Estimation and Welfare Analysis from Mixed Logit Recreation Demand Models with Large Choice Sets

Adam Domanski

Roger H. von Haefen

North Carolina State University

April 2010

We thank Ray Palmquist, Laura Taylor, and Xiayong Zhang for helpful comments. All remaining errors are our own.

1. Introduction

Environmental economists frequently use discrete choice models to analyze recreation site selection decisions. These models have proven valuable when modeling scenarios where agents make extensive margin decisions from large choice sets and substitution is important. However, when the individual'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 a conditional logit framework can obviate computational difficulties and produce consistent estimates. His approach has been widely used throughout the literature (e.g., 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, identically distributed type I extreme value. The assumption implies that the odds ratio for any two alternatives does not change with the addition of a third alternative. This property is known as the independence of irrelevant alternatives, or IIA. Although necessary for consistent estimation under sampling of alternatives, it is well known that IIA is often a restrictive and inaccurate characterization of choice.

In recent years, applied researchers have developed several innovative models that exploit recent computational advances to relax IIA. Perhaps the most notable and widely used is the mixed logit model. 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 a powerful and attractive tool for discrete choice modeling. However, adopting a mixed logit approach comes with a significant cost – there is no proof that the sampling of alternatives approach within the mixed logit framework will generate 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 asymptotic bias and more flexible substitution patterns with mixed logit.[1]

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. In practice the normal distribution is often employed, 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 remains a concern (see Hess and Rose, 2006).

Both of these 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 approach 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. 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 a more feasible 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 produce consistent parameter estimates (von Haefen and Jacobsen, unpublished). The implications of this estimation strategy for welfare analysis have not been investigated in the recreation demand context previously.

In this paper, we empirically evaluate the welfare implications of this novel estimation strategy for large choice set problems using a recreational dataset of Wisconsin anglers. The Wisconsin dataset is attractive for this purpose because it includes a large

---

[1] See McConnell and Tseng (200) and Nerella and Bhat (2004) for a discussion and empirical evaluation of sampling with mixed logit.

number of recreational destination alternatives (569 lakes in total) that raises difficult but not insurmountable computational challenges for estimation with the full choice set. By comparing estimates generated with the full choice set to estimates generated with samples of alternatives of different sizes, we can gauge the benefits and costs of sampling of alternatives in terms of reduced estimation run time and efficiency loss. Our strategy for quantifying these benefits and costs involves repeatedly running latent class models on random samples of alternatives of different sizes (in particular, 50%, 25%, 12.5%, 5%, 2%, and 1% sample sizes). Our results suggest that for our preferred latent class specification, using a sample of alternatives that is 12.5% of the full choice set 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 5% of the full set of alternatives may be sufficiently informative for policy purposes, and that smaller sample sizes often generate point estimates with very large confidence intervals.

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 our empirical application with the Wisconsin angler dataset. Section six 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 (2003) 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 ε_{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 β_n are preference parameters that may vary across individuals. ε_{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:

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

2.1 Conditional Logit

Different distributional specifications for β_n and ε_{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 ε_{ni} is an independent and identically distributed (iid) draw from the type I extreme value distribution with scale parameter μ. The probability that individual n prefers 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 β and μ are not separately identified and thus, with no loss in generality, μ can be normalized to one.

The conditional logit model embodies the IIA property (which means that the odds ratio for any two alternatives is unaffected by the inclusion of any third alternative). To see this, consider the ratio of probabilities for sites i and k:

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

Because P_{ni} and P_{nk} share the same denominator, the observable attributes for all other alternatives drops out, and thus the odds ratio will not change with the addition of any third alternative. This property of the conditional logit model is a direct result of the independent type I extreme value assumption.

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, it seems plausible that the individuals who previously chose the closed site have a strong preference for catching trout and would thus resort to visiting 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 along 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 β_n, $f(\beta_n \mid \theta)$, where θ 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 β_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 β_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:

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

When the dimension of β_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 $f(\beta_n \mid \theta)$, 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 about a factor of three. In addition, although the mean welfare estimates were similar across the three tested 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:

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

where $S_{nc}\left(\delta, z_n\right)$ is the probability of being in class c (c = 1,…,C) and δ and z_n are parameters and observable demographics 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 in environmental economics, however, the class membership probabilities depend on observable demographics which parsimoniously introduces additional preference heterogeneity. In these cases, the class probabilities typically assume a logit structure:

$$S_{nc}\left(\delta, z_n\right) = \frac{\exp\left(\delta_c z_n\right)}{\sum_{l=1}^{C} \exp\left(\delta_l z_n\right)}.$$

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

3. Large Choice Sets

The specification of the choice set is vital to the effective implementation of any discrete choice model. Choice set definition deals with specifying the objects of choice that enter an individual's preference ordering. In practice, defining an individual's choice set is influenced by the limitations of available data, the nature of the policy questions addressed, the analyst's judgment, and economic theory (von Haefen 2008). In recreation demand applications, 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.[2] This section reviews commonly used strategies for

[2] 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 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.

addressing the computational issues raised by large choice set specifications with a detailed treatment of the sampling of alternatives method.

3.1 Alternative Strategies

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 make the assumption that alternatives can be grouped into representative choice options. For a recreational demand context, similar recreation sites can be treated as one; in housing, a group of homes in a given sub-development can be aggregated. This methodology can be effective but is problematic in that the success of estimation is entirely dependent on the assumptions made in the aggregation. McFadden (1978) and Ben-Akiva and Lerman (1985) have both shown that this technique can produce biased estimates if the utility variance and composite size within aggregates is not accounted for. Although the composite size is commonly observed or easily proxied in recreation demand applications, the utility variance depends on unknown parameters and is therefore unknown and difficult to proxy by the analyst. Kaoru amd 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 cases, they find large differences between disaggregated and aggregated models, but their results suggest no clear direction of bias from aggregation. 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.

Separability assumptions allow the researcher to selectively remove alternatives based on whether recreation sites support the type of recreation considered. 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 frequently focus on day trips, which implies a geographic boundary to sites that can be accessed with a day trip. 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.

3.2 Sampling of Alternatives

The third common solution is to use a sample of alternatives the decision maker faces in estimation. By selecting a random subset of the relevant alternatives, the researcher is left with a much more computationally tractable choice set. McFadden (1978) proved the approach will produce consistent estimates as long as the resulting choice probability ratios do not change due to the elimination of choice alternatives. This is feasible within the standard logit model due to the IIA assumption. Sampling has been successfully utilized and demonstrated in the literature (Parsons and Kealy 1992; Sermons and Koppelman 2001; Waddell 1996; Bhat et al. 1998; Guo and Bhat 2001; Ben-Akiva and Bowman 1998, von Haefen and Jacobsen, unpublished).

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

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

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

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

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

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

McConnell and Tseng (2000) perform an empirical analysis on beach use and recreational fishing to evaluate sampling of alternatives. Since there is no theoretical foundation for sampling of alternatives in a continuous distribution mixture model, they seek to broaden the understanding of the mixed logit model through empirical evidence. Their results with a recreation data set consisting of a relatively small choice set suggest that sampling in a continuous distribution mixture model does not alter the results significantly or systematically, although their results should be interpreted cautiously given the same number of sampling replications they consider. Nerella and Bhat (2004) perform a similar analysis with simulated data. They analyze the effect of sample size on the empirical accuracy and efficiency of multinomial and mixed multinomial models. Their results suggest that analysts can, in practice, use a 12.5% sample in the conditional logit model and down to a 25% sample in the mixed logit model.

3.3 Sampling in a Mixture Model

Utilizing the latent class model estimated via the recursive expectation-maximization (EM) algorithm, sampling of alternatives can generate theoretically consistent estimates (von Haefen and Jacobsen, unpublished). This section describes use of the EM algorithm to estimate a latent class model and addresses the issues of model selection, and computation of standard errors.

3.4 EM Algorithm

The EM algorithm (Dempster et al. 1977) is an alternative estimation method for recovering parameter estimates from likelihood-based models where maximum likelihood estimation is computationally difficult. The EM algorithm has become a popular tool with estimation problems involving incomplete data (McLachlan and Krishnan 1997) as well as mixture estimation (Bhat 1997; Train 2008). The method also facilitates the consistent sampling of alternatives as shown by von Haefen and Jacobsen (unpublished).

Assuming that an unknown parameter (in this case the latent class membership) is represented as a value in some parameterized probability distribution, the EM algorithm is a recursive procedure which begins with the expectation or "E" step: specifying the expected value of unknown parameters (class membership) given some known parameters (our current guess of the model parameters). The maximization or "M" step follows: the known parameters are then updated given the expected values of the unknown parameters. The steps are then repeated until convergence, defined as a pre-determined small change in the parameter estimates between iterations (Train 2008). This methodology represents an improvement over gradient-based methods by its ability to transform the computationally difficult maximization of a log of sums into a simpler recursive maximization of the sum of logs.

Given our log-likelihood function:

$$LL_n = \ln\left(\sum_c S_{nc}(\delta_c) L_n(\beta_c^t)\right)$$

and some fixed set of starting values for the parameters $\overline{\phi}^{t}=(\beta_c^t,\delta^t)$, the EM algorithm lets us iteratively calculate a new value for the parameters:

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

where t represents the iteration number and N is the number of observations. Since the right hand portion of the equation can be rewritten as

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

the maximization can be performed independently for each set of parameters. Beginning with the "E" step at iteration t, using various starting values, the probability (weight) of individual n belonging to class c conditional on the parameters β_c^t and δ^t is:

$$h_{nc}(\phi^t)=\frac{S_{nc}\left(\delta^t\right)L_n(\beta_c^t)}{\sum_{l=1}^{C}S_{nl}\left(\delta^t\right)L_n(\beta_l^t)}$$

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

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

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

$$\beta_c^{t+1}=\arg\max_{\beta}\sum_{n=1}^{N}h_{nc}(\phi^t)\ln L_n(\beta_c)$$

The weights are then recalculated using the new parameter values, and the entire process is repeated until convergence.[3] Each successive maximization uses the prior parameters ϕ^t and individual-specific class probabilities to form the weights used in the maximization the new parameter values. The previously computationally burdensome estimation has now been

[3] Note that when sampling, the full choice set is still used when updating the weights while the fixed sampled choice sets are used when maximizing the conditional likelihood function.

transformed into a recursive conditional logit estimation for each class and choice probability. By breaking the mixed logit non-IIA model into a series of conditional logit IIA models, sampling of alternatives can be reintroduced at each recursive step. The maximization procedure calculates a conditional logit (IIA) likelihood function for each class independently, keeping the individual weights fixed from the previous step.

It should be noted that when using this method, as with any mixture model, convergence may be at a local instead of a global maximum because the unconditional likelihood is not globally concave. To address this, it is often necessary to use multiple starting values.

3.5 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. Throughout the latent class literature a variety of information criteria statistics have been used. In general form (Hurvich and Tsai 1989), the information criteria statistic is specified as $IC = -2*LL + A*\gamma$ where LL is the log likelihood of the model at convergence, A is the number of estimated parameters in the model, and γ 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 γ.

[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 implausibly large. In our subsequent empirical exercise, parameter estimates for specific classes diverged from estimates for other classes by three orders of magnitude in some cases, while at the same time being coupled with a very small latent class probability. 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.

3.6 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 on 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:

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

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.[4]

4. 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.

4.1 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]

---

[4] A limitation with this approach to estimating standard errors is that because it does not estimate the empirical Hessian directly and thus cannot be used to construct robust standard errors.

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 equal to 50%, 25%, 12.5%, 5%, 2%, and 1%[5] 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.

4.2 Conditional Logit Results

Estimation code was written and executed in Matlab and Gauss. In contrast to the latent class model, the likelihood function for the conditional logit model is globally concave so starting values will effect 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. Two hundred random samples were generated for each sample size.

[Figure 1 – Estimation Time: Conditional Logit Model]

The primary benefit from sampling of alternatives is to reduce the computational burden of estimation. An analysis of sampling's effect on estimation run times suggests an almost linear relationship with diminishing returns at very small samples. Figure 1 shows the average estimation time of the sampled model relative to the estimation time of the full model. Cutting the sample by an additional 50% in any model roughly equates to a 56% reduction in estimation time, however the marginal time saved decreases with sample size.

[Table 3 –Parameter Estimates: Conditional Logit Model]

---

[5] 285, 142, 71, 28, 11, and 6 alternatives respectively

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-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 is 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,[6] 2) an increase in entry fees,[7] 3) an urban watershed management

---

[6] Supposing that a boat ramp was constructed at each Wisconsin lake that did not have one (27% of sites).
[7] $5 increase in entry fees at all state-managed sites (defined by being in a state forest or wildlife refuge); approximately 23% of sites.

program, 4) an agricultural runoff management program,[8] and 5) a fish stocking program.[9] 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 4 – Welfare Scenarios]

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

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

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

[Figure 2 – Welfare Results: Conditional Logit Model]

Figure 2 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

---

[8] 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.

[9] Fish stocking program where the catch rate of trout is increased by 25% across all sites that currently contain trout.

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.

[Table 5 – 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 2 comes with a significant benefit – a reduction in run time. To quantify the tradeoff between precision and run time, Table 5 reports 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 did exhibit similar behavior. The variation in CI ranges across the five policy scenarios is relatively small, so Table 5 only reports mean CI ranges. The results strongly suggest that for samples as small as 12.5%, the time savings are substantial while the precision losses are modest. For example, the 50% sample size estimates were generated with a 56% time savings and resulted in 6% larger CIs. Similarly, the 12.5% sample generated results with a 90% time savings while CIs were 33% larger. By contrast for sample sizes below 12.5%, the marginal reductions in run times are small while the loss in precision is substantial. For example, moving from a 12.5% to a 5% sample of alternatives reduces run times by less than ten percent but more than doubles the loss in precision. More strikingly, moving from a 5% to a 1% sample of alternatives generates a one percent run time savings but increases CI ranges more than threefold.

4.3 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 ten starting values were used for each fixed sample, the largest SLL of which was determined to be the global maximum.[10] The travel cost parameter was fixed across all classes but the remaining site characteristic parameters were random.

[Figure 3 – Estimation Time: Latent Class Model]

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. The average computation time is shown in Figure 3 and the estimation time increases substantially with each additional class. Sampling provides a decrease in relative runtime on a similar scale as in the conditional logit model, with diminishing time savings at much smaller sample size.

Additionally, sampling did not seem to decrease runtimes uniformly as in the conditional logit models. Convergence problems lead to increased estimation times when sampling in an overspecified model (specifically, the five and six class models). In a correctly specified model (three classes), each successive sample reduces overall runtime. Model selection issues are described in the next section.

For relatively small sample sizes with large numbers of classes, convergence was sometimes elusive. This may be the result of a specific random sample chosen having an 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 are sample selection concerns with these results. The properties of the random samples that

---

[10] 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 ten 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.

do not converge were not examined and the existence, source, and magnitude of any bias remain an avenue of further study.

[Table 6 – 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.[11] 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 sets 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 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 5-class model attempting to use a single class to account for a handful of anomalous outlying observations. Thus, in our view, the more appropriate decision criterion is the crAIC which incorporates the greatest penalty for an increased number of parameters.[12]

Recall that all the LL values are calculated with the parameter estimates from each sampled model using the full choice set. This methodology simplifies comparison between models that use different numbers of classes as well as sample sizes. We previously found that the three-class model is preferred to the conditional logit model when using the full choice set. We also know that reducing the sample size while keeping the number of classes fixed reduces computation time dramatically.

[Figure 4 – 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 4. As can be seen, the latent class model can

[11] The AIC and BIC indicated the same number of optimal classes as the CAIC.
[12] Scarpa and Thiene (2005) found similar results and drew similar conclusions.

recover preference distributions that do not necessarily resemble commonly used continuous parametric distributions (e.g., normal, uniform).

[Figure 5 – Welfare Results: Latent Class Model]

The stability of the WTP estimates across sampling in the mixed model is analyzed in Figure 4 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 summed together to produce a single value for each run. In the sampled models, the welfare estimates reported are the weighted average of ten random samples.

Interpreted, 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 between $0.73 and $0.99.[13] The researcher could conversely have run a 5% random sample of alternatives 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.

[Table 7 – Increase in Range of 95% CI of WTP Estimates Compared to Full Choice Set: Latent Class Model]

Table 7 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 5% level.[14] Relative to using the full choice set, the 50% 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 25%, 12.5%, and 5% samples respectively.[15] At the 2% and 1% levels, the CIs for some of the samples are extremely large. In this model, WTP estimates at sample sizes below 5% could be

[13] The CI represents the interval that is likely to include the parameter estimate.
[14] This value is 1% for the conditional logit model.
[15] These values are 6%, 16%, 33%, and 76% respectively for the conditional logit model.

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.

5. 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 of alternatives as low as 12.5% of the full choice set. Smaller sample of alternative sizes generate small reductions in run times at the cost of substantial increases in the range of confidence intervals. Depending on the needs of the researcher however, results may be useful down to the 1% sample size.

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 alternatives as low as 25%. Although sample sizes may be useful down to the 5% level, they come with relatively small time savings and substantial losses in precision. Estimates from sample sizes below 5% were found to be unreliable in our application.

Although a broad selection of samples sizes were tested, it is unclear whether the percent sample or absolute sample size play the predominant role in the efficiency loss and time savings of sampling. Other research (von Haefen and Jacobson, unpublished) indicates through a Monte Carlo exercise that absolute sample size may be more important. Additional tests performed with an alternative dataset may be necessary to recover this result.

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 5%) should be avoided. See other work by von Haefen and Jacobsen (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. Additional research analyzing bias in sampling techniques would also be valuable.

References

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

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

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

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

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

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

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

Boersch-Supan, A. and V. Hajvassiliou. (1990). "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models." Journal of Econometrics, 58, pp. 347-368.

Dempster, A., N. Laird, and D. Rubin. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm." Journal of the Royal Statistical Society. Series B, 39(1), pp. 1-38.

Desarbo, W., M. Wedel, M. Vriens, and V. Ramaswamy. (1992). "Latent Class Metric Conjoint Analysis." Marketing Letters, 3(3), pp. 273-288.

Feather, P. (1994). "Sampling and Aggregation Issues in Random Utility Model Estimation." American Journal of Agricultural Economics, 76(4), pp. 772-780.

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

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

Geweke, J., M. Keane, and D. Runkle (1994). "Alternative Computational Approaches to Inference in the Multinomial Probit Model." Review of Economics and Statistics, 76, pp. 609–632.

Greene, W. and D. Hensher (2003). "A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit." Transportation Research B: Methodological 37(8), pp 681-698.

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

Haab, T. and R. Hicks (1997). "Accounting for Choice Set Endogeneity in Random Utility Models of Recreation Demand." Journal of Environmental Economics and Management, 34, pp. 127-147.

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

Heckman, J. and B. Singer (1984). "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." Econometrica, 52, pp. 271-330.

Hensher, D. and W. Greene (2003). "The Mixed Logit Model: The State of Practice." Transportation, 30(2), pp 133-176.

Hess, S. and J.M. Rose (2006), "Effects of Distributional Assumptions on Conditional Estimates from Mixed Logit Models." ETH Working Paper, Arbeitsberichte Verkehrs- und Raumplanung, 423, IVT, ETH Zurich, Zurich.

Horowitz, J. and J. Louviere (1995). "What is the Role of Consideration Sets in Choice Modeling?" International Journal of Research in Marketing, 12(1), pp. 39-54.

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

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

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

Kling, C. and C. Thomson (1996). "The Implications for Model Specification for Welfare Estimation in Nested Logit Models." American Journal of Agricultural Economics, 78, pp. 103-114.

Landry, C. and H. Liu (2009). "A Semi-Parametric Estimator for Revealed and Stated Preference data - An Application to Recreational Beach Visitation." Journal of Environmental Economics and Management, 57(2), pp. 205-218.

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

Louis, T. (1982). "Finding the Observed Information Matrix when Using the EM Algorithm." Journal of the Royal Statistical Society. Series B, 44(2), pp. 226-233.

Lupi, F. and P. Feather (1998). "Using Partial Site Aggregation to Reduce Bias in Random Utility Travel Cost Models." Water Resources Research, 34(12), pp. 3595-3603.

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

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

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

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

McFadden, D. and P. Ruud (1994). "Estimation by Simulation." Review of Economics and Statistics, 76(4), pp. 591-608.

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

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

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

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

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

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

Parsons, G. and A. B. Hauber (1998). "Spatial Boundaries and Choice Set Definition in a Random Utility Model of Recreation Demand." Land Economics, 74(1), pp. 32-48.

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

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

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

Revelt, D. and K. Train (1998). "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level." Review of Economics and Statistics, 80, pp. 1–11.

Rigby, D., K. Balcombe, and M. Burton (2009). "Mixed Logit Model Performance and Distributional Assumptions: Preferences and GM Foods." Environmental and Resource Economics 42(3), pp. 279-295.

Roeder, K., K. Lynch, and D. Nagin (1999). "Modeling Uncertainty in Latent Class Membership: A Case Study in Criminology." Journal of the American Statistical Association, 94, pp. 766-776.

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

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

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

Shaw, W. and M. Ozog (1999). "Modeling Overnight Recreation Trip Choice: Application of a Repeated Nested Multinomial Model." Environmental and Resource Economics, 13(4), pp. 397-414.

Small, K. and H. Rosen (1981). "Applied Welfare Economics with Discrete Choice Models." Econometrica, 49(1), pp. 105-130.

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

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

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

Train, K. and G. Sonnier (2003). "Mixed Logit with Bounded Distribution of Partworths." Working paper. University of California, Berkeley and Los Angeles.

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

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

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

von Haefen, R. and M. Jacobsen (unpublished). "Sampling of Alternatives in a Mixture Model." Working Paper, Department of Agricultural and Resource Economics, North Carolina State University.

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

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

Table 1
Information Criteria Statistics

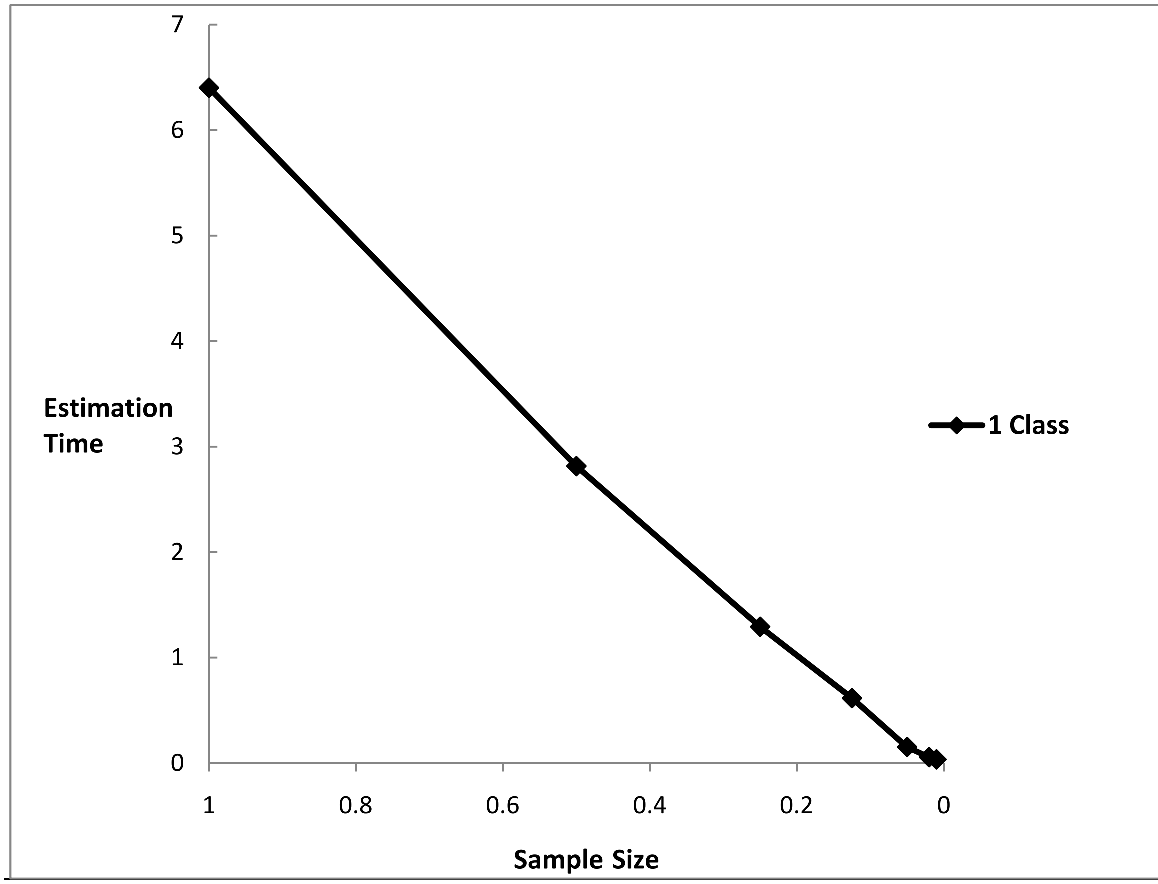
Information Criteria	Penalty Constant γ
Akaike Information Criteria	2
Bayesian Information Criteria	$\ln(N)$
Consistent Akaike Information Criteria	$1+\ln(N)$
Corrected Akaike Information Criteria	$2 + 2(A + 1)(A + 2)/(N - A - 2)$

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

Table 2
Summary Statistics

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

Figure 1
Estimation Time
Conditional Logit Model



* Estimation time represents the average time to convergence (in minutes) for 200 Monte Carlo runs on a 3 GHz 64 bit dual processor with 8 GB of RAM.

Table 3
Parameter Estimates: Conditional Logit Model

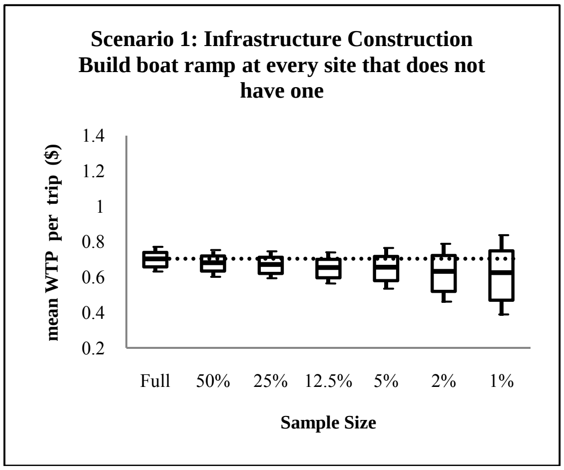
Sample Size	**50% [285]**	**25% [142]**	**12.5% [71]**	**5% [28]**	**2% [11]**	**1% [6]**	**50% [285]**
SLL	-13257	-10901	-8640	-6568	-4172	-2314	-1414
se		(42)	(56)	(65)	(75)	(73)	(63)
NLL	-13257	-13264	-13274	-13294	-13344	-13432	-13542
se		(4)	(6)	(11)	(27)	(53)	(105)
Variable							
tcost	**-10.07**	**-10.026**	**-9.907**	**-9.7**	**-9.303**	**-8.84**	**-8.6**
	(0.412)	**(0.407)**	**(0.4)**	**(0.393)**	**(0.39)**	**(0.401)**	**(0.434)**
ramp	**0.421**	**0.407**	**0.397**	**0.38**	**0.371**	0.349	0.343
	(0.176)	**(0.177)**	**(0.178)**	**(0.179)**	**(0.184)**	(0.195)	(0.217)
refuge	0.165	0.169	0.177	0.19	0.202	0.229	0.24
	(0.194)	(0.195)	(0.196)	(0.199)	(0.211)	(0.233)	(0.27)
forest	0.152	0.142	0.131	0.124	0.128	0.174	0.229
	(0.171)	(0.174)	(0.177)	(0.182)	(0.185)	(0.202)	(0.227)
urban	-0.068	-0.084	-0.092	-0.094	-0.08	-0.024	0.032
	(0.117)	(0.118)	(0.119)	(0.122)	(0.134)	(0.157)	(0.186)
restroom	0.149	0.149	0.144	0.143	0.147	0.174	0.211
	(0.13)	(0.131)	(0.131)	(0.13)	(0.134)	(0.148)	(0.168)
river	-0.013	-0.015	-0.02	-0.032	-0.07	-0.18	-0.266
	(0.297)	(0.299)	(0.305)	(0.319)	(0.348)	(0.399)	(0.458)
small lake	**-0.789**	**-0.789**	**-0.776**	**-0.765**	**-0.724**	**-0.707**	**-0.702**
	(0.161)	**(0.164)**	**(0.168)**	**(0.175)**	**(0.185)**	**(0.209)**	**(0.236)**
trout	**1.651**	**1.674**	**1.728**	**1.781**	**1.975**	**2.49**	**2.972**
	(0.566)	**(0.571)**	**(0.584)**	**(0.605)**	**(0.668)**	**(0.797)**	**(0.933)**
smallmouth	**0.943**	**0.948**	**0.972**	**0.996**	**1.05**	**1.149**	**1.174**
	(0.359)	**(0.362)**	**(0.371)**	**(0.376)**	**(0.369)**	**(0.385)**	**(0.427)**
walleye	**2.69**	**2.652**	**2.605**	**2.54**	**2.457**	**2.433**	**2.364**
	(0.379)	**(0.376)**	**(0.374)**	**(0.38)**	**(0.405)**	**(0.479)**	**(0.561)**
northern	**2.659**	**2.536**	**2.328**	2.013	1.505	0.908	0.658
	(0.935)	**(0.955)**	**(0.998)**	(1.07)	(1.215)	(1.472)	(1.746)
musky	**5.361**	**6.136**	**6.809**	**7.417**	**8.375**	**9.158**	**9.769**
	(1.346)	**(1.752)**	**(2.069)**	**(2.324)**	**(2.585)**	**(2.803)**	**(3.376)**
salmon	**7.733**	**7.852**	**7.968**	**8.043**	**8.139**	**7.848**	**7.545**
	(1.384)	**(1.405)**	**(1.441)**	**(1.503)**	**(1.635)**	**(1.88)**	**(2.132)**
panfish	**0.763**	**0.769**	**0.78**	**0.789**	**0.804**	**0.814**	**0.814**
	(0.189)	**(0.189)**	**(0.191)**	**(0.194)**	**(0.201)**	**(0.223)**	**(0.253)**

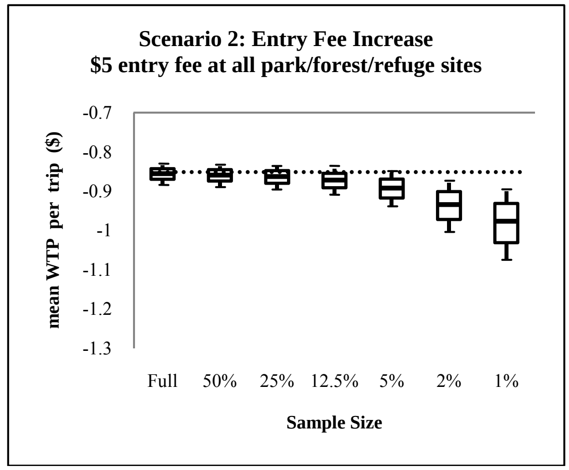
* Results for the sampled models represent the mean of 200 random samples; clustered non-robust standard errors in parentheses; **bold** indicates significance at the 5% level; "NLL" is the log-likelihood calculated at the parameter values for the entire choice set for comparison purposes.

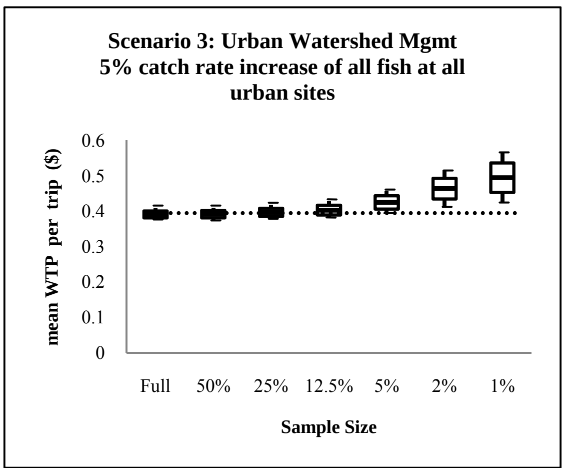
Table 4
Welfare Scenarios

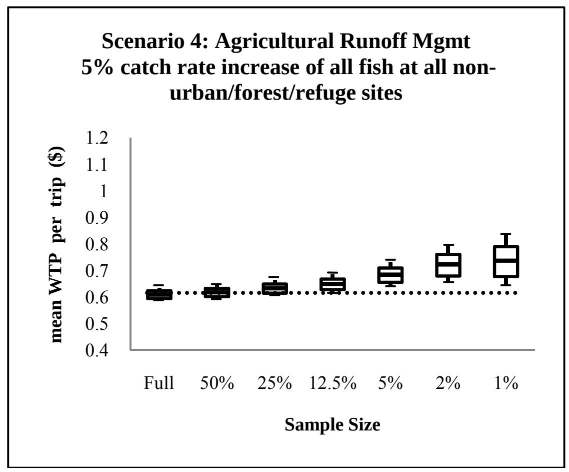
Scenario	Impacted Characteristics	Affected Sites
Infrastructure Construction	Build boat ramp at every site that does not have one	27.4%
Entry Fee Increase	$5 entry fee at all park/forest/refuge sites	23.4%
Urban Watershed Management	5% catch rate increase for all fish at all urban sites	17.9%
Agricultural Runoff Management	5% catch rate increase for all fish at all non-urban/forest/refuge sites	30.1%
Fish Stocking Program	25% increase in Trout catch rate across all sites	98.4%

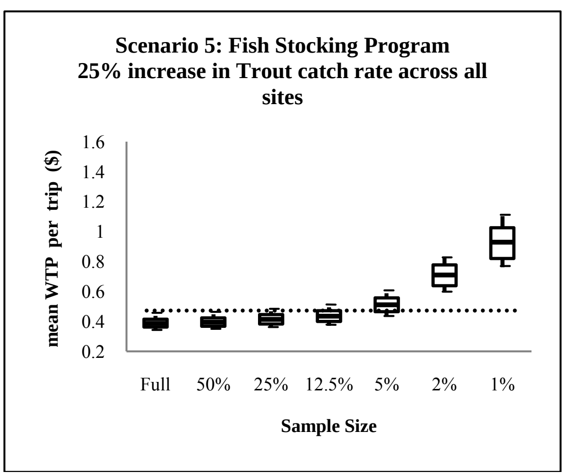
Figure 2
Welfare Results
Conditional Logit Model











* 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.

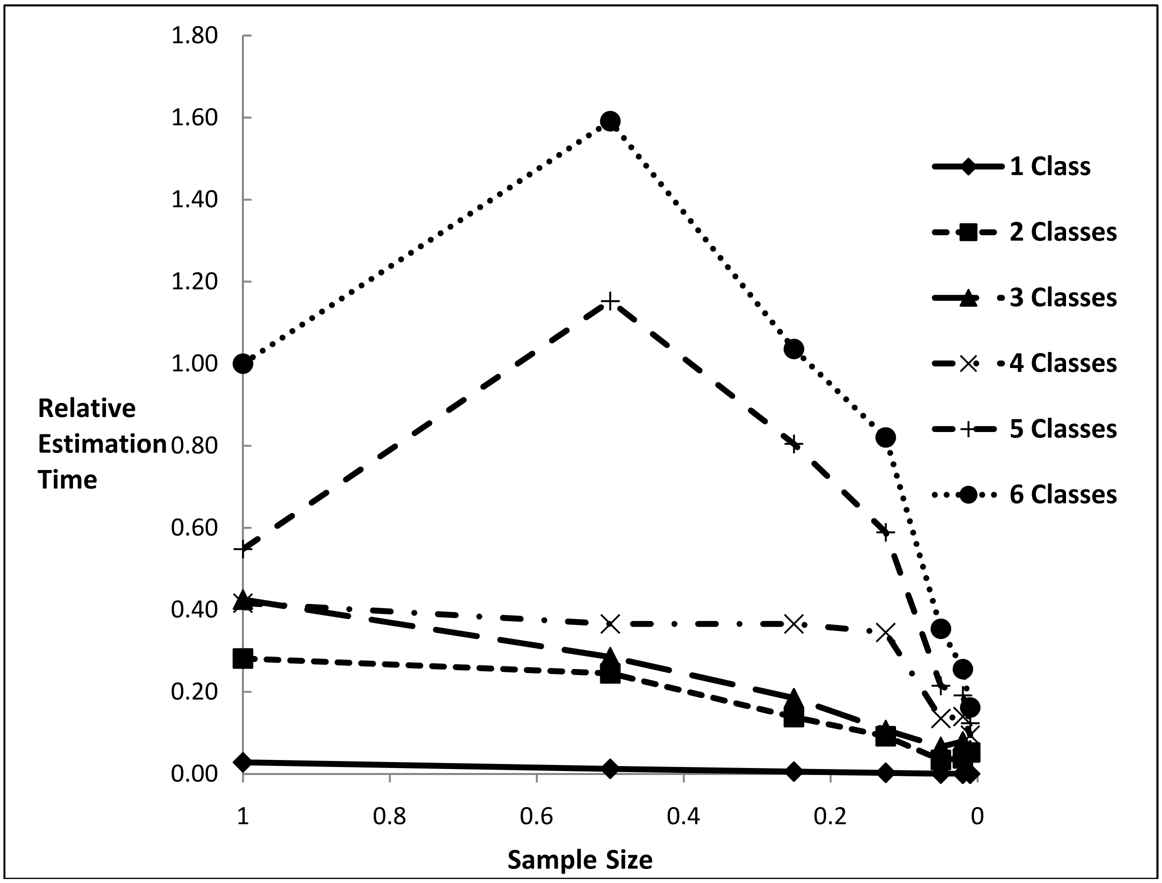
* The dashed line represents the mean WTP estimate for the full sample model.

Table 5
Increase in Range of 95% CI of WTP Estimates
Conditional Logit Model

	Sample Size					
	50% [285]	**25% [142]**	**12.5% [71]**	**5% [28]**	**2% [11]**	**1% [6]**
Efficiency Loss	6%	16%	33%	76%	165%	272%
Percent Error	1%	3%	6%	13%	28%	42%
Time Savings	56%	80%	90%	98%	99%	99%

* Compared to the full choice set. Absolute sample size in brackets. Efficiency Loss is the percent increase in the range of the 95% CI. Percent Error is the percentage deviation of the mean WTP.

Figure 3
Estimation Time
Latent Class Model



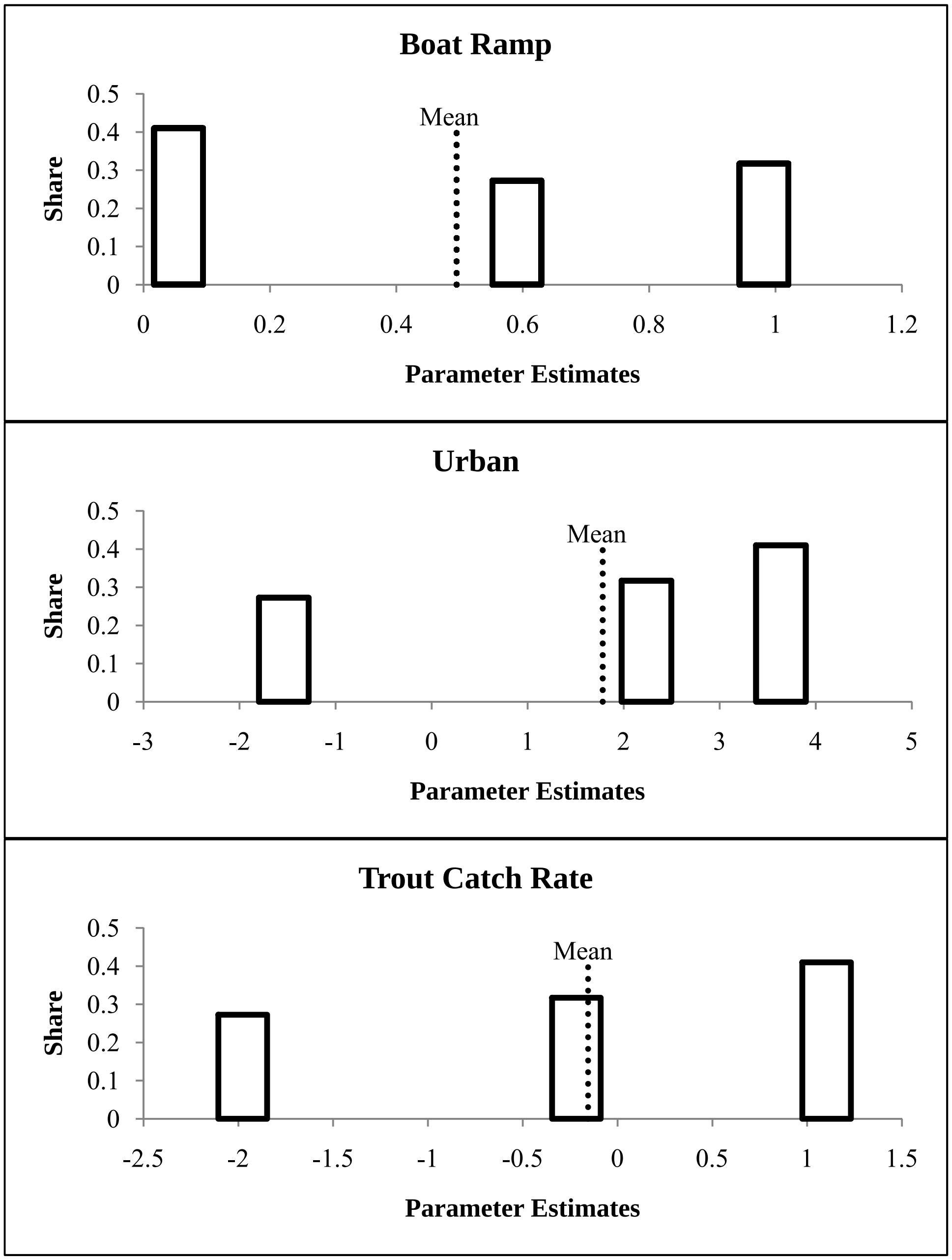
* Estimation Time represents the time to convergence (in minutes) for one random sample with one set of starting values on a 3 GHz 64 bit dual processor with 8 GB of RAM.

Table 6
Information Criteria

	# of classes	Sample Size Full	50%	25%	12.5%	5%	2%	1%
NLL	**1**	-13257	-13264	-13274	-13294	-13344	-13432	-13542
	2	-12416	-12425	-12447	-12471	-12586	-12905	-13264
	3	-12025	-12042	-12066	-12084	-12220	-12526	-12872
	4	-11724	-11734	-11757	-11805	-12010	-12311	*DNC*
	5	-11432	-11486	-11532	-11592	-11748	-12232	*DNC*
CAIC	**1**	26623	26637	26657	26697	26797	26973	27193
	2	25071	25089	25132	25180	25411	26049	26767
	3	24420	24453	24500	24538	24809	25422	**26113**
	4	23947	23967	24013	24110	24518	25122	*DNC*
	5	**23494**	**23601**	**23693**	**23813**	**24127**	**25093**	*DNC*
crAIC	**1**	26560	26574	26594	26634	26734	26910	27130
	2	25063	25081	25124	25172	25403	26040	26759
	3	**24765**	**24798**	**24846**	**24883**	**25154**	**25767**	**26458**
	4	25141	25161	25207	25304	25712	26316	*DNC*
	5	26260	26367	26459	26579	26892	27859	*DNC*

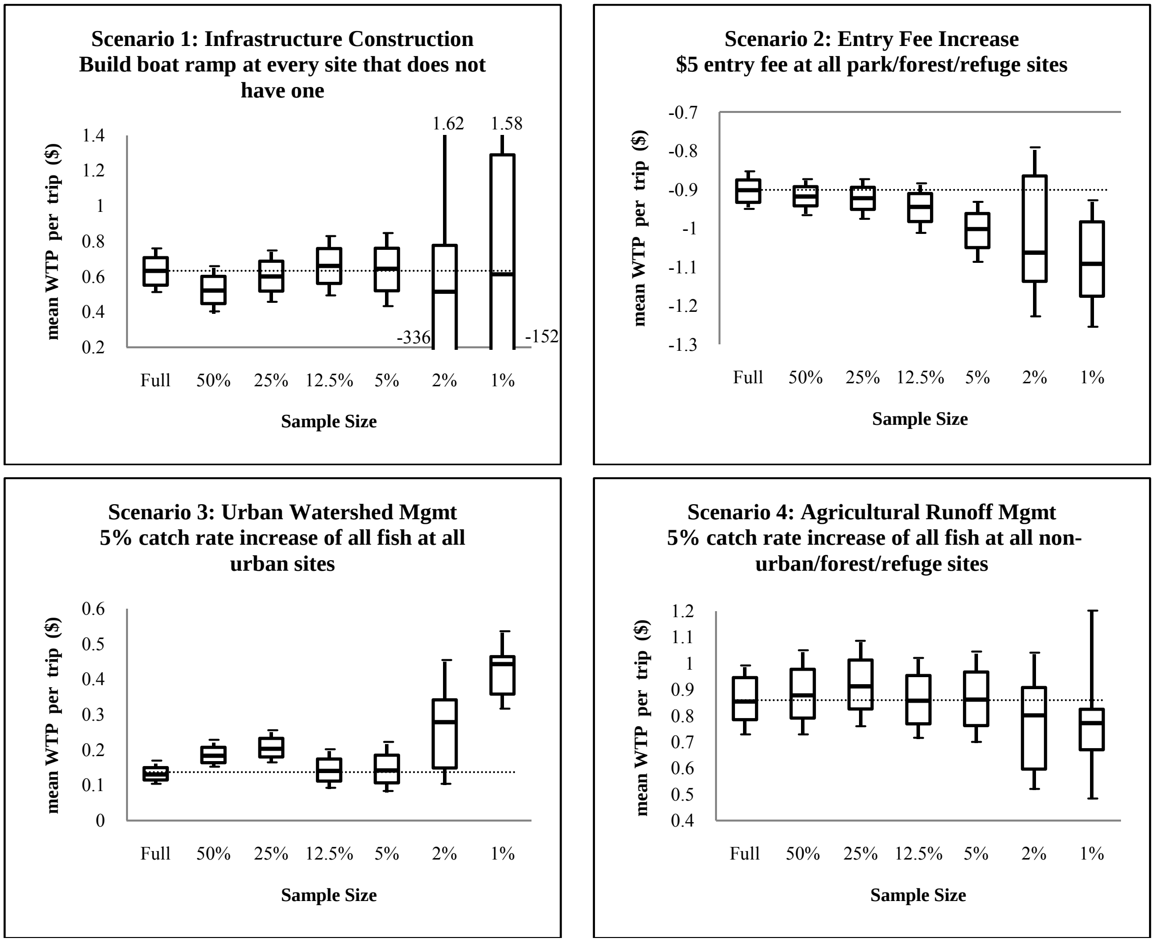
* CAIC, and crAIC calculated using the LL calculated with the full choice set; Mean of 200 and ten random samples reported for the one and multiple class models respectively; Optimal # of classes outlined and in **bold** defined by the minimum of the information criteria; *DNC* = model did not converge.

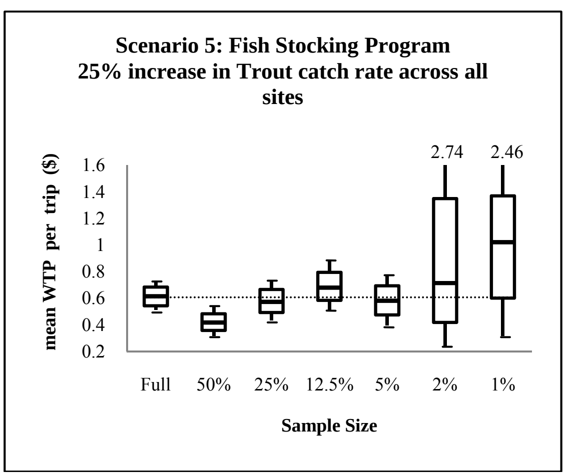
Figure 4
Parameter Estimate Distributions



* 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 mean.

Figure 5
Welfare Results
Latent Class Model (crAIC)





* 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.

Table 7
Increase in Range of 95% CI of WTP Estimates
Latent Class Model

	Sample Size					
	50% [285]	**25% [142]**	**12.5% [71]**	**5% [28]**	**2% [11]**	**1% [6]**
Efficiency Loss	10%	28%	51%	76%	84632%	18360%
Percent Error	19%	15%	6%	6%	34%	60%
Time Savings	33%	56%	75%	84%	81%	86%

* Compared to the full choice set. Absolute sample size in brackets. Efficiency Loss is the percent increase in the range of the 95% CI. Percent Error is the percentage deviation of the mean WTP.