Estimating Mixed Logit Models with Large Choice Sets

Roger H. von Haefen and Adam Domanski

April 2013

Word Count: 9,904

Abstract

We use the expectation-maximization (EM) algorithm to consistently estimate a latent class, mixed logit model with a sample of alternatives. Our approach represents an econometrically tractable strategy for estimating discrete choice models with large choice sets that do not rely on the Independence of Irrelevant Alternatives (IIA) assumption. We present Monte Carlo evidence suggesting our approach generates consistent estimates, and then apply the method to a data set of Wisconsin anglers. Of interest to applied researchers, our empirical results quantify the tradeoff between model run-time and the efficiency/precision of welfare estimates associated with samples of different sizes.

1. Introduction

Researchers frequently apply discrete choice methods to revealed preference data to estimate consumer preferences for the characteristics of quality-differentiated goods such as automobiles (Bento et al. 2009), housing (Bayer et al. 2007) and recreation sites (von Haefen and Phaneuf 2008). A large and growing empirical literature suggests that these methods are well suited for representing extensive margin decisions from large choice sets where substitution is important. However, when the agent’s choice set becomes very large (on the order of hundreds or thousands of alternatives), computational limitations can make estimation with the full choice set difficult, if not intractable. McFadden (1978) suggested that using a sample of alternatives in estimation can obviate these difficulties and produce consistent parameter estimates. His
approach has been widely used in several empirical settings (e.g., Bayer et al., 2007; Parsons and Kealy 1992; Feather 1994; Parsons and Needelman 1992). When implementing the sampling of alternatives approach, researchers typically assume that unobserved utility is independently and identically distributed type I extreme value. Independence implies that the odds ratio for any two alternatives does not change with the addition of a third alternative. This property, known as the independence of irrelevant alternatives (IIA), is necessary for consistent estimation under the sampling of alternatives approach, but is often a restrictive and inaccurate characterization of choice.

In recent years, applied researchers have developed several innovative models that relax IIA by exploiting recent computational advances. Perhaps the most notable and widely used is the mixed logit model (McFadden and Train 1998). Mixed logit models generalize the conditional logit model by introducing unobserved preference heterogeneity through the parameters (Train 1998). This variation allows for richer substitution patterns and thus makes the mixed logit model an attractive tool for discrete choice modeling. However, adopting it comes at a significant cost – there is no proof that sampling of alternatives within the mixed logit framework generates consistent parameter estimates. Consequently, researchers adopting the sampling of alternatives approach are forced to choose either asymptotic unbiasedness and restrictive substitution patterns with conditional logit or potential asymptotic bias and more flexible substitution patterns with mixed logit.

Additionally in a mixed logit model, preference heterogeneity is often introduced through analyst-specified parametric distributions for the random parameters. The researcher's choice of error distribution thus becomes an important modeling judgment. The normal distribution is often employed in practice, although its well-known restrictive skewness and kurtosis properties raise the possibility of misspecification. Alternative parametric mixing distributions have been proposed (e.g., truncated normal, log normal, triangular, uniform), but in each case misspecification of the underlying distribution of preferences remains a concern (see Hess and Rose, 2006; Dekker and Rose, 2011).

We propose that both problems can be overcome through the use of a finite mixture (or latent class) model estimated via the expectation-maximization (EM) algorithm. The latent class framework probabilistically assigns individuals to classes, where preferences are heterogeneous across – but homogeneous within – classes. This approach allows the researcher to recover separate preference parameters for each consumer type without assuming a parametric mixing distribution. As demonstrated by Swait (1994), latent class models can be conveniently 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 recursive estimation of IIA conditional logit models. By reintroducing the IIA property at each maximization step of the recursion, sampling of alternatives can be used to generate consistent parameter estimates.

4 Consumer types can be driven by any combination of attitudinal, spatial, demographic, or other variation in the population.
In this paper, we report results from a detailed Monte Carlo simulation that strongly suggest the consistency of the approach. Using the simulation results as guidance, we then empirically evaluate the welfare implications of this novel estimation strategy using a recreational dataset of Wisconsin anglers. The Wisconsin dataset is attractive for this purpose because it includes a large number of recreational destination alternatives (569 in total) that allows us to test the sampling of alternatives approach against estimation using the full choice set. By comparing estimates generated with the full choice set to estimates generated with samples\(^5\) of alternatives of different sizes, we can compare the benefits and costs of sampling of alternatives in terms of estimation run time, small sample bias, and efficiency loss. Our strategy involves repeatedly running latent class models on random samples of alternatives of different sizes. In particular, we examine the effects of using sample sizes of 285 (50%), 142 (25%), 71 (12.5%), 28 (5%), 11 (2%), and 6 (1%) alternatives. Our results suggest that for our preferred latent class specification, using a 71-alternative sample size will generate on average a 75% time savings and 51% increase in the 95% confidence intervals for the five willingness-to-pay measures we construct. We also find that the efficiency losses for sample sizes as small as 28 alternatives may be sufficiently informative for policy purposes, but that smaller sample sizes often generate point estimates with very large confidence intervals.

These results provide useful guidance for researchers, policymakers, and other practitioners interested in estimating models with large choice sets. The overall time saved during estimation can allow researchers to model a broader and more complex set of specifications. Additionally, time-saving techniques enable practitioners to explore alternative specifications before dedicating limited resources to estimating final models. This flexibility can be especially useful in the early stages of data analysis when the researcher’s goal is to quickly identify promising specifications deserving further study. And while processor speed is constantly improving, this method will always be available to estimate models at or beyond the frontier of computing power.

The paper proceeds as follows. Section two summarizes the conditional and mixed logit models. Section three describes large choice set problems in discrete choice modeling. Section four details the latent class model estimated via the EM algorithm as well as sampling of alternatives in a mixture model. Section five presents the results of our Monte Carlo simulation. Section six presents our empirical application with the Wisconsin angler dataset. Section seven concludes with a discussion of directions for future research.

2. The Discrete Choice Model

This section reviews the conditional logit model, the IIA assumption, and the mixed logit model with continuous and discrete mixing distributions. We begin by briefly discussing the generic structure of discrete choice models. Economic applications of discrete choice models employ the random utility maximization (RUM) hypothesis and are widely used to model and predict qualitative choice outcomes (McFadden 1974). Under the RUM hypothesis, utility maximizing agents are assumed to have complete knowledge of all factors that enter preferences

\(^5\) Throughout this paper, “sample size” will refer to the sample of alternatives, not the sample of observations.
and determine choice. However, the econometrician’s knowledge of these factors is incomplete, and therefore preferences and choice are random from her perspective. By treating the unobserved determinants of choice as random draws from a distribution, the probabilities of choosing each alternative can be derived. These probabilities depend in part on a set of unknown parameters which can be estimated using one of many likelihood-based inference approaches.

More concretely, the central building block of discrete choice models is the conditional indirect utility function, $U_{ni}$, where $n$ indexes individuals and $i$ indexes alternatives. A common assumption in empirical work is that $U_{ni}$ can be decomposed into two additive components, $V_{ni}$ and $\varepsilon_{ni}$. $V_{ni}$ embodies the observable determinants of choice, $x_{ni}$, as well as preference parameters, $\beta_n$, and is typically assumed to take a linear functional form, i.e., $V_{ni} = \beta_n x_{ni}$. $\varepsilon_{ni}$ captures those factors that are unobserved and idiosyncratic from the analyst’s perspective. Under the RUM hypothesis, individual $n$ selects recreation site $i$ if it generates the highest utility from the available set of $J$ alternatives:

$$\text{alternative } i \text{ chosen iff } \beta_n x_{ni} + \varepsilon_{ni} > \beta_n x_{nj} + \varepsilon_{nj}, \forall j \neq i.$$

2.1 Conditional Logit

Different distributional specifications for $\beta_n$ and $\varepsilon_{ni}$ generate different empirical models. One of the most widely used models is the conditional logit which arises when $\beta_n = \beta$, $\forall n$ and each $\varepsilon_{ni}$ is an independent and identically distributed (iid) draw from the type I extreme value distribution with scale parameter $\mu$. The probability that individual $n$ chooses alternative $i$ takes the well-known form (McFadden 1974):

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

where the second equality follows from the fact that $\beta$ and $\mu$ are not separately identified and thus, with no loss in generality, $\mu$ is normalized to one.

As stated in the introduction, the conditional logit model embodies the IIA property, meaning that the odds ratio for any two alternatives is unaffected by the inclusion of any third alternative. IIA’s restrictive implications for behavior can best be appreciated in terms of substitution patterns resulting from the elimination of a choice alternative. Consider, for example, the case of a recreational site closure due to an acute environmental incident. Assume the closed site has unusually high catch rates. Intuitively, individuals who previously chose the closed site likely have a strong preference for catching fish and would thus substitute to other sites with relatively high catch rates when their preferred site is closed. IIA and the conditional logit model predict, however, that individuals would shift to other sites in proportion to their selection probabilities. In other words, those sites with the highest selection probabilities would see the largest increase in demand even if they do not have relatively high catch rates.

2.2 Mixed Logit

To relax IIA and introduce non-additive unobserved preference heterogeneity, applied researchers frequently specify a mixed logit model (Train 1998; McFadden and Train 2000).

4
Mixed logit generalizes the conditional logit by introducing unobserved taste variations for attributes through the coefficients. This is accomplished by assuming a probability density function for $\beta$, $f(\beta \mid \theta)$, where $\theta$ is a vector of parameters. Introducing preference heterogeneity in this way results in correlation in the unobservables for alternatives with similar attributes and thus relaxes IIA. Conditional on $\beta$, the probability of selecting alternative $i$ in the mixed logit is:

$$
P_m(\beta) = \frac{\exp(\beta x_i)}{\sum_j \exp(\beta x_j)}.
$$

The probability densities for $\beta$ can be specified with either a continuous or discrete mixing distribution. With a continuous mixing distribution, the unconditional probability of selecting alternative $i$ is:

$$
P_n = \int P_m(\beta) f(\beta \mid \theta) d\beta.
$$

When the dimension of $\beta$ is moderate to large, good analytical or numerical approximations for the above integral are generally not possible; however, $P_n$ can be approximated via simulation (Geweke et al. 1994; McFadden and Ruud 1994). This involves generating several pseudo-random draws from $f(\beta \mid \theta)$, calculating $P_n(\beta)$ for each draw, and then averaging across draws. By the law of large numbers, this simulated estimate of $P_n$ will converge to its true value as the number of simulations grows large.

In practice, a limitation with the continuous mixed logit model is that the mixing distribution often takes an arbitrary parametric form. Several researchers have investigated the sensitivity of parameter and welfare estimates to the choice of alternative parametric distributions (Revelt and Train 1998; Train and Sonnier 2003; Rigby et al. 2008; Hess and Rose 2006). The consensus finding is that distribution specification matters. For example, Hensher and Greene (2003) studied the welfare effect of a mixed logit model with lognormal, triangular, normal, and uniform distributions. Although the mean welfare estimates were very similar across the normal, triangular, and uniform distributions, the lognormal distribution produced results that differed by roughly a factor of three. And although the mean welfare estimates were similar across the triangular, normal, and uniform distributions, the standard deviations varied by as much as 17 percent.

Concerns about arbitrary distributional assumptions have led many researchers to specify discrete or step function distributions that can readily account for unknown features of the data. The unconditional probability is the weighted sum of logit kernels:

$$
P_n = \sum_c S_m(z_n, \delta) P_n(\beta_c),
$$

where $S_m(z_n, \delta)$ is the probability of being in class $c$ ($c = 1, \ldots, C$) while $z_n$ and $\delta$ are observable demographics and parameters that influence class membership, respectively. If the class membership probabilities are independent of $z_n$, then the mixing distribution has a nonparametric or “discrete-factor” interpretation (Heckman and Singer 1984). More commonly,
however, the class membership probabilities depend on observable demographics that parsimoniously introduce additional preference heterogeneity. In these cases, the class probabilities typically assume a logit structure:

$$S_{mc} (z_n, \delta) = \frac{\exp(\delta z_n)}{\sum_{i=1}^{C} \exp(\delta_i z_n)}.$$  

where $\delta = [\delta_1, ..., \delta_C].$

3 Large Choice Sets

The specification of the choice set is a critical modeling judgment with the 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). The combination of these factors in a given application can lead to large choice set specifications (Parsons and Kealy 1992, Parsons and Needelman 1992) that raise computational issues in estimation.6

There are three generic strategies for addressing the computational issues raised by large choice sets: 1) aggregation, 2) separability, and 3) sampling. Solutions (1) and (2) require additional assumptions about preferences or price and quality movements within the set of alternatives. Even when one or both of the first two approaches are adopted, the third approach – sampling – may still be appealing due to computational considerations.

As McFadden (1978) has shown, estimation with a sample of alternatives works as follows. The analyst combines the chosen alternative with a random sample of alternatives within traditional maximum likelihood estimation. This simplifies computational burden and produces consistent estimates as long as the uniform conditioning property7 holds. Moreover, other sampling schemes can produce consistent estimates (e.g., importance sampling as in Feather 1994) as long as the site selection mechanism is properly controlled for.

A maintained assumption in McFadden’s (1978) consistency proof is that the ratio of choice probabilities for any two alternatives does not change with the elimination of a third alternative, or IIA. Despite this restriction, sampling of alternatives has been widely utilized in

---

6 Here we are abstracting from the related issue of consideration set formation (Manski 1977), or the process by which individuals reduce the universal set of choice alternatives down to a manageable set from which they seriously consider and choose. Consideration set models have received increased interest in recent environmental applications despite their significant computational hurdles (Haab and Hicks 1997; von Haefen 2008). Nevertheless, to operationalize these models the analyst must specify the universal set from which the consideration set is generated as well as the choice set generation process. In many applications, the universal set is often very large.

7 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.
the applied literature with fixed parameter logit and nested logit models (Parsons and Kealy 1992; Bhat et al. 1998; Guo and Bhat 2001; Ben-Akiva and Bowman 1998; Bayer et al. 2007). However, because random parameter models relax IIA, the consistency of parameter estimation via maximum likelihood with a sample of alternatives has not been established.

This gap in theoretical understanding raises a practical question: how bad is it to use a sample of alternatives with a non-IIA model? McConnell and Tseng (2000) and Nerella and Bhat (2004) explore this issue using real or synthetic data. Using two recreational data sets with 10 sites each, McConnell and Tseng (2000) find that samples of four, six, and eight alternatives generate parameter and welfare estimates that on average are qualitatively similar to estimates based on the full choice set of ten alternatives. However, their conclusions are based on only 15 replications and thus should be interpreted cautiously. Nerella and Bhat (2004) perform a similar analysis with synthetic data. Based on simulations with 200 alternatives and 10 replications, they find small bias for sample sizes greater than 50 alternatives.

4 Sampling in a Mixture Model

We now introduce our proposed sampling of alternatives approach to estimating discrete choice models that exploits a variation of the expectation-maximization (EM) algorithm. We describe the estimator below.

4.2 EM Algorithm

The EM algorithm (Dempster et al. 1977) is an estimation framework for recovering parameter estimates from likelihood-based models when traditional maximum likelihood is computationally difficult. It is a popular tool for estimating models with incomplete data (McLachlan and Krishnan 1997) and mixture models (Bhat 1997a; Bhat 1997b; Train 2008). The method also facilitates consistent estimation with a sample of alternatives as we describe below.

The EM algorithm is a recursive procedure with two steps. The first is the expectation or “E” step, whereby the expected value of the unknown variable (class membership in our case) is constructed using Bayes’ rule and the current estimate of the model parameters. The maximization or “M” step follows: the model parameters are then updated by maximizing the expected log-likelihood which is constructed conditional on the probabilities from the “E” step. The E and M steps are then repeated until convergence (Train 2008). This recursion is often an attractive estimation strategy relative to gradient-based methods because it transforms the computationally difficult maximization of a “log of sums” into a simpler recursive maximization of the “sum of logs.”

More concretely, the EM algorithm works as follows in the latent class context. Given parameter values \( \hat{\theta}' = (\beta', \delta') \) and log-likelihood function:

\[
LL_n = \ln \left( \sum_c S_{nc} (\delta) L_n(\beta') \right),
\]

Bayes’ rule is used to construct the conditional probability that individual \( n \) is a member of class \( c \):
These probabilities serve as weights in the construction of the expected log-likelihood which is then maximized to generate updated parameter values:

$$\phi^{t+1} = \arg\max_\phi \sum_{n=1}^{N} \sum_{c=1}^{C} h_{nc}(\phi') \ln\left( S_{nc}(\delta') L_n(\beta_c') \right)$$

where \( N \) is the number of observations. Since

$$\ln\left( S_{nc}(\delta)L_n(\beta_c) \right) = \ln\left( S_{nc}(\delta) \right) + \ln\left( L_n(\beta_c) \right),$$

the maximization can be performed separately for the \( \delta \) and \( \beta \) parameters. In other words, the class membership parameters \( \delta \) can be updated with one maximization:

$$\delta^{t+1} = \arg\max_\delta \sum_{n=1}^{N} \sum_{c=1}^{C} h_{nc}(\phi') \ln S_{nc}(\delta),$$

and the \( \beta \) parameters entering the conditional likelihoods are updated with a second:

$$\beta^{t+1} = \arg\max_\beta \sum_{n=1}^{N} h_{nc}(\phi') \ln L_n(\beta_c).$$

These steps are repeated until convergence, defined as a small change in parameter values across iterations. It should be noted that this recursion may converge at a local optimum because the likelihood function is not globally concave. To address this possibility, researchers typically employ multiple starting values and chose the parameter values that imply the highest likelihood.

For large choice set problems, it is important to recognize that the M step involves (weighted) logit estimation. Therefore, IIA holds and a sample of alternatives can generate consistent parameter estimates and reduce the computational burden. Our Monte Carlo evidence strongly suggests that this modified version of the EM algorithm generates consistent parameter estimates.

Two details associated with implementation are worth emphasizing here. First, the outlined strategy implies using the sample of alternatives at the M step, but the full set of alternatives at the E step. Second, to avoid simulation chatter, the sample of alternatives should be held constant across iterations in the recursion. Fixing the sample of alternatives is akin to the standard procedure of fixing the random draws in maximum simulated likelihood estimation.

5 Monte Carlo Analysis

To better understand the properties of our proposed and maximum likelihood-based estimation strategies with samples of alternatives, we conducted a detailed Monte Carlo investigation with alternative logit models. Our simulations expand upon past research by McConnell and Tseng (2000) and Nerella and Bhat (2004) who explore the performance of

---

8 The E step is a calculation which updates the latent class probabilities. Using the full choice set here is not computationally burdensome relative to the maximization which occurs at the M step.
random parameter logit models estimated with a sample of alternatives via maximum likelihood. We investigate a wider set of empirical specifications – i.e., normal and discrete (i.e., latent class) mixing distributions – and employ significantly more Monte Carlo replications which should generate more reliable inference. We also investigate the performance of our proposed EM-based estimator, and although our simulations do not formally prove its consistency, they strongly suggest that this is indeed the case.

The setup of our Monte Carlo is as follows. We simulate $N = 500, 1000, 2000$ individuals each making a single discrete choice from a choice set of 100 alternatives. Four attributes that vary independently across individuals and alternatives characterize each alternative. We assume individual choices are generated according to one of three models: 1) a fixed-parameter, conditional logit model with four parameters; 2) a random coefficient model with two fixed and two continuously distributed (i.e., normal) random parameters; and 3) a two-class, latent class model with two fixed and two random parameters. For the latent class model, we assume the class membership probabilities are functions of two parameters that interact with a constant and an individual-specific variable. For each specification and sample size, we simulate 250 independent samples that form the basis of our analysis. We then estimate models with a random sample of alternatives of different sizes (5, 10, 25, 50 and 100) with each sample and assess how the maximum likelihood and (in the case of the latent class specification) EM algorithm estimators perform.

We begin with a discussion of our fixed-parameter, conditional logit results. With this specification, IIA is assumed, so McFadden’s consistency proof suggests that maximum likelihood estimators should perform well. As summarized in Figure 1, our Monte Carlo results confirm this expectation. Figure 1 graphically presents two summary statistics: 1) mean parameter bias (MPB), or the parameter estimate divided by its true value; and 2) proportional standard error (PSE), or the standard deviation across the 250 replications divided by the parameter’s true value and rescaled by 100. For compactness, Figure 1 only reports the average values across the four fixed parameters. MPB (PSE) correspond to the black (gray) solid ($N = 2000$), dashed ($N = 1000$), and dotted ($N = 500$) lines. A well-performing estimator should generate MPB estimates near one (implying minimal bias), and for all sample sizes, we generally find this to be the case across all sample of alternatives sizes. The only exception is for

---

9 To our knowledge such a consistency proof for parameter estimates estimated via the EM algorithm does not exist, and the fact that we are employing a sample of alternatives in estimation adds a level of complexity that has not been theoretically explored in the context of models estimated via the EM algorithm.

10 In the results presented in this section, we generated the independent variables by simulating independent and identically distributed (iid) draws from the standard normal distribution. Although not reported here, we also generated the attributes by simulating iid draws from non-normal distributions, and the conclusions we draw from these alternative runs were qualitatively similar.

11 Stated differently, the fixed parameters are equal across the two classes whereas the random parameters are class-specific.

12 We report the individual parameter values in an appendix available upon request. The average and individual parameter values are qualitatively similar, so in the interest of brevity, we report only the averages in this section.
the 500 individual, five alternative specification, where the relatively small sample size likely explains the modestly larger bias. Consistent with our a priori expectations, the PSE results suggest that smaller samples of alternatives imply larger standard errors, with the largest standard errors corresponding to the smallest sample of alternatives. These findings serve as the benchmark against which we can compare our results from non-IIA models.

Results from the continuously distributed mixed logit model estimated via maximum likelihood are presented in Figure 2. Here we graphically present the results in three panels where the top, middle, and bottom panels report results for the fixed parameters, the normally distributed random parameters, and the standard deviations of the random parameters, respectively. Two main findings arise from these graphs. First, there appears to be significant attenuation bias with the mean and standard deviation parameters for the random parameters, and this bias grows as the size of the sample of alternatives declines. Second, the proportional standard errors rise dramatically for the 500 individual, 5 alternative models, suggesting that models estimated with limited data are likely to be imprecise.

Results from the continuously distributed mixed logit model estimated via maximum likelihood are presented in Figure 2. Here we graphically present the results in three panels where the top, middle, and bottom panels report results for the fixed parameters, the normally distributed random parameters, and the standard deviations of the random parameters, respectively. Two main findings arise from these graphs. First, there appears to be significant attenuation bias with the mean and standard deviation parameters for the random parameters, and this bias grows as the size of the sample of alternatives declines. Second, the proportional standard errors rise dramatically for the 500 individual, 5 alternative models, suggesting that models estimated with limited data are likely to be imprecise.

Figure 3 presents the latent class model results estimated via maximum likelihood. Here we graphically present results for three distinct parameter groups: 1) fixed parameter means, 2) random parameter means (i.e., the parameters entering the conditional indirect utility functions that are class-specific); and 3) the class membership parameters. Similar to Figure 2, we find attenuation bias with the random parameters that increases as the sample size decreases and relatively large imprecision with the 500 individual, 5 alternative model. We also find upward bias with the class membership parameters estimated with small samples of alternatives. Combined, the results presented in Figures 2 and 3 suggest the significant pitfalls arising from maximum likelihood estimation of non-IIA models with samples of alternatives.

Figure 4 presents the Monte Carlo results for the latent class models estimated with our proposed EM algorithm. Compared to Figure 3, these results generally suggest that sampling of alternatives combined with the EM algorithm produce unbiased parameter estimates for sample sizes with at least 1000 observations. Smaller sample sizes (N = 500) apparently do not have sufficient information to estimate accurately and efficiently all parameters with only 5 alternatives. In all other cases, however, our proposed EM algorithm appears to generate reliable estimates that improve with increases in observations and sampled alternatives. In sum, these findings suggest that our proposed estimator is consistent.

6 Empirical Investigation

This section describes the empirical example, including information about the data set along with a comparison of results from the conditional logit and latent class models.

13 We used three sets of randomly generated starting values around the true parameters values during the estimation of all latent class models. Experimentation with more starting values suggested that employing only three sets in our Monte Carlo was sufficient.
6.2 Data

An empirical illustration is performed with data from the Wisconsin Fishing and Outdoor Recreation Survey. Collected in 1998 by Triangle Economic Research, this dataset has been investigated 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. In many ways this is an ideal dataset to evaluate the consistency of sampling of alternatives: it is large enough that a researcher might prefer to work with a smaller choice set to avoid computational difficulties, but small enough that estimation of the full choice set is still feasible for comparison. Table 1 presents summary statistics.

| Table 1 about here |

The full choice set is estimated with both a conditional logit model and several latent class specifications. The parameter results are evaluated and information criteria are used to compare improvements in fit across specifications. The same estimation is then also performed on randomly sampled choice sets of 285, 142, 71, 28, 11, and 6\(^{14}\) of the non-selected alternatives. The sampling properties of the conditional logit model will be used to benchmark the latent class results. Since we use the outer product of the gradient formula to recover standard errors in the latent class model, we will use the same method with the conditional logit model.

6.3 Conditional Logit Results

Estimation code was written and executed in Gauss and duplicated in Matlab. In contrast to the latent class model, the likelihood function for the conditional logit model is globally concave so starting values will affect run times but not convergence. A complicating factor with our dataset is that individuals make multiple trips to multiple destinations, i.e., our data has a panel structure. For consistency of the parameter estimates, it is necessary to generate a random sample of alternatives for each individual-site visited pair. For a sample size of \(M\), \(M-1\) alternatives were randomly selected and included with the chosen alternative. 200 random samples were generated for each sample size.

| Table 2 about here |

The parameter estimates and standard errors for each of the sample sizes are shown in Table 2. The means of the estimates and means of the standard errors from the 200 random samples are reported. Two log-likelihood values are reported in this table: the “sampled log-likelihood” (SLL) and the “normalized log-likelihood” (NLL). In any sampled model, a smaller set of alternatives will generally result in a larger log-likelihood. This number, however, is not useful in comparing goodness-of-fit across different sample sizes. The NLL is reported for this reason. After convergence is reached in a sampled model, the parameter estimates are used with

\(^{14}\) 50%, 25%, 12.5%, 5%, 2%, and 1% sample sizes, respectively.
the full choice set to compute the log-likelihood. A comparison of the NLL across samples shows that, when sampling, the reduction in information available in each successive sample reduces goodness of fit, as expected. A decrease in the sample size also increases the standard errors of the NLL reflecting the smaller amount of information used in estimation.

The parameters themselves are sensible (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 an efficiency loss when less data is used. In the smallest samples, this decrease in fit is enough to make parameters that are significant with the full choice set insignificant.

Table 2 suggests that parameter estimates are somewhat sensitive to sample size, but the welfare implications of these differences are unclear. To investigate this issue, welfare estimates for five different policy scenarios are constructed from the parameter estimates summarized in Table 2. The following policy scenarios are considered (see Table 3): 1) infrastructure construction, 2) an increase in entry fees, 3) an urban watershed management program, 4) an agricultural runoff management program, and 5) a fish stocking program. Note that general equilibrium congestion effects are not considered here (Timmins and Murdock 2007), but these scenarios can be augmented or modified to fit any number of policy proposals.

Table 3 – Welfare Scenarios

The methodology used to calculate WTP is the log-sum formula derived by Hanemann (1978). Given our constant marginal utility of income \( \beta_p f(y - p_j) = \beta_p (y - p_j) \) and a price and attribute change from \((p_0, q_0)\) to \((p_1, q_1)\), the expected compensating surplus is

\[
E(CS) = \frac{1}{\beta_p} \left[ \ln \left( \sum_j \exp(-\beta_p p_j + \beta_q q_j) \right) - \ln \left( \sum_j \exp(-\beta_p p_0 + \beta_q q_0) \right) \right].
\]

The full choice set is used for computation of WTP estimates.

Figure 5 summarizes the performance of the welfare estimates across different sample sizes using box-and-whisker plots. To construct these plots, mean WTP, 95%, and 75% confidence intervals (CIs) for each unique sample were first calculated. Note that all CIs were constructed using the parametric bootstrapping approach suggested by Krinsky and Robb (1986). The plots contain the mean estimates of these summary statistics across the 200 random samples that were run. As the plots suggest, there is a loss of precision and efficiency with smaller sample sizes. Depending on the welfare scenario, there are modest upward or downward deviations relative to the full sample specification. However, there is no consistent trend across scenarios.

For concreteness, consider the welfare effects of building a boat ramp at every site that does not have one (scenario one). The results from the full choice set model indicate that the average recreational fisherman in the dataset would be willing to pay an additional $0.70 per trip to fund the construction of a boat ramp at the 156 sites without one, with the 95% CI between $0.63 and $0.76 per trip. A researcher could have similarly run one eighth of the sample size
and would expect to find a mean WTP of $0.65 per trip with the 95% CI between $0.56 and $0.73 per trip.

[Figure 6 about here]

The loss in precision from sampling identified in Figure 5 comes with a significant benefit – a reduction in run time. To quantify the tradeoff between precision and run time, Figure 6 shows the change in the range of the 95% CI across sample sizes in comparison to that of the model utilizing the full choice set. The 75% CI range is not reported, but exhibited similar behavior. The ‘Percent Error’ reported is the relative deviation of the sampled mean WTP estimate as compared to the full sample model. It can be interpreted as a measure of how effective the sampled model is at predicting the mean WTP of the full choice set. The ‘Time Savings’ is measured in relation to estimation of the full set of alternatives.

The variation in CI ranges across the five policy scenarios is relatively small, so Figure 6 only shows the mean CI ranges. The results strongly suggest that for samples as small as 71 alternatives, the time savings are substantial while the precision losses are modest. For example, the 286 sample size estimates were generated with a 56% time savings and resulted in 6% larger CIs. Similarly, the 71 alternative sample generated results with a 90% time savings while CIs were 33% larger. By contrast for sample sizes below 71 alternatives, the marginal reductions in run times are small while the loss in precision is substantial. For example, moving from a 71 to a 28 alternative sample reduces run times by less than 10 percent but more than doubles the loss in precision. More strikingly, moving from a 28 to a 6 alternative sample generates a one percent run time savings but increases CI ranges more than threefold. The practitioner much choose their optimal tradeoff between estimation time and parameter efficiency.

6.4 Latent Class Results

A similar evaluation of sampling was conducted with the latent class model. For these models, convergence was achieved at the iteration in the EM algorithm where the parameter values did not change. Since the likelihood function is not globally concave and there is the possibility of convergence on a local minimum, a total of 10 starting values were used for each fixed sample, the largest SLL of which was determined to be the global maximum. The travel cost parameter was fixed across all classes, but the remaining site characteristic parameters were random.

To provide a useful comparison to the conditional logit results presented earlier, an equivalent sample selection process was used. Ten independent samples were taken for each successive sample size, using the same randomization procedure as in the conditional logit model.

For relatively small sample sizes with large numbers of classes, convergence was sometimes elusive. This may be the result of a chosen random sample having insufficient

\[15\] It may be advantageous in some situations to use more starting values to ensure convergence on a global minimum, but due to the computational burden in estimation and the large number of runs conducted, we limited ourselves to just 10 starting values. This may be defensible in our situation because we do have good starting values from our full choice set model where we considered 25 sets of starting values.
variation in the site characteristics data to facilitate convergence. Additional runs were able to eventually produce random samples that were able to converge, however, there may be sample selection concerns with these results. The properties of the random samples that did not converge were not examined and remain an avenue of further study.

Model selection was performed using several information criteria and the NLL statistic. The results were somewhat mixed as to which model was best, but due to overfitting concerns (Hynes et al. 2008), we ultimately put more weight on the rankings implied by the crAIC statistic, which incorporates the greatest penalty for an increased number of parameters.\footnote{Scarpa and Thiene (2005) found similar results and drew similar conclusions.}

[Figure 7 about here]

The latent class model delivers three sets of parameter estimates for each of the indirect utility parameters. The distributions of select parameter estimates (boat ramp, urban, and trout catch rate) are shown in Figure 7. As can be seen, the latent class model can recover preference distributions that do not necessarily resemble commonly used continuous parametric distributions (e.g., normal, uniform).

[Figure 8 about here]

The stability of the WTP estimates across sampling in the mixed model is analyzed in Figure 8 using the results from the optimal model as determined by the crAIC. Using the same policy scenarios as in the conditional logit model, WTP estimates (mean, 95%, and 75% CIs) are constructed for each individual in each class. They are then weighted by the individual latent class probabilities and averaged to produce a single value for each run. In the sampled models, the welfare estimates reported are the average of 10 random samples.

The results indicate that the average Wisconsin fisherman would be willing to pay $0.85 per trip for the proposed agricultural runoff management program (and postulated 5% increase in catch rates), with the 95% CI on average between $0.73 and $0.99.\footnote{The CI represents the interval that is likely to include the parameter estimate.} The researcher could conversely have run a 28 alternative sample and recovered a mean WTP of $0.86 per trip, with the 95% CI being between $0.70 and $1.04. In comparison with the conditional logit model, the latent class WTP estimates are larger in magnitude for scenarios two, four, and five, while smaller for scenarios one and three.

[Figure 9 about here]

Figure 9 shows the change in the range of the mean 95% CIs across sample sizes in comparison to that of the model utilizing the full choice set. The results show that sampling can produce reasonably reliable WTP estimates down to the 28 alternative level.\footnote{This value is 1% for the conditional logit model.} Relative to using the full choice set, the 285 alternative choice set model’s 95% CI is, on average for all considered policy scenarios, 10% wider, and this value becomes 28%, 51%, and 76% for the 142, 71, and 28 alternative samples, respectively.\footnote{These values are 6%, 16%, 33%, and 76%, respectively, for the conditional logit model.} At the 11 and 6 alternative levels, the CIs for some of the samples are extremely large. In this model, WTP estimates at sample sizes under 28 could be considered unreliable. Once again, dependent on the needs of the researcher, an
improvement in computation time is traded off with a lack in precision. Ultimately of course, a researcher’s total run time is conditional on starting values, convergence criteria, and the number of random samples estimated.

7 Conclusion

In this paper we investigated the welfare implications of sampling of alternatives in a mixed logit framework. By employing the EM algorithm, we show how estimation of latent class mixed logit models can be broken down into recursive conditional logit estimation. Within each class, IIA holds and thus allows for sampling of alternatives. We then empirically investigated the performance of the conditional logit and latent class models under various sample sizes in a recreational demand application. Our empirical results suggest that there is modest efficiency loss and significant time savings for the conditional logit models estimated with samples with as few as 71 alternatives. Smaller sample sizes generate relatively modest reductions in run times at the cost of substantial increases in the range of confidence intervals in our application, suggesting that researchers should be cautious when using such small samples. Nevertheless, our results suggest a computationally tractable and empirically useful strategy of estimating flexible mixed logit models with large data sets.
References


Figure 1
Monte Carlo Results
Conditional Logit Model

![Graph showing Fixed Coefficients with varying sample sizes (2000, 1000, 500 observations) for Mean Parameter Bias (MPB) and Proportional Standard Error (PSE).]
Figure 2
Monte Carlo Results
Continuous Distribution Mixed Logit Model

![Graphs showing Monte Carlo results for fixed and random coefficients with different sample sizes.](image-url)
Figure 3
Monte Carlo Results
Latent Class Model Estimated via Maximum Likelihood

Fixed Coefficients
Random Coefficient Means
Latent Class Probability Coefficient Means
Monte Carlo Results
Latent Class Model Estimated via EM Algorithm

Figure 4

Fixed Coefficients

Random Coefficient Means

Latent Class Probability Coefficient Means
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mean</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual Summary Statistics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>trips</td>
<td>day trips during 1998 season</td>
<td>6.994</td>
<td>(7.182)</td>
</tr>
<tr>
<td>boat</td>
<td>dummy = 1 if household owns boat</td>
<td>0.514</td>
<td>-</td>
</tr>
<tr>
<td>kids</td>
<td>dummy = 1 if children under 14 in household</td>
<td>0.414</td>
<td>-</td>
</tr>
<tr>
<td>income</td>
<td>personal income</td>
<td>$28,991</td>
<td>(12,466)</td>
</tr>
<tr>
<td><strong>Site Summary Statistics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tcost</td>
<td>round trip travel time x opp. cost of time +$0.15 x round trip miles</td>
<td>$100.70</td>
<td>(58.28)</td>
</tr>
<tr>
<td>ramp</td>
<td>dummy = 1 if site has at least one paved boat launch</td>
<td>0.726</td>
<td>-</td>
</tr>
<tr>
<td>refuge</td>
<td>dummy = 1 if site is inside a wildlife area or refuge</td>
<td>0.056</td>
<td>-</td>
</tr>
<tr>
<td>forest</td>
<td>dummy = 1 if site is in a national, state, or county forest</td>
<td>0.178</td>
<td>-</td>
</tr>
<tr>
<td>urban</td>
<td>dummy = 1 if urban area on shoreline</td>
<td>0.179</td>
<td>-</td>
</tr>
<tr>
<td>restroom</td>
<td>dummy = 1 if restroom available</td>
<td>0.580</td>
<td>-</td>
</tr>
<tr>
<td>river</td>
<td>dummy = 1 if river fishing location</td>
<td>0.313</td>
<td>-</td>
</tr>
<tr>
<td>small lake</td>
<td>dummy = 1 if inland lake surface area &lt;50 acres</td>
<td>0.172</td>
<td>-</td>
</tr>
<tr>
<td>trout</td>
<td>catch rate for brook, brown, and rainbow trout</td>
<td>0.094</td>
<td>(0.170)</td>
</tr>
<tr>
<td>smallmouth</td>
<td>catch rate for smallmouth bass</td>
<td>0.200</td>
<td>(0.205)</td>
</tr>
<tr>
<td>walleye</td>
<td>catch rate for walleye</td>
<td>0.125</td>
<td>(0.145)</td>
</tr>
<tr>
<td>northern</td>
<td>catch rate for northern pike</td>
<td>0.085</td>
<td>(0.057)</td>
</tr>
<tr>
<td>musky</td>
<td>catch rate for muskellunge</td>
<td>0.010</td>
<td>(0.022)</td>
</tr>
<tr>
<td>salmon</td>
<td>catch rate for coho and chinook salmon</td>
<td>0.009</td>
<td>(0.048)</td>
</tr>
<tr>
<td>panfish</td>
<td>catch rate for yellow perch, bluegill, crappie, and sunfish</td>
<td>1.579</td>
<td>(0.887)</td>
</tr>
</tbody>
</table>
### Table 2
Parameter Estimates: Conditional Logit Model

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Full</th>
<th>285</th>
<th>142</th>
<th>71</th>
<th>28</th>
<th>11</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLL</td>
<td>-13257</td>
<td>-10901</td>
<td>-8640</td>
<td>-6568</td>
<td>-4172</td>
<td>-2314</td>
<td>-1414</td>
</tr>
<tr>
<td></td>
<td>(42)</td>
<td>(56)</td>
<td>(65)</td>
<td>(75)</td>
<td>(73)</td>
<td>(63)</td>
<td></td>
</tr>
<tr>
<td>NLL</td>
<td>-13257</td>
<td>-13264</td>
<td>-13274</td>
<td>-13294</td>
<td>-13344</td>
<td>-13432</td>
<td>-13542</td>
</tr>
<tr>
<td></td>
<td>(4)</td>
<td>(6)</td>
<td>(11)</td>
<td>(27)</td>
<td>(53)</td>
<td>(105)</td>
<td></td>
</tr>
</tbody>
</table>

**Variable**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.412)</td>
<td>(0.407)</td>
<td>(0.4)</td>
<td>(0.393)</td>
<td>(0.39)</td>
<td>(0.401)</td>
<td>(0.434)</td>
</tr>
<tr>
<td>ramp</td>
<td>0.421</td>
<td>0.407</td>
<td>0.397</td>
<td>0.380</td>
<td>0.371</td>
<td>0.349</td>
<td>0.343</td>
</tr>
<tr>
<td></td>
<td>(0.176)</td>
<td>(0.177)</td>
<td>(0.178)</td>
<td>(0.179)</td>
<td>(0.184)</td>
<td>(0.195)</td>
<td>(0.217)</td>
</tr>
<tr>
<td>refuge</td>
<td>0.165</td>
<td>0.169</td>
<td>0.177</td>
<td>0.190</td>
<td>0.202</td>
<td>0.229</td>
<td>0.240</td>
</tr>
<tr>
<td></td>
<td>(0.194)</td>
<td>(0.195)</td>
<td>(0.196)</td>
<td>(0.199)</td>
<td>(0.211)</td>
<td>(0.233)</td>
<td>(0.27)</td>
</tr>
<tr>
<td>forest</td>
<td>0.152</td>
<td>0.142</td>
<td>0.131</td>
<td>0.124</td>
<td>0.128</td>
<td>0.174</td>
<td>0.229</td>
</tr>
<tr>
<td></td>
<td>(0.171)</td>
<td>(0.174)</td>
<td>(0.177)</td>
<td>(0.182)</td>
<td>(0.185)</td>
<td>(0.202)</td>
<td>(0.227)</td>
</tr>
<tr>
<td>urban</td>
<td>-0.068</td>
<td>-0.084</td>
<td>-0.092</td>
<td>-0.094</td>
<td>-0.080</td>
<td>-0.024</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>(0.117)</td>
<td>(0.118)</td>
<td>(0.119)</td>
<td>(0.122)</td>
<td>(0.134)</td>
<td>(0.157)</td>
<td>(0.186)</td>
</tr>
<tr>
<td>restroom</td>
<td>0.149</td>
<td>0.149</td>
<td>0.144</td>
<td>0.143</td>
<td>0.147</td>
<td>0.174</td>
<td>0.211</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(0.131)</td>
<td>(0.131)</td>
<td>(0.13)</td>
<td>(0.134)</td>
<td>(0.148)</td>
<td>(0.168)</td>
</tr>
<tr>
<td>river</td>
<td>-0.013</td>
<td>-0.015</td>
<td>-0.020</td>
<td>-0.032</td>
<td>-0.070</td>
<td>-0.180</td>
<td>-0.266</td>
</tr>
<tr>
<td></td>
<td>(0.297)</td>
<td>(0.299)</td>
<td>(0.305)</td>
<td>(0.319)</td>
<td>(0.348)</td>
<td>(0.399)</td>
<td>(0.458)</td>
</tr>
<tr>
<td>small lake</td>
<td>-0.789</td>
<td>-0.789</td>
<td>-0.776</td>
<td>-0.765</td>
<td>-0.724</td>
<td>-0.707</td>
<td>-0.702</td>
</tr>
<tr>
<td></td>
<td>(0.161)</td>
<td>(0.164)</td>
<td>(0.168)</td>
<td>(0.175)</td>
<td>(0.185)</td>
<td>(0.209)</td>
<td>(0.236)</td>
</tr>
<tr>
<td>trout</td>
<td>1.651</td>
<td>1.674</td>
<td>1.728</td>
<td>1.781</td>
<td>1.975</td>
<td>2.490</td>
<td>2.972</td>
</tr>
<tr>
<td></td>
<td>(0.566)</td>
<td>(0.571)</td>
<td>(0.584)</td>
<td>(0.605)</td>
<td>(0.668)</td>
<td>(0.797)</td>
<td>(0.933)</td>
</tr>
<tr>
<td>smallmouth</td>
<td>0.943</td>
<td>0.948</td>
<td>0.972</td>
<td>0.996</td>
<td>1.050</td>
<td>1.149</td>
<td>1.174</td>
</tr>
<tr>
<td></td>
<td>(0.359)</td>
<td>(0.362)</td>
<td>(0.371)</td>
<td>(0.376)</td>
<td>(0.369)</td>
<td>(0.385)</td>
<td>(0.427)</td>
</tr>
<tr>
<td>walleye</td>
<td>2.690</td>
<td>2.652</td>
<td>2.605</td>
<td>2.540</td>
<td>2.457</td>
<td>2.433</td>
<td>2.364</td>
</tr>
<tr>
<td></td>
<td>(0.379)</td>
<td>(0.376)</td>
<td>(0.374)</td>
<td>(0.38)</td>
<td>(0.405)</td>
<td>(0.479)</td>
<td>(0.561)</td>
</tr>
<tr>
<td>northern</td>
<td>2.659</td>
<td>2.536</td>
<td>2.328</td>
<td>2.013</td>
<td>1.505</td>
<td>0.908</td>
<td>0.658</td>
</tr>
<tr>
<td></td>
<td>(0.935)</td>
<td>(0.955)</td>
<td>(0.998)</td>
<td>(1.07)</td>
<td>(1.215)</td>
<td>(1.472)</td>
<td>(1.746)</td>
</tr>
<tr>
<td></td>
<td>(1.346)</td>
<td>(1.752)</td>
<td>(2.069)</td>
<td>(2.324)</td>
<td>(2.585)</td>
<td>(2.803)</td>
<td>(3.376)</td>
</tr>
<tr>
<td>salmon</td>
<td>7.733</td>
<td>7.852</td>
<td>7.968</td>
<td>8.043</td>
<td>8.139</td>
<td>7.848</td>
<td>7.545</td>
</tr>
<tr>
<td></td>
<td>(1.384)</td>
<td>(1.405)</td>
<td>(1.441)</td>
<td>(1.503)</td>
<td>(1.635)</td>
<td>(1.88)</td>
<td>(2.132)</td>
</tr>
<tr>
<td>panfish</td>
<td>0.763</td>
<td>0.769</td>
<td>0.780</td>
<td>0.789</td>
<td>0.804</td>
<td>0.814</td>
<td>0.814</td>
</tr>
<tr>
<td></td>
<td>(0.189)</td>
<td>(0.189)</td>
<td>(0.191)</td>
<td>(0.194)</td>
<td>(0.201)</td>
<td>(0.223)</td>
<td>(0.253)</td>
</tr>
</tbody>
</table>

* Results for the sampled models represent the mean of 200 random samples; means of the standard errors in parentheses; **bold** indicates significance at the 5% level; “NLL” is the log-likelihood calculated at the parameter values for the entire choice set for comparison purposes.
<table>
<thead>
<tr>
<th>Scenario</th>
<th>Impacted Characteristics</th>
<th>Percentage of Affected Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infrastructure Construction</td>
<td>Build boat ramp at every site that does not have one</td>
<td>27.4%</td>
</tr>
<tr>
<td>Entry Fee Increase</td>
<td>$5 entry fee at all park/forest/refuge sites</td>
<td>23.4%</td>
</tr>
<tr>
<td>Urban Watershed Management</td>
<td>5% catch rate increase for all fish at all urban sites</td>
<td>17.9%</td>
</tr>
<tr>
<td>Agricultural Runoff Management</td>
<td>5% catch rate increase for all fish at all non-urban/forest/refuge sites</td>
<td>30.1%</td>
</tr>
<tr>
<td>Fish Stocking Program</td>
<td>25% increase in Trout catch rate across all sites</td>
<td>98.4%</td>
</tr>
</tbody>
</table>

*569 total sites*
Welfare Results
Conditional Logit Model

Scenario 1: Infrastructure Construction
Build boat ramp at every site that does not have one

Scenario 2: Entry Fee Increase
$5 entry fee at all park/forest/refuge sites

Scenario 3: Urban Watershed Management
5% catch rate increase of all fish at all urban sites

Scenario 4: Agricultural Runoff Management
5% catch rate increase of all fish at all non-urban/forest/refuge sites

Scenario 5: Fish Stocking Program
25% increase in Trout catch rate across all sites

* Mean WTP of 200 unique samples, the mean of the mean, 95th, and 75th CIs of which are reported.

* Method: Small and Rosen (1981); Hanemann (1978) using the parameter estimates from the sample size specified, constructed with the full choice set.

* Outer-product-of-the-gradient method used for calculating SEs.

* The dashed line represents the mean WTP estimate for the full sample model.
Figure 6
Increase in Range of 95% CI of WTP Estimates
Conditional Logit Model

* Compared to the full choice set. Efficiency Loss is the percent increase in the average range of the 95% CI. Percent Error is the percentage deviation of the mean WTP from the full sample. Time Savings is relative to the full sample model.
Figure 7
Parameter Estimate Distributions

* Select parameters. Results from the three-class model with the full choice set; best of ten starting values. Class share is the mean of individual class shares calculated using the individual specific parameters. The dashed line represents the parameter expected value.
Figure 8
Welfare Results
Latent Class Model (crAIC)

Scenario 1: Infrastructure Construction
Build boat ramp at every site that does not have one

Scenario 2: Entry Fee Increase
$5 entry fee at all park/forest/refuge sites

Scenario 3: Urban Watershed Management
5% catch rate increase of all fish at all urban sites

Scenario 4: Agricultural Runoff Management
5% catch rate increase of all fish at all non-urban/forest/refuge sites

Scenario 5: Fish Stocking Program
25% increase in Trout catch rate across all sites

* Mean WTP of ten unique samples, the mean of the mean, 95th, and 75th CIs of which are reported.

* Method: Small and Rosen (1981); Hanemann (1978) using the parameter estimates from the sample size specified, constructed with the full choice set.

* The dashed line represents the mean WTP estimate for the full sample model.
Figure 9
Increase in Range of 95% CI of WTP Estimates
Latent Class Model

* Compared to the full choice set. Efficiency Loss is the percent increase in the average range of the 95% CI. Percent Error is the percentage deviation of the mean WTP from the full sample. Time Savings is relative to the full sample model.