RURAL ECONOMY

Understanding Heterogeneous Preferences in Random Utility Models: The Use of Latent Class Analysis

Peter C. Boxall and Wiktor L. Adamowicz

Staff Paper 99-02

STAFF PAPER

Department of Rural Economy
Faculty of Agriculture, Forestry
and Home Economics
University of Alberta
Edmonton, Canada

Understanding Heterogeneous Preferences in Random Utility Models:
The Use of Latent Class Analysis

Peter C. Boxall and Wiktor L. Adamowicz

Staff Paper 99-02

The authors are, respectively, Associate Professor and Professor, Department of Rural Economy, University of Alberta, Edmonton.

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. Thanks also to Bonnie McFarlane, David Watson, Joffre Swait, and Jeffrey Englin for assisting with this research.

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 as well as for forecasting demand by including individual characteristics and providing a broader picture of the distribution of resource use decisions or policy impacts. However, many empirical economic analyses assume homogeneous preferences among consumers. Alternatively, previous analyses considered 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 demographic variables.

There is empirical evidence that these methods identify sources of heterogeneity. For example, Famulari (1995) showed that stratifying households by demographic categories significantly improved tests of consistency with the axioms of revealed preference. Boxall et al. (1996) studied recreation demand in a traditional travel cost model framework[1] and found that stratification reduced the percentage and mean error of violations of the choice axioms. These studies examined traditional demand analysis, rather than considering the individual choice behaviour typified by the random utility framework. Also there are few economic studies which examine individual-specific variables other than sociodemographic factors.

[1] The traditional travel cost model considers price quantity data gathered from a sample of recreationists treating these as if this data were generated by a single consumer maximizing a constrained utility function.

Heterogeneity is particularly difficult to examine in the random utility model because an individual's characteristics are invariant among a set of choices. In an econometric sense this means that the effect of individual characteristics are not identifiable in the probability of choosing commodities. In essence, the model parameters are the same for each sampled individual implying that different people have the same tastes over model components. These features have been examined 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 in a sense for heterogeneity by allowing model parameters to vary randomly over individuals (e.g. Layton 1996; Train 1997, 1998). While these procedures explicitly 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 factor. Cameron and Englin (1997) explain heterogeneity by "parameterizing" this 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 the approach of Cameron and Englin, in which individual characteristics enter as affecting scale, is more appropriate than an alternative approach in which these features 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.[2] Furthermore, while theory may provide an understanding of sources of heterogeneity, it would also be desirable to incorporate heterogeneity in the estimation of economic choice parameters. These features point to joint estimation of the explanators of heterogeneity and the explanators associated with attributes of choices.

---

[2] These individual features are commonly used in the marketing, transportation, and tourism literatures to define various market segments (Wind 1978). These types of features are usually referred to as psychographics.

A promising avenue for tackling these problems involves the use of latent variables. Latent variable approaches involve the logic that some unobserved or latent variable explains the behaviour of interest, but that one can only observe indicators (in essence other variables) which are functionally related to the latent variable. An analyst using this approach must assume that covariation among a set of observed variables is due to each variable's relationship to the latent variable. That is, the latent variable is actually the true source of the covariation.

The strategy of use of latent variables can be extended to consideration of latent classes. In this approach the latent construct 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 a set of discrete observed measures such as attitudinal scales, or they 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.

The Latent Segmentation Approach

McFadden (1986) recognized the prospect of using latent variables in understanding choice behaviour. 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 behaviour, 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 behaviour and then simulate this choice behaviour 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 psychographic or socioeconomic effects, but also elements of observed behaviour. This type of model may have considerable relevance to decision-makers in that it allows a degree of 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.

This 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 aspects of this study use information from a choice experiment involving wilderness park choice. In this experiment five environmental and managerial attributes were varied in the design. 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.

THE LATENT SEGMENTATION MODEL

In deriving the latent segmentation model, random utility theory is first employed to model choices among a set of substitutes or alternatives on a given choice occasion with each choice occasion assumed to be independent of the others. 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, $U_{ni}=V_{ni}+\epsilon_{ni}$ 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:

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

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 take the form:

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

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 $V_{in|s} = \beta_s X_{in} + \epsilon_{in|s}$. In this expression the utility parameters are now segment specific and equation (2) becomes:

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

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 can classify individuals into one of the S segments. The classification variables that influence 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 psychographic 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 psychographic 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:

$$M^*_{ns} = \lambda_s Z_n + \zeta_{ns}, \; s = 1, \ldots, S \tag{5}$$

where Z_n is a vector of both the psychographic constructs (P_n) and socioeconomic characteristics (S_n), and λ_s is a vector of parameters. This classification mechanism allows n to be placed in s if and only if :

$$M^*_{ns} = \max\{M^*_{nj}\}, \; j \neq s, \; s = 1, \ldots, S. \tag{6}$$

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:

$$\pi_{ns} = \frac{\exp \alpha (\lambda_s Z_n)}{\sum_{s=1}^{S} \exp \alpha (\lambda_s Z_n)}, \qquad \textbf{(7)}$$

This form 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, the following constraints must be met:

$$\sum_{s=1}^{S} \pi_{ns} = 1; \quad \text{and} \quad 0 \leq \pi_{ns} \leq 1. \qquad \textbf{(8)}$$

To further develop the latent segment model 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 (7): $\pi_{ns}(i) = \pi_{ns}$

$\pi_{n|s}(i)$. Thus, the probability that a randomly chosen individual n chooses i is given by:

$$\pi_n(i) = \sum_{s=1}^{S} \pi_{ns} \ \pi_{n|s}(i) \, , \qquad \textbf{(9)}$$

and substituting the equations for the choice (equation 3) and membership (equation 7) probabilities yields the expression:

$$\pi_n(i) = \sum_{s=1}^{S} \left[\frac{\exp \alpha(\lambda_s Z_n)}{\sum_{s=1}^{S} \exp \alpha(\lambda_s Z_n)}\right]\left[\frac{\exp \mu_s(\beta_s X_i)}{\sum_{k \in C} \exp \mu_s(\beta_s X_k)}\right]. \qquad \textbf{(10)}$$

This model allows the use of both choice attribute data and individual consumer characteristics to simultaneously explain choice behaviour. Note that the expression contains two types of logit formulations; one is a multinomial logit model which includes the segment membership parameters and the other is a conditional logit model which contains the segment specific utility parameters. Because these two formulations are mixed together this model is considered to be a mixed logit model in the literature (e.g. Titterington et al. 1985).

A number of features of this model are noteworthy. First, the observation that the ratio of probabilities of selecting any two alternatives (equation 10) would contain arguments that include the systematic utilities from 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 factor 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 used here since a researcher would **not** want to impose utility parameter equality. Therefore, utilizing this model in empirical estimation requires that all of the scale factors in (10) are set equal to one.

Third, as Swait (1994) points out when $\lambda_s = 0$, $\beta_s = \beta$, and $\mu_s = \mu$ for each segment, equation (10) 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 in a set of data can be considered a segment. Under this condition each respondent behaves as if their behaviour 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 models (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. On the other end is the case where each individual is considered a segment in which heterogeneity of preferences is, in a sense, completely accounted for. The potential advantage of the latent segment model in this series of approaches is its potential to explain and account for heterogeneity to some degree.

AN APPLICATION - WILDERNESS PARK CHOICE IN CENTRAL CANADA

A Framework for Wilderness Recreation Decisions

In understanding the selection of wilderness areas for recreation trips, a framework of choice 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 psychographic 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 (7).

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 behaviour 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 psychographic and sociodemographic characteristics together with the objective and subjective wilderness attributes enables the implementation of the latent segmentation model. Thus, the decision protocol is represented through equation (10). 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.

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; Englin et al. 1998) 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. They cater to a relatively heterogeneous market 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.[3] 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 five 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 which 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

[3] Registrations from the Boundary Waters Canoe Area were not available and thus visitors to this park could not be included in the sample.

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 “branded” choice) where the two Manitoba parks were combined into one and the Boundary Waters Canoe Area was available as one of the choices. Five 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 1.

A choice scenario (Figure 2) consisted of six alternatives (five parks and the stay at home option).[4] Statistical design methods 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 attributes and levels for these four parks were constructed from a 4^5 x 4^5 x 4^5 x 4^5 x 2 orthogonal main-effects design, yielding 64 possible combinations of the levels (or choice sets). 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

[4] Note that this design incorporates two choice questions. Only the first choice was analyzed in this study.

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 1.

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 the analysis of the motivational indicators. This entailed a factor analysis of the 20 statements on reasons for taking a trip. The factor analysis provided estimates of the latent motivational constructs which enter the membership likelihood function. Because these statements were developed *a priori* to assess motivations, this involved a confirmatory approach. 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 shown in Table 2. The first component was called “challenge and freedom” because statements relating to this factor loaded highly in this factor (shaded grey in Table 2). 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 (10). 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 degree of estimation required limited the set of variables for inclusion. However, to explore the role of sociodemographic features in segment membership, a posterior analysis of the characteristics of latent segments was performed. This will be described below. Thus, in summary five variables and an intercept were included in the Z_n vector.

The X_i vector consisted of the attributes associated with the parks presented in the choice task. These variables entered the latent segment model through their impact on the utility function. Recall that there are five attributes, each with four levels. The attributes were effects coded as described by Louviere (1988) and Adamowicz et al. (1994).

Estimation of the λ_s and β_s parameter vectors was performed via maximum likelihood in GAUSS using the BFGS algorithm. 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:

$$\ln L(\lambda, \beta \,|\, S) \;=\; \sum_{n=1}^{N} \sum_{\forall m} \delta_{ni} \ln\Big[\sum_{s=1}^{S} \frac{\exp \alpha(\lambda_s Z_n)}{\sum_{s=1}^{S} \exp \alpha(\lambda_s Z_n)} \times \frac{\exp \mu_s(\beta_s X_i)}{\sum_{i=1}^{6} \exp \mu_s(\beta_s X_i)}\Big] \qquad \textbf{(11)}$$

where N refers to the 620 individuals who provided complete information, m represents the 4892 choice sets for which choice data were provided, i represents the alternatives from the choice experiment, and δ_{ni} 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) and Swait (1994) three criteria were used to assist in determining the size of S. These were: the minimum Akaike Information Criterion (AIC), the minimum Bayesian Information Criterion (BIC) (Allenby 1990), and the maximum of a modification of McFadden's ρ^2 called the Akaike Likelihood Ratio index (or ρ bar^2)(see Ben-Akiva and Swait 1986). Their calculation is shown in the first row of Table 3. 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.

RESULTS AND DISCUSSION

Choosing the Number of Segments

In estimating the models, 1, 2, 3, 4, 5 and 6-segment solutions were attempted. Table 3 summarizes the aggregate statistics for these models as well as a single segment 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. This information supports the hypothesis of the existence of latent segments, but does not suggest how many segments are in the data. The other statistics in Table 3 must be inspected to answer this question.

Inspection of columns 5 to 7 in Table 3 support four segments as the optimal solution in the data. First, while the AIC values grow smaller as the number of segments increases, the change in AIC is markedly smaller for the 4- to 5-segment and 5- to 6-segment solutions than the 2- to 3- and 3- to 4-segment solutions. Second, the ρbar^2 statistics exhibit a similar pattern in that improvement in the values is reduced beyond the 4-segment model. Finally, 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.

Characterizing the Segments

The segment membership (λ_s) parameters for the 4-segment solution are displayed in Table 4. Note that the parameters for the first segment are equal to 0 which results from 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 4-segment solution are displayed in Table 5. Also shown are the parameters for a 1 segment solution for comparison. The parameters on entry fees are negative for each segment which is consistent with economic theory. Parameters for the chances of entry are variable across the 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 can be compared to the single segment model in which suggests all individuals would prefer areas with high chances of entry.

The utility parameters associated with the campsite type, levels of development and encounter variables were plotted in Figure 3 to show the differences among the segments. These plots identify the sensitivity of segments to changes in the levels of these three sets of variables. Campsite type seems not to be a large influence on park choice for any of the four segments. Yet it is noteworthy that in the single segment model, two of the campsite parameters are not statistically significant while these parameters are significant in each of the four segments. However, development and encounter levels appear to have an important effect on choice behaviour. Recreationists in segment 3 (nature nuts) would be strongly negatively affected by higher levels of development. Wilderness trippers (segment 4) would be more negatively affected by higher levels of encounters than the other 3 segments.

A final set of utility parameters result from the alternative specific constants (ASCs) used to identify the 5 parks (brands) in the choice experiment. These parameters are shown in Table 5. Recreationists in segment 1 strongly prefer Quetico followed by the

BWCA and Woodland Caribou parks. Segment 2 individuals, the weekend challengers, only prefer the Manitoba parks and would tend to avoid the other four parks, all other things being equal. Segment 3 individuals exhibit negative parameters for all five parks suggesting that they may prefer parks not included in the choice experiment or are more likely to participate in other activities. This seems to be an odd result and will be addressed further below. 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. Once again, comparison of the alternative specific constants with the single-segment model suggests that the simpler model would not capture sources of heterogeneity associated with the latent segment model.

Since segment membership parameters (λ_s) were jointly estimated with the utility parameters (β_s), one should expect consistent behavioural relationships among the two parameter vectors. These features appear to be present in this dataset. For example, the trip choices of weekend challengers should be 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 should 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.

The role of sociodemographic characteristics in explaining latent class membership (from Fig. 1) was explored by computing segment membership probabilities for each individual and then regressing these four probabilities against the individuals' characteristics.[5] In this procedure, the method of Bucklin and Gupta (1992) was used in which the probabilities were transformed by the following formula: $\ln(\pi_s /1-\pi_s)$. It must be recognized that the error terms for the four regressions would be correlated and that seemingly unrelated regression (SUR) methods should be used to improve the efficiency of the parameter estimates. However, this would require the addition of parameter restrictions which requires theoretical justification. This is a topic for future research and thus equation by equation OLS methods were used to provide an illustration of the role of these individual characteristics.

The results of these regressions (Table 6) identify 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 the weekend challengers. Other factors such as household size, age, education levels also are associated with the various membership probabilities.

In order to complete the application of the framework proposed in Fig.1 for the wilderness recreation data, 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 (7). Each individual was assigned to one of the four segments based on the largest probability. This assignment method determined that 41.4% of the sample were

---

[5] 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 resources available. Thus in this example it is assumed that these individual features are correlated with the variables included in the membership function (Table 4).

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 the sample of wilderness recreationists.

Second, the levels of the attributes of the five parks were determined using individual's perceptions of their attributes as outlined in Figure 1. For this, indicators of perceptions of campsite types, levels of development, and numbers of encounters were utilized from the questionnaire. The questions used to collect this information, and a summary of the results, are shown in Appendix 1. For fees and chances of entry, the objective levels of these were used. For the majority of the 620 recreationists, the objective and perceived levels of these two variables were identical. These indicators provide linkage to the latent perceptions as outlined in the model (Fig.1).

Figure 4 displays the predicted distribution of wilderness recreation choices among the five parks by segment. Individuals in segments 1 and 3 would be more likely to visit Quetico than the other two segments. Those in segment 2 would be more likely to visit the Manitoba Parks. The other parks appear to be less attractive to these members of segment 2, particularly the Boundary Waters and Wabakimi. These findings have implications for identifying the relevant choice sets across the segments, but these results are not as strong as those reported by Swait (1994).

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. McFadden (1996) 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, in a sense, 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 segment membership probabilities and summed to compute a total welfare measure.

McFadden (1996) and Hanemann (1982) show 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:

$$CV_n = \frac{1}{\gamma}\left[\ln\left(\sum_{k \in C} \exp(\beta X_k^{0})\right) - \ln\left(\sum_{k \in C} \exp(\beta X_k^{1})\right)\right] \quad \textbf{(12)}$$

where CV_n is the compensating variation for individual n, γ is the marginal utility of income, βX_k represents the indirect utility function 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:

$$CV_{n|s} = \frac{1}{\gamma_s}\left[\ln\left(\sum_{k \in C} \exp(\beta_s X_k^{0})\right) - \ln\left(\sum_{k \in C} \exp(\beta_s X_k^{1})\right)\right]. \qquad \textbf{(13)}$$

Employing this further to generate an aggregate welfare measure, weighted by segment membership, can be calculated by:

$$CV_n = \sum_{s=1}^{S} \pi_s \left(\frac{1}{\gamma_s}\left[\ln\left(\sum_{k \in C} \exp(\beta_s X_k^{0})\right) - \ln\left(\sum_{k \in C} \exp(\beta_s X_k^{1})\right)\right]\right) \qquad \textbf{(14)}$$

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

The parameters on fees (Table 5) 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 exploratory analyses of the choice experiment data. In turn, this was probably the result of the alternative specific constants in the model confounding the distance parameter. Thus, fees were considered the most appropriate price variable associated with a trip in this sample.

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 can be completely 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, objective assessments by park managers of the chances of entry, and the modal perceptions of the three wilderness attributes used in estimating the park choices shown in Figure 4.

The welfare impacts of these changes are shown in Table 7 for a representative recreationist in the sample for the single segment model and in each segment for the 4 segment model. In these simulations equation (12) was used for the former and equation (13) for the latter. The results highlight the limitations of single segment models in understanding the distribution of welfare impacts. For example, the closure of Quetico has a larger impact on members of segments 1 and 3, and a relatively minor impact on members of segment 2 in comparison to the single segment case.

Simulated increases in congestion also suggest distributional effects not captured by the simple model. 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 two 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).

The weighted welfare measure (equation (14)) 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[6]. 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 probably underestimate the value of the loss to the wilderness recreationists.

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 class analysis in concert with the

[6] 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.

conventional random utility structure to explain choices. This latent segment model simultaneously groups individuals into relatively homogeneous segments and explains the choice behaviour of the segments. A major advantage of this latent segment approach may be its ability to enrich the traditional economic choice model by including psychological factors. However, this integrated modeling strategy also offers an opportunity to merge various social psychological and economic theories in explaining behaviour.

To illustrate these features, a latent segment model was developed and applied to recreation demand in a set of wilderness parks by a sample of 620 people. The theoretical basis for this involved the incorporation of sociodemographics and latent constructs relating to motivations in describing segment membership. These constructs were integrated using indicators derived from survey responses related to reasons for taking a wilderness trip. The development of the utility function involved recreation site choice attributes which were examined in a choice experiment.

The results from this integrated approach provided a much richer interpretation of wilderness recreation site choice behaviour 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 behaviour.

The latent segment model may be at a considerable advantage in adding to understanding the distribution of the effects of management policies among members of a population. To illustrate this, three welfare measures were developed. The first is the form frequently used in the empirical literature (Hanemann 1982) in which homogeneous preferences are assumed. This welfare measure is relevant for the single segment case. The other two welfare measures were variants of this case and explicitly included segment differences. One measure was employed to assess welfare impacts in each segment and the results were used to examine the distributional impacts of policies across segments. The other welfare measures utilized the probability of segment membership and used this probability to adjust the weights of the segment welfare impacts and generated a single welfare measure. The resulting welfare measures from these latter two approaches were quite different than the single segment case.

The empirical application of latent segmentation to the wilderness data suggests that this method holds considerable promise in understanding recreation choice behaviour. 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.[7] Regardless of the application, however, the underlying theory which incorporates latent psychographic 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

[7] 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 behaviour.

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.

A major challenge in the use of choice models incorporating psychographic 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 behaviour.

In the case of broader issues in which prediction to a more general population is of interest, the use of psychographic 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 considerable 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 behaviour.

LITERATURE CITED

Adamowicz, W.L., 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. Swait, P.C. Boxall, 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.

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., W.L. Adamowicz, and T. Tomasi. 1996. Nonparametric tests of the traditional travel cost model. *Canadian Journal of Agricultural Economics* 44:183-193.

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.

Famulari, M. 1995. A household-based, nonparametric test of demand theory. *Review of Economics and Statistics* :372-382.

Hanemann, W.M., 1982. Applied welfare analysis with qualitative response models, Working Paper. No. 241. University of California, Berkeley, 26 pp.

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.

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.

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

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In: P. Zarembka (Editor), *Frontiers in Econometrics*. Academic Press, New York, pp. 105-142.

McFadden, D. 1986. The choice theory approach to market research. *Marketing Science* 5:275-297.

McFadden, D. 1996. Computing willingness to pay in random utility models. *Energy and Resource Economics* (forthcoming).

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*. Oxford Univ. Press, New York.

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.

Shocker, A.D., M. Ben-Akiva, B. Boccara, and P. Nedungadi. 1997. Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing Letters* 2:181-197.

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*. John Wiley & Sons, New York, NY.

Train, K.E. 1997. Mixed logit models for recreation demand. In: C. Kling and J. Herriges (editors) *Valuing the Environment using Recreation Demand Models*, Edward Elgave, Lyme New Hampshire. (Forthcoming)

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

Wind, Y. 1978. Issue and advances in segmentation research. *Journal of Marketing Research* 15:317-337.

TABLE 1
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

TABLE 2

Distribution and Factor Analysis of Attitudinal Statements reflecting Motivations for Wilderness Recreation in the System of Wilderness Parks

	Factor Loadings			
Variable	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	0.1
To be in charge of a situation	0.635	0.056	0.039	0.19
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	0.05	0.66	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	0.14	0.12	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	0.09
To escape from the pressures of work	0.043	0.125	0.153	0.708
To relieve my tensions	0.25	0.132	0.049	0.667
To get away from my everyday routine	0.08	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	1.18

TABLE 3

Latent Segment Models with Factor Scores, Intercepts and Trip Dummy as Segment Membership Variables [a]

Number of Segments	Number of Parameters (P)	Log Likelihood at Convergence (LL)	Log Likelihood evaluated at 0	ρ^2	AIC [b]	ρbar^2 [c]	BIC [d]
1	16	-7040.37	-8765.30	0.197	14112.74	0.195	7091.81
2	38	-6931.97	-8765.30	0.209	13939.90	0.205	7054.13
3	60	-6775.50	-8765.30	0.227	13671.04	0.220	6968.39
4	82	-6693.75	-8765.30	0.236	13551.50	0.227	6957.37
5	104	-6641.62	-8765.30	0.242	13491.24	0.230	6975.97
6	126	-6625.25	-8765.30	0.244	13502.50	0.230	7030.32

[a] Sample size is 4,892 choices from 620 individuals (N)

[b] AIC (Akaike Information Criterion) is calculated using {-2(LL-P)}

[c] ρbar^2 is calculated using {1-AIC/2LL(0)}

[d] BIC (Bayesian Information Criterion) is calculated using {-LL + [P/2)*ln(N)]}

TABLE 4

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

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

TABLE 5

Parameters (*t* statistics) on the Variables for Two Recreation Site Choice Models

Variable	1 Segment Model	4 Segment Model			
		Segment 1 (*Escapists*)	Segment 2 (*Weekend Challengers*)	Segment 3 (*Nature Nuts*)	Segment 4 (*Wilderness Trippers*)
Fee	-0.0562 (-15.964)	-0.0714 (-8.102)	-0.0594 (-5.360)	-0.0931 (-6.006)	-0.0813 (-7.099)
Chance of Entry	0.4130 (6.981)	-0.1934 (-5.141)	1.7413 (37.774)	-0.3731[1] (-1.170)	1.1772 (10.518)
Campsite 1	-0.0012[1] (-0.052)	–0.0843 (-2.088)	0.1380 (3.770)	1.2275 (24.498)	-0.2220 (-4.448)
Campsite 2	-0.0190[1] (-0.670)	0.2150 (7.455)	-0.0408[1] (-1.001)	-0.9867 (-10.761)	0.0612[1] (1.487)
Campsite 3	-0.2301 (-8.903)	-0.1718 (-4.701)	-0.4488 (-17.199)	0.2991 (6.758)	-0.4686 (-13.999)
Level of development 1	0.2419 (10.611)	0.2058 (4.280)	-0.2214 (-7.054)	2.8070 (24.624)	0.0979 (2.993)
Level of development 2	-0.4287 (-24.032)	-0.2133 (-6.449)	-0.0976 (-3.189)	-2.7642 (-36.340)	-0.5581 (-14.714)
Level of development 3	-0.6188 (-20.244)	-0.5242 (-10.508)	-0.2217 (-6.668)	-3.6791 (10.371)	-0.8593 (-21.121)
Level of encounters 1	0.5820 (23.521)	0.5257 (17.613)	0.5561 (13.073)	0.4391 (10.371)	1.2153 (21.013)
Level of encounters 2	-0.5097 (-17.697)	0.0415[1] (0.840)	-0.2968 (-8.756)	-0.9491 (-16.085)	-1.4854 (-24.131)
Level of encounters 3	-1.0289 (-30.215)	-0.6214 (-15.729)	-0.7639 (-25.327)	-0.7857 (-26.510)	-2.1524 (-44.715)
Woodland Caribou ASC	-0.1405 (-2.229)	0.7537 (6.885)	-0.3239 (-7.755)	-4.5748 (-11.759)	0.3804 (12.092)
Quetico ASC	0.8298 (12.842)	2.3845 (20.726)	-0.9966 (-30.451)	-1.2438 (-3.374)	0.8408 (25.567)
BWCA ASC	0.2029 (3.245)	1.4675 (15.724)	-2.1021 (-22.297)	-2.4624 (-28.112)	-1.3463 (-20.896)
Wabakimi ASC	-0.4917 (-7.903)	-0.1798[1] (-1.148)	-0.9760 (-19.059)	-2.9857 (-8.764)	-0.1883 (-3.578)
Manitoba Parks ASC	-0.4724 (-7.328)	-2.8797 (-23.760)	0.1172 (2.423)	-3.0494 (-9.650)	-0.2970 (-10.077)

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

TABLE 6

Posterior Analysis of the Segment Characteristics: OLS Coefficients Resulting from Regressing Segment Membership Probabilities on Individual Characteristics.

Variable	Segment 1 (*Escapists*)	Segment 2 (*Weekend Challengers*)	Segment 3 (*Nature Nuts*)	Segment 4 (*Wilderness Trippers*)
Intercept	-1.9001[1]	-0.3425	-3.8337[1]	-6.9163[1]
Years of experience	-0.0129[1]	0.0339[1]	-0.0134[1]	-0.0419[1]
Level of specialization	0.0351	-0.6413[1]	0.2731[1]	0.8613[1]
Canadian	0.0121	-0.6852[2]	0.0997	0.1394
Household size	0.0060	0.3072[1]	-0.1097[1]	-0.2080[1]
Years of age	0.0144[1]	-0.0253	0.0152[1]	0.0393[1]
Level of education	0.0401	-0.1064	0.0927[1]	0.2247[1]
Income	0.0193	-0.1138[2]	0.0186	0.0141

[1] Coefficient is significant at the 5% level or better
[2] Coefficient is significant at the 10% level or better

TABLE 7
Compensating Variation ($/trip) for Some Hypothetical Changes in Recreation Quality for a Representative Individual in the Sample[1]

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

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

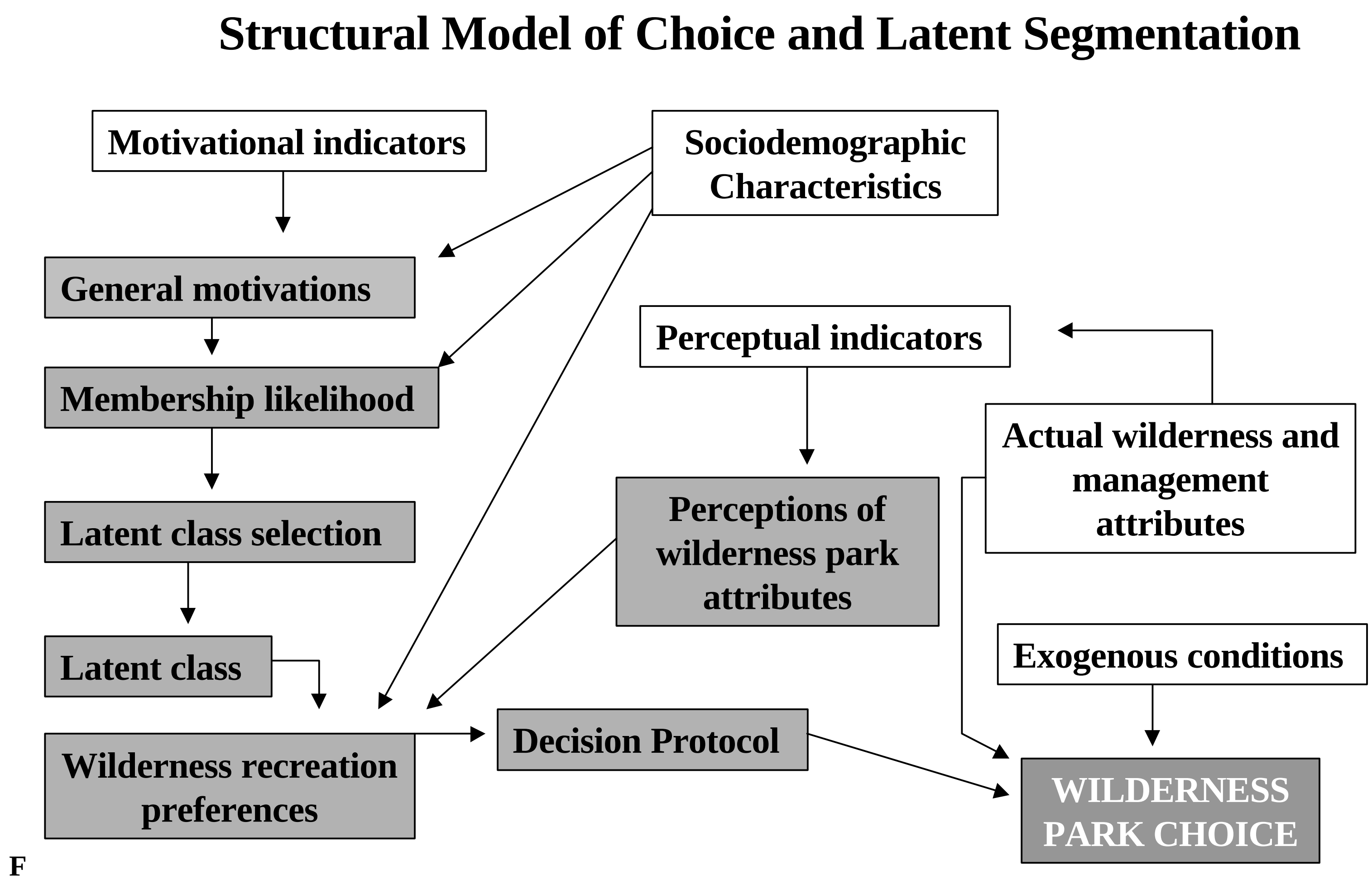

Figure 1.
A path diagram outlining the application of the latent segmentation choice model to backcountry recreation in a set of wilderness parks in the Canadian Shield region . Shaded boxes refer to the latent constructs utilized in the model.

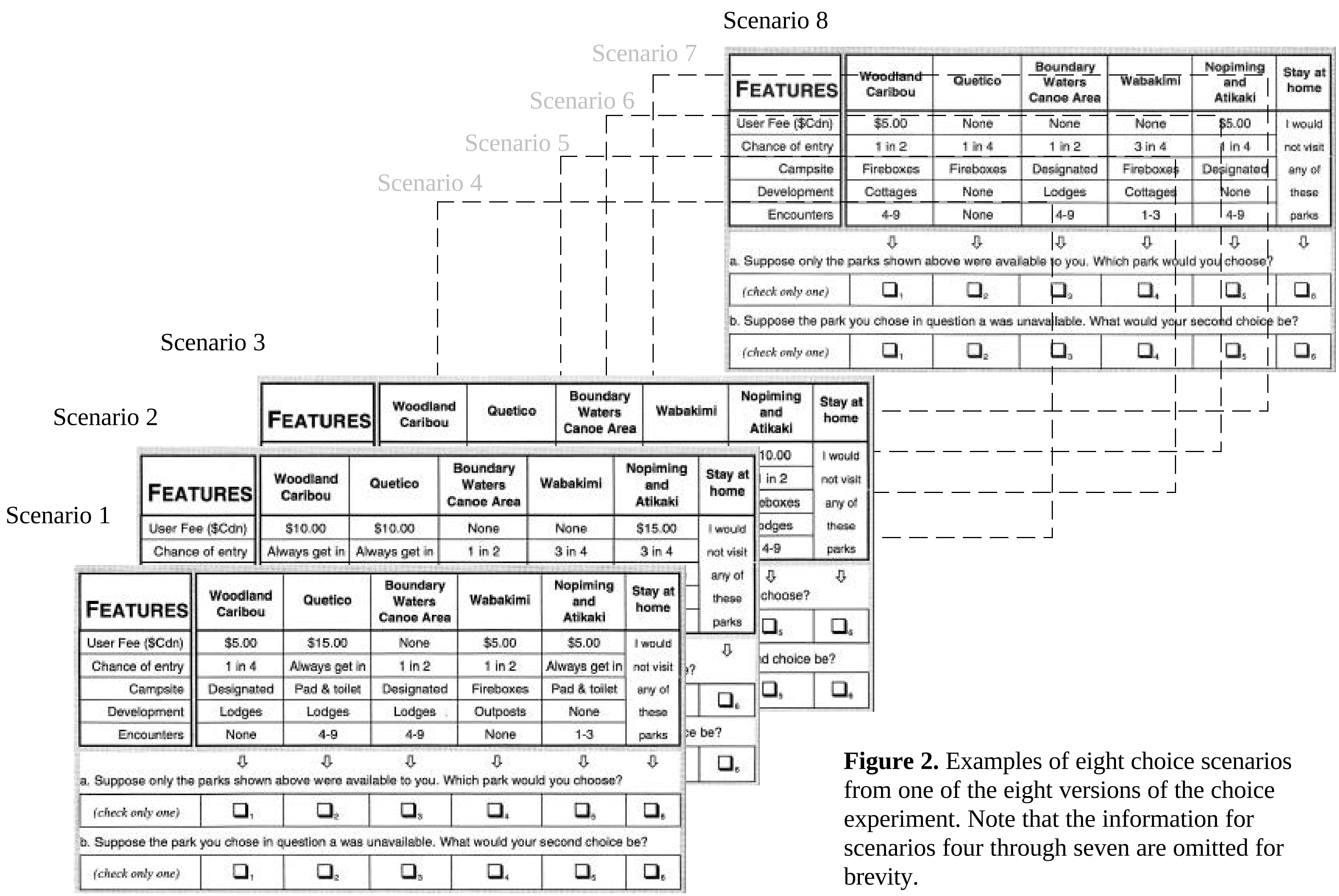

Figure 2. Examples of eight choice scenarios from one of the eight versions of the choice experiment. Note that the information for scenarios four through seven are omitted for brevity.

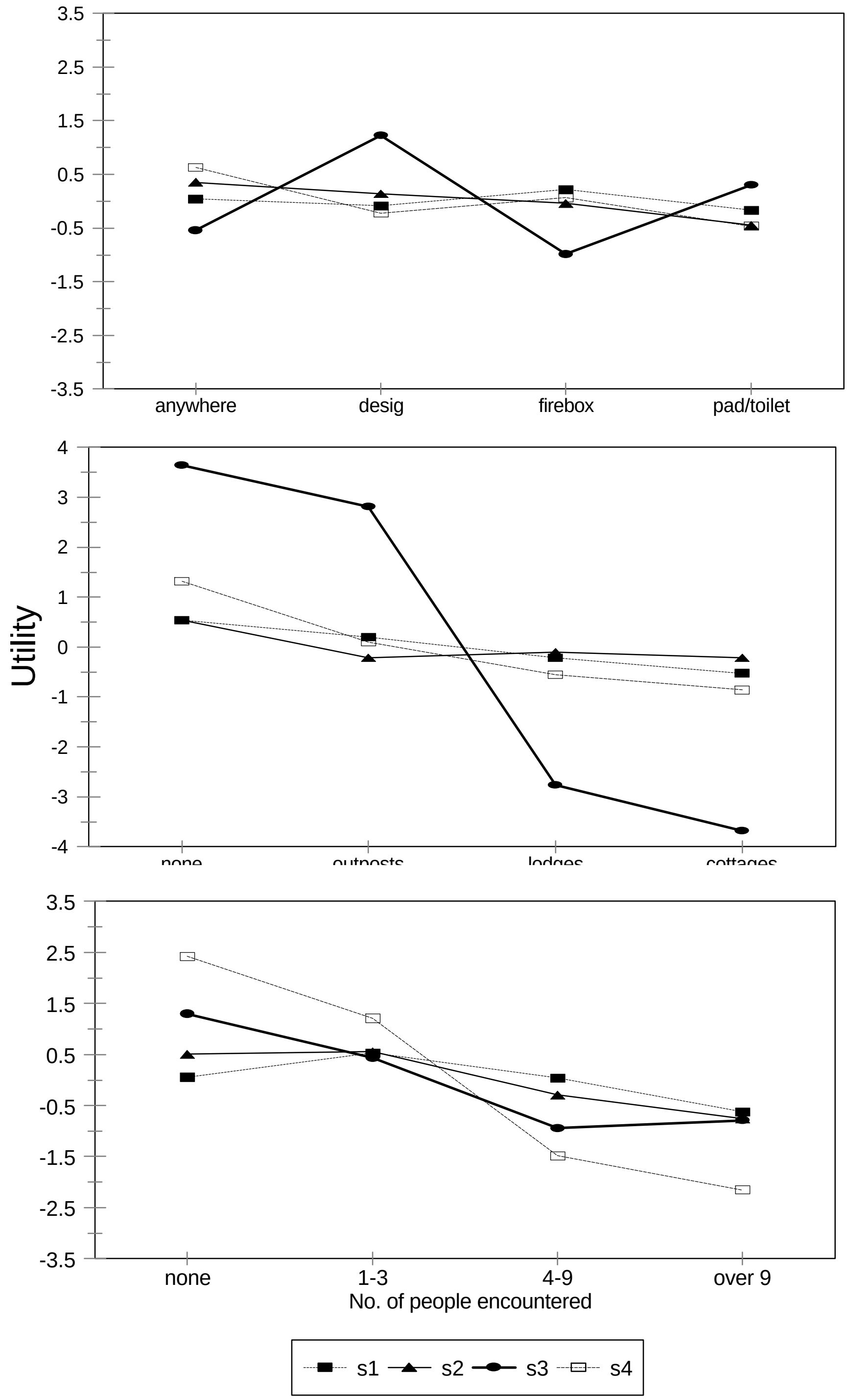

Figure 3. Plots of parameters associated with four levels of three attributes in the utility function over wilderness park choice. Note that the parameters on these levels were effects coded in the development of the model.

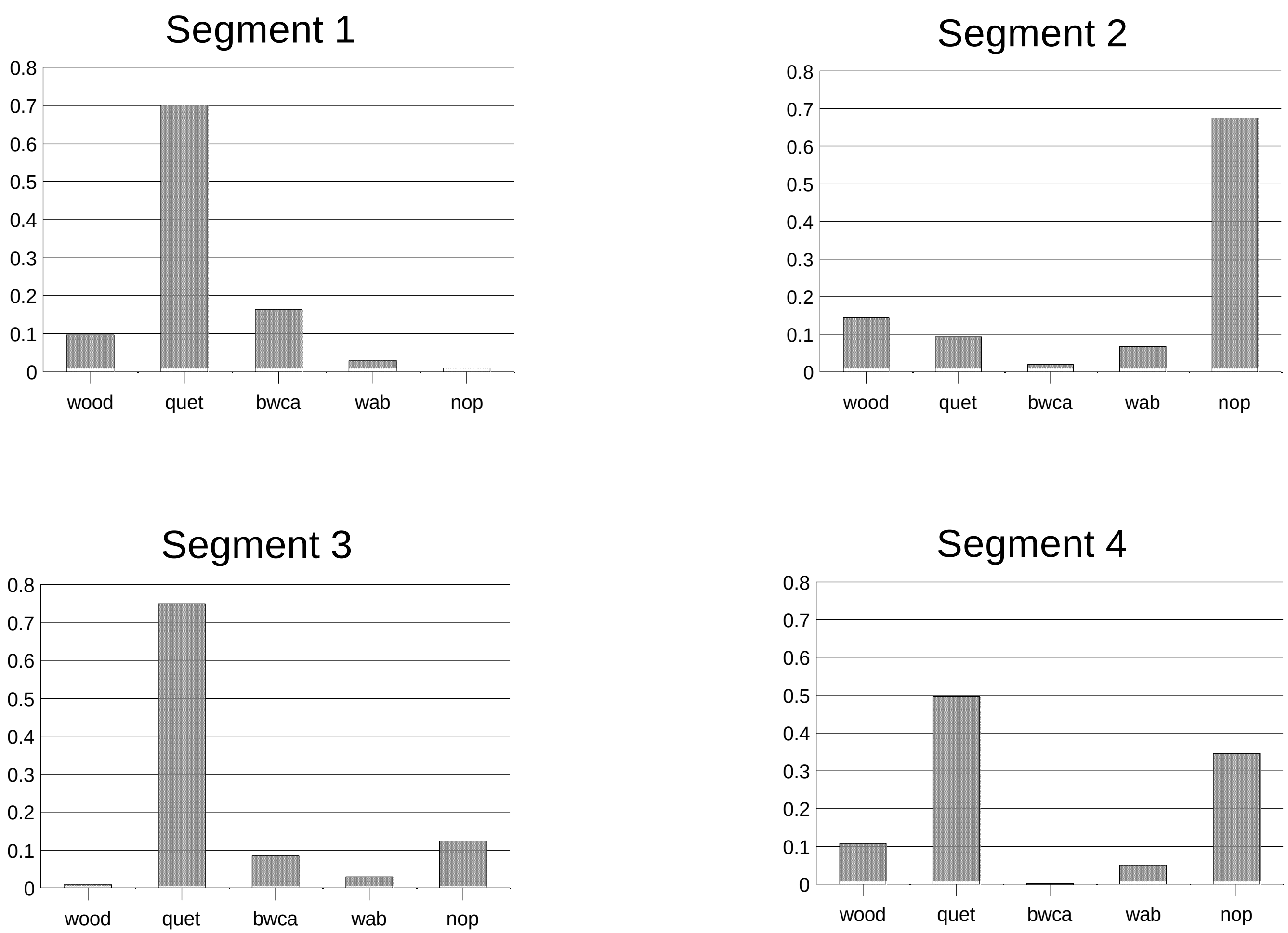

Figure 4. The estimated probability distribution of trips among the five parks for each of the four segments in the latent segment model.