Understanding Heterogeneous Preferences in Random Utility Models: A Latent Class Approach

Article *in* Environmental and Resource Economics · December 2002
DOI: 10.1023/A:1021351721619 · Source: RePEc

2 authors:

Wiktor Adamowicz
University of Alberta

Peter Boxall
University of Alberta

Environmental and Resource Economics **23:** 421–446, 2002.

Understanding Heterogeneous Preferences in Random Utility Models: A Latent Class Approach

PETER C. BOXALL and WIKTOR L. ADAMOWICZ
Department of Rural Economy, University of Alberta, Edmonton, AB, Canada T6G 2H1
(E-mail: peter.boxall@ualberta.ca)

Accepted 6 March 2002

Abstract. A finite mixture approach to conditional logit models is developed in which latent classes are used to promote understanding of systematic heterogeneity. The model is applied to wilderness recreation in which a branded choice experiment involving choice of one park from a demand system was administered to a sample of recreationists. The basis of membership in the classes or segments in the sample involved attitudinal measures of motivations for taking a trip, as well as their stated preferences over wilderness park attributes. The econometric analysis suggested that four classes of people exist in the sample. Using the model to examine welfare measures of some hypothetical policy changes identified markedly different welfare effects than the standard single segment model, and provided insight into the differential impact of alternative policies.

Key words: choice experiments, latent class, environmental valuation, preference heterogeneity

JEL classification: Q260, Q280

Abbreviations: BWCA – Boundary Waters Canoe Area; ASC – Alternative specific constant; RPL – Random parameters logit; AIC – Aikaike Information Criterion; BIC – Baysian Information Criterion

1. Introduction

Consumer preferences for goods and services are characterized by heterogeneity. Accounting for this heterogeneity in economic analysis will be useful in estimating unbiased models and for forecasting demand by including individual characteristics. Incorporating and understanding heterogeneity will provide information on the distributional effects of resource use decisions or policy impacts. However, some empirical economic analyses assume homogeneous preferences among consumers, while others consider preference heterogeneity *a priori* by: (1) including demographic parameters in demand functions directly or through the utility function (e.g., Pollack and Wales 1992); or (2) by stratifying consumers into various segments and estimating demands separately on each stratum. For these analyses, economists traditionally focus on sociodemographic variables.

Heterogeneity is difficult to examine in the random utility model because an individual's characteristics are invariant among a set of choices. This means that the effect of individual characteristics are not identifiable in the probability of choosing commodities. These limitations have been relaxed by interacting individual-specific characteristics with various attributes of the choices (e.g., Adamowicz et al. 1997). Morey et al. (1993) take advantage of knowledge of income levels by explicitly incorporating them into the indirect utility function of their respondents. These methods are limited because they require *a priori* selection of key individual characteristics and attributes and only involve a limited selection of individual specific variables (e.g., income).

Another set of approaches called random parameter logit/probit models explicitly account for heterogeneity by allowing model parameters to vary randomly over individuals (e.g., Layton 1996; Train 1997, 1998). While these procedures incorporate and account for heterogeneity, they are not well-suited to *explaining* the sources of heterogeneity. In many cases these sources relate to the characteristics of individual consumers.

Two streams of research point to a role for individual-specific characteristics in explaining heterogeneity in choice. The first highlights the possible role of individual characteristics in affecting tastes. For example, in Salomon and Ben-Akiva (1983) choice model development was preceded by multivariate cluster analysis of sociodemographic characteristics to determine relatively homogeneous segments of individuals. In this process, the series of choice models estimated separately for each cluster was statistically superior to a model which pooled the clusters.

A second avenue in explaining heterogeneity involves the scale parameter. Cameron and Englin (1997) explain heterogeneity by "parameterizing" scale in binary logit models. In this case parameterizing heterogeneity in the choice model with demographic variables exhibited superior statistical properties over models which imposed homogeneity. However, it is not clear if individual characteristics should enter as affecting scale, or if these features should influence tastes (i.e., utility parameter differences). In the present paper it is assumed that heterogeneity affects tastes.

In any approach to incorporate heterogeneity into demand analysis there must be *a priori* knowledge of the elements of heterogeneity. Ideally, an effective procedure should utilize theory to provide a foundation for possible sources of heterogeneity. While these sources may include sociodemographics, theory may also point to other characteristics of individuals such as attitudes, perceptions, social influences, and past experiences.[1] Furthermore, while theory may provide an understanding of sources of heterogeneity, it would be desirable to incorporate heterogeneity in the estimation of economic choice parameters. This points to joint estimation of the explanators of heterogeneity and the explanators associated with attributes of choices.

A promising avenue for tackling these problems involves an extension of latent variable approaches to consider latent classes or segments. The latent construct in this strategy represents a typology, classification, or series of segments which are constructed from a combination of observed constituent variables (McCutcheon 1987). Thus, latent class methods involve characterizing segments from discrete observed measures such as attitudinal scales, or can involve empirically testing whether a theoretically posed typology adequately fits a set of data (McCutheon 1987: 8). This framework, when coupled with information on preferences relating to consumer choice, offers an opportunity to both understand and incorporate preference heterogeneity in consumer demand analysis.

2. The Latent Segmentation Approach

McFadden (1986) recognized the prospect of using latent variables in understanding choice behavior. He posed an integration of information from choice models with attitudinal, perceptual and socioeconomic factors using a latent variable system. While the observable outputs using this approach are predictions of choice or market behavior, the underlying constructs of the choice decision process are more elaborate than traditional consumer demand theory. McFadden (1986) mentions that "the critical constructs in modeling the cognitive decision process are *perceptions* or beliefs regarding the products, generalized *attitudes* or values, *preferences* among products, decision protocols that map preferences into choices, and *behavioral intentions* for choice" (McFadden 1986: 276). Thus, the problem for an analyst using this approach is to gather psychometric data to quantify the theoretical or latent constructs underpinning choice behavior, and then simulate this choice behavior using attributes associated with the products of interest.

Swait (1994) utilized McFadden's idea to understand preferences for beauty aids. In this application latent segments were characterized by different degrees of sensitivity to product attributes. Swait utilized brand image ratings from a sample of consumers along eight psychometric dimensions as individual-specific information, and a set of repeated choices of preferred products from among five brands was taken as the choice information. Swait's (1994) model simultaneously conducted market segmentation and predicted choice of beauty product for the sample. This model, called a finite-mixture model in the statistical literature (Titterington et al. 1985), allows market segments to be related to characteristics of individual consumers such as psychometric or socioeconomic effects, but also elements of observed behavior. The finite mixture model may have considerable relevance to decision-makers in that it permits some understanding of preference heterogeneity through incorporation of individual characteristics. It also accounts for preference heterogeneity to a degree by simultaneously estimating segment specific membership and choice parameters.

Our paper applies this latent segmentation approach to a set of wilderness recreation park choice data. The foundation of this application is a model which incorporates motivations towards wilderness recreation and perceptions of environmental quality. The behavioral components come from a choice experiment in which five environmental and managerial attributes of individual parks in a recreation demand system were varied. The analysis will assess simultaneously the influence of individual characteristics, motivational aspects, and the influence of choice-based attributes in the estimation of latent segments. To examine the argument that heterogeneity explained may be more useful in some policy contexts than heterogeneity captured, we estimate a random parameters model and compare welfare impacts between the two approaches. Finally, while explaining heterogeneity is useful in understanding the incidence of impacts (e.g., welfare impacts), accounting for heterogeneity may be important econometrically as well. In discrete choice models of the type we employ, errors that in simple linear models only generate inefficient estimates, can result in bias (Yatchew and Griliches, 1984). Thus accounting for heterogeneity is especially important in these random utility based models.

3. The Latent Segmentation Model

An individual (n) receives utility, U, from choosing an alternative (i) equal to $U_{ni} = U(X_{ni})$, where X_{ni} is a vector of the attributes of i. Utility is modelled as two components, where one portion is deterministic and depends on the attributes of the alternative, and the remainder is not. Thus, $U_{ni} = V_{ni} + \varepsilon_{ni}$ where $V_{ni} = f(X_{ni})$ is the deterministic component and ε_{ni} a random component of the utility function.

In this model, individual n faces a choice of one alternative from a finite set C of sites. The probability (π) that alternative i will be visited is equal to the probability that the utility gained from its choice is greater than or equal to the utilities of choosing another alternative in C. Thus, the probability of choosing i is:

$$\pi_n(i) = \text{Prob}\{V_{ni} + \varepsilon_{ni} \geq V_{nk} + \varepsilon_{nk};\, i \neq k, \forall k \in C\}. \tag{1}$$

The conditional logit model, developed by McFadden (1974), can be utilized to estimate these probabilities if the random terms are assumed to be independently distributed Type-I extreme value variates. Substituting the attributes associated with each alternative into the deterministic portion of utility (V) and selecting a linear functional form allows the choice probabilities to take the form:

$$\pi_n(i) = \frac{\exp(\mu\beta X_i)}{\sum_{k\in C}\exp(\mu\beta X_k)} \tag{2}$$

where μ is a scale parameter that is assumed to equal 1, and β is a vector of parameters. Note that in this model the vector β is not specific to an individual. Now assume the existence of S segments in a population and that individual n belongs to segment s (s = 1, ..., S). The utility function can now be expressed $V_{ni|s} = \beta_s X_{ni} + \varepsilon_{ni|s}$. In this expression the utility parameters are now segment specific and equation (2) becomes:

$$\pi_{n|s}(i) = \frac{\exp(\mu_s\beta_s X_i)}{\sum_{k\in C}\exp(\mu_s\beta_s X_k)} \tag{3}$$

where β_s and μ_s are segment-specific utility and scale parameters respectively.

Following Swait (1994), consider an unobservable or latent membership likelihood function M^* that classifies individuals into one of the S segments. The classification variables influencing segment membership are related to latent general attitudes and perceptions, as well as socioeconomic characteristics of the individuals. For a specific individual n, this function can be described by the following set of equations:

$$\begin{aligned} M_{ns}^* &= \Gamma_{ps} P_n^* + \Gamma_s S_n + \zeta_{ns} \\ P_n^* &= \beta_P P_n + \zeta_{nP} \end{aligned} \tag{4}$$

where M_{ns}^* is the membership likelihood function for n and segment s; P_n^* is a vector of latent psychometric constructs held by n; S_n is a vector of observed sociodemographic characteristics of individual n; P_n is a vector of observed indicators of the latent psychometric constructs held by n; Γ and β_p are parameter vectors to be estimated; and the ζ vectors represent error terms. Relating this function to the classical latent variables approach where observed variables are related to the latent variable, M^* can be expressed at the individual level as:

$$M_{ns}^* = \lambda_s Z_n + \zeta_{ns}, \quad s = 1, \ldots, S \tag{5}$$

where Z_n is a vector of both the psychometric constructs (P_n) and socioeconomic characteristics (S_n), and λ_s is a vector of parameters.

As Swait (1994) points out, these membership likelihood functions are random variates and one must specify the distribution of their error terms in order to use them in practice. Thus, following Swait (1994), Gupta and Chintagunta (1993), and Kamakura and Russell (1989), the error terms are assumed to be independently distributed across individuals and segments with Type I extreme value distribution and scale factor α. Incorporating these assumptions allows the probability of membership in segment s to be characterized by:

$$\pi_{ns} = \frac{\exp(\alpha \lambda_s Z_n)}{\sum_{s=1}^{S} \exp(\alpha \lambda_s Z_n)}, \tag{6}$$

This is the multinomial logit model used by Schmidt and Strauss (1975) in which individual-specific characteristics rather than attributes of choices produce choice probabilities. Other functional forms could be chosen to represent the probability of segment membership. However, regardless of the form chosen, $\sum_{s=1}^{S} \pi_{ns}$ must equal 1, and $0 \leq \pi_{ns} \leq 1$. Now define $\pi_{ns}(i)$ as the joint probability that individual n belongs to segment s and chooses alternative i. This can be expressed as the following product of the probabilities defined in equations (3) and (6): $\pi_{ns}(i) = \pi_{ns}\pi_{n|s}(i)$. Thus the probability that a randomly chosen individual n chooses i is given by:

$$\pi_n(i) = \sum_{s=1}^{S} \pi_{ns} \pi_{n|s}(i). \tag{7}$$

and substituting the equations for the choice equation (3) and membership equation (6) probabilities yields the expression:

$$\pi_n(i) = \sum_{s=1}^{S} \left[\frac{\exp(\alpha\lambda_s Z_n)}{\sum_{s=1}^{S} \exp(\alpha\lambda_s Z_n)} \right] \left[\frac{\exp(\mu_s \beta_s Z_i)}{\sum_{k\varepsilon C} \exp(\mu_s \beta_s X_k)} \right] \quad (8)$$

This model permits choice attribute data and individual consumer characteristics to simultaneously explain choice behavior.

The joint distribution of choice probability and segment membership probability is specified and estimated this model. The membership function determining the structure of the latent classes is not a behavioural relation, but a statistical classification process. Thus, one can ignore the correlation between the error in the utility functions and the "error" in the classification function. If one placed a behavioural interpretation on the classification function, it may be possible to examine the possibility of error correlation between the two equations in this model.

This model can be compared to a 'random parameters logit" or "mixed logit model" with a finite number of support points (and thus this approach is also often called a finite mixture model). In a mixed logit model the probability of choosing an alternative (basic conditional logit model) is modified by integrating over all possible values of tastes. This integration requires assumptions about the structure of the tastes (e.g., normal, log-normal, etc.) but once the structure is assumed, the probability can be estimated (see Train 1998). In the latent class model we multiply the conditional distribution (site choice probability conditional on being in a specific segment) times the probability of being in a segment where the segments are the finite analogue to the random parameters distributions. The distributions are well specified and thus the estimation of the joint distribution occurs.

A number of features of this model are noteworthy. First, the observation that the ratio of probabilities of selecting any two alternatives (equation 8) would contain arguments that include the systematic utilities of other alternatives in the choice set is of note. This is the result of the probabilistic nature of membership in the elements of S. The implication of this result is that independence from irrelevant alternatives need not be assumed (Shonkwiler and Shaw 1997).

Second, there are two types of scale factors which cannot be estimated simultaneously. The α scale parameter represents the scale across the segment membership function and as such is not identifiable. The μ_s terms denote the scale for the sth segment's utility function and in theory can be used to test hypotheses about scale and utility parameter equality across segments (Swait and Louviere 1993). These scale factors are only identifiable under conditions where the segment specific utility parameters are constrained to be equal (e.g., Adamowicz et al. 1997). However, this assumption of parameter equality across segments is contrary to the spirit of the latent segment model since a researcher would **not** want to impose utility parameter equality. Therefore, utilizing this model in empirical estimation requires that all the scale factors in (8) are set equal to one.

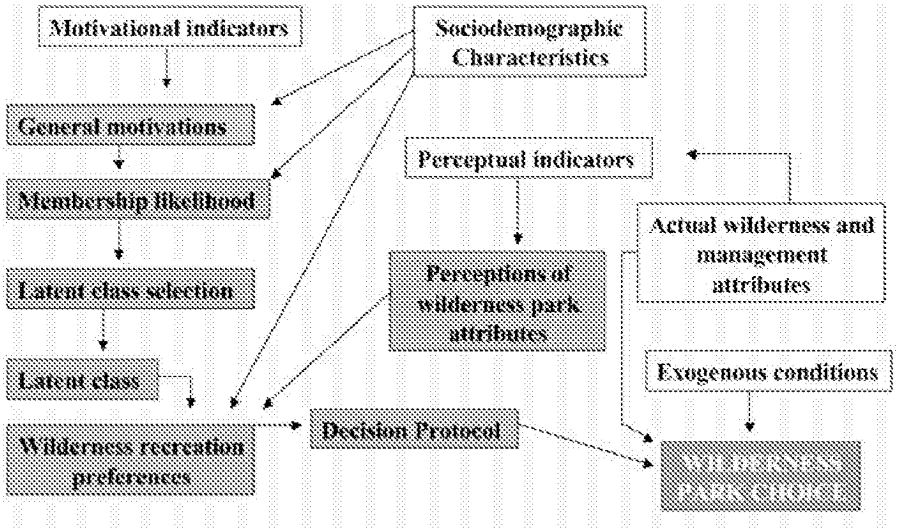


Figure 1. A path diagram outlining the application of the latent segmentation choice model to recreation in a set of wilderness parks in the Canadian Shield region. Shaded boxes refer to the latent constructs utilized in the model.

Third, as Swait (1994) points out when $\lambda_s = 0$, $\beta_s = \beta$, and $\mu_s = \mu$ for each segment, equation (8) reduces to the conditional logit model in shown in (2). These conditions essentially impose homogeneity of preferences and are represented by the case in which there exists only one segment in which every individual in the data holds membership. Conversely, one could consider the case where each individual could be considered a segment. Under this condition each respondent behaves as if their behavior is consistent with a conditional logit model, but each individual has their own set of parameters. This situation can be represented by the random parameter logit/probit model (e.g., Layton 1996; Train 1997, 1998). Thus, the latent segmentation model represents a model located within a range of approaches. On one end of the range is the single segment case which assumes perfect homogeneity of preferences, and at the other end is the case where each individual is considered a segment. The potential advantage of the latent segment model in this series of approaches is its potential to both explain and account for heterogeneity to some degree.

4. An Application – Wilderness Park Choice in Central Canada

4.1. A FRAMEWORK FOR WILDERNESS RECREATION DECISIONS

A framework of choice for wilderness recreation and segmentation was developed based on the path diagrams in McFadden (1986) and Swait (1994). The framework, shown in Figure 1, incorporates latent constructs in boxes shaded with grey while the white boxes represent observable variables. This model utilizes psychometric features that relate to motivations for taking a wilderness recreation trip. The observable motivational indicators are related to latent motivations, and these, in concert with an individual's sociodemographic characteristics, influence the likelihood of membership in one or more latent classes or segments. When observable motivational indicators are available, this part of the framework can be represented by equation (6).

The other components in Figure 1 are related to the attributes of the available wilderness choices and consist in part of actual or objective characteristics of the places one could choose to go. However, some of these characteristics may be influenced by past visits, contact with media, levels of wilderness experience etc., and these elements may result in the formation of perceptions of wilderness features. Perceptions of attribute qualities have been revealed as an important influence in choice behavior by Adamowicz et al. (1997). Both objective and subjective components of wilderness choice attributes, along with sociodemographic characteristics, may influence wilderness recreation preferences. This part of the framework is represented by equation (3).

Putting the psychometric and sociodemographic characteristics together with the objective and subjective wilderness attributes enables the implementation of the latent segmentation model. The result of the model is the probability of choosing a wilderness area from available wilderness choices. A final set of influences on this choice, however, result from exogenous features such as the closure of wilderness areas due to forest fires or other stochastic events.

4.2. EMPIRICAL APPLICATION

This framework for understanding wilderness park choice was applied to recreationists who use a set of five wilderness parks in eastern Manitoba (Nopiming and Atikaki Provincial Parks), western Ontario (Woodland Caribou, Quetico and Wabakimi Provincial Parks) and northern Minnesota (Boundary Waters Canoe Area (BWCA)). Recreational use of these parks has been considered a demand system in previous research (Boxall et al. 1999), indicating that a sample of visitors to these areas would consider them as elements of a recreation choice set. The parks represent a range of development, entry restrictions, congestion levels, and management intervention and have a number of management issues which require knowledge about the characteristics of people who use them and the "products" or features desired for recreation trips. Thus, the application of the latent segment model to visitors in these areas would be of considerable value to park managers.

During 1995 a sample of 1000 visitors to Nopiming and Atikaki Provincial Parks in Manitoba, and Woodland Caribou, Quetico, and Wabakimi Provincial Parks in Ontario were drawn from park registrations or on-site registrations administered by the Canadian Forest Service.[2] About 71% of individuals in this sample were from Quetico, about 18% were from Woodland Caribou, 10% were from both Manitoba parks, and about 1% were from Wabakimi. This distribution was selected because it approximately represented the levels of visitation across the four parks (see Boxall et al. 1999).

A questionnaire was developed that gathered information about opinions of wilderness management, levels of past visitation to all of the parks, descriptions of a typical wilderness trip, and sociodemographic characteristics. Three additional pieces of information were collected that were used in the latent segment model. The first involved a series of 20 statements (see Appendix and Boxall and Adamowicz 1999) that represented reasons why the individual visited backcountry or wilderness areas. Respondents were asked to rate the level of importance of each statement on a 5 point Likert scale ranging from "Not at all important" to "Very important." The statements used for this purpose were derived from research by Crandall (1980) and Beard and Ragheb (1983) on leisure motivations. The scores of the respondents were used to derive a scale to measure motivations for visiting wilderness areas.

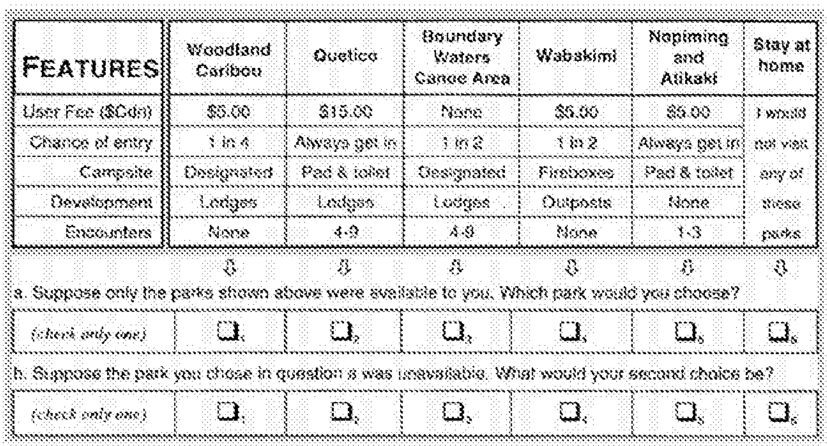
The second was the application of a choice experiment which required respondents to consider choosing among five wilderness areas for a trip next season, or the option of not taking a trip. The choice experiment employed the actual park names as choice options (hence a "labelled or branded" choice) where the two Manitoba parks were combined into one and the Boundary Waters Canoe Area was available as one of the choices. Branded alternatives were employed because the set of alternative wilderness canoe areas used by the experienced wilderness recreationists is well-known, and we felt that using generic alternatives would not be considered realistic by the respondents. Five wilderness attributes each consisting of four levels were developed based on three years of previous research on wilderness recreation in the area, and discussions with park managers, recreationists, and academics. These attributes were: (1) the fee per day per person; (2) the chances of entry into the park as a result of entry or quota restrictions; (3) the type of campsite available; (4) the level of development related to human habitation and access; and (5) the total number of encounters with other wilderness recreation groups per day. These attributes and their levels are described in Table I.

A disadvantage of the branded choice design is that respondents, given their knowledge of the alternatives offered, may consider other attributes of the sites or may not consider the presentation of attributes in each alternative realistic. One way to address this concern is to incorporate alternative specific constants in the analysis of the choice experiment to capture knowledge about the sites that is not reflected in the attributes presented. However, this raises the possibility of specification error and subsequent bias in the coefficients if one has not captured the missing attribute perception adequately. We assume that respondents are able to evaluate our representations of the sites (using attributes that are not the current attributes) and will not confuse the current site attributes with the ones they are asked to evaluate. These assumptions regarding the use of branded alternatives in a recreation context are worthy of investigation in future research.

Table I. A summary of the attributes and levels used in the park level branded choice experiment

Attribute	Level
User fee	None: no fees $5.00 per day per person $10.00 per day per person $15.00 per day per person
Chance of entry due to management restrictions such as quotas	Always get in 3 in 4 1 in 2 1 in 4
Campsite type	Anywhere In designated areas only In designated areas with fireboxes In designated areas with fireboxes, tent pad and pit toilets
Level of development	None: no development in the park and no roads directly to or in the park and no motor boats Outposts: unstaffed outpost cabins in places and a road exists to the boundary of the park, but not inside, some motor boats may be present Lodges: fishing or hunting lodges present with motor boats and a road goes through the park Cottages: some places have cottage sub-divisions and there is a network of roads that allows improved access; motor boats will be present
Encounters with other wilderness groups	None: no other groups will be encountered 1–3 groups are encountered each day 4–9 groups are encountered each day over 9 groups are encountered each day

A choice scenario (see Figure 2) consisted of six alternatives (five parks and the stay at home option).[3] Statistical design methods (see Louviere et al. 2000) were used to structure the presentation of the levels of the five attributes in the scenario. In this presentation the levels of attributes of one alternative (the BWCA) were held constant while those of the other four parks were varied in the design. The BWCA attributes were fixed because management at this park is relatively constant over time and we were not able to obtain a sample of visitors to this park who would be sensitive to management changes there. The attributes and levels for the other four parks were constructed from a $4^5 \times 4^5 \times 4^5 \times 4^5 \times 2$ orthogonal



FEATURES	Woodland Caribou	Quetico	Boundary Waters Canoe Area	Wabakimi	Nopiming and Atikaki	Stay at home
User Fee ($Cdn)	$5.00	$15.00	None	$5.00	$5.00	I would
Chance of entry	1 in 4	Always get in	1 in 2	1 in 2	Always get in	not visit
Campsite	Designated	Pad & toilet	Designated	Fireboxes	Pad & toilet	any of
Development	Lodges	Lodges	Lodges	Outposts	None	these
Encounters	None	4-9	4-9	None	1-3	parks

a. Suppose only the parks shown above were available to you. Which park would you choose?

(check only one)	☐1	☐2	☐3	☐4	☐5	☐6

b. Suppose the park you chose in question a was unavailable. What would your second choice be?

(check only one)	☐1	☐2	☐3	☐4	☐5	☐6

Figure 2. An example of one choice scenario from the choice experiment.

main-effects design, yielding 64 possible combinations of the levels (or choice sets). Based on the choice experiment literature (e.g., Louviere et al. 2000), this number was considered too large a task for a respondent to complete so the 64 combinations were blocked into 8 versions of the questionnaire with eight choice scenarios presented in each version.

The third body of information from the questionnaire was the answers to a series of questions aimed at gathering respondents' current perceptions of the levels of the attributes in each park. These levels and attributes were the same as those described in Table I.

The questionnaire was mailed to the sample of 1000 recreationists. After adjusting for non-deliverables, the response after one post card reminder and a second follow-up questionnaire was 80%. Further adjustment of the respondents for item non-response resulted in a final sample of 620 individuals who provided complete data for the measurement of motivations, sociodemographic characteristics, and information on 4892 choices.

Econometric Model

The first step in developing the latent segmentation model involved a factor analysis of the 20 statements (motivational indicators) to provide estimates of the latent motivational constructs which enter the membership likelihood function. Because these statements were developed *a priori* to assess motivations, this involved an exploratory approach which is common in the recreation literature. The scores from the 20 statements were factor analyzed using principal component analysis with varimax rotation. Components were extracted until eigenvalues were less than or equal to 1.0.

The factor analysis identified four components of motivations for taking a wilderness trip which accounted for virtually all of the variation. These motivational components were labeled based on magnitudes of the loadings of individual statements (see Appendix and Boxall and Adamowicz 1999). The first component was called "challenge and freedom" because statements relating to this factor loaded highly in this factor. The second factor was labeled "nature appreciation". The third factor involved statements relating to family and friends and was labeled "social relationships". The fourth factor was called "escape from routine".

Scores for the four factors were then calculated for each individual in the sample yielding four variables to be included in the Z_n vector in equation (8). An additional variable added to this vector was a dummy variable which equaled one if a respondent's trip length typically was 3 or less days. This variable was selected to capture sociodemographic effects that may influence trip characteristics and that may not be related to the factor scores. Other sociodemographic features of respondents could have been chosen for inclusion in this vector, but the complexity of the model and the estimation effort required limited the set of variables for inclusion.[4] Thus, five variables and an intercept were included in the Z_n vector.

The X_i vector consisted of the levels of five attributes associated with the parks presented in the choice task. For the None option, the levels of the attributes were set equal to 0. These variables entered the latent segment model through their impact on the utility function. The attributes were effects coded as described by Louviere (1988) and Boxall and MacNab (2000). In addition, alternative specific constants (ASCs) for each park in the choice instrument were included in the utility function.

Estimation of the λ_s and β_s parameter vectors was performed via maximum likelihood methods using GAUSS. For the λ_s vector, the parameters for one of the segments must be normalized to zero to permit identification of segment membership parameters for the other segments. The log likelihood function was:

$$\ln L(\alpha, \beta|S) = \sum_{n=1}^{N} \sum_{m} \sum_{i\varepsilon C_m} \delta_{nmi} \ln \left(\sum_{s=1}^{S} \pi_{n|s}(i) \times \pi_{ns} \right) \tag{9}$$

where N refers to the 620 individuals, m represents the 4892 choice sets for which choice data were provided, i represents the alternatives from the choice experiment, and δ_{nmi} equals 1 if individual n chose i and 0 otherwise. The other symbols are described above. In this procedure independence was assumed across the set of choices from each respondent, and the scale parameters (α and μ_s) were set equal to 1.

In estimating latent segment models the number of segments, S, cannot be defined. Thus, S must be imposed by the investigator and statistical criterion must be used to select the "optimal" number of segments in a set of estimations where the number of segments imposed varies in each estimation. At issue in this process is that while one expects improvement in the log likelihood values as additional segments are added to the model, the model fits must be "penalized" for the increase in the number of parameters that are added due to additional segments. Thus, following Kamakura and Russell (1989), Gupta and Chintagupta (1994),

Table II. Information on the converged latent segment models and a random parameters logit model[a]

Number of segments	Number of parameters (P)	Log likelihood at convergence (LL)	Log likelihood evaluated at 0 (LL0)	ρ^{2b}	AIC[c]	BIC[d]
1	16	−7040.37	−8765.3	0.2	14112.74	7091.81
2	38	−6931.97	−8765.3	0.209	13939.94	7054.13
3	60	−6775.5	−8765.3	0.227	13671.04	6968.39
4	82	−6693.75	−8765.3	0.236	13551.5	6957.37
5	104	−6641.62	−8765.3	0.242	13491.24	6975.97
6	126	−6625.25	−8765.3	0.244	13502.50	7030.32
RPL model[e]	31	−6978.16	−8765.3	0.204	14018.32	7077.82

[a]Sample size is 4,892 choices from 620 individuals (N).
[b]ρ^2 is calculated as 1-(LL)/LL(0).
[c]AIC (Akaike Information Criterion) is calculated using {-2(LL-P)}.
[d]BIC (Bayesian Information Criterion) is calculated using {-LL+[P/2]$*$ln(N)]}.
[e]RPL represents the random parameters logit model.

Swait (1994), and Bhat (1999) two criteria were used to assist in determining the size of S. These were: the minimum Akaike Information Criterion (AIC), and the minimum Bayesian Information Criterion (BIC) (Allenby 1990). Their calculation is shown in the first row of Table II. As Swait (1994) describes, these criteria should be used as a guide to determine the size of S; conventional rules for this purpose do not exist and judgement and simplicity play a role in the final selection of the size of S.

For comparative purposes a random parameters logit (RPL) model was estimated using the information and software provided by Train (1998; http://elsa.berkely.edu/~train). For this model, we held the parameter on the variable Fees fixed (equal among individuals), and assumed the other parameters were normally distributed across the individuals in the sample.

5. Results and Discussion

5.1. CHOOSING THE NUMBER OF SEGMENTS

In estimating the latent segment models, 1, 2, 3, 4, 5 and 6-segment solutions were attempted. Table II summarizes the aggregate statistics for these models as well as the random parameters model. The log likelihood values at convergence (column 3) reveal improvement in the model fit as segments are added to the procedure, particularly with the 2, 3, and 4 segment models. This is evident in the ρ^2 values which increase from the base of 0.197 to 0.244 with the 6 segment model. The same statistics for the RPL model also suggest improvement over the single segment

Table III. Parameters (t statistics) on the segment membership variables for the four segment model

Variables	Segment 1 (*Escapists*)	Segment 2 (*Weekend Challengers*)	Segment 3 (*Nature Nuts*)	Segment 4 (*Wilderness Trippers*)
Intercept	0	−3.2688	−0.5596	0.1758
		(−15.224)	(−6.969)	(4.285)
Short Trip Dummy	0	7.0756	−0.7204	−4.1779
		(16.254)	(−4.704)	(−9.060)
Factor 1	0	1.8772	−0.3169	−0.0682
(challenge and freedom)		(14.611)	(−8.046)	(−2.544)
Factor 2	0	−0.0322[a]	0.6742	0.4701
(nature appreciation)		(−0.964)	(14.222)	(20.361)
Factor 3	0	1.1751	−0.6713	−0.6987
(social relationships)		(8.322)	(−19.593)	(−17.207)
Factor 4	0	−0.6633	−0.0501[a]	−0.0287[a]
(escape from routine)		(−18.168)	(−1.439)	(−0.826)

[a]Indicates that the parameter is **not** significantly different than 0 at the 5% level.

model. This information supports the existence of heterogeneity in the data. It also suggests the existence of latent segments, but does not suggest how many segments there are. The other statistics in Table II must be inspected to answer this question.

The following information supports four segments as the optimal solution in these data. First, the minimum BIC statistic is clearly associated with the 4 segment model. It is noteworthy that the BIC values rise when additional segments beyond four are added. Second, the AIC values grow smaller as the number of segments increases up to the 5th, suggesting that 5 segments is optimal. However, the change in AIC for the 4- to 5-segment solutions is markedly smaller than for the 3- to 4-segment solutions, suggesting that adding an additional segment beyond the 4th may not be gaining much improvement.

5.2. CHARACTERIZING THE SEGMENTS AND COMPARING THE MODELS

The segment membership (λ_s) parameters for the 4-segment solution are displayed in Table III. Note that the parameters for the first segment are equal to 0 due to their normalization during estimation. Thus, the other three segments must be described relative to this first segment. Segment 2 was labeled "weekend challengers" because the dummy variable on short or weekend-long trips was relatively large and positive, and the parameter on motivations relating to challenge and freedom was the same. For segment 3 the short trip dummy was close to zero, but the variable on motivations relating to nature appreciation was positive and was the largest over all 4 of the segments. For this reason, this segment was labeled "nature nuts". Segment 4 was classified as "wilderness trippers" because the short trip dummy variable was large and negative. Finally segment 1 was labeled "escapists" due to the fact that the motivational factor on escape from routine was negative for the other 3 segments. Despite the labels, however, the diversity of influences on segment membership is striking. For example, motivations relating to social relationships are positive for one segment, but negative for two others. Only for escape from routine and nature appreciation are the directions of the effects similar across segments.

The utility function parameters (β_s) for the 1-segment model, the RPL model, and the 4-segment model are displayed in Table IV.[5] The parameters on entry fees are negative for every model presented which is consistent with economic theory. Parameters for the chances of entry are variable across the models and segments, suggesting that this effect is characterized by heterogeneity. The 4-segment model implies that weekend challengers and wilderness trippers would seek parks with high chances of entry, while escapists and nature nuts prefer areas with low chances of entry. Individuals in these latter segments might choose places with lower chances of entry because these areas may offer the experiences they are seeking due to the restrictions on the number of visitors. These effects are absent in the single segment model which suggests that individuals would prefer areas with high chances of entry. For the RPL model, the parameter on the mean is close to that of the 1-segment model, but not close to any of those of the 4-segment one. It is noteworthy that the RPL model parameter on the SD of this variable is high which indicates high variation among individuals. This finding is congruent with the large variation in the parameters for Chance of Entry among the four segments (–0.1934 to 1.7413).

All three models suggest that individuals have preferences for the management related attributes, in particular the development and encounter attributes. The RPL and the 4-segment model, however, identify significant heterogeneity in the sample for these preferences. For example, the 1-segment model suggests that utility declines as encounter levels increase. The RPL parameters suggest the same pattern, but provide more information. For example, the insignificant parameter on the SD of the first development level indicates that every individual gains positive utility with that level. The SD parameters on the other levels suggest a large degree of variability around the utility levels with some perhaps having preferences over higher (rather than lower) encounter levels. The 4-segment model, however, provides information on the relative impacts of changing encounter levels. This model suggests that the Wilderness Trippers (segment 4) would be more negatively affected by higher levels of encounters than the other 3 segments.

The final set of utility parameters for the three models result from the ASCs used to identify the 5 parks (brands) in the choice experiment. Holding the attribute

Table IV. Parameters (t statistics) on the park attributes for three recreation site choice models

Variable	1 segment model	Random paramenters model		4 segment model			
		Mean	SD	Segment 1 (*Escapists*)	Segment 2 (*Weekend Challengers*)	Segment 3 (*Nature Nuts*)	Segment 4 (*Wilderness Trippers*)
Fee	–0.0562	–0.0851		–0.0714	–0.0594	–0.0931	–0.0813
	(–15.964)	(–10.378)		(–8.102)	(–5.360)	(–6.006)	(–7.099)
Chance of entry	0.4130	0.6899	1.6724	–0.1934	1.7413	–0.3731[1]	1.1772
	(6.981)	(4.949)	(4.976)	(–5.141)	(37.774)	(–1.170)	(10.518)
Campsite 1	–0.0012[1]	0.0494[1]	0.4986	–0.0843	0.1380	1.2275	–0.2220
	(–0.052)	(0.736)	(2.218)	(–2.088)	(3.770)	(24.498)	(–4.448)
Campsite 2	–0.0190[1]	–0.0068[1]	0.5481[1]	0.2150	–0.0408[1]	–0.9867	0.0612[1]
	(–0.670)	(–0.090)	(2.233)	(7.455)	(–1.001)	(–10.761)	(1.487)
Campsite 3	–0.2301	–0.4099	0.8095	–0.1718	–0.4488	0.2991	–0.4686
	(–8.903)	(–4.446)	(2.998)	(–4.701)	(–17.199)	(6.758)	(–13.999)
Level of	0.2419	0.5049	0.4361[1]	0.2058	–0.2214	2.8070	0.0979
development 1	(10.611)	(6.391)	(1.524)	(4.280)	(–7.054)	(24.624)	(2.993)
Level of	–0.4287	–0.6089	1.0165	–0.2133	–0.0976	–2.7642	–0.5581
development 2	(–24.032)	(–7.110)	(5.688)	(–6.449)	(–3.189)	(–36.340)	(–14.714)
Level of	–0.6188	–1.1142	1.3089	–0.5242	–0.2217	–3.6791	–0.8593
development 3	(–20.244)	(–8.698)	(5.470)	(–10.508)	(–6.668)	(10.371)	(–21.121)
Level of	0.5820	1.0008	0.0737[1]	0.5257	0.5561	0.4391	1.2153
encounters 1	(23.521)	(10.130)	(0.307)	(17.613)	(13.073)	(10.371)	(21.013)
Level of	–0.5097	–0.7904	0.9149	0.0415[1]	–0.2968	–0.9491	–1.4854
encounters 2	(–17.697)	(–6.527)	(5.021)	(0.840)	(–8.756)	(–16.085)	(–24.131)
Level of	–1.0289	–1.6838	0.8491	–0.6214	–0.7639	–0.7857	–2.1524
encounters 3	(–30.215)	(–8.783)	(2.089)	(–15.729)	(–25.327)	(–26.510)	(–44.715)
Woodland	–0.1405	–0.5441	0.1308[1]	0.7537	–0.3239	–4.5748	0.3804
Caribou ASC	(–2.229)	(–2.711)	(0.107)	(6.885)	(–7.755)	(–11.759)	(12.092)
Quetico ASC	0.8298	0.8077	1.2488	2.3845	–0.9966	–1.2438	0.8408
	(12.842)	(5.276)	(2.768)	(20.726)	(–30.451)	(–3.374)	(25.567)
BWCA ASC	0.2029	0.0177[1]	0.0642[1]	1.4675	–2.1021	–2.4624	–1.3463
	(3.245)	(0.083)	(0.177)	(15.724)	(–22.297)	(–28.112)	(–20.896)
Wabakimi ASC	–0.4917	–1.0892	0.0016[1]	–0.1798[1]	–0.9760	–2.9857	–0.1883
	(–7.903)	(–5.130)	(0.010)	(–1.148)	(–19.059)	(–8.764)	(–3.578)
Manitoba Parks	–0.4724	–1.0580	0.3816[1]	–2.8797	0.1172	–3.0494	–0.2970
ASC	(–7.328)	(–5.040)	(1.268)	(–23.760)	(2.423)	(–9.650)	(–10.077)

[1]Indicates that the parameter is **not** significantly different than 0 at the 5% level.

levels constant, the 1-segment model suggests positive preference for Quetico and the BWCA and negative preferences for the other 3 parks. The RPL model provides a different interpretation of the ASCs. Woodland Caribou, Wabakimi and the Manitoba Parks are unambiguously negative, while Quetico is positive but the effect is highly variable among individuals as evidenced by the large parameter on the SD. The mean and SD parameters for the BWCA ASC are not significant. The 4-segment model suggests that recreationists in segment 1 strongly prefer Quetico followed by the BWCA and Woodland Caribou parks. Weekend Challengers appear to prefer the Manitoba parks and not the other four parks. Segment 3 individuals exhibit negative parameters for all five parks, suggesting they may prefer parks not included in the choice experiment or are more likely to participate in other activities. Finally, individuals in segment 4 exhibit higher utility, all else being equal, for the two Ontario parks. These individuals also exhibit a negative association with the Boundary Waters. The significant negative effect of the BWCA in three of the four segments and the insignificance of the BWCA in the RPL model, suggests that the 4-segment model provides improved precision of the effect of that alternative on behavior.

Since segment membership parameters (λ_s) in the 4-segment model were jointly estimated with the utility parameters (β_s), one should expect consistent behavioral relationships among the two parameter vectors.[6] These features appear to be present in this dataset. For example, the trip choices of weekend challengers are positively influenced by recreation areas with higher chances of gaining entry; nature nuts are more likely to avoid areas with high levels of development; and wilderness trippers would seek areas where few other recreationists would be encountered. Thus, in this empirical example, the latent segment model appears to have identified sources of heterogeneity in recreation site choice and to have incorporated this by identifying different utility functions.

In order to complete the application of the framework proposed in Figure 1, estimates of park choices were calculated. This required knowledge of the attributes of the parks and placement of individuals in the segments. First, segment membership probabilities were computed for each of the 620 individuals using equation (6) and individuals were assigned to one of the four segments based on the largest probability.[7] This assignment method determined that 41.4% of the sample were members of segment 1, 7.3% were members of segment 2, 0.8% were members of segment 3 and the remaining 50.4% were assigned to segment 4. Thus, escapists and wilderness trippers dominate this sample of wilderness recreationists. However, the segment assignment is probabilistic and every respondent has a positive probability of being a member of each of the four segments. In the case of segment 3, for example, only 0.8% of the respondents had this segment as their highest probability of membership.

Second, the levels of the attributes of the five parks were determined using individual's perceptions of their attributes. For this, indicators of perceptions of campsite types, levels of development, and numbers of encounters were utilized from the questionnaire. These indicators provide linkage to the latent perceptions as outlined in the model (Figure 1). For fees and chances of entry the actual or objective levels of these were used for two reasons: (1) the objective and perceived levels of these two variables were identical for the majority of the 620 recreationists; and (2) Fees and entry restrictions are under the control of park managers who use them as policy measures to limit entry. These objective measures provide linkage to some of the actual park conditions shown in Figure 1.

5.3. WELFARE MEASURES IN THE LATENT SEGMENT MODEL

One of the roles of recreation economic choice models is to examine the welfare implications of environmental or management changes. Hanemann (1982) outlines the theory required for deriving welfare measures using conditional logit models. In what follows, this theory is applied to the latent segment model in two ways. The first involves the derivation of welfare measures on a segment by segment basis. In this case, the distributional impacts of policies can be understood. However, computing these welfare measures requires that respondents be assigned to a segment. The second way of applying this theory involves correcting the standard aggregate procedure, which assumes homogeneous preferences, for heterogeneity. Using this method, welfare measures are computed segment by segment for each individual, and these are then weighted by the individual's segment membership probabilities and summed to compute a welfare measure.

Hanemann (1982) shows that the expected utility on any given choice occasion is the sum of utility gained from each choice times its respective probability of being chosen. Thus, measuring a change in welfare associated with decreasing some attribute in the indirect utility function involves estimating the amount individuals must be compensated to remain at the same utility level as before the decrease. The following formula from Hanemann (1982) shows this calculation under the assumption of no income effects:

$$CV_n = \frac{1}{\gamma}\left[\ln\left(\sum_{k\in C}\exp(\beta X_k^0)\right) - \ln\left(\sum_{k\in C}\exp(\beta X_k^1)\right)\right] \qquad (10)$$

where CV_n is the compensating variation for individual n, γ is the marginal utility of income, βX_k represents the indirect utility over k choices, the 0 superscript refers to the initial state and the 1 superscript refers to the new state following some change in an attribute in X in at least one of the choices in k. Applying this formula to understand the distribution of welfare effects across segments necessitates the incorporation of segment-specific utility parameters and the assignment of individuals to segments. Hence:

$$CV_{n|s} = \frac{1}{\gamma_s}\left[\ln\left(\sum_{k\in C}\exp(\beta_s X_k^0)\right) - \ln\left(\sum_{k\in C}\exp(\beta_s X_k^1)\right)\right]. \qquad (11)$$

Table V. Compensating variation ($/trip) for some hypothetical changes in recreation quality for a representative individual in the sample[a]

Change	1 segment model	RPL model (at means)	4 segment model			
			Segment 1 (*Escapists*)	Segment 2 (*Weekend Challengers*)	Segment 3 (*Nature Nuts*)	Segment 4 (*Wilderness Trippers*)
Closure of Quetico Provincial Park	−12.86	−11.89	−17.68	−1.86	−9.49	−8.40
Increase congestion by 1 level:						
At Woodland Caribou Park	−1.52	−0.98	−0.53	−1.50	−0.07	−1.51
At Quetico Park	−4.58	−8.88	−4.86	−1.04	−6.24	−7.62
At BWCA	−0.58	−0.29	−1.05	−0.12	−0.18	−0.01
At Wabakimi Park	−0.94	−0.47	−0.18	−0.67	−0.28	−0.70
At Nopiming/Atikaki Parks	−2.61	−1.90	−0.03	−8.03	−1.19	−4.36
At all parks	−18.36	−20.36	−7.13	−14.21	−8.25	−33.05

[a]These estimates used the modal perception of campsite type, development, and number of encounters as the base case.

Extending this further to generate a welfare measure weighted by segment membership, equation (11) can be changed to:

$$CV_n = \sum_{s=1}^{S} \pi_s \left(\frac{1}{\gamma_s} \left[\ln \left(\sum_{k \in C} \exp(\beta_s X_k^0) \right) - \ln \left(\sum_{k \in C} \exp(\beta_s X_k^1) \right) \right] \right), \quad (12)$$

where π_s refers to the probability of membership is segment s.

Two welfare simulations were conducted. The first involved the hypothetical closure of Quetico Provincial Park. This scheme, while hypothetical, is not far-fetched as the portions of the park can be closed during severe forest fires, and in some cases entry to the entire park is prevented. The second simulation involved increasing congestion levels at each park, one at a time, and at all parks simultaneously. This scenario is related to the possibility that demand for experiences in these areas is increasing (Boxall et al. 1999) and would result in increasing levels of visitation and encounters between recreation parties in the backcountry. In both scenarios the base levels for the attributes in the utility function involved the actual levels of fees,[8] objective assessments by park managers of the chances of entry, and the modal perceptions of the three wilderness attributes.

The welfare impacts of these changes are shown in Table V for a representative recreationist in the sample for the single segment model, a representative recreationist for the RPL model using the means of the parameters, and representative recreationist in each segment for the 4 segment model. In these simulations, equation (10) was used for the single segment and RPL models and equation (11) for the 4-segment model. The 1-segment and RPL models suggest that the closure of Quetico has a fairly large negative impact. However, the latent class model shows that the closure has a larger impact on members of segments 1 and 3, and a relatively minor impact on members of segment 2. These results highlight the usefulness of the 4-segment model in understanding the distribution of welfare impacts.

Simulated increases in congestion also suggest distributional effects not captured by the simple model or the RPL model when evaluated at the means.[9] Increasing congestion at individual parks illustrates the effect of segments and substitution among the parks in the choice set. As a result the welfare differences between the 1 and 4 segment models are not remarkable except for those segments which exhibit preference for the park in which congestion is changed. However, the simulation for all parks highlights the effects of segmentation alone. In this case impacts are estimated at $-18.36/trip for the simple model, but the latent segmentation model suggests that the negative impacts of this scenario on wilderness trippers would be almost twice as much ($-33.05/trip). It would be about half as much for escapists ($-7.13) and nature nuts ($-8.25). It is noteworthy that the RPL welfare measure for the Quetico case is much higher than the 1-segment model and for any of the 4 segments as well. This is probably the result of the level and significance of the Quetico ASC in the RPL model. In addition, the implicit probability structure assumed in the RPL model (normal) and the probabilistic segment membership function generate very different descriptions of the distributions of the taste parameters. This difference is also likely to generate differences in welfare measures.

The weighted welfare measure (equation (12)) was examined by extracting a sub-sample of 17 individuals from the sample who provided complete information on the perceptions of campsite type, development, and congestion for all five parks.[10] The mean welfare loss for the closure of Quetico was estimated to be $-9.05/trip/person in this sub-sample and the individual welfare measures ranged from $-21.20 to $-3.47. The single segment welfare measure estimated the welfare loss for the same group of individuals at $-8.80/trip and the range was $-18.67 to $2.25. Thus, in this empirical examination failure to incorporate heterogeneity in the welfare measure associated with the closure policy would underestimate the value of the loss to the wilderness recreationists.

6. Conclusions

This paper was motivated by the need to simultaneously incorporate and explain sources of heterogeneity in random utility models. Current approaches in the literature involve simple parameterizations of the scale factor in conditional logit models or the random coefficients logit method proposed by Train (1998) and others. An alternative model proposed here involves the use of latent classes in concert with conventional random utility structure to explain choices. This latent segment model simultaneously groups individuals into relatively homogeneous segments and explains the choice behavior of segment members. A major advantage of this latent segment approach may be its ability to enrich the traditional economic choice model by including psychological factors. This integrated modeling strategy also offers an opportunity to merge various social psychological and economic theories in explaining behavior.

The results from this integrated approach provided a much richer interpretation of wilderness recreation site choice behavior than a traditional single segment model (which assumed homogeneity of preferences). For these data the latent segment approach suggested that heterogeneity was related to the motivational constructs underlying wilderness trips, sociodemographic characteristics, preferences for specific wilderness parks (holding changes in their characteristics constant), and perceptions of managerial attributes and congestion levels at the five parks. These findings support both economic and social psychological constructs related to wilderness recreation behavior. The RPL model estimated on these same data identified heterogeneity, but captured it in a different way. An example of this difference is the effect of the ASC on the BWCA in the two models. In the RPL case the parameter was insignificant, but in the latent segment model the parameter was significant in 3 of the 4 segments. This example highlights the advantage the latent segment model in understanding the distribution of the effects of management policies among members of a population.

The empirical application of latent segmentation to the wilderness data suggests that this method is promising in understanding recreation choice behavior. The method may be even more useful when applied to other types of recreation data, for example those in which the participants are more heterogeneous than are the wilderness recreationists examined in this study.[11] Regardless of the application, however, the underlying theory which incorporates latent psychometric constructs must be relevant to the activity being studied, and the indicator variables used to describe these constructs have acceptable explanatory power (see Ben-Akiva et al. 1997). The recreation literature abounds with theoretical and empirical studies on attitudes, perceptions, and motivations and would appear to offer fertile ground for further applications of the latent segment approach. For example, the success of the empirical application in this paper was related to prior existence of suitable instruments (see Beard and Ragheb 1983) to measure motivations for taking a trip.

This approach, however, has some difficulties. A major challenge in the use of choice models incorporating psychometric information is out-of-sample prediction. This is a problem because one generally does not know nor can predict the answers to attitudinal questions from those outside of the sample. In recreation contexts involving managed areas like parks, however, there is usually considerable information on the number and types of visitors who visit these areas. In these cases it may be possible to construct attitudinal instruments which, in concert with socioeconomic and experiential information, may be generalized to the recreation population of interest. In essence what is required is reasonable confidence in allocating out-of-sample individuals to segments and then using the segment-specific choice parameters to predict their behavior.

In the case of broader issues in which prediction to a more general population is of interest, the use of psychometric information may be problematic. Successful out-of-sample prediction in these instances will require the development of attitudinal questions and sufficient understanding of the answers to these before out-of-sample individuals can be allocated to segments with confidence. This represents a challenge to social science research agendas. In the absence of this kind of knowledge analysts may have to rely on the traditional sets of socioeconomic variables to understand membership in segments and their behavior.

Despite these issues, however, the research and methods reported in this paper are useful to policy makers because it outlines "who" comprises the population of interest and who would have been affected by policy changes. Our approach also confirms that understanding preference heterogeneity requires more information about respondents than the simple socio-demographic variables typically collected by researchers.

Acknowledgements

Funding from the Canada-Manitoba Partnership Agreement in Forestry, the Northern Ontario Development Agreement, and the Canadian Forest Service's core research budget is gratefully acknowledged. This manuscript was completed while Adamowicz was a Gilbert White visiting Fellows at Resources for the Future in Washington D.C. and he gratefully acknowledges their support. Thanks to Bonnie McFarlane, David Watson, Joffre Swait, Michel Haener, and Jeffrey Englin for assisting with the research, and Edward Morey and participants of the RTI seminar series for comments and suggestions.

Notes

1. These individual features are commonly used in the marketing, transportation, and tourism literatures to define various market segments (Wind 1978) and are usually referred to as psychometrics.

2. Registrations from the Boundary Waters Canoe Area were not available and thus visitors to this park could not be included in the sample.
3. We use "stay at home" as an alternative to allow respondents to choose none of the parks. In a study of branded choice alternatives in which the brands represent essentially the entire market, this is likely the best way to represent the "none" alternative. See Adamowicz et al. (1998) or Louviere et al. (2000) for further discussion on modeling the "none" alternative in choice experiments.Note also that this design incorporated a first and second choice question. Only the first choice was analyzed in this study. In addition, the order of the alternatives was the same in each scenario. We do not think there is an order effect in this presentation, but this is a topic for future research.
4. To explore the role of additional characteristics such as sociodemographic features in segment membership a posterior analysis of the characteristics of latent segments can be performed. This involves regressing the probabilities of segment membership against these characterisitics (e.g., Bucklin and Gupta 1992). This procedure was performed using our data and the results are reported in Boxall and Adamowicz (1999).
5. Since the choice experiment data we employ have a panel data structure some inefficiency may arise from not estimating this model in a panel framework. As a test, we examined conditional logit models in a panel framework and compared the results to the non-panel equivalent. We found very little difference in the coefficients or standard errors.
6. The posterior analysis described in footnote 4 and in Boxall and Adamowicz (1999) identified that high levels of experience in wilderness recreation are associated with the escapists, but that low levels of specialization in canoeing are associated with the Weekend Challengers. Residency in the USA is associated with them. Other factors such as household size, age, education levels also are associated with the various membership probabilities. An argument can be made for including these characteristics in the membership likelihood function. However, the computational effort required for adding these variables was beyond the scope of the computing resources available. Thus, in our empirical example it is assumed that these individual features are correlated with the variables included in the membership function.
7. It is important to recall that estimation of the membership function parameters occurred jointly with the estimation of the utility function or taste parameters. This means that the choice experiment information influenced the segment membership probabilities.
8. The parameters on fees (Table V) were chosen as the marginal utility of income parameter (γ). This choice was based on the fact that the distances between recreationists' homes and each of the five parks were not significant in explaining park choice in preliminary analyses of the choice experiment data. In turn, this was probably the result of incorporating ASCs in the model which confounded the estimation of the distance parameter.
9. Train (1998) describes other ways to generate welfare measures from the RPL model which capture distributional effects. These methods involve random draws from the parameter distributions.
10. These 17 individuals were chosen because they reported complete information for all of the required explanatory variables. These people did not appear to be a unique group in the sample. The mean (SD) probabilities of membership in each of the 4 segments over these 17 people were: 0.32 (0.10), 0.12 (0.17), 0.18 (0.08), and 0.38 (0.12) respectively. The max/min probabilities for each segment were 0.53/0.13, 0.64/0.001, 0.37/0.06, and 0.59/0.14.
11. An example of this heterogeneity may be participation in automobile camping in which equipment preferences, social and environmental settings, and facilities may drive site choice behavior.

Appendix

A summary of the factor analysis of attitudinal statements reflecting motivations for wilderness recreation in the system of wilderness parks.

Statement	Factor loadings			
	Factor 1 *(Challenge and freedom)*	Factor 2 *(Nature appreciation)*	Factor 3 *(Social relationships)*	Factor 4 *(Escape from routine)*
To challenge my skills and abilities	0.714	0.153	0.118	0.031
To develop my skills	0.636	0.145	0.206	0.100
To be in charge of a situation	0.635	0.056	0.039	0.190
To feel independent	0.573	0.267	0.091	0.094
To feel free from society's restrictions	0.501	0.071	0.093	0.434
To challenge nature	0.418	0.031	0.162	0.086
To be alone	0.395	0.188	–0.271	0.312
To feel close to nature	0.345	0.669	–0.001	–0.001
To observe the beauty of nature	0.050	0.660	0.014	0.142
To obtain a feeling of harmony with nature	0.329	0.632	0.037	0.011
To find quiet places	0.076	0.579	0.003	0.26
To enjoy the sights, sounds, and smells of nature	0	0.567	0.006	0.103
To be with my friends or family	–0.024	0.023	0.746	0.063
To strengthen relationships with friends or family	0.140	0.120	0.666	0.059
To do things with other people	0.183	–0.109	0.665	0.109
To be with people with similar interests	0.304	0.029	0.533	0.090
To escape from the pressures of work	0.043	0.125	0.153	0.708
To relieve my tensions	0.250	0.132	0.049	0.667
To get away from my everyday routine	0.080	0.101	0.221	0.649
To be away from other people	0.278	0.225	–0.239	0.431
Eigenvalues	4.619	1.989	1.263	1.180

References

Adamowicz, W. L., J. J. Louviere and M. Williams (1994), 'Combining Stated and Revealed Preference Methods for Valuing Environmental Amenities', *Journal of Environmental Economics and Management* **26**, 271–296.

Adamowicz, W. L., J. D. Swait, P. C. Boxall, J. J. Louviere and M. Williams (1997), 'Perceptions Versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation', *Journal of Environmental Economics and Management* **32**, 65–84.

Adamowicz, W. L., J. Louviere and J. Swait (1998), Introduction to Attribute-based Stated Choice Methods. Final Report to Resource Valuation Branch, Damage Assessment Center, NOAA, U.S. Department of Commerce. January. 44 pp.

Allenby, G. (1990), 'Hypothesis Testing with Scanner Data: The Advantage of Bayesian Methods', *Journal of Marketing Research* **27**, 379–389.

Beard, J. G. and M. G. Ragheb (1983), 'Measuring Leisure Motivation', *Journal of Leisure Research* **15**, 219–228.

Ben-Akiva, M. and J. Swait (1986), 'The Akaike Likelihood Ratio Index', *Transportation Science* **20**, 133–136.

Ben-Akiva, M., J. Walker, A. T. Bernardino, D. A. Gopinath, T. Morikawa and A. Polydoropoulou (1997), Integration of Choice and Latent Variable Models. Paper presented at the 1997 IATBR, University of Texas at Austin.

Boxall, P. C. and W. L. Adamowicz (1999), Understanding Heterogeneous Preferences in Random Utility Models: The Use of Latent Class Analysis. Staff Paper 99-02, Department of Rural Economy, University of Alberta, Edmonton, Alberta.

Boxall, P. C. and B. MacNab (2000), 'Exploring The Preferences of Wildlife Recreationists for Features of Boreal Forest Management: A Choice Experiment Approach', *Canadian Journal of Forest Research* **30**, 1931–1941.

Boxall, P. C., J. Englin and D. O. Watson (1999), 'Valuing Backcountry Recreation in Wilderness Parks: A Demand Systems Approach in The Canadian Shield', *Information Report* NOR-X-361, Northern Forestry Centre, Canadian Forest Service, Edmonton, Alberta.

Bucklin, R. E. and S. Gupta (1992), 'Brand Choice, Purchase Incidence and Segmentation: An Integrated Modeling Approach', *Journal of Marketing Research* **20**, 201–215.

Cameron, T. C. and J. Englin (1997), 'Respondent Experience and Contingent Valuation of Environmental Goods', *Journal of Environmental Economics and Management* **33**, 296–313.

Crandall, R. (1980), 'Motivations for Leisure', *Journal of Leisure Research* **12**, 45–54.

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

Hanemann, W. M. (1982), Applied Welfare Analysis With Qualitative Response Models. Working Paper No. 241. University of California, Berkeley, 26 pp.

Kamakura, W. and G. Russell (1989), 'A Probabilistic Choice Model for Market Segmentation and Elasticity Structure', *Journal of Marketing Research* **26**, 379–390.

Layton, D. F. (1996), Rank-Ordered, Random Coefficients Multinomial Probit Models for Stated Preference Surveys. Paper Presented at the 1996 Association of Environmental and Resource Economists Workshop, June 2–4, Tahoe City, California.

Louviere, J. J., D. A. Hensher, and J. D. Swait (2000), *Stated Choice Methods: Analysis and Application.* Cambridge, UK: Cambridge University Press.

McCutcheon, A. L. (1987) *Latent Class Analysis.* Sage University Papers Series: Quantitative Applications in the Social Sciences Number 07-064. Newbury Park, CA: Sage Publications.

McFadden, D. (1974), 'Conditional Logit Analysis Of Qualitative Choice Behavior', in P. Zarembka, ed., *Frontiers in Econometrics.* New York, NY: Academic Press, pp. 105–142.

McFadden, D. (1986), 'The Choice Theory Approach to Market Research', *Marketing Science* **5**, 275–297.

Morey, E. R., R. D. Rowe, and M. Watson (1993), 'A Repeated Nested-Logit Model of Atlantic Salmon Fishing', *American Journal of Agricultural Economics* **75**, 578–592.

Pollack, R. A. and T. J. Wales (1992), *Demand System Specification and Estimation.* New York, NY: Oxford University Press.

Salomon, I. and M. Ben-Akiva (1983), 'The Use of the Life-Style Concept in Travel Demand Models', *Environmental and Planning A* **15**, 623–638.

Schmidt, P. and R. Strauss (1975), 'The Predictions of Occupation Using Multinomial Logit Models', *International Economic Review* **16**, 471–486.

Shonkwiler, J. S. and W. D. Shaw (1997), Shaken, Not Stirred: A Finite Mixture Approach To Analysing Income Effects in Random Utility Models. Paper Presented at the 1997 Annual Meeting of the American Agricultural Economics Association. August 2–4, Toronto, Ontario.

Swait, J. R. (1994), 'A Structural Equation Model of Latent Segmentation and Product Choice for Cross-Sectional Revealed Preference Choice Data', *Journal of Retailing and Consumer Services* **1**, 77–89.

Swait, J. R. and J. J. Louviere (1993), 'The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models', *Journal of Marketing Research* **30**, 305–314.

Titterington, D. M., A. F. M. Smith and U. E. Makov (1985), *Statistical Analysis of Finite Mixture Distributions*. New York, NY: John Wiley & Sons.

Train, K. E. (1997), 'Mixed Logit Models for Recreation Demand', in C. Kling and J. Herriges, eds., *Valuing the Environment Using Recreation Demand Models*. Edward Elgave: Lyme New Hampshire.

Train, K. E. (1998), 'Recreation Demand Models with Taste Differences over People', *Land Economics* **74**, 230–239.

Wind, Y. (1978), 'Issue and Advances in Segmentation Research', *Journal of Marketing Research* **15**, 317–337.

Yatchew, A. and Griliches, Z. (1984), 'Specification Error in Probit Models', *Review of Economics and Statistics* **66**, 134–139.