Full information maximum likelihood estimation of heterogeneous preferences with choice data and other preference statements: a joint latent-class model

William Breffle *

School of Business and Economics, Michigan Technological University

Edward Morey †

Department of Economics, University of Colorado-Boulder

Jennifer Thacher

Department of Economics, University of New Mexico

September 8, 2008

*Corresponding author: Michigan Technological University; 104 Academic Office Building; 1400 Townsend Dr; Houghton, MI 49931-1295. Tel: 906.487.1959. Email: wsbreffl@mtu.edu

†All are equal authors.

Abstract

Preference surveys ask choice questions and direct questions about preferences (e.g., the extent of agreement with a statement). Most econometric models that estimate preferences ignore preference-statement data and rely solely on choice data. We develop and estimate a FIML model that uses all the preference data to simultaneously explain both the observed choices and statements of preference. Using the E-M algorithm, preferences and preference heterogeneity are estimated in the context of a discrete-choice, random-utility, latent-class model, simultaneously estimating the: (1) number of preference classes, (2) probability that an individual belongs to a particular class, (3) attribute parameters in the conditional, indirect-utility function for each class, and (4) probability for each preference statement that an individual in a particular class will give a particular response. The joint model generally tells the same qualitative story as a model using choice data alone. However, while the weighted mean WTP does not differ between the two models, there are significant differences in estimated WTP for each class. Incorporating preference-statement data improves efficiency and the ability to identify parameters.

Key words: Latent class, choice data, FIML, preference heterogeneity, preference statements, Likert-scale questions

JEL code: C51,Q51,D12

In addition to choice questions (revealed and stated choices), preference surveys typically include other questions that provide information about preferences. A common type of question assesses the relative importance the individual places on different attributes of a good, providing evidence of the individual's preferences towards those attributes. Another type of question presents a preference statement to the respondent and asks the extent to which she agrees or disagrees. Below are three examples of this latter type of question. The first is from a survey of anglers [1], the second from a survey of depressed individuals about possible treatment side-effects [2], and the third from a survey of mountain bikers [3].

> For the fish you would like to fish for in the waters of Green Bay, how much would it bother you, if at all, if PCBs resulted in the following fish consumption advisory: "Do not eat". (1='Not at all bothersome', ..., 5='Very bothersome')
>
> How much would little or no interest in sex bother you? (1='Not at all', 2='Slightly', 3='Some', 4='A fair bit', 5='A lot')
>
> I hate trying to keep up with riders faster than me. (1='Definitely agree', 2='Somewhat agree', 3='Neither agree nor disagree', 4='Somewhat disagree', 5='Definitely disagree')

In each question the respondent chooses the *response category* most consistent with his or her preferences. What distinguishes these questions from conventional choice questions is that the individual is not choosing between states of the world, actual or hypothetical (e.g., choice questions or a referendum CVM). Rather the individual is choosing her level of agreement with an expression of preference or choosing a response category that best answers a direct question about his preferences. We call this other type of preference data *preference statements*. The response categories in preference statements are often Likert scales, such as those shown above, but do not have to be.

We proceed assuming preferences are latent (unobservable) and both choice data (actual and hypothetical) and preference statements are manifestations of those latent preferences. If correct, there is a strong argument for incorporating the latter data into the estimation of preferences and preference heterogeneity [4, 5, 6]. This approach is analogous to combining revealed preference and stated preference data in order to increase estimation efficiency.[1]

[1]See, for example, Ben-Akiva and Morikawa [7], Cameron [8], Hensher and Bradley [9], and Adamowicz *et al.* [10].

We assume a discrete-choice, latent-class, random-utility model. By assumption, there is some small but unknown number of preference classes. Everyone within a class has the same preferences but preferences vary across the classes. Class membership is latent. Using the E-M (expectation-maximization) algorithm [11], we implement full-information maximum likelihood (FIML) estimation to find the model parameters that maximize the likelihood of observing both the choice data and the preference statements. To our knowledge, this is the first specification and estimation of a FIML model with these two types of data and a rare application of the E-M algorithm with environmental data.

We estimate a latent-class joint model for the preferences over the fishing characteristics of Green Bay, a large bay on Lake Michigan. While Green Bay is a popular fishing area, it is contaminated by PCBs and has fish-consumption advisories in place. The observed data include the choices made over hypothetical fishing alternatives, the attributes of these alternatives, and preference statements. Looking ahead, preference heterogeneity is adequately described with three classes. The model fits the data well for both types of data. Willingness to pay (WTP) estimates and other summary statistics from our joint model are presented, discussed, and compared with restricted forms of the model. We find that for this particular application, qualitative results are similar between all the models examined; however, there are significant quantitative differences. In addition, the joint model shows significant gains in efficiency and estimation.

1 Background and terminology

In order to distinguish our model from previous work, we first identify three types of latent-class models. A *latent-class choice-only model* is a latent-class model estimated with only choice data. The choice data is used to estimate the number of classes, the probability of class membership, and the preference parameters in each class's conditional indirect-utility function. No preference-statement data is used in latent-class choice-only models. Economic examples of latent-class choice-only models are multiplying and include Provencher *et al.* [12], Greene and Hensher [13], Scarpa and Thiene [14], Scarpa *et al.* [15], Kemperman and Timmermans [16], Colombo and Hanley [17], and Patunru *et al.* [18].

Alternatively, one can estimate a latent-class model of preferences using only preference statements - a *latent-class preference-statement model.* These models estimate the number of preference classes, the probabilities of class membership, and probability of answering level s to preference statement q; they do not estimate preference parameters, by class, in utility functions. No choice data is used. Examples of latent-class preference-statement models include include McCutcheon [19], Thacher *et al.* [2], Morey *et al.* [20], Aldrich *et al.* [21], Owen and Videras [22], and Morey *et al.* [23].

Here, we build on previous work [20] and integrate choice data and preference statements in a FIML latent-class model. Designate this a *latent-class joint model.*[2] It is a joint model because it simultaneously utilizes two very different types of preference data. Combining the choice data and preference-statement data results in more efficient estimates than would be obtained by either data type alone. Restricted forms of our joint model include the latent-class preference-statement model,

[2]Morey *et al.* [20] proposed but did not estimate a sequential model using both types of data.

the latent-class choice-only model, and a standard logit model of choice estimated with choice data.

It is important to distinguish our model from the latent-class models of Swait and Sweeney [24], Boxall and Adamowicz [6], Owen and Videras [22], and Patunru *et al.* [18]. These models also combine choice data and preference statements but use this data differently from our model.

Swait and Sweeney [24] and Boxall and Adamowicz [6] use preference statements as exogenous variables to explain observed choices, whereas our model explains both the choices and the preference statements. Owen and Videras [22] and Patunru *et al.* [18] both are sequential estimators. Owen and Videras [22] estimate a latent-class preference-statement model, then impose each respondent's estimated class-membership probabilities from that model on a probit model of environmental choice. Rather than estimating a latent-class choice model with these probabilities imposed (the sequential estimator suggested in Morey *et al.* [20]), they make the probability of making an environmental choice a direct function of each respondents class-membership probabilities. That is, in the second step, they do not estimate a latent-class choice model. Patunru *et al.* [18] estimate a latent-class preference-statement model, then deterministically assign each respondent to the class associated with their highest estimated class-membership probability. A separate choice model is then estimated for each class, so class membership is deterministic at the second step.

2 The latent-class joint model

Assume the population consists of C different preference classes. The researcher observes, for each individual, the data $(\mathbf{x}_i, \mathbf{y}_i)$; the matrix $\mathbf{x}_i$ is the set of individual i's answers to the preference statements, where $x_{iqs} = 1$ if individual i's answer to statement q is level s and 0 otherwise.

The vector $\mathbf{y}_i$ is individual i's discrete choice data (revealed preference, stated preference, or some combination), where $y_{ijk} = 1$ if individual i chose alternative j in choice pair k and 0 otherwise.

Latent-class models assume that individuals in the same class respond and behave similarly to one another: the response patterns of individuals from the same class are more correlated with each other than with individuals in other classes. A standard assumption in latent-class models is that once one has conditioned on class, all responses are independent, both across questions and across individuals. Put simply, the correlation is completely induced by the latency of class membership; an individual's answers to all of the stated-choice and preference statements are independent of one another once one conditions on class.

If one observes $\mathbf{x}_i$, $\mathbf{y}_i$, and class is unobserved, the likelihood function is:

$$L = \prod_i \left[\Pr\left(\mathbf{x}_i, \mathbf{y}_i\right) \right] = \prod_i \left[\sum_{c=1}^{C} \Pr(c) \Pr\left(\mathbf{x}_i, \mathbf{y}_i | c\right) \right], \tag{1}$$

where $\Pr(c)$ is the unconditional probability of belonging to class c. $\Pr\left(\mathbf{x}_i, \mathbf{y}_i | c\right)$ is a conditional probability and represents the probability of observing the individual's responses, conditional on belonging to class c.

Given the independence induced by conditioning on class, the likelihood function can be rewritten as:

$$L = \prod_i \left[\sum_{c=1}^{C} \Pr(c) \Pr\left(\mathbf{x}_i | c\right) \Pr\left(\mathbf{y}_i | c\right) \right], \tag{2}$$

where

$$\Pr\left(\mathbf{x}_i | c\right) = \prod_{q=1}^{Q} \prod_{s=1}^{S} (\pi_{qs|c})^{x_{iqs}} \tag{3}$$

and

$$\Pr\left(\mathbf{y}_i | c\right) = \prod_{k=1}^{K} \prod_{j=1}^{J} (P_{jk|c})^{y_{ijk}}. \tag{4}$$

$\pi_{qs|c}$ is the probability that an individual in class c answers level s to statement q. $P_{jk|c}$ is the probability of choosing alternative j in discrete-choice set k, conditional on being a member of class

c. Each $\pi_{qs|c}$ is estimated as a separate parameter subject to the constraint that $\sum_{s=1}^{S} \pi_{qs|c} = 1$.[3] The $P_{jk|c}$ are functions of the parameters in the class-specific conditional-indirect utility functions, the β_c parameters. That is, $P_{jk|c}$ can be a probit or logit probability of choosing alternative j from discrete-choice set k, conditional on being a member of class c.

Estimation is with the E-M algorithm, a Bayesian process.[4] The E-M algorithm works as follows. First, one selects initial values - guesses are fine - for each of the conditional probabilities, $\Pr(c\,|\mathbf{x}_i, \mathbf{y}_i)$. These initial values are used to calculate the unconditional class membership probabilities, $Pr(c)$, which are simply the averages of the conditional class-membership probabilities for all individuals:

$$\Pr(c) = \frac{1}{N} \sum_{i=1}^{N} \Pr(c\,|\mathbf{x}_i, \mathbf{y}_i)\,. \tag{5}$$

The conditional probabilities, unconditional class-membership probabilities, and preference statements are then used to calculate the $\pi_{qs|c}$ (probability of answering level s to statement q, if in class c). The formula, obtained by maximizing the likelihood function and solving the first order conditions, is:

$$\pi_{qs|c} = \frac{\sum_{i=1}^{N} \Pr(c_i\,|\mathbf{x}_i, \mathbf{y}_i)\, x_{iqs}}{\Pr(c)N}. \tag{6}$$

The ratio is an estimate of the proportion of times individuals in class c answer level s to statement q. Then the estimated $\pi_{qs|c}$ are substituted into Equation 3 to calculate the conditional probability of observing the individual's responses, $\Pr(\mathbf{x}_i|c)$.

The above steps are an application of the E-M algorithm: one finds the values of the $\Pr(c)$ and

[3]This specification does not restrict the response categories to have an ordering, so is consistent with both nominal and ordinal response categories. Imposing ordinality on the response categories reduces the number of parameters to estimate but is restrictive.

[4]In theory, estimation can also be done by directly maximizing the full likelihood function. However, given the large number of parameters, this is often not possible in practice.

the $\Pr(\mathbf{x}_i|c)$ that maximize the expectation of the joint likelihood function. It is an "expected" likelihood function because one is using the expected values of the conditional membership probabilities as if they were the true values. After plugging the unconditional class membership probabilities and the conditional response category probabilities into the likelihood function (Equation 2), a maximization algorithm (such as Optimum or Maxlik in GAUSS [25]) can be used to maximize the likelihood function with respect to the choice-based preference parameters, β_c.

At this point, better estimates of the conditional class membership probabilities ($\Pr(c|\mathbf{x}_i, \mathbf{y}_i)$) can be obtained using Bayes theorem:

$$\Pr(c|\mathbf{x}_i, \mathbf{y}_i) = \frac{\Pr(c) \prod_{q=1}^{Q} \prod_{s=1}^{S} (\pi_{qs|c})^{x_{iqs}} \prod_{k=1}^{K} \prod_{j=1}^{J} P_{jkc}^{y_{ijk}}}{\Pr(\mathbf{x}_i, \mathbf{y}_i)}. \tag{7}$$

This completes one iteration of the E-M algorithm.

The process is repeated with these new, better estimates of the conditional class-membership probabilities. Continue iterating until the conditional likelihood function increases by less than some predetermined amount. The result is FIML estimates of all of the parameters.[5]

2.1 The latent-class choice-only model and the latent-class preference-statement model

For purposes of comparison, consider two sub-models of the joint model: the latent-class choice-only model and the latent-class preference-statement model.

The likelihood function for the latent-class choice-only model is:

$$L_{choice} = \prod_i \left[\sum_{c=1}^{C} \Pr(c) \Pr(\mathbf{y}_i|c) \right]. \tag{8}$$

It is a function of only the choice data. The preference statements are ignored. The estimates of C, the $\Pr(c)$, and the β_c are those that maximize Equation 8. The software Latent Choice [26] has

[5] Gauss code to estimate the latent-class joint model is available by request from the first author.

a package to maximize Equation 8, or one can program the likelihood function in software such as GAUSS [25] or R [27].

The likelihood function for the latent-class preference-statement model is:

$$L_{PS} = \prod_i \left[\sum_{c=1}^{C} \Pr(c) \Pr(\mathbf{x}_i | c) \right]. \tag{9}$$

It is a function of only the preference statements. The choice data is ignored. The estimates of C, the $\Pr(c)$, and the $\pi_{qs|c}$ are those that maximize Equation 9. One can maximize Equation 9 using the *LC Cluster* package in Latent Gold [28] or software such as GAUSS [25] or R [27].

The choice-only model and the preference-statement model will likely estimate different C and $\Pr(C)$. In both cases, although results are consistent, efficiency is not achieved because neither method uses all of the data.

Morey *et al.* [20] suggested but did not implement a sequential estimator for all the choice data and preference statements: first estimate a latent-class preference-statement model, use it to estimate conditional class-membership probabilities ($\Pr(c\,|\mathbf{x}_i)$) for all of the respondents, then estimate a latent-class choice-only model with these estimated class-membership probabilities imposed. This sequential estimator is inefficient relative to our joint FIML estimator, so not presented.

3 Application: preferences of Green Bay anglers

To demonstrate the approach, we apply the latent-class joint model to estimate preferences over the fishing characteristics of Green Bay. The goal is to simultaneously use the choice data and preference statements to characterize the preferences, and heterogeneity in those preferences, of anglers for the fishing characteristics of Green Bay. The site characteristics examined are launch fees, catch rates by species (yellow perch, walleye, salmon, and bass), and fish consumption advisory (FCA) levels for PCBs.

The target population is active Green Bay anglers who purchase Wisconsin fishing licenses in eight Wisconsin counties near Green Bay; most Green Bay fishing days are by these anglers. The sample consists of 640 anglers. See Breffle *et al.* [1] for a full description of the data.

Anglers answered fifteen preference statements and eight stated-preference choice-questions. The choice-questions were of the type: Would you rather fish Green Bay under conditions A or B? (See Figure 1.)

The fifteen preference statement are:

1. On a scale from 1 to 7 where 1 means "Much Worse" and 7 means "Much Better", how do you rate the quality of fishing on the water of Green Bay compared to other places you fish?

2. On a scale from 1 to 5 where 1 means "Not at all Bothersome" and 5 means "Very Bothersome", answer the following question. For the fish you would like to fish for in the waters of Green Bay, how much would it bother you, if at all, if PCBs resulted in the following fish consumption advisories:

   (a) Eat not more than one meal a week.

   (b) Eat not more than one meal a month.

   (c) Do not eat.

3. On a scale from 1 to 5 where 1 means "Strongly Disagree" and 5 means "Strongly Agree", how do you feel about each of the following statements about boat launch fees? If you don't fish from a boat, please think of the daily boat launch fee as a

fee you would have to pay to fish the waters of Green Bay.

   (a) I would be willing to pay higher boat launch fees if catch rates were higher on the waters of Green Bay.
   (b) I would be willing to pay higher boat launch fees if the fish had no PCB contamination.

4. On a scale from 1 to 5 where 1 is "Not at all important" and 5 is "Very Important", when you were making your choices how important were each of the following?
   (a) The average catch rate for yellow perch
   (b) The fish consumption advisory for yellow perch
   (c) The average catch rate for trout/salmon
   (d) The fish consumption advisory for trout/salmon
   (e) The average catch rate for walleye
   (f) The fish consumption advisory for walleye
   (g) The average catch rate for smallmouth bass
   (h) The fish consumption advisory for smallmouth bass
   (i) Your share of the boat launch fee (or daily access fee if not fishing from a boat)

Note that fourteen of the preference statements have five response categories and one has seven response categories.

We assume that the deterministic part of the conditional-indirect utility function for a Green Bay fishing day is a linear function of the average time to catch each species of fish (i.e., the reciprocal of the catch rate), the cost of a trip to Green Bay, and dummy variables for eight of the nine FCA configurations. Each of the FCA configurations specified the level ('Do not eat', 'Eat only once a month', 'Eat only once a week', or 'No advisory') for all of the four species. Level one indicates PCB levels for which there is no fish consumption advisory. Level nine is the most restrictive of the FCA configurations: do not eat for all four species. Level four corresponds to current FCA conditions on Green Bay.

We assume the simple linear specification,

$$\begin{aligned} V_{jk|c} &= \beta_{Fee}(Fee) \\ &\quad + \beta_{Tbass}(Tbass) + \beta_{Tsalmon}(Tsalmon) + \beta_{Tperch|c}(Tperch) + \beta_{Twalleye|c}(Twalleye) \\ &\quad + B_{FCA1|c}(FCA1) + B_{FCA2|c}(FCA2) + ... + B_{FCA9|c}(FCA9), \end{aligned}$$

where Fee is the fee to fish and $FCAx$ is level x of the FCA levels.[6] $Tsalmon$, for example, is the average amount of time to catch a salmon. By assumption, the marginal utility of income, $-\beta_{fee}$, does not vary by class. Preliminary analysis suggested that the effects of catch times for bass and salmon do not vary by class, so they were restricted to be equal across the classes.

A logit specification is assumed for the $P_{jk|c}$ so the probability of choosing alternative A in each choice pair k is:

$$P_{Ak|c} = \frac{e^{V_{Ak|c}}}{e^{V_{Ak|c}} + e^{V_{Bk|c}}}.$$

3.1 Estimated number of classes and model fit

In this section we explain the criteria used to determine the number of latent classes, assess fit, and discuss how well our model fared on these measures. There is no classical statistical test to determine whether increasing the number of classes significantly improves model fit. For example, a likelihood-ratio test is not appropriate because one cannot show that the likelihood-ratio statistic for adding a class has a chi-square distribution - the discrete nature of adding a class violates assumptions

[6] The omitted level (1) is no FCAs.

needed to prove the statistic is chi-square distributed [29]. What then is the appropriate way to determine the number of classes?

In the latent-class choice-only literature, a number of papers have used information criteria to determine the number of classes [14, 16, 18]. These criteria do not provide an absolute measure of fit but do suggest which of two non-nested models provides a superior fit. However, because they are not statistical tests, the analyst is faced with choosing among many possible information criteria, each of which may lead to different conclusions. In the latent-class literature, especially in the areas of education and psychology, the approach has been to find the most parsimonious model with a good fit. In that literature, Pearson and Read-Cressie statistics are often used to examine how well the model fits the data. These statistics compare the expected and actual frequencies of responses [30]. Once a minimum number of classes has been identified that 'fits' the data, information criteria are then used to examine the trade-off between parsimony and fit. However, as has been frequently noted [31, 32], it is problematic to implement these statistics for survey data because of the problem of sparse data (i.e., the number of possible response patterns is large relative to the sample size). In fact, many possible response patterns are never observed, meaning that the chi-squared approximation for the Pearson and Read-Cressie statistics will not be valid.[7] Alternate statistical tests have been proposed for the case of binary response data [34, 35, 36].

Given these issues, we determined the number of classes and assessed the model on the basis of two criteria: (1) information criteria and (2) competing measures of how well a candidate model explains the data. We estimated models with one to four classes and on the basis of these criteria

[7] One potential solution to this implemented in the latent-class literature is bootstrapping the Pearson and Read-Cressie statistics [33]. This is typically not a straightforward procedure and is made even more computationally complex in the case of a joint model. Furthermore, given that the Pearson and Read-Cressie of statistics are not used in latent-choice models, it is unclear whether they should be applied in the case of a joint model.

determined that the preferences of the Green-Bay anglers are adequately characterized with three classes.

Information criteria compare the (ln) likelihood function for c versus $c + 1$ classes to see if the increase in the likelihood is sufficient to warrant the addition of a class. Adding a class increases the likelihood score but also increases the number of parameters, so there is a trade-off. Numerous information criteria have been proposed to determine whether a class should be added [37, 38, 39, 40]. These information criteria are essentially log-likelihood scores with a penalty for sample size and number of parameters. Yang and Yang [32] examine the ability of information criteria, including those with sample size adjustments, to differentiate between latent-class models. Table 1 reports the scores of each of the information criteria examined in Yang and Yang [32]. The three-class model is preferred under all but one criteria (CAIC). In general, the CAIC tends to favor stricter models and imposes a stronger parameter penalty.

The second criteria examined was how well the model explained the data using a number of possible indicators of fit. Since there are two types of data (choice data and preference statements) one can assess the fit in terms of only the choice data, the preference statements, or simultaneously in terms of both. Here we do all three.

The simplest measure of fit is the percentage of responses that the model correctly predicts. A individual's response is defined as correctly predicted if the individual chooses the response category with the highest estimated probability of being chosen. The three-class model correctly predicts 75% of the SP choice pairs answered and 41% of the preference statements. Thus, overall it correctly predicts 53% of all the responses. A random allocation would correctly predict 50% of the choice pairs, 20% of the preference statements, and 30% of all of the responses. There are two possible issues with this measure of fit. First, discrete choice models are probabilistic: one is predicting the percent of time that an alternative would be chosen if the experiment was replicated many times, not which alternative is more likely to be chosen [41]. Secondly, this measure of fit ignores the underlying premise of latent-class modeling that one is estimating a *response pattern* as opposed to individual responses.

Another measure of fit for the choice data is the relationship, by choice pair, between the number of times alternative A was chosen and the model prediction of how many times it would be chosen. Eighty choice pairs were presented (8 pairs in each survey version and 10 versions) and each pair was answered approximately 64 times. For each pair one can compare the percentage of times alternative A was chosen to the estimate of the percentage of times alternative A would be chosen. The correlation between these percentages is 97%, suggesting a very good fit in terms of predicting the responses to the choice questions. The model predicts alternative A will be chosen 2800 times and it was picked $2,829$ times.[8]

To examine fit of the preference statements, we ran Pearson chi-square tests on each of the individual questions, testing the null hypothesis that there is no difference between the number of observed responses for each level and the expected number of responses for each level.[9] In all cases, we are unable to reject the null hypothesis (p-value ≥ 0.99) that the number of expected observations in each response category is significantly different from the number of observed observations. Thus, we conclude that at least for an individual question, there is a good fit. Again, because of the problem of sparse data, we cannot do a chi-square on all of the preference statements simultaneously.

Finally, we examined how precisely the chosen model assigns respondents to classes. In general,

[8]The estimated number is calculated as follows. For each respondent who was asked and answered pair k ($k = 1, 2, ..80$), one calculates the probability that they will chose alternative A. The estimated times A is chosen in pair k is these probabilities summed over all of the respondents who were asked pair k.

[9]For seven of the statements, there were too few observations in the missing category to calculate a valid Pearson chi-square test. Thus, for purposes of the test, we equally apportioned missing observations across all other response categories.

for each respondent, of the three estimated conditional class-membership probabilities, one is much higher than the other three, and often close to one. For example, the maximum of the three conditional class-membership probabilities is 90% or greater for 79% of the sample and 95% or greater for 71% of the sample. Thus, for a given response pattern, individuals are predicted to belong to one particular class with a very probability. This implies that there are notable differences in classes; in the case of classes that did not vary, one would expect probabilities closer to one-third.

3.2 Model results and interpretation

Table 2 reports the estimated β_c. As expected, utility is negative for both fee and the amount of time it takes to catch each fish species. Class 2 cares the most about the time to catch perch while Class 3 cares the least.[10] Class 1 cares a lot about the FCA level; disutility for this class increases as the the level increases. Class 2 also cares about FCA levels but not nearly as much as Class 1. Those in Class 3 generally do not care about the FCA level; except for levels 3 and 6, all of their β_{FCAx} are not significantly different from zero at the five percent level.

For a reduction from current FCA levels (level 4) to no FCAs (level 1), expected WTP is \$22.66 for Class 1, \$11.63 for Class 2, and zero for Class 3.[11] The class-weighted probability average is \$11.00. Summarizing, the estimated $\boldsymbol{\beta}_c$ suggest a class that feels strongly about FCAs (Class 1), a class with no WTP for reducing FCA levels (Class 3), and a class that cares about perch and some about FCAs (Class 2).

[10]Because the average real perch catch time is low (0.75 hours), all perch catch time parameters reported in the paper represent the marginal utility of a change of one-tenth of an hour (i.e., six minutes). Average real catch times for the other three species are multiple hours. So the marginal utilities correspond to a change in catch time of one hour.

[11]This calculation restricts β_{FCA4} to zero, as it is only significantly different from zero at the 10 percent level. For Class 3 WTP is $-\$2.56$ if one uses $\beta_{FCA4} = .13$.

The 74 estimated $\pi_{qs|c}$ for the preference statements tell a similar story but with some interesting differences. These estimates are summarized in two ways. Table 3 reports the predicted average response level, by class, for each of the fifteen preference statements.[12] Figure 2 summarizes the results with a class-difference heat map.

Table 3 shows that the classes vary in the amount that they are bothered by FCAs. Consistent with the estimated preference parameters, Class 1 is most bothered, followed by Class 2 and Class 3. As would be expected, across all groups, individuals care most about the do-not-eat FCAs followed by monthly restrictions. Consistent with these statements, Class 1 is most likely to agree and Class 3 least likely to agree that they would be willing to pay higher boat launch fees if the fish had no PCB contamination. Class 2 attaches more importance to the perch catch times than do the other classes, and they have the highest agreement with the statement that they would be willing to pay more for lower catch times. Class 3 attaches the most importance to the fee. Class 3 is less interested in fishing Green Bay than are the other two classes. They have the lowest average response level on fourteen of the fifteen preference statements, the exception being the importance they attach to the fee.

Figure 2, a class-difference heat map, shows how response-level probabilities for each preference statement vary by class - the darker the color the more the probability of answering at that level differs between the two classes specified. Specifically, for each preference-statement q, the class-difference map shows the difference between the probability that a member of class r chooses response level s and the probability that a member of Class c chooses that level. For example, the top third of the map is the Class 1 probability minus the Class 2 probability. In each third of the heat map, the rows, $1-5$, are the response levels.[13] Each preference statement is a column and the

[12] The estimated $\pi_{qs|c}$ are reported in the appendix in Tables 8 and 9.

[13] For ease of comparison, no-response and the seven-level 'How does Green Bay compare' question are suppressed in the heat map.

columns are sorted by the type of preference statement. Shades of red signal a positive difference while shades of blue signal a negative difference. The darker the cell color, the larger the difference between classes w and r in the estimated probability that they will choose level s. For example, the dark red cell in the fifth row, fifth column indicates that members of Class 1 are much more likely to find the FCA level for walleye to be extremely important than are members of Class 2. The class-difference heat map shows the same FCA ranking indicated by the estimated $\boldsymbol{\beta}_c$: Class 1 attaches the most importance, is most bothered by, and indicates the greatest WTP to eliminate FCAs while Class 3 attaches the least importance, is least bothered by, and indicates the lowest WTP to eliminate FCAs.

There are subtle differences in the stories told by the estimated β_c and the $\pi_{qs|c}$. Consider Class 3. The $\boldsymbol{\beta}$ parameters in Table 2 generally indicate that members of Class 3 are not negatively affected by FCAs. The $\pi_{qs|3}$ indicate that Class 3 is negatively affected, but only by the most stringent FCA levels: "do-not-eat" for all species. One interpretation is that members of Class 3 would prefer a world with no "do-not-eat" FCAs. They have a slight preference for eliminating FCAs but not enough to give up money or other site attributes to achieve it; thus they have no WTP to reduce FCAs.

Based on the responses to the choice pairs and the preference statements, and given that the policy emphasis is on the FCAs, we characterize the three classes as follows: Class 1 is the 'Don't Like FCAs' Class, Class 2 is the 'Perch Important' Class, and Class 3 is the 'Won't Pay To Reduce FCAs' Class.

3.3 Comparison of latent-class joint model with choice-only and preference-statement models

A natural question is how the joint model compares with a model that uses just choice data or just preference-statement data. We compare across the three models on three criteria: estimation, efficiency, and characterization of the classes.[14] In general, the joint model tells the same qualitative story as the choice-only and preference-statement models. However, the joint model provides gains in efficiency and model estimation compared to the choice-only.

3.3.1 Estimation and efficiency

Estimation of the joint model was done with the E-M algorithm. It resulted in smooth convergence and inversion of the estimated Hessian. This was not the case with the choice-only model. The likelihood function for the choice-only model converged to a potential maximum but the Hessian would invert only after $\boldsymbol{\beta}_c$ parameters that were insignificant in the joint model were fixed at zero in the choice-only model. We conjecture that by combining the preference statements with the choice data, the maximum of the likelihood becomes more distinct. In the choice-only model there may be likelihood flatness in the $\boldsymbol{\beta}_c$ and $\Pr(c)$ dimensions; in other words, different combinations of the $\boldsymbol{\beta}_c$ and $\Pr(c)$ values can produce almost equal likelihood values. Adding the preference statement data weakens the multicollinearity link and improves identification for the numerical search process: the $\Pr(c)$ now have to explain class membership for both the choice responses and the preference statements.

Table 4 reports the parameter estimates from the choice-only model. For almost all the pa-

[14]We do not report comparisons with a standard logit model of pair-wise choices, as it is dominated by the three-class choice-only model.

rameters, the t-statistics on the joint model are larger than on the choice-only model (Table 4), illustrating the efficiency gains of the joint model over the choice-only model.

3.3.2 Characterization of classes

Table 5 compares the characterization of the classes from the joint and choice-only models using marginal WTP, as it controls for scale effects. For example, in the joint model, the marginal WTP for Class 1 to reduce FCAs from level 2 to level 1 is $12.08. In the joint model, the marginal WTP for Class 2 to reduce average time to catch a walleye by one hour is $0.75 and the marginal WTP for class 2 to reduce average time to catch a perch by ten minutes is $1.87.

In general, the choice-only model tells the same qualitative story as was reported for the joint model in Section 3.2: Class 1 cares the most about FCAs while Class 2 cares the most about perch catch times. One difference between the choice-only and joint models is the importance of perch catch times for Class 3; while both models identify Class 3's marginal WTP to reduce perch catch times to be the lowest across the classes, the choice-only model identifies it to be negative (and significant at the 5% level). The most notable difference between the choice-only and joint models, however, is in the magnitude of marginal WTP estimates for Classes 1 and 2 across the two models.

Compared to the joint model, the choice-only model results in higher marginal WTP estimates for Class 1 and lower marginal WTP estimates for Class 2. For example, the ratio of estimated WTP in Class 1 versus Class 2 for reducing FCA levels from the current levels (level 4) to no FCAs is two-to-one in the joint model but ten-to-one in the choice-only model. Weighted mean marginal WTP, however, is very similar across the joint and choice-only models.

We compare the preference-statement results in the joint model (Table 3) and preference-statement model (Table 6) by examining predicted average response levels.[15] Again, results are

[15] In comparison to Morey *et al.* [20], Table 6 presents predicted average response level.

generally similar with some small differences.

There is some change in the characterization of the classes based on species-specific catch times and FCAs. In brief, compared to the joint model, the preference-statement model provides higher average responses for Class 3 for the general FCA questions and lower average response levels for questions pertaining to trout and bass. The preference-statement model provides lower estimates for Class 2 for the general FCA questions and questions pertaining to perch. In addition, the differences between Classes 2 and 3 as identified by the joint and preference-statement models are the largest. This tells us that the results from the joint and preference-statement models are differing the most in how Classes 2 and 3 are composed.

We summarize the differences across the models by providing summary characterizations of each of the classes in Table 7. A strong FCA class exists in all three models. An important difference across models is in the characterization of Class 3: while in the choice-only model FCA levels do not matter, in the preference-statement model they matter at the highest levels, just less than in the other classes. The other cross-model distinction is in Class 2. For the choice data (included in the the joint model and the choice-only model), Class 2 responses indicate strong preferences for perch relative to other species. In contrast, in the preference-statement data (included in the joint and preference-statement models), model Class 2 has a strong preference for perch and walleye, both relative to the other species.

Finally, looking at the probability of belonging to each class, some general observations can be drawn. While the characterization of Class 1 is quite consistent across the three models, the probability of belonging to this group is significantly larger for the preference-statement model. Perhaps because Class 2 is identified as a perch and walleye group, individuals who are identified as belonging to the more general perch group are also getting lumped with Class 1. Similarly, because the choice-only model identifies Class 3 as considering lower perch catch rates to be a bad, individuals who like perch but may be less strong in their feelings about FCAs are being assigned to Class 2. This may be lowering the parameter estimates for FCAs in the choice-only model, compared to the joint.

4 Conclusion

Latent-class choice-only models and latent-class preference-statement models are increasingly appearing in the environmental economics literature. What distinguishes preference statements from conventional choice questions is that the individual is not choosing between states of the world but rather choosing a response category that best answers a direct question about their preferences.

The contribution of this paper is the specification and FIML estimation of a latent-class joint model, a model that uses both choice data and statements of preference. Similar in spirit to the practice of combining stated and revealed preference data, this paper provides a method for combining preference-statement and choice data to obtain better estimates of preference parameters and preference heterogeneity. The joint model is based on the premises that preferences are latent and that both choice data and preference statements are manifestations of those unobserved preferences. This model simultaneously estimates the number of latent classes, class-membership probabilities, utility parameters by class, and responses to preference statements by class. We compare the results from the joint model with the choice-only and preference-statement models.

We demonstrate the joint model with an application to angler preferences over the fishing characteristics of Green Bay, a large PCB-contaminated fishing site. Each survey respondent answered eight pair-wise choice questions and fifteen preference statements. The joint model identifies three classes that best explain all of the data: a class whose choices are driven by FCAs levels, regardless of species, a class whose choices are driven by the availability of perch and their FCA level, and a class that only cares about the most severe FCA levels and is unwilling to pay to reduce FCAs. This heterogeneity implies that policies designed to remove PCBs, thus reducing FCAs, will have varying support among recreational anglers.

Preference-statement data has strengths that complement choice data. Survey respondents are generally very familiar with preference statement questions. Likert questions, a common form for preference statements, allow individuals to indicate intensity of preference and nuances in their preferences. Preference-statement data is often discarded, however, as it has been unclear how to incorporate preference-statement data into economic decision-making. Adding preference-statement data to choice data increases the richness of the data. It allows for models that are potentially more nuanced than those obtained from choice data alone.

We identify three primary conclusions from this study. First, while qualitative results are generally very similar across the three models examined, quantitative results for each class differ. In this particular application, including the preference-statement data resulted in less extreme variation among groups. For example, in the joint model, estimated WTP for removing PCBs for the class that cares most about FCAs is twice the WTP for the class that care mostly about perch; this same ratio is ten-to-one in the choice-only model. The weighted mean marginal WTP results, however, are remarkably robust over both the joint and choice-only specifications and match well to more traditional RUM specifications (see Breffle *et al.* [1]). Second, there are efficiency gains from incorporating both types of data. Estimates from a model that uses only choice data are less efficient than those from a model that incorporates both choice and preference-statement data. The t-statistics from the joint model are generally much larger than the t-statistics from the choice-only model. Finally, there are potential estimation gains from incorporating both types of data. The choice-only model had difficulty inverting. Adding the preference-statement data improved identification.

Figure 1: Example choice pair

Figure 5-1
Example Choice Question

If you were going to fish the waters of Green Bay, would you prefer to fish the waters of Green Bay under Alternative A or Alternative B? *Check one box in the last row*

	Alternative A ▽	Alternative B ▽
Yellow Perch		
Average catch rate for a typical angler......................	40 minutes per perch	30 minutes per perch
Fish consumption advisory.......	No more than one meal per week	No more than one meal per week
Trout and Salmon		
Average catch rate for a typical angler......................	2 hours per trout/salmon	2 hours per trout/salmon
Fish consumption advisory.......	Do not eat	No more than one meal per month
Walleye		
Average catch rate for a typical angler......................	8 hours per walleye	4 hours per walleye
Fish consumption advisory.......	Do not eat	No more than one meal per month
Smallmouth bass		
Average catch rate for a typical angler......................	2 hours per bass	2 hours per bass
Fish consumption advisory.......	No more than one meal per month	Unlimited consumption
Your share of the daily launch fee..............................	Free	$3
Check the box for the alternative you prefer..	❑	❑

Table 1: Information Criteria

Classes	Ln L	AIC-3	CAIC	CAIC*	HT-AIC	BIC	BIC*	Parameters
2	-2318	5068	5710	5253	5009	5566	5109	144
3	-2088	4821	5780	5097	4827	5565	4882	215
4	-1984	4825	6101	5193	5008	5815	4907	286

See Yang and Yang [32] for the underlying formulas and the following authors for the original derivation: AIC-3 = Andrews and Currim [42]; Consistent AIC (CAIC) = Bozdogan [38]; CAIC with sample size adjustment (CAIC*)= Yang and Yang [32]; HT-AIC = Hurvich and Tsai [39]; BIC = Schwarz [40]; BIC with sample size adjustment (BIC*)= Yang and Yang [32].

Table 2: Estimated utility parameters (β_c) from joint model

Parameters	**Class 1**		**Class 2**		**Class 3**	
	Est	**t-stat**	**Est**	**t-stat**	**Est**	**t-stat**
β_{fee}	-0.051	-15.80***	-0.051	-15.80***	-0.051	-15.80***
$\beta_{Tsalmon}$	-0.030	-18.19***	-0.030	-18.19***	-0.030	-18.19***
β_{Tbass}	-0.032	-9.33***	-0.032	-9.33***	-0.032	-9.33***
β_{Tperch}	-0.040	-4.19***	-0.096	-12.80***	-0.016	-2.06**
$\beta_{Twalleye}$	-0.034	-4.45***	-0.038	-6.91***	-0.046	-7.47***
β_{FCA2}	-0.62	-4.95***	-0.15	-1.54*	0.074	0.75
β_{FCA3}	-0.64	-4.93***	-0.42	-4.59***	0.20	1.96**
β_{FCA4}	-1.2	-9.14***	-0.59	-6.49***	0.13	1.29*
β_{FCA5}	-1.5	-10.86***	-0.79	-8.23***	-0.069	-0.67
β_{FCA6}	-1.1	-8.73***	-0.65	-7.09***	0.16	1.63**
β_{FCA7}	-1.7	-12.74***	-0.92	-10.12***	-0.12	-1.27*
β_{FCA8}	-2.2	-14.79***	-1.4	-14.43***	-0.081	-0.79
β_{FCA9}	-2.3	-15.6***	-1.6	-15.18***	-0.12	-1.23

Based on logit specification with 640 observations.

***,**, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 3: Joint model: predicted average response level, by class, for preference statements

Amount bothered statements[a]	**Class** 1	**Class** 2	**Class** 3
FCA of eat no more than one meal/week	4.00	2.87	2.32
FCA of eat no more than one meal/month	4.43	3.81	2.73
Do-not-eat FCA	4.79	4.57	3.42
Agreement statements[b]	**Class** 1	**Class** 2	**Class** 3
WTP for higher catch rates	2.86	3.04	2.49
WTP for no PCB contamination	3.76	3.54	2.63
Importance statements[c]	**Class** 1	**Class** 2	**Class** 3
Catch rate: perch	3.53	3.89	3.04
FCA: perch	4.59	3.91	2.63
Catch rate: trout/salmon	3.23	2.60	2.59
FCA: trout/salmon	4.46	2.79	2.40
Catch rate: walleye	3.66	3.56	3.28
FCA: walleye	4.75	3.67	2.77
Catch rate: bass	3.46	2.99	2.92
FCA: bass	4.48	2.67	1.96
Fee	3.04	3.04	3.17
Comparison statement[d]	**Class** 1	**Class** 2	**Class** 3
Green Bay compared to other sites	3.74	3.72	3.88

Calculated by multiplying each predicted response probability to the corresponding Likert level. The full wording of the statements can be found beginning on Page 11.

[a] Scale: 1 =Not at All, ..., 5 = Very Bothersome

[b] Scale: 1 =Strongly Disagree, ..., 5 = Strongly Agree

[c] Scale: 1 =Not at All Important, ..., 5 = Very Important

[d] Scale: 1 =Much Worse, ..., 7 = Much Better

Figure 2: Preference statements: differences in predicted response probabilities by class

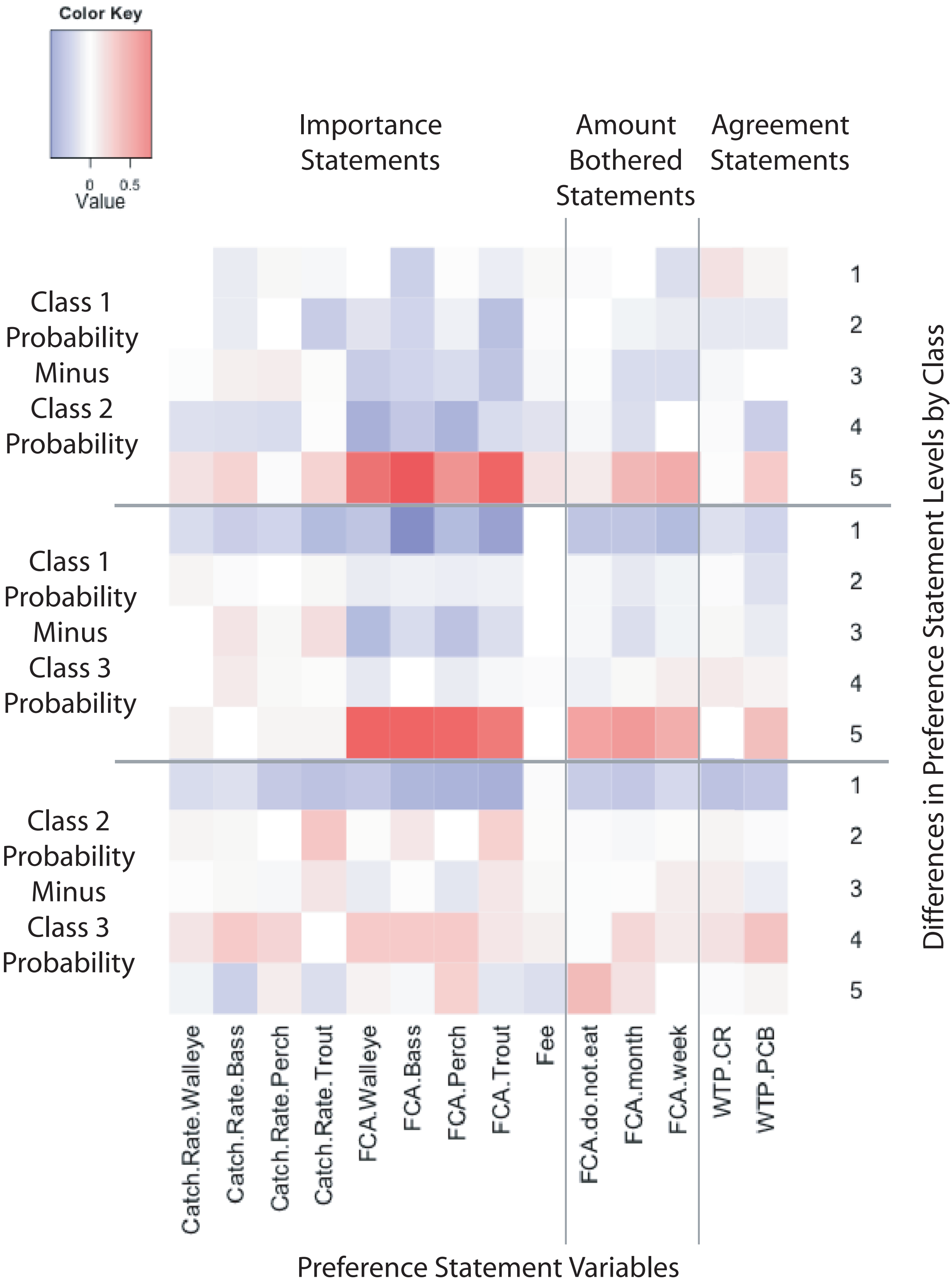

Table 4: Estimated utility parameters (β_c) from the choice-only model

	Class 1		**Class 2**		**Class 3**	
	Est	**t-stat**	**Est**	**t-stat**	**Est**	**t-stat**
β_{fee}	-0.062	-15.21*	-0.062	-15.21*	-0.062	-15.21*
$\beta_{Tsalmon}$	-0.033	-7.92*	-0.033	-7.92*	-0.033	-7.92*
β_{Tbass}	-0.034	-8.69*	-0.034	-8.69*	-0.034	-8.69*
β_{Tperch}	-0.057	-4.24	-0.187	-7.53*	0.024	1.94*
$\beta_{Twalleye}$	-0.038	-3.47*	-0.044	-3.81*	-0.048	-6.33*
β_{FCA2}	-1.01	-4.97	0.0^a	-	0.0^a	-
β_{FCA3}	-1.34	-6.20	-0.195	-1.44*	0.262	2.91
β_{FCA4}	-1.88	-8.03*	-0.191	-1.21*	0.0^a	-
β_{FCA5}	-2.53	-8.91*	-0.269	-1.71 *	0.0^a	-
β_{FCA6}	-2.11	-8.23*	-0.146	-0.92*	0.0^a	-
β_{FCA7}	-3.03	-9.65*	-0.328	-2.15*	0.0^a	-
β_{FCA8}	-3.95	-10.37*	-0.926	-5.29*	0.0^a	-
β_{FCA9}	-4.25	-10.21*	-1.02	-6.07*	0.0^a	-

Based on logit specification with 640 observations.

* indicates that the joint model has a larger t-statistic than this model.

[a] Fixed at zero.

Table 5: Marginal WTP from the joint and choice-only models

	Latent-class joint				**Latent-class choice-only**			
	Class 1	**Class 2**	**Class 3**	**Mean**	**Class 1**	**Class 2**	**Class 3**	**Mean**
Tsalmon	\$ 0.58	\$ 0.58	\$ 0.58	\$ 0.58	\$ 0.53	\$ 0.53	\$ 0.53	\$ 0.53
Tbass	\$ 0.63	\$ 0.63	\$ 0.63	\$ 0.63	\$ 0.55	\$ 0.55	\$ 0.55	\$ 0.55
Tperch	\$ 0.77	\$ 1.87	\$ 0.32	\$ 1.12	\$ 0.92	\$ 3.01	\$ -0.40	\$ 1.73
Twalleye	\$ 0.66	\$ 0.75	\$ 0.9	\$ 0.77	\$ 0.61	\$ 0.71	\$ 0.77	\$ 0.70
FCA2	\$ 12.08	\$ 2.92	\$ -1.45	\$ 3.99	\$ 16.30	\$ 0[a]	\$ 0[a]	\$ 4.40
FCA3	\$ 12.43	\$ 8.15	\$ -3.83	\$ 5.67	\$ 21.77	\$ 3.14	\$ -4.23	\$ 6.58
FCA4	\$ 22.66	\$ 11.63	\$ -2.56	\$ 10.24	\$ 30.39	\$ 3.08	\$ 0[a]	\$ 9.80
FCA5	\$ 28.78	\$ 15.51	\$ 1.34	\$ 14.71	\$ 40.93	\$ 4.33	\$ 0[a]	\$ 13.30
FCA6	\$ 22.27	\$ 12.81	\$ -3.17	\$ 10.47	\$ 34.03	\$ 2.35	\$ 0[a]	\$ 10.41
FCA7	\$ 33.08	\$ 18.02	\$ 2.39	\$ 17.25	\$ 48.79	\$ 5.30	\$ 0[a]	\$ 15.93
FCA8	\$ 43.08	\$ 27.43	\$ 1.58	\$ 23.75	\$ 63.66	\$ 14.94	\$ 0[a]	\$ 24.95
FCA9	\$ 45.24	\$ 31.35	\$ 2.42	\$ 26.28	\$ 68.61	\$ 16.57	\$ 0[a]	\$ 27.14

[a] Underlying parameter estimate fixed at zero to ensure convergence.

Table 6: Preference-statement only model: predicted average response level, by class, for preference statements

Amount bothered statements[a]	**Class** 1	**Class** 2	**Class** 3
FCA of eat no more than one meal/week	3.99	2.69	2.54
FCA of eat no more than one meal/month	4.35	3.51	3.14
Do-not-eat FCA	4.72	4.25	3.85
Agreement statements[b]	**Class** 1	**Class** 2	**Class** 3
WTP for higher catch rates	2.77	2.97	2.58
WTP for no PCB contamination	3.64	3.41	2.75
Importance statements[c]	**Class** 1	**Class** 2	**Class** 3
Catch rate: perch	3.68	3.59	3.25
FCA: perch	4.68	3.61	2.76
Catch rate: trout/salmon	3.19	2.78	2.2
FCA: trout/salmon	4.4	2.9	2.02
Catch rate: walleye	3.83	3.5	3.1
FCA: walleye	4.84	3.54	2.62
Catch rate: bass	3.53	3.13	2.47
FCA: bass	4.5	2.68	1.6
Fee	3.2	3.01	3.07
Comparison statement[d]	**Class** 1	**Class** 2	**Class** 3
Green Bay compared to other sites	3.62	3.91	3.65

Calculated by multiplying each predicted response probability to the corresponding likert level. The full wording of the statements can be found beginning on Page 11.

[a] Scale: 1 =Not at All, ..., 5 = Very Bothersome

[b] Scale: 1 =Strongly Disagree, ..., 5 = Strongly Agree

[c] Scale: 1 =Not at All Important, ..., 5 = Very Important

[d] Scale: 1 =Much Worse, ..., 7 = Much Better

Table 7: Class characterization and membership probability, by model

Class		Joint	Choice-only	Preference-statement
1	Description	Strong FCA	Strong FCA	Strong FCA
	$Pr(c)$	26%	27%	38%
2	Description	Perch	Perch	Perch and Walleye
	$Pr(c)$	44%	52%	30%
3	Description	Won't pay to reduce FCAs; Severe FCAs matter some	FCAs do not matter; perch a bad	Severe FCAs matter some
	$Pr(c)$	30%	22%	32%

References

[1] Breffle W, Morey E, Rowe R, Waldman D, Wytinck S. Recreational fishing damages from Fish Consumption Advisories in the waters of Green Bay. Technical report, Prepared by Stratus Consulting for US Fish and Wildlife Service, 1999.

[2] Thacher J, Morey E, Craighead E. Using patient characteristics and attitudinal data to identify treatment preference groups: A latent-class model. *Depression and Anxiety* 2005; **21**(2): 47–54.

[3] Kritzberg D, Morey E. It's not where you do it, but who you do it with? A companion and their relative ability as characteristics in site-specific recreational demand models, 2008. Working Paper.

[4] McFadden D. The choice theory approach to market research. *Marketing Science* 1986; **5**(4): 275–297.

[5] Ben-Akiva M, Walker M, Bernardino A, Gopinath D, Morikawa T, Polydoropoulos A. Integration of choice and latent variable models. In H Mahmassani, editor, *Perpetual Motion: Travel Behavior Research Opportunities and Application Challenges*. Pergamon, 2002; .

[6] Boxall P, Adamowicz W. Understanding heterogeneous preferences in random utility models: a latent class approach. *Environmental and Resource Economics* 2002; **23**(4): 421–446.

[7] Ben-Akiva M, Morikawa T. Estimation of travel demand models from multiple data sources. In M Koshi, editor, *Transportation and Traffic Theory*. Elsevier, New York, 1990; .

[8] Cameron T. Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods. *Land Economics* 1992; **68**(3): 302–317.

[9] Hensher D, Bradley M. Using stated response choice data to enrich revealed preference discrete choice models. *Marketing Letters* 1993; **4**(2): 139–151.

[10] Adamowicz W, Louviere J, Williams M, of Rural Economy D, of Alberta U. Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities. *Journal of Environmental Economics and Management* 1994; **26**: 271–271.

[11] Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete observations. *Journal of the Royal Statistical Society: Series B* 1977; **39**: 1–38.

[12] Provencher B, Baerenklau K, Bishop R. A finite mixture logit model of recreational angling with serially correlated random utility. *American Journal Of Agricultural Economics* 2002; **84**(4): 1066–1075.

[13] Greene W, Hensher D. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological* 2003; **37**(8): 681–698.

[14] Scarpa R, Thiene M. Destination choice models for rock-climbing in the North-East Alps: a latent-class approach investigating intensity of preferences. *Land Economics* 2005; **81**(3): 426–444.

[15] Scarpa R, Willis K, Acutt M. *Individual-specific welfare measures for public goods: a latent class approach to residential customers of Yorkshire Water*. Edward Elgar Publisher, Aldershot, 2005.

[16] Kemperman A, Timmermans H. Preferences, benefits, and park visits: A latent class segmentation analysis. *Tourism Analysis* 2006; **11**(4): 221–230.

[17] Colombo S, Hanley N. Modelling preference heterogeneity in stated choice data for environ-

mental goods: a comparison of random parameter, covariance heterogeneity and latent class logit models. *EAERE Annual Conference, Thessalonica, Greece* 2007; 27–30.

[18] Patunru A, Braden J, Chattopadhyay S. Who Cares about Environmental Stigmas and Does It Matter? A Latent Segmentation Analysis of Stated Preferences for Real Estate. *American Journal of Agricultural Economics* 2007; 1–15. DOI:10.1111/j.1467-8276.2007.00988.

[19] McCutcheon A. Sexual morality, pro-life values, and attitudes toward abortion - a simultaneous latent structure analysis for 1978-1983. *Sociological Methods and Research* 1987; **16**(2): 256–275.

[20] Morey E, Thacher J, Breffle W. Using angler characteristics and attitudinal data to identify environmental preference classes: A latent-class model. *Environmental and Resource Economics* 2006; **34**(1): 91–115.

[21] Aldrich G, Grimsrud K, Thacher J, Kotchen M. Relating environmental attitudes and contingent values: How robust are methods for identifying heterogeneous groups? *Environmental and Resource Economics* 2006; DOI: 10.1007/s10640-006-9054-7.

[22] Owen A, Videras J. Culture and public goods: The case of religion and the voluntary provision of environmental quality. *Journal of Environmental Economics and Management* 2007; **54**(2): 162–180.

[23] Morey E, Thiene M, De Salvo M, Signorello G. Using attitudinal data to identify latent classes that vary in their preference for landscape preservation. *Forthcoming in Ecological Economics* 2008; .

[24] Swait J, Sweeney J. Perceived value and its impact on choice behavior in a retail setting. *Journal of Retailing and Consumer Services* 2000; **7**(2): 77–88.

[25] GAUSS. *Manual*. Maple Valley, WA: Aptech Systems Inc, 2000.

[26] Vermunt J, Magidson J. *Latent GOLD Choice*. Statistical Innovations, 2003.

[27] R Development Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2005. Http://www.R-project.org.

[28] Vermunt J, Magidson J. *Latent GOLD*. Statistical Innovations, 2005.

[29] Wedel M, Kamakura W. *Market Segmentation: Conceptual and Methodological Foundations*. Boston: Kluwer Academic Publishers, second edition, 2000.

[30] Formann A. Latent class model diagnostics - a review and some proposals. *Computational Statistics and Data Analysis* 2003; **41**: 549–559.

[31] Eid M, Langeheine R, Diener E. Comparing typological structures across cultures by multigroup latent class analysis - a primer. *Journal of Cross-Cultural Psychology* 2003; **34**(2): 195–210.

[32] Yang CC, Yang CC. Separating latent classes by information criteria. *Journal of Classification* 2007; **24**: 183–203.

[33] Langeheine R, Pannekoek J, van der Pol F. Bootstrapping goodness-of-fit measures in categorical data analysis. *Sociological Methods and Research* 1996; **24**: 492–516.

[34] Reiser M, Lin Y. A goodness of fit test for the latent class model when expected frequencies are small. *Sociological Methodology* 1999; 81–111.

[35] Bartholomew D, Leung S. A goodness of fit test for sparse 2p contingency tables. *British Journal of Mathematical and Statistical Psychology* 2002; **55**(1): 1–15.

[36] Maydeu-Olivares A, Joe H. Limited-and Full-Information Estimation and Goodness-of-Fit Testing in [2. Sup. N] Contingency Tables: A Unified Framework. *Journal of the American Statistical Association* 2005; **100**(471): 1009–1021.

[37] Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974; **19**: 716–723.

[38] Bozdogan H. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika* 1987; **52**: 345–370.

[39] Hurvich M, Tsai C. Regression and time series model selection in small samples. *Biometrika* 1989; **76**(2): 297–307.

[40] Schwarz G. Estimating the dimension of a model. *The Annals of Statistics* 1978; **6**: 461–464.

[41] Train K. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press, first edition, 2003.

[42] Andrews R, Currim I. A Comparison of Segment Retention Criteria for Finite Mixture Logit Models. *Journal of Marketing Research* 2003; **40**(2): 235–243.

A Appendix

Table 8: Estimated Response Probabilities by Class

Comparison statement: quality of Green Bay fishing compared to other places[a]

	1	**2**	**3**	**4**	**5**	**6**	**7**
Class 1	2%	13%	19%	24%	21%	9%	4%
Class 2	2%	10%	22%	31%	18%	11%	1%
Class 3	4%	7%	12%	28%	23%	12%	6%

Numbers do not sum to 100% as no response is not reported.

[a] 1 = Much Worse ,...., 7 = Much Better, 8 = No response

Table 9: Estimated Response Probabilities by Class (continued)

	Class 1					**Class 2**					**Class 3**				
	1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
Amount bothered statements[a]															
FCA of eat no more than one meal a week	10%	8%	11%	15%	56%	23%	16%	25%	16%	18%	39%	13%	16%	6%	19%
FCA of eat no more than one meal a month	4%	1%	8%	15%	70%	5%	6%	23%	28%	37%	28%	10%	21%	11%	24%
Do-not-eat FCA	0%	0%	2%	1%	94%	3%	1%	4%	5%	84%	24%	4%	6%	7%	53%
Agreement statements[b]															
WTP for higher catch rates	27%	7%	31%	20%	14%	14%	16%	35%	23%	12%	39%	10%	26%	10%	15%
WTP for no PCB contamination	16%	1%	21%	14%	47%	10%	10%	22%	34%	24%	33%	13%	29%	7%	18%
Importance statements[c]															
Catch rate: perch	8%	13%	28%	13%	36%	3%	12%	19%	27%	39%	25%	13%	23%	9%	30%
FCA: perch	4%	3%	4%	4%	85%	2%	9%	19%	35%	35%	33%	10%	29%	12%	15%
Catch rate: trout/salmon	13%	11%	34%	16%	25%	17%	32%	31%	14%	6%	42%	6%	19%	13%	19%
FCA: trout/salmon	7%	1%	5%	9%	77%	14%	27%	29%	24%	5%	47%	7%	18%	13%	15%
Catch rate: walleye	3%	13%	30%	19%	34%	2%	13%	32%	31%	21%	17%	7%	30%	19%	26%
FCA: walleye	1%	2%	3%	5%	88%	2%	13%	24%	38%	23%	25%	10%	32%	14%	16%
Catch rate: bass	6%	12%	33%	19%	28%	14%	20%	25%	32%	9%	26%	15%	21%	9%	28%
FCA: bass	3%	5%	7%	5%	79%	22%	22%	24%	28%	3%	52%	11%	22%	5%	7%
Fee	19%	14%	28%	10%	27%	15%	17%	32%	21%	14%	18%	14%	28%	13%	27%

Numbers do not sum to 100% as no response is not reported.

[a] 1=Not at all bothersome,..., 5=Very Bothersome

[b] 1=Strongly Disagree,..., 5=Strongly Agree

[c] 1=Not at All Important, ..., 5= Very Important