Environ Resource Econ (2010) 47:477–493
DOI 10.1007/s10640-010-9389-y

Combining Discrete and Continuous Representations of Preference Heterogeneity: A Latent Class Approach

Angel Bujosa · Antoni Riera · Robert L. Hicks

Accepted: 17 June 2010 / Published online: 6 July 2010

Abstract Unobserved preference heterogeneity has been widely recognized as a critical issue not only for modelling choice behaviour, but also for policy analysis. This paper examines alternative approaches for incorporating heterogeneity in recreational demand. We apply a hybrid model combining discrete and continuous heterogeneity representations of tastes to capture the defining features of both the latent class and the random parameter logit specifications. This model allows for the joint estimation of discrete segments and within segment heterogeneity providing a richer interpretation of preference heterogeneity. A database of recreational trips to forest sites in Mallorca has been used to compare the empirical performance of this hybrid approach with common specifications such as the conditional logit, the random parameter logit, and the latent class model.

Keywords Travel cost method · Recreation demand · Random parameter model · Latent class model · Forests

Abbreviations

CL	Conditional logit
RPL	Random parameter logit
LC	Latent class
LC-RPL	Latent class-random parameter logit
WTP	Willingness-to-pay

A. Bujosa (✉) · A. Riera
Centre de Recerca Econòmica (UIB · Sa Nostra), Ctra. Valldemossa km. 7.5,
07122 Palma de Mallorca, Spain
e-mail: angel.bujosa@uib.es

R. L. Hicks
Department of Economics, College of William and Mary, 129 Morton Hall,
Williamsburg, VA 23187, USA

1 Introduction

The study of preference heterogeneity has become an important focus in the recreation demand literature based on choice modeling (Train 1998; Morey et al. 2006). Researchers have found that when heterogeneous preferences are not properly accounted for, valuable information is discarded and inconsistent estimates and biased welfare measures are obtained (Provencher and Bishop 2004; Hynes et al. 2008).

In this context, the computing revolution of the last 20 years, and the consequent generalization of simulation methods (e.g. simulated maximum likelihood estimation), has allowed researchers to estimate models with more flexible specifications overcoming the potentially restrictive assumptions implicit with conventional specifications such as conditional or nested logit models (Morey and Rossmann 2003; Train 2003). Two predominant approaches for dealing with preference heterogeneity have been developed and are often compared in recreation demand studies: the Random Parameter Logit (RPL) and the Latent Class model (LC) (Provencher et al. 2002; Hynes et al. 2008).

The RPL provides an elegant way to accommodate preference heterogeneity using a continuous distribution for individual tastes by assuming that each member in the sample has a different set of utility parameters (Revelt and Train 1998; McFadden and Train 2000; Provencher and Moore 2006). While the appeal of handling underlying heterogeneity is an attractive feature of the RPL, the model may be inadequate in the presence of different groups of individuals with different group-specific tastes (see Lenk and DeSarbo 2000; Scarpa and Thiene 2005). In these cases other solutions such as the LC model are superior. The LC (finite mixture) approach accounts for preference heterogeneity in a different way by assuming that the sample of respondents arises from a given number of groups, sometimes referred to as classes or segments (Gupta and Chintagunta 1994; Boxall and Adamowicz 2002; Shonkwiler and Shaw 2003). Although this discrete representation of taste variation is not well designed to handle within-group heterogeneity (Andrews et al. 2002), its intuitive interpretation of variation across segments in the population has turned the LC model into a convenient tool to obtain useful information about the distribution of welfare effects associated with policy changes (Provencher et al. 2002; Greene and Hensher 2003; Ouma et al. 2007).

Unfortunately, empirical evidence shows that the use of the LC specification might oversimplify the preferences of the population, especially when a small number of classes is defined and the underlying distribution of preferences is, in fact, continuous within classes (Allenby and Rossi 1998; Wedel et al. 1999). In this situation, it is unlikely that all individuals with similar socioeconomic characteristics will have the same preferences and, hence, the assumption of within-group homogeneity is too restrictive to represent adequately the preferences of recreators (Wedel and Kamakura 2002). Underscoring this point, Allenby et al. (1998) remark that the extent of heterogeneity is much greater than that measured by LC and indicates that further research is needed to improve the modeling of taste heterogeneity.

This paper suggests a hybrid modeling approach combining discrete and continuous representations of preferences to overcome the limitations of the conventional RPL and LC models. Following the work of Lenk and DeSarbo (2000), a random distribution of taste coefficients is integrated over the segments of a LC specification. In this way, the Latent Class—Random Parameter Logit model (LC-RPL) is first applied in a recreational demand setting to account for taste heterogeneity in two ways: (1) identifying different behavioural groups as a function of socioeconomic characteristics and (2) considering taste diversity among individuals in the same group (within-group heterogeneity). The LC-RPL captures the defining features of both the LC and the RPL models becoming more parsimonious than the former and more flexible than the later. Consequently, by allowing for taste variation across individuals in the same group, this new approach provides an additional insight on preference heterogeneity and a richer interpretation of the distribution of welfare effects of policy changes across the population. Furthermore, the LC and RPL models are special cases of the more general LC-RPL model.

A database of recreational trips to forest sites in Mallorca (Spain) has been used to compare the empirical performance of this new approach with common estimation approaches in recreational demand modelling such as the conditional logit (CL), the RPL, and the LC model. Comparison of goodness-of-fit measures and in-sample forecasts across specifications will be used to provide information on model performance. Finally, the welfare effects of two policy scenarios, a quality increase and site closures, will be compared to illustrate whether Willingness-To-Pay (WTP) and its distribution across individuals significantly differs between alternative representations of taste heterogeneity in choice modelling.

The paper is structured as follows. The theoretical background of random utility model underlying the LC-RPL is developed in the next section. The data and the rationale behind preference heterogeneity amongst Mallorcan forest recreators are presented in Sect. 3. Section 4 outlines model estimates. Results of WTP for various policy scenarios are provided in Sect. 5 and, finally, some conclusions and directions for future research are presented in Sect. 6.

2 Model and Welfare Measures

Following the development of the conventional LC specification, the LC-RPL model assumes the existence of K segments or groups in the sample of N respondents, where K is exogenously defined by the analyst. The individual's utility functions can vary between these segments (Boxall and Adamowicz 2002). However, as these classes are latent, not observable by the analyst, a probabilistic equation explaining the assignment of each individual n into the K segments has to be defined.

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 (Boxall and Adamowicz 2002; Provencher et al. 2002; Hynes et al. 2008).[1] Using a multinomial logit formulation, the probability that individual n belongs to segment k can be written as a function of its socioeconomic variables z_n and the vector of estimated coefficients θ_k related to that specific segment (Bhat 1997):

$$\pi_{nk} = \frac{e^{\theta_k' z_n}}{\sum_{m=1}^{K} e^{\theta_m' z_n}} \tag{1}$$

Given the membership to group k, the LC-RPL site-choice probabilities follow the random utility framework. The individual chooses the alternative yielding the highest level of utility from a set of $i = 1, \ldots, I$ known and mutually exclusive possibilities on a given choice occasion (Ben-Akiva and Lerman 1985). Assuming that the error components of this equation are independent and identically distributed as Type I extreme value, the probability that individual n chooses alternative i, conditional on belonging to taste group k, takes the familiar logit form (Hensher and Greene 2003):

[1] We have estimated this semi-parametric form for both the LC and LC-RPL models. The results of these models indicate that the membership class probabilities were not well defined with this specification. For the sake of brevity, these results are not reported here but are available from the authors.

$$\pi_{ni|k} = \frac{e^{\beta'_{nk} x_{ni}}}{\sum_{j=1}^{I} e^{\beta'_{nk} x_{nj}}} \tag{2}$$

where x_{ni} represent the vector of attributes associated with each alternative and β_{nk} the vector of estimated coefficients. Note that, instead of assuming a fixed vector of coefficients for all subjects in segment k, a set of individual specific coefficients (β_{nk}) is used in each segment, accommodating preference heterogeneity across individuals belonging to that group (Lenk and DeSarbo 2000). Following the specification of the conventional RPL, utility coefficients vary randomly across individuals within the same segment according to a specific distribution defined by the researcher. However, as the analyst does not observe these coefficients, it is necessary to integrate the logit formula in expression (2) over all possible values of β_{nk} (Train 2003):

$$\pi_{ni|k} = \int \frac{e^{\beta'_{k} x_{ni}}}{\sum_{j=1}^{I} e^{\beta'_{k} x_{nj}}} f(\beta_k) d\beta_k \tag{3}$$

At this point, the researcher has to assume a distribution for β_{nk} and estimate the parameters of such distribution that, in most applications, has been specified as normal $\beta \sim N(b, W)$ or lognormal $\ln\beta \sim N(b, W)$ with mean b and covariance W (Train 1998, 1999; McFadden and Train 2000; Meijer and Rouwendal 2006). Finally, the unconditional probability that individual n chooses i can be written from Eqs. (1) and (3):

$$\pi_{ni} = \sum_{k=1}^{K} \pi_{nk} \pi_{ni|k} \tag{4}$$

Therefore, the log-likelihood function reduces to a weighted average of the log-likelihoods of the K latent classes:

$$LL = \sum_{n=1}^{N} \ln \left[\sum_{k=1}^{K} \pi_{nk} \left(\prod_{i=1}^{I} \left(\pi_{ni|k} \right)^{y_{ni}} \right) \right] \tag{5}$$

where π_{nk} and $\pi_{ni|k}$ are the membership and site-choice probabilities from Eqs. (1) and (3) and y_{ni} equals one when the nth individual chooses alternative i and 0 otherwise. As the solution to expression (5) involves the evaluation of a multiple-dimensional integral which does not have a closed-form, the estimation of such model requires the use of simulation methods (Bhat 1998; Revelt and Train 1998).[2]

Note that the LC-RPL with only one class is equivalent to the conventional RPL model. At the same time, if several classes are implemented assuming within-group homogeneity, the model collapses to the traditional LC specification. If only one class characterized by homogeneous preferences is considered, the LC-RPL model collapses to the CL model.

[2] Model identification has been an issue of major concern since the work of Dillon and Kumar (1994) in the context of mixture specifications. In this line, different measures have been implemented in the literature to avoid identifiability problems (see Lenk and DeSarbo 2000; Frühwirth-Schnatter et al 2004). In this paper, following Wedel and Kamakura (2000), an analysis of the Hessian matrix of second-order partial derivatives has been performed to guarantee model identifiability under the mixture specifications (LC and LC-RPL).

Willingness-To-Pay (WTP) measures based upon Small and Rosen (1981) have been defined for the CL (Hanemann 1982, 1999), the RPL (Train 1998), the LC (Boxall and Adamowicz 2002) and the LC-RPL. While the WTP measures of the four models are compared in the paper, only the equation for the LC-RPL is included. In this model, heterogeneity enters in two ways: (a) the expected WTP is conditional on individual preferences β_n, found by integrating over the estimated preference distribution of the population within a segment, and (b) the expected WTP has to be weighted by segment membership for each individual. Consequently, we can write expected willingness to pay for individual n as:

$$E(WTP_n) = \sum_{k=1}^{K} \pi_{nk} \int \left(\frac{1}{\beta_{TC}} \left[\ln \left(\sum_{i=1}^{I^0} e^{\beta_k' x_{ni}^0} \right) - \ln \left(\sum_{i=1}^{I^1} e^{\beta_k' x_{ni}^1} \right) \right] \right) f(\beta_k) d\beta_k \quad (6)$$

where β_{TC} is the travel cost coefficient associated with the marginal utility of income and the attributes before and after the policy change are denoted by x_{ni}^0 and x_{ni}^1, and the available sites pre and post-policy by I^0 and I^1, respectively.

3 Data

Revealed-preference data based on one-day recreational trips to forests in the Island of Mallorca (Spain) has been used for this study. The following sections are devoted to review this information on both, the environmental characteristics of forest sites and the recreational behaviour of Mallorcan residents. This data comes from a previously published study where the aggregate value of forest recreation is estimated in a regional context (Bujosa and Riera 2009), to which the reader can address for a more detailed illustration of the data.

3.1 Site Attribute Data

A geographical information system and ArcGIS 9.2 software have been used to provide data on the environmental characteristics and recreational facilities characterizing recreational sites in Mallorca. Data has been collected from fieldwork inventories and existing data sets (e.g., the Balearic Islands topographic map from the Balearic Island Government, the National Institute of Meteorology and the National Forest Inventory collected by the Ministry of Environment) covering the 153,115 hectares of the study area. Overall, the data set on Mallorcan forests includes attributes describing forest composition (broad-leaved, mixed, etc.), land use of the site and its surroundings (agricultural and urban areas), land fragmentation, visibility, recreational facilities (playgrounds, parking, picnic areas, etc.) and public infrastructures (roads, reservoirs, sewage treatment plants, etc.).

There is no doubt that site characteristics, specially those attributes related to recreational facilities such as picnic areas or hiking trails, determine to a great extent the type of activities visitors can practice in any site. In this sense, it is important to note that, thanks to their environmental and recreational diversity, Mallorcan forests constitute an ideal environment for investigating a wide range of leisure activities such as hiking, picnicking, going for a walk, camping, adventure sports (biking, climbing, etc.) and observing the flora and fauna. In order to consider all recreational alternatives in the island, highly disaggregated site definitions have been used identifying up to 59 forest sites. Table 1 presents some descriptive statistics characterizing the environmental and recreational attributes of these 59 sites.

Table 1 Site characteristics

Variable	Description	Mean	SD	Min	Max
Size	Size of the site in hectares	594.01	827.73	0.18	3791.99
Picnic facilities	Dummy variable indicating whether there are picnic facilities available at the site	0.36	0.48	0.00	1.00
Hiking trails	Kilometres of marked trails for hiking	16.32	19.87	0.00	80.72
Kilometres of roads	Kilometres of roads accessible to cars in the site	1.56	2.12	0.00	9.92
Distance to coast	Euclidean distance from the site to the nearest coastline	5.80	5.55	0.03	22.02
Farming area	Size in hectares of farming area	16.38	85.01	0.00	626.30
Broad-leaved forests	Size in hectares of broad-leaved forests	41.65	113.59	0.00	680.29
Mixed forests	Size in hectares of mixed forests	38.24	126.07	0.00	895.32
Arboreal cover	Percentage of forest covered with trees	0.48	0.20	0.00	0.85

Source: Own elaboration

3.2 Respondents Data

For data on recreation behavior, we use a regional-wide survey randomly administrated to 1,043 Mallorcan residents surveyed at their homes by trained interviewers. Only those respondents who had taken one or more trips to forests in the 12 months previous to the survey, 841 individuals, were included in the sample of this application. In this way, forest recreation trips are characterized by visitors with distinct socioeconomic profiles and types of forest recreation activity providing for an excellent setting for investigating preference heterogeneity and to examine whether diversity of tastes for on-site attributes may lead to different WTP for various forest policies.

After testing the questionnaire in a pilot survey, the final version was administrated from April to July 2006 with a response rate slightly over 60%. It was divided in different sections collecting data regarding number of trips, visited sites, activities undertaken in the site (hiking, picnicking, going for a walk, camping, observing the flora and the fauna, adventure sports as biking, climbing, etc.) and socioeconomic information about the respondent (income, age, place and year of birth, attained level of studies, occupation, etc.).

To compute the travel cost variable from each trip origin to the 59 available sites, data on means of transport, party size, on-site time and other costs associated with the visit was also gathered. In addition, travel time and distance have been calculated from the Mallorcan road map at scale 1:25,000 and Teleatlas digital data. When more than one route was available for a specific individual, it has been assumed that the shortest one was chosen. The mileage cost and the opportunity cost of driving time have been jointly considered to estimate the travel cost.[3]

Overall, visitors took an average of 10 trips each and going for a walk was the most popular activity in forests (40.90%), followed by hiking (24.85%), picnicking (22.95%), adventure sports (6.66%) and other activities (4.64%). The mean age in the sample was 42 and the average monthly income was 985 euros. 73.49% of sampled residents were born in Mallorca and 16.88% in Mainland Spain. While 29.72% had completed primary studies, 39.36% second-

[3] The mileage cost has been set to €0.19 per kilometre according to the official cost per kilometre dictated by the Spanish Government in 2005. For opportunity cost of time, we use the lower bound often used in literature, consisting of one-third of the individuals wage (Englin and Shonkwiler 1995; Phaneuf and Smith 2005).

Table 2 Visitors descriptive statistics

Variable	Description	Mean	SD
Income	Household monthly income (in euros)	985.43	636.86
Age	Age (years)	41.95	14.11
Gender	Gender (1 = male, 0 = female)	0.49	0.50
Birthplace	Mallorca	0.73	0.44
	Mainland Spain	0.17	0.37
	European country	0.03	0.18
	Non-European country	0.06	0.24
Completed studies	Primary studies	0.30	0.46
	Secondary studies	0.39	0.49
	Tertiary studies	0.31	0.46
Employment status	Employed	0.68	0.47
	Unemployed	0.04	0.20
	Housewife or househusband	0.10	0.30
	Retired	0.10	0.30
	Student	0.07	0.26
Resident in Palma	= 1 if the individual lives in Palma (the capital of the province)	0.40	0.49
Natural areas	= 1 if the individual considers adequate the natural areas available	0.35	0.48
Trips	Number of trips in the prior 12 months	10.86	17.34

Source: Own elaboration

ary studies and 30.92% tertiary studies. As for employment status, 68.13% were employed, 4.28% were unemployed, 9.87% were housewife or househusband, 10.23% were retired and, lastly, 7.49% were students. For a more detailed description of visitors characteristics see Table 2.

4 Results

The econometric results for the LC-RPL model defined in section two are presented in Table 3. For comparison, the CL, the RPL and the LC models have also been estimated and included in Table 3 as benchmark cases. Equation (5) has been estimated by simulated maximum likelihood (or maximum likelihood) using MATLAB. For simulated log-likelihood functions we use 1,000 quasi-random Halton draws.[4]

For model specification, we consider travel cost and a large set of environmental attributes and recreational facilities characterizing forest sites (see Table 1 for more details). While alternative specifications with more site-specific attributes were estimated, the final model includes only the key attributes being the most significant determinants of choice overcoming significant collinearity issues. The site-specific variables are 'travel cost', the availability of 'picnic facilities' in the site, the 'kilometres of roads' in the site (used as a

[4] Our code benefitted greatly from MATLAB code from the Workshop 'Revealed Preferences Outside Markets: Micro-econometrics in Environmental Economics' organized by Dan Phaneuf and Kerry Smith. A sensitivity analysis for starting values has been performed to examine the convergence properties of the models. No convergence problems have been found for the models presented in the paper.

Table 3 Coefficients estimates (t-statistics in parentheses)

Variable	CL	RPL	LC	LC-RPL
Site-choice equation (class 1)				
Travel cost	−0.2234	−0.2303	−0.8667	−0.9292
	(−20.40)	(−20.46)	(−9.06)	(−8.70)
Picnic facilities	0.2432	0.2411	0.4435	0.4964
	(3.24)	(3.18)	(2.61)	(2.73)
Kilometres of roads	0.1559	0.1657	0.3247	0.3581
	(10.21)	(10.57)	(8.30)	(7.60)
Distance to coast	−0.3471	−0.3535	0.0220+	0.0796+
	(−4.91)	(−5.03)	(0.15)	(0.51)
Arboreal cover (mean)	0.7742	1.0554	2.0055	2.3497
	(3.89)	(4.31)	(3.50)	(3.08)
Arboreal cover (SD)	–	2.9255	–	3.2254*
		(6.14)		(2.10)
Site-choice equation (class 2)				
Travel cost			−0.0754	−0.0850
			(−4.50)	(−5.00)
Picnic facilities			0.1579+	0.1316+
			(1.40)	(1.16)
Kilometres of roads			0.0604	0.0680
			(2.70)	(2.98)
Distance to coast			−0.6887	−0.7089
			(−5.52)	(−5.76)
Arboreal cover (mean)			0.5193**	0.8291*
			(1.82)	(2.38)
Arboreal cover (SD)			–	3.2655
				(5.07)
Membership equation (class 1)				
Intercept			1.6080	1.5326
			(3.97)	(3.90)
Income			−0.0616	−0.0614
			(−3.28)	(−3.33)
City			−1.9575	−1.9185
			(−6.67)	(−6.62)
High-education			−1.0427	−1.0420
			(−3.70)	(−3.71)
Natural areas			0.4662**	0.4673**
			(1.83)	(1.84)
Averaged membership probability class 1			38.88	37.87
Averaged membership probability class 2			61.12	62.13
Log-likelihood function	−3140.0525	−3133.2525	−2974.3673	−2965.6274
Restricted log-likelihood	−3429.2090	−3429.2090	−3429.2090	−3429.2090
McFadden-R^2	0.0843	0.0863	0.1326	0.1352

Table 3 continued

Variable	CL	RPL	LC	LC-RPL
Adjusted McFadden-R^2	0.0829	0.0846	0.1283	0.1302
In-sample forecasts mean-squared error	0.0002698	0.0002575	0.0002027	0.0001910

All estimated coefficients are statistically significant at a 1% level except those denoted by * and ** which are significant at 5 and 10% level, respectively. Non-significant coefficients are denoted by $^+$
Source : Own elaboration

proxy measure of accessibility inside the forest area), the 'distance to coast' and an 'arboreal cover' index.

The signs and magnitudes of coefficients conform to expectations and, in general, their interpretation across models is similar. While 'travel cost' and the 'distance to coast' variable have a negative effect on the probability of site-choice as shown by the negative sign of their coefficients, the availability of 'picnic facilities', the 'kilometres of roads' in the site and the 'arboreal cover' are desirable characteristics for recreationists and, hence, increase site-choice probability ceteris paribus.

For the RPL and the LC-RPL specifications, alternative distributions (normal, lognormal, uniform and triangular) have been investigated for the random parameter coefficients. However, for our data, only the introduction of 'arboreal cover' as a random parameter significantly improves the model fit.[5] The values related to the 'arboreal cover' variable provided in Table 3 correspond to the estimated mean and standard deviation parameters of the normal distribution associated to this variable. The rationale behind allowing for randomness in the arboreal cover coefficient is motivated by the varying attractiveness of this attribute depending on the recreational interests of individuals (i.e. the recreational activity that they undertake in the forest) and their environmental attitudes. Overall, the coefficients estimated from the RPL specification are similar to those of the CL model.

Although the RPL model allows the identification of heterogeneity in preferences for 'landscape quality', this specification does not provide any information on the source of such diversity of tastes.[6] In contrast, the LC model allows for the identification of different behavioral groups within the sample of respondents. A sequential estimation and comparison of models with an alternative number of classes has been carried out to identify the optimal number of segments with heterogeneous preferences in the sample. However, as the conventional specification tests such as the likelihood ratio or the Wald tests do not satisfy the regularity conditions for a limiting chi-square distribution under the null within this context, the choice of the number of classes has to be based on some information criteria statistics

[5] As noted by one reviewer, the use of a fixed price coefficient can become a strong assumption leading to a constant marginal utility of income and a fixed scale parameter. However, in spite of its unrealistic interpretation, the use of a fixed price coefficient is still a customary practice among researchers because it avoids a number of severe problems associated with specifying a random travel cost (Colombo et al. 2007; Olsen 2009). Furthermore, the implementation of alternative modelling approaches intended to overcome the difficulties associated with the use of a fixed price coefficient (e.g. the WTP space estimation) is still yielding mixed results and needs further attention (Balcombe et al. 2009). The investigation of these issues is, in any case, beyond the purpose of this paper.

[6] As correctly pointed by one anonymous reviewer, the random parameter model can identify the sources of preference heterogeneity by specifying the parameters of the random variables as a function of individual-specific covariates. However, although this approach does have the benefits of not having to worry with the problem of finding the optimal number of classes, to be completely analogous to the LC-RPL specification it would require a significant number of parameters to be implemented leading to an approach with much fewer degrees of freedom than the LC-RPL.

developed by Hurvich and Tsai (1989).[7] Similarly, the entropy index[8] have commonly been used to provide some guidance to the analyst to decide the number of segments in the model (Thacher et al. 2005). However, it is important to note that the choice of number of classes must also consider other issues as the significance of the estimated parameters, the interpretability of the model (in terms of the signs and magnitudes of the parameters) and the prior information regarding existent groups of recreationists (Scarpa and Thiene 2005; Morey et al. 2006; Ruto et al. 2008).

The statistical analysis of our data shows as recreationists split into two groups with a clear differentiated behavioral profile. If more than two classes are included in the model, the additional groups represent only a small portion of the total respondents and the lack of significance of their parameters (based on the evaluation of *t*-ratios) precludes their association to a specific behaviour. In addition, the optimization process of the simulated log-likelihood function of the LC-RPL with more than two classes for our data, quite often, has failed to converge. For all these reasons and to facilitate comparison between models, only two classes have been included in the LC and LC-RPL specifications.

In looking across all models, some general results emerge. First, recreationists prefer sites that are less costly to access, with better amenities, closer to coast and with a higher arboreal cover regardless of model or tier.[9]

From the results of the LC model, class 1 shows a high sensitivity to travel expenses, the presence of roads and the level of arboreal cover as compared with class 2. Furthermore, the presence of picnic facilities is an attractive feature for individuals in class 1, but not for class 2. Conversely, larger distances to coast are undesirable for respondents in class 2, while people in class 1 do not exhibit any preference for this attribute. In sum, the pattern of tastes of class 1 is associated to those individuals looking for forest areas close to home equipped with recreational facilities to undertake recreational activities such as picnicking usually carried out at heavily frequented natural areas with a considerable degree of arboreal cover. In contrast, class 2 is representative of those individuals with a more naturalistic attitude, looking for a direct contact with the natural beauty of forests and undertaking recreational activities such as hiking, going for a walk, observing the flora and fauna, etc. Accordingly, such individuals prioritize landscape quality to accessibility and do not care about recreational facilities relative to class 1. Those individuals prefer sites close to coast characterized by a larger diversity of landscapes and with lower degrees of arboreal cover.

The estimated coefficients of the membership equation are also reported in Table 3 providing information about the sources of taste heterogeneity across both segments. The membership coefficients for the second group have been normalized to zero to be able to identify the remaining coefficients of the model and, hence, the membership equation for class one

[7] Given the log-likelihood of the model at convergence LL, the number of parameters included in the specification J and a penalty constant δ, the information criteria statistic C is defined as $C = -2LL + J\delta$ (Scarpa et al. 2007; Hynes et al. 2008). Consequently, different statistics can be derived under different values of the penalty constant. For $\delta = 2$ the Akaike Information Criteria (AIC) is obtained, $\delta = \ln(N)$ the Bayesian Information Criteria (BIC) is derived, and for $\delta = 2 + 2(J+1)(J+2)/(N-J-2)$, the corrected AIC (crAIC) is obtained.

[8] The entropy index is a measure of good segregation across groups that takes the form (Wedel and Kamakura 2000; Morey et al. 2006):

$$\varepsilon = 1 - \frac{\sum_{n=1}^{N}\sum_{k=1}^{K} -\pi_{nk}\ln(\pi_{nk})}{N\ln(K)}$$

[9] There are two exceptions to this general rule: the positive but not significant signs for picnic facilities (in tier 2) and distance to coast (in tier 1).

has to be evaluated relative to group two. The final specification of the membership equation includes a constant term and variables describing the socioeconomic background of individuals. As pointed out by Bhat (1997), the constant in the membership equation do not have any substantive interpretation beyond its contribution in the probability mass assigned to each segment (see also Scarpa and Thiene 2005).

For the membership probabilities, the negative sign of the dummy variables ‘income’, ‘city’ and ‘high-education’ indicate that those individuals with higher income, a higher level of education or living in a city (instead of in a small village) are more likely to belong to group 2. These results are reasonable when compared with the behavioral profiles identified in the site-choice equation of both classes: higher income and educational level of people in the second group is consistent with the lower sensitivity of these respondents to travel costs and their higher interest for landscape quality. Similarly, the positive sign of the ‘natural areas’ dummy variable shows as respondents who considered that the provision of natural areas was acceptable are more likely to be in group 1. For the estimated LC model the membership probabilities averaged across all individuals in the sample place 38.88% weight on group 1 and the remaining 61.12% to group 2.

Finally, we turn to the coefficient estimates for the LC-RPL specification. Note that, while the LC specification restricts within group heterogeneity, the LC-RPL model allows for preference heterogeneity through the ‘arboreal cover’ random parameter. Then, although LC-RPL coefficients are similar to those of the LC model, the estimated variance for the ‘arboreal cove’ variable in both tiers suggests that, even though significant variation can be explained by socioeconomic data, an important part of the variation remains unexplained. Concerning the class membership probabilities, they are quite similar to those of the LC model with a 37.87% for group 1 and 62.13% for group 2.

Three primary conclusions can be drawn about the models presented in Table 3. First, the magnitudes and signs of the coefficient estimates do not show big differences across models. Second, based on likelihood ratio tests, the LC-RPL statistically dominates the other specifications in terms of goodness-of-fit.[10] Third, from mean-squared errors calculated for in-sample forecasts, the LC-RPL specification provides the best prediction of observed site-choice probabilities outperforming the CL, RPL, and LC models.[11]

5 Welfare Estimates

Two different scenarios, that could be plausible targets of policy actions, have been selected to analyze the effects of two hypothetical policies for illustrative purposes using welfare measures as defined in Eq. (6). First, we investigate the welfare effects of a 25% increase in arboreal cover at all sites in the choice set to illustrate how the treatment of heterogeneity might impact the willingness to pay for policies that have widespread impacts on the attractiveness for recreation of forest areas (e.g. varying forest composition using more leafy species, recovering burned forests through reforestation projects, etc.). Second, given that the impact caused by recreation demand on some forest ecosystems is significant, we analyze the welfare effects of closing the six most visited recreational sites to illustrate the case where managers may need to close heavily impacted sites to allow for natural recovery.

[10] Two likelihood ratios tests are provided, the so-called McFadden-R^2 or Pseuso-R^2 defined by McFadden (1974) and the adjusted McFadden-R^2 suggested by Ben-Akiva and Lerman (1985). We also note that based on AIC and BIC criteria, the LC-RPL is the preferred model.

[11] Mean-squared errors have been computed as the mean of the squared differences between the true and the estimated site-choice probabilities under each model specification.

Table 4 Mean, median (in parenthesis) and 95% confidence intervals (in brackets) expected WTP estimates*

Scenario	Model	Class 1	Class 2	Average
(1) Quality increase	CL	-	-	0.4518
				(*0.4493*)
			-	[0.2165 to 0.7212]
	RPL	-	-	0.9033
				(*0.8985*)
			-	[0.5405 to 1.2898]
	LC	0.3216	0.8942	0.6721
		(*0.3202*)	(*0.8477*)	(*0.6350*)
		[0.1369 to 0.5074]	[−0.0278 to 2.1799]	[0.1199 to 1.3950]
	LC-RPL	0.3986	2.4803	1.6790
		(*0.3857*)	(*2.3479*)	(*1.6081*)
		[0.1340 to 0.7131]	[0.8556 to 4.6592]	[0.6993 to 2.9413]
(2) Site closures	CL	–	–	−0.8956
				(*−0.8946*)
				[−0.9651 to −0.8296]
	RPL	–	–	−0.8521
				(*−0.8508*)
				[−0.9259 to −0.7816]
	LC	−0.8279	−1.7007	−1.3296
		(*−0.8268*)	(*−1.6051*)	(*−1.2798*)
		[−0.9188 to −0.7428]	[−2.7922 to −1.1676]	[−1.9065 to −1.0284]
	LC-RPL	−0.8382	−1.4324	−1.1758
		(*−0.8393*)	(*−1.3887*)	(*−1.1512*)
		[−0.9253 to −0.7559]	[−2.1274 to −1.0321]	[−1.5642 to −0.9411]

* WTP values expressed in euros per choice occasion
Source: Own elaboration

WTP simulated population distributions have been constructed as described in Train (1998) and von Haefen (2003) by generating 1,000 pseudo-random draws from the unconditional distribution of the estimated parameters and calculating the simulated estimates for each draw. The welfare results of both policy scenarios are summarized in Table 4. As expected, mean and median WTP (and confidence intervals) vary significantly among the estimated models and much of this variation depends on the treatment of heterogeneity. Results show that models allowing for preference heterogeneity including random parameters in their specifications (RPL and LC-RPL), in general, lead to higher mean WTP estimates when the evaluated scenario increases the well-being of individuals. For example, in the quality increase scenario, the mean WTP in the LC-RPL model (1.68 euros) is 2.5 times higher than in the LC (0.67 euros) and the mean WTP in the RPL model (0.90 euros) is 1.34 times higher than in the LC model. However, under a simulated scenario where the well-being of individuals is reduced (i.e. the site closure scenario), the use of random parameter to capture preference heterogeneity lead to similar or lower WTP measures. Thus, while the mean WTP in the LC model is −1.33 euros, it is only −0.85 euros in the RPL model and −1.18 euros in the LC-RPL.

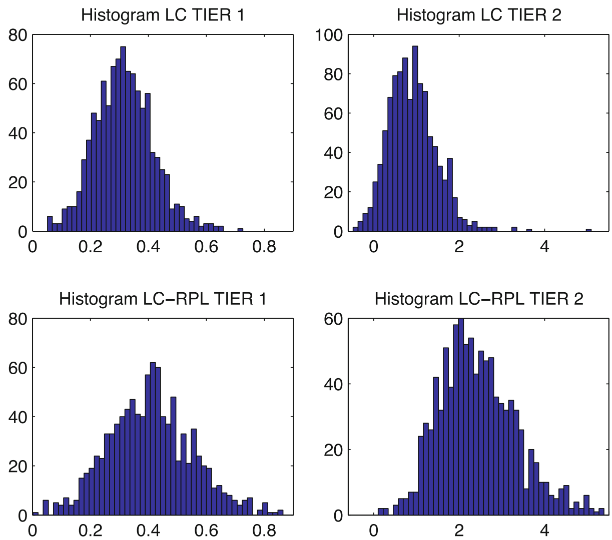

Fig. 1 Histograms of the estimated individual WTP for a 25% increase in arboreal cover

Taken together these findings show that the choice of modeling approach for heterogeneity is a key issue when policy guidance is needed for interventions that may impact attributes where heterogeneous preferences are thought to exist. To get insight into exactly how the choice of model impacts WTP, we provide histograms of the simulated distribution of WTP and compare the LC to the LC-RPL model.

Focusing first on the quality increase scenario, Fig. 1 shows how for each tier (reported in columns) the distribution of WTP differs by model (reported in rows). The distributional effects captured by the LC and LC-RPL are consistent with the findings from previous studies (Armstrong et al. 2001) where individuals with higher income are willing to pay more for site-specific attributes. The estimated distributions are quite different and these results show that allowing for heterogeneity shifts the mass of the distribution right-ward for the LC-RPL, compared to the LC model. For example, in the second tier of the LC model only a 5.1% of the mean simulated WTP lie to the right of 2 euros, while in the LC-RPL over 65% are above 2 euros.[12] Figure 2 shows the distributions for welfare measures related to the site closure policy under both models. Again, significant differences can be appreciated between tiers and models.

We have also formally tested for differences in distribution intra-model (across tiers) and intra-tier (across models) using the Wilcoxon non-parametric test. The results of testing for differences can be found in Table 5. The results show that significant differences exist when moving across tiers within the same model for either policy change, even for the case of closures. When moving beyond the mean welfare effect, the treatment of heterogeneity provides insight into how the population is distributed, and in the case of the LC-RPL information on the within-group distribution of preferences. Perhaps as importantly, particularly for the case of the LC-RPL, is the finding that for all cases, the estimated distributions with-in are statistically distinct. In the case of the quality change scenario, this is visually evident.

Beyond the statistical comparisons found in WTP between groups, the LC-RPL model confirms the hypothesis of intra-group heterogeneity by showing that the WTP is not con-

[12] Experiments with other distributions for the random coefficients show that the same patterns hold for other specifications. These results are available from the authors.

Fig. 2 Histograms of the estimated individual WTP for closure of six most visited sites

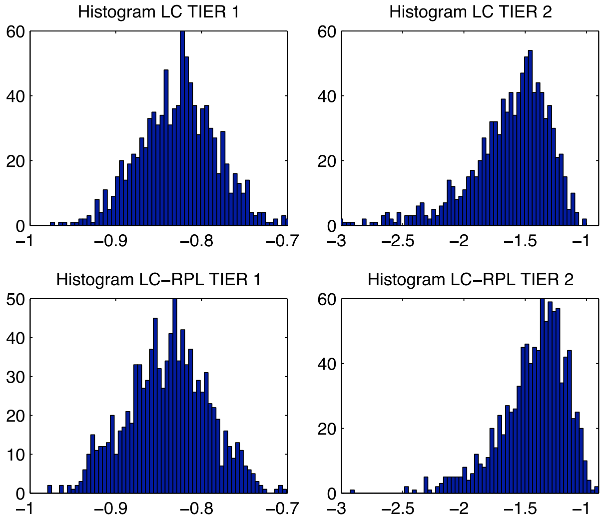

Table 5 Testing for differences in simulated WTP distributions

	Distribution comparison			
Scenario	LC_1-LC_2	LC RPL_1-LC RPL_2	LC_1-LC RPL_1	LC_2-LC RPL_2
Quality increase	Reject	Reject	Reject	Reject
Site closure	Reject	Reject	Reject	Reject

Source: Own elaboration

stant across individuals in the same group. That is, people in the same group and, hence, with similar socioeconomic characteristics, have different WTP for changes in environmental attributes. While the recreationist groups identified in the LC model is useful information from an environmental management standpoint, the LC-RPL model uncovers even more information about important differences in the preferences of forest recreators. Consequently, for this study the LC model is over-simplifying the structure of preferences of individuals and likely an underestimation of the WTP for forest amenities.

6 Conclusion

This papers examines alternative approaches for incorporating heterogeneity in models of recreational demand. The results reveal the existence of heterogeneous preferences for environmental attributes related to socioeconomic characteristics (income, education, place of residence, etc.) and two behavioural groups with different socioeconomic profiles are identified in the empirical application. Specifications accounting for preference heterogeneity demonstrate better performance for both, goodness-of-fit and in-sample forecasts. Models with discrete (LC) and discrete-continuous (LC-RPL) representations of heterogeneity, outperform those specifications based exclusively on continuous distributions of tastes (RPL). The LC-RPL model outperforms all models for goodness-of-fit and best in-sample predictions.

We investigate two policies –an island-wide quality change and the closure of the 6 most visited sites—finding that WTP measures vary significantly among the estimated models depending on the treatment of heterogeneity. Models allowing for preference heterogeneity by means of random parameters (RPL and LC-RPL) lead to higher mean WTP estimates when the evaluated scenario increases the well-being of individuals. Additionally, considerable differences have been found between the two groups of recreationists identified in both models, the LC and LC-RPL, with individuals in the second group having higher income and higher WTP being an evidence of their greater concern towards environmental issues as arboreal cover or distance to coast. In contrast, individuals in the first group are less sensitive to the policy changes investigated in this application showing a lower WTP.

Beyond merely identifying these behavioural groups, the LC-RPL has relaxed restrictions by including preference heterogeneity beyond the mean welfare effect for individuals within the same group. Such treatment of heterogeneity provides an additional insight into how the population is distributed even within the same tier and, hence, it leads to more useful and accurate WTP estimates. In fact, the results from the LC-RPL model give evidence of the heterogeneous tastes of those individuals which, sharing a similar socioeconomic profile, have been grouped in the same segment. Nevertheless, the application of our model to other study sites is needed before reaching definitive conclusions.

In sum, the LC-RPL approach developed in this paper has the potential for significantly enhancing the effectiveness of policy decisions by analysing the heterogeneous preferences of individuals in a context of recreational destination choice. Results show the potential of this hybrid model in those situations not well represented by continuous preferences where significant heterogeneity remains within discrete groups of individuals. Then, the ability of the LC-RPL model to identify these groups, based on their socioeconomic characteristics, at the same time that allows for within group taste heterogeneity towards different environmental site attributes, can provide useful information to policy-makers in different contexts. However, more simulated and empirical studies are needed to apply this method in other datasets and, in this way, to fully understand its strengths and weaknesses for estimating choice models when heterogeneity in preferences is present.

Acknowledgments The authors gratefully acknowledge the funding support from the Department of the Environment of the Balearic Islands Government (Contract No. 1211).

References

Allenby GM, Rossi PE (1998) Marketing models of consumer heterogeneity. J Econom 89(1–2):57–78

Allenby GM, Arora N, Ginter JL (1998) On the heterogeneity of demand. J Mark Res 35(3):384–389

Andrews RL, Ainslie A, Currim IS (2002) An empirical comparison of logit choice models with discrete versus continuous representations of heterogeneity. J Mark Res 39(4):479–487

Armstrong P, Garrido R, Ortzar JdD (2001) Confidence intervals to bound the value of time. Transp Res Part E Logist Transp Rev 37(2–3):143–161

Balcombe K, Chalak A, Fraser I (2009) Model selection for the mixed logit with bayesian estimation. J Environ Econ Manage 57(2):226–237

Ben-Akiva M, Lerman SR (1985) Discrete choice analysis: theory and application to travel demand. MIT Press series in transportation studies, The MIT Press, Cambridge

Bhat CR (1997) An endogenous segmentation mode choice model with an application to intercity travel. Transp Sci 31(1):34–48

Bhat CR (1998) Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. Transp Res Part A Policy Pract 32(7):495–507

Boxall PC, Adamowicz WL (2002) Understanding heterogeneous preferences in random utility models: a latent class approach. Environ Resour Econ 23(4):421–446

Bujosa A, Riera A (2009) Estimating the aggregate value of forest recreation in a regional context. J Forest Econ. doi:10.1016/j.jfe.2009.11.005
Colombo S, Calatrava-Requena J, Hanley N (2007) Testing choice experiment for benefit transfer with preference heterogeneity. Am J Agric Econ 89(1):135–151
Dillon W, Kumar A (1994) Latent structure and other mixture models in marketing: an integrative survey and overview. In: Bagozzi RPAdvanced methods in marketing research. Blackwell, Cambridge
Englin J, Shonkwiler JS (1995) Modeling recreation demand in the presence of unobservable travel costs: toward a travel price model. J Environ Econ Manag 29(3):368–377
Frühwirth-Schnatter S, Tchler R, Otter T (2004) Bayesian analysis of the heterogeneity model. J Bus Econ Stat 22(1):2–15
Greene WH, Hensher DA (2003) A latent class model for discrete choice analysis: contrasts with mixed logit. Transp Res Part B Methodol 37(8):681–698
Gupta S, Chintagunta PK (1994) On using demographic variables to determine segment membership in logit mixture models. J Mark Res 31(1):128–136
Hanemann WM (1982) Applied welfare analysis with qualitative response models. Cudare Working Papers
Hanemann WM (1999) Welfare analysis with discrete choice models. In: Herriges JA, Kling CLValuing recreation and the environment. Edward, Cheltenham
Hensher D, Greene W (2003) The mixed logit model: the state of practice. Transp 30(2):133–176
Hurvich CM, Tsai CL (1989) Regression and time series model selection in small samples. Biometrika 76(2):297–307
Hynes S, Hanley N, Scarpa R (2008) Effects on welfare measures of alternative means of accounting for preference heterogeneity in recreational demand models. Am J Agric Econ 90(4):1011–1027
Lenk P, DeSarbo W (2000) Bayesian inference for finite mixtures of generalized linear models with random effects. Psychometrika 65(1):93–119
McFadden D (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) Frontiers in econometrics. Academic Press, New York pp 105–142
McFadden D, Train KE (2000) Mixed MNL models for discrete response. J Applied Econom 15(5):447–470
Meijer E, Rouwendal J (2006) Measuring welfare effects in models with random coefficients. J Appl Econom 21(2):227–244
Morey E, Rossmann KG (2003) Using stated-preference questions to investigate variations in willingness to pay for preserving marble monuments: Classic heterogeneity, random parameters, and mixture models. J Cult Econ 27(3):215–229
Morey E, Thacher J, Breffle W (2006) Using angler characteristics and attitudinal data to identify environmental preference classes: a latent-class model. Environ Resour Econ 34(1):91–115
Olsen SB (2009) Choosing between internet and mail survey modes for choice experiments surveys considering non-market goods. Environ Resour Econ 44(4):591–610
Ouma E, Abdulai A, Drucker A (2007) Measuring heterogeneous preferences for cattle traits among cattle-keeping households in East Africa. Am J Agric Econ 89(4):1005–1019
Phaneuf DJ, Smith VK (2005) Recreation demand models. In: Mäler KG, Vincent J (eds) Handbook of environmental economics: valuing environmental changes. Elsevier, Amsterdam, pp 671–761
Provencher B, Bishop RC (2004) Does accounting for preference heterogeneity improve the forecasting of a random utility model? A case study. J Environ Econ Manag 48(1):793–810
Provencher B, Moore R (2006) A discussion of using angler characteristics and attitudinal data to identify environmental preference classes: a latent-class model. Environ Resour Econ 34(1):117–124
Provencher B, Baerenklau KA, Bishop RC (2002) A finite mixture logit model of recreational angling with serially correlated random utility. Am J Agric Econ 84(4):1066–1075
Revelt D, Train K (1998) Mixed logit with repeated choices: households choices of appliance efficiency level. Rev Econ Stat 80(4):647–657
Ruto E, Garrod G, Scarpa R (2008) Valuing animal genetic resources: a choice modeling application to indigenous cattle in Kenya. Agric Econ 38(1):89–98
Scarpa R, Thiene M (2005) Destination choice models for rock climbing in the Northeastern Alps: a latent-class approach based on intensity of preferences. Land Econ 81(3):426–444
Scarpa R, Thiene M, Tempesta T (2007) Latent class count models of total visitation demand: days out hiking in the Eastern Alps. Environ Resour Econ 38(4):447–460
Shonkwiler JS, Shaw WD (2003) A finite mixture approach to analyzing income effects in random utility models: reservoir recreation along the Columbia river. In: Hanley N, Shaw WD, Wright RE (eds) The new economics of outdoor recreation. Edward, Cheltenham
Small KA, Rosen HS (1981) Applied welfare economics with discrete choice models. Econom J Econome Soc 49(1):105–130

Thacher JA, Morey E, Craighead WE (2005) Using patient characteristics and attitudinal data to identify depression treatment preference groups: a latent-class model. Depress Anxiety 21(2):47–54

Train KE (1998) Recreation demand models with taste differences over people. Land Econ 74(2):230–239

Train KE (1999) Mixed logit models for recreation demand. In: Herriges JA, Kling CL (eds) Valuing recreation and the environment: revealed preference methods in theory and practice, New Horitzons in Environmental Economics. Edward, Cheltenham

Train KE (2003) Discrete choice methods with simulation. Cambridge University Press, Cambridge

von Haefen RH (2003) Incorporating observed choice into the construction of welfare measures from random utility models. J Environ Econ Manag 45(2):145–165

Wedel M, Kamakura WA (2000) Market segmentation: conceptual methodological foundations. Kluwer, Boston

Wedel M, Kamakura W (2002) Introduction to the special issue on market segmentation. Int J Res Mark 19(3):181–183

Wedel M, Kamakura W, Arora N, Bemmaor A, Chiang J, Elrod T, Johnson R, Lenk P, Neslin S, Poulsen CS (1999) Discrete and continuous representations of unobserved heterogeneity in choice modeling. Mark Lett 10(3):219–232