Combining attitudinal and choice data to improve estimates of preferences and preference heterogeneity: a FIML, discrete-choice, latent-class model

William Breffle

Stratus Consulting Inc., Boulder CO.

Edward Morey *

Department of Economics, University of Colorado-Boulder

Jennifer Thacher †

Department of Economics, University of New Mexico

December 2, 2005

Abstract

This paper shows how attitudinal data can be combined with choice data to more efficiently estimate preference parameters and preference heterogeneity. Most surveys

*All are equal authors.

†Corresponding author: jthacher@unm.edu; Department of Economics, University of New Mexico, MSC05 3060, Albuquerque NM 87131.

collect both types of data, but the majority of econometric models of preferences rely solely on choice data. Two types of data, answers to choice questions and likert-scale attitudinal questions, are used to simultaneously estimate (1) the probability that an individual belongs to a particular preference class, (2) the parameters in each classes' conditional, indirect-utility function, and (3) for each attitudinal question, the probability that an individual in a particular class will give a particular response. Estimation is with the expectation-maximization (E-M) algorithm. FIML (full-information maximum likelihood) estimates are obtained by finding those values of the parameters in the model that maximize the likelihood of simultaneously observing both the attitudinal and choice data. The parameter estimates from FIML estimation are compared with those from sequential estimation, which are not asymptotically efficient.

Key words: Latent class, attitudinal data, choice data, FIML, preference heterogeneity

JEL code C51,Q51,D12

Surveys often include a significant number of attitudinal questions. Attitudinal questions often assess the relative importance the individual places on different attributes of a good and indicate how the individual ”feels” about those attributes. In other words, attitudinal questions can reveal how much an individual likes or dislikes a particular attribute.[1]

Consider an example Likert-scale attitudinal question from a survey of Green Bay anglers:

On a scale from 1 to 5 where 1 means ”Not at all Bothersome” and 5 means

[1]Attitudinal questions differ from questions that ask the individual to indicate his or her perceived level of an attribute.

"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 advisory: "Do not eat".

Or, from a survey of depressed individuals about the possible side effects of treatment alternatives:

> How much would little or no interest in sex bother you? ("Not at all, slightly, some, a fair bit, a lot")

While many economists do not view answers to attitudinal questions as data one uses in an econometric model to estimate preferences, we believe that attitudinal data can provide significant information about the existence of different preference classes and how preferences vary across those classes.[2] Here we assume that the answers to attitudinal questions are expressions of exogenous well-behaved preferences: individuals can rank states of the world. Preferences are latent (unobservable), and both choices (actual and hypothetical) and answers to attitudinal questions are manifestations of those unobserved preferences. Given these assumptions, it would seem derelict to estimate preferences and preference heterogeneity without using attitudinal data, if it is available. Including both attitudinal and chocie data in estimation results in more efficient estimation.

The intent of this paper is to identify and estimate preference heterogeneity for environmental amenities in terms of a small number of preference classes. Class membership and the preferences of each class are latent. What is observed are the choices made, the attributes

[2]Economists who have viewed answers to attitudinal answers as data include Ben-Akiva et al. (2002), McFadden (1986), and Boxall and Adamowicz (2002).

of the alternatives in the choice sets, and the answers to attitudinal questions about those attributes.

Using the E-M (expectation-maximization) algorithm (Dempster et al., 1977), we implement full-information maximum likelihood (FIML) estimation to find those values of the parameters in the model that maximize the likelihood of observing both the attitudinal and choice data. Deriving this joint likelihood function, developing an algorithm to find the values of the parameters that maximize it, implementing that algorithm, and comparing the FIML results to those obtained from sequential estimation are the main accomplishments of this paper.

Section 1 provides a brief background on latent-class models. Section 2 presents a model with both attitudinal and choice data and explains two alternate methods for estimating this model: FIML and seqential estimation. In section 3, we show how this model can be implemented by applying it to fishing preferences.

1 Background: the LC_A, LC_C, and LC_{AC} models

The raison d'être of latent-class models is to model preference heterogeneity among discrete groups without assuming some observable deterministic explanation for that heterogeneity. We build on previous work by us and others in this area. Morey et al. (2005), ignoring the available choice data, estimate a latent-class model of preferences using only the answers to attitudinal questions - a LC_A model.[3] Examples of LC_A models outside of economics include Clogg and Goodman (1984), McCutcheon (1987), McCutcheon and Nawojcyzk (1995), De Menezes and Bartholomew (1996), Yamaguchi (2000).

While an LC_A model can provide a lot of information about preferences, many economists equate estimating preferences with estimating the values of preference parameters in utility functions. In a latent class context, once can estimate an LC_C model, a discrete choice, latent-class model estimated with only choice data. One assumes some number of classes and specifies a conditional, indirect-utility function that allows the preferences parameters to vary by class. 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 attitudinal data is used. Examples of LC_C models include: Gupta and Chintagunta (1994), Kamakura and Russell (1989), Greene and Hensher (2002), Provencher et al. (2002), Hu et al. (2004), and Scarpa and Thiene (2005).

In this paper, we estimate a combined LC_A model and LC_C model: the LC_{AC} model. We estimate both the probability of class membership and the parameters in an indirect utility function using both the choice and attitudinal data.[4] We would not be surprised if others have already estimated a LC_{AC} model, but know of no example.[5]

[3]LC denotes "latent-class" and the subscript(s) denote what type or types of data are used to estimate the model.

[4]Also related to, but different from the LC_A, LC_C, and LC_{AC} models are discrete-choice models where some of the attributes of alternatives in the choice sets are latent variables. Ben-Akiva et al. (2002) provides empirical examples. These models deal with varying perceptions with respect to attribute levels; our LC_{CA} model does not. Most of these models are not latent-class models.

[5]Our LC_{AC} model is similar in appearance to but fundamentally different from the models of Boxall and Adamowicz (2002) and Swait (1994). Boxall and Adamowicz (2002) assumes class membership is a

2 A latent-class model of choice and attitudinal data: the LC_{AC} model

Assume the population consists of C different preference classes. An individual's preference class is latent. The researcher observes, for each individual, the data $(\mathbf{x}_i, \mathbf{y}_i)$; $\mathbf{x}_i$ is the set of individual i's answers to the attitudinal questions (the individual's attitudinal response pattern) and $\mathbf{y}_i$ represents individual i's answers to a set of stated preference (SP) choice pairs. An example of a SP question is included in the appendix.

If one observes $\mathbf{x}_i$, $\mathbf{y}_i$, and class membership, the likelihood function for the sample can be written as:

$$L = \prod_{i}^{N} Pr(\mathbf{x}_i, \mathbf{y}_i, c_i). \tag{1}$$

But, since class membership is unobserved, the best one can do is to model:

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

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 attitudinal and stated preference responses, conditional on belonging to class c.[6]

function of latent psychological variables, and that the answers to the attitudinal questions are indicators for these unobserved psychological variables. Put simply, Boxall and Adamowicz (2002) make the probability of class membership a function of the answers to attitudinal questions, whereas the LC_{AC} model assumes class membership is exogenous and determines how one will answer attitudinal questions.

[6] One could also model the probability of belonging to a class as a function of a number of observable characteristics of the respondents, such as gender and age.

Because 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. Latent-class models assume that once one has conditioned on class, an individual's answers to all of the stated-choice and attitudinal questions are independent of one another. Accepting this assumption, the likelihood function can then be written:

$$L = \prod_i \left[\sum_{c=1}^{C} \Pr(c) \Pr(\mathbf{x}_i|c) \Pr(\mathbf{y}_i|c) \right], \tag{3}$$

where

$$\Pr(\mathbf{x}_i|c) = \prod_{q=1}^{Q} \prod_{s=1}^{S} (\pi_{qs|c})^{x_{iqs}} \tag{4}$$

and

$$\Pr(\mathbf{y}_i|c) = \prod_{k=1}^{K} \prod_{j=1}^{J} (P_{jk|c})^{y_{ijk}}. \tag{5}$$

$\pi_{qs|c}$ is the probability that an individual in class c answers level s to attitudinal question q; $x_{iqs} = 1$ if individual i's answer to attitudinal question q is level s and 0 otherwise. $P_{jk|c}$ is the probability of choosing alternative j in SP-choice pair k, conditional on being a member of class c; $y_{ijk} = 1$ if individual i choose alternative j in choice pair k and 0 otherwise.

$P_{jk|c}$ are functions of the parameters in the class-specific conditional-indirect utility functions, the β_c parameters. That is, $P_{jk|c}$ is the probit or logit probability of choosing alternative j from SP-choice set k, conditional on being a member of class c and takes the typical form. For example, if a logit model is assumed, the probability of choosing alternative j is,

$$P_{jk|c} = \frac{\exp(\beta_c' \mathbf{z}_{jk})}{\sum_{j=1}^{J} \exp(\beta_c' \mathbf{z}_{jk})} \quad c = 1, 2, ..., C, \tag{6}$$

where $\mathbf{z}_{jk}$ is the vector of characteristics of the good in alternative j of choice-pair k.

The goal of estimation is to find the β_c, the $\pi_{qs|c}$, and the $\Pr(c)$ that maximize Equation 2. Two conditional class membership probabilities will be useful for estimating these parameters. The first is the probability that an individual is a member of class c conditional on her answers to the attitudinal questions. By Bayes Theorem, this probability is:

$$\Pr(c\,|\mathbf{x}_i\,) = \frac{\Pr(c)\prod_{q=1}^{Q}\prod_{s=1}^{S}(\pi_{qs|c}\,)^{x_{iqs}}}{\Pr(\mathbf{x}_i)}, \tag{7}$$

where

$$\Pr(\mathbf{x}_i) = \sum_{c=1}^{C}\Pr\,(c)\,\Pr(\mathbf{x}_i|c) = \sum_{c=1}^{C}\Pr(c)\prod_{q=1}^{Q}\prod_{s=1}^{S}(\pi_{qs|c})^{x_{iqs}}. \tag{8}$$

The second useful probability, the probability that an individual is a member of class c conditional on **both** her answers to the attitudinal questions and her answers to the SP choice questions, is

$$\Pr(c|\mathbf{x}_i,\mathbf{y}_i) = \frac{\Pr(c)\prod_{q=1}^{Q}\prod_{s=1}^{S}(\pi_{qs|c}\,)^{x_{iqs}}\prod_{k=1}^{K}\prod_{j=1}^{J}P_{jkc}^{y_{ijk}}}{\Pr(\mathbf{x}_i,\mathbf{y}_i)}. \tag{9}$$

where $\Pr(\mathbf{x}_i,\mathbf{y}_i)$ is individual i's contribution to the likelihood function (the bracketed term in Equation 3).

We next lay out two alternative methods for estimating the LC_{AC} model: FIML and sequential estimation. The FIML estimates are consistent and asymtotically efficient; the sequential estimates are only consistent.

With both types of estimation, we utilize a variant of the E-M algorithm, a technique to do maximum-likelihood estimation with incomplete information (Dempster et al. (1977), Arcidiacono and Jones (2003)). The missing pieces of information in the LC_{AC} model is class membership and the preference parameters for each class.

Put simply, the E-M algorithm replaces unobserved information with its expected value and then conducts maximum likelihood estimation as if these expectations were correct. The maximum likelihood estimates can be then used to update the original expectations. The E-M algorithm consists of two steps: an expectation step and a maximization step. In the expectation step, one calculates the expected value of the unobserved information. In the maximization step, one conducts maximum likelihood estimation as if the true value of the unobserved information was the expected value of the unobserved information. Based on the results of the maximization step, one then updates the expected value of the unobserved information. The process continues until the change in the log-likelihood function becomes very small.

2.1 FIML estimation of the LC$_{AC}$ model

In this section, we describe in more detail how one can use the E-M algorithm to do FIML estimation. Estimates obtained from this approach will be efficient as they use all of the data.

1. Guess or estimate the $N \times C$ initial values of $\Pr(c\,|\mathbf{x}_i, \mathbf{y}_i)$, denoted $\Pr(c\,|\mathbf{x}_i, \mathbf{y}_i)^{\{0\}}$ where $^{\{d\}}$ refers to iteration d. These initial guesses at the conditional probabilities, Equation 9, could be from a sequential estimation of the parameters in the model. Sequential estimation is discussed below.

2. Use the $\Pr(c\,|\mathbf{x}_i, \mathbf{y}_i)^{\{0\}}$ to calculate the unconditional membership probabilities, $\Pr(c)^{\{1\}}$. They are obtained by maximizing the likelihood function and solving the first order condi-

tions:

$$\Pr(c) = \frac{1}{N}\sum_{i=1}^{N}\Pr(c\,|\mathbf{x}_i, \mathbf{y}_i)\,. \tag{10}$$

Equation 10 is simply the average of the conditional class-membership probabilities for all individuals in class c. So, at this point one has calculated $\Pr(c_i)^{\{1\}}$ for each respondent in the sample, conditional on the $\Pr(c_i\,|\mathbf{x}_i, \mathbf{y}_i)^{\{0\}}$.

3. Then use the $\Pr(c)^{\{1\}}$, the $\Pr(c_i\,|\mathbf{x}_i, \mathbf{y}_i)^{\{0\}}$, and the attitudinal data to calculate the $\pi_{qs|c}^{\{1\}}$. The formula, obtained by maximizing the likelihood function and solving the first order conditions, is:

$$\pi_{qs|c} = \frac{\sum_{i=1}^{N}\Pr(c_i\,|\mathbf{x}_i, \mathbf{y}_i)\,x_{iqs}}{\Pr(c)N}. \tag{11}$$

The denominator in Equation 11 is an estimate of the number of people in class c. The numerator, $\sum_{i=1}^{N}\Pr(c\,|\mathbf{x}_i, \mathbf{y}_i)x_{iqs}$, is the number of times individuals in the sample answered level s to question q, each weighted by the conditional probability that the individual is in c. That is, the numerator is an estimate of the number of times individuals in class c answer level s to question q. The ratio is therefore an estimate of the proportion of times individuals in class c answer level s to question q.

4. Now use the $(\pi_{qs|c})^{\{1\}}$ and Equation 4 to calculate the $\Pr(\mathbf{x}_i|c)^{\{1\}}$.

Summarizing to here, based on our initial "guesses" for the $\Pr(c\,|\mathbf{x}_i, \mathbf{y}_i)$ and the data, we have come up with the estimates of the $\Pr(c)^{\{1\}}$ and the $\Pr(\mathbf{x}_i|c)^{\{1\}}$. Steps $2-3$ are an application of the E-M algorithm. One is finding the values of the $\Pr(c)$ and the $\pi_{qs|c}$ 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, the $\Pr(c\,|\mathbf{x}_i, \mathbf{y}_i)^{\{d\}}$, as if they were the true values.

5. Plugging in the $\Pr(c)$ and $\pi_{qs|c}$, the likelihood function, conditional on these estimates, is:

$$L_r^{\{1\}} = \prod_i \left[\sum_{c=1}^{C} \Pr(c)^{\{1\}} \Pr(\mathbf{x}_i|c)^{\{1\}} \Pr(\mathbf{y}_i|c) \right] \tag{12}$$
$$= \prod_i \left[\sum_{c=1}^{C} \Pr(c)^{\{1\}} \prod_{q=1}^{Q} \prod_{s=1}^{S} (\pi_{qs|c}^{\{1\}})^{x_{iqs}} \prod_{k=1}^{K} \prod_{j=1}^{J} P_{jk|c}^{y_{ijk}} \right].$$

Use a maximization algorithm (such as Optimum or Maxlik in Gauss) to maximize $\ln L_r$ in terms of the β_c. Denote these parameter estimates $\beta_c^{\{1\}}$. The subscript r indicates that the likelihood function is conditioned/restricted.

6. Now plug the $\beta_c^{\{1\}}$, the $\Pr(c)^{\{1\}}$, the $(\pi_{qs|c}^{\{1\}})$, along with the attitudinal and choice data into Equation 9 to calculate $\Pr(c\,|\mathbf{x}_i, \mathbf{y}_i)^{\{1\}}$. $\Pr(c\,|\mathbf{x}_i, \mathbf{y}_i)^{\{1\}}$ is an expected value. This is the end of iteration 1.

Return to step 1 but start with new best estimate of the $\Pr(c\,|\mathbf{x}_i, \mathbf{y}_i)$, the $\Pr(c\,|\mathbf{x}_i, \mathbf{y}_i)^{\{1\}}$. Continue iterating until the conditional likelihood function, Equation 12, increases by less than some predetermined amount.

2.2 Sequential estimation of the LC$_{AC}$ model

Sequential estimation can be viewed as an alternative to FIML estimation or a way to get good initial estimates of the conditional class membership probabilities, the $\Pr(c\,|\mathbf{x}_i, \mathbf{y}_i)$, for the start of FIML estimation.

Sequential estimation, as defined here, first obtains maximum likelihood estimates of the $\Pr(c)$ and the $\pi_{qs|c}$ using only the attitudinal data. These estimates are not as efficient as the FIML estimates because not all of the information/data is used in their estimation. Denote these estimates $\Pr(c)^{s1a}$ and $\pi_{qs|c}^{s1a}$ where the superscript "$s1a$" denotes stage 1 sequential estimates based solely on the attitudinal data. These can then be plugged into Equation 7 to obtain stage 1 maximum likelihood estimates of the conditional class-membership probabilities, $\Pr(c\,|\mathbf{x}_i\,)$. Denote these $\Pr(c\,|\mathbf{x}_i\,)^{s1a}$.

At stage 2 one obtains the maximum likelihood estimates of the β_c, taking as given the $\Pr(c\,|\mathbf{x}_i\,)^{s1a}$, and using only the SP-choice data. At the end of sequential estimation one has estimates of all of the parameters in the joint model (the $\Pr(c)$, $\pi_{qs|c}$, and β_c) but these estimate are only consistent. They are not asymptotically efficient because none were estimated using all of the data. Put simply, they are not the parameter estimates that maximize the joint likelihood function, Equation 3.

In more detail, consider first the likelihood function for the attitudinal data:

$$L_a(\Pr(c), \pi_{qs|c}) = \prod_i \left[\sum_{c=1}^{C} \Pr(c) \prod_{q=1}^{Q} \prod_{s=1}^{S} (\pi_{qs|c})^{x_{iqs}} \right]. \tag{13}$$

This likelihood function is developed, estimated, and explained in Morey et al. (2005). One maximizes Equation 13 to obtain the $\Pr(c)^{s1a}$ and $\pi^{s1a}_{qs|c}$. These estimates are then used to calculate conditional class-membership probabilities using Equation 7. Note these conditional probabilities are conditional on only the attitudinal data.

The likelihood function for the β_c parameters, taking as given the $\Pr(c\,|\mathbf{x}_i\,)^{s1a}$, and using only the SP-choice data, is

$$L_{sp}(\beta_c) = \prod_i \left[\sum_{c=1}^{C} \Pr(c\,|\mathbf{x}_i\,)^{s1a} \prod_{k=1}^{K} \prod_{j=1}^{J} P_{jk|c}^{y_{ijk}} \right] \tag{14}$$

That is, each individual's probability of choosing alternative j in choice-pair k, conditional on being a member of class c, is weighted by the stage 1 best estimate of the probability that the individual is in class c. The estimated stage 2 estimates of the β_c, β_c^{s2sp} are obtained by using Gauss to maximize $\ln(L_{sp}(\beta_c))$.

Note that one can use the sequentially estimated $\Pr(c)^{s1}$, $\pi^{s1}_{qs|c}$, and β^{s2}_{c} along with all of the data in Equation 9 to obtain initial estimated values for the conditional membership probabilities, $\Pr(c\,|\mathbf{x}_i, \mathbf{y}_i\,)$. Denote these $\Pr(c\,|\mathbf{x}_i, \mathbf{y}_i\,)^{\{0\}}$ because they can be used to as initial estimates for $\Pr(c\,|\mathbf{x}_i, \mathbf{y}_i\,)$ in the first step in the first iteration of FIML estimation.[7]

3 Application: preferences of Green Bay anglers

To show how the LC_{AC} model can be implemented in practice, we apply the model to estimate preferences over the fishing characteristics of Green Bay, a large bay on Lake Michigan that is contaminated by PCBs. The goal is 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, bass), and fish consumption advisory (FCA) levels. Anglers answered 15 likert-scale attitudinal questions and eight SP questions of the type: Would you rather fish Green Bay under conditions A or B? The attitudinal questions and an example choice question are included in the appendix. 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.

We first estimate the model using sequential estimation and then compare these results to those obtained using FIML.

[7] The sequential model could also be estimated in the opposite order with the indirect utility parameters estimated first.

3.1 Sequential estimation

3.1.1 Stage one

We start with sequential estimation because we use the sequential estimates to calculate starting values for the FIML estimation. To keep things simple, we assume only two classes; therefore, $\Pr(class\ 2) = 1 - \Pr(class\ 1)$. The first stage estimates of the $\Pr(1)^{s1a}$ and $\pi_{qs|c}^{s1a}$ are obtained by maximizing Equation 13.

Table 1: Average Response to Attitudinal Questions by Class: Sequential vs FIML

	Sequential		FIML	
	FCA	**Perch/Walleye**	**FCA**	**Perch/Walleye**
Attribute Importance (5=Very Important)				
Catch:Bass	3.42	2.96	3.36	2.92
Catch:Perch	3.59	3.52	3.64	3.48
Catch:Trout/Salmon	3.17	2.59	3.13	2.52
Catch:Walleye	3.75	3.40	3.75	3.34
FCA:Bass	4.33	2.35	4.05	2.19
FCA:Perch	4.58	3.35	4.47	3.20
FCA:Trout/Salmon	4.34	2.60	4.08	2.46
FCA:Walleye	4.74	3.25	4.56	3.10
Fee	3.13	3.05	3.05	3.09
Amount Bothered (5=Very Bothersome)				
FCA: 1/week	3.94	2.61	3.81	2.46
FCA: 1/month	4.34	3.37	4.34	3.19
FCA: Don't eat	4.68	4.13	4.72	4.00
Agreement (5=Strongly Agree)				
WTP Higher Fees: Higher Catch	2.84	2.83	2.92	2.78
WTP Higher Fees: No PCBs	3.69	3.17	3.75	3.04
Comparison to Other Sites (7=Green Bay Much Better)				
Green Bay Quality	3.63	3.83	3.68	3.84

Columns 1 and 2 in Table 1 reports the average responses to the attitudinal questions for the anglers most likely to belong to each class. These average responses indicate an FCA class and a Perch/Walleye class. Table 1 shows that those in the FCA class are "more bothered" by FCA levels than those in the Perch/Walleye class. In addition, those in the FCA class stated that the FCA levels for all species were the most important factors in their choices; those in the Perch/Walleye class stated that the most important factors for them were the catch and FCA levels for Perch and Walleye. The probability of class membership for the FCA class using only the attitudinal data, $\Pr(1)^{s1a}$, is 0.296; that is, 29.6% of Green Bay anglers are predicted to be in the FCA class. This estimate is imposed as an assumed fixed value at the second stage of sequential estimation.

The 124 estimated response probabilities from stage 1 ($\pi_{qs|c}^{s1a}$) can be combined with the unconditional class membership probabilities and each individual's attitudinal data to estimate each angler's class membership probability conditional on his answers to the attitudinal questions, the $\Pr(1\,|\mathbf{x}_i)$ (Equation 7).[8] Most of these conditional estimates of membership put each individual into one of the classes with high probability; the maximum of the probabilities for the two classes is 90% or greater for 89% of the sample and effectively 100% for 69% of the sample.

3.1.2 Stage two

To estimate the second stage and obtain estimates for parameters in the indirect utility function, we assumed that the functional form of the deterministic part of the conditional-

[8]For a Q level likert scale question, if $Q-1$ levels are estimated, the last level is implicitly known. Therefore, since there are 14 questions with four estimated levels and one question with six estimated levels, and two classes, there are 124 response probabilities total.

indirect utility function for a Green Bay fishing day is a linear function of the catch times for the different species, FCA levels, and the cost of a trip to Green Bay. Cost simplifies to only the launch fee because travel cost, for an angler, is always the same constant. In the Green Bay SP-choice pairs there were nine possible configurations of FCA levels. Each specified the level ("do not eat", "once a month", "once a week", no advisory") for each of the four species. Level one indicates PCB levels for which there is no health risk from consumption. Level nine is the most restrictive. Level four corresponds to current conditions on Green Bay. FCAs were considered the important policy variables and thus were allowed to vary among the classes. The perch and walleye species are also more important from a policy perspective and thus catch rates for these two species were allowed to vary between classes; catch parameters on bass and salmon were assumed to not be different. We assumed a logistic probability.

The second-stage estimates of the indirect utility parameter estimates, β, are reported in columns 2 and 6 in Table 2. These estimates are consistent with the first stage results; this provides evidence, and supports our assumption, that both the choice and attitudinal responses are manifestations of the same underlying preferences. In explanation, those in the Perch/Walleye class care more about the Perch and Walleye catch rates than do those in the FCA class; they get more disutility from increased catch-times. And, at every FCA level, the FCA class is more concerned about that FCA level than is the Perch/Walleye class.

At this point, we could stop: we have consistent estimates of all of the parameters and they could be used, for example, to obtain expected compensating variations for changes in FCA levels. However these estimates are not asymptotically efficient: all of the information/data was not used to simultaneously estimate all of the parameters.

One can also use the sequential estimates to calculate for each angler the probability of being in class 1 conditional on answers to both the attitudinal and choice data ($\Pr(1\,|\mathbf{x}_i, \mathbf{y}_i\,)$) using Equation 9. For 411 of the anglers, $\Pr(1\,|\mathbf{x}_i, \mathbf{y}_i\,)$ puts them in one of the classes with at least 90% certainty - a high degree of separation. It is of interest to compare the membership probabilities conditional on full information with those based on only the attitudinal data. Summarizing, for 629 anglers (98% of the sample), both of the conditional probabilities predict the same class (181 in the PCB class and 448 in the Perch/Walleye class). For these 629 anglers, the probability of class membership increases for 276 anglers when all of the data and preference parameters are used to estimate class membership. Many of those that did not improve had $\Pr(1\,|\mathbf{x}_i\,)$ of effectively zero or one, so there was no room for improvement. There are no examples where $\Pr(1\,|\mathbf{x}_i\,)$ predicted class membership with high certainty and $\Pr(1\,|\mathbf{x}_i, \mathbf{y}_i\,)$ predicted membership in the other class with high certainty.

Table 2: Sequential and FIML Parameter Estimates (Est/SE)

Parameters	Sequential	FIML			Sequential	FIML		
		Iter 1	Iter 10	Converged		Iter 1	Iter 10	Converged
	FCA Class				**Perch/Walleye Class**			
Pr(class)	0.296	0.301	0.346	0.404				
Perch	−0.291	−0.311	−0.363	−0.378	−0.630	−0.632	−0.623	−0.623
	(−3.245)	(−3.646)	(−4.459)	(−5.029)	(−11.805)	(−11.883)	(−11.356)	(−10.884)
Walleye	−0.032	−0.033	−0.032	−0.031	−0.038	−0.038	−0.039	−0.04
	(−4.466)	(−4.785)	(−4.849)	(−5.261)	(−9.516)	(−9.449)	(−9.424)	(−9.401)
FCA2	−0.523	−0.513	−0.494	−0.484	−0.060	−0.057	−0.055	−0.03
	(−4.395)	(−4.540)	(−4.605)	(−4.802)	(−0.874)	(−0.841)	(−0.778)	(−0.404)
FCA3	−0.603	−0.572	−0.603	−0.602	−0.143	−0.141	−0.111	−0.084
	(−4.947)	(−4.970)	(−5.552)	(−6.008)	(−2.159)	(−2.125)	(−1.605)	(−1.168)
FCA4	−1.046	−1.02	−1.038	−1.051	−0.283	−0.281	−0.247	−0.178
	(−8.578)	(−8.972)	(−9.636)	(−10.448)	(−4.251)	(−4.209)	(−3.578)	(−2.468)
FCA5	−1.373	−1.367	−1.392	−1.325	−0.435	−0.422	−0.379	−0.345
	(−10.719)	(−11.313)	(−12.080)	(−12.424)	(−6.299)	(−6.127)	(−5.355)	(−4.687)
FCA6	−1.031	−1.028	−1.07	−1.035	−0.275	−0.268	−0.225	−0.178
	(−8.469)	(−8.901)	(−9.721)	(−10.128)	(−4.233)	(−4.127)	(−3.358)	(−2.545)
FCA7	−1.485	−1.467	−1.509	−1.456	−0.558	−0.539	−0.485	−0.451
	(−11.802)	(−12.615)	(−13.579)	(−14.124)	(−8.570)	(−8.303)	(−7.271)	(−6.463)
FCA8	−1.951	−1.956	−1.976	−1.949	−0.806	−0.78	−0.722	−0.639
	(−13.846)	(−15.031)	(−16.167)	(−17.439)	(−11.699)	(−11.419)	(−10.239)	(−8.706)
FCA9	−2.098	−2.088	−2.171	−2.205	−0.9	−0.889	−0.812	−0.699
	(−15.144)	(−16.063)	(−17.305)	(−18.891)	(−12.906)	(−12.777)	(−11.387)	(−9.428)
	Same Parameters for Both Classes							
Fee	−0.477	−0.478	−0.481	−0.48				
	(−15.324)	(−15.362)	(−15.396)	(−15.366)				
Salmon/Trout	−0.028	−0.028	−0.028	−0.028				
	(−7.913)	(−7.853)	(−7.759)	(−7.891)				
Bass	−0.032	−0.032	−0.032	−0.032				
	(−9.415)	(−9.341)	(−9.300)	(−9.341)				

3.2 FIML estimation

The $\Pr(1\,|\mathbf{x}_i, \mathbf{y}_i)$ from sequential estimation were used as starting values for FIML estimation using the E-M algorithm (Section 2.1). The six steps involved in each iteration were programmed in Gauss (step 5 using MaxLik).[9]. Convergence was achieved at 30 iterations.[10] Assuming the assumptions of our model are correct, the parameters from FIML estimation are asymptotically efficient, whereas those from sequential estimation are not. Two questions arise. (1) How does the characterization and sizes of the two classes estimated jointly with the attitudinal and choice data differ from the two classes estimated with only the attitudinal data? And (2), how, if at all, have the estimated preference parameters changed?

As can be seen by comparing columns 1 and 2 with columns 3 and 4 in Table 1, the qualitative characterizations of the two classes in terms of the average predicted responses to the attitudinal questions are the same whether those responses are estimated with the FIML or sequential approach: there is a FCA class and a Perch/Walleye class.

Table 2 reports estimates of the unconditional probability of belonging to the FCA class and the preference parameters from both sequential and FIML estimation. For FIML estimation, the results are reported after one iteration, after 10 iterations, and at convergence.

The first and most important thing to notice is that in going from sequential estimation to FIML estimation the probability of belonging to class 1, the FCA class, has increased from 29.6% to 40.4%. This is a 37% increase and quite substantial.

[9]The code and data are available at xxxx

[10]Convergence was assumed when the probability of membership in Class 1 remained the same, rounded to the nearest percentage, for three iterations, and simaltaneously the estimates of the $\boldsymbol{\beta}$ parameters all changed by less than 1% of their value.

The second major observation is that the separation between the two classes increases in terms of the preference parameters: the FCA class remains effectively the same in terms of the preference parameters and the Perch/Walleye class's disutility from FCAs declines.[11] In explanation, the Perch/Walleye class has effectively the same catch-time parameters as in sequential estimation, but, for every FCA level, the estimated disutility had declined; in other words, the FIML Perch/Walleye class cares less about FCA levels than does the sequential Perch/Walleye class. Comparing the sequential and FIML estimates for the FCA class, FIML estimates indicate more disutility associated with increased perch catch-times (still much less than for those in the Perch/Walleye class) and no change in the estimate of the preference parameter for Walleye catch-times. The disutility associated with the different FCA levels change little, except for level nine where the estimated disutility is marginally higher.

The FIML parameters on fee and catch times for Trout/Salmon and Bass are the same as the sequential estimates.

Thinking in terms of willingness-to-pay for the absence of PCBs, and speaking loosely, for the FCA class, the FIML and sequential parameter estimates would generate almost the same willingness to pay per Green Bay fishing day for eliminating the need for the current FCA levels, but FIML estimates predicts a much larger FCA class. For the Perch/Walleye class both FIML estimated damages and the size of the class are smaller than those based on the sequential estimates.

Interestingly, the estimated asymptotic t statistics for all of the FCA class-specific pref-

[11] Since the fee parameter essentially is the same in all models, this is equivalent to an analysis in terms of MRS.

erence parameters have all increased in absolute value, appearing to demonstrate that FIML estimation with all of the data has made these parameter estimates more efficient. However, the opposite has happened with all of the Perch/Walleye class-specific preference parameters, suggesting that the accuracy of the parameter estimates for this class have decreased. Perhaps the increased t statistics for the FCA class have more to do with its increased size than with FIML versus sequential estimation.

3.3 Discussion and possible extensions

This paper has two objectives. The first objective is to combine standard choice data with the answers to attitudinal survey questions to estimate preferences and preference heterogeneity. While both types of data are standard in surveys, they are not both typically used to estimate preferences. Our contribution is to jointly model answers to attitudinal and choice questions in the context of a discrete-choice, latent-class, random-utility model of preference heterogeneity. The results in our application indicate that answers to both the choice and attitudinal questions are coming from the same data generating process, supporting the underlying assumption of our proposed model.

The second objective is to compare a FIML model of attitudinal and choice data with a sequential model. In our application, the FIML and sequential estimation approaches give similar qualitative results. There are significant quantitative differences between the results of the two methods, however. The class sizes, the parameters in the indirect utility functions, and the significance levels all change, which impact elasticities, predicted choices, and willingness to pay measures. While sequential estimation is certainly much easier to perform, a joint FIML model is preferred on efficiency grounds. In addition, if one only oestimates a sequential model, one doesn't konw if the FIML estimates differ.

What lessons can we draw from this analysis? First, if you have two types of data generated by the same process, use both sets of data to get the best estimates. Second, modeling both types of data simultaneously will result in more efficient estimates than modeling both types of data sequentially.

There are a number of possible extensions that can be done with the presented model. It can be modified to use revealed-preference choice data or combined stated and revealed preference choice data. The number of classes can be estimated rather than assumed. The probability of class membership can be modeled as a function of observable covariates such as age or gender. Finally, the results from a LC_{AC} can be compared with those of a L_C model in order to examine the added value of attitudinal data.

Code and data to replicate, or extend, our results is available at xxxxx.

A Attitudinal questions

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 in $Q15$ through $Q34$, 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)

Figure 1: Example choice question

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...	❑	❑

References

Arcidiacono, P. and J. Jones (2003). 'Finite mixture distributions, sequential likelihood, and the EM algorithm'. *Econometrica*, 71(3), 933–946.

Ben-Akiva, M., M. Walker, A. Bernardino, D. Gopinath, T. Morikawa, and A. Polydoropoulos (2002). 'Integration of choice and latent variable models'. In H. Mahmassani, editor, *Perpetual Motion: Travel Behavior Research Opportunities and Application Challenges*. Pergamon.

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

Clogg, C. and L. Goodman (1984). 'Latent structure analysis of a set of multidimensional contingency tables'. *Journal of the American Statistical Association*, 79(388), 762–771.

De Menezes, L. and D. Bartholomew (1996). 'New developments in latent structure analysis applied to social attitudes'. *Journal of the Royal Statistical Society: Series A*, 159(2), 213–224.

Dempster, A., N. Laird, and D. Rubin (1977). 'Maximum likelihood from incomplete observations'. *Journal of the Royal Statistical Society: Series B*, 39, 1–38.

Greene, W. H. and D. A. Hensher (2002). 'A latent-class model of Discrete Choice Analysis: Contrasts with Mixed Logit'. Working Paper.

Gupta, S. and P. K. Chintagunta (1994). 'On using Demographic Variables to Determine Segment Membership in Logit Mixture models'. *Journal of marketing research*, 31, 128–136.

Hu, W., A. Hunnemeyer, M. Veeman, W. Adamowicz, and L. Srivastava (2004). 'Trading off health, environmental and genetic modification attributes in food'. *European Review of Agricultural economics*, 31(3), 389– 405.

Kamakura, W. A. and G. J. Russell (1989). 'A probabilistic choice model for market segmentation and elasticity structure'. *Journal of Marketing Research*, 26, 379–390.

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

McCutcheon, A. and M. Nawojcyzk (1995). 'Making the break - Popular sentiment toward legalized abortion among American and Polish Catholic laities'. *International Journal of Public Opinion Research*, 7(3), 232–252.

McFadden, D. (1986). 'The Choice Theory Approach to Market Research'. *Marketing Science*, 5(4), 275–297.

Morey, E., J. Thacher, and B. Breffle (2005). 'Using Angler Characteristics and Attitudinal-Data to Identify Environmental Preference Classes: A Latent-Class Model'. *Environmental and Resource Economics*, in press.

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

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

Swait, J. (1994). ‘A structural equation model of latent segmentation and product choice for cross-sectional revealed preference choice data’. *Journal of retailing and consumer services*, 1(2).

Yamaguchi, K. (2000). ‘Multinomial logit latent-class regression models: an analysis of the predictors of gender-role attitudes among Japanese women’. *American Journal of Sociology*, 105(6), 1702–1740.