Combining discrete and continuous representations of preference heterogeneity: a latent class approach

Angel Bujosa Bestard* and Antoni Riera Font†

Centre de Recerca Econòmica (UIB ∘ Sa Nostra)
Ctra. Valldemossa km. 7.5, 07122 Palma de Mallorca, Spain

Robert L. Hicks

College of William and Mary, Department of Economics
129 Morton Hall, Williamsburg, VA 23187

March 2, 2009

Abstract

This paper investigates heterogeneity in preferences for forest recreators in Mallorca, Spain. We develop a latent class approach combining discrete and continuous representations of tastes and compare it with the conventional latent class and random parameter logit approaches. We investigate the performance of the discrete-continuous model by comparing welfare estimates and predictive accuracy. While the discrete-continuous model shows the best performance in terms of goodness-of-fit, the conventional latent class model has a higher prediction power when in-sample forecasts are calculated. The estimation of models with alternative means of accounting for preference heterogeneity confirms the existence of such variation among individuals' tastes and, hence, considerable differences across WTP estimates are found.

Keywords: Travel Cost Method, latent class model, random parameter model, recreation demand, forests.
JEL Classification: C25, Q23, Q26, Q51.

---

*Corresponding author. Tel.: +34 971 171381; fax. +34 971 172389
E-mail address: angel.bujosa@uib.es.

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

1 Introduction

Over the years, the random utility model (McFadden, 1974) has become the predominant approach for describing individual preferences through the analysis of discrete choices (Hicks and Strand, 2000; Haab and McConnell, 2002; Phaneuf and Smith, 2005). In this context, the conventional logit specifications (e.g. conditional and nested logit models) have dominated the recreational demand literature. As a result, and given the behavioural restrictions of these formulations (Hensher and Greene, 2003), preference homogeneity has been a common assumption among discrete choice applications (Ouma *et al.*, 2007; Hynes *et al.*, 2008).

However, since the publication of recreational demand studies proving the existence of taste heterogeneity across individuals (Provencher *et al.*, 2002), the investigation of such diversity of preferences and its appropriate treatment has become one of the greatest challenges explored in the choice modelling literature (Train, 1998; Provencher *et al.*, 2002; Shonkwiler and Shaw, 2003; Provencher and Bishop, 2004; 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 (Bhat, 1997; Bhat and Castelar, 2002; Hess *et al.*, 2005; Hynes *et al.*, 2008).

In this context, the computing revolution of the last 20 years, and the consequent generalisation of simulation methods (e.g. simulated maximum likelihood estimation), has provided researchers with more flexible specifications overcoming the restrictive assumptions of conventional logit models (Morey and Greer Rossmann, 2003; Train, 2003). Two predominant approaches for dealing with preference heterogeneity have been developed: the Random Parameter Logit (RPL) and the Latent Class model (LC) (Andrews *et al.*, 2002; Adamowicz, 2004; Ouma *et al.*, 2007).

On the one hand, the RPL introduces taste variation by assuming that each member in the sample has a different set of utility parameters (Revelt and Train, 1998; McFadden and Train, 2000; Phaneuf and Smith, 2005). However, although this specification provides an elegant way to accommodate preference heterogeneity using a continuous distribution for individual tastes, it is not without shortcomings (Morey and Greer Rossmann, 2003; Provencher and Moore, 2006). Following Hensher and Greene (2003), the choice of a tractable parametric form for the distribution of each random coefficient and the consequent model estimation constitute two challenges for researchers. In addition, Lenk and DeSarbo (2000) and Scarpa and Thiene (2005), point out that the RPL may be inadequate in the presence of different groups of individuals with different tastes.

On the other hand, the LC or finite mixture approach accounts for preference heterogeneity by assuming that the sample of respondents arises from a given number of groups, also called classes or segments (Gupta and Chintagunta, 1994; Boxall and Adamowicz, 2002; Shonkwiler and Shaw, 2003; Provencher and Bishop, 2004). Although this discrete representation of taste variation is not well designed to handle within-group heterogeneity (Andrews *et al.*, 2002), it provides an intuitive interpretation of variation across segments in the population, presenting attractive information about the distribution of welfare effects associated with policy changes (Milon and Scrogin, 2006; Patunru *et al.*, 2007). Unfortunately, empirical evidence proves that the use of the LC specification might over-simplify the real preferences of the population, especially when a small number of classes is defined and the underlying distribution is, in fact, continuous (Allenby and Rossi, 1998; Wedel *et al.*, 1999).

In sum, beyond the advantages and disadvantages of both specifications, the debate over continuous (RPL) versus discrete (LC) representation of preferences is still open within discrete choice literature (Wedel *et al.*, 1999; Andrews *et al.*, 2002). While RPL model has proved to be more powerful in accommodating taste variation, its limitations to identify the sources of unobserved heterogeneity has turned the LC model into the preferred approach of researchers (Provencher *et al.*, 2002; Greene and Hensher, 2003; Ouma *et al.*, 2007).[1]

However, in spite of the popularity and increasing amount of LC applications published in last years, the underlying assumption of within-group homogeneity has remained as an important aspect not adequately addressed within literature (Lenk and DeSarbo, 2000). Undoubtedly, it is unlikely to expect that all individuals with identical socioeconomic characteristics will have the same preferences and, hence, the assumption of within-group ho-

[1] The sources of taste heterogeneity are often related to socioeconomic characteristics, attitudes, perceptions, social influences and past experiences (Boxall and Adamowicz, 2002).

mogeneity is too restrictive to represent adequately the complex preferences of society (Wedel and Kamakura, 2002). In the same way, 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 modelling of taste heterogeneity.

This paper suggests a combination of 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) accounts for taste heterogeneity in two simultaneous ways by (1) identifying different behavioural groups of people as a function of socioeconomic characteristics and (2) considering taste diversity among individuals in the same group (within-group heterogeneity). So, the LC-RPL captures the best features of both the LC and the RPL models becoming more parsimonious than the former and more flexible than the later. Consequently, by expanding the consideration of taste variation across individuals in the same group, this new approach provides an additional insight on preference heterogeneity and a much richer interpretation of the distribution of welfare effects of policy changes across the population.

A database of recreational trips to forest sites in Mallorca (Spain) has been used to compare the empirical performance of this new approach against the most common alternative means of accounting for heterogeneous preferences in recreational demand modelling (RPL and LC models). Comparison of goodness-of-fit measures and in-sample forecasts across specifications will provide information on the level of improvement achieved by different models. At the same time, the welfare effects of two policy scenarios, a quality increase and the provision of new picnic facilities, will be calculated 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 as well as the more traditional RPL and LC models is presented in the next section. The data and the rationale behind preference heterogeneity amongst individuals undertaking recreational activities in Mallorca are commented in section three. Section four presents model estimates. WTP and their distribution across individuals are provided in section five and, finally, some conclusions and directions for future research are presented in section six.

2 Methodology

Following the conventional framework of the Random Utility Model (RUM), a population of N individuals chooses from $i = 1, ..., I$ known and mutually exclusive alternatives on a given choice occasion (McFadden, 1974; 1977). In this context, the total utility perceived by an individual n from choosing alternative i is assumed to be given by the indirect utility function that, following a linear-in-parameters specification, can be expressed as (Hensher and Greene, 2003; Patunru *et al.*, 2007):

$$U_{ni} = \beta' x_{ni} + \varepsilon_{ni} \tag{1}$$

This function is derived from two different components: the (observable) deterministic portion of utility that depends on the attributes x_{ni} associated with each alternative and the vector β of estimated coefficients, and the remainder (non-observable) stochastic portion ε_{ni} that is treated by the researcher as random (Train, 2003). At this point, as the individual is supposed to choose the alternative yielding the highest level of utility, the probability of observing individual n choosing alternative i becomes (Ben-Akiva and Lerman, 1985):

$$\pi_{ni} = Pr\left(\beta' x_{ni} + \varepsilon_{ni} > \beta' x_{nj} + \varepsilon_{nj}\right) \forall j \neq i \tag{2}$$

However, as the analyst cannot observe the stochastic portion of the utility, a distribution for ε_{ni} has to be specified. Following the original development of McFadden's RUM and, hence, assuming that ε_{ni} is independent and identically distributed according to the Gumbel or type I extreme value cumulative distribution function $F(\varepsilon_{ni}) = e^{-e^{-\varepsilon_{ni}}}$ the Conditional Logit (CL) model is derived and the site-selection probability in equation (2) can be expressed as (McFadden, 1978; Train, 2003):

$$\pi_{ni} = \frac{e^{\beta' x_{ni}}}{\sum_{j=1}^{I} e^{\beta' x_{nj}}} \tag{3}$$

Unfortunately, although easy to implement and estimate, site-choice probability in the CL is not affected by individual's characteristics and, hence, the model is not well suited to examine heterogeneous preferences (Milon and Scrogin, 2006). However, a mixture of the logit function evaluated at different $\beta's$ has recently been used to overcome the limitations of the traditional conditional logit model (Train, 2003). Such model, known as the mixed logit model, can be derived under a variety of different specifications according to the mixing distribution $f(\beta)$ used.

2.1 The random parameter logit model

The RPL is derived when a continuous distribution is assigned to the mixing distribution $f(\beta)$. More precisely, the RPL is a generalization of the conditional logit model that allows the coefficients of observable variables to vary randomly over people rather than being a fixed constant across the sample (Hynes *et al.*, 2008). Following this formulation, the choice probability conditional on β_n becomes:

$$\pi_{ni} = \frac{e^{\beta_n' x_{ni}}}{\sum_{j=1}^{I} e^{\beta_n' x_{nj}}} \tag{4}$$

Unfortunately, the researcher does not know β_n. Consequently, it is necessary to integrate the logit formula in expression (4) over all possible values of β_n and, in this way, estimate the unconditional choice probability for individual n visiting site i (Train, 2003):

$$\pi_{ni} = \int \frac{e^{\beta' x_{ni}}}{\sum_{j=1}^{I} e^{\beta' x_{nj}}} f(\beta) d\beta \tag{5}$$

At this point, the researcher has to assume a distribution for β_n and estimate the parameters of such distribution that, in most applications, has been specified as a normal $\beta \sim N(b, W)$ or a lognormal $ln\beta \sim N(b, W)$ with mean b and covariance W (Train, 1998, 1999; McFadden and Train, 2000; Meijer and Rouwendal, 2006). As a result, the log-likelihood function for a given value of the parameter vector β takes the form:

$$LL = \sum_{n=1}^{N} \sum_{i=1}^{I} y_{ni} ln \left(\pi_{ni}\right) \tag{6}$$

where N represents the number of individuals in the sample, π_{ni} are the choice probabilities from equation (5) and y_{ni} equals one when the nth individual chooses alternative i and 0 otherwise. As the solution to expression (6) 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).

2.2 The latent class model

In contrast to the RPL, the use of a discrete distribution for $f(\beta)$ leads to the LC specification. The LC model accounts for preference heterogeneity to a given degree by assuming the existence of K homogenous segments or groups in the sample of respondents (Bhat, 1997). In this way, although preferences are assumed to be homogeneous within each segment, tastes and, hence, utility functions and welfare measures can vary between segments (Hynes *et al.*, 2008). More precisely, the derivation of the LC model is based on two different probability equations. On the one hand, the membership equation explains the assignment of individual n into the K segments, where K is exogenously defined by the analyst. Although a semi-parametric form based only on a constant term can be used to define 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). 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 λ_k related to the segment k:

$$\pi_{nk} = \frac{e^{\lambda'_k z_n}}{\sum_{m=1}^{K} e^{\lambda'_m z_n}} \tag{7}$$

On the other hand, and given the membership to group k, the site-choice probability equation is used to explain the choice across alternatives following the traditional framework of the conditional logit model presented above. That is, the probability that individual n chooses alternative i, conditional on belonging to taste group k, takes the form:

$$\pi_{ni|k} = \frac{e^{\beta_k' x_{ni}}}{\sum_{j=1}^{I} e^{\beta_k' x_{nj}}} \tag{8}$$

where β_k represent the vector of estimated coefficients associated to segment k. Finally, the unconditional probability that a randomly chosen individual n chooses i can be written from equations (7) and (8):

$$\pi_{ni} = \sum_{k=1}^{K} \pi_{nk} \pi_{ni|k} \tag{9}$$

Therefore, following Greene and Hensher (2003), Thacher *et al.* (2005) and Aldrich *et al.* (2007), the log-likelihood function reduces to a weighted average of the log-likelihoods of the K latent classes and, hence, simulation is not needed:

$$LL = \sum_{n=1}^{N} ln \left[\sum_{k=1}^{K} \pi_{nk} \left(\prod_{i=1}^{I} \left(\pi_{ni|k} \right)^{y_{ni}} \right) \right] \tag{10}$$

2.3 The latent class - random parameter model

Finally, and following the specification of the conventional LC model, the LC-RPL assumes a discrete distribution for $f(\beta)$. Again K groups of respondents are defined and individuals are assigned to one of them on the basis of the membership equation defined in expression (7). However, instead of assuming a fixed vector of coefficient β_k 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). In this way, and following the specification of the RPL, utility coefficients vary randomly across individuals within the same segment according to a specific distribution defined by the researcher. Consequently, the site-choice equation (conditional on being in group k) of the LC-RPL model is the integral of the well-know logit formula over all values of β_{nk}:

$$\pi_{ni|k} = \int \frac{e^{\beta_k' x_{ni}}}{\sum_{j=1}^{I} e^{\beta_k' x_{nj}}} f(\beta_k) d\beta_k \tag{11}$$

At this point, the unconditional probability that a randomly chosen individual n chooses alternative i in the LC-RPL model can be written from equations (7) and (11) leading to the same equation represented in expression (9). Accordingly, equation (10) denotes the log-likelihood function of the LC-RPL which estimation, likewise the RPL, requires the use of simulation methods.

3 Data

Preference heterogeneity has been investigated using a dataset of destination choices concerning one-day recreational trips in the Island of Mallorca (Spain). As the 59 forest sites that have been identified within the 153,115 hectares of the study area are quite diverse from both a geographical and recreational perspective, a geographical information system has been assembled using ArcGIS 9.2 software. In this way, information regarding the environmental characteristics as well as the recreational facilities that provide these sites have been collected from fieldwork inventory and existing datasets such as the Balearic Islands topographic map (Balearic Island Government), the National Institute of Metereology and the National Forest Inventory (Ministry of Environment).

In addition, the regional-wide survey administrated to 1043 individuals has provided a heterogeneous sample of Mallorcan residents, not only in terms of socioeconomic characteristics, but also in terms of the kind of outdoor recreation that they were undertaking. As a result, the different direction and magnitude that some environmental features can have on utility, especially amongst individuals with a distinct socioeconomic profile and/or with different recreational interests, becomes very convenient for investigating preference heterogeneity in a context of recreational demand modelling. Undoubtedly, such diversity of tastes will lead to different WTP across recreationists depending on the activities that they undertake, their concern for environmental attributes, their demand for specific recreational facilities, their socioeconomic characteristics, etc.

The residents included in this population-specific sampling scheme were randomly chosen and surveyed at home by trained interviewers. After testing the questionnaire in a pilot survey, the final version was administrated from April to July 2006. 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.[2] Regarding the opportunity cost of driving time, the traditional lower bound proposed by recreation literature has been used, consisting of one-third of the individuals wage (Englin and Shonkwiler, 1995; Phaneuf and Smith, 2005).

In the prior 12 months previous to the survey, 80.63% of the sample (841 respondents) had taken one or more trips to forests. Additionally, those individuals who had visited forests took an average of 10 trips each. 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%). With regard to the socioeconomic information of respondents, the mean age in the sample was 44 and the average monthly income was 950 euros. 92.43% of sampled residents were Spanish and 48.21% men. Concerning the educational level of the sample, 35.57% had completed primary studies, 37.87% secondary studies and 26.56% tertiary studies. Regarding the occupation, 63.57% were employed, 4.03% were unemployed, 10.45% were housewife or househusband, 15.44% were retired and, lastly, 6.51% were students.

[2]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.

4 Results

The econometric results for the RPL, LC and LC-RPL models defined in section two are presented in Table 1. For completeness, the CL model is also provided as a benchmark case for comparison with the models accounting for heterogeneity in preferences. MATLAB software has been used to maximize the log-likelihood function, for the CL and LC models, and the simulated log-likelihood function with 200 replications per observation, for the RPL and LC-RPL specifications.[3]

Beyond the consideration of travel costs, a large set of environmental attributes and recreational facilities characterizing forest sites has been used as explanatory variables for choice probabilities. Four variables to capture environmental attributes of sites in the final specifications have been chosen. 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 collinearity issues. The site-specific variables are the 'travel cost', the availability of 'picnic facilities' in the site, the 'kilometres of roads' in the site (used as a proxy measure of accessibility inside the forest area) and a 'landscape quality' index that brings into consideration the biotic (forest composition, arboreal cover classification, burned forest areas, etc.) and anthropogenic elements (infrastructures, urban areas, farms, etc.) present in the site and its surroundings.

The signs and magnitudes of coefficients conform to expectations and, in general, their interpretation across models is similar. In this way, while 'travel cost' has a negative effect on the probability of site-choice as shown by the negative sign of its coefficient, the availability of 'picnic facilities', the 'kilometres of roads' in the site and the 'landscape quality' are desirable characteristics for recreationists and, hence, increase their site-choice probability.

Concerning the estimation search for models accounting for random taste parameters, that is, the RPL and the LC-RPL specifications, alternative distributions (normal, lognormal, etc.) have been investigated for the ran-

[3] A sensitivity analysis for starting values has been performed to guarantee the convergence of the models to a global optimum. Only when the starting values have been set sufficiently far away from the solution (and with opposed signs) the model has failed to converge. Such converge problems have been found in both, the RPL and the LC-RPL models.

Table 1: Coefficients estimates (standard errors in parentheses)

Variable	CL		RPL		LC		LC-RPL	
Site-choice equation (class 1)								
Travel cost	-0.2309	(0.0113)	-0.2362	(0.0118)	-0.8237	(0.0859)	-0.8287	(0.0860)
Picnic facilities	0.2983	(0.0783)	0.2952	(0.0783)	0.4329$^{(*)}$	(0.1681)	0.4745	(0.1673)
Kilometres of roads	0.1809	(0.0157)	0.1741	(0.0157)	0.2954	(0.0359)	0.2889	(0.0358)
Landscape quality (mean)	0.1395	(0.0302)	-0.7675	(0.5350)	0.1042$^{(+)}$	(0.0723)	-3.1525	(1.0140)
Landscape quality (s.d.)	-		1.5636	(0.4032)	-		2.0262	(0.5065)
Site-choice equation (class 2)								
Travel cost					-0.0580	(0.0195)	-0.0541	(0.0202)
Picnic facilities					0.1616$^{(+)}$	(0.1226)	0.1323$^{(+)}$	(0.1244)
Kilometres of roads					0.1057	(0.0232)	0.0964	(0.0232)
Landscape quality (mean)					0.1669	(0.0487)	-2.4384	(0.6108)
Landscape quality (s.d.)					-		1.3992	(0.4654)
Membership equation (class 1)								
Intercept					1.7800	(0.4379)	1.8910	(0.4589)
Income					-0.0647	(0.0189)	-0.0627	(0.0191)
City					-1.8799	(0.2819)	-1.9019	(0.2839)
High-education					-1.0504	(0.2930)	-1.1162	(0.3038)
Natural areas					0.4456$^{(**)}$	(0.2607)	0.4592$^{(**)}$	(0.2641)
Log-likelihood function	*-3146.1239*		*-3144.3707*		*-2994.4493*		*-2990.1738*	
Restricted log-likelihood	*-3429.2090*		*-3429.2090*		*-3429.2090*		*-3429.2090*	
McFadden-R^2	*0.0826*		*0.0831*		*0.1268*		*0.1280*	
Adjusted McFadden-R^2	*0.0814*		*0.0816*		*0.1230*		*0.1237*	
In-sample forecasts mean-squared error	*0.0002692*		*0.0002716*		*0.0002034*		*0.0002173*	

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

dom parameter coefficients. However, the evidence suggests that, in the present dataset, only the consideration of the 'landscape quality' variable as a random parameter significantly improves the model fit. The rationale behind such pattern of heterogeneity is based on both, the wide range of landscape quality levels existent in the Mallorcan forests (in contrast, to the low variation that characterize other variables such as, for instance, the presence/absence of picnic facilities) and the variability in the perception of quality that can be found across individuals.[4]

All coefficients have been included as fixed in both models with the exception of the 'landscape quality' attribute, specified as random following a lognormal distribution.[5] The values related to the 'landscape quality' variable provided in Table 1 correspond to the estimated mean and standard deviation parameters of a lognormal distribution. Consequently, implicit in this lognormal distribution is the assumption that while preferences vary over individuals, everyone prefers more to less 'landscape quality' in the visited site. In fact, the highly significant standard deviation parameter of the 'landscape quality' variable indicates that preferences towards this attribute do indeed vary across the population. Overall, the coefficients derived under the RPL specification are in the same line that those of the CL model.

Although the RPL model allows the identification of heterogeneity in preferences through the consideration of taste variation towards landscape quality, this specification do not provide any information on the source of such diversity of tastes. In contrast, as explained above, the LC model has allowed the identification of different behavioural groups within the sample of respondents. Unfortunately, the determination of the optimal number of classes or segments with different preferences is not considered in the estimation process and no valid statistical test exists to decide the number of classes (Thacher *et al.*, 2005; Hynes *et al.*, 2008).[6] In this context, although

[4]The concept of quality can be based on different environmental features depending on the recreational interests of each individual, its environmental attitudes, education, etc.

[5]The lognormal distribution has traditionally been used in variables which coefficient is expected to have the same sign for all individuals in the sample (Revelt and Train, 1998; Train, 1999). In this case, it is quite reasonable to assume that 'landscape quality' is a desirable feature with a positive effect on utility for all respondents.

[6]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 (Scarpa *et al.*, 2007).

different statistics based on the information criteria developed by Hurvich and Tsai (1989),[7] as well as the entropy index,[8] have commonly been used to provide some guidance to the analyst to decide the number of segments or classes (Thacher *et al.*, 2005), the choice of number of classes must also consider other issues as the significance of the estimated parameters and the meaningfulness of the parameters and their signs (Scarpa and Thiene, 2005; Morey *et al.*, 2006).

The statistical analysis of our data shows as recreationists split into two groups with a clear differentiated behavioural 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 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, 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.

From the results of the LC model, respondents in class 1 show a high sensitivity to travel expenses and the presence of roads if compared with individuals in class 2. Furthermore, the presence of picnic facilities is an attractive feature for individuals in class 1, but not fore people in class 2. Conversely, the landscape quality of the forest site is desirable for respondents in class 2, while people in class 1 do not care about it. 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 intensive recreational activities such as picnic. In contrast, class 2 is representative of those individuals with a more naturalistic attitude, looking

[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. In this way, when $\delta = 2$ the Akaike Information Criteria (AIC) is obtained, when $\delta = ln(N)$ the Bayesian Information Criteria (BIC) is derived and when $\delta = 2{+}2(J{+}1)(J{+}2)/(N{-}J{-}2)$, 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):

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

for a direct contact with the natural beauty of forests and undertaking non-intensive 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.

The estimated coefficients of the membership equation are also reported in Table 1 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 has to be evaluated relative to group two. More precisely, the final specification of the membership equation includes a constant term and some variables describing the socioeconomic background of individuals. Following Bhat (1997), the constant in the membership equation do not have any substantive interpretation beyond its contribution in the size of the segments.

With regards to the negative sign of the dummy variables 'income', 'city' and 'high-education', they 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 behavioural profiles identified in the site-choice equation of both classes. In this way, the higher income and educational level of people in the second group can explain the lower sensitivity of these respondents to travel costs and their higher interest for landscape quality. In the same line, 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, that includes people living in small villages and, hence, more in contact with nature. Additionally, the probabilities of membership to groups 1 and 2 can be calculated from the membership equation. As a result, the 41.48% of the respondents belong to group 1 and the remaining 58.52% to group 2.

Finally, the coefficients estimates for the LC-RPL specification are quite similar to those of the LC model. However, the major difference is the consideration of intra-group heterogeneity through the 'landscape quality' random parameter. In this way, while the LC specification restricts within group heterogeneity, in the LC-RPL model the intensity of this attraction can vary across individuals within the same group. The estimated variance for the 'landscape quality' variable 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 42.93% for group 1 and 57.07% for group 2.

So far, and comparing the estimates of all models, three conclusions can be drawn about their general performance. First, the magnitudes and signs of the coefficient estimates do not show big differences among models. Second, based on likelihood ratio tests, the LC-RPL statistically dominates the other specifications in terms of goodness-of-fit.[9] Third, following the mean-squared error calculated for in-sample forecasts, the LC specification provides the best way of predicting site-choice probabilities. Overall, taking into consideration that both LC models dominate the mixed specification and that the LC-RPL accounting for within group heterogeneity is only outperformed by the LC in a 6.83%, it seems quite reasonable to choose the LC-RPL as the most preferable approach to model unobserved preference heterogeneity.

In addition, as the estimation of the welfare effects related to environmental policy changes is an important goal facing environmental economists today, the analysis of such measures derived under different ways of accounting for taste variation becomes of interest. For this reason, welfare measures are defined and calculated in next section to investigate the implications of the choice of method for incorporating preference heterogeneity.

5 Welfare estimates

Based upon Small and Rosen (1981) and Hanemann (1982; 1999) measures of welfare change for random utility models, the Willingness-To-Pay (WTP) is defined as the measure of the welfare change associated to an increase (decrease) of some attribute present in the indirect utility function of an individual. Accordingly, in a context of no income effects, the individual-specific WTP becomes:

[9]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).

$$WTP_n = \frac{1}{\beta^{TC}} \left[ln \left(\sum_{i=1}^{I} e^{\beta' x_{ni}^0} \right) - ln \left(\sum_{i=1}^{I} e^{\beta' x_{ni}^1} \right) \right] \tag{12}$$

where β_{TC} is the travel cost coefficient associated with the marginal utility of income and the 0 and 1 superscripts refer, respectively, to the initial and the new state following some change in the set of attributes x_{ni}.

Although the WTP measure in equation 12 corresponds to the conditional logit model, such measure has been also estimated for the RPL (Hynes *et al.*, 2008), the LC (Boxall and Adamowicz, 2002) and the LC-RPL. However, only the equation of the LC-RPL is included here in the interest of brevity. In this case, the expected WTP is conditional on individual tastes β_n and, hence, it has to be calculated by integrating over the taste distribution of the population. Additionally, to represent the distribution of welfare effects across segments, the expected WTP has to be weighted by segment membership becoming:

$$WTP_n = \sum_{k=1}^{K} \pi_{nk} \int \left(\frac{1}{\beta^{TC}} \left[ln \left(\sum_{i=1}^{I} e^{\beta_k' x_{ni}^0} \right) - ln \left(\sum_{i=1}^{I} e^{\beta_k' x_{ni}^1} \right) \right] \right) f(\beta_k) d\beta_k \tag{13}$$

Given the impact caused by recreation demand on some forest ecosystems, specially when visitation rates exceed site carrying capacity, the reallocation of visitors from more crowded areas to less visited sites becomes a legitimate policy to guarantee both, the quality of the recreational experience and the preservation of natural areas. In this context, six recreational sites with visitation rates under the mean have been chosen two test the development of two policies intended to improve their environmental attributes and recreational facilities and, in this way, increase the number of trips to these areas. More precisely, the first policy assumes a 50% increase in the landscape quality of those sites, requiring important investments to enhance the conservation of forests, increasing the diversity of species and landscapes and reducing the negative effects of anthropogenic elements. The second policy simulation involves the provision of additional picnic facilities in the same set of recreational sites that, currently, lack in such infrastructure.

The welfare results of both policy scenarios are summarized in Table 2. In addition, the Krinsky-Robb simulation method have been implemented to estimate the 95% confidence intervals (Krinsky and Robb, 1986).[10] As expected, mean annual WTP (and their confidence intervals) vary significantly among the estimated models. However, the variation in such welfare estimates depends not only on the consideration of heterogeneous preferences, but also on the assumptions of the model concerning the representation of such heterogeneity in tastes. In this way, the empirical results show that those models considering higher degree of preference heterogeneity in their specifications lead to higher mean WTP estimates. Therefore, while the individual WTP for a 50% increase of landscape quality in the 6 chosen sites is 0.0887 euros in the RPL model, it becomes more than three times as much in the LC-RPL model (0.2998 euros per person and year).

Table 2: Mean expected WTP (95% confidence intervals in brackets)*

Scenario	Model	Class 1	Class 2	Average
1	CL	-	-	0.0685
				[0.0394 – 0.0978]
	RPL	-	-	0.0887
				[0.0565 – 0.1345]
	LC	0.0113	0.4072	0.1809
		[-0.0037 – 0.0302]	*[0.1386 – 0.9802]*	*[0.0575 – 0.4781]*
	LC-RPL	0.0326	0.5104	0.2998
		[0.0118 – 0.0714]	*[0.1920 – 1.4086]*	*[0.1274 – 0.7759]*
2	CL	-	-	0.1232
				[0.0601 – 0.1891]
	RPL	-	-	0.1165
				[0.0577 – 0.1781]
	LC	0.0406	0.3698	0.1874
		[0.0116 – 0.0757]	*[-0.1673 – 0.8774]*	*[-0.0332 – 0.4391]*
	LC-RPL	0.0401	0.2511	0.1590
		[0.0106 – 0.0701]	*[-0.2232 – 0.9906]*	*[-0.1048 – 0.5366]*

(*) Confidence intervals have been computed following the Krinsky-Robb method with 1,000 repetitions. WTP values expressed in euros per year.
Source: own elaboration

Consequently, the choice of modelling approach to account for heterogeneous tastes is a key issue when WTP measures are required to guide policy decision-making. In this context, the use of the mean expected annual WTP

[10] More precisely, 1,000 random draws from a multivariate normal distribution, with means given by the estimated coefficients and covariance given by the estimated covariance matrix of the coefficients, have been used to simulate values of WTP (Armstrong *et al.*, 2001). At this point, the 0.025 and 0.975 percentile has been used as the limits of the confidence interval at the 95% level (Hole, 2007).

to evaluate the change in welfare under different policy scenarios, can lead to very different policy recommendations. For instance, and following the results of the CL and RPL models, recreationists put a higher value on the second policy scenario (the provision of new picnic facilities). However, when a LC model is considered, both policies show a similar WTP and when the LC-RPL estimates are analysed, the policy intended to increase landscape quality becomes the most valued one.

Concerning the behavioural groups identified in the LC and the LC-RPL specifications, specific welfare estimates have been provided for each segment. WTP estimates illustrate that individuals belonging to the second group, on average, experience larger welfare impacts from both policy scenarios compared to respondents in group 1. Taking into consideration that people with higher income belong to group 2, 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, show an expenditure rate higher than lower income individuals. Moreover, as income increases, the range of values of the intervals increases considerably.

However, beyond the differences found in WTP between groups, the LC-RPL model confirms the hypothesis of intra-group heterogeneity by showing that the WTP is not constant across individuals in the same group. That is, people in the same group and, hence, with similar socioeconomic characteristics, have different welfare measures. To illustrate this, kernel density distributions of WTP have been estimated and plotted in Figure 1 to establish the empirical profile of welfare estimates for both policy scenarios. From these distributions, it is clear that, while the conventional LC model proves only the existence of two differentiated groups including people with similar WTP, the LC-RPL shows a larger variability in the WTP measure, not only between both groups, but also across individuals in the same group. Such heterogeneity in WTP is especially relevant across people in group 2 showing a considerable range in the values of WTP for an increase in landscape quality.

Overall, although recreationists groups identified by the LC model are very compelling from a managerial standpoint, the LC-RPL model shows that those segments estimated in the LC model can be over-simplifying the real preferences of individuals leading to an underestimation of the real WTP of recreationists.

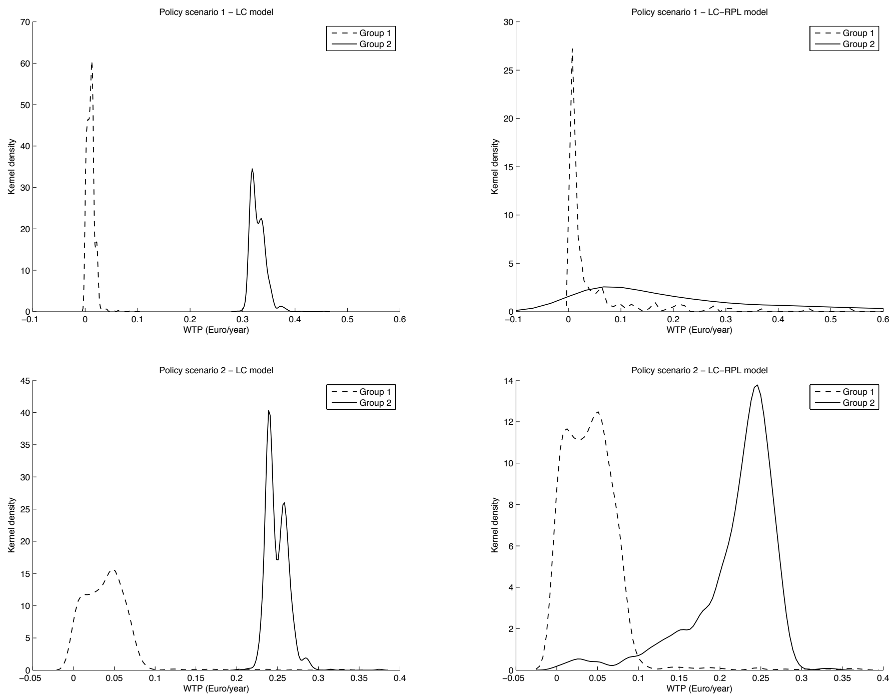

Figure 1: Estimates kernel densities of individual WTP

6 Conclusion

In recent years, the presence of unobserved preference heterogeneity has been widely recognized as a critical issue not only for modelling choice behaviour (Allenby and Rossi, 1998; Wedel *et al.*, 1999), but also for policy analysis (Christie and Hanley, 2008). Examining alternative approaches for incorporating such heterogeneity in models of recreational demand has been the central issue of this paper. We have combined discrete and continuous heterogeneity representations of tastes to capture the best features of both the LC and the RPL models. Unlike previous LC approaches, the LC-RPL model has allowed the joint consideration of discrete segments and within segment heterogeneity providing a richer interpretation of preference heterogeneity. In addition, the CL, the RPL, and the LC model have also been estimated for comparison reasons.

The results have revealed the existence of heterogeneous preferences towards environmental attributes and recreational facilities mainly related to socioeconomic characteristics (income, education, place of residence, etc.).

More precisely, two behavioural groups with a different socioeconomic profile have been identified in the empirical application. Moreover, a broad characterization of both classes in terms of their relative preferences over environmental attributes and recreational facilities of Mallorcan forests has been undertaken. On the one hand, a group of respondents interested in visiting forests to undertake picnic-related activities, specially concerned with accessibility issues, travel expenses and picnic facilities. On the other hand, a group of individuals characterized by a higher income and educational level and more inclined to consider landscape quality as an integral part of their recreational experience.

Overall, specifications accounting for preference heterogeneity have demonstrated a higher performance in terms of both, goodness-of-fit and in-sample forecasts. In fact, only the RPL has showed a lower prediction power when compared with the conditional logit model. More precisely, and in contrast to the findings reached by Wedel *et al.* (1999), models with discrete (LC) and discrete-continuous (LC-RPL) representations of heterogeneity, appear to be doing better than those specifications based exclusively on continuous distributions of tastes (RPL). The LC-RPL model has become the best performance in terms of goodness-of-fit and has only been narrowly outperformed by the conventional LC specification for best in-sample predictions with the lower mean squared error.

Concerning welfare estimates, the RPL produces welfare estimates similar to those of the conditional logit model for a 50% increase in the landscape quality of 6 sites, while WTP measures computed from the LC and the LC-RPL specifications become three times bigger. Considerable differences have been found between the two recreationists groups 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 landscape quality. 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 relayed restrictions by including preference heterogeneity for individuals within the same group, and it has provided an additional insight to find and understand the implications of preference heterogeneity and, hence, it has lead to more useful and accurate WTP estimates. In fact, the results from the LC-RPL model have given 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. There is no doubt that the ability of the LC-RPL model to identify different groups of users, based on their socioeconomic characteristics, at the same time that allows for within group taste heterogeneity towards different environmental site attributes, can become very useful for 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.

7 References

Adamowicz, W. L. (2004). What's it worth? An examination of historical trends and future directions in environmental valuation. *Australian Journal of Agricultural & Resource Economics*, 48(3), 419-443.

Aldrich, G., Grimsrud, K., Thacher, J., and Kotchen, M. (2007). Relating environmental attitudes and contingent values: how robust are methods for identifying preference heterogeneity? *Environmental and Resource Economics*, 37(4), 757-775.

Allenby, G. M., Arora, N., and Ginter, J. L. (1998). On the Heterogeneity of Demand. *Journal of Marketing Research*, 35(3), 384-389.

Allenby, G. M., and Rossi, P. E. (1998). Marketing models of consumer heterogeneity. *Journal of Econometrics*, 89(1-2), 57-78.

Andrews, R. L., Ainslie, A., and Currim, I. S. (2002). An Empirical Comparison of Logit Choice Models with Discrete Versus Continuous Representations of Heterogeneity. *Journal of Marketing Research*, 39(4), 479-487.

Armstrong, P., Garrido, R., and Ortúzar, J. d. D. (2001). Confidence intervals to bound the value of time. *Transportation Research Part E: Logistics and Transportation Review*, 37(2-3), 143-161.

Ben-Akiva, M., and Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand.* Cambridge, Massachusetts: The MIT Press.

Bhat, C. R. (1997). An Endogenous Segmentation Mode Choice Model with an Application to Intercity Travel. *Transportation Science*, 31(1), 34-48.

Bhat, C. R. (1998). Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. *Transportation Research Part A: Policy and Practice*, 32(7), 495-507.

Bhat, C. R., and Castelar, S. (2002). A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area. *Transportation Research Part B: Methodological*, 36(7), 593-616.

Boxall, P. C., and Adamowicz, W. L. (2002). Understanding Heterogeneous Preferences in Random Utility Models: A Latent Class Approach. *Environmental and Resource Economics*, 23(4), 421-446.

Christie, M., and Hanley, N. (2008). Evaluation of heterogeneous preferences for forest recreation in the UK using choice experiments. In E. Birol & P. Koundouri (Eds.), *Choice experiments informing environmental policy*. Cheltenham, UK: Edward Elgar.

Englin, J., and Shonkwiler, J. S. (1995). Modeling Recreation Demand in the Presence of Unobservable Travel Costs: Toward a Travel Price Model. *Journal of Environmental Economics and Management*, 29(3), 368-377.

Greene, W. H., and Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*, 37(8), 681-698.

Gupta, S., and Chintagunta, P. K. (1994). On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models. *Journal of Marketing Research*, 31(1), 128-136.

Haab, T. C., and McConnell, K. E. (2002). *Valuing environmental and natural resources: the econometrics of non-market valuation*. Cheltenham, UK: Edward Elgar.

Hanemann, W. M. (1982). Applied welfare analysis with qualitative response models. CUDARE Working Papers. Berkeley: Department of Agricultural and Resuorce Economics, University of California.

Hanemann, W. M. (1999). Welfare analysis with discrete choice models. In J. A. Herriges & C. L. Kling (Eds.), *Valuing recreation and the environment*. Cheltenham, UK: Edward Elgar.

Hensher, D., and Greene, W. (2003). The Mixed Logit model: The state of practice. *Transportation*, 30(2), 133-176.

Hess, S., Bierlaire, M., and Polak, J. W. (2005). Capturing correlation and taste heterogeneity with mixed GEV models. In R. Scarpa & A. Alberini (Eds.), *Applications of simulation methods in environmental economics*. Dordrecht, The Netherlands: Springer.

Hicks, R. L., and Strand, I. E. (2000). The Extent of Information. *Land Economics*, 76(3), 374-385.

Hole, A. R. (2007). A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Economics*, 16(8), 827-840.

Hurvich, C. M., and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2), 297-307.

Hynes, S., Hanley, N., and Scarpa, R. (2008). Effects on Welfare Measures of Alternative means of Accounting for Preference Heterogeneity in Recreational Demand Models. *American Journal of Agricultural Economics*, in press.

Krinsky, I., and Robb, A. L. (1986). On Approximating the Statistical Properties of Elasticities. *The Review of Economics and Statistics*, 68(4), 715-719.

Lenk, P., and 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 P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105-142). New York: Academic Press.

McFadden, D. (1977). Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments. Unpublished Discussion Papers. Cowles Foundation Discussion Papers.

McFadden, D. (1978). Modelling the choice of residential location. In A. K. e. al. (Ed.), *Spatial Interaction Theory and Planning Models* (pp. 75-96). Amsterdam: North-Holland Publishing.

McFadden, D., and Train, K. E. (2000). Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics*, 15(5), 447-470.

Meijer, E., and Rouwendal, J. (2006). Measuring welfare effects in models with random coefficients. *Journal of Applied Econometrics*, 21(2), 227-244.

Milon, J. W., and Scrogin, D. (2006). Latent preferences and valuation of wetland ecosystem restoration. *Ecological Economics*, 56(2), 162-175.

Morey, E., and Greer Rossmann, K. (2003). Using Stated-Preference Questions to Investigate Variations in Willingness to Pay for Preserving Marble Monuments: Classic Heterogeneity, Random Parameters, and Mixture Models. *Journal of Cultural Economics*, 27(3), 215-229.

Morey, E., Thacher, J., and Breffle, W. (2006). Using Angler Characteristics and Attitudinal Data to Identify Environmental Preference Classes: A Latent-Class Model. *Environmental and Resource Economics*, 34(1), 91-115.

Ouma, E., Abdulai, A., and Drucker, A. (2007). Measuring Heterogeneous Preferences for Cattle Traits among Cattle-Keeping Households in East Africa. *American Journal of Agricultural Economics*, 89(4), 1005-1019.

Patunru, A. A., Braden, J. B., and Chattopadhyay, S. (2007). Who Cares about Environmental Stigmas and Does It Matter? A Latent Segmentation Analysis of Stated Preferences for Real Estate. *American Journal of Agricultural Economics*, 89(3), 712-726.

Phaneuf, D. J., and Smith, V. K. (2005). Recreation Demand Models. In K.-G. Mäler & J. Vincent (Eds.), *Handbook of Environmental Economics: Valuing Environmental Changes* (pp. 671-761). Amsterdam: Elsevier.

Provencher, B., Baerenklau, K. A., and Bishop, R. C. (2002). A Finite Mixture Logit Model of Recreational Angling with Serially Correlated Random Utility. *American Journal of Agricultural Economics*, 84(4), 1066-1075.

Provencher, B., and Bishop, R. C. (2004). Does accounting for preference heterogeneity improve the forecasting of a random utility model? A case study. *Journal of Environmental Economics and Management*, 48(1), 793-810.

Provencher, B., and Moore, R. (2006). A Discussion of Using Angler Characteristics and Attitudinal Data to Identify Environmental Preference Classes: A Latent-Class Model. *Environmental and Resource Economics*, 34(1), 117-124.

Revelt, D., and Train, K. (1998). Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level. *The Review of Economics and Statistics*, 80(4), 647-657.

Scarpa, R., and Thiene, M. (2005). Destination Choice Models for Rock Climbing in the Northeastern Alps: A Latent-Class Approach Based on Intensity of Preferences. *Land Economics*, 81(3), 426-444.

Scarpa, R., Thiene, M., and Tempesta, T. (2007). Latent class count models of total visitation demand: days out hiking in the eastern Alps. *Environmental and Resource Economics*, 38(4), 447-460.

Shonkwiler, J. S., and Shaw, W. D. (2003). A finite mixture approach to analyzing income effects in random utility models: reservoir recreation along the Columbia river. In N. Hanley, W. D. Shaw & R. E. Wright (Eds.), *The new economics of outdoor recreation.* Cheltenham, UK: Edward Elgar Publishing Limited.

Small, K. A., and Rosen, H. S. (1981). Applied Welfare Economics with Discrete Choice Models. *Econometrica*, 49(1), 105-130.

Thacher, J. A., Morey, E., and Craighead, W. E. (2005). Using patient characteristics and attitudinal data to identify depression treatment preference groups: a latent-class model. *Depression and Anxiety*, 21(2), 47-54.

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

Train, K. E. (1999). Mixed logit models for recreation demand. In J. A. Herriges & C. L. Kling (Eds.), *Valuing recreation and the environment: revealed preference methods in theory and practice.* Cheltenham: Edward Elgar.

Train, K. E. (2003). *Discrete choice methods with simulation.* Cambridge: Cambridge University Press.

Wedel, M., and Kamakura, W. (2002). Introduction to the Special Issue on Market Segmentation. *International Journal of Research in Marketing*, 19(3), 181-183.

Wedel, M., Kamakura, W., Arora, N., Bemmaor, A., Chiang, J., Elrod, T., *et al.* (1999). Discrete and Continuous Representations of Unobserved Heterogeneity in Choice Modeling. *Marketing Letters*, 10(3), 219-232.

Wedel, M., and Kamakura, W. A. (2000). *Market segmentation: conceptual methodological foundations.* Boston: Kluwer.