ESTIMATING MIXED LOGIT RECREATION DEMAND MODELS WITH LARGE CHOICE SETS

ADAM DOMANSKI[1]

APRIL 30, 2009

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2009 AAEA & ACCI Joint Annual Meeting, Milwaukee, WI, July 26-28, 2009.

[1] Ph.D. Candidate – Department of Economics, North Carolina State University, Raleigh, NC, USA.
Contact: Adam Domanski; amdomans@ncsu.edu.
The author would like to thank Roger von Haefen, Ray Palmquist, Xiaoyong Zheng, Laura Taylor and Ryan Bosworth for their comments, suggestions, and guidance, as well as Chris Timmins and Jennifer Murdock for graciously granting access to their excellent data. All other errors are my own.

INTRODUCTION

Discrete choice models are widely used in studies of recreation demand. They have proven valuable when modeling situations where decision makers face large choice sets and site substitution is important. However, when the choice set faced by the individual becomes very large (on the order of hundreds or thousands of alternatives), computational limitations make estimation with the full choice set intractable. McFadden (1978) shows that sampling of alternatives in a conditional logit framework is an effective method to limit computational burdens while still producing consistent estimates. His approach has been widely used throughout the literature (Parsons and Kealy 1992; Feather 2003; Parsons and Needelman 1992). To implement the sampling approach researchers typically assume that unobserved utility is independently, identically distributed extreme value. The assumption implies that the relative probabilities of any two alternatives do not change with the addition of a third alternative. This is known as the independence of irrelevant alternatives, or IIA, which is necessary for consistent estimation under sampling of alternatives. Unfortunately, the IIA assumption is oftentimes a restrictive and inaccurate method for modeling behavior. Recent discrete choice innovations relax this assumption, however eliminating the reliance on IIA implies that sampling methods cannot be used.

The random parameters mixed logit model is one of the more attractive and widely used innovations of discrete choice modeling. It generalizes the conditional logit model by introducing unobserved preference heterogeneity across consumers through the model parameters (Train 1998). This makes the random parameters model a powerful and effective discrete choice tool but in doing so does not permit sampling of alternatives. Unfortunately, when a researcher is faced with analyzing a model with a very large choice set (e.g. a large number of lakes in a region or access points to the ocean), they must choose between a more accurate model (with mixed logit) and a computationally feasible one (with a sampled conditional logit).

Additionally in a random parameter model, preference heterogeneity is often introduced through analyst-specified parametric distributions for the coefficients. The researcher's choice in error distribution thus becomes an important step in the estimation procedure. The normal distribution is often employed but the well known restrictive skewness and kurtosis properties of such a specification raises the possibility that misspecification may be present. Other parametric distributions may be used, but in each case misspecification is a concern.

Both of these problems can be overcome through the use of a finite mixture model (latent class model) estimated via the expectation-maximization (EM) algorithm. The latent class approach probabilistically assigns individuals to certain classes, each with homogeneous preferences within class. This approach allows the researcher to recover separate preference parameters for each type of consumer without assumption of a parametric error distribution.

This type of model can be estimated with the recursive EM algorithm. Doing so transforms estimation of the non-IIA mixed logit model from a one-step computationally intensive estimation into a more feasible recursive estimation of an IIA conditional logit model. By reintroducing the IIA assumption at each step of the recursion, sampling of alternatives can be used to produce consistent estimates (von Haefen and Jacobsen unpublished). This estimation strategy has not been exploited in the recreation literature before.

This paper begins by describing how the latent class method can be used to account for preference heterogeneity while dealing with large choice sets without assuming a restrictive error distribution. Section one introduces the conditional and mixed logit models. Section two describes large choice set problems in discrete choice modeling. Section three details the latent class model estimated via the EM algorithm. Section four presents an empirical example using a recreation dataset of Wisconsin lake visits. Section five finishes with a conclusion and discussion of further research.

THE DISCRETE CHOICE MODEL

Discrete choice models are a set of powerful estimation techniques used to predict qualitative choice outcomes (McFadden 1974). Economic applications of discrete choice models assume that decision makers are utility maximizing. An individual, when presented with a set of alternatives, will select the outcome which generates the highest utility. From a recreation demand standpoint, the decision maker can be an individual who makes a trip to one of many possible recreation sites (in our case lakes), potentially many times a year.

Following Train (2003), the discrete choice model is based on the assumption that consumers (indexed by n, $n = 1, \dots, N$) choose recreation destination i if it gives them greater utility over all other alternative lakes $j = 1 \dots J,\ j \neq i$. Regardless of the purpose of visit (fishing, viewing, swimming, boating, etc), on that choice occasion, the lake chosen provides the best recreation opportunity for that individual. The consumer might enjoy visiting any other alternative lake j, but will choose lake i because $U_{ni} > U_{nj}\ \forall\ j \neq i$. We can relate the utility received from visiting lake i to a set of observable attributes relating to that choice, x_{ni}, (e.g. travel cost, water quality, amenities such as restrooms or boat ramps, catch rates, scenery, etc.) and the decision maker z_n (e.g. income, # of kids, whether they own a boat, etc.) Average or representative utility can be characterized as $V_{nj} = V(x_{nj}, z_n)\ \forall j$.

There is also an idiosyncratic component of utility which is captured by the error term ε_{ni}. This component is known to the individual but unknown to the researcher so utility and choice are random from her perspective. The resulting total utility from alternative i is represented by

$$U_{ni} = V_{ni} + \varepsilon_{ni}, \quad n = 1,\ldots, N, i \in J,$$

If the researcher assumes that each error is independently, identically distributed (iid) extreme value with the scale of utility normalized to the variance of unobserved utility by the scale parameter μ, the probability P_{ni} that individual n prefers alternative i is given by the multinomial logit formula (McFadden 1974):

$$P_{ni} = \frac{\exp(V_{ni}^* / \mu)}{\sum_j \exp(V_{nj}^* / \mu)}$$

Specifying representative utility to be linear in parameters $V_{ni} = \beta' x_{ni}$ where $\beta = \beta^* / \mu$ gives a choice probability that can be represented as:

$$P_{ni} = \frac{\exp(\beta' x_{ni})}{\sum_j \exp(\beta' x_{nj})}$$

The expected demand for alternative i is: $D_i = \sum_{n=1}^{N} P_{ni}$.

Independence of Irrelevant Alternatives

The logit model exhibits the IIA principle which states that the relative probabilities of any two alternatives must be independent of all other alternatives. Using the probability of choosing an alternative in the logit framework, the relative probability of choosing alternative *i* over *i** is:

$$\frac{P_{ni}}{P_{ni*}} = \frac{\dfrac{\exp(V_{ni})}{\sum_j \exp(V_{nj})}}{\dfrac{\exp(V_{ni*})}{\sum_j \exp(V_{nj})}} = \frac{\exp(V_{ni})}{\exp(V_{ni*})}$$

As can be seen, the ratio of the probabilities is independent of all other alternatives. This may be a plausible assumption in some cases, but in many, it is not behaviorally accurate.

The behavioral weakness of this assumption is well explained throughout the literature using the classic "red bus/blue bus" example (McFadden 1974). Suppose a decision maker has the option of commuting either by car or by red bus. Assume the probability of choosing either alternative is one half. The probability ratio is thus one and all probabilities sum to one. Now suppose a blue bus alternative is introduced. The relative probability of red bus to car must still be equal to one and we can assume the decision maker would be indifferent between buses making the red bus/blue bus probability ratio equal to one. Since the relative probabilities between any two alternatives are equal to one and the sum of all probabilities must equal to one, the conditional probability of any one alternative is now one-third. However, this now indicates that by introducing the blue bus, we have increased the probability of taking any bus to two thirds while reducing the car probability to one third. In reality, we should not expect the introduction of a second type of bus to change the car/bus relative probability, thus illustrating the behavioral weakness of IIA.

In a recreational demand model where an individual is choosing between two lakes, the construction of an identical second access point at one lake should not change the relative probability of choosing one lake over the other. However, including the new access point as a separate alternative will change the odds ratios in an IIA-restricted model.

Nested logit models have been used to overcome the limitations of IIA (Ben-Akiva 1973; Train et al. 1987; Forinash and Koppelman 1993; and Lee 1999). The fundamental decision made by the consumer can be represented as a series of decisions made in sequence. In a recreational context, a consumer would choose whether to fish at a lake or a river, then choose which specific lake to visit, etc. Within each decision step, the IIA assumption holds. However, across different steps, the ratio of probabilities can depend on the attributes of other alternatives in those nests and IIA does not hold.

An additional limitation of the conditional logit model is its limited ability to account for unobserved or random preference heterogeneity. The estimated coefficients represent an average for all decision makers. In some cases these results may present a weak fit for the data or may misrepresent preferences of the population, especially in situations where those preferences are diverse or polarized.

Observed preference heterogeneity can be accounted for by interacting observed individual demographic characteristics with the attributes of the alternatives. As a result, every one of the preference parameters is a function of the vector of observed socioeconomic characteristics. However, even after accounting for differences in observed individual characteristics it is likely the heterogeneity may still remain.

Mixed Logit Model

To relax the IIA assumption and account for preference heterogeneity, the mixed logit model can be used (Train et al. 1987; Ben-Akiva et al 1993; McFadden and Train, 2000). It is an extension of the standard logit model that allows the coefficients to vary across individuals. Theoretically, the utility of person *n* for alternative *i* is

$$U_{ni} = \beta_n x_{ni} + \varepsilon_{ni}$$

with ε_{ni} ~ iid extreme value and $\beta_n \sim f(\beta_n \mid \theta)$ where θ is a vector of parameters. As with the conditional logit model, it is assumed that every individual chooses the alternative which maximizes their utility. Conditional on β_n, the choice probability is now:

$$L_{ni}(\beta_n) = \frac{\exp(\beta_n x_{ni})}{\sum_j \exp(\beta_n x_{nj})}$$

where the probability of individual *n* choosing alternative *i* is the product of the logit probabilities over the density of β_n. These densities (and resulting probabilities) can be expressed either as a continuous or discrete mixing distribution.

Continuous Distribution Random Parameter Model

In the continuous mixing distribution the unconditional probability of selecting alternative *i* is expressed as:

$$P_{ni} = \int L_{ni}(\beta_n) f(\beta_n \mid \theta) d\beta$$

However, for estimation purposes, the researcher usually specifies the distribution of θ parametrically. She can then test various distributions and choose the one which provides the best fit. The estimation procedure involves simulation of the choice probabilities and estimation of parameters by drawing pseudo-random realizations from this underlying error distribution (Boersch-Supan and Hajivassiliou 1990; Geweke et al. 1994; McFadden and Ruud 1994).

There has been much discussion concerning the choice of distributions (Revelt and Train K 1998; Train and Sonnier 2003; Rigby et al 2008). Hensher and Greene (2003) analyzed the welfare effect of a mixed logit model with lognormal, triangular, normal, and uniform distributions. Although the mean welfare estimate was very similar across the normal, triangular, and uniform distributions, the lognormal distribution produced a result that differed by about a factor of three. Even though the mean estimate was similar amongst the normal, triangular, and uniform distributions, the standard deviation varied by as much as 17 percent. The functional form chosen can have a major impact on resulting WTP estimates and associated inferences as well as can be a determining factor for the accuracy of the estimation.

Latent Class Model

An alternative and more flexible specification relative to the parametric random parameters model is the latent class model. In this case the logit probability is expressed as a discrete mixing distribution:

$$P_{ni} = \sum_{c} s_{cn} L_{ni}(\beta_c),\ c = 1,\ldots,C.$$

where s_{cn} is the probability that $\beta_c = \beta$ and $s_{cn} = f(\beta_c \mid \theta)$. The subscript c represents each discrete class of parameters. Membership in each class is unobservable but can be predicted. Conditional on the agent's choices, the probability that an agent is a member of class c (and has coefficients β_c) is $h_{nc} = s_{cn} L_n(\beta_c)/P_n$ (Train 2008).

This is similar in functional form to the discrete factor method (DFM) where the distribution of classes is approximated with a step function and integrated out through a weighted sum of step levels where the weights are given by empirically estimated probabilities (Heckman and Singer 1984; Landry and Liu, forthcoming). Whereas DFM nonparametrically specifies the class weights, the above latent class model introduces individual demographic data into the latent class probabilities.

Computationally, the log-likelihood function for the sample to be maximized is

$$LL = \sum_n \ln(P_{nc}) = \sum_n \ln\left(\sum_c s_{cn} L_{ni}(\beta_c)\right)$$

Where

$$s_{cn} = \frac{\exp(\delta_c z_n)}{\sum_{c=1}^{C} \exp(\delta_c z_n)}$$

$$L_{ni}(\beta_c) = \prod_k^K \left(\frac{\exp(\beta_c x_i)}{\sum_{j=1}^{J} \exp(\beta_c x_j)} \right)^{1_k}$$

where k represents the choice occasion. Gradient-based maximization of the log-likelihood is possible, albeit computationally burdensome. The calculation involves the maximization of the log of the sum of the conditional probabilities weighted by the class probability. This maximization becomes even more complex with a large number of variables or classes. Additionally, the likelihood is not necessarily strictly concave implying the existence of several local maxima. In light of this, researchers typically employ several sets of starting values which can significantly increase the estimation run time.

LARGE CHOICE SETS

The selection of the choice set is vital to the effective implementation of any discrete choice model. Choice set definition deals with specifying the objects of choice that enter an individual's preference ordering. In practice, defining an individual's choice set is influenced by the limitations of available data, the nature of the policy questions addressed, the analyst's judgment, and economic theory (von Haefen 2008).

Train (2003) describes the three requirements of a choice set: 1) alternatives must be mutually exclusive, 2) the choice set must be exhaustive, and 3) the number of alternatives must be finite. In a recreational site-demand context, at any one moment, a consumer can only visit one site (satisfying requirement #1). However, when constructing the model, requirements two and three lend themselves to large choice set problems. All feasible alternatives must be included. To model a one-day recreation trip, all sites within one day's travel must be incorporated. When analyzing a multi-day trip, the choice set becomes much more complicated as more sites must be included in order to satisfy the exhaustive and finite requirements. These large choice set issues have been confronted throughout the literature (McFadden 1978, Parsons and Kealy 1992, Feather 2003, Parsons and Needelman 1992).

There are three common solutions to large choice set problems: 1) aggregation of alternatives, 2) assumptions of separability, and 3) sampling of alternatives. Solutions (1) and (2) are used in the formation of consideration sets and require the analyst to make additional assumptions about the behavior of the decision maker.

Many researchers believe that when consumers are confronted with choices from a large set of quality-differentiated goods, they may only seriously consider and choose from a subset of available alternatives. An explanation of this is that decision makers have limited information about all available goods and incur costs in the acquisition of additional information. They rationally search for more information whenever the marginal benefit of doing so exceeds the marginal cost and at any point in time, have detailed information for only a subset of the available goods (e.g. their 'consideration set'). There is growing empirical evidence supporting the behavioral foundation of consideration sets (Shocker et al. 1991; Sethuraman et al. 1994) and the concept has been employed in the environmental economic literature (Horowitz 1991; Peters et al. 1995; and Parsons et al. 2000). However empirically, there is the practical question of how to identify the objects of choice that enter each individual's consideration set from the universal set of relevant alternatives. This information is not revealed by consumer choice and the analyst may be forced to make potentially restrictive assumptions on the search process of the consumer.

Aggregation methods make the assumption that alternatives can be grouped into representative choice options. For a recreational demand context, similar recreation sites can be treated as one; in housing, a group of homes in a given sub-development can be aggregated. This methodology can be effective but is problematic in that the success of estimation is entirely dependent on the assumptions made in the aggregation. McFadden

(1978) and Ben-Akiva and Lerman (1985) have both shown that this technique can produce biased estimates if the utility variance within aggregates is not accounted for.

Separability assumptions allow the researcher to selectively remove alternatives from the choice set based on a perceived very low probability that those options would be chosen by the individual. This restricts preferences to an analyst-determined consideration set. This method is common in discrete choice housing models where homes are excluded from the choice set based upon distance or price. Again, the success of this method is only as good as the assumptions made.

The third common solution is to sample the alternatives the decision maker faces. Fundamentally, this can be done as long as the resulting choice probability ratios do not change due to the elimination of choice alternatives. This is feasible within the standard logit model due to the IIA assumption.

McFadden (1978) proves that within the context of a discrete choice logit model sampling of alternatives provides consistent model parameters. This has been successfully utilized and demonstrated in the literature (Sermons and Koppelman 2001; Waddell 1996; Bhat et al. 1998; Guo and Bhat 2001; Ben-Akiva and Bowman 1998, von Haefen and Jacobsen unpublished).

Sampling of Alternatives

When faced with a very large choice set which exhibits IIA, sampling from alternatives can simplify the computational process while still producing consistent estimates as long as the uniform conditioning property holds (McFadden 1978). This property (which is necessary for sampling) states that the resulting choice probabilities will be the same regardless of the sample chosen. More formally, uniform conditioning states that if there are two alternatives, *i* and *j* which are both members of the full set of alternatives *C* and both have the possibility of being an observed choice, the probability of choosing a sample of alternatives *D* (which contains the alternatives *i* and *j*) is equal, regardless of whether *i* or *j* is the chosen alternative.

Both the continuous distribution random parameter and latent class models, as shown earlier, can account for preference heterogeneity and in some cases provide for an improvement in fit over the conditional logit model. However, when faced with a large choice set, the continuous distribution method cannot provide consistent estimates when sampling from alternatives. Recall that the mixed logit probability is represented by:

$$P_{ni} = \int L_{ni}(\beta) f(\beta \mid \theta) d\beta$$

The relative probability of choosing alternative *i* over *i** is:

$$\frac{P_{ni}}{P_{ni*}} = \frac{\int \frac{\exp(\beta_n x_{ni})}{\sum_j \exp(\beta_n x_{nj})} f(\beta \mid \theta) d\beta}{\int \frac{\exp(\beta_n x_{ni*})}{\sum_j \exp(\beta_n x_{nj})} f(\beta \mid \theta) d\beta}$$

The denominators of the integral are inside the logit formula and therefore do not cancel. The resulting relative choice probabilities do depend on the other alternatives and IIA does not hold.

Nerella and Bhat (2004) evaluate sampling of alternatives in a continuous distribution model. However, they sample by assuming uniform conditioning without a theoretical basis for doing so. Nevertheless, they analyzed the effect of sample size on the empirical accuracy and efficiency of multinomial and mixed multinomial models. Their results suggest (for a standard model) using an eighth of the size of the full choice set as a minimum, and suggest a fourth of the full choice set as a desirable target. Empirical testing for the mixed model suggests using a fourth of the full choice set to one-half of the full choice set as a desirable target.

McConnell and Tseng (2000) perform a similar analysis on beach use and recreational fishing. They found that sampling in a continuous distribution model does not alter the results significantly or systematically. The theoretical infeasibility however remains. The inability to sample alternatives has proven a difficult question in the literature and has been deemed an area requiring further study.

SAMPLING IN A MIXTURE MODEL

However, utilizing the latent class model maximized via the recursive expectation-maximization (EM) algorithm, sampling of alternatives can achieve theoretically consistent estimates (von Haefen and Jacobsen, unpublished).

The EM algorithm (Dempster et al., 1977) is a method for maximizing a computationally complicated maximum likelihood function given the values of some correlated, known parameters. The EM algorithm has become a popular tool in statistical estimation problems involving incomplete data (McLachlan and Krishnan 1997) or problems which can be posed in a similar form, such as mixture estimation (Bhat 1997; Train 2008). The method also facilitates the consistent sampling of alternatives as shown by von Haefen and Jacobsen (unpublished).

Assuming that an unknown parameter (in this case the latent class probability) is represented as a value in some parameterized probability distribution, the EM algorithm is a recursive procedure which begins by 1) specifying the expected value of unknown parameters given some known parameters. 2) The parameters of the known values are then re-estimated given the expected values of the unknown parameters. The steps are then repeated until convergence, defined by a pre-determined small change in the parameter estimates between iterations (Train 2008). This methodology is an improvement over gradient-based methods by its ability to transform the maximization of a log of sums into a recursive maximization of the sum of logs.

In our recreational demand context, given some set of starting values for the parameters θ_t (representing individual and alternative specific characteristics), the EM algorithm lets us calculate a new value for the parameters:

$$\theta_{t+1} = \arg\max_{\theta} \sum_{n} \sum_{c} h_{nc}(\theta_t) \ln\left(s_c L_n(\beta_c)\right)$$

where t represents the iteration number. Since the right hand side of the equation can be rewritten as $\ln\left(s_{cn} L_n(\beta_c)\right) = \ln\left(s_{cn}\right) + \ln\left(L_n(\beta_c)\right)$, the maximization can be performed independently for each set of parameters. Using various starting values, the probability (weight) of individual n belonging to class c is calculated keeping the choice probabilities fixed:

$$h_{cn} = \frac{s_{cn} L_{ni}(\beta_c)}{\sum_{c} s_{c*n} L_{ni}(\beta_{c*n})}$$

A maximization is then performed to update the individual class probability dependent on individual specific variables using the weights from the previous step as given:

$$LL_n = \sum_{c}^{C} h_{cn} \ln\left(s_{cn}\right)$$

Another maximization is performed to update the conditional probability parameters, again using the weights as fixed; independently for each class:

$$LL_n = \sum_c^C h_{cn} \ln\left(L_{ni}\left(\beta_c\right)\right)$$

The weights are then re-estimated using the new parameter values, and the entire process is repeated until convergence. Each successive maximization takes the prior parameters and individual-specific class probabilities as fixed for the maximization of the new values. The previously computationally burdensome estimation has now been transformed into a recursive conditional logit estimation for each class and choice probability.

By breaking the mixed logit non-IIA model into a series of standard logit IIA models, sampling of alternatives can be reintroduced at each recursive step while still effectively modeling a mixing distribution. Recall that the two functions to be maximized are:

$$s^{i+1} = \arg\max_s \sum_n \sum_c h_{nc}(\theta^i) \ln s_{cn}$$

which finds the expected value of the unknown variable (individual class probability) and

$$\beta_c^{\ i+1} = \arg\max_{\beta_b} \sum_n h_{nc}(\theta^i) \ln L_{ni}(\beta_c)$$

which re-estimates the likelihood function. The maximization procedure calculates a standard logit (IIA) likelihood function for each class independently, keeping the individual weights fixed from the previous step. In this way IIA holds within-class and for each estimation step. The within-class estimates are independent of other alternatives in the choice set, and thus via the EM algorithm, sampling of alternatives will generate consistent estimates.

It should be noted that use of the EM algorithm to solve a nonparametric log-likelihood function has two notable drawbacks. First, convergence may be at a local instead of a global maximum because the unconditional likelihood is not globally concave. To address this, it is often necessary to use multiple starting values. Second, recursive estimation with multiple starting values may be time consuming. The procedure provides the greatest computational benefits when used with a very large dataset where traditional maximum likelihood is not feasible and lesser benefits with marginally smaller datasets.

Model Selection

The researcher must choose the number of latent classes to be used. This may seem similar to the dilemma of choice of error distribution in the mixed logit model; however the implications are much less troublesome. Traditional specification tests (likelihood ratio, Lagrange multipliers, and Wald tests) do not satisfy the regularity conditions for a limiting chi-square distribution so alternative tests must be used. Throughout the latent class literature a variety of information criteria statistics have been used. In general form (Hurvich and Tsai 1989), the information criteria statistic is specified as $-2\ln(L) + P{*}\delta$ where $\ln(L)$ is the log likelihood of the model at convergence, P is the number of estimated parameters in the model, and δ is a penalty constant. There are a number of different types of information criteria statistics that depends on the value of the penalty constant δ.

Information Criteria	Penalty Constant δ =
Akaike Information Criteria	2
Bayesian Information Criteria	$\ln(N)$
Consistent Akaike Information Criteria	$1+\ln(N)$
Corrected Akaike Information Criteria	$2 + 2(P+1)(P+2)/(N-P-2)$

In each case, the optimal model is the one which gives the minimum value of the respective information criteria. Roeder et al. (1999) and Greene and Hensher (2003) suggest using the Bayesian Information Criteria (BIC). One advantage of the BIC over traditional hypothesis testing is that it has good properties under weaker regularity conditions than the likelihood ratio test (Roeder et al., 1999). Alternatively, many past papers (e.g. Meijer and Rouwendal 2006; Desarbo et al 1992; Morey et al. 2006) have used the AIC (Akaike 1974). Other papers have compared the various information criteria (Thacher et al 2005; Scarpa and Thiene 2005; and Hynes et al 2008), however there is no general consensus in the literature for using one test over the others. The improved fit of a given model will be taken in comparison with the possibility of over-fitting.

Standard Errors

Calculation of the standard errors of parameter estimates can be cumbersome since there is no direct method for evaluating the information matrix. There is a wide expanse of statistical literature addressing various methods of calculating standard errors based upon the observed information matrix, the expected information matrix, or on resampling methods (Baker 1992, Jamshidian and Jennrich 2002, Meng and Rubin 1991). The method used in this paper is a simple one based upon Ruud (1991). At convergence of the EM algorithm, the matrix of scores and the numerical hessian matrix are calculated for each independent maximization step and then inverted to calculate robust clustered standard errors for the parameter estimates.

EMPIRICAL EXAMPLE

An empirical illustration is performed with the Wisconsin Fishing and Outdoor Recreation Survey. Taken in 1998 by Triangle Economic Research, this dataset has been used previously by Murdock (2006) and Timmins and Murdock (2007). A random digit dial of Wisconsin households produced a sample of 1,275 individuals who participated in a telephone and diary survey of their recreation habits over the summer months of 1998. 513 individuals reported taking a single day trip to one or more of 569 sites in Wisconsin (identified by freshwater lake or, for large lakes, quadrant of the lake). Of the 513 individuals, the average number of trips was 6.99 with a maximum of 50. Each of the 569 lake sites had an average of 6.29 visits, with a maximum of 108. Table one presents summary statistics.

[Table 1 – Summary Statistics]

This is an ideal dataset to evaluate the consistency of sampling of alternatives with a nonparametric approach because it is large enough so that a researcher would ideally prefer to work with a smaller choice set, however it is small enough so that estimation of the full choice set is still feasible for comparison.

The full choice set is estimated with both a standard logit model and a multiple class latent class model. The parameter results are evaluated and the various information criteria are used to compare improvements in fit by allowing for preference heterogeneity. The same estimation is then also performed using three random sets of starting values on five randomly sampled choice sets equal to 50%, 25%, 12.5%, 5%, 2%, and 1% of the non-selected alternatives.

To provide complete analysis of sampling in the latent class method, first discussion on the sampling properties of the conditional logit model will be presented. The results of the latent class model will then be compared using these results as a baseline.

Conditional Logit Results

All estimation was coded and performed in Matlab using the fminunc optimization toolbox with an analytically coded gradient and numerical hessian. In the conditional logit model, the likelihood function is globally concave so starting value choice is of minimal importance (as opposed to the mixture model). The dataset contains multiple trip occasions; however the estimation was performed on a per-trip basis. In the sampled models, a unique and random choice set was generated for each choice occasion. With a sample size of N, N-1 alternatives were randomly selected and included with the chosen alternative. Five random samples were run for each sample size.

[Table 2 – Estimation Time: Conditional Logit Model]

The primary reason for sampling from alternatives is to reduce the computational burden of estimation. An analysis of sampling's effect on estimation time shows a negative slope with diminishing returns at very small samples. Estimation was performed on multiple computers with varying processor speeds, so all times were normalized at the 50% one-class model. Estimation time for the full model was approximately 31 minutes. Table two shows the average estimation time of the five random samples relative to the estimation time of the full model. Cutting the sample by an additional 50% in any model roughly equates to an 80% reduction in estimation time.

[Table 3 –Parameter Estimates: Conditional Logit Model]

Table three shows the parameter estimates and standard errors for each of the sample sizes. Five random samples were run and the means of the estimates and standard errors are reported. Two log likelihood values are reported in this table. The true log likelihood (LL) and the “normalized log likelihood” (NLL). In any sampled model, a smaller set of alternatives will necessarily result in a smaller LL. This number, however, is not useful in comparing goodness of fit across sample sizes. For this reason the NLL is used. After convergence is reached in a sampled model, the parameter estimates are then used with the full choice set to compute the LL as if they were the results of the full model. Although this is marginally more time consuming than using the sampled data, it does not drastically increase overall computation time, as the optimization routine is the most computationally intensive process.

A comparison of the LL and NLL show that, when sampling, less information is available and each successive sample provides a marginal reduction in goodness of fit, as expected, however this decrease in fit is very small. A decrease in the sample size also increases the standard errors of the NLL reflecting the increased variability of estimates in smaller samples.

The parameters themselves are rational (in terms of sign and magnitude) in the full model and relatively robust across sample sizes. Travel cost and small lake are negative and significant, while all fish catch rates and the presence of boat ramps are positive and significant, as expected. The standard errors for the parameters generally increase as the sample size drops, reflecting decreased ability to produce as strong of a fit with less information. In the smallest samples, this decrease in fit is enough to make some previously significant parameters insignificant. The parameters themselves move across sample sizes; however the most useful method for analyzing these is via a comparison of several welfare scenarios.

Five different policy scenarios are considered and welfare estimates for all specifications are calculated. Welfare changes are measured independently for each class and then aggregated across all classes within a specification, dependant on the class share. The following policy scenarios are considered: 1) infrastructure construction, 2) an increase in entry fees, 3) and urban watershed management program, 4) an agricultural runoff management program, and 5) a fish stocking program. Note that crowding considerations are not considered here, but these scenarios can be augmented or modified to fit any number of policy proposals.

The infrastructure construction program simulates an augmentation of current man-made infrastructure across sites. From the parameter estimates, it is clear that boat ramps are desired amenities, but only approximately 73% of sites have one. Supposing that a boat ramp was constructed at each Wisconsin lake that did not have one, the estimates reflect the average willingness to pay (WTP) per participant per trip.

The second policy scenario replicates the impact of a $5 increase in entry fees at all state managed sites (defined by being in a state forest or wildlife refuge); approximately 23% of sites. The estimates reflect the average WTP per participant per trip.

The third and fourth policy scenarios are related by the assumption that lakes are negatively affected by either urban or agricultural runoff. Sites that are near an urban area are affected by urban runoff, while sites that are not in an urban area but also not in a state forest or wildlife refuge, are affected by agricultural runoff. It is assumed that all sites affected by one of these scenarios are impacted uniformly. This decrease in water quality is assumed to cause a decrease in the aquatic life in a lake and by association, a decrease in catch rate. The policy scenarios suppose that a storm water or non-point source pollution management policy could improve the quality of water and increase the catch rate by a uniform 5% across all fish species at affected sites. The estimates reflect the average WTP per participant per trip.

The final policy scenario replicates a fish stocking program where the catch rate of trout is increased by 25% across all sites. Not every site contains trout, so the stocking program only takes place in locations where trout are indigenous.

The methodology used to calculate WTP is the log-sum formula derived by Hanemann (1978) and Small and Rosen (1981). Assuming a constant marginal utility of income $\beta_p^* f(y - p_j) = \beta_p^*(y - p_j)$ and an attribute improvement from q^0 to q^1, the compensating surplus is

$$CS = \frac{1}{\beta_p}\left(\max_j\left(-\beta_p p_j + \beta_q q_j^1 + \varepsilon_j\right) - \max_j\left(-\beta_p p_j + \beta_q q_j^0 + \varepsilon_j\right)\right)$$

and for the iid type 1 extreme value case is

$$E(CS) = \frac{1}{\beta_p}\left(\ln\left(\sum_j^J \exp(-\beta_p p_j + \beta_q q_j^1) \right) - \ln\left(\sum_j^J \exp(-\beta_p p_j + \beta_q q_j^0) \right) \right)$$

[Table 4 – Welfare Scenarios: Conditional Logit Model]

This table shows the performance of the welfare estimates across sample size. Recall that five unique samples were run. The parameters from each sample were then used with the full choice set. As in the case with the NLL, using the full choice set does take more time than simply using the sampled set, however this provides for a stronger comparison and does not drastically increase total computation time since no optimization is occurring.

The mean WTP for each unique sample is calculated. The table shows the mean of the five mean WTP estimates, as well as the standard deviation of the five mean WTP estimates. As can be seen across the various welfare results, there is an increased variation in WTP estimates as the sample size becomes smaller, reflecting the reduced information available. Depending on the welfare scenario, there is also a slight upward or downward bias, however this is not consistent across scenarios and might not be predictable. Considering the full choice set estimates as the “true” value, it can be seen that progressively smaller samples lead to a smaller probability that the estimated WTP will match the true value. Generally however, sampling performs very well down to the 12.5% sample size and, depending on the needs of the researcher, can be useful down to the 1% level.

[Table 5 –Welfare Confidence Intervals: Conditional Logit Model]

Table five shows the mean 95% and 75% confidence intervals across the samples. This illustrates the variance of the individual WTP estimates within each sample. A similar conclusion can be drawn. Variability of the estimates increases as the sample size decreases and is dependent on the specific welfare scenario being modeled, but can be considered very useful down to the 1% level depending on the needs of the researcher.

Latent Class Results

A similar evaluation of the performance of sampling in a mixture model will now be performed. The main goal will be to show how the proposed procedure performs in direct comparison to the conditional logit model.

Estimation was performed in Matlab with analytically coded gradients and numerical hessians. Convergence in the EM algorithm was defined as the point at which the maximum change in parameters between iterations was less than 1/100. Since the likelihood function is not globally concave and there is the possibility of convergence on a local minimum, multiple starting values were used. To speed up computation, initial starting values were determined by a method used by Train (2008). For a model with C classes, the full dataset is partitioned into C segments and a conditional logit model is performed on each, the resulting parameters are used as the starting values for each of C classes. Alternative starting values were chosen by making random variations to those initial values. A total of three starting values were used on the same fixed sample, the smallest log likelihood of which was determined to be the global minimum. It would be advantageous to use more than three starting values (for instance, 10) to ensure convergence on a global minimum, but for the purposes of this research, since many different models were being run, only three starting values were used.

Fifteen site-specific parameters are estimated along with four individual-specific parameters. The price coefficient is held fixed across classes and the individual-specific parameters are normalized for the first class, thus leaving only C-1 individual specific coefficients.

[Table 6 – Estimation Time: Latent Class Model]

Five independent samples were taken for each successive sample size, using the same procedure as in the conditional logit model. The average computation time is shown in the graph above. Estimation time approximately doubles with each additional class, and convergence in the full sample six class model was reached after approximately one week. Sampling provides a decrease in relative runtime on a similar scale as in the conditional logit model, with diminishing returns to scale at much smaller sample size.

As in the conditional logit model, the NLL is calculated and used for comparison purposes. However, in this model, the NLL is more appropriate than the LL even without a cross-sample comparison since the full choice set is used for constructing the latent class probabilities.

[Table 7 – Information Criteria]

The NLL is used to calculate the various information criteria to determine the correct number of classes. The AIC, BIC, and CAIC all indicate use of the same number of classes (six with the full choice set), however, evaluation of the parameter estimates suggest an over-fitting of the model. At six classes, the model is attempting to fit 105 parameters to 569 alternatives. At a large number of classes, the parameters for one class for certain variables (specifically catch rates for trout, musky, and salmon) diverge dramatically. This is the result of the model trying to incorporate a handful of anomalous outlying observations, and thus the more appropriate decision criteria is the crAIC which incorporates the greatest penalty for an increased number of parameters.

When sampling, the information criteria produce different conclusions based upon the sample size. A smaller sample size in some cases leads to the result of a fewer number of classes in the optimal model. This is certainly true in the crAIC, which suggests the use of three classes in all sample sizes except the 1% level, where two classes are optimal. This result points towards a possible over-specification and the failure of the uniform conditioning property to hold at extremely small samples. The 1% sample uses only six out of 569 alternatives, and the probability that any random sample will realistically represent the entire choice set is low.

[Table 8 –Parameter Estimates: Two Class Model]

In a latent class model, a full set of parameters is estimated for each class. Recall that in our construction, individuals are homogeneous within class and heterogeneity is captured by the results of multiple classes. Each individual is assigned to each class with a probability that can be constructed out of the individual specific parameters. Table eight shows the results from a single two class model with multiple sample sizes estimated with three sets of starting values. When analyzing the full sample results, it can be seen that owning a boat will lead to a larger probability of being in class one, which corresponds to a greater preference for boat ramps than class two. Class one can be described as the 'boat owner' class. Recall that when analyzing the results for multiple random samples in the conditional logit model we took the mean of the parameter estimates and parameter standard errors. This type of comparison is not possible in a latent class model because, although class one happened to be the 'boat owner' class, using alternative starting values or sampled choice sets may result in class two being the 'boat owner' class. Thus, for the purposes of comparison, mean parameter estimates for the optimal number of classes are computed and presented in a similar format as the conditional logit model above. Using the crAIC decision rule, the parameters for each class of the optimal model are weighted by the average class share (determined by the average latent class probability across individuals) and averaged to present a mean parameter estimate for the entire population. This sort of analysis would be useful to a researcher seeking estimates from a better-fitting model, but not if a cross-parameter comparison within classes is required. The mean parameter estimates from the optimal model is presented in table nine. Parameter standard errors are calculated by the same methodology. This weighted mean of parameter estimates is useful only when running multiple random samples. In a practical application where only one set of estimates is used, the full set of results as seen in table eight provides more information.

[Table 9 –Parameter Estimates: Latent Class Model]

As can be seen across sample sizes, the parameter estimates are, for the most part, robust for those that are significant at the 95% level. Standard errors also are fairly consistent across samples and generally increase as the sample size decreases – a similar result to the conditional logit model. Greater precision can be gained by running additional starting values and using more random samples.

[Table 10 – Welfare Scenarios: Latent Class Model]

Finally, the stability of the WTP estimates across sampling in the mixed model is analyzed. Using the same policy scenarios as in the conditional logit model, WTP estimates are constructed for each individual in each class. They are then weighted by the individual class probability and averaged together. In the sampled models, the mean WTP across individuals within a class is compared with that of four alternate random samples, the mean and standard deviation of which is reported in table 10. The standard deviation shows how the mean estimates vary across different samples. The results are not entirely clean or consistent to the smallest samples, but as in the conditional logit model, are good down to the 12.5% level, and depending on the needs of the researcher, can be useful down to the 1% level. The lower stability of the estimates (as compared to the conditional logit model) is a result of using a small number of starting values (again, only three were used while 10 would be ideal) and the demands of extracting a large amount of information (in terms of the number of parameters to be estimated) from a relatively small amount of information – an effect which is exacerbated at the smallest sample sizes.

CONCLUSION

This paper has demonstrated the theoretical and practical uses of sampling of alternatives in a mixed logit non-parametric framework. By employing the EM algorithm, non-IIA estimation can be broken down into a recursive conditional logit estimation for each class. Within each class, IIA holds and thus allows for sampling of alternatives.

At the current state of research, this paper has demonstrated the theoretical foundation for sampling of alternatives in a discrete choice mixture model. By running several specifications on a recreation dataset, the applicability of the method has been illustrated as well. Future research could include a comparison against the nested logit model and continuous random parameter mixed logit model. Additionally, a full Monte Carlo analysis of the model would be useful for analyzing bias in sampling techniques and to fully explore the practical extent of the model. Using the greater number of runs, a meta-regression can be conducted to analyze trends present when varying the sample size or number of latent classes following Banzhaf and Smith (2007).

Finally, since the current state of computing power is constantly and drastically improving, it would seem that this method will soon be out-of-date and could have rather provided the greatest benefit ten to fifteen years ago. Although computational limitations were the prime motivational force behind McFadden's 1978 paper introducing sampling, this method is not yet antiquated as evidenced by its still widespread use in the academic literature as well as the sometimes practical need for producing quick consistent estimates for policy analysis or legal cases. Most importantly, the method will always allow us to "push the envelope" of research and innovation by running larger and more complicated models.

REFERENCES

Akaike, H. (1974). "A New Look at the Statistical Model Identification." IEEE Transactions on Automatic Control 19, pp. 716-723.

Baker, S. (1992). “A Simple Method for Computing the Observed Information Matrix When Using the EM Algorithm with Categorical Data.” Journal of Computational and Graphical Statistics 1(1), pp. 63-76.

Banzhaf, S. and K. Smith. (2007). “Meta-analysis in Model Implementation: Choice Sets and the Valuation of Air Quality Improvements.” Journal of Applied Econometrics 22, pp 1013-1031.

Ben-Akiva, M. (1973). "Structure of Passenger Travel Demand Models." Ph.D. Dissertation, Department of Civil Engineering, MIT.

Ben-Akiva, M. and J. Bowman. (1998). "Integration of an Activity-based Model System and a Residential Location Model." Urban Studies 35(7), pp. 1131-1153.

Ben-Akiva, M., and S.R. Lerman. (1985). "Discrete Choice Analysis: Theory and Application to Travel Demand." Boston, MA: MIT Press.

Ben-Akiva, M., D. Bolduc, and M. Bradley. (1993). "Estimation of Travel Choice Models with Randomly Distributed Values of Time," Papers 9303, Laval - Recherche en Energie.

Bhat, C. (1997). "Recent Methodological Advances Relevant to Activity and Travel Behavior Analysis." In H.s. Mahmassani, ed., Recent Developments in Travel Behavior Research. Pergamom, Oxford.

Bhat, C., A. Govindarajan, and V. Pulugurta. (1998). Disaggregate Attraction-End Choice Modeling: Formulation and Empirical Analysis." Transportation Research Record 1645, pp. 60-68.

Boersch-Supan, A. and V. Hajvassiliou. (1990). "Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models." Journal of Econometrics 58, pp. 347–368.

Dempster, A., N. Laird, and D. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm." Journal of the Royal Statistical Society. Series B 39(1) pp. 1-38.

Desarbo, W., M. Wedel, M. Vriens, and V. Ramaswamy. "Latent Class Metric Conjoint Analysis." Marketing Letters 3(3), pp. 273-288.

Feather, P. (2003). "Valuing Food Store Access: Policy Implications for the Food Stamp Program." American Journal of Agricultural Economics 85(1), pp. 162-172.

Forinash, C. and F. Koppelman. (1993). "Application and Interpretation of Nested Logit Models of Intercity Mode Coice." Transportation Research Record 1413, pp. 98-106.

Geweke, J., M. Keane, and D. Runkle. (1994). "Alternative computational approaches to inference in the multinomial probit model." Review of Economics and Statistics LXXVI, pp. 609–632.

Greene, W. and D. Hensher. (2003). "A latent class model for discrete choice analysis: contrasts with mixed logit." Transportation Research B.

Guo, J.Y., and C.R. Bhat. (2001). "Residential Location Choice Modeling: A Multinomial Logit Approach." Technical Paper, Department of Civil Engineering, The University of Texas at Austin.

Hanemann, W. (1978). "A Methodological and Empirical Study of the Recreation Benefits from Water Quality Improvement." Ph.D. dissertation, Department of Economics, Harvard University.

Heckman, J. and B. Singer. (1984). "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." Econometrica 52, pp. 271-230.

Horowitz, J. (1991). "Modeling the choice of choice set in discrete-choice random-utility models." Environment and Planning A 23(9), pp. 1237–1246.

Hurvich, C. and C. Tsai. (1989). “Regression and Time Series Model Selection in Small Samples." Biometrika 76 (2), pp. 297-307.

Hynes, S., N. Hanley, and R. Scarpa. (2008). "Effects on Welfare Measures of Alternative Means of Accounting for Preference Heterogeneity in Recreational Demand Models." American Journal of Agricultural Economics 90 (4), pp. 1011-1027.

Jamshidian, M. and R. Jennrich. (2002). “Standard Errors for EM Estimation.” Journal of the Royal Statistical Society: Series B 62(2), pp. 257-270.

Landry, C. and H. Liu. forthcoming. "A Semi-parametric Estimator for Revealed and Stated Preference data - An Application to Recreational Beach Visitation." Journal of Environmental Economics and Management forthcoming.

Lee, B. (1999). "Calling Patterns and Usuage of Residential Toll Service under Self Selecting Tariffs." Journal of Regulatory Economics 16(1), pp. 45-82.

McConnell, K. and W. Tseng. (2000). "Some Preliminary Evidence on Sampling of Alternatives with the Random Parameters Logit." Marine Resource Economics 14, pp. 317-332.

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

McFadden, D. (1974). "Conditional Logit Analysis of Qualitative Choice Behavior." In P. Zarembka, ed., Frontiers in Econometrics. New York: Academic Press.

McFadden, D. (1978). "Modeling the Choice of Residential Location," in Spatial Interaction Theory and Planning Models, ed. by A. Karlgvist, et al. Amsterdam: North-Holland.

McFadden, D. and P. Ruud. (1994). "Estimation by simulation." Review of Economics and Statistics LXXVI(4) pp. 591–608.

McLachlan, GJ and T. Krishnan. (1997). The EM Algorithm and Extensions. Hoboken, New Jersey: John Wiley & Sons Inc.

Meijer, E. and J. Rouwendal. (2006). "Measuring Welfare Effects in Models with Random Coefficients." Journal of Applied Econometrics 21(2), pp. 227-244.

Meng, X. and D. Rubin. (1991). "Using EM to Obtain Asymptotic Variance-Covariance Matrices: The SEM Algorithm." Journal of the American Statistical Association 86(416), pp. 899-909.

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

Murdock, J. (2006). "Handling Unoberserved Site Characteristics in Random Utility Models of Recreation Demand." Journal of Environmental Economics and Management 51(1), pp. 1-25.

Nerella, S. and C. Bhat. (2004). "Numerical Analysis of Effect of Sampling of Alternatives in Discrete Choice Models." Transportation Research Record 1894, pp. 11-19.

Parsons, G., A. Plantinga, and K. Boyle. (2000). "Narrow Choice Sets in a Random Utility Model of Recreation Demand." Land Economics 76(1) pp. 86-99.

Parsons, G. and M. Kealy. (1992). "Randomly Drawn Opportunity Sets in a Random Utility Model of Lake Recreation." Land Economics 68(1), pp. 93-106.

Parsons, G. and M. Needelman. (1992). "Site Aggregation in a Random Utility Model of Recreation." Land Economics 68(4) pp. 418-433.

Peters, T., W. Adamowic, and P. Boxall. (1995). “Influence of Choice Set Considerations in Modeling the Benefits from Improved Water Quality,” Water Resources Research 31(7) pp. 1781-1787.

Revelt, D. and K. Train. (1998). "Mixed Logit with repeated choices: households’ choices of appliance efficiency level." Review of Economics and Statistics 80, pp. 1–11.

Rigby, D., K. Balcombe, and M. Burton. (2008). "Mixed Logit Model Performance and Distributional Assumptions: Preferences and GM Foods." Environmental and Resource Economics, forthcoming.

Roeder, K., K. Lynch, and D. Nagin. (1999). "Modeling Uncertainty in Latent Class Membership: A Case Study in Criminology." Journal of the AMerican Statistical Association 94, pp 766.

Ruud, P. (1991) . "Extensions of Estimation Methods Using the EM Algortihm." Journal of Econometrics 49(3), pp. 305-341.

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

Sermons, M. and F. Koppelman. (2001). Representing the Differences Between Female and Male Commute Behavior in Residential Location Choice Models." Journal of Transport Geography 9(2), pp. 101-110.

Sethuraman, R., V. Srinivasan, and D. Kim. (1994). "Assymetric and Neighborhood Cross-Price Effects: Some Empirical Generalizations." Marketing Science 18(1), pp. 23-41.

Shocker, A., M. Ben-Akiva, B. Boccara, and P. Nedungadi. (1991). "Consideration Set Influences on Consumer Decision-making and Choice: Issues, Models, and Suggestions." Marketing Letters 2(3), pp. 181-197.

Thacher, J., E. Morey, and E. Craighead (2005). “Using Patient Characteristics and Attitudinal Data to Identify Depression Treatment Preference Groups: A Latent-Class Model.” Depression and Anxiety 21(2), pp.47-54.

Timmins, C. and J. Murdock. (2007). "A Revealed Preferenec Approach to the Measurement of Congestion in Travel Cost Models." Journal of Environmental Economics and Management 53(2), pp. 230-249.

Train, K. (2008). "EM Algorithms for Nonparametric Estimation of Mixing Distributions." Journal of Choice Modelling 1(1), pp. 40-69.

Train, K. and Sonnier, G. (2003). "Mixed logit with bounded distribution of partworths." Working paper. University of California, Berkeley and Los Angeles.

Train, K., D. McFadden, and M. Ben-Akiva. (1987). "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices." Rand Journal of Economics 18(1), pp. 109-123.

Train K (2003), Discrete Choice Methods with Simulation. Cambridge: Cambridge University Press.

Train, K. (1998). "Recreation Demand Models with Taste Differences over People." Land Economics 74(2), pp. 230-239.

von Haefen, R. and M. Jacobsen (unpublished). "Sampling of Alternatives in a Mixture Model."

von Haefen, R. (2008). "Latent Consideration Sets and Continuous Demand System Models," Environmental and Resource Economics 41(3), pp. 363-379.

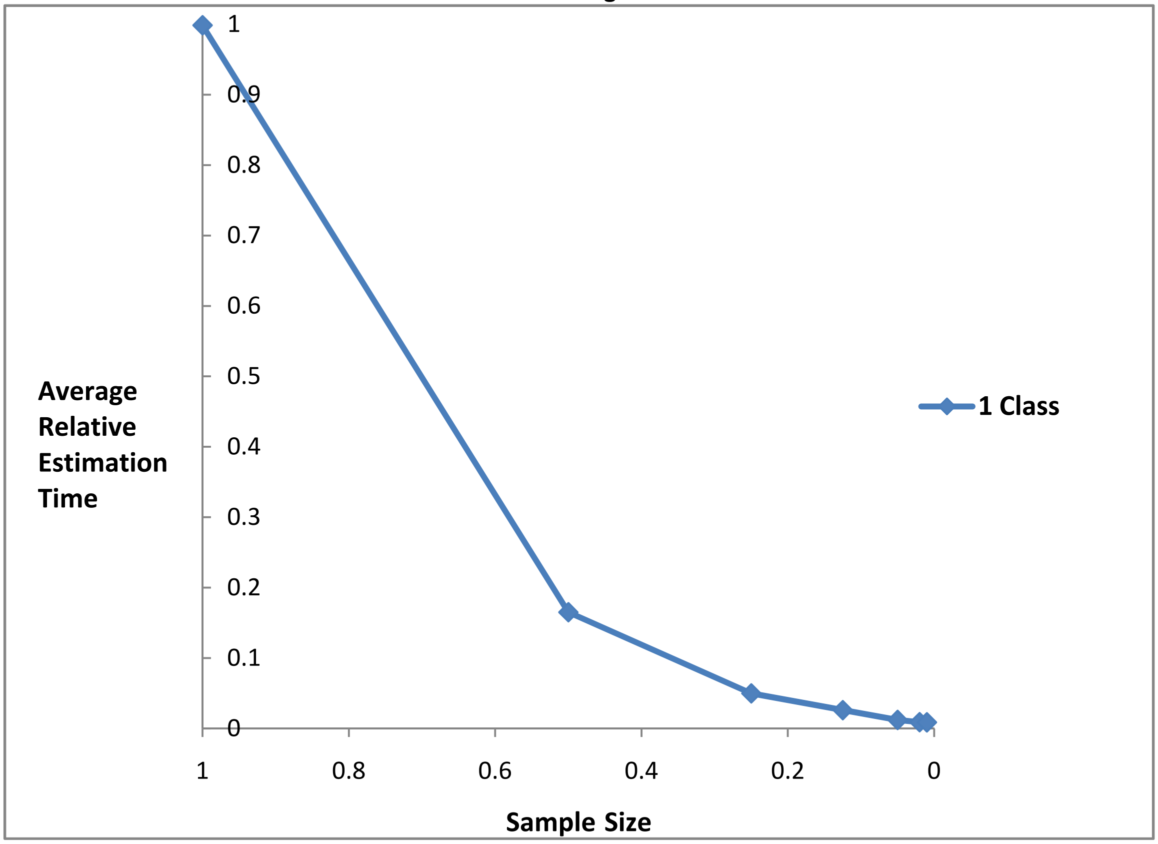
Waddell, P. Accessibility and Residential Location: The Interaction of Workplace, Residential Mobility, Tenure, and Location Choices. Presented at the Lincoln Land Institute TRED Conference, 1996 (http://www.odot.state.or.us/tddtpau/modeling.html).

TABLES AND FIGURES

Table 1
Summary Statistics

Variable	Description	Mean	St. Dev.
	Individual Summary Statistics		
trips	day trips during 1998 season	6.994	*7.182*
boat	dummy = 1 if household owns boat	0.514	-
kids	dummy = 1 if children under 14 in household	0.414	-
income	personal income	$28,991	*12,466*
	Site Summary Statistics		
tcost	round trip travel time x opp. cost of time +$0.15 x round trip miles	$100.70	*58.28*
ramp	dummy = 1 if site has at least one paved boat launch ramp	0.726	-
refuge	dummy = 1 if site is inside a wildlife area or refuge	0.056	-
forest	dummy = 1 if site is in a national, state, or county forest	0.178	-
urban	dummy = 1 if urban area on shoreline	0.179	-
restroom	dummy = 1 if restroom available	0.580	-
river	dummy = 1 if river fishing location	0.313	-
small lake	dummy = 1 if inland lake surface area <50 acres	0.172	-
trout	catch rate for brook, brown, and rainbow trout	0.094	*0.170*
smallmouth	catch rate for smallmouth bass	0.200	*0.205*
walleye	catch rate for walleye	0.125	*0.145*
northern	catch rate for northern pike	0.085	*0.057*
musky	catch rate for muskellunge	0.010	*0.022*
salmon	catch rate for coho and chinook salmon	0.009	*0.048*
panfish	catch rate for yellow perch, bluegill, crappie, and sunfish	1.579	*0.887*

Table 2
Estimation Time
Conditional Logit Model



* 1 ≈ 31 minutes

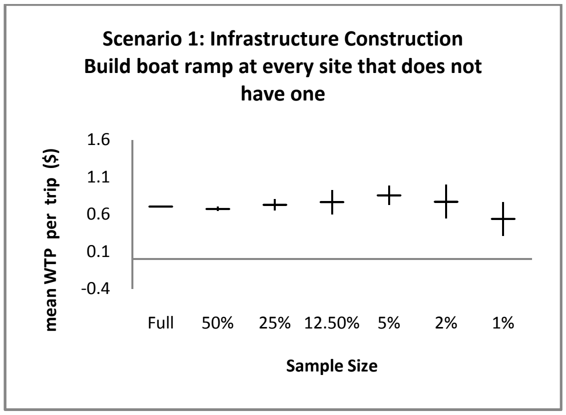
Table 3

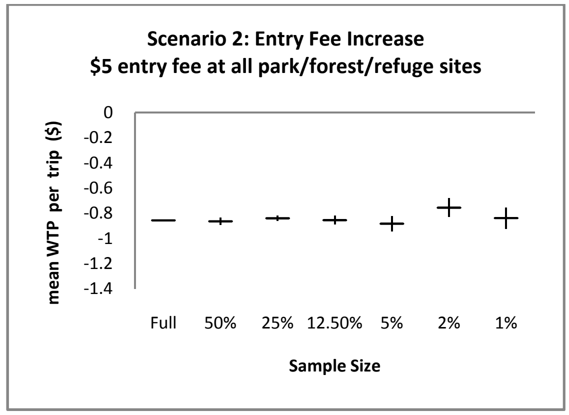
Parameter Estimates: Conditional Logit Model

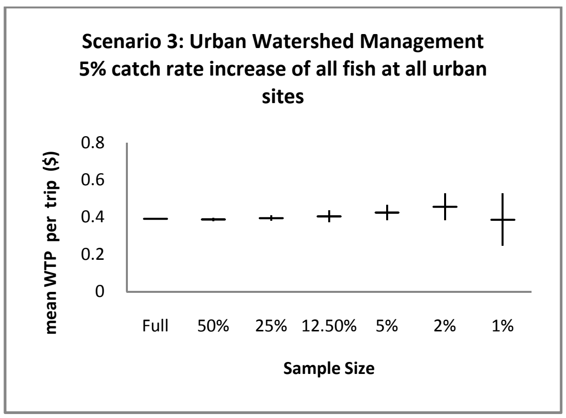
Sample Size	Full	50%	25%	12.50%	5%	2%	1%
LL	-13257	-10901	-8640	-6568	-4172	-2314	-1414
se		*42*	*56*	*65*	*75*	*73*	*63*
Normalized LL	-13257	-13264	-13274	-13294	-13344	-13432	-13542
se		*4*	*6*	*11*	*27*	*53*	*105*
Site Specific Variable							
tcost	**-10.070**	**-10.026**	**-9.907**	**-9.700**	**-9.303**	**-8.840**	**-8.600**
	0.412	***0.407***	***0.400***	***0.393***	***0.390***	***0.401***	***0.434***
ramp	**0.421**	**0.407**	**0.397**	**0.380**	**0.371**	0.349	0.343
	0.176	***0.177***	***0.178***	***0.179***	***0.184***	*0.195*	*0.217*
refuge	0.165	0.169	0.177	0.190	0.202	0.229	0.240
	0.194	*0.195*	*0.196*	*0.199*	*0.211*	*0.233*	*0.270*
forest	0.152	0.142	0.131	0.124	0.128	0.174	0.229
	0.171	*0.174*	*0.177*	*0.182*	*0.185*	*0.202*	*0.227*
urban	-0.068	-0.084	-0.092	-0.094	-0.080	-0.024	0.032
	0.117	*0.118*	*0.119*	*0.122*	*0.134*	*0.157*	*0.186*
restroom	0.149	0.149	0.144	0.143	0.147	0.174	0.211
	0.130	*0.131*	*0.131*	*0.130*	*0.134*	*0.148*	*0.168*
river	-0.013	-0.015	-0.020	-0.032	-0.070	-0.180	-0.266
	0.297	*0.299*	*0.305*	*0.319*	*0.348*	*0.399*	*0.458*
small lake	**-0.789**	**-0.789**	**-0.776**	**-0.765**	**-0.724**	**-0.707**	**-0.702**
	0.161	***0.164***	***0.168***	***0.175***	***0.185***	***0.209***	***0.236***
trout	**1.651**	**1.674**	**1.728**	**1.781**	**1.975**	**2.490**	**2.972**
	0.566	***0.571***	***0.584***	***0.605***	***0.668***	***0.797***	***0.933***
smallmouth	**0.943**	**0.948**	**0.972**	**0.996**	**1.050**	**1.149**	**1.174**
	0.359	***0.362***	***0.371***	***0.376***	***0.369***	***0.385***	***0.427***
walleye	**2.690**	**2.652**	**2.605**	**2.540**	**2.457**	**2.433**	**2.364**
	0.379	***0.376***	***0.374***	***0.380***	***0.405***	***0.479***	***0.561***
northern	**2.659**	**2.536**	**2.328**	2.013	1.505	0.908	0.658
	0.935	***0.955***	***0.998***	*1.070*	*1.215*	*1.472*	*1.746*
musky	**5.361**	**6.136**	**6.809**	**7.417**	**8.375**	**9.158**	**9.769**
	1.346	***1.752***	***2.069***	***2.324***	***2.585***	***2.803***	***3.376***
salmon	**7.733**	**7.852**	**7.968**	**8.043**	**8.139**	**7.848**	**7.545**
	1.384	***1.405***	***1.441***	***1.503***	***1.635***	***1.880***	***2.132***
panfish	**0.763**	**0.769**	**0.780**	**0.789**	**0.804**	**0.814**	**0.814**
	0.189	***0.189***	***0.191***	***0.194***	***0.201***	***0.223***	***0.253***

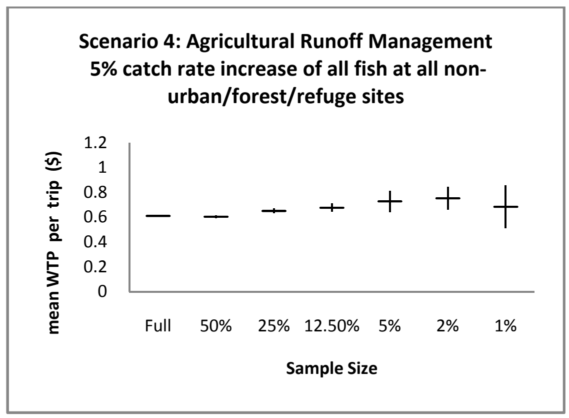
*results for the sampled models represent the mean of five random samples; robust clustered standard errors in italics; **bold** indicates significance at the 5% level; "Normalized LL" is the log-likelihood calculated at the parameter values for the entire choice set for comparison purposes.

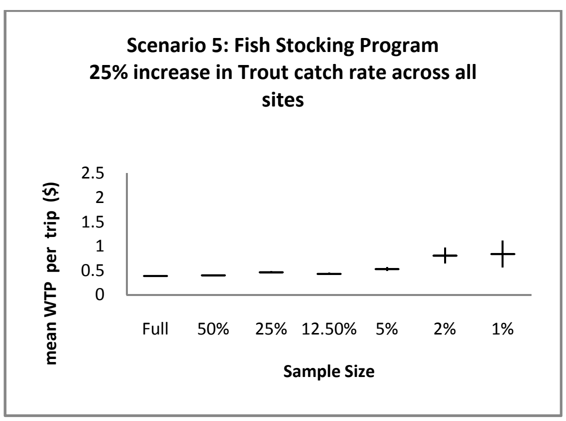
Table 4
Welfare Scenarios
Conditional Logit Model





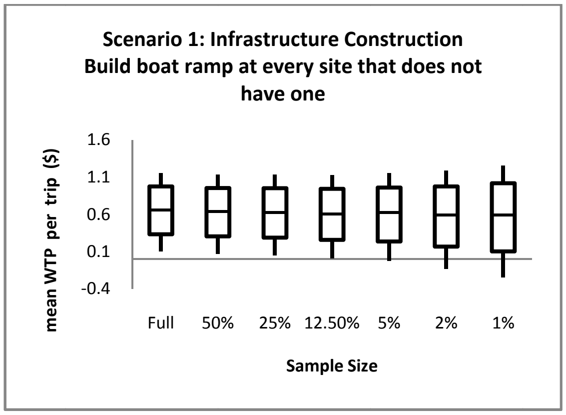


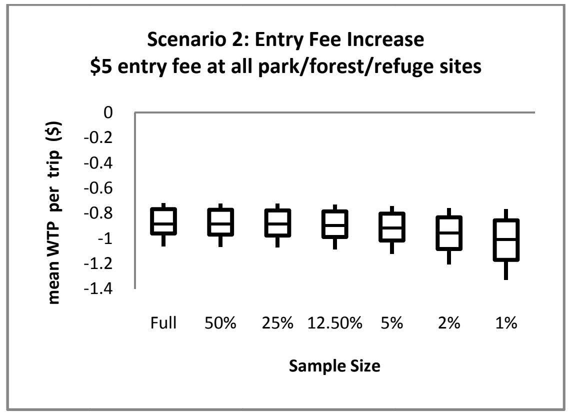


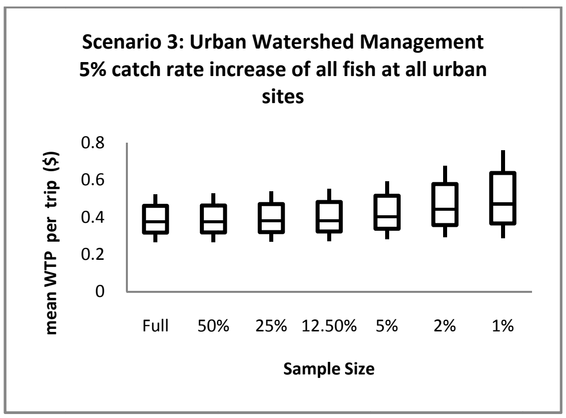


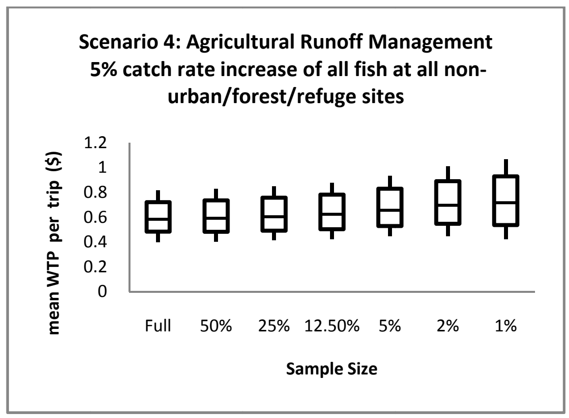
* Mean WTP of five unique samples, the mean and standard deviation of which are reported. Method: Small and Rosen (1981); Hanemann (1978) performed using the parameter estimates from the sample size specified, calculated using the full choice set.

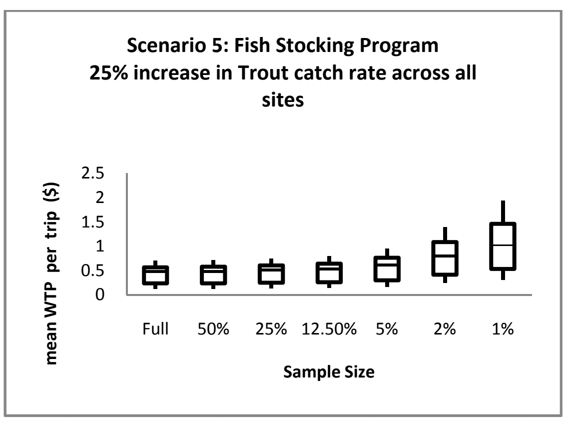
Table 5
Welfare Confidence Intervals
Conditional Logit Model





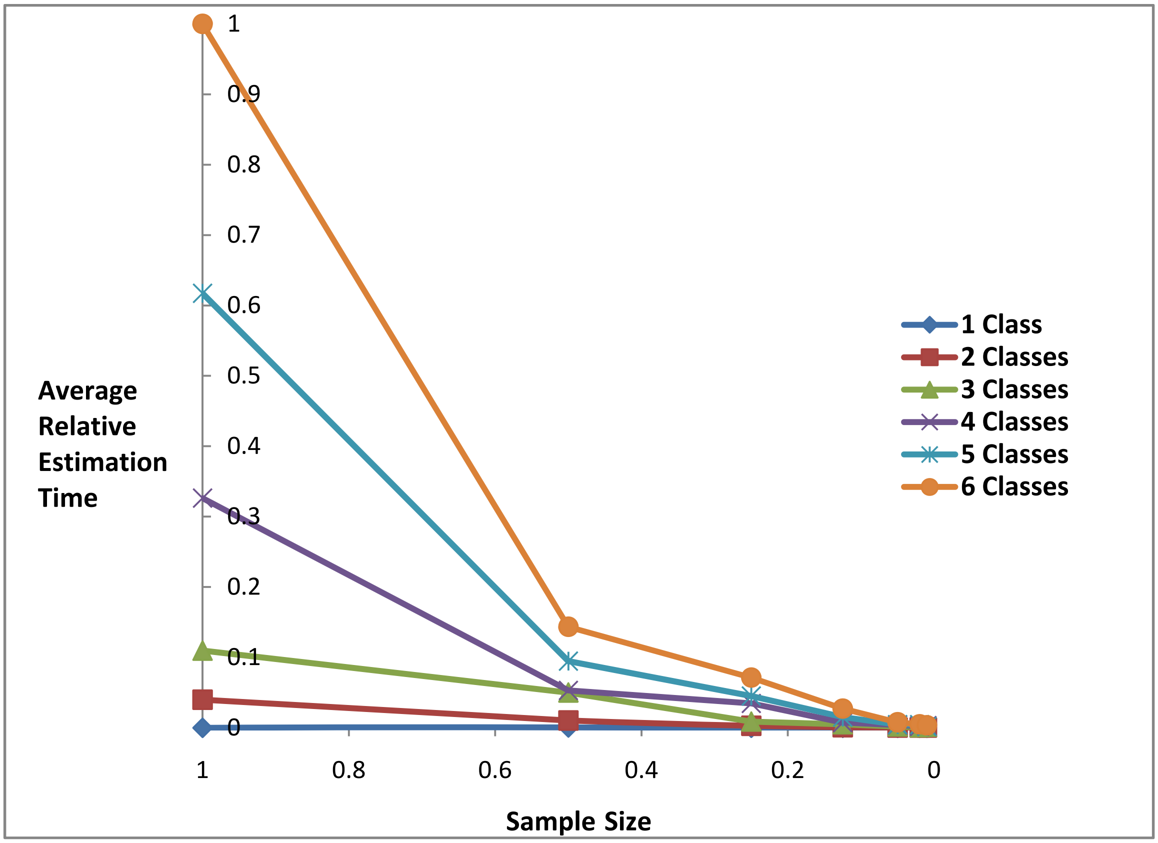






* Mean WTP of five unique samples, the mean of the mean, 95th, and 75th confidence intervals of which are reported. Method: Small and Rosen (1981); Hanemann (1978) performed using the parameter estimates from the sample size specified, calculated using the full choice set.

Table 6
Estimation Time
Latent Class Model



* 1 ≈ 6.99 days

Table 7
Information Criteria

	# of classes	Sample Size Full	50%	25%	12.5%	5%	2%	1%
LL	**1**	-13257	-13264	-13274	-13294	-13344	-13432	-13542
	2	-12417	-12438	-12526	-12570	-12633	-12918	-13416
	3	-12094	-12103	-12147	-12172	-12284	-12455	-13328
	4	-11793	-11817	-11879	-12019	-12024	-12262	-13310
	5	-11521	-11554	-11581	-11644	-12010	-12245	-13299
	6	-11236	-11275	-11356	-11456	-11995	-12234	-13294
CAIC	**1**	26623	26637	26657	26697	26797	26973	27193
	2	25073	25114	25291	25378	25505	26075	27071
	3	24557	24576	24662	24713	24938	25279	**27024**
	4	24085	24134	24256	24538	**24548**	**25023**	27119
	5	23672	23738	23791	23919	24649	25120	27229
	6	**23232**	**23309**	**23472**	**23672**	24750	25228	27348
crAIC	**1**	26560	26574	26594	26634	26734	26910	27130
	2	25065	25106	25283	25370	25497	26067	**27063**
	3	**24902**	**24921**	**25007**	**25058**	**25283**	**25624**	27370
	4	25279	25328	25450	25732	25742	26217	28313
	5	26438	26504	26557	26684	27415	27886	29994
	6	28563	28640	28803	29003	30081	30559	32679

* CAIC, and crAIC calculated using the "Normalized LL"; Mean of 200 and five random samples reported for the one and multiple class models respectively; Optimal # of classes in **bold** defined by the minimum of the information criteria.

Table 8

Parameter Estimates: Two Class Model

Sample Size	**Full**		**50%**		**25%**		**12.50%**		**5%**		**2%**		**1%**	
Normalized LL	-12416		-12426		-12431		-12464		-12506		-12890		-12944	
	Class 1	**Class2**	**Class 1**	**Class2**	**Class 1**	**Class2**	**Class 1**	**Class2**	**Class 1**	**Class2**	**Class 1**	**Class2**	**Class 1**	**Class2**
Individual Specific Variable														
intercept	**-1.138**	-	**-1.024**	-	**-1.140**	-	**-0.956**	-	**-1.008**	-	**-2.726**	-	**-0.518**	-
boat	**0.377**	-	**0.332**	-	**0.370**	-	**0.200**	-	**0.274**	-	**0.361**	-	**0.455**	-
kids	**-0.106**	-	-0.142	-	**-0.114**	-	**-0.136**	-	**-0.080**	-	**-0.136**	-	**-0.089**	-
income	**0.057**	-	0.059	-	**0.058**	-	**0.042**	-	**0.080**	-	**0.171**	-	**0.166**	-
Fixed Site Specific Variable														
tcost	**-10.287**	-	**-10.160**	-	**-10.102**	-	**-10.051**	-	**-9.730**	-	**-9.038**	-	-9.283	-
Site Specific Variable														
ramp	**0.716**	**0.161**	**0.615**	**0.137**	**0.776**	**0.159**	**0.597**	**0.216**	**0.965**	**0.042**	**0.766**	0.399	**0.828**	-1.253
refuge	**-1.857**	**0.950**	**-1.315**	**0.945**	**-2.141**	**0.945**	**-1.369**	**1.021**	**-0.567**	**0.851**	**0.697**	**0.531**	**-0.332**	1.203
forest	**-0.669**	**0.482**	**-0.762**	**0.499**	**-0.927**	**0.556**	-0.971	**0.633**	**-0.715**	**0.661**	**-1.545**	**0.521**	**-0.316**	0.758
urban	**-1.491**	**0.406**	**-1.393**	**0.388**	**-1.497**	**0.353**	**-1.284**	**0.399**	**-1.480**	**0.494**	**-0.827**	-0.023	**-1.979**	1.359
restroom	**-0.659**	**0.784**	**-0.613**	**0.828**	**-0.675**	**0.802**	**-0.601**	**0.947**	**-0.378**	**0.913**	**0.423**	**0.222**	**-0.159**	1.547
river	**3.460**	**-1.866**	**3.139**	**-2.116**	**3.421**	**-2.009**	**3.627**	**-2.147**	**3.187**	**-2.301**	**-0.39**	0.047	**0.605**	0.844
small lake	**-1.139**	**-0.465**	**-1.151**	**-0.376**	**-1.079**	**-0.442**	**-0.990**	**-0.438**	**-0.696**	**-0.539**	**-1.704**	**-0.567**	**-0.588**	-1.326
trout	**-0.796**	**2.398**	**-0.493**	**2.485**	**-0.724**	**2.530**	**1.318**	**2.503**	**-0.651**	**3.470**	**-24.625**	**2.54**	**1.662**	3.250
smallmouth	**1.851**	**0.809**	**1.380**	1.095	**1.417**	**0.995**	**0.758**	**1.221**	**1.040**	**1.367**	**5.524**	**0.434**	**0.947**	1.759
walleye	**4.527**	1.721	**4.294**	1.512	**4.768**	1.596	4.722	**1.700**	**3.325**	1.403	**5.96**	**0.907**	3.632	1.893
northern	**5.294**	**1.959**	**4.578**	**1.405**	**4.442**	1.495	**0.817**	**1.957**	**3.235**	1.523	**-8.625**	**0.957**	**3.953**	**-8.328**
musky	**4.169**	**7.587**	**4.369**	**9.136**	**1.490**	**7.453**	**6.291**	**7.932**	**6.820**	**9.281**	-2.549	**9.363**	**12.702**	6.522
salmon	**8.966**	**1.320**	**8.466**	**0.613**	**7.854**	**0.991**	8.170	**1.261**	**6.369**	**0.317**	**21.625**	**8.042**	**13.252**	9.639
panfish	**2.907**	**-0.474**	**2.723**	**-0.614**	**2.867**	**-0.524**	**2.970**	-0.477	**2.636**	**-0.445**	-1.564	**0.778**	**1.247**	1.474

*robust clustered standard errors not reported; **bold** indicates significance at the 5% level; “Normalized LL” is the log-likelihood calculated at the parameter values for the entire choice set for comparison purposes.

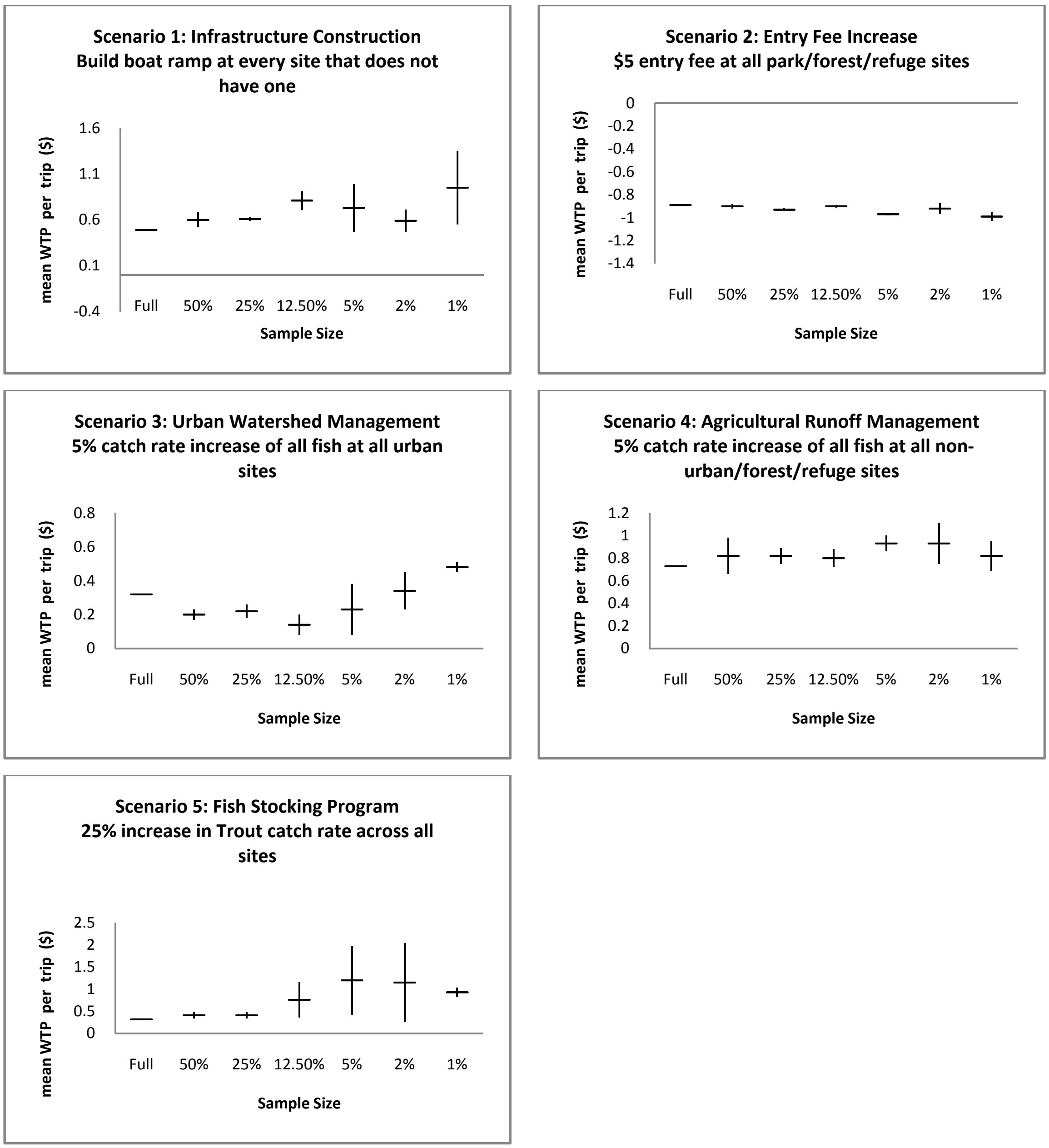
Table 9

Parameter Estimates: Latent Class Model

Sample Size	Full	50%	25%	12.50%	5%	2%	1%
Normalized LL	-12094	-12103	-12147	-12172	-12284	-12455	-13416
se		*6*	*17*	*90*	*81*	*48*	*96*
Site Specific Variable							
tcost	**-10.315**	**-10.397**	**-10.285**	**-9.887**	**-9.473**	**-9.151**	**-8.867**
	0.378	***0.288***	***0.312***	***0.313***	***0.487***	***0.439***	***2.274***
ramp	0.334	0.259	0.340	0.341	0.421	0.305	**1.940**
	0.239	*0.225*	*0.195*	*0.213*	*0.231*	*0.297*	***0.781***
refuge	0.131	-0.025	0.150	0.117	0.323	0.305	0.412
	0.278	*0.256*	*0.275*	*0.264*	*0.298*	*0.282*	*3.222*
forest	0.096	-0.059	-0.008	0.152	0.099	-0.157	0.061
	0.347	*0.366*	*0.343*	*0.313*	*0.343*	*0.353*	*0.231*
urban	-0.383	-0.260	-0.301	-0.269	-0.249	-0.196	0.098
	0.242	*0.305*	*0.239*	*0.231*	*0.270*	*0.304*	*0.221*
restroom	0.202	0.323	0.153	**0.353**	0.357	0.388	**0.484**
	0.199	*0.189*	*0.171*	***0.174***	*0.239*	*0.260*	***0.168***
river	-0.159	**-0.636**	**-0.637**	**-0.897**	**-0.837**	-0.310	0.112
	0.196	***0.191***	***0.185***	***0.194***	***0.206***	*0.244*	*0.151*
small lake	**-0.971**	**-1.086**	**-1.178**	**-0.752**	**-1.658**	**-1.436**	**-1.832**
	0.473	***0.417***	***0.443***	***0.505***	***0.528***	***0.564***	***0.409***
trout	**0.555**	**0.909**	**0.895**	**1.029**	**2.049**	**2.407**	0.943
	0.279	***0.260***	***0.271***	***0.266***	***0.311***	***0.348***	*1.783*
smallmouth	0.588	0.949	1.066	1.184	1.438	0.841	0.903
	0.987	*0.968*	*1.134*	*1.088*	*1.284*	*1.083*	*1.057*
walleye	**2.391**	**2.347**	**2.313**	**2.038**	**2.189**	**1.699**	**2.140**
	0.507	***0.557***	***0.502***	***0.527***	***0.509***	***0.617***	***0.417***
northern	**2.961**	**2.749**	**1.893**	**1.697**	0.104	**1.401**	**1.029**
	0.475	***0.511***	***0.509***	***0.579***	*0.623*	***0.658***	***0.438***
musky	**4.348**	**5.861**	**7.851**	**5.526**	**6.469**	**8.519**	**9.041**
	1.456	***1.434***	***1.571***	***1.579***	***2.010***	***2.231***	***1.536***
salmon	**7.309**	**4.850**	4.843	-4.182	2.936	4.787	**7.989**
	2.445	***2.346***	*2.937*	*3.206*	*3.928*	*4.183*	***3.232***
panfish	0.609	0.441	0.457	0.364	0.468	0.746	0.859
	1.905	*2.276*	*2.287*	*3.651*	*2.641*	*2.788*	*1.910*

*results for the sampled models represent the mean of 5 random samples using the optimal number of classes as defined by the crAIC and weighted parameter means; robust clustered standard errors in italics; **bold** indicates significance at the 5% level; "Normalized LL" is the log-likelihood calculated at the parameter values for the entire choice set for comparison purposes.

Table 10
Welfare Scenarios
Latent Class Model (crAIC)



* Mean WTP of five unique samples using the best of three starting values, the mean and standard deviation of which are reported. Method: Small and Rosen (1981); Hanemann (1978) performed using the parameter estimates from the sample size specified, calculated using the full choice set.