, J. and Mellman J., (1980). The effect of fuel economy standards on the U.S. automotive market: A hedonic demand analysis, *Transportation Research A* 14: 367-378.
- Cardell, S. and Dunbar F., (1980). Measuring the societal impact of automobile downsizing. *Transportation Research A* 14: 423-434.
- Carlsson, F. Frykblom, P. and Liljenstolpe C., (2003). Valuing wetland attributes: an application of choice experiments, *Ecological Economics* 47: 95-103.
- Chen, H.Z. and Cosslett, S., (1998). Environmental quality preference and benefit estimation in multinomial probit models: A simulation approach. *American Journal of Agricultural Economics*. 80(3): 512-520.

- Colombo, S., Hanley N. and Calatrava-Requena J., (2005). Designing policy for reducing the off-farm effects of soil erosion using Choice Experiments. *Journal of Agricultural Economics* 56(1): 80-96.
- Colombo S and Hanley N (2008). What determines prediction errors in benefits transfer models? *Land Economics*, forthcoming.
- Cumulus Consultants Ltd., Institute for European Environmental Policy and the Countryside and Countryside Research Unit, (2005). Assessment of the impact of CAP Reform and other key policies on upland farms and land use implications in both Severely Disadvantaged & Disadvantaged Areas of England, report to the Department for Environment, Food and Rural Affairs.
- DEFRA, (2006). Rural Development Programme for England 2007-2013: uplands rewards structure consultation document. London: Department for the Environment, Food and Rural Affairs.
- De Shazo, R. and Fermo, G., (2002). Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *Journal of Environmental Economics and Management* 44: 123–143
- Dellaert, B., Brazell J. and Louviere, J., (1999). The effect of Attribute Variation on Consumer Choice Consistency. *Marketing Letters* 10(2): 139-147.
- EFTEC (2006). Economic evaluation of environmental impacts in the severely disadvantaged areas of England. Report to DEFRA, London.
- Ferrini S. and Scarpa R. (2007) “Designs with a priori information for non-market valuation with choice experiments: a Monte Carlo study” *Journal of Environmental Economics and Management*, 53, 342-363.
- Green, W. and Hensher, D., (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B* 37:681-698.

- Hanemann, W, (1984). Welfare Evaluations in Contingent Valuation Experiment with Discrete Responses. *American Journal of Agricultural Economics* 66: 332-341.
- Hanley, N., Wright R., and Koop G., (2002). .Modelling Recreational Demand Using Choice Experiment: Climbing in Scotland. *Environmental and Resource Economics* 22: 449-466.
- Hanley N, Ryan M and Wright R., (2003). Estimating the monetary value of health care: lessons from environmental economics *Health Economics*, 12 (1): 176.189.
- 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.
- Hanley N, Colombo S, Johns H and Mason P (2007) The reform of support mechanisms for upland farming: paying for public goods in the Severely Disadvantaged Areas of England *Journal of Agricultural Economics* 58(3): 433-453.
- Hensher, D (2001). The valuation of commuter travel time savings for car drivers: evaluating alternative model specifications. *Transportation* 28: 101-118.
- Hurvich, M. and C. Tsai. (1989). Regression and time series model selection in small samples. *Biometrika* 76: 297-307.
- Kamakura, W., & Russell, G. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, 26: 379–390.
- Kling, C. and Thomson C. (1996). Implications of Model Specification for Welfare Estimation in Nested Logit Models. *American Journal of Agricultural Economics*, 78: 103-14.
- Krinsky, I. and Robb, A.L.(1986) “On Approximating the Statistical Properties of Elasticities”. *Review of Economics and Statistics*, 68: 715-719.
- Layton, D.F, (2000). Random coefficient models for stated preference surveys, *Journal of Environmental Economics and Management* 40:21-36.

- Louviere, J.J., Hensher, D.A. and Swait, J, (2000). *Stated Choice Methods, Analysis and Application*. Cambridge University Press.
- Louviere, J.J., Street, D., Carson, R., Ainslie, A., Deshazo, J., Cameron, T., Hensher, D., Kohn, R. and Marley, T., (2002). Dissecting the random component of utility. *Marketing Letters* 13(3): 177-193.
- Louviere, J. and Eagle T, (2006). Confound it! That Pesky Little Scale Constant Messes Up . CenSoC Working Paper No. 06-002.
- Magidson J. and Vermunt K. J. (2007). Removing the Scale Factor Confound in Multinomial Logit Choice Models to Obtain Better Estimates of Preference. Paper presented at 2007 Sawtooth Symposium, Conference, Santa Rosa, California, October 15-19.
- Mansky, C. (1977). The Structure of Random Utility Models. *Theory and Decision* 8:229-254.
- McFadden, D, (1973): Conditional logit analysis of qualitative choice behaviour. En, P. Zarembka (Ed.), *Frontiers in econometrics*, New York: Academic Press, p. 105-142.
- McFadden, D. (2000). Economic choice. In T. Persson (Ed.), *Nobel lectures, Economics 1996-2000*, 330-364. Singapore. World Scientific Publishing.
- McFadden, D. and Train K, (2000). Mixed MNL models of discrete response. *Journal of Applied Econometrics*, 15: 447-470.
- Macmillan, D.C. Philip, L, Hanley, N.D. and Alvarez-Farizo, B (2002). Valuing the non-market benefits of wild goose conservation: A comparison of interview and group based approaches. *Ecological Economics* 43(1); 49-59.
- Milon, J. and Scrogin, D., (2006). Latent preferences and valuation of Wetland ecosystem restoration. *Ecological Economics* 56: 162-175.
- Morey, E. Thacher J. and Breffle W., (2006). Using Angler Characteristics and Attitudinal Data to Identify Environmental Preference Classes: A Latent-Class Model. *Environment and Resource Economics* 34: 91-115.

- Persson, T.H, (2002). Welfare calculations in models of the demand for sanitation. *Applied Economics* 34(12): 1509-1518.
- Poe, G. L., Giraud, K. L. and Loomis, J. B. (2005). Computational methods for measuring the difference of empirical distributions. *American Journal of Agricultural Economics* 89(2):353-65.
- Provencher, B. and Bishop, R., (2004). Does accounting for preference heterogeneity improve the forecasting of a random utility model? A case study. *Journal of Environmental Economics and Management* 48: 793-810.
- Randall, A. (1997). The NOAA Panel Report: A New Beginning or the End of an Era? *American Journal of Agricultural Economics* 79: 1489-1494
- Revelt, D. and Train, K, (1998). Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level, *Review of Economics and Statistics*, 80:1-11.
- Rigby D. and Burton M, (2006). Modelling disinterest and dislike: a bounded Bayesian mixed logit model of the UK market for GM food. *Environmental and Resource Economics*, 33(4): 485-510.
- Ruto E., Garrod G. and Scarpa R. (2008) "Valuing animal genetic resources: a choice modelling application to indigenous cattle in Kenya" *Agricultural Economics*, 38 (1), 89-98.
- Ruud, P.A, (1996). Simulation of the multinomial probit model: An analysis of covariance matrix estimation. Working paper, Department of Economics, University of California.
- Scarpa, R. and M. Thiene. (2005). Destination choice models for rock-climbing in the North-Eastern Alps: a latent-class approach based on intensity of participation. *Land Economics* 81: 426-444.
- Scarpa, R., K. G. Willis, and Acutt M. (2005). Individual-specific welfare measures for public goods: a latent class approach to residential customers of Yorkshire Water. Chapter 14 in: *Econometrics Informing Natural Resource Management*, edited by Phoebe Koundouri, Edward Elgar Publisher. Pages 316-337.

- Scarpa, R., K. G. Willis, and Acutt M. (2007). Valuing externalities from water supply: Status quo, choice complexity and individual random effects in panel Kernel logit analysis of choice experiment. *Journal of environmental economics and management* 50(1):449-466.
- Termansen, M., McLean, C. and Scarpa R. (2004). Economic valuation of Danish forest recreation combining mixed logit models and GIS. Paper presented at the 2004 EAERE conference Budapest, 25-28 June.
- Shen, J. Sakata, Y. and Hashimoto Y, (2006). A comparison between Latent Class Model and Mixed Logit Model for Transport Mode Choice: evidences from Two Datasets of Japan. Discussion paper 06-05 Graduate School of Economics, Osaka University, Japan.
- Street, D., L., Burgess, and J. Louviere. (2005). Quick and easy choice sets: Constructing optimal and nearly optimal stated choice experiments. *International Journal of Research in Marketing* 22 (4): 459-470.
- Street, D. and L., Burgess (2007). The construction of optimal stated choice experiments: Theory and Methods. Wiley.
- Strumse, E. (1994). Perceptual Dimensions in the Visual Preferences for Agrarian Landscapes in Western Norway, *Journal of Environmental Psychology*, 14: 281-292.
- Swait, J. and Adamowicz, W., (2001). Choice Environment, Market Complexity and Consumer Behaviour: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice. *Organizational Behaviour and Human Decision Processes* 86(2): 141-167.
- Van den Berg, A. E., et al. (1998). Group differences in the aesthetic evaluation of nature development plans: A multilevel approach, *Journal of Environmental Psychology*, 18: 141-157.
- Train, K., (1998). Recreation Demand Models with Taste Differences over People, *Land Economics*, 74: 230-239.

Train, K., (2003). *Discrete Choice Methods with Simulations*, Cambridge University Press.

Train, K. and Weeks M., (2005). Discrete Choice Models in Preference Space and Willingness-to-Pay Space, in *Applications of Simulation Methods in Environmental Resource Economics*, in A. Alberini and R. Scarpa, eds., Springer Publisher: Dordrecht, The Netherlands.

Willis, K. G., et al. (1995). Benefits of Environmentally Sensitive Area Policy in England: A Contingent Valuation Assessment, *Journal of Environmental Management*, 44: 105-125.

Yatchew A, Griliches Z. (1984). Specification Error in Probit Models. *Review of Economics and Statistics* 65: 134-139.