Estimating Mixed Logit Models with Large Choice Sets

Roger H. von Haefen and Adam Domanski

January 2013

Abstract
We develop an expectation-maximization (EM) algorithm to 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

1 We thank Ray Palmquist, Laura Taylor, Xiayong Zhang, Chris Parmeter and seminar participants at Oregon State, Appalachian State, Maryland, East Carolina, the University of Miami, and Ohio State for helpful comments. All remaining errors are our own. The views expressed in this paper are those of the authors and do not necessarily reflect those of the National Oceanic and Atmospheric Administration.

2 Department of Agricultural and Resource Economics, North Carolina State University, NCSU Box 8109, Raleigh, NC 27695-8109; and National Bureau of Economic Research, Cambridge, MA 02138; rhhaefen@ncsu.edu.

3 NOAA/IM Systems Group, 1305 East West Highway, SSMC4 Station 10356, Silver Spring, MD 20910; adam.domanski@noaa.gov.
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 sampling and misspecification 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.

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

---

4 Consumer types can be driven by any combination of attitudinal, spatial, demographic, or other variation in the population.
full choice set to estimates generated with samples 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

---

5 Throughout this paper, “sample size” will refer to the sample of alternatives, not the sample of observations.
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 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 (see Train (2009) for a detailed discussion).

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 determinants of choice such as travel cost, site characteristics, and demographic/site characteristic interactions that the econometrician observes as well as preference parameters. In most empirical applications, a linear functional form is assumed, i.e., \( V_{ni} = \beta_n'x_{ni} \) where \( x_{ni} \) are observable determinants of choice and \( \beta_n \) are preference parameters that may vary across individuals. \( \varepsilon_{ni} \) 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 (indexed by \( j \)). This structure implies the individual’s decision rule can be succinctly stated as: 
alternative i 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 the case of a recreational site closure due to an acute environmental incident. Assume the closed site has unusually high catch rates for trout. Intuitively, individuals who previously chose the closed site likely have a strong preference for catching trout and would thus substitute to other sites with relatively high trout 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 trout catch rates.

To generate more realistic substitution patterns, environmental economists have frequently used nested logit models where those sites with common features (e.g., high trout catch rates) are grouped into common nests that exhibit greater substitution effects.
(Ben-Akiva 1973; Train et al. 1987; Parsons and Kealy 1992; Jones and Lupi 1997; Parsons and Hauber 1998; Shaw and Ozog 1999; and Parsons et al. 2000). With these models, the angler’s decision can be represented as a sequence of choices. For example, an angler’s choice could be modeled as an initial decision of lake or river fishing, a conditional choice of target species, and a final choice of recreation site. Although the nested logit assumes that within each nest the IIA assumption holds, it relaxes IIA across different nests.

Despite its ability to relax IIA, the nested logit suffers from significant shortcomings. One shortcoming is the sensitivity of parameter and welfare estimates to different nesting structures (Kling and Thomson 1996). Another arises because all unobserved heterogeneity enters preferences through an additive error term, a characteristic shared by the conditional logit model. Although observed preference heterogeneity can be introduced by interacting observable demographic data with site attributes, the fact that unobserved heterogeneity enters preferences additively limits the analyst’s ability to allow for general substitution patterns across multiple dimensions (e.g., catch rates, boat ramps, and water quality). This point seems especially relevant for situations where preferences for attributes are diverse or polarized.

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). 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 $\beta_n$, $f(\beta_n | \theta)$, where $\theta$ is a vector of parameters. Introducing preference heterogeneity in this way results in correlation in the unobservables for sites with similar attributes and thus relaxes IIA. Conditional on $\beta_n$, the probability of selecting alternative $i$ in the mixed logit is:
\[ P_{ni}(\beta_n) = \frac{\exp(\beta_n x_{ni})}{\sum_j \exp(\beta_n x_{nj})}. \]

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

\[ P_{ni} = \int P_{ni}(\beta_n) f(\beta_n | \theta) d\beta_n. \]

When the dimension of \( \beta_n \) is moderate to large, analytical or numerical solutions for the above integral are generally not possible; however, \( P_{ni} \) can be approximated via simulation (Boersch-Supan and Hajivassiliou 1990; Geweke et al. 1994; McFadden and Ruud 1994). This involves generating several pseudo-random draws from \( f(\beta_n | \theta) \), calculating \( P_{ni}(\beta_n) \) for each draw, and then averaging across draws. By the law of large numbers, this simulated estimate of \( P_{ni} \) 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 environmental economists to specify discrete or step function distributions that can readily account for
different features of the data. The unconditional probability is the sum of logit kernels weighted by class membership probabilities:

$$P_n = \sum_{c} S_{nc}(z_n, \delta) P_n(\beta_c).$$

where $S_{nc}(z_n, \delta)$ 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; Landry and Liu 2009). 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:

$$S_{nc}(z_n, \delta) = \frac{\exp(\delta_z z_n)}{\sum_{c} \exp(\delta_z z_n)}.$$

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

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, Feather 2003) that raise computational issues in estimation.6

6 Here we are abstracting from the related issue of consideration set formation (Manski 1977; Horowitz 1991), 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
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 the analyst to make additional assumptions about preferences or price and quality movements within the set of alternatives.

Aggregation methods assume that alternatives can be grouped into composite choice options. In the recreational demand context, similar recreation sites are combined into a representative site, and in housing, homes within a subdivision are aggregated into a typical home. McFadden (1978) has shown that this approach generates consistent estimates if the utility variance and composite size within aggregates is controlled for. Although the composite size is commonly observed or easily proxied in recreation demand applications, the utility variance depends on unknown parameters and is thus difficult to proxy by the analyst. Kaoru and Smith (1990), Parsons and Needleman (1992) and Feather (1994) empirically investigate the bias arising from ignoring the utility variance with a recreation data set. In some instances, these authors find large differences between disaggregated and aggregated models, although the direction of bias is unclear.7

Separability assumptions allow the researcher to selectively remove alternatives based on a variety of criteria. For example, it is common in recreation demand analysis to focus on just boating, fishing, or swimming behavior. In these cases, sites that do not support a particular recreation activity are often eliminated. Likewise, recreation studies increased interest in recent environmental applications despite their significant computational hurdles (Haab and Hicks 1997; Parsons et al. 2000; 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 Similarly, Lupi and Feather (1998) consider a partial site aggregation method where the most popular sites and those most important for policy analysis will enter as individual site alternatives, while the remaining sites are aggregated into groups of similar sites. Their empirical results suggest partial site aggregation can reduce but not eliminate aggregation bias. Haener et al. (2004) utilize precise site data to test the welfare effects of spatial resolution in site-aggregation recreation demand models. They introduce an aggregation size correction in a nested structure to account for part of the aggregation bias. However their disaggregate model is not nested making full comparison difficult.
frequently focus on day trips, which imply a geographic boundary to sites that can be accessed by day. Empirical evidence by Parsons and Hauber (1998) on the spatial boundaries for choice set definition suggests that after some threshold distance, adding more alternatives has a negligible effect on estimation results. Nevertheless, even if the separability assumptions that motivate shrinking the choice set are valid, the remaining choice set can still be intractably large, particularly when sites are defined in a disaggregate manor.

The third common solution to large choice set problems is to employ a sample of alternatives the decision maker faces in estimation. As McFadden (1978) has shown, combining a random sample of alternatives within traditional maximum likelihood estimation techniques simplifies the computational burden and produces consistent estimates as long as the uniform conditioning property\(^8\) 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 the applied literature with fixed parameter logit and nested logit models (Parsons and Kealy 1992; Sermons and Koppelman 2001; Waddell 1996; 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

\(^8\) 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.
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. 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.

3 Sampling in a Mixture Model

This section describes a sampling of alternatives approach to estimating discrete choice models that exploits a variation of the expectation-maximization (EM) algorithm. We describe the estimator as well as two practical issues associated with its implementation – model selection and computation of standard errors.

3.1 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 1997; 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 with 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 \( \phi^t = (\beta^t, \delta^t) \) and log-likelihood function:

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

Bayes’ rule is used to construct the conditional probability that individual \( n \) is a member of class \( c \):

\[
h_{nc}(\phi^t) = \frac{S_{nc}(\delta^t)L_n(\beta^t)}{\sum_{c=1}^{C} S_{nc}(\delta^t)L_n(\beta^t)}, \ c = 1, ..., C.
\]

These probabilities serve as weights in the construction of the expected log-likelihood which is then maximized to generate updated parameter values:

\[
\phi^{t+1} = \arg \max_{\phi} \sum_{n=1}^{N} \sum_{c=1}^{C} h_{nc}(\phi^t) \ln \left( S_{nc}(\delta)L_n(\beta^c) \right)
\]

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:

\[
\delta^{t+1} = \arg \max_{\delta} \sum_{n=1}^{N} \sum_{c=1}^{C} h_{nc}(\phi^t) \ln S_{nc}(\delta),
\]

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

\[
\beta^{t+1} = \arg \max_{\beta} \sum_{n=1}^{N} h_{nc}(\phi^t) \ln L_n(\beta^c).
\]

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 best model fit.

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\(^9\). 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.

3.2 Model Selection

A practical issue with latent class models is the selection of the number of classes. Traditional specification tests (likelihood ratio, Lagrange multiplier, and Wald tests) are inappropriate in this context because increasing the number of classes also increases the number of variables to be estimated. These tests ignore the potential of overfitting the model. A variety of information criteria statistics have been used throughout the latent class literature. In general form (Hurvich and Tsai 1989), the information criteria statistic is specified as \( IC = -2*LL + A*\gamma \) where \( LL \) is the log likelihood of the model at convergence, \( A \) is the number of estimated parameters in the model, and \( \gamma \) is a penalty constant. There are a number of different information criteria statistics that differ in terms of the penalty associated with adding parameters, represented by the penalty constant \( \gamma \).

\(^{\text{9}}\) 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.
[Table 1 – Information Criteria Statistics]

In each case, the optimal model is that 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 performs better under weaker regularity conditions than the likelihood ratio test (Roeder et al. 1999). Alternatively, many past papers (Meijer and Rouwendal 2006; Desarbo et al. 1992; Morey et al. 2006) have used the Akaike Information Criteria (AIC) (Akaike 1974). Other papers have compared the various information criteria (Thacher et al. 2005; Scarpa and Thiene 2005), but there is no general consensus in the literature for using one test over the others.

As observed previously in Hynes et al. (2008), it is possible in practice for the analyst to select overfitted models when using only the AIC or BIC. For example, specified models with many parameters can generate parameter estimates and standard errors that are unstable or implausibly large. In our subsequent empirical exercise, parameter estimates for specific classes diverged wildly from estimates for other classes while at the same time being coupled with very small latent class probabilities. Results like these suggest overfitting and the need for a more parsimonious specification. Since the CAIC or crAIC penalize the addition of parameters more severely, they may be more useful to applied researchers if evidence of overfitting arises.

### 3.3 Standard Errors

Calculation of the standard errors of parameter estimates from the EM algorithm can be cumbersome since there is no direct method for evaluating the information matrix (Train 2008). There is a large statistical literature addressing various methods of calculating standard errors based upon the observed information matrix, the expected information matrix, or resampling methods (Baker 1992, Jamshidian and Jennrich 2002, Meng and Rubin 1991). Train (2008) provides additional discussion on the different strategies available for calculating standard errors.
An aspect of the EM algorithm that can be exploited is the fact that the score of the log-likelihood is solved at each maximization step. Ruud (1991) uses this observed log-likelihood at the final step of the iteration to compute estimates of the standard errors. Derived by Louis (1982), the information matrix can be approximated with the outer-product-of-the-gradient formula:

$$
\hat{I} = N^{-1} \sum_{n=1}^{N} g(\hat{\phi})g(\hat{\phi})'
$$

where $g$ is the score vector generated from the final step of the EM algorithm. This estimation of the information matrix is a common method for recovering standard errors and is the simplest method for doing so with the EM algorithm.\(^{10}\)

4 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 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,\(^{11}\) they strongly suggest that this is indeed the case.

\(^{10}\) A limitation with this approach to estimating standard errors is that it does not estimate the empirical Hessian directly and thus cannot be used to construct robust standard errors.

\(^{11}\) 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.
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.\(^{12}\) 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.\(^{13}\) 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.

[Figure 1 – Conditional Logit Model – Monte Carlo Results]

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

\(^{12}\) 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.

\(^{13}\) Stated differently, the fixed parameters are equal across the two classes whereas the random parameters are class-specific.
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 samples sizes, we generally find this to be the case across all sample of alternatives sizes. The only exception is for 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.

[Figure 2 – Continuous Distribution Mixed Logit Model – Monte Carlo Results]

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 – Latent Class Model Estimated via Maximum Likelihood – Monte Carlo Results]

---

14 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.
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 model 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. This finding suggests that our proposed estimator is consistent.

The Monte Carlo simulations that support Figures 1-4 can also be used to address a practical issue that arises in applications involving exceptionally large (e.g., scanner) data sets. In these instances the researcher has to choose how best to allocate a fixed amount of memory between more observations or more sampled alternatives. Ben-Akiva

---

15 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.
and Lerman (1985) speculated that including more observations and less sampled alternatives will lead to more precise estimates, and Banzhaf and Smith (2004) present Monte Carlo evidence confirming their intuition in the conditional logit context. Figure 5 provides further evidence in the latent class context where estimation proceeds with our proposed EM algorithm. We consider three alternative data configurations: 1) 2000 observations and 25 sampled alternatives; 2) 1000 observations and 50 sampled alternatives; and 3) 500 observations and 100 sampled alternatives. The results reported in Figure 5 strongly suggest that more observations and fewer alternatives lead to more precise estimates and no sacrifice in bias.

5 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.

5.1 Data

An empirical illustration is performed with data from the Wisconsin Fishing and Outdoor Recreation Survey. Conducted 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 2 presents summary statistics.

[Table 2 – Summary Statistics]
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 \(6^{16}\) 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 to recover standard errors in the latent class model, we will use the same method with the conditional logit model. Although not the focus of the paper here, we also performed a preliminary estimation of a sampled continuous distribution mixed logit model with a normal distribution and 30 replications, however the results indicated a high degree of parameter variability across sample sizes.

5.2 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. 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 3 –Parameter Estimates: Conditional Logit Model]

The parameter estimates and standard errors for each of the sample sizes are shown in Table 3. 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-

\(^{16}\) 50\%, 25\%, 12.5\%, 5\%, 2\%, and 1\% sample sizes, respectively.
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 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 3 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 3. The following policy scenarios are considered (see Table 4): 1) infrastructure construction,17 2) an increase in entry fees,18 3) an urban watershed management program, 4) an agricultural runoff management program,19 and 5) a fish stocking program.20 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.

17 Supposing that a boat ramp was constructed at each Wisconsin lake that did not have one (27% of sites).
18 $5 increase in entry fees at all state-managed sites (defined by being in a state forest or wildlife refuge); approximately 23% of sites.
19 Supposing 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.
20 Fish stocking program where the catch rate of trout is increased by 25% across all sites that currently contain trout.
The methodology used to calculate WTP is the log-sum formula derived by Hanemann (1978) and Small and Rosen (1981). 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 compensating surplus is

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

and for our iid type 1 extreme value case, the expected consumer surplus has a closed form

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

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

Figure 6 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)s 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 7 – Increase in Range of 95% CI of WTP Estimates Compared to Full Choice Set: Conditional Logit Model]

The loss in precision from sampling identified in Figure 6 comes with a significant benefit – a reduction in run time. To quantify the tradeoff between precision and run time, Figure 7 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 7 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.

5.3 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 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.

[Table 5 – Information Criteria]

Model selection is performed using the information criteria described earlier. Using the NLL, the various decision rules produce different results. The CAIC indicates that five or more classes were optimal with the full choice set. However, a careful inspection of the parameter estimates suggests possible over-fitting of the model. With five classes, the EM algorithm is attempting to estimate 87 parameters from 513 observations of individual choices. With a large number of classes, the parameters for one class for certain variables diverge dramatically. For example, one class in the five-class model has parameters of -35, -86, and 29 for the forest, trout, and musky variables,

---

21 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.

22 The AIC and BIC indicated the same number of optimal classes as the CAIC.
respectively. By contrast, the mean parameter estimates across the other four classes are 0.71, 0.50, and 4.70, respectively. This empirical finding may be the result of the five-class model attempting to use a single class to account for a handful of anomalous observations. Thus, in our view, the more appropriate decision criterion is the crAIC, which incorporates the greatest penalty for an increased number of parameters.23

[Figure 8 – Distribution of Parameter Estimates]

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 8. 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 9 – Welfare Results: Latent Class Model]

The stability of the WTP estimates across sampling in the mixed model is analyzed in Figure 9 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 $0.99.24 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.

23 Scarpa and Thiene (2005) found similar results and drew similar conclusions.
24 The CI represents the interval that is likely to include the parameter estimate.
Figure 10 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.\textsuperscript{25} 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.\textsuperscript{26} 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.

6 Conclusion

This paper has investigated the welfare implications of sampling of alternatives in a mixed logit framework. By employing the EM algorithm, estimation of latent class mixed logit models 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.

We have empirically investigated the performance of the conditional logit and latent class models under various sample sizes in a recreational demand application. Our 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 small reductions in run times at the cost of substantial increases in

\textsuperscript{25} This value is 1% for the conditional logit model.

\textsuperscript{26} These values are 6%, 16%, 33%, and 76%, respectively, for the conditional logit model.
the range of confidence intervals. Depending on the needs of the researcher, however, results may be useful down to 6 alternatives.

For the latent class models, the results reported in this paper suggest that sampling of alternatives significantly reduces run times and performs very well for samples of as few as 142 alternatives. Although sample sizes may be useful down to the 28 alternative level, they come with relatively small time savings and substantial losses in precision. Estimates from sample sizes below 28 alternatives were found to be unreliable in our application.

Certain lessons for the practitioner should be noted. As with any mixture model, the latent class model may be sensitive to starting values. We considered 10 sets of starting values for each specification due to computational limitations, but a greater number of starting values would be preferred to ensure estimation of the global minima. More starting values will increase total computation time so the researcher will have to exercise judgment in this regard. Additionally, although the consistency and efficiency of estimates at small samples will depend on the data set, our results suggest that extremely small samples (below 28 alternatives) should be avoided. See other work by von Haefen (unpublished) for a further discussion and Monte Carlo analysis.

At the current state of research, this paper has demonstrated the practicality of 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.
References


Figure 1
Monte Carlo Results
Conditional Logit Model
Figure 2
Monte Carlo Results
Continuous Distribution Mixed Logit Model

- Fixed Coefficients
- Random Coefficient Means
- Random Coefficient Standard Deviations
Figure 3
Monte Carlo Results
Latent Class Model Estimated via Maximum Likelihood
Figure 4
Monte Carlo Results
Latent Class Model Estimated via EM Algorithm

Fixed Coefficients

Random Coefficient Means

Latent Class Probability Coefficient Means
Figure 5
Monte Carlo Results
Latent Class Model Estimated via EM Algorithm

![Diagram showing Observations versus Alternatives.](image)

Table 1
Information Criteria Statistics

<table>
<thead>
<tr>
<th>Information Criteria</th>
<th>Penalty Constant $\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akaike Information Criteria</td>
<td>2</td>
</tr>
<tr>
<td>Bayesian Information Criteria</td>
<td>$\ln(N)$</td>
</tr>
<tr>
<td>Consistent Akaike Information Criteria</td>
<td>$1 + \ln(N)$</td>
</tr>
<tr>
<td>Corrected Akaike Information Criteria</td>
<td>$2 + 2(A + 1)(A + 2)/(N - A - 2)$</td>
</tr>
</tbody>
</table>

* General formula: $IC = -2*LL + A*\gamma$, where $LL$: log likelihood, $A$: # of parameters, $\gamma$: penalty constant
### Table 2
Summary Statistics

<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 ramp</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>
<tr>
<td>Sample Size</td>
<td>Full</td>
<td>285</td>
<td>142</td>
</tr>
<tr>
<td>-------------</td>
<td>-------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>SLL</td>
<td>-13257</td>
<td>-10901</td>
<td>-8640</td>
</tr>
<tr>
<td></td>
<td>(42)</td>
<td>(56)</td>
<td>(65)</td>
</tr>
<tr>
<td>NLL</td>
<td>-13257</td>
<td>-13264</td>
<td>-13274</td>
</tr>
<tr>
<td></td>
<td>(4)</td>
<td>(6)</td>
<td>(11)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimates: Conditional Logit Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>tcost</td>
<td>-10.07</td>
</tr>
<tr>
<td></td>
<td>(0.412)</td>
</tr>
<tr>
<td>ramp</td>
<td>0.421</td>
</tr>
<tr>
<td></td>
<td>(0.176)</td>
</tr>
<tr>
<td>refuge</td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td>(0.194)</td>
</tr>
<tr>
<td>forest</td>
<td>0.152</td>
</tr>
<tr>
<td></td>
<td>(0.171)</td>
</tr>
<tr>
<td>urban</td>
<td>-0.068</td>
</tr>
<tr>
<td></td>
<td>(0.117)</td>
</tr>
<tr>
<td>restroom</td>
<td>0.149</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
</tr>
<tr>
<td>river</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td>(0.297)</td>
</tr>
<tr>
<td>small lake</td>
<td>-0.789</td>
</tr>
<tr>
<td></td>
<td>(0.161)</td>
</tr>
<tr>
<td>trout</td>
<td>1.651</td>
</tr>
<tr>
<td></td>
<td>(0.566)</td>
</tr>
<tr>
<td>smallmouth</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td>(0.359)</td>
</tr>
<tr>
<td>walleye</td>
<td>2.690</td>
</tr>
<tr>
<td></td>
<td>(0.379)</td>
</tr>
<tr>
<td>northern</td>
<td>2.659</td>
</tr>
<tr>
<td></td>
<td>(0.935)</td>
</tr>
<tr>
<td>musky</td>
<td>5.361</td>
</tr>
<tr>
<td></td>
<td>(1.346)</td>
</tr>
<tr>
<td>salmon</td>
<td>7.733</td>
</tr>
<tr>
<td></td>
<td>(1.876)</td>
</tr>
<tr>
<td>panfish</td>
<td>0.763</td>
</tr>
<tr>
<td></td>
<td>(0.189)</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*
Figure 6
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 7
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.
Table 5
Information Criteria

<table>
<thead>
<tr>
<th># of classes</th>
<th>Sample Size</th>
<th>NLL</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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NLL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-13257</td>
<td>-13264</td>
<td>-13274</td>
<td>-13294</td>
<td>-13344</td>
<td>-13432</td>
<td>-13542</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-12416</td>
<td>-12425</td>
<td>-12447</td>
<td>-12471</td>
<td>-12586</td>
<td>-12905</td>
<td>-13264</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-12025</td>
<td>-12042</td>
<td>-12066</td>
<td>-12084</td>
<td>-12220</td>
<td>-12526</td>
<td>-12872</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-11724</td>
<td>-11734</td>
<td>-11757</td>
<td>-11805</td>
<td>-12010</td>
<td>-12311</td>
<td>DNC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-11432</td>
<td>-11486</td>
<td>-11532</td>
<td>-11592</td>
<td>-11748</td>
<td>-12232</td>
<td>DNC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAIC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>26623</td>
<td>26637</td>
<td>26657</td>
<td>26697</td>
<td>26797</td>
<td>26973</td>
<td>27193</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>25071</td>
<td>25089</td>
<td>25132</td>
<td>25180</td>
<td>25411</td>
<td>26049</td>
<td>26767</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>24420</td>
<td>24453</td>
<td>24500</td>
<td>24538</td>
<td>24809</td>
<td>25422</td>
<td>26113</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>23947</td>
<td>23967</td>
<td>24013</td>
<td>24110</td>
<td>24518</td>
<td>25122</td>
<td>DNC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>23494</td>
<td>23601</td>
<td>23693</td>
<td>23813</td>
<td>24127</td>
<td>25093</td>
<td>DNC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>crAIC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>26560</td>
<td>26574</td>
<td>26594</td>
<td>26634</td>
<td>26734</td>
<td>26910</td>
<td>27130</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>25063</td>
<td>25081</td>
<td>25124</td>
<td>25172</td>
<td>25403</td>
<td>26040</td>
<td>26759</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>24765</td>
<td>24798</td>
<td>24846</td>
<td>24883</td>
<td>25154</td>
<td>25767</td>
<td>26458</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>25141</td>
<td>25161</td>
<td>25207</td>
<td>25304</td>
<td>25712</td>
<td>26316</td>
<td>DNC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>26260</td>
<td>26367</td>
<td>26459</td>
<td>26579</td>
<td>26892</td>
<td>27859</td>
<td>DNC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* CAIC, and crAIC calculated using the LL calculated with the full choice set; Mean of 200 and ten random samples reported for the one and multiple class models respectively; Optimal # of classes outlined and in **bold** defined by the minimum of the information criteria; **DNC** = model did not converge.
* 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 9
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 10
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.