Environ Resource Econ (2009) 44:379–395
DOI 10.1007/s10640-009-9291-7

Deriving and Testing Efficient Estimates of WTP Distributions in Destination Choice Models

Mara Thiene · Riccardo Scarpa

Accepted: 27 April 2009 / Published online: 16 May 2009
© Springer Science+Business Media B.V. 2009

Abstract Estimation of welfare measures is often a dominant driver in the empirical literature on nonmarket valuation. To this end, qualitative choice models based on random utility theory have been widely employed in outdoor recreation studies. A frequent goal of applied studies has been the estimation of welfare changes associated with site attribute changes at recreation sites in order to inform regulatory policy and resource management. We review the evolution of the methodology of random utility theory in this field with a focus on taste heterogeneity models and then focus on the recent proposal of specifying utility in the WTP-space (Train K, Weeks M (2005) Discrete choice models in preference space and willing-to-pay space. In: Scarpa R, Alberini A (eds) Applications of simulation methods in environmental and resource economics, chapter 1. Springer, Dordrecht, pp 1–16). Our empirical application is on outdoor alpine recreation data. We emphasize the efficiency and direct testing that using the maximum simulated likelihood estimator affords to practitioners using the WTP-space approach, and illustrate these with examples.

Keywords Mixed logit · Random parameters · Random willingness to pay · Travel cost · Destination site selection

JEL Classification H4 · Q51 · C12

M. Thiene
Dipartimento Territorio e Sistemi Agroforestali, University of Padua, Viale dell'Università, 16, 35020 Legnaro, PD, Italy
e-mail: mara.thiene@unipd.it

R. Scarpa (✉)
Department of Economics, Waikato Management School, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand
e-mail: rscarpa@waikato.ac.nz

1 Introduction

Discrete choice models addressing preference heterogeneity are very widespread not only in market research, health economics, food economics, and transport studies but also in non-market valuation of environmental goods. Most of the applications of discrete choice analysis in applied environmental economics are geared towards the estimation of structural models of qualitative choice for the purpose of deriving estimates of non-market values. An area of frequent application of these models is that of outdoor recreation. This line of enquiry is driven by the need to provide measures of economic values associated with changes in amenities recreation sites in order to inform regulatory policy and resource management. Since the pioneering work by Bockstael et al. (1987), which introduced random utility theory in this research programme, the statistical methodology for modelling recreation site selection has revolved around the use of multinomial logit models, which still dominates the practice.

The use of logit models in outdoor recreation has evolved over the years, reflecting the advances in discrete choice modelling. In the early 1990s, Morey et al. (1993) introduced the use of nested logit models to account for different substitution patterns between destinations. In the late 1990s the mixed logit methodology was introduced by Train (1998, 1999) and the previously restrictive assumptions of taste fixity, independence of irrelevant alternatives and independence of choice probabilities in panels of choices by the same agent were overcome. In the early to mid 2000s further refinements of mixed logit were illustrated in the context of flexible substitution patterns via suitably defined structures of joint error components by Herriges and Phaneuf (2002). Provencher et al. (2002) championed the advantages of finite, rather than continuous mixing by illustrating the desirable properties of latent class models, which were further extended by Scarpa and Thiene (2005) in terms of the implied welfare estimates at the individual level to differentiate the policy consequences for various groups of users with homogeneous taste. Further comparisons of alternative ways of addressing taste variation in choice-based travel cost modelling are to be found in Provencher and Bishop (2004) and in Hynes et al. (2008).

The whole evolution of this methodology has a common denominator in specifications that are linear in income and expressed in the so-called "preference space". Such an approach requires a two-step estimation process to derive WTP estimates. In the first step the researcher estimates the preference parameters and the marginal utility of income (or the hyper-parameters of their distributional laws in continuous mixed logit), and in the second step one combines such estimates to derive WTP and/or consumer surplus estimates. Such a two-stage process is not one that provides complete information on maximum likelihood and it is hence a statistically inefficient estimator of WTP distributions.

The two-step estimation is in somewhat strident contrast with the specifications in the "expenditure function space" (or valuation function, or WTP-space or money-space), that have been commonly employed in the practice of referendum contingent valuation data analysis since the seminal paper by Cameron and James (1987). After all, the mathematical formulation of "utility" is only a tool to achieve an ordering of preferences and can easily be re-scaled to any metric, including a money-metric. So, there is no apparent reason as to why such an approach should not be extended to the multinomial choice context from the (admittedly less complicated) binomial one.

The first to directly address this issue were Train and Weeks (2005) who proposed a solution and illustrated it on a stated-preference (SP) dataset of car choice. These authors proposed, as an alternative to the commonly employed preference-space utility specification, a model specified directly in the WTP-space and showed, on the same data, how this model produces WTP distributions with spreads that are more plausible than those obtained by conventional and analogous models in the preference-space. In that study the authors discussed how the presence of the scale parameter of the Gumbel error in multinomial logit models induces correlation across WTP estimates "even when taste intensities are assumed to be uncorrelated" in the preference specification. They note that it is unlikely that many researchers are aware of this being an implication of the derived WTP distributions, yet it inescapably follows from the algebra.

The first application of a recreational destination choice model with utility specified in the WTP-space is by Scarpa et al. (2008). That paper is also the first to (1) compare estimates from hierarchical Bayes (HB) and maximum (simulated) likelihood, (2) report results indicating that the WTP-space specification fits the observed data better, and (3) clearly show via examples that the use of bounded distributions in the preference-space *does not* ensure boundedness of WTP estimates, which instead can be achieved by using these in the WTP-space.

The full implications of utility specifications in the money-space have recently been investigated in other papers, with applications outside outdoor recreation modelling. Balcombe et al. (2009) use SP data to investigate consumer preferences for bread produced with reduced levels of various pesticides, and focus on the issue of mis-reporting. They find that the statistical evidence of mis-reporting disappears once heterogeneity of preferences is addressed, and find evidence supporting a preference-space specification. Sonnier et al. (2007) also use SP data and a slightly different terminology as they call the WTP-space specification the "consumer' surplus model". Similarly to Train and Weeks (2005), they find more plausible WTP distribution features associated with their "surplus model", but noted a lower statistical fit than with the conventional preference space model. In a study of relative importance, despite being somewhat different because it was limited to the binary case, Das et al. (2007) use a maximum (simulated) likelihood random parameters version of a censored logistic regression model to estimate the preferences of Rhode Island residents for landfill sites using contingent choice survey data. They find this model fits their binary data better than that obtained using a standard random utility version, and emphasize the efficiency advantages that MSL estimation affords the analyst, which is instead lost in the two step estimation. Finally, Willis and Scarpa (2009) use WTP-space specifications to explain choices made between different renewable energy microgeneration solutions by UK households.

The point of departure of our study is the observation that the testing of hypotheses on WTP distributions can be carried out more directly using WTP-space specifications and we expand on this, providing examples and discussing the ensuing implications for practitioners. Such an advantage seems to have escaped the attention of analysts so far. Instead, we find it useful to emphasize the advantage of WTP-space specifications because it allows researchers to control the features of WTP distributions directly, rather than indirectly, through the choice of the distributions of the preference parameters. Our results and discussion should encourage analysts to use WTP-space specifications each time they intend to directly test hypotheses implying restrictions on WTP distributions. The practical scope for such testing is vast, and the approach we propose advantageous. With the conventional preference-space approach even the simple test of no-difference between underlying WTP distributions requires some form of parametric bootstrapping (Poe et al. 2005).

The rest of the paper is organized as follows: Sect. 2 illustrates the importance of testing the properties of WTP distributions and the advantages of WTP-space specifications. Section 3 illustrates the method of WTP space estimation while Sect. 4 illustrates the dataset used. The following WTP-space maximum simulated likelihood estimates are discussed in Sect. 5, which also offers an empirical illustration of the testing of hypotheses concerning

WTP in this framework. Section 6 concludes by providing policy implications and possible extension of the research.

2 WTP Distributions and Hypothesis Testing in Mixed Logit Models

From the viewpoint of nonmarket valuation studies, one drawback of mixed logit models specified in the preference space is the lack of control that analysts have over the features of the implied WTP distributions, such as median, quantiles and mean. In order to provide the reader with practical motivation for this study we start with an example. Assume that in the estimation of a random utility model specified in the preference space one coefficient attribute α is normal and the mean is estimated to be equal to 2, with a standard deviation estimate of 2, so that the population distribution for the coefficient is $\alpha \sim \mathrm{N}(2,4)$. In order to derive the attribute marginal WTP one must also have an estimate for marginal utility of income, which in a linear income specification is the negative of the travel cost coefficient. Define this as β, which is obviously negative and to enforce this property on a distribution assume that the negative of the travel cost coefficient is distributed log-normal, as suggested in the seminal literature (Train 1998, 1999) and used in many applications since. Suppose the underlying normal distribution of this log-normal variate is estimated to have a mean equal to -1 and a standard deviation of 1.2, so that $\ln(\beta) \sim \mathrm{N}(-1, 1.44)$. It has long been known that the distribution of WTP in the form of a ratio $\mathrm{WTP} = \alpha/\beta$ does not have a closed form and must be approximated by simulation. That is, by drawing a large number of variates from both of the estimated distributions, and combining them in pairs so as to compute for each replicate r the values of $\mathrm{WTP}^r = \alpha^r/\beta^r$. We illustrate this point in practice by using the statistical package R, which is an open-source freeware software system anyone can download from the web to replicate these results.[1] Consider the following series of commands:

```
set.seed(123456, kind=NULL)           # sets the random seed
alpha <- rnorm(10000)*2+(2)           # draws 10,000 alphas
beta <- exp(rnorm(10000)*1.2+(−1))    # draws 10,000 betas
wtp <- alpha/beta                     # computes 10,000 WTPs
mean(wtp)                             # print the mean of the WTPs
quantile(wtp)                         # print the q.les of the WTPs
hist(wtp,br=100)                      # produces a histogram
```

The results indicate that the WTP distribution has a mean of 11.36 and a median of 4.08, showing extreme and potentially unreasonable skeweness due to long and "fat" tails in the distribution. This despite the fact that original distributions of coefficients appeared quite plausible.

There have been numerous attempts to deal with this issue, which have either involved the use of bounded distributions (e.g., Johnson S_b, constrained triangular, etc.) or forced the analyst to make heroic assumptions, such as assuming a non-random cost coefficient, which implies an untenable fixed marginal utility of money across agents. Scarpa et al. (2008) show that, when the range of the cost coefficient has any positive density in proximity of zero, boundedness of distributions for the utility coefficients in the preference space does not imply boundedness of the implied WTP. Because of the properties of the distribution of ratios of random variables there is always a non-zero probability of such a ratio having implausibly high values, thereby affecting the mean of the simulated distribution. This also

[1] http://www.r-project.org.

creates problems in defining the upper quantiles of such implied distributions. The resulting lack of control over WTP distribution is very limiting in the estimation phase.

It is often important to be able to test directly the distributional features of marginal WTP, yet preference space specification only allows this testing to take place *indirectly* via the estimated preference coefficients. For example, it may be policy relevant to test that no more than a given share of the population has a WTP larger than a given amount. For politicians, who are always interested in maintaining consensus, this might be a very valuable piece of information. Suppose a socially-inclined policy maker wants to implement an access fee that ninety percent of the visitors are willing to pay, so as to limit controversy. This might not be optimal in an economic sense, but the politician might consider the implementation to show her willingness to go "some way towards" economically optimal solutions, while maintaining broad consensus (or minimizing controversy). Then the hypothesis to be tested can be framed around the 90th quantile of the WTP distribution. Using the properties of the indirect cumulative distribution function, given a mean value for the WTP and a range of values for the 90th quantile, it is easy to derive the standard deviation values corresponding to each quantile threshold. If the analyst can impose restrictions on a mixed logit model directly via the utility specification then the hypotheses concerning each assumed quantile can be tested *directly in estimation* using the properties of the ML estimator (e.g., $\chi^2 = 2 \times (\ln L^u - \ln L^r)$). Regardless of the additional scope for testing, the unconstrained estimates of the hyper-parameters of the WTP distributions will be *efficient* estimates under the correct specification.

3 Method

To illustrate the utility specification in the money space we follow the set-up proposed by Train and Weeks (2005) and adapt it to our alpine site selection data. Day visitors are indexed by n, destination sites by j, and choice situations by t. We specify utility as separable in price, p, and a vector of non-price attributes, x because this can help the discussion about the role of the scale parameter:

$$U_{njt} = -\alpha_n p_{njt} + \beta_n' x_{njt} + \eta_{njt} \tag{1}$$

where p_n and β_n vary randomly over visitors and η_{njt} is i.i.d. an extreme value. The variance of η_{njt} is visitor-specific: Var $\left(\eta_{njt}\right) = \mu_n^2 \left(\pi^2/6\right)$, where μ_n is the scale parameter for visitor n.

To understand the circumstances under which the scale parameter can be expected to vary across visitors one has to focus on the conceptual difference existing between a random scale parameter and random values in the specification of the utility function. α (marginal utility of money) and β (taste intensities for alpine site attributes) represent the tastes (or partworths) of the visitor and these parameters vary because of personal taste differences across the population of visitors.

However, because of differences in scale of the Gumbel error μ_n, visitors with the same taste structure may have utilities with different variance. The scale parameter μ_n is not a term within the utility function but rather the standard deviation of utility over different destination choices. By allowing the scale parameter to be random, the researcher gives variability to the error variance of individual utility valuations across visitors (Breffle and Morey 2000).

Focussing on the alpine outdoor recreation study, several reasons can be outlined for random variation in the scale parameter across day visitors. First, visitors may share the same taste structure but have different knowledge of the territory and different past experiences at mountain sites. This can affect the variability of utility evaluations across visitors.

For example, day visitors who already have past experience of most of the destination sites and have done most of the activities they enjoy at most of the sites listed in the consideration set of destinations may have smaller variance than those who have not. In other words their utility evaluations are less dispersed than those of other visitors. Second, the physical conditions and degree of fitness of each visitor are important factors in enjoying mountain visits. Some sites may offer wonderful scenery from specific vantage points and alpine shelters. Yet, some visitors may not be in adequate physical condition to reach them. One would expect those in a good state of fitness to have larger variance in the error since they enjoy a larger spectrum of potential activities.

Because utility is ordinal, one can scale Eq. 1 by the scale parameter to obtain its scale free equivalent. Dividing utility in Eq. 1 by the individual scale parameter does not affect behavior and yet it results in a new error term that has the same variance for all decision-makers:

$$U_{njt} = -(\alpha_n/\mu_n)p_{njt} + (\beta_n/\mu_n)' p_{njt} + \eta_{njt} \tag{2}$$

where η_{njt} is i.i.d. type-one extreme value, with constant variance $\pi^2/6$. The utility coefficients are defined $\lambda_n = (\alpha_n/\mu_n)$ and $c_n = (\beta_n/\mu_n)$, such that utility may be written:

$$U_{njt} = -\lambda_n p_{njt} + c_n' x_{njt} + \eta_{njt} \tag{3}$$

Note that if μ_n varies randomly, then the utility coefficients are correlated, since μ_n enters the denominator of each coefficient. Specifying the utility coefficients to be independent implicitly constrains the μ_n to be constant. If the scale parameter varies and α_n and β_n are fixed, then the utility coefficients vary with *perfect* correlation. If the utility coefficients have a correlation less than unity, then α_n and β_n will necessarily vary in addition to, or instead of, the scale parameter.

Equation 3 is called the model in 'preference space' by Train and Weeks. From this specification, the implied marginal WTP for a site attribute is the ratio of the attribute's coefficient to the price coefficient: $\omega_n = c_n/\lambda_n = \beta_n/\alpha_n$. Using this definition, utility can be rewritten as:

$$U_{njt} = -\lambda_n p_{njt} + (\lambda_n \omega_n)' x_{njt} + \eta_{njt}, \tag{4}$$

Train and Weeks have called this 'utility in WTP space', while Sonnier et al. (2007) called it the 'surplus model'. In our alpine destination choice context the scale is expected to be variable, hence this specification is very useful to identify WTP variation independent of variation in scale. The reason for using the travel cost coefficient as the base to incorporate scale is simply that it can be easy interpreted as WTPs. Of course, any other coefficient could be used. A relevant issue in ML estimation is the resulting non-linearity of this utility specification, which complicates the computation of the gradient.[2]

The two expressions of utility (3) and (4) are behaviourally equivalent. The conventional preference space approach involves the following steps. First the analyst specifies distributions for random coefficients of the preference Eq. 3. Then estimates of the hyper-parameters of the underlying distributions are obtained. And finally, the distributions of WTPs are derived by simulation from the estimated distributions of the random coefficients in preference space. As many authors have mentioned, distributive assumptions in preference space do not imply convenient distributions for WTP, and vice-versa. Furthermore, this two-stage estimation method does not produce efficient estimates of WTP, which is arguably the main purpose of many nonmarket valuation studies.

[2] On the other hand, nonlinearity in the parameters does not affect estimation time in Hierarchical Bayes estimation.

The panel mixed logit specification with continuous taste distributions was introduced in the seminal work by Revelt and Train (1998). In our Alpine destination choice context, visitor n faces a choice among J destination alternatives in each of T_n trips taken over an outdoor season. J in our instance is 18 while we have a maximum of $T = 40$ which represents a reasonable maximum number of days out over a year. However, the number of one day long recreational trips to the Alps varies across individuals, hence the subscript n.

For some coefficients the assumption of normality for β_n is untenable, as in the case of the travel cost coefficient λ_n, which is expected to have a negative sign, implying a zero probability of a positive draw and hence incompatible with a normal distribution that spans the entire line. To ensure a negative price coefficient, it is expressed as $\lambda_n = -\exp(\upsilon_n)$, where υ_n can be considered the latent random normal factor underlying the price coefficient. Let n denote the random terms entering utility, which are υ_n and c_n for the model in preference space, (Eq. 3), and λ_n and ω_n for the model in WTP space, (Eq. 4). Similarly, let utility be written $U_{njt} = V_{njt}(\beta_n) + \eta_{njt}$, with $V_{njt}(\beta_n)$ being defined by either Eqs. 3 or 4, depending on the parameterization.

Visitor n chooses destination i in period t if $U_{nit} > U_{njt} \forall_j \neq i$. Denote the visitor's chosen destination in choice occasion t as y_{nt}, the visitor's sequence of choices over the T_n choice occasions as $y_n = \langle y_{n1}, \ldots, y_{nTn} \rangle$. Conditional on β_n, the probability of visitor n's sequence of choices is the product of standard logit formulas:

$$P(y_n \,|\beta_n) = \prod_{t=1}^{t=T_n} \frac{e^{V_{nit}(\beta_n)}}{\sum_j e^{V_{njt}(\beta_n)}} \tag{5}$$

The unconditional probability is the integral of $L(y_n \,|\beta_n)$ over all values of β_n weighted by its density:

$$P_n(y_n) = \int P(y_n \,|\beta_n) g(\beta_n|\theta) d\beta_n \tag{6}$$

where $g(\cdot)$ is the density of β_n, which depends on the hyper-parameters θ to be estimated. This unconditional probability is called the mixed logit choice probability, since it is a product of logit probabilities mixed over a density of random factors reflecting taste variation. Because the integral has no closed form, in estimation it is approximated via simulation by averaging out over a large number of draws with good equidispersion properties (we used Latin hypercube sampling, as recommended by Hess et al. (2006), and 250 draws):

$$\tilde{P}_n(y_n) = \int P(y_n \,|\beta_n) g(\beta_n|\theta) d\beta_n \approx \frac{1}{R} \sum_{r=1}^{R} P\left(y_n \,|\beta_n^r\right) \tag{7}$$

The resulting simulated sample log-likelihood to maximize is therefore:

$$\ln L = \sum_{n=1}^{N} \ln\left(\tilde{P}_n(y_n)\right) \tag{8}$$

Because it is necessary to use a suitable algorithm to deal with non-linearity in V_{njt} during the search for a maximum for ln$L$ we used BIOGEME (Bierlaire (2002, 2003)) and employed the algorithm CFSQP (Lawrence et al. 1997) which avoids the problem of local optima.

Table 1 Descriptive statistics of excursions per head by destination

Site attribute		Mean	SD	Visits	%
1	Vette Feltrine	0.7	1.5	642	7.0
2	P.Dolomiti-Pasubio	2.1	4.0	1,808	19.6
3	Alpago-Cansiglio	0.5	1.7	414	4.5
4	Asiago	1.5	2.8	1,318	14.3
5	Grappa	0.9	2.1	757	8.2
6	Baldo-Lessini	1.2	3.6	1,045	11.3
7	Antelao	0.3	0.7	244	2.6
8	Pelmo	0.3	0.6	243	2.6
9	Cortina	0.3	0.8	220	2.4
10	Duranno-Cima Preti	0.1	0.3	44	0.5
11	Sorapis	0.1	0.5	128	1.4
12	Agner-Pale S.Lucano	0.1	0.5	112	1.2
13	Tamer-Bosconero	0.2	0.6	188	2.0
14	Marmarole	0.2	0.7	161	1.7
15	Tre Cime-Cadini	0.6	1.2	547	5.9
16	Civetta-Moiazza	0.7	1.3	561	6.1
17	Pale S.Martino	0.7	1.3	564	6.1
18	Marmolada	0.3	0.7	225	2.4

4 Data

The data used for this study are revealed-preference travel-cost based data. This sets this contribution apart from many of the others that have employed stated-preference data. The dataset has already been used in Scarpa et al. (2008) to which we address the reader for a more detailed illustration of the data and for further results. We report here only some essential information.

4.1 Respondents Data

The data for our estimates were collected using a survey and consists of a sample of 858 members of the local (Veneto Region) chapter of the CAI, who reported on their mountain visits for the year 1999. The total number of trips reported was 9,221 and some descriptive statistics are reported in Table 1. The most visited sites are Piccole Dolomiti, Asiago, Lessini-Baldo, which are located in the pre-Alps, and Civetta, Pale S. Martino and Tre Cime, all of which are in the Dolomites. Unsurprisingly, the most frequently attended sites are those located closest to lowland urban centres. The interviewers contacted the CAI members at club meetings taking place in the municipalities of the Veneto region. The various parts of the questionnaire were explained to a small group of respondents, and then each member of the group was asked fill out the questionnaire on their own. Respondents were asked questions about their mountaineering abilities and experience (i.e., when they started mountain recreation, whether they attended mountaineering training courses, and the kind of activities they usually undertook at the sites etc.). Importantly for this application they were asked the total number of day outs they took to each of the 18 sites in the previous twelve months. Finally, they provided the interviewers with socio-economic information about themselves and their households. The round-trip distance from own residence to each of the destinations in the choice set was calculated using the software package "Strade d'Italia e d'Europa". These data were used to estimate the individual travel cost for each trip. Distance costs were converted into monetary values using a figure of 0.35 Euro per km, which was the approximate car running cost at the time. Each reported trip was a day trip, as is customary for this form of local outdoor recreation. The 18 mountain destinations differ substantially from both a morphological and mountaineering point of view, but they can provide both specialist and non-specialist outdoor recreation, and so are all destinations for local visitors.

4.2 Site Attribute Data

Data on the attributes of mountain destinations have been mostly provided by means of a Geographical Information System and some of them were coded according to expert knowledge of the hiking features. Two broad geographically-determined groups can be distinguished. Destinations 1–6 (Table 2) belong to the pre-Alps, which are mountains with gentler slopes and lower peaks separating the plain from the proper Alps. Because of their proximity to the main urban centres, the pre-Alps are the final destination for many local excursions. Destinations 7–18 are in the northeastern Alps, in the mountain chain of the Dolomites, which is an extended rocky area mostly made of dolomite rocks. This rare and characteristic rock type is geologically well-defined as it originates from coral reefs. Mountains made of this rock are quite attractive to the eye, as they tend to take on orange-pink hues at sunset.

The relevance of reliable estimates of welfare measures in this context of nonmarket valuation is mainly driven by the need to inform policy makers and institutions involved in land management. The possibility of directly testing alternative hypotheses on marginal WTP distributions is particularly important in the Alpine region. Tourist visitation choices in this region are strongly dependent on site characteristics. This is because site characteristics are very diverse across sites and they determine the type of activities visitors can practice. Accordingly, we focused the analysis on a larger number of site attributes than those used in the original AJAE paper.[3] The inclusion of further variables is also an attempt to compensate for another issue, which is common in destination choice models. That is, with the exclusion of travel cost, site attributes do not vary between recreationists. In the identification of these site attributes there were no site-specific constants included in the estimation. This implies that no unmeasured attributes were left out of the specification with mean valuation different from zero, and furthermore, if there were any excluded attributes that correlated with attributes included in the specification then we would have had an omitted variable bias.[4] So, by enlarging the number of site attributes included in the specification the intention is to reduce exposure to such a bias.

Some of the attributes describe the land use of the sites and some others provide specific information about site features by means of an index. *Hike difficulty* is a three category scored index describing the average degree of challenge of a hiking itinerary. That is, taking into account not only the length of a trail but also the average degree of adversity of the mountain environment. The higher this value, the more arduous is the expected hiking experience. Other activities that visitors may have the opportunity to take part in include climbing, so we also took features of this amenity into account: *climb difficulty* is a four scored category

[3] With respect to the paper published in The American Journal Agricultural Economics, according to the aims of the study, six additional explanatory variables were employed in the estimation. The remaining variables are the same.

[4] We are grateful to an anonymous reviewer for point out to us this issue of relevance.

Table 2 Site characteristics

	Sites	Hike difficulty (score)	Climb difficulty (score)	Glacier (dummy)	Broadleaves (hectar)	Conifers (hectar)	Shelters (n.)	Total trails (km)	Ferrate (n.)	Hard trails (%)	Easy trails (%)	Climbing routes (score)	Prealps (dummy)
1	Vette Feltrine	3	2	0	24,208	10,950	25	739.1	3	6.6	61.1	2	1
2	P.Dolomiti-Pasubio	1	3	0	9,890	1,560	13	344.5	4	17.4	54.2	5	1
3	Alpago-Cansiglio	3	1	0	10,883	4,439	10	312.9	4	7.7	86.0	1	1
4	Asiago	1	1	0	11,754	27,808	13	815.6	0	0.0	99.7	1	1
5	Grappa	1	1	0	18,960	3,450	5	533.4	1	0.8	99.2	1	1
6	Baldo-Lessini	1	1	0	30,914	3,409	18	801.7	2	1.6	76.3	1	1
7	Antelao	3	3	1	121	4,180	6	103.5	0	7.8	68.0	2	0
8	Pelmo	3	3	0	400	8,953	9	188.1	0	4.0	66.2	3	0
9	Cortina	2	3	0	1,039	10,560	32	480.3	22	10.6	53.4	5	0
10	Duranno-Cima Preti	3	3	0	4,704	3,102	4	143.8	0	9.1	32.6	2	0
11	Sorapis	3	3	0	166	2,162	9	107.7	4	22.6	36.1	2	0
12	Agner-Pale S.Lucano	3	3	0	2,785	5,669	7	196.2	2	13.8	51.4	3	0
13	Tamer-Bosconero	3	2	0	2,682	8,366	6	218.6	0	5.7	30.2	3	0
14	Marmarole	2	1	0	439	7,959	9	231.1	1	6.5	50.9	2	0
15	Tre Cime-Cadini	2	4	0	1,048	3,547	9	177.8	4	8.4	59.9	5	0
16	Civetta-Moiazza	2	4	0	1,477	6,121	16	202.8	4	11.4	34.3	5	0
17	Pale S.Martino	2	3	1	1,175	4,673	14	288.7	11	13.9	46.4	5	0
18	Marmolada	3	4	1	279	604	13	29	2	24.6	21.0	5	0

index of the average degree of difficulty of climbing routes in the site, whereas *climbing routes* is a five scored category index providing information of the number of climbing itineraries available in the area. The data shows that the most challenging and most commonly attempted climbs are at those sites located in the Dolomites. *Alpine shelter* is the number of equipped alpine shelters accessible in the area and *ferrata* is the number of via ferratas. These are technically difficult trails set up during World War I specifically to allow soldiers to reach military vantage points; the climber proceeds by anchoring the body to iron cables secured on the mountain face. Land use of the sites is described on the basis of the amount of *broadleaf* and *conifer* forest cover measured in hectares. The reason for including both types of forest lies in the different type of landscape they provide and their seasonal variation. *Broadleaf* forests are more extensive in the pre-Alps than in the Dolomites. The provision of trails is typically a feature of the service provided by a management organization, which may be the Parks Commission or a local environmental and/or mountaineering association, such as the Alpine Club. As an attribute trails are characterized by different degrees of difficulty and on the amenability of the particular trail network to diverse itineraries. Accordingly, site attributes such as the total length of trails in km (*total trails*), percentage (in terms of total length) of ‘easily’ walkable trails and carriage roads (*% of easy trails*) and the percentage of hard walkable trails (*% of hard trails*) out of the total of the network are included in the specification. The presence of a *glacier* is taken into account by including a dummy variable. The pre-Alps also have more shelters than the Dolomites and two of them (Asiago and Grappa) display the highest percentage of trails classified as ‘easy’ in the study. The Dolomite sites of Cortina and Pale S. Martino were found to have the greatest number of via ferrata. Because Dolomites and pre-Alps differ so much, we also include a common ASC for destinations in the latter group of mountains (*Prealps*).

5 An Empirical Illustration

As an illustration we use the above data to derive WTP-space estimates of outdoor recreation models of choice selection between Alpine sites. Because conventional preference-space models have already been estimated from these data using both HB and MSL estimators and are available elsewhere (Scarpa et al. 2008), we present here only MSL WTP-space estimates. In the first part we compare and discuss results of two different WTP-space specifications. The first model is based on a minimal specification where only the travel cost coefficient is random. The second is the outcome of a more extended specifications search that identified which site attributes have marginal WTPs that are random, i.e., are a source of taste heterogeneity. The search involved introducing randomness into coefficients one coefficient at a time and retaining as randomly specified in the final model 2 only those with significant standard deviations. In the second part we illustrate the direct testing of restrictions on marginal WTP distributions on this second specification. In particular, we test three different sets of hypotheses. We select three mountain site attributes for which our data show evidence of taste heterogeneity and that can be plausible targets of policy actions. On these WTP distributions we test restrictions on the mean value and the standard deviation first separately and then jointly.

5.1 Basic WTP-Space Models

Table 3 reports the estimates obtained using the maximum simulated likelihood estimator specified in WTP space. Recall that estimated parameters in WTP space specifications are the parameters of the implied distribution of marginal WTP (hyper-parameters) and under

Table 3 MSL estimates. Coefficients for WTP space models

WTP parameters	Mean	SE	SD	SE	Shares < 0
ln(λ)	−1.412	0.05	0.983	0.05	
Broadleaves forest (ln Ha)	−0.134	0.11			
Conifers forest (ln Ha)	−0.561	0.10			
Hike difficulty	−0.755	0.12			
Climb difficulty	−0.473	0.25			
Ferrata	−0.325	0.02			
Glaciers	1.126	0.21			
% of hard trails (×100)	−2.287	0.26			
% of easy trails (×100)	6.491	0.62			
Alpine shelters	0.369	0.02			
Total trails (ln)	−1.320	0.35			
Climbing routes	1.730	0.19			
Model 1: simulated log-lik. at convergence −20,801.47					
ln(λ)	−1.03	0.03	0.74	0.03	
Broadleaves forest (ln Ha)	−0.29	0.10	0.85	0.09	0.63
Conifers forest (ln Ha)	−0.21	0.11	1.06	0.08	0.58
Hike difficulty	−0.86	0.12	1.64	0.10	0.70
Climb difficulty	−0.09	0.15	1.65	0.07	0.52
Ferrata	−0.26	0.02			
Glaciers	−0.72	0.22	2.57	0.26	0.61
%of hard trails (×100)	−2.34	1.86	5.77	1.44	0.66
%of easy trails (×100)	7.48	0.51			
Alpine shelters	0.27	0.01	0.10	0.01	0.00
Total trails (ln)	−0.36	0.28			
Climbing routes	1.46	0.13			
Prealps Error Comp.	0.00	0.00	3.79	0.27	0.50
Model 2: simulated log-lik. at convergence −19,642.89					

the correct specification such MSL estimates are efficient (i.e., they have minimum variance). Two models are reported. Model 1 is a simple model in which only (the negative of) the coefficient of travel cost is assumed log-normally distributed, while all other taste intensities are assumed to be fixed. In this specification variation in scale *and* in marginal utility of income are all accounted for by this random parameter, while all attribute WTPs are fixed, and hence the only heterogeneity concerns scale and the travel cost coefficient. All other coefficients are interpretable as marginal WTP, without further transformation. The attributes that show positive marginal WTP are the presence of a Glacier at destination, which is valued at one Euro per choice occasion, the percentage (multiplied by 100) of trails classified as easy in the trail network at destination, which if it is increased by 10% increases the WTP by 65 cents per choice occasion. In addition we found that Alpine shelters were valued at 37 cents per one extra shelter, and finally the no. of climbing routes, resulted in a marginal WTP of 1.7 Euro per choice occasion. Interestingly, both types of tree coverage (conifers and broadleaves) are valued negatively at the margin, and negative values were also recorded for the classification of difficulty of hike and climb, as well as percentage of trails classified as “hard” and the (log of) total length of trail network at destination.

The second (model 2) reports the outcome of a specifications search and defines as random all attributes for which evidence of heterogeneity was found in the search, including the presence of a zero-mean error component that induces correlations between utilities of sites in the Pre-Alps (Herriges and Phaneuf 2002). The assumed distributions for marginal WTP are all independent normals, while (the negative of) the travel cost is assumed to be log-normal. The specification for Model 2 allows for a much more flexible specification by allowing for heterogeneity of marginal WTP across visitors. As a result the value of the simulated sample log-likelihood at a maximum is increased from $-20, 801.47$ to $-19, 642.89$. The estimated mean values of the WTP distributions in Model 2 mostly show the same signs and magnitudes as those of Model 1. Both signs and magnitudes appear plausible, while the significant estimates of the standard deviations provide evidence in favor of heterogeneity for a number of attributes. Such heterogeneity describes how those attributes with negative mean WTPs found in Model 1 are in fact the outcome of a mixture of visitors with negative and positive marginal WTPs. The implied shares of the population of visitors showing a negative marginal WTP are reported in the last column of the bottom part of Table 3. For example, the mean and standard deviation of the marginal WTP distribution for broadleaf coverage shows that for over 37% of visitors such a site attribute has a positive, but small marginal WTP, and that 95% of the WTP values are contained in the interval $[-1.96, 1.38]$, with less than 2.5% showing a marginal WTP larger than 1.38 Euro per choice occasion. We note that with this specification the marginal WTP for the presence of a glacier is negative, but it is associated with a large variation, with 39% of the visitors showing a positive WTP. The attribute with the largest valuation variance is the percentage of trails classified as “hard” at destination, a fact that we find plausible in a population of unspecialized mountain visitors.

This sample of visitors appears to place some emphasis on the length of trails classified as “easily walkable” since the estimated mean WTP is 0.075 Euro for a one percent increase in this category in the total length of the network. Almost all of the visitors show a positive marginal WTP for alpine shelters and the mean of the marginal values of an extra shelter is 0.30 Euro. Mountain destinations with many Alpine shelters are appreciated for both providing refuge in case of sudden weather changes, and because shelters are a popular destination for those wanting to enjoy typical Alpine food. The significant estimate for the standard deviation of the zero-mean additional error component for the Prealps sites indicates that these destinations share a different substitution pattern from the rest.

Although the majority of the visitors dislike hiking along technically challenging routes, 30% of them are willing to pay little less than one Euro for this amenity (see also Fig. 1 for a box-percentile plot representation of this estimated distribution of WTP). Unspecialized visitors seem to dislike ferrata and show no taste heterogeneity about this feature. For those visitors who, besides hiking, are interested in climbing or mountaineering, the availability of climbing routes becomes a relevant issue. Visitors are willing to pay 1.5 Euro for sites with a high density of climbing routes. In addition, slightly less than 40% shows a positive WTP (0.72 Euro) for hiking in mountain sites with a glacier. This result indicates the presence of a group of keen and sufficiently well-experienced hikers or mountaineers in this broad and unspecialized sample from the Alpine Club.

5.2 Testing Hypothesis in WTP Space

MSL estimates in WTP space are powerful tools because they directly estimate WTP values. Under the correct specification they are efficient and the estimator also allows researchers

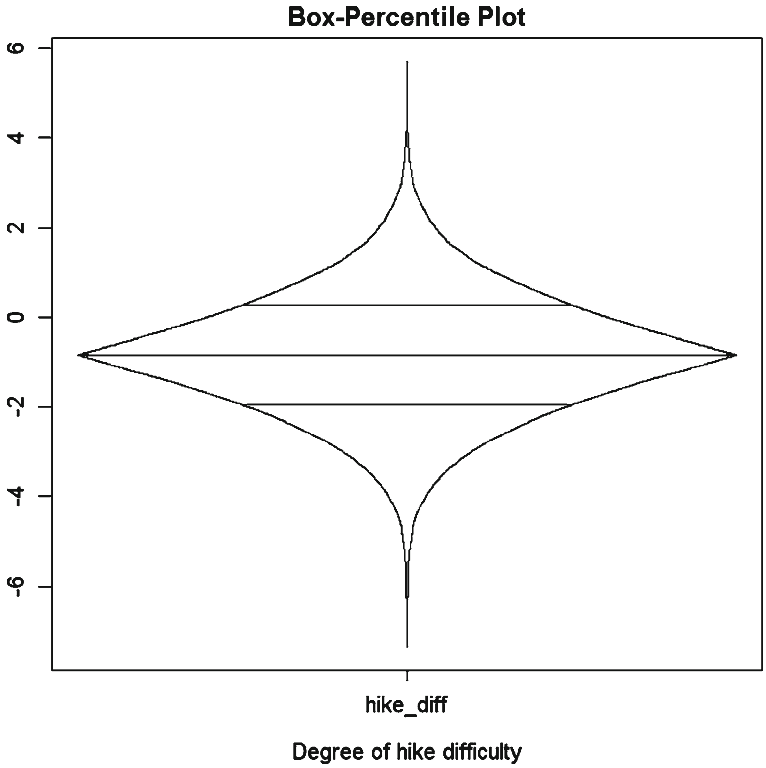

Fig. 1 Distribution of WTP for degree of hike difficulty (SD = 3)

to test the maintained hypothesis directly in the WTP space in an easy fashion. As far as we know, similar testing is not possible, or not as convenient, with models in the preference space when marginal utility of income and scale of error are both random. This versatility turns out to be a powerful analytical advantage in applied work, particularly when focussing on policy implications. Policy makers are usually interested in implementing policies aiming at increasing the benefits of a large fraction of the population. That is, increasing the population share of winners. Suppose the standard deviation of the WTP distribution for a specific site attribute is very large (as for example, in the case of Model 2 for the attribute "hard trails"). That implies that a larger proportion of people are willing to pay vastly different amounts of money to avoid or encourage a specific policy action on hard trails. Hence, a large spread implies a more controversial policy outcome because of the large values at stake and the large population shares underlying each range of values. Institutions involved in land management might be interested in testing the null of a smaller variance of WTP for this attribute. In this case, one might be interested in testing whether restricting the spread of the standard deviation to a smaller value would be statistically supported by the data. This constitutes the basis for our first illustrative hypothesis test. We tested the restriction that the standard deviation value is equal to 3. To appreciate the implications of such a restriction one can examine the values in Table 4 and the box-percentile plots in Fig. 2, where the estimated and the hypothesized distributions are compared. The maintained restriction implies a decrease in the log-likelihood value of 1.66 which in turns implies a χ^2 value of 3.32 and hence the null cannot be rejected at the 5% confidence level (p-value 0.07), but it can at 10%.

Table 4 Comparison of WTP estimates from unrestricted and restricted (SD = 3) models for percentage of hard trails

Quantile (%)	Unrestricted model (€)	Restricted model (€)
1	−15.8	−9.4
5	−11.9	−7.5
50	−2.5	−2.6
95	7.2	2.4
99	10.8	4.3

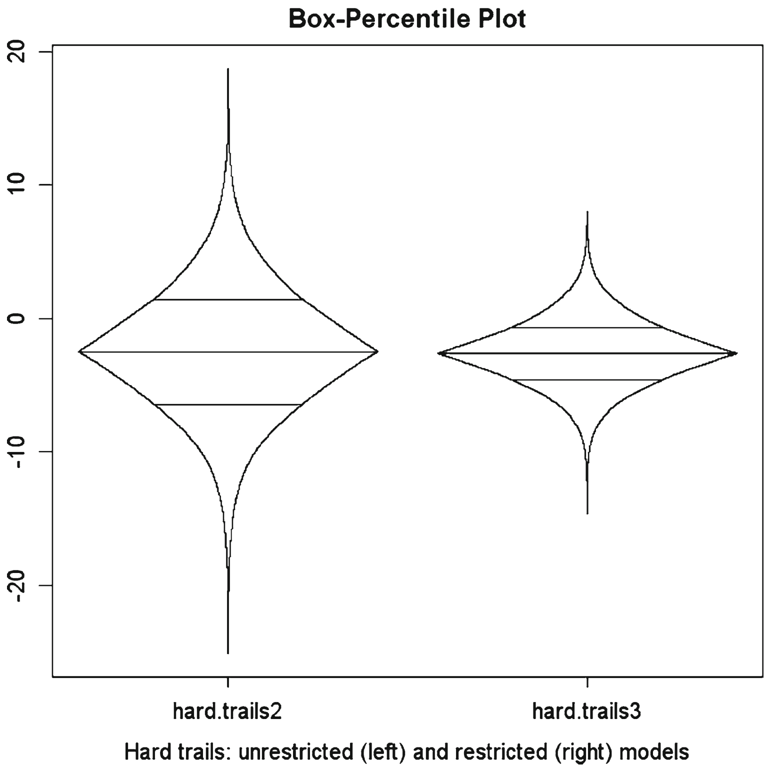

Fig. 2 Distribution of WTP for hard trails from unrestricted and restricted model (SD = 3)

Restrictions can also be formulated in terms of location of the mean of the marginal WTP distribution. In our second illustrative example of a hypothesis test we test the null that the mean of the marginal WTP for the degree of difficulty of the climb is positive and equal to 0.50 Euro rather than the currently estimated value of −0.09 Euro. Although most of the one day trips are taken by visitors going hiking, some occasionally engage in mountaineering, and climbing is growing in popularity. The application of the log-likelihood test provides evidence that this null hypothesis is strongly rejected (p-value 0.01).

Our third and final illustrative example concerns restrictions in both the mean and the standard deviation of marginal WTP and pertains to landscape quality attributes, such as the effect of the relative area covered by broadleaves. We focused on this attribute because we want to concoct an example in which virtually the entire population of visitors shows a positive marginal WTP within the framework of a normally distributed WTP. So, we estimated a restricted model in which the standard deviation is particularly small (equal to 0.1) and the mean is positive, but small (equal to 0.30 Euro). Of course, given the difference in values with the unconstrained estimates we did not expect this restriction to be supported by the data, and it is indeed rejected with a p-value <0.001.

Other hypotheses that can be tested using this approach include different subsets of fixed parameters imposed, or combinations of restrictions having implications for quantiles, with the adequate choices of values for means and standard deviations. The limit, we feel, is only in the imagination of the researcher. Altogether the specification of utility in WTP space represents a major advance in the degree of control the researcher can exercise in formulating testable hypotheses directly on the money space, and hence becomes immediately relevant for policy guidance. The additional cost of a lengthier estimation time in MSL estimation is a relatively low cost to pay considering the ease of testing that it affords. As computer power increases, this cost will inevitably decrease and this approach will become increasingly affordable from the computational viewpoint.

6 Conclusions

WTP estimation is the primary objective of much of the empirical literature on discrete choice modelling for nonmarket valuation in market research, health, transport and environmental economics. In the subfield of applications on outdoor recreation much use has been made of qualitative choice models based on random utility theory to capture preference heterogeneity. We review the evolution of this literature and focus on the recent proposal (Train and Weeks 2005) of choice models in which utility is specified in the WTP space and allows for randomness of scale values across visitors thereby overcoming the issue of interpersonal variance heterogeneity highlighted by Swait and Louviere (1993). We illustrate a method based on maximum simulated likelihood estimation with which hypotheses about WTP distributions can be directly tested, and highlight the efficiency features of WTP estimates derived in this fashion. This study adds to the existing literature providing yet more rationale for the use of WTP-space specifications, and some novel and advantageous implications of using the maximum simulated likelihood estimator with which the testing of such hypotheses is less complex than would be feasible with the hierarchical Bayes estimator. We note that the latter estimator maintains other advantages (e.g., speed) and hence the choice of estimator will remain an issue to be considered on a case by case basis.

Acknowledgments We are grateful to two anonymous reviewers whose comments greatly improved the original manuscript and to participants in the 2008 Summer School on Discrete Choice Modeling at the University of Bologna, Italy for useful comments on the original idea. All remaining errors, as usual are ours.

References

Balcombe K, Fraser I, Chalak A (2009) Model selection in the Bayesian mixed logit: misreporting or heterogeneous preferences? J Environ Econ Manag 57(2):219–225

Bierlaire M (2002) An introductory tutorial to BIOGEME. Federal Polytechnic School of Lausanne, Switzerland. http://rosowww.epfl.ch/mbi/biogeme. Accessed June 2008

Bierlaire M (2003) BIOGEME: a free package for the estimation of discrete choice models. 3rd Swiss Transport Research Conference Monte Verita, March 2003

Bockstael NE, Hanemann WM, Kling CL (1987) Estimating the value of water quality improvements in a recreational demand framework. Water Resour Res 23:951–960

Breffle WS, Morey ER (2000) Investigating preference heterogeneity in a repeated discrete-choice recreation demand model of Atlantic salmon fishing. Marine Resour Econ 15:1–20

Cameron T, James M (1987) Efficient estimation methods for closed-ended contingent valuation survey data. Rev Econ Stat 69(2):269–276

Das C, Anderson CM, Swallow SK (2007) Direct estimation of distributions of willingness to pay for heterogeneous populations. Working Paper

Herriges JA, Phaneuf DJ (2002) Inducing patterns of correlation and substitution in repeated logit models of recreation demand. Am J Agric Econ 84(4):1076–1090

Hess S, Train KE, Polak JW (2006) On the use of a modified latin hypercube sampling (MLHS) method in the estimation of a mixed logit model for vehicle choice. Transp Res B 40:147–167

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

Lawrence C, Zhou J, Tits A (1997) User's guide for CFSQP version 2.5: AC code for solving (large scale) constrained nonlinear (minimax) optimization problems, generating iterates satisfying all inequality constraints. Technical Report TR-94-16r1, Institute for Systems Research, University of Maryland, College Park

Morey E, Rowe R, Watson M (1993) A repeated nested-logit model of Atlantic salmon fishing. Am J Agric Econ 75(3):578–592

Poe GL, Giraud KL, Loomis JB (2005) Computational methods for measuring the difference of empirical distributions. Am J Agric Econ 87(2):353–365

Provencher B, Bishop R (2004) Does accounting for preference heterogeneity improve the forecasting of a random utility model? A case study. J Environ Econ Manag 48:793–810

Provencher B, Barenklau K, Bishop RC (2002) A finite mixture 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. Rev Econ Stat 80(4):647–657

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

Scarpa R, Thiene M, Train K (2008) Utility in WTP space: a tool to address confounding random scale effects in destination choice to the Alps. Am J Agric Econ 90(4):994–1010

Sonnier G, Ainslie A, Otter T (2007) Heterogeneity distributions of willingness-to-pay in choice models. Quant Mark Econ 5(3):313–331

Swait J, Louviere J (1993) The role of the scale parameter in the estimation and use of multinomial logit models. J Mark Res 30:305–314

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

Train K (1999) Mixed logit models for recreation demand. In: Herriges J, Kling C (eds) Valuing recreation and the environment. Edward Elgar, Northampton

Train K, Sonnier G (2005) Mixed logit with bounded distributions of correlated partworths. In: Scarpa R, Alberini A (eds) Applications of simulation methods in environmental and resource economics, chapter 7. Springer, Dordrecht, pp 117–134

Train K, Weeks M (2005) Discrete choice models in preference space and willing-to-pay space. In: Scarpa R, Alberini A (eds) Applications of simulation methods in environmental and resource economics, chapter 1. Springer, Dordrecht, pp 1–16

Willis K, Scarpa R (2009) Willingness-to-pay for renewable energy: Primary and discretionary choice of British households' for micro-generation technologies. 2009 International Energy Workshop, Venice