Incorrectly accounting for preference heterogeneity in choice experiments: what are the implications for welfare measurement?

Catalina M. Torres, Sergio Colombo, Nick Hanley

October 2014

Abstract

Gains from the incorporation of monetary values for changes in environmental goods and services within cost-benefit analysis depend on how well researchers can estimate these values. One key problem in both stated and revealed preference approaches is how best to model preference heterogeneity. Researchers have implemented several approaches to represent this heterogeneity, and have shown that the choice of approach can have an effect on welfare estimates. However, the question as to the degree of error in welfare measurement from an inappropriate choice of approach has not been addressed. We use Monte Carlo analysis to investigate this issue in the context of choice modelling of coastal water quality changes, when the researcher chooses between a random parameters and latent class model for representing heterogeneity. This allows us to quantify the errors that emerge from using the wrong model in estimating the benefits of water quality improvements. Our overall conclusion is smaller welfare errors are likely to come from use of a latent class model.

Keywords: choice experiments; cost-benefit analysis; Monte Carlo analysis; non-market goods; preference heterogeneity; welfare measurement.

***JEL classification**:* C15, C52, Q51.

INTRODUCTION

Choice modelling (CM) has emerged as a flexible and informative method for estimating non-market values in a range of fields of application, including agricultural, environmental, transport and health economics (Hensher et al 2005). The method can be applied to both stated preference (SP) and revealed preference (RP) data. Its advantages are now well-known: the ability to estimate values for the characteristics or attributes of a range of goods, services and policy designs; to produce estimates of compensating or equivalent surplus for a range of outcomes specified in terms of changes in multiple attribute levels; and to measure both use and non-use values if an SP approach is employed. Dating from Train (1998), choice modellers have become increasingly interested in how to represent heterogeneity in preferences, a research direction foreseen by Ben-Akiva and Lerman (1985, p. 367) in one of the earliest works on discrete choice analysis. A range of empirical approaches to representing preference heterogeneity have emerged in CM, and we review these in the following section. Researchers have been able to explore the differences that selection of a particular approach makes to welfare measures in a particular dataset, and indeed have been able to implement a number of tests for which approach best fits a particular set of data (Colombo et al 2009; Hynes et al 2008).

However, the question as to the degree of error in welfare estimation which results from an inappropriate choice of empirical approach to represent preference heterogeneity in a particular empirical setting – in terms of the difference between *estimates* of the money metric measure of welfare change and the true, underlying money metric welfare change – has not been addressed. This is because, of course, in most situations we are unable to observe this underlying, true measure of welfare change for non-market goods (Johansson 1993). Any systematic over- or under-estimation of benefits due to an inappropriate selection of modelling approach is of particular relevance for many environmental applications of CM, where the main goal is to inform benefit-cost analysis (Hanley and Barbier 2009).

In this article, we use Monte Carlo (MC) analysis to address this question. MC analysis allows the researcher to start with a particular "true" utility function and a particular distribution of preferences across a population of consumers, and then to simulate choices to a set of choice alternatives based on these preferences. A variety of models with alternative treatments of preference heterogeneity can be estimated based on these simulated choices, and welfare estimates calculated. Since the true utility functions underlying these choices are known to the researcher, including the true, underlying pattern of preference heterogeneity which generates the data, we can then quantify both the relative and absolute magnitudes of errors in welfare estimates in relation to the true, underlying money metric measure of compensating or equivalent surplus. Thus, welfare error quantification enables us to comment on the likely consequences for benefit-cost analysis of incorrect assumptions about how to model preference heterogeneity in survey-based applications of the method.

The structure of the rest of this article is as follows. The next section provides a review of how heterogeneity has been modelled in stated choice data. It shows little attention has been paid to the implications for the accuracy and efficiency of welfare measures of mistaken assumptions about preference heterogeneity. Section 3 discusses the methodology used and the data employed for the MC experiments. Results are reported in section 4, where the sensitivity of welfare measures to mistaken assumptions about the nature of preference heterogeneity is analysed. Discussion and conclusions follow in section 5.

MODELLIG HETEROGENEITY IN STATED CHOICE DATA

The random parameter logit (RPL) and latent class (LC) model have emerged as popular approaches to preference heterogeneity (Hensher and Greene, 2003; Scarpa and Thiene, 2005). A range of papers in transport, leisure and environmental economics compare the performance of RPL and LC approaches to determine which fits the data better and to examine differences in welfare estimates (Birol et al 2006; Boxall and Adamowicz 2002; Broch and Vedel 2012; Greene and Hensher 2003; Hynes et al 2008; Outma et al 2007; Provencher and Bishop 2004; Strazzera et al., 2013). In the RPL or "mixed logit" model, the utility function for respondent *n* choosing over *J* alternatives is augmented with a vector of parameters that incorporate individual preference deviations with respect to the mean. In LC models, heterogeneity is captured 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. The two approaches can also be combined, as shown by Bujosa et al (2010). Empirical results show that there is no clear pattern of which approach (RPL, LC) is superior to the other (Greene and Hensher, 2003; Scarpa and Thiene 2005), with the "best" choice of approach seeming to depend on the nature of the underlying data generating process, as would be expected.

The fact that RPL and LC models assume that the variance of the error term of utility is constant, and consequently that the scale parameter is also constant, has led to the emergence of alternative approaches which focus on modelling scale heterogeneity. The main reason is pointed out by Louviere (2006; 1999; 2002), Louviere and Eagle (2006), Meyer (2007) and Louviere et al (2008): all statistical models in which the dependent variable is latent are likely to confound estimates of the model's parameters with error variability, and as such the parameter estimates do not represent mean tastes but the means multiplied by a scale factor. One of the approaches for dealing with scale heterogeneity is the covariance heterogeneity (Cov-Het) model, which includes heterogeneity in the *stochastic* part of utility by allowing the scale parameter to be a function of choice attributes and respondents' socioeconomics characteristics. Colombo et al (2009) compare the performance of the RPL, LC and the Cov-Het models and conclude the LC approach best fits the data although the three models perform equally well in terms of out-of-sample predictions.

More recently, there has been an emerging literature which aims to combine the modelling of taste and scale heterogeneity. Fiebig et al (2010) compare a RPL model with a scale heterogeneity multinomial logit model (S-MNL) where only scale heterogeneity is allowed, and a generalized multinomial logit

(G-MNL) model which considers both taste and scale heterogeneity. They conclude that the S-MNL and G-MNL models outperform the RPL model especially in datasets that involve more complex choices. Greene and Hensher (2010) find however that accommodating only scale heterogeneity (i.e. neglecting taste heterogeneity) may be of limited empirical interest, resulting in a statistically inferior model, whereas the inclusion of both scale and taste heterogeneity results in an improvement over the standard RPL model. Importantly, they observe that compared to failure to include for taste heterogeneity, failure to account for scale heterogeneity may not be of such great empirical consequence especially when WTP measures are of primary interest. The reason is that the effect of confounding between scale and taste cancels out in the estimation of the WTP, because this is calculated by dividing the estimated coefficients by the price parameter (i.e. making the estimation of WTP scale-free). Although this is not always generally applicable[1], most CM applications in environmental valuation aim primarily on providing information to decisions makers about non-market values of environmental goods, and in particular to produce estimates of compensating or equivalent surplus for a range of outcomes specified in terms of changes in attribute levels for public goods. When the analyst is primarily interested in WTP measures, the more parsimonious model approach which considers taste heterogeneity alone can thus be adequate.[2]

As can be seen then, research on heterogeneity in choice modelling is extensive. However, little attention has been paid to examine the effects on welfare estimates of mistaken assumptions about the nature of preference heterogeneity. As pointed out by Torres et al. (2011), the interest in analysing the bias and efficiency of welfare estimates within non-market valuation has been mainly centred on investigating, through MC analysis, issues such as the (i) specification of the recreation demand function in travel cost models, ii) WTP elicitation in the contingent valuation approach, and iii) experimental design under different utility specifications in choice experiments (Kling 1987; Kling 1988; Kling 1989; Adamowicz et al 1989). Investigations of the effects of decisions over appropriate nesting structures in multiple site recreation demand models represent a related area of concern (Herriges and Kling 1997; Kling and Thomson 1996), which also makes use of MC analysis. In the contingent valuation field, papers focusing on the analysis of welfare bias and variance through the MC approach mostly deal with the advantages of combining RP and SP data (Kling 1997), and the efficiency gains from using double-bounded discrete choice model relative to a bivariate probit model

[1] Flynn et al (2010) point out that such normalization is not always possible, as for instance in the medical field where often there is not a monetary attribute used in the design. In this case, it is paramount to take into account both taste and scale heterogeneity to obtain unbiased estimates of the parameter of interest. At the same time they warn that there may be different variance-scale factors by attribute and the traditional solution of dividing the attribute coefficients by the price coefficient may be wrong.

[2] Hess and Rose (2012) argue that gains in fit obtained in models accounting for scale heterogeneity are the results of using more flexible distributions, rather than an ability to capture scale heterogeneity. Indeed, they argue that recent work aimed at providing separate and uncorrelated stochastic treatments of ‘scale’ and ‘taste’ sensitiveness’, such as Fiebig et al (2010) and Greene and Hensher (2010), ignores the existence of scale/taste sensitivity confound and hence interpretation from their results is wrong.

(Alberini 1995) and different elicitation formats and bid designs (Scarpa and Bateman, 2000). In choice experiments (CE), the main concern of analysts has been directed towards examining through the use of MC methods the implications for welfare measurement of different experimental design strategies (Carlsson and Martinsson 2003; Ferrini and Scarpa 2007; Lusk and Norwood 2005; Scarpa and Rose 2008).

It is thus easy to see that most of the literature concerned about the analysis of welfare bias and efficiency through the use of MC analysis has paid little attention to the question of how important the way in which preference heterogeneity is modelled in CM is for welfare measurement. To our knowledge, only Torres et al (2011) attempt to examine the errors from mistaking the way of explaining heterogeneity in CEs. In particular, and with a focus on different attribute specifications, they analyze the effects on welfare estimates from i) correctly assuming RPL taste heterogeneity but mistaking parameter distributional assumptions, and ii) incorrectly assuming RPL taste heterogeneity when it is driven by the scale factor. However, Torres et al (2011) do not investigate the significance of analytical errors resulting from differences between latent class (finite mixture) and random parameter (continuous mixture) utility functions.

In this paper, we contribute to the literature by examining the errors from mistaken empirical approaches to account for the nature of underlying preference heterogeneity, when choice is only affected by variations in tastes across people and not by variations in the scale of the error. In other words, we focus on the implications of mistaken assumptions about the underlying utility function capturing taste heterogeneity in CEs. The feature of preference heterogeneity on which we focus is this distinction between finite- and continuous-mixing models. As discussed, there are indeed more options for the researcher focusing on preference heterogeneity to choose from now than just RPL or LC, but we exclude them from our analysis for two reasons. First, our focus is on the two most widely-used approaches to date in modeling preference heterogeneity in the environmental valuation literature, namely RPL and LC models (Beharry-Borg et al 2009; Bujosa et al 2010; Colombo et al 2009; Hess and Beharry-Borg 2012; Hynes et al 2008; Provencher et al 2002). Second, the debate around how best to model scale heterogeneity is still inconclusive. In this context, we think it is of considerable interest for cost-benefit analysts to know how much of an error can be made in estimating welfare measures by specifying the "wrong" model under either of these two simple model specifications.

METHODS

The Experimental Design

We base our Monte Carlo analysis on an actual CE study of recreational beach use in Santa Ponça Bay, a small Mallorcan tourism area.[3] We consider three site quality attributes, two representing measures of water quality (X_1, X_2), an indicator of congestion at the beach (X_3), and a cost attribute (X_4). Each attribute takes three possible levels. The features of the experimental design used in this paper are

[3] See Torres et al (2009) for more details on the study.

explained in Torres et al (2011), who base their MC analysis on the same recreational study. We use a D-efficient design allowing for main effects only.[4] The main features of the design are shown in table 1.

[Table 1]

Underlying Taste Heterogeneity and True Compensating Surplus

Given preference heterogeneity in the systematic part of utility has been commonly understood on the basis of RPL and LC models in environmental valuation, we focus on representations of these two types of taste heterogeneity at the first stage of our MC analysis. Thus, for both types of taste heterogeneity we consider two underlying linear-in-attributes utility functions with the same explanatory variables (X_1, X_2, X_3 and X_4).

For simplicity reasons, differences in preferences across individuals have only been assumed for the two environmental attributes X_1 and X_2. Thus, when true preferences are best described using an RPL model, each individual has been assigned their own parameters for X_1 and X_2, which represent mean attribute weights plus person-specific deviations from those means, as shown in Equation (1):

$$U_{jit} = (\alpha + \eta_i) X_{1jt} + (\beta + \psi_i) X_{2jt} + \gamma X_{3jt} + \omega X_{4jt} + \varepsilon_{jit} \quad (1)$$

where U_{jit} is the indirect utility of alternative *j* for individual *i* and choice occasion *t*, α, β, γ and ω are the known parameters of the attributes (i.e. mean attribute weights), η_i and ψ_i are individual-specific standard deviation parameters for α and β, respectively, and ε_{jit} is the error term associated with alternative *j* and individual *i* and choice occasion *t* (Train 1998; Train 2009)

We recognize that assuming a non-random coefficient for the cost attribute is a strong assumption as it implies the assumption of a constant marginal utility of income (Lanz and Provins 2013) which has implications for WTP estimates (Daily et al 2012). However, there is still a debate in the literature about the relative advantages of using a random or a non-random cost coefficient. Indeed, many authors state that a random cost parameter is associated with problems of identification of WTP values (Colombo et al 2007; Hensher 2005; Olsen 2009; Rigby et al 2009). Besides, the development of modeling approaches intended to overcome those problems, such as the WTP space estimation, is still yielding mixed results (Balcombe et al 2009).[5] According to this, and especially considering that in this

---

[4] Although one could argue that a WTP efficient design would be better when the focus is on welfare estimates, using a D-efficient design allows us to relate the simulations to real data collected using the same design.

[5] This helps to explain why many of the papers using RPL models that have been published in top journals in the last years still opt to use a fixed parameter for the price term both in the RP (Frondel and Vance 2013; Massey et al 2006; Moeltner and

paper we focus on mistaken assumptions about taste heterogeneity in a context where we control for the true underlying utility specification, we opt for specifying a non-random cost parameter to make our results clearer. However, we recognize that assuming constant the marginal utility of income is unrealistic and represents a limitation of this work for the use of the welfare estimations in real policy settings.

When true preferences are best described using a LC model, we consider heterogeneity for X_1 and X_2 is explained by the fact that individuals are assigned to two behavioural groups or latent (i.e. unobserved) segments on the basis of three LC covariates. The covariates considered to probabilistically determine membership of the two segments are two continuous variables, namely Age (Z_1) and Education (Z_2), and one dummy variable indicating if the individual belongs or not to some environmental organization (Z_3). Although it is not uncommon that only two latent classes in a given population best explain preferences (Bearry-Borg and Scarpa 2010; Birol et al 2006), our main motivation for using a simple set up with only two segments is to consider in comparison the most extreme case between the discrete and continuous distribution of the preferences. Indeed, the higher the number of latent classes, the closer the discrete and continuous distributions become. LC taste heterogeneity is then driven by the individual probability of membership in a latent class *s* (Equation 2) in such a way that preferences are assumed homogeneous within each class (Equation 3) but heterogeneous between segments (Train 2009).

$$P_{i/s} = \exp(\lambda_{1s}Z_{1i} + \lambda_{2s}Z_{2i} + \lambda_{3s}Z_{3i} + \xi_{is}) \Big/ \sum_{s=1}^{2} \exp(\lambda_{1s}Z_{1i} + \lambda_{2s}Z_{2i} + \lambda_{3s}Z_{3i} + \xi_{is}) \qquad (2)$$

where $P_{i|s}$ is the probability for individual *i* of membership in segment *s*, Z_{1i}, Z_{2i} and Z_{3i} are the covariates for individual *i*, λ_{1s}, λ_{2s} and λ_{3s} are the known parameters of the covariates for segment *s*, and ξ_{is} is a Gumbel distributed error term associated to individual *i* and segment *s* (Bhat 1997). Although a semi-parametric form based only on a constant term can be used to specify the membership probability (Scarpa and Thiene 2005), the most common specification is implemented with a set of socioeconomic covariates (Bujosa et al 2010).

Conditional on belonging to segment *s*, the utility of individual *i* for alternative *j* is specified as:

$$U_{jits} = \alpha_s X_{1jt} + \beta_s X_{2jt} + \gamma X_{3jt} + \omega X_{4jt} + \varepsilon_{jits} \qquad (3)$$

Shonkwiler 2005; Murdock 2006; Provencher and Bishop 2004) and the SP fields (Birol et al 2006; Beharry-Borg et al 2009; Boxall and Adamowicz 2002; Burton and Rigby 2009; Colombo et al 2007; Foster and Mourato 2003; Hensher et al 2005; Kaye-Blake et al 2009; Kipperberg and Larson 2012; Olsen 2009; Olsen et al 2011; Rigby et al 2009; Rolfe and Windle 2013).

where U_{jits} is the indirect utility of alternative *j* for individual *i*, choice occasion *t* and segment *s*, α_s and β_s are the known parameters of X_1 and X_2 for segment *s*, γ and ω are known parameters of X_3 and X_4 being constant for both segments, and ε_{jits} is the error term associated with alternative *j*, individual *i* , choice occasion *t* and segment *s*.

Note that when generating RPL simulated choices, we have not made the random parameters of X_1 and X_2 to depend on the LC socio-demographics. Comparing RPL and LC models in such a way is not an uncommon practice in the literature (Boxall and Adamowicz 2002; Broch and Vedel 2012; Outma et al 2007). In fact, as stated by Broch and Vedel (2012), the LC model is believed to be able to provide a different dimension for describing taste heterogeneity, where individuals are expected to have different motivations and aims for their choices, and therefore potentially belong to discrete groups based on latent variables. In the case of RPL, the heterogeneity is described by allowing the preference to vary according to a random distribution. We want to keep this difference in the data generation process.

With the purpose of measuring the difference between estimated and true compensating surplus (CS), we perform asimulation exercise using an experimental design based on a real empirical application involving an improvement in the good being valued (beach quality) described by changes in three of its attributes, namely water quality (X_1), the duration of an algal bloom (X_2) and crowding at the beach (X_3). Following Hanemann (1984), the CS at the individual level, defined as the WTP for a change in the attributes from the business-as-usual (BAU) scenario, has been calculated for the RPL and LC heterogeneity contexts as shown in Equations (4) and (5), respectively:

$$CS_i = -(1/\omega)(\alpha_i \Delta X_1 + \beta_i \Delta X_2 + \gamma \Delta X_3) \qquad (4)$$

$$CS_i = P_{i/s} CS_s + P_{i/s*} CS_{s*} \qquad (5)$$

$$P_{i/s*} = 1 - P_{i/s}$$

$$CS_{class} = -(1/\omega)(\alpha_{class} \Delta X_1 + \beta_{class} \Delta X_2 + \gamma \Delta X_3), class = s, s^*$$

where $\Delta X_1, \Delta X_2, \Delta X_3$ represent the changes in X_1, X_2 and X_3, respectively, from the policy-off to the policy-on context, CS_s and CS_{s*} are the CS corresponding to segment 1 and segment 2, respectively, being constant across individuals within each segment, and ω is the parameter for the cost attribute X_4 (or the marginal utility of income).

Table 2 shows the known parameters used to calibrate the utility function (equations 1 and 3 for RPL and LC model, respectively) and to calculate compensating surplus (equation 4 and 5). The scenario considered was a hypothetical change in X_1, X_2 and X_3 from a status quo level (a value of 60, 8 and 10, respectively, as shown in Table 1) to a situation in which they take the levels 20, 6 and 20, respectively, indicating a reduction in water pollution and in the congestion level at the beach. The values for all the parameters but α used in the RPL DGP are those used for the linear utility specification in Torres et al (2011). The values for the attribute parameters of segment 1 and 2 in the LC DGP have been chosen in such a way that individuals belonging to segment 1 are more sensitive to impacts on beach quality.[6] Besides, we consider individuals of segment 1 to be younger, have higher education levels and belong to an environmental organization.[7] This means that younger, better educated individuals belonging to an environmental organization will be more sensitive to beach quality. The mean and standard deviation of the three covariates used in the analysis are also reported in Table 2. The values assumed for the covariates and their parameters allow us to assume that the true probability of belonging to segment 1 is 55.15% and that of belonging to segment 2 is 44.85% We base the percentages on the findings by Birol et al. (2006), Beharry-Borg and Scarpa (2010) and Bujosa et al (2010), who, after showing that a LC model with two segments best fit the data, report segment shares ranging from 40% to 60%.[8]

[Table 2]

MC Experiments and Quantification of Errors in Welfare Estimates

At the second stage of the analysis, MC experiments have been undertaken to simulate choices for each of the two types of true taste heterogeneity when attribute values change in the way specified above. The utility of each alternative for each choice occasion has been calculated by combining the known parameters of the utility function (in table 2) with the attribute levels and an error term. These error terms have been generated from a Gumbel distribution, and a unique error has been randomly drawn not only for each alternative but also for each observation in the sample.

This procedure generates 2 datasets, one for each type of underlying true taste heterogeneity or Data Generating Process (DGP). In each dataset, for each choice task the simulated choice has been assigned to the alternative providing the highest utility level. In the simulation, 240 individuals have been considered. As each hypothetical individual faces 6 choice tasks (as might be the case in a typical empirical study), 1,440 (240x6) observations have been created by this process for each DGP. These observations represent the two underlying true forms of preference heterogeneity. Then, using these

---

[6] For comparability reasons, the values for the parameters representing homogeneous preferences among individuals (γ and ω) are the same in both DGPs.

[7] To set the parameter values for the covariates and attributes, we undertook LC model estimations on different sets of choices generated through LC schemes built on different parameter values. We chose the values for which the LC model showed a better performance.

[8] In particular, the class probabilities they find are 57.24% vs. 42.76% in Birol et al (2006), 61% vs. 39% in Beharry-Borg and Scarpa (2010), and 38.88% vs. 61.12% in Bujosa et al (2010). To set the values for the true segment shares, we run different LC models on LC choices generated under different values for the segment shares ranging from 40% to 60%. Again, we chose the values for which the LC model showed a better performance.

simulated samples of responses, RPL and LC models have been estimated in the usual manner, and welfare estimates calculated. By estimating different models under different DGPs it is possible to determine the errors that the analyst would incur when he/she mistakenly estimates an LC model when the true underlying preferences are distributed continuously and vice-versa.

The errors in welfare measurement from mistaken assumptions on the part of the analyst about the nature of taste heterogeneity have then been calculated for different scenarios, as shown in Table 3. First, a scenario 1 in which the analyst assumes preference heterogeneity for X_1 and X_2 is driven by the existence of two latent classes in the population when true preferences are lognormally-distributed (i.e. by erroneously estimating a LC model when the true DGP is characterised by an RPL). Second, a scenario 2 where the parameters for X_1 and X_2 are assumed to vary across individuals according to a lognormal distribution when true preference heterogeneity is driven by the existence of two latent classes (i.e. erroneously estimating a RPL model when the true DGP is characterised by a LC DGP). Third, to examine the implications of assuming a parameter distribution other than the lognormal one, scenarios 3 and 4 in which RPL models assuming triangular-distributed parameters have been estimated under the two types of DGPs stated above (LC and RPL (log-normal)), respectively. Finally, an additional analysis where the analyst assumes preference homogeneity for X_1 and X_2 (i.e. erroneously estimating a Multi-Nomial Logit model when preferences are indeed heterogeneous) has also been undertaken (scenarios 5 and 6). [Table 3]

Taking into account the 2 types of underlying true DGPs and the four analytical scenarios (LC, RPL-Log, RPL-Triang, MNL), 8 different MC experiments (2x4) have thus been undertaken. The individual CS values for the same change in X_1 , X_2 and X_3 have been estimated for each MC experiment following Equations (4) and (5) according to the type of estimated model (i.e. RPL or LC). This process has been repeated 1,000 times. Next, the importance of using the correct model to account for taste heterogeneity has been examined by quantifying the individual errors in the estimated CS values. To do this, the mean squared proportional error (MSPE) has been calculated at the individual level for each MC replication, according to equation (6):

$$MSPE_r = (1/I)\left[\sum_{i=1}^{I}\left(\left(CS_{ir}^{e} - CS_i^{t}\right)/CS_i^{t}\right)^2\right] \qquad (6)$$

where *r* is a specific repetition of the MC experiment, *I* is the total number of simulated individuals, CS_{ir}^{e} is the estimated CS of individual *i* in repetition *r* and CS_i^{t} is the true CS of individual *i*. The MSPE represents the square of the ratio between bias (the difference between the estimated and true CS) and the true CS. The MSPE is the most appropriate measure relative to other accuracy measures typically used in the literature such as Bias, Relative Bias or Mean Square Error for two reasons: 1) it is a relative measure and, hence, it is independent of the magnitude of the true CS, thus making comparable the results from the MC experiments; and 2) it gives an idea of not only the accuracy but also the efficiency (the variance) of welfare estimates. At each MC repetition, the MSPE has been calculated as the average over 240 individual welfare measures. After 1,000 MC repetitions, a distribution of MSPE mean values for the change in X_1 , X_2 and X_3 has been obtained for each experiment. The values for MSPE reported for each MC experiment have been calculated as the average of the sum of the mean values obtained in each MC replication (equation 6) over the 1,000 repetitions.

RESULTS

The results of the MSPE in the estimated CS for each MC experiment are reported in Table 4. As stated above, these values refer to a hypothetical change in the attributes X_1 , X_2 and X_3 from the baseline levels of $X_1 = 60$, $X_2 = 8$ and $X_3 = 10$ to the levels of $X_1 = 20$, $X_2 = 6$ and $X_3 = 20$. MSPE measures are shown in terms of the two DGPs (i.e. true RPL and true LC preferences) and the estimation model (MNL, RPL and LC) used in the simulations.[9]

[Table 4]

When true preferences are characterised by continuous mixing as in the RPL-Log case, mistakes about the correct distribution within the RPL model seem irrelevant; both RPL-Log and RPL-Triang provide the same MSPE. This is probably due to the high adaptability of the RPL model to fit the underlying true random distribution. A mistaken assumption of LC preferences produces a relatively small increase in the error in welfare estimation, as the MPSE increases by a factor of 1.7 (from 0.037 to 0.056). As it may be expected, mistakenly assuming homogeneous preferences by specifying a multi-nomial logit leads to much larger increase in the error, a factor of 4.7 (from 0.037 to 0.143). A different situation is observed when true preferences are characterised by finite mixing with a 2-class structure. The smallest error is achieved when analysts correctly "guess" the true underlying heterogeneity. However, a larger relative error results from mistakenly specifying the model as RPL, as the MPSE increases by a factor of 5 (from 0.026 to 0.106), relative to the previous case. Again, the distributional form of RPL model is of little consequence, and a multi-nomial logit specification leads the error to increase by a factor of 21, that is, it produces the biggest relative error (from 0.026 to 0.428).

The main result to this juncture is thus that errors in welfare estimation from getting the preference heterogeneity specification wrong are most important when true preferences result in a LC DGP. In contrast, if the analyst mistakenly assumes that the preference heterogeneity is discrete when in reality

---

[9] Note that, for comparability reasons, results from using the correct model to account for taste heterogeneity have also been reported.

it follows a continuous distribution, then this produces a smaller error in consumers' surplus than the opposite assumption about taste heterogeneity.

An interesting follow-up question which arises is: why does the LC specification do relatively well when the true utility function is RPL-log? One possible reason concerns the way in which "errors" are considered here, that is, with respect to welfare measures (CS) rather than preference parameters.[10] In this case what matters is the similarity between the estimated and the true underlying CS distributions more than the similarity between the true and estimated preference distributions. In LC, although preferences are discrete, generating welfare estimates requires preference parameters to be combined with a continuous membership probability function. As such, even when the true CS distribution results from a set of preferences which vary continuously, the "mixing" between the utility parameters and class membership probability draws a continuous distribution of CS that approximates well to the true underlying CS distribution. Thus the MPSE error measures can be small.

These results suggest that it would be fruitful to examine the effects on welfare error magnitudes of changing the distribution of true CS when this originates from either a discrete or a continuous preference structure. Thus, starting from the assumption that the true CS distribution originates from a discrete preference structure (LC DGP), we change the distribution of variables which determine LC membership and calculate MSPE errors for the four model specifications described above (LC-2 seg, RPL-Log, RPL-Triang, MNL). Put another way, we change the true share of the two segments leading to a given true CS distribution without changing the segment covariate and attribute parameters.[11] This allows us to consider the same two segments in terms of their environmental preferences and socioeconomic features, thus making results comparable. As the CS distribution results from the combination of discrete mixing preferences and the continuous distribution of the class membership probability function, it enables us to use a wider variety of true CS distributions facilitating the analysis of CS distribution similarities. Thus, we consider three cases which are labelled A, B and C in Table 5. In Case A, class one probabilistically contains 28% of respondents and class two, 72%, whilst in case C the figures are (78%, 22%), respectively.[12] . Case B represents an intermediate situation whose results

[10] We argue that in the empirical application of a CE aiming at disclosing the social value of a good or service, the leading measure to feed benefit-costs analysis is the compensating surplus rather than the preferences towards the attributes.

[11] For example, in this application this is achieved by simply modifying the distribution of the Age, Education and "Belonging to an environmental organization" variables in the simulated sample.

[12] To set the values for representative cases of low, medium and high segment shares, we tried to estimate LC models on LC choices generated under the assumptions that the true share for segment 1 was 25%, 50% and 75% (and 75%, 50% and 25% for segment 2, respectively). However, LC models collapsed for those values. Thus the values we report in table 5 are those for which the LC model showed the best performance. We did not consider segment shares lower than 25% as this would have implied to assign a higher than 75% share to the second class, which, as earlier said, would have moved us away from common findings in the literature (Beharry-Borg and Scarpa 2010; Birol et al 2006; Bujosa et al 2010).

are described above (in Table 4) .[13]. In Table 5, we show the MSPE for the three cases considering the four model specifications used above under a LC DGP.

[Table 5]

As expected, in all cases the LC model provides the lowest error in terms of matching the estimated model with the true underlying DGP. Comparing Case A with Case B it may be seen that the MSPE value from (erroneously) using an RPL model rather than a LC model increases solely as a result of this change in the class membership distribution. Case C gives a similar outcome for specifying RPL rather than the (correct) LC model as B, but a much smaller error from specifying a MNL model. The RPL model performs much better in Case A. The reason can be easily spotted by considering the distribution of the true CS in the three cases investigated. These are shown in Figures 1-3

[Figures 1-3]

As can be seen a log-normal distribution fits better with a true distribution of CS values as in case A, relative to case B or case C. Because of this, the RPL model which assumes a log-normal distribution for its parameters provides smaller errors in the MSPE of case A relative to the other cases.[14] However, for cases B and C, the RPL model performs poorly because the resulting true underlying distribution of CS differs notably from that estimated using the RPL models.

We now further investigate the robustness of the LC model in terms of welfare measurement by assuming now that the true CS distribution comes from an RPL DGP. For this purpose, we simulate three types of RPL-Log DGP scenarios differing only in the variance of the random parameters – labelled *Small Var*, *Medium Var* and *High Var*. That is, we consider a true CS distribution which originates from a set of preferences which are continuously distributed according to a log-normal distribution with different variances. For the three cases, we consider the variance of the random parameters is about 13%, 45% and 75% of their means according to common findings in the literature. Changing the known value for the variance of the random parameters rather than the known values for

[13] To set the values for representative cases of low, medium and high segment shares, we tried to estimate LC models on LC choices generated under the assumptions that the true share for segment 1 was 25%, 50% and 75% (and 75%, 50% and 25% for segment 2, respectively). However, LC models collapsed for those values. Thus the values we report in table 5 are those for which the LC model showed the best performance. We did not consider segment shares lower than 25% as this would have implied to assign a higher than 75% share to the second class, which, as earlier said, would have moved us away from common findings in the literature (Beharry-Borg and Scarpa 2010; Birol et al 2006; Bujosa et al 2010).

[14] Clearly an RPL model with lognormally-distributed parameters and a fixed cost term generates a distribution of WTP which is lognormal.

their means allows us to keep the analysis focussed on the same preferences towards the environment, thus making results more comparable. The MSPE values are described in Table 6.

[Table 6]

Again, on average the model which provides the most accurate measures is the one which is in accordance to the true underlying DGP (RPL-Log in this case), and assuming the wrong distribution of the random parameter has little effect. However, it is worth noting that when the analyst mistakenly assumes discrete preference heterogeneity when the true underlying source of preference heterogeneity is continuous, the MSPE values are lower than in the opposite case. As a general finding here, the LC model seems to be more robust than the RPL model to mis-specification of the preference heterogeneity. The suggested reason is the large flexibility of the LC model to adapt to the true underlying distribution of CS by means of combining the utility parameter and class probability distributions.

DISCUSSION AND CONCLUSIONS

How best to model preference heterogeneity for environmental goods has been of on-going interest to choice modellers working with both revealed and stated preference data for many years. In this paper, we have compared the two most usual approaches to date in the environmental literature, namely RPL and LC, exploring the impact on welfare measures. Our paper complements earlier work by investigating the relative errors from mis-specifying the model of preference heterogeneity when the true DGP is unknown to the researcher. This is a relevant question, since the "best" choice of econometric model to fit to a given data set of choices is often a hard question to answer: trade-offs often exist between different measures of model performance such as fit, flexibility, consistency with underlying theory, parsimony and simplicity of interpretation. We do this using a MC approach, focussing on the deterministic element of utility within a random utility set-up (that is, ignoring scale heterogeneity). This approach has the great merit that it enables us to measure the true, underlying (money-metric) utility change from a change in environmental quality, and then compare this "true" measure with the estimated welfare change under different econometric modelling assumptions.

Our main findings are that, in presence of taste heterogeneity, the smallest welfare errors are always found when analyst estimates a model whose treatment of heterogeneity mirrors the underlying true heterogeneity. However, the form of this underlying true preference heterogeneity is typically unknown to the analyst, whilst the conventional tests based on model fitting and log-likelihood are not always conclusive due to the different nature of model specifications. When uncertainty exists about how to model preference heterogeneity and the main interest of the analyst rests in the generation of welfare measures for use in cost-benefit analysis, we find that the LC modelling seems to be a more robust and flexible approach in the sense of generally resulting in smaller errors. This model provides CS measures which are not so distant from the true measures, even if the true DGP is described by continuous (RPL) distribution. This result has been observed for the specific case of a 2-latent class model which represents the "most extreme" case of preference difference between the assumed continuous and discrete distributions. Having more latent classes[15] would offer a higher flexibility to the model and further reduce the gap between the true and estimated CS. Changes in the distribution of people across latent classes have effects on the error in welfare measurement from mis-specified models, although the LC maintains a higher degree of precision and accuracy. The performance of an RPL model is impacted by the correct selection of the distributions for the random parameters. We also find that the errors from using a LC model when the true DGP is continuous mixing depend on the variance of this distribution across the population, with the LC modelling proving to be a robust approach even in the presence of high variances.[16] Our overall conclusion is then that smaller welfare errors are likely to come from use of a latent class model, which is in line with the findings by Birol et al. (2006), Colombo et al, (2009), Hynes et al. (2008), Shen (2009) and Sagebiel (2011).

In this context, one can argue that with real (as opposed to simulated) data, the analyst never knows the true DGP and hence MC analysis plays a limited role in drawing conclusions about the most desirable empirical strategy, as its results will always be restricted to the assumptions made about the DGP. Generalising the findings from a MC study in this context can be challenging. However, assumptions about the underlying individual choice behaviour have also to be made when working with real data: the analyst has to specify a model according to her assumptions about individual preferences. Uncertain knowledge about true preferences means that any econometric model will be mis-specified to a degree. After all, models are by definition only approximations to reality (Fiebig et al 2010). The impossibility of a "perfect" model specification –or, equivalently, the acceptance that all models are mis-specified to a degree– makes necessary the use of tools aimed at analyzing the performance of alternative models for a given data set. In this setting, MC analysis plays a crucial role. Indeed, in contrast to common tests based on model fitting and log-likelihood, it allows analysts to generate data and hence examine model performance through quantification of welfare errors from mistaken model assumptions. Findings from studies such as this can have important implications for the use of benefit-cost analysis in policy evaluation, if they point to the likely occurrence of large errors in benefits estimation from following particular empirical strategies.

In this paper, in the RPL approach we have assumed two of the most common distributions for random parameters: lognormal and triangular.[17] However, several other distributions are available in the literature (although many are not implemented in the standard statistical software available to analysts)

[15] Subject to criteria such as a minimum class membership probability, for example.

[16] The RPL which uses a triangular distribution drastically increases the MSPE error when the variances of the random parameters are large.

[17] The triangular distribution has been used instead of the more common normal distribution to avoid the effect of the tails of the normal in the CS measures.

and the sensitivity of results to the use of these alternative distributions needs to be tested in future research. Additionally, it would be interesting to examine how results change when estimating the standard RPL or LC model when heterogeneity in true preferences is not driven by taste heterogeneity alone but by a combination of taste and scale heterogeneity, or only by the latter. An obvious extension of the current research is to apply the simulation to a wider context of models accounting for preference heterogeneity, including those that combine aspects of both finite and continuous mixing (Greene and Hensher 2013), and boundedly-rational decision making rules such as attribute non-attendance (Scarpa et al 2009).

REFERENCES

Adamowicz, W.L., J.J. Fletcher and T. Graham-Tomasi. 1989. Functional form and the statistical properties of welfare measures. *American Journal of Agricultural Economics* 71(2): 414-421.

Alberini, A. 1995. Efficiency vs. bias of willingness-to-pay estimates: Bivariate and interval-data models. *Journal of Environmental Economics and Management* 29(2): 169-180.

Bearry-Borgh, N., D.A. Hensher and R. Scarpa. 2009. An analytical framework for joint vs. separate decisions by couples in choice experiments. The case of coastal water quality in Tobago. *Environmental and Resource Economics* 43: 95-117

Bearry-Borg, N. and R. Scarpa. 2010. Valuing quality changes in Caribbean coastal wàters for heterogeneous beach visitors. *Ecological Economics* 69: 1124-1139

Ben-Akiva, M. and S. R. Lerman .1985. *Discrete choice analysis: Theory and application to travel demand*. Massachussetts: The MIT Press.

Bhat, C.R. 1997. An endogenous segmentation mode choice model with an application to intercity travel. *Transportation Science* 31(1): 34–48.

Birol, E., K. Karousakis and P. Koundouri. 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. and W. Adamowicz. 2002. Understanting heterogeneous preferences in random utility models: A latent class approach. *Environmental and Resource Economics* 23: 421-446

Broch, S.W., S.E. Vedel. 2012. Using Choice Experiments to Investigate the Policy Relevance of Heterogeneity in Farmer Agri-Environmental Contract Preferences. *Environmental and Resource Economics* 51(4): 561-581.

Bujosa, A., A. Riera and R.L. Hicks. 2010. Combining discrete and continuous representations of preference heterogeneity: a latent class approach. *Environmental and Resource Economics* 47: 477-493

Burton, M. and D. Rigby. 2009. Hurdle and latent class approaches to serial non-participation in choice models. *Environmental and Resource Economics* 42: 211-226

Carlsson F. and P. Martinsson. 2003. Design techniques for stated preference methods in health economics. *Health Economics* 12: 281-294

Claassen J., D. Hellerstein and S.G. Kim 2013. Using mixed lògit in land use models: can expectation-maximization (EM) algorithms facilitate estimation? *American Journal of Agricultural Economics* 95(2): 419-425.

Colombo S., J. Calatrava-Requena and N. Hanley. 2007. Testing choice experiments for benefit transfer with preference heterogeneity. *American Journal of Agricultural Economics* 89(1), 135-151

Colombo S., N. Hanley and J.J. Louviere. 2009. Modeling preference heterogeneity in stated choice data: an analysis for public goods generated by agriculture. *Agricultural Economics* 40: 307-322

Daily, A.J., S. Hess and E. Train. 2012. Assuring finite moments for willingness-to-pay in random coefficients models. *Transportation* 39(1): 19-31

Ferrini S. and R. Scarpa. 2007. Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study. *Journal of Environmental Economics and Management* 53(3): 342-363

Fiebig, D.G., M.P. Keane, J.J. Louviere and N. Wasi. 2010. The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Science* 29(3): 393-421

Flynn, T.N., J.J. Louviere, T.J. Peters and J. Coast. 2010. Using discrete choice experiments to understand preferences for quality of life. Variance-scale heterogeneity matters. *Social Science and Medicine* 70: 1957-1965.

Foster V. and S. Mourato. 2003. Elicitation format and sensitivity to scope. Do contingent valution and choice experiments give the same results? *Environmental and Resource Economics* 24: 141-160.

Frondel, M. and C. Vance. 2013. Heterogeneity in the Effect of Home Energy Audits: Theory and Evidence. *Environmental and Resource Economics* 55: 407-418

Greene, W. and D. Hensher. 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B* 37: 681-698

Greene, W.H. and D.A. Hensher. 2010. Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. *Transportation* 37 (3): 413-428

Greene, W.H. and D.A. Hensher. 2013. Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. *Applied Economics* 45(14): 1897-1902

Hanemann, W.M. 1984. Applied welfare analysis with quantitative response models. Working paper no. 241. University of California, Berkeley.

Hanley, N. and E.B. Barbier. 2009. *Pricing Nature: Cost-Benefit Analysis and Environmental Policy Making*. Cheltenham: Edward Elgar.

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

Hensher, D., J. Rose and W. Greene. 2005. *Applied Choice Analysis: a primer*. Cambridge: Cambridge University Press.

Hensher, D., N. Shore and K. Train. 2005. Household's willingness to pay for water service attributes. *Environmental and Resource Economics* 32:509-531.

Herriges, J.A. and C.L. Kling. 1997. The performance of nested logit models when welfare estimation is the goal. *American Journal of Agricultural Economics* 79(3): 792-802.

Hess, S. and N. Beharry-Borg. 2012. Accounting for Latent Attitudes in Willingness-to-Pay Studies: The Case of Coastal Water Quality Improvements in Tobago. *Environmental and Resource Economics* 52 (1): 109-131

Hess, S. and J. Rose. 2012. Can scale and coefficient heterogeneity be separated in random coefficients models? *Transportation* 39(6): 1225-1239.

Hynes, S., N. Hanley and R. Scarpa. 2008. Effects on welfare measures of alternative means of accounting for preference heterogeneity in recreational demand models. *American Journal of Agricultural Economics* 90(4): 1011-1027.

Johansson, P.O. 1993. Cost-benefit analysis of environmental change. Cambridge: Cambridge University Press.

Kipperberg, G. and D.M. Larson. 2012. Heterogeneous preferences for community recycling programs. *Environmental and Resource Economics* 53 (4): 577-604.

Kling, C.L. 1987. A simulation approach to comparing multiple site recreation demand models using Chesapeake Bay survey data. *Marine Resource Economics* 4: 95-109.

Kling, C.L. 1988. The reliability of estimates of environmental benefits from recreation demand models. *American Journal of Agricultural Economics* 70(4): 892-901.

Kling, C.L. 1989. The importance of functional form in the estimation of welfare. *Western Journal of Agricultural Economics* 14(1): 168-174.

Kling, C.L. 1997. The gains from combining travel cost and contingent valuation data to value nonmarket goods. *Land Economics* 73(3): 428-439.

Kling, C.L. and C.J. Thomson. 1996. The implications of model specification for welfare estimation in nested logit models. *American Journal of Agricultural Economics* 78(1): 103-114.

Lanz, B. and A. Provins. 2013. Valuing Local Environmental Amenity with Discrete Choice Experiments: Spatial Scope Sensitivity and Heterogeneous Marginal Utility of Income. *Environmental and Resource Economics* 56: 105-130.

Louviere, J.J. and T. Eagle. 2006. Confound it! that pesky little scale constant messes up our convenient assumptions. Paper presented at the Sawtooth Software Conference, Washington, USA.

Louviere, J.J., R.J. Meyer, D.S. Bunch et al. 1999. Combining sources of preference data for modelling complex decision processes. *Marketing Letters* 10(3): 205-217.

Louviere, J.J., D. Street, L. Burgess et al. 2008. Modeling the choices of individual decision makers by combining eficient choice experiment designs with extra preference information. *Journal of Choice Modelling* 1(1): 128-163.

Louviere, J.J., D. Street, R. Carson et al. 2002. Dissecting the random component of utility. *Marketing Letters* 13: 177-193.

Louviere, J.J., D.A. Hernsher and J.D. Swait. 2000. Stated choice methods: analysis and application. Cambridge: Cambridge University Press.

Lusk, J.L. and F.B. Norwood. 2005. Effect of experimental design on choice-based conjoint valuation estimates. *American Journal of Agricultural Economics* 87(3): 771-785.

Massey, D.M., S.C. Newbold and B. Gentner. 2006. Valuing water quality changes using a bioeconomic model of a coastal recreational fishery. *Journal of Environmental Economics and Management* 52 (1):482-500.

Meyer, R.J. 2007. Formal choice models of informal choices: what choice modelling research can (and can't) learn from behavioral theory. In *Review of Marketing Research* (vol.4) edited by N.K. Malhotra. New York: Emerald Group Publishing Limited.

Milon, J.W. and D. Scrogin. 2006. Latent preferences and valuation of wetland ecosystem restoration. *Ecological Economics* 56: 162-175.

Moeltner, K. and J.S. Shonkwiler. 2005. Correcting for on-site sampling in random utility models. *American Journal of Agricultural Economics* 87(2): 327-339.

Murdock, J. 2006. Handling unobserved site characteristics in random utility models of recreation demand. *Journal of Environmental Economics and Management* 51(1): 1-25.

Olsen, S.B., T.H. Lundhede, J.B. Jacobsen et al. 2011. Tough and easy choices: testing the influence of utility difference on stated certainty-in-choice in choice experiments. *Environmental and Resource Economics* 49: 491-510.

Ouma, E., A. Abdulai and A. Drucker. 2007. Measuring heterogeneous preferences for cattle traits among cattle-keeping households in East Africa. *American Journal of Agricultural Economics* 89 (4): 1005-1019.

Provencher, B. and R. Bishop. 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.

Rigby, D., K. Balcombe and M. Burton. 2009. Mixed Logit Model Performance and Distributional Assumptions: Preferences and GM foods. *Environmental and Resource Economics* 42: 279-295.

Rolfe, J and J. Windle. 2013. Distance Decay Functions for Iconic Assets: Assessing National Values to Protect the Health of the Great Barrier Reef in Australia. *Environmental and Resource Economics* 53(3): 347-365.

Sagebiel, J. (2011). Comparing the latent class model with the random parameters logit. A choice experiment analysis of highly heterogeneous electricity consumers in Hyderabad, India. II *International Choice Modeling Conference, Leeds, 2011.*

Scarpa, R. and I. Bateman. 2000. Efficiency gains afforded by improved bid designs versus follow-up valuation questions in discrete-choice CV studies. *Land Economics* 76(2): 299-311.

Scarpa, R., T. Gilbride, D. Campbell et al. 2009. Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics* 36(2): 151-174.

Scarpa, R. and J. Rose. 2008. Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. *Australian Journal of Agricultural and Resource Economics* 52(3): 253-282.

Scarpa, R. and M. Thiene. 2005. Destination choice models for rock climbing in the Northeastern Alps: a latent class approach based on intensity of preferences. *Land Economics* 81(3): 426–444.

Shen, J. and Saijo, T. (2009). Does an energy efficiency label alter consumers' purchasing decisions? A latent class approach based on a stated choice experiment in Shanghai. *Journal of Environmental Management* 90: 3561-3573.

Strazzera, E., Contu, D. and Ferrini, S. (2013), Check it out! A Monte Carlo analysis of the performance of selection criteria and tests for Choice Experiments models, International Choice Modelling Conference 2013, The Sebel Pier One Sydney, 3 July 2013 – 5 July 2013.

Torres C., N. Hanley and A. Riera. 2011. How wrong can you be? Implications of incorrect utility function specification for welfare measurement in choice experiments. *Journal of Environmental Economics and Management* 62: 111-121.

Train, K.E. 1998. Recreation demand models with taste differences over people. *Land Economics* 74(2): 230-239.

Train, K.E. 2009. Discrete Choice Methods with Simulation. Cambridge: Cambridge University Press, second edition.

TABLES

Table 1. Features of the experimental design

Experimental design factors		Design[a]
Attribute levels	X_1: Loss of water transparency (%)	20 40 60*
	X_2: Duration of algal bloom (weeks)	3 6 8*
	X_3: Congestion (meters)	3 10* 20
	X_4: Cost (€)	3 10.5 24 (0*)
Alternatives		2+BAU
Choice sets per individual		6
Blocks		6
Block replications		40
Total observations[b]		1,440

[a]Single asterisk (*) denotes the levels for the BAU option.

[b]Total observations are the number of choice sets x the number of blocks x the number of block replications.

Table 2. Known parameters and true consumer surplus (CS)			
	Taste heterogeneity scenarios[a]		
		LC-2 seg	
Parameters	RPL-Log	Segment 1	Segment 2
α	-1.8	-3	-1.5
β	-0.7	-1.9	-0.1
γ	0.4	0.4	0.4
ω	-0.8	-0.8	-0.8
λ_1		-0.5	0
λ_2		1	0
λ_3		5	0
Z_1_mean *(std. dev.)*[b]		44.90 *(12.97)*	
Z_2_mean *(std. dev.)*[b]		2.23 *(0.69)*	
Z_3_mean *(std. dev.)*[b]		0.51 *(0.50)*	
True shares		0.5515	0.4485
True CS[c]	15.45	19.37	

[a] RPL-Log represents the RPL preference scenario, where α and β are lognormally-distributed, with 1.8 mean and 0.45 variance for α, and 0.7 mean and 0.18 variance for β. Note that we assume the variance being a 25% of the mean. LC-2 seg represents the LC preference scenario where two latent segments exist in the population with shares 0.55 and 0.45.

[b] The minimum and maximum values of Age are 18 and 83, respectively. Education has only 3 levels meaning low (1), medium (2) and high (3) education levels. EO takes value 1 when the individual belongs to an environmental organization. Therefore, the assumed population consists of mid-life, educated, environmentally-aware individuals.

[c] The true CS has been obtained by averaging the individual CS values over all the simulated individuals.

Table 3. Description of scenarios considered to measure welfare errors		
Scenarios	True DGP[a]	Estimated model[b]
Scenario 1	RPL-log	LC-2seg
Scenario 2	LC-2seg	RPL-Log
Scenario 3	RPL-log	RPL-Triang
Scenario 4	LC-2seg	RPL-Triang
Scenario 5	RPL-log	MNL
Scenario 6	LC-2seg	MNL

[b] DGP means Data Generating Process, which can follow either a RPL-Log scheme with lognormally-distributed parameters for X_1 and X_2 or a LC-2seg scheme with 2 segments.

[b] LC-2seg means estimating a LC with 2 segments, whilst RPL-Log and RPL-Triang refer to a RPL with lognormally- and triangular distributed parameters for X_1 and X_2, respectively. MNL refers to the Multinomial Logit Model.

Table 4. Mean square proportional error (MSPE) in the estimated value of a hypothetical change in the attributes (over 1,000 repetitions)

True DGP	Estimation model[a]	MSPE
RPL-Log	RPL-Log	0.037
	RPL-Triang	0.037
	LC-2seg	0.056
	MNL	0.143
LC-2seg	LC-2seg	0.026
	RPL-Log	0.106
	RPL-Triang	0.100
	MNL	0.428

[a]RPL-Log means estimating a RPL assuming lognormally-distributed parameters for X_1 and X_2, whilst RPL-Triang means estimating a RPL model assuming triangular-distributed parameters for these attributes. LC-2seg means estimating a LC model with 2 segments. DGP is Data Generating Process.

Table 5. MSPE in the estimated value of a hypothetical change in the attributes under different segment shares (over 1,000 repetitions)

		MSPE for different true segment shares[b]		
		Case A	Case B	Case C
True DGP	Estimation model[a]	S1: 0.28 S2: 0.72	S1: 0.55 S2: 0.45	S1: 0.78 S2: 0.22
LC-2seg	LC-2seg	0.030	0.026	0.024
	RPL-Log	0.057	0.106	0.099
	RPL-Triang	0.088	0.100	0.071
	MNL	0.137	0.428	0.262

[a] RPL-Log means estimating a RPL assuming lognormally-distributed parameters for X_1 and X_2, whilst RPL-Triang means estimating a RPL model assuming triangular-distributed parameters for these attributes. LC-2seg means estimating a LC model with 2 segments.

[b] S1 means true share for segment 1 and S2 means true share for segment 2.

Table 6. MSPE in the estimated value of a hypothetical change in the attributes under different levels of variances for the RPL-Log DGP (over 1,000 repetitions)

Estimation model[a]	RPL LOG DGP[b]		
	Small var	Medium var	High var
RPL-Log	0.031	0.037	0.048
RPL-Triang	0.028	0.037	0.405
LC-2seg	0.042	0.056	0.079
MNL	0.075	0.143	0.392

[a] RPL-Log means estimating a RPL assuming lognormally-distributed parameters for X_1 and X_2, whilst RPL-Triang means estimating a RPL model assuming triangular-distributed parameters for these attributes. LC-2seg means estimating a LC model with 2 segments.

[b] In Small var, α and β are lognormally-distributed, with 1.8 mean and 0.23 variance for α, and 0.7 mean and 0.09 variance for β; in Medium var, α and β have 1.8 mean and 0.45 variance, and 0.7 mean and 0.18 variance, respectively; and in High var, they have 1.8 mean and 1.35 variance, and 0.7 mean and 0.53 variance, respectively.

FIGURES

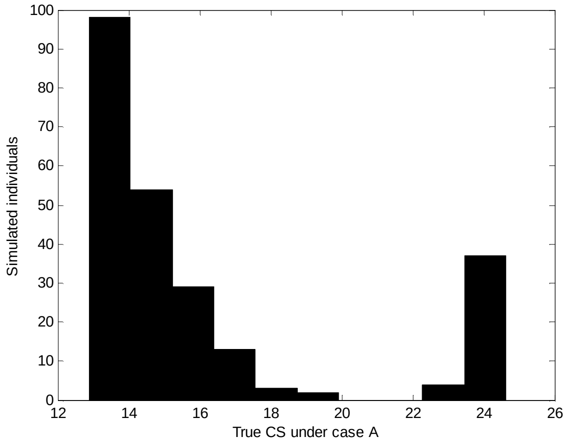

Fig.1. Distribution of true consumer surplus (CS) under case A. In Case A, class one probabilistically contains 28% of respondents and class two, 72%.

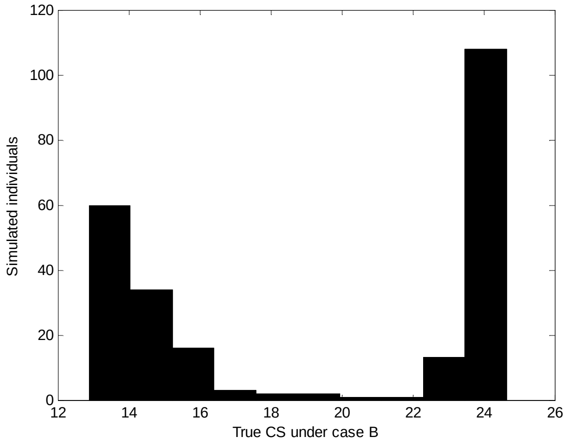

Fig.2. Distribution of true consumer surplus (CS) under case B. In Case B, class one probabilistically contains 55% of respondents and class two, 45%.

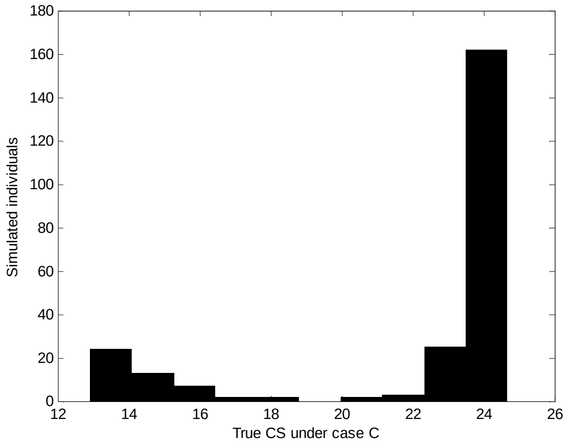

Fig.3. Distribution of true consumer surplus (CS) under case C. In Case C, class one probabilistically contains 78% of respondents and class two, 22%.