UNIVERSITY OF STIRLING

Stirling Management School

Incorrectly accounting for taste heterogeneity in choice experiments: Does it really matter for welfare measurement?

Cati Torres

Nick Hanley

Sergio Colombo

Stirling Economics Discussion Paper 2011-02

February 2011

Online at
http://www.management.stir.ac.uk/research/economics/working-papers

Incorrectly accounting for taste heterogeneity in choice experiments: Does it really matter for welfare measurement?

Cati Torres

Centre de Recerca Econòmica (UIB· SA NOSTRA), 07122-Palma de Mallorca, Spain.

Nick Hanley

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

Sergio Colombo

Agricultural Economics Department, IFAPA, Junta de Andalusia, Spain.

Abstract

A range of empirical approaches to representing preference heterogeneity have emerged in choice modelling. Researchers have been able to explore the differences which 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. However, the question as to the degree of error in welfare estimation from an inappropriate choice of empirical approach has not been addressed. In this paper, we use Monte Carlo analysis to address this question. Given the high popularity of both the random parameter logit (RPL) and latent class models among choice modellers, we examine the errors in welfare estimates from using the incorrect model to account for taste preference heterogeneity. Our main finding is that using an RPL specification with log-normally distributed preferences seems the best bet.

Keywords: preference heterogeneity, welfare measurement, accuracy, efficiency, choice experiments, Monte Carlo analysis.

***JEL classification**:* C51, D69, C99, C15.

1. 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 environmental, transport and health economics (Louviere et al., 2000; Hensher et al., 2005). It can be applied to both stated preference (SP) and revealed preference (RP) data. Its advantages are 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 which 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 (Hynes et al., 2008; Colombo et al., 2009).

However, the question as to the degree of error in welfare estimation from an inappropriate choice of empirical approach –in terms of the difference between *estimates* of the money metric measure of welfare change from choice data 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 measure of welfare change (Johansson, 1993).

In this paper, we use Monte Carlo (MC) analysis to address this question. MC analysis allows the researcher to start with a particular utility function and a particular distribution of preferences across a population of agents, and then to simulate choices to a particular set of 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.

The structure of the rest of this paper is as follows. The next section provides a review of the environmental valuation literature focusing on modelling preference heterogeneity in stated choice data to show the lack of studies examining bias and variance of welfare measures. Section 3 discusses the methodology used and the data employed for the experiments. Results are reported in section 4, where the sensitivity of welfare measures to mistaken assumptions about the nature of taste heterogeneity is analysed. Conclusions are drawn in section 5.

2. Analyzing the accuracy and efficiency of welfare estimates when modelling preference heterogeneity in stated choice data: a gap in the literature.

As a response to the weaknesses of the conditional logit approach to represent preference heterogeneity, alternative approaches have grown in popularity among discrete choice modellers. In this sense, random parameter logit (RPL) and latent class (LC) models have emerged to account for heterogeneity in the systematic component of utility, that is, when heterogeneity is thought to affect tastes (i.e. taste heterogeneity). Several authors in the field of transport, leisure and environmental economics have compared the performance of RPL and LC approaches to choice data to determine which one fits the data better and to examine differences in welfare estimates. Boxall and Adamowicz (2002), Green and Hensher (2003), Provencher and Bishop (2004), Birol el al. (2006) and Hynes et al. (2008) carry out this comparison. The empirical results show that there is no clear pattern of which approach is superior to the other, although some other authors find that LC model performs better than the RPL model.

In addition to modelling taste heterogeneity, research has recently recognised the importance of modelling scale heterogeneity. The main reason has been pointed out by Louviere several times (1999; 2002; 2006); all statistical models in which the dependent variable is latent confound estimates of model parameters with error variability, and as such the parameter estimates do not represent mean tastes but the means multiplied by the scale factor. As a result a growing literature which aims specifically to model scale heterogeneity alone, or taste and scale heterogeneity jointly, has emerged. Fiebig et al. (2009) compare a RPL model with a scale heterogeneity multinomial logit model where only scale heterogeneity is allowed, and a generalized multinomial logit (G-MNL) model where both taste and scale heterogeneity are allowed. They conclude that models which allow for both scale heterogeneity and G-MNL model outperform the RPL model especially in datasets that involve more complex choices. These findings are confirmed by a recent similar investigation (Greene & Hensher, 2010) which find that accommodating only scale heterogeneity (i.e. neglecting preference heterogeneity) may be of limited empirical interest, resulting in a statistically inferior model, whereas the inclusion of both scale and preference heterogeneity results in an improvement over the standard RPL model. Importantly, Green and Hensher (2010) observe that compared to failure to include for preference heterogeneity that is consequential, failure to account for scale heterogeneity may not be of such great empirical consequence especially when willingness to pay (WTP) measures are of 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 coefficient (i.e. making scale free the estimation). Although this is not always generally applicable[1], most CM applications in environmental economics 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 multiple attribute levels. When the analyst is interested in WTP measures, the more parsimonious model approach which considers preference heterogeneity alone can be adequate.

In this context, little attention has been paid to the analysis of the effects on welfare estimates of mistaken assumptions about the nature of this preference heterogeneity. Surprisingly, in a context of uncertain knowledge about the *true* type of preference heterogeneity where the analyst can choose among many models to account for it, little discussion has been in the literature about the implications for the accuracy and efficiency of welfare estimates of using the incorrect model.

Indeed, the interest in analyzing the bias and variance of welfare estimates has been mainly centred on investigating, through MC analysis, issues such as the specification

[1] Flynn et al. (2010) point out that it is not always possible such normalization, as for instance in the medical field where often there is not a monetary attribute. 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.

of i) the recreation demand function in travel cost (TC) models, ii) the WTP elicitation in the contingent valuation (CV) approach, and iii) the type of experimental design under different utility specifications in CM. Early studies concerned with the factors affecting welfare measurement dealt with the implications of different approaches to TC modelling such as different site recreation demand models (Kling, 1987) and functional forms for the demand function (Kling, 1988; Kling, 1989). Likewise, Adamowicz et al (1989) focus on effects on the variance of welfare estimates comparing consumer surplus functions for linear, semilog, log-log and restricted Box-Cox forms. Investigations of the effects of decisions over appropriate nesting structures in multiple site recreation demand models represent a related area of concern (Kling & Thomson, 1996; Herriges & Kling, 1997) which makes use of MC analysis.

In the CV field, papers using a MC approach have focussed on the precision and efficiency of welfare estimates, and have mostly dealt with the advantages of combining TC and CV data (Kling, 1997), and the efficiency gains from using i) alternative models such as a double-bounded discrete choice model relative to a bivariate probit model (Alberini, 1995) and ii) different elicitation formats and bid designs (Scarpa & Bateman, 2000). In CM, the main concern of analysts using MC methods has been directed towards examining the implications for welfare measurement of different experimental design strategies (Carlsson & Martinsson, 2003; Lusk & Norwood, 2005; Ferrini & Scarpa, 2007; Scarpa & Rose, 2008).

In the light of this review, it is easy to see that the question of how important the nature of preference heterogeneity is for welfare measurement in CM has received little attention in the MC literature. To our knowledge, only Torres et al. (In press) have attempted to examine the errors from mistaking the way of explaining heterogeneity in CM. 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. In our paper, we want to contribute to this issue by examining the errors from mistaken empirical approaches to account for 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 choice experiments (CE).

3. Designing MC experiments to examine the implications of mistaken assumptions about the nature of taste heterogeneity

3.1 The experimental design

The attribute data employed to create the experimental design used in this paper come from a CE study of recreational beach use in Santa Ponça Bay, a small Mallorcan tourism area.[2] We consider three non-monetary 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 design was generated under a D-efficiency criterion[3] allowing for main effects (ME) only. According to Louviere et al. (2000), a ME only design typically explains about 70-90% of the variance in choice. The final design consisted of 36 pairs of attribute combinations. These were blocked

[2] For a description of the CE study, see Torres et al. (2009).

[3] D-efficiency is a common measure of design efficiency representing a function of the geometric mean of the Eigen values of $(X'X)^{-1}$. It is formally given by $100 \cdot 1/N\left|(X'X)^{-1}\right|^{1/A}$, where *X* is the matrix of attributes used in the design, *N* is the number of observations in a design and *A* is the number of attribute x levels in the design (Lusk & Norwood, 2005).

into different versions each of 6 choice sets of 2 alternatives plus a business-as-usual (BAU) option. The main features of the design are shown in Table 1.[4]

Table 1. Features of the experimental design

Experimental design factors		Design
Attribute levels	X_1	2 4 6*
	X_2	3 6 8*
	X_3	0.3 1* 2
	X_4	3 10.5 24 (0*)
Alternatives		2+BAU
Choice sets per individual		6
Blocks		6
Block replications		40
Total observations[a]		1,440

*Starred numbers correspond to the levels for the BAU option.
[a]Total observations are the number of choice sets x the number of blocks x the number of block replications.

3.2 Underlying taste heterogeneity and true compensating surplus

In CEs, preference heterogeneity in the systematic part of utility has been commonly understood on the basis of both RPL and LC models. Thus, at the first stage of the MC analysis, two underlying generic utility functions with the same explanatory variables (X_1, X_2, X_3 and X_4) have been considered for each type of true taste heterogeneity scenario. Differences in preferences across individuals have only been assumed for X_1 and X_2. For a RPL heterogeneity context, each individual has been assigned their own parameters for X_1 and X_2, which represent mean attribute weights plus person-specific deviations from that means, as shown in Equation (1):

[4] The number of pair combinations (36) used was the result from an application of SAS design software for a ME effects design maximizing D-efficiency. Given the BAU levels have been considered constant across the choice sets, only pair combinations have been optimized when creating the design. The BAU alternative has been added to the generated choice sets after the optimization process. However, the BAU levels (except €0 level for the *Cost* attribute) have not been for the exclusive use of the BAU option. Therefore, they have also been employed to generate the optimized pair combinations, this leading to a 3^4 experimental design for each of the two alternatives.

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

where U_{ji} is the indirect utility of alternative *j* for individual *i*, 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_{ji} is the error term associated with alternative *j* and individual *i*.

For a true LC preference scenario, 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.[5] 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.

$$P_{i|s} = \frac{\exp(\lambda_{1s} Z_{1i} + \lambda_{2s} Z_{2i} + \lambda_{3s} Z_{3i} + \xi_{is})}{\sum_{s=1}^{2} \exp(\lambda_{1s} Z_{1i} + \lambda_{2s} Z_{2i} + \lambda_{3s} Z_{3i} + \xi_{is})} \quad (2)$$

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 the error term associated to individual *i* and segment *s*.

$$U_{jis} = \alpha_s X_{1j} + \beta_s X_{2j} + \gamma X_{3j} + \omega X_{4j} + \varepsilon_{jis} \quad (3)$$

[5] The covariates considered to construct the two segments consist of two continuous variables, namely Age and Education, and one dummy variable indicating if the individual belongs or not to some environmental organization.

where U_{jis} is the indirect utility of alternative *j* for individual *i* 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_{jis} is the error term associated with alternative *j*, individual *i* and segment *s*.

Following Hanemann (1984), the true compensating surplus (CS) at the individual level, defined as the WTP for a change in the attributes from the BAU scenario, has been calculated for the RPL and LC heterogeneity scenarios as shown in Equations (4) and (5), respectively:

$$CS_i = -\frac{1}{\omega}\left(\alpha_i \Delta X_1 + \beta_i \Delta X_2 + \gamma \Delta X_3\right) \qquad (4)$$

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

$$P_{i|s'} = 1 - P_{i|s}$$
$$CS_{class} = -\frac{1}{\omega}\left(\alpha_{class}\Delta X_1 + \beta_{class}\Delta X_2 + \gamma \Delta X_3\right),\ class = s, s'$$

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 and the true CS values for a hypothetical change in X_1, X_2 and X_3 from the BAU level (a value of 6, 8 and 1, respectively, as shown in

Table 1) to a situation in which they take the levels 2, 6 and 2, respectively, indicating a reduction in both water pollution and congestion level at the beach.

Table 2. Known parameters and true CS

	Taste heterogeneity scenarios		
		LC-2 seg[a]	
Parameters	**RPL-Log[a]**	**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		0.2	2.3
λ_2		1.7	0.2
λ_3		1.5	0.05
True CS[b]	**15.45**	**16.41**	

[a] RPL-Log represent 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 0.175 variance for β. LC-2 seg represent the LC preference scenario where two latent segments exist in the population.
[b] The true CS has been obtained by averaging the individual CS values over all the simulated individuals.

3.3 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 scenario 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.[6]

[6] Note that for the LC preference scenario, two utility levels have been calculated for each alternative in each choice set: one by using segment 1 attribute parameters and another one by using segment 2 attribute parameters. The utility of each alternative for each choice occasion is then obtained by weighting the segment-specific alternative utility level by the individual probability of membership in each segment. Membership probability is derived from combining the known segment-specific parameters for the covariates with the individual covariate levels and a segment-specific Gumbel-distributed error term.

This procedure has generated 2 sets of simulated choices (one for each type of true taste heterogeneity). For each choice task, the simulated choice has been assigned to that alternative in the choice set providing the highest utility level. In the simulation, 240 individuals have been considered. Given each individual faces 6 choice tasks, 1,440 (240x6) observations have been created by this process for each of the 2 data generating processes (DGP), corresponding to the two underlying true forms of preference heterogeneity. Then, using these simulated samples, RPL and LC models have been estimated, and welfare estimates calculated in the usual way.

The errors in welfare measurement from mistaken assumptions about the nature of taste heterogeneity have then been calculated for different scenarios. First, a scenario 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 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).[7] Third, and to examine the implications of assuming a parameter distribution other than the lognormal one, a context in which RPL models assuming triangular-distributed parameters have been estimated under the two types of DGPs stated above (LC and RPL (log-normal)).[8] An additional analysis assuming

[7] Under both scenarios of true preferences, LC models considering 3 segments have also been estimated to analyze the contribution to the errors in welfare measurement of a number of classes other than 2. However, assuming 3 segments has led the LC models to stop after a specific number of MC replications. So, only results from LC models with 2 segments have been reported in the paper.

[8] Like the lognormal distribution, the triangular distribution can be constrained to have the same sign for the parameter of interest.

preference homogeneity for X_1 and X_2 (i.e. erroneously estimating a MNL model) has also been undertaken.

Taking into account the 2 types of DGPs and the before-mentioned scenarios, 12 different MC experiments (2x6) have thus been undertaken.[9] 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, mean proportional errors (MPE), or relative biases, and mean squared proportional errors (MSPE) have been calculated at the individual level at each MC replication.[10] MPE or relative bias is defined as the ratio between bias (difference between the estimated and true CS) and the true CS and indicates the degree of accuracy of welfare estimates, whilst MSPE represents the square of the MPE and gives an idea of their efficiency (i.e. variance), as seen in Equations (6) and (7).

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

[9] Although not being the focus of the paper, the analysis of the errors in welfare measurement from estimating both RPL and LC models under true homogeneous preferences has also been undertaken. Thus, choices have also been simulated under a DGP following a MNL scheme. Whilst for the RPL model the resulting errors in welfare estimates have been very small –given the MNL model is a particular case of the RPL one–, the LC models with 2 and 3 segments have not converged. These results are available from the authors upon request.

[10] Although bias and mean squared error (MSE) values have also been calculated, only the values of the relative measures are reported in the paper as these are independent of the magnitude of the true CS, thus making comparable the results from all the MC experiments. The values of bias and MSE are available from the authors upon request.

$$MSPE_r = \frac{1}{I}\left[\sum_{i=1}^{I}\left((CS_{ir}^{e} - CS_i^t)/CS_i^t\right)^2\right] \qquad (7)$$

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*.

As seen, at each MC repetition both MPE and MSPE have been defined as the average over 240 individual accuracy measure values. After 1,000 MC repetitions, a distribution of MPE and MSPE mean values for the change in X_1, X_2 and X_3 has been obtained for each experiment. The values for MPE and MSPE reported for each MC experiment have been calculated as the average of the sum of the mean values obtained in each MC replication over the 1,000 repetitions.

4. Results

The results of MPE and MSPE in the estimated CS for each MC experiment are reported in Table 3. As stated above, these values refer to a hypothetical change in X_1, X_2 and X_3 from the BAU levels of $X_1 = 6$, $X_2 = 8$ and $X_3 = 1$ to the levels of $X_1 = 2$, $X_2 = 6$ and $X_3 = 2$. MPE and 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.[11]

[11] Note that, for comparability reasons, results from using the correct model to account for taste heterogeneity (i.e. estimating a RPL model assuming lognormally-distributed parameters under true RPL preferences and estimating a LC with 2 segments under true LC preferences) have also been reported.

Table 3. MPE and MSPE in the estimated value of a hypothetical change in the attributes (over 1,000 repetitions)

True DGP	Estimation model[a]	MPE	MSPE
RPL-Log	**RPL-Log**	0.0145	0.0356
	RPL-Triang	0.0137	0.0369
	LC-2seg	0.0336	0.0558
	MNL	0.2096	0.1338
LC-2seg	**LC-2seg**	0.0420	0.0266
	RPL-Log	0.0766	0.1048
	RPL-Triang	0.1051	0.0616
	MNL	0.2722	0.1635

[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.

As seen in Table 3, when true preferences for X_1 and X_2 randomly vary across individuals according to a lognormal distribution, mistaking parameter distributional assumptions is not a relevant issue. Indeed, if the analyst assumes the correct model to account for taste heterogeneity by estimating an RPL model, assuming either lognormally-distributed or triangular-distributed parameters leads to similar values of the MPE (0.0145 vs. 0.0137) and MSPE (0.0356 vs. 0.0369). In contrast, erroneously assuming LC taste heterogeneity diminishes the accuracy of the estimates (0.0336 vs. 0.0145) and worsens their efficiency (0.0558 vs. 0.0356). As expected, the major errors in welfare measurement come from erroneously assuming homogeneous preferences for all the attributes (estimating a MNL model).

If preference heterogeneity in the systematic part of utility is driven by the existence of two latent segments in the population, using the incorrect model to account for taste heterogeneity not only leads again to more biased and less efficient estimates but also makes parameter distributional assumptions more important. Indeed, assuming the parameters for X_1 and X_2 are lognormally-distributed leads to more accurate but less efficient estimates than considering they vary according to a triangular distribution (0.0766 vs. 0.1051, and 0.1048 vs. 0.0616, respectively). Again, the worst scenario comes when the analyst think preferences are homogeneous.

Interestingly, when comparing MPEs from mistaken assumptions with those from the correct models for both types of true DGPs, one observes the relevance of the nature of taste heterogeneity in terms of accuracy relies on parameter distributional assumptions. Indeed, MPE from erroneously estimating a LC model increases by a factor of 2.31 (0.0336 vs. 0.0145), whilst MPE from erroneously estimating a RPL increases by a factor of 1.82 under lognormally-distributed parameters (0.0766 vs. 0.0420) and by a factor of 2.50 under triangular-distributed parameters (0.1051 vs. 0.0420). Thus, it seems to be less risky to erroneously estimate a RPL than to erroneously estimate a LC if RPL parameters are assumed to be lognormally-distributed. In terms of accuracy, and taking into account true preferences are unknown, results seems to suggest that opting for an RPL-Log normal specification is a good option. However, if looking at the MSPE, the LC model should be preferred (i.e. the increases in MSPE from using the incorrect model are higher in a context of true LC preferences: 3.94 for RPL-Log and 2.31 for RPL-Triang vs. 1.57 for LC-2seg).

To analyze the sensitivity of results to the magnitude of welfare change, the MC experiments have been repeated for a hypothetical change in X_1, X_2 and X_3 from the BAU levels (6 , 8 and 1 respectively, see Table 1) to the levels of 4, 6 and 2, respectively. Results are reported in Table 4.

Table 4. MPE and MSPE in the estimated value of a hypothetical smaller change in the attributes (over 1,000 repetitions)

True DGP	Estimation model	MPE	MSPE
RPL-Log	**RPL-Log**	0.0166	0.0425
	RPL-Triang	0.0072	0.0440
	LC-2seg	0.0386	0.0681
	MNL	0.2395	0.1652
LC-2seg	**LC-2seg**	0.0643	0.0448
	RPL-Log	0.1084	0.5432
	RPL-Triang	0.1107	0.1112
	MNL	0.3400	0.3033

Although general conclusions remain when comparing Table 4 with Table 3, results for a smaller welfare change show higher values of MPE and MSPE in almost all the MC experiments. Thus, a smaller welfare change makes more critical the issue of using the correct model to account for taste heterogeneity. Interestingly, in contrast to a higher welfare change, the relevance of the nature of taste heterogeneity in accuracy terms does not depend anymore on RPL parameter distributional assumptions. Indeed, erroneously estimating a LC model increases the MPE by a factor of 2.32 with respect to using the correct model (0.0386 vs. 0.0166), whilst erroneously estimating a RPL model leads to increases of 1.69 (0.1084 vs. 0.0643) and 1.72 (0.1107 vs. 0.0643) under lognormally- and triangular-distributed parameters, respectively. In other words, a smaller welfare change reinforces the idea that using the RPL model is a good option, regardless of the parameter distributional assumptions, under uncertain knowledge about true preferences. However, as for a higher welfare change, these results also reinforce the appropriateness of the LC model in terms of efficiency (i.e. the increase in MSPE from using the incorrect model is much higher in a context of true LC preferences: 12.125 for RPL-Log and 2.48 for RPL-Triang vs. 1.60 for LC-2seg).

5. Conclusions

How to model underlying preference heterogeneity has been of growing interest to choice modellers working with both RP and SP data. This research has compared different approaches such as RPL (mixed logit), LC and Covariance Heterogeneity for modelling a given choice data set, exploring the implications for welfare measures and prediction (Provencher & Bishop, 2004; Colombo et al., 2009). Our paper complements this earlier work by investigating the relative errors from mis-specifying the model of preference heterogeneity, when the true DGP is known to the researcher. 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 underlying (money-metric) utility change from a change in environmental quality, and then compare this “true” measure with the estimated welfare change under different assumptions.

Our main findings are that (i) when the true DGP is described by a log-normal RPL, then small degrees of bias emerge whatever assumption is made about preference heterogeneity so long as the researcher does not assume that preferences are homogenous (ii) however, when the true DGP is described by latent classes, then larger errors arise from assuming an RPL with a triangular distribution: again, the largest error comes from assuming homogeneity. We also find that these results seem to be sensitive to the size of the change in attribute values; and that there are trade-offs between the two measures of mis-specification used here (MPE and MSPE). Overall, though, using an RPL specification with log-normally distributed preferences seems the best bet.

However, these results are subject to the data employed in the MC experiments, that is, to the experimental design, the functional form of the utility function, the known parameters, their distribution in the true RPL preference context, the number of true latent classes in the population and the error structure. Thus, the analysis under different assumptions from the above ones would be of interest to examine the robustness of the conclusions drawn in this paper. 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 but either by taste and scale heterogeneity, or only by the scale of the error. This would contribute to testing issues that, despite the high popularity of RPL and LC models in a context where knowledge about the true type of preference heterogeneity is uncertain, have been largely overlooked in the literature.

6. References

Adamowicz, W. L., Fletcher, J. J., & Graham-Tomasi, T. (1989). "Functional form and the statistical properties of welfare measures." *Amer. J. Agr. Econ. 71(2)*: 414-421.

Alberini, A. (1995). "Efficiency vs bias of willingness-to-pay estimates: Bivariate and interval-data models." *J. Environ. Econ. Manag. 29(2)*: 169-180.

Ben-Akiva, M., & Lerman, S. R. (1985, p.367). *Discrete choice analysis: Theory and application to travel demand*. Massachussetts, The MIT Press.

Birol, E., Karousakis, K., & 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). "Understanting heterogeneous preferences in random utility models: A latent class approach. ." *Environmental and Resource Economics 23*: 421-446.

Carlsson, F., & Martinsson, P. (2003). "Design techniques for stated preference methods in health economics." *Health Econ. 12*: 281-294.

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

Ferrini, S., & Scarpa, R. (2007). "Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study." *J. Environ. Econ. Manag. 53(3)*: 342-363.

Fiebig, D. G., Keane, M. P., Louviere, J., & Wasi, N. (2009). "The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity." *Marketing Science. Articles in Advance*: 1-29.

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

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

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

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

Herriges, J. A., & Kling, C. L. (1997). "The performance of nested logit models when welfare estimation is the goal." *Amer. J. Agr. Econ. 79(3)*: 792-802.

Hynes, S., Hanley, N., & Scarpa, R. (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.

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

Kling, C. L. (1988). "The reliability of estimates of environmental benefits from recreation demand models." *Amer. J. Agr. Econ. 70(4)*: 892-901.

Kling, C. L. (1989). "The importance of functional form in the estimation of welfare." *Western J. Agr. Econ. 14(1)*: 168-174.

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

Kling, C. L., & Thomson, C. J. (1996). "The implications of model specification for welfare estimation in nested logit models." *Amer. J. Agr. Econ. 78(1)*: 103-114.

Louviere, J., & Eagle, T. (2006). Confound it! that pesky little scale constant messes up our convenient assumptions. 2006 Sawtooth Software Conference Washington, USA, Sawtooth Software, Sequem.

Louviere, J. J., arson, R. T., Ainslie, A., Cameron, T. A., DeShazo, J. R., Hensher, D. A., Kohn, R., Marley, T., & Street, D. J. (2002). "Dissecting the random component of utility." *Marketing Letters 13*: 177-193.

Louviere, J. J., Hernsher, D. A., & Swait, J. D. (2000). *Stated choice methods: analysis and application.* Cambridge, UK, Cambridge University Press.

Louviere, J. J., Meyer, R. J., Bunch, D. S., Carson, R., Dellaert, B., Hanemann, W. M., Hensher, D. A., & Irwin, J. (1999). "Combining sources of preference data for modelling complex decision processes." *Marketing Letters 10(3)*: 205-217.

Lusk, J. L., & Norwood, F. B. (2005). "Effect of experimental design on choice-based conjoint valuation estimates." *Amer. J. Agr. Econ. 87(3)*: 771-785.

Provencher, B., & 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.

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

Scarpa, R., & Rose, J. M. (2008). "Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why." *Australian J. Agr. Resource Econ. 52(3)*: 253-282.

Torres, C. M., Hanley, N., & Riera, A. (In press). "How wrong can you be? Implications of incorrect utility function specification for welfare measurement in choice experiments." *Journal of Environmental Economics and Management*.

Torres, C. M., Riera, A., & García, D. (2009). "Are preferences for water quality different for second-home residents?" *Tourism Econ. 15(3)*: 629-651.

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