Environ Resource Econ (2011) 50:83–110
DOI 10.1007/s10640-011-9463-0

A Joint Latent-Class Model: Combining Likert-Scale Preference Statements With Choice Data to Harvest Preference Heterogeneity

William S. Breffle · Edward R. Morey · Jennifer A. Thacher

Accepted: 26 February 2011 / Published online: 30 March 2011

Abstract In addition to choice questions (revealed and stated choices), preference surveys typically include other questions that provide information about preferences. *Preference-statement* data include questions on the importance of different attributes of a good or the extent of agreement with a particular statement. The intent of this paper is to model and jointly estimate preference heterogeneity using stated-preference choice data and preference-statement data. The starting point for this analysis is the belief that the individual has preferences, and both his/her choices and preference statements are manifestations of those preferences. Our modeling contribution is linking the choice data and preference-statement data in a latent-class framework. Estimation is straightforward using the E-M algorithm, even though our model has hundreds of preference parameters. Our estimates demonstrate that: (1) within a preference class, the importance anglers associate with different Green Bay site characteristics is in accordance with their responses to the preference statements; (2) estimated across-class utility parameters for fishing Green Bay are affected by the preference-statement data; (3) estimated across-class preference-statement response probabilities are affected by the inclusion of the choice data; and (4) both data sets influence the number of classes and the probability of belonging to a class as a function of the individual's type.

William S. Breffle, Edward R. Morey, Jennifer A. Thacher have equally contributed.

W. S. Breffle (✉)
School of Business and Economics, Michigan Technological University, 126 Academic Office Building, 1400 Townsend Dr, Houghton, MI 49931-1295, USA
e-mail: wsbreffl@mtu.edu

E. R. Morey
Department of Economics, University of Colorado-Boulder, Boulder, CO, USA
e-mail: Edward.Morey@Colorado.edu

J. A. Thacher
Department of Economics, University of New Mexico, Albuquerque, NM, USA
e-mail: jthacher@unm.edu

Keywords Latent class · E-M algorithm · Choice data · Preference statements · Likert-scale · Preferences · Heterogeneity

Abbreviations

E-M algorithm	Expectation-maximization algorithm
WTP	Willingness-to-pay
FCA	Fish consumption advisory
MWTP	Marginal willingness-to-pay
SP	Stated preference
RP	Revealed preference

1 Introduction

The viewpoint underlying this paper is that heterogeneity in preferences is important and should be fully explored using all of the available data. If possible, the extent of heterogeneity should be determined and used to identify and explain similarities and differences between the preferences of individuals.

We pursue a latent-class approach to heterogeneity: there is a finite number of preference classes, and class membership is latent or unobserved from the researcher's perspective. With latent-class models, while the individual knows his/her preferences, the task of the researcher is to estimate the number of classes of preferences and the probability that an individual belongs to a specific class. The goal of this paper is to develop a model of preference heterogeneity that helps predict why an individual is more or less likely to belong to a particular preference class.

We believe that exploring heterogeneity using latent-class techniques is important for a number of reasons. First, the exposition of heterogeneity can help policymakers, lawyers, and judges understand the concept of WTP and economic valuation. Understanding and policy making is further enhanced if the researcher can show how preferences differ systematically among individuals with observable characteristics. For example, showing that lower-income individuals are more likely to belong to preference class x and women rarely belong to preference class y can help policy makers have faith in WTP estimates.[1] Second, identifying heterogeneity can help generate acceptance of WTP estimates from the public. For example, if an individual knows that his/her WTP for PCB removal is \$50, he/she may reject an estimate of WTP of \$10. On the other hand, suppose the researcher explains that three preference classes dealing with PCB removal have been identified: a class like this individual who is willing to pay \$50, a smaller class willing to pay \$100, and, the majority, a class that will pay nothing. In this case, the individual can understand that average WTP is \$10 even though it differs from his/her own WTP. Finally, even if heterogeneity cannot be explained, there is something to be said for mapping its scope. There is joy in knowing we are not all the same, but that your preferences are more like hers than his.

Our model is motivated by the intent to use all the data collected in a survey: both traditional choice data and what we call *preference-statement* data. 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 importance the

[1] Alternatively, one could model preference heterogeneity assuming preferences are continuously distributed, a *random-parameters model*, implying no one has exactly the same preferences. We find a finite number of preferences classes easier to articulate to policy makers.

individual places on different attributes of a good, providing evidence of the individual's preferences towards those attributes. Consider two examples. The first is from our survey of anglers (Breffle et al. 1999) and the second is from a survey of depressed individuals about possible treatment side-effects (Thacher et al. 2005):

> 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 (FCA): "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',....,5='A lot'*)

Another type of question describes a strong preference in first-person terms and asks the extent to which the respondent agrees or disagrees. For example, from a survey of mountain bikers (Kritzberg and Morey 2008):

> I hate trying to keep up with riders faster than me.
> *(1='Definitely agree',....,5='Definitely disagree')*

In these types of questions the individual chooses a response category. What distinguishes these questions from conventional choice questions is that the individual is neither choosing between states of the world, actual or hypothetical (e.g., choice questions or a referendum contingent valuation question), nor different actions. Rather the individual is choosing his/her level of agreement with an expression of preference or choosing a response category that best answers a direct question about that person's preferences. The response categories in preference statements are often Likert scales, as in the above examples, but do not have to be.

The intent of this paper then is to model and estimate preference heterogeneity jointly with choice data and preference-statement data. Our primitive is that when presented with different states (actual or hypothetical), an individual chooses the preferred state. When asked about how he/she feels or would feel about a state, the individual provides an answer consistent with choices over states. That is, the individual has preferences, and both choices and preference statements are consistent with those preferences. Calls advocating the use of preference-statement data and combining choice data with preference-statement data under this premise go back to McFadden (1986), Ben-Akiva et al. (2002), Boxall and Adamowicz (2002), and Morikawa et al. (2002). For example, McFadden (1986) notes:

> It is common in market research to present the results of conjoint studies, or of scaling exercises for attitudes or perceptions, as useful direct information on new products and marketing programs. However, the data from these experiments can also be treated as added material for the choice theory models traditionally used by econometricians... More detailed information on preferences, obtained from ratings or rankings of alternatives and self-explicated scales, can be used to sharpen the choice model representation.

Ben-Akiva et al. (2002) states,

> ... indicators [i.e., preference statements] are helpful in model identification and increase the efficiency of the estimated choice model parameters.

We start with the assumption that underlying preferences are latent. We assume that there are C *classes* or groups of individuals: everyone in the same class has the same underlying preferences but preferences differ by class. The number of classes and their sizes are estimated. Thus, the latent-class model places few *a priori* restrictions on the degree or form of the preference heterogeneity.

The site chosen to demonstrate the model is Green Bay, a large bay on Lake Michigan, one of the Great Lakes. Green Bay is a much studied site.[2] It is heavily fished, has FCAs in place due to PCB contamination, and was the site of a litigious natural resource damage assessment.

Two types of preference parameters are estimated: the utility parameters on the site characteristics in a conditional indirect-utility function for fishing Green Bay, and the probabilities associated with answering level s to preference-question q about fishing on Green Bay. Both the marginal-utility parameter on site characteristic k ($\beta_{k|c}$) and the probability that an individual answers level s to question q ($\pi_{qs|c}$) are conditioned on class. Both are preference parameters. Recreation-demand modelers are accustomed to assuming parameters on site characteristics are preference parameters but might not, at first, recognize the $\pi_{qs|c}$ as preference parameters. Both types of parameters relate how the underlying preferences convert into choices: the first case deals with choices over states while the second case deals with choice of response category given a statement about one's preferences.

These two types of preference parameters, $\beta_{k|c}$ and $\pi_{qs|c}$, are linked together by the estimated number of classes, C, and the estimated probability that an individual is in class c as a function of his/her *type*; individuals of a certain type share a set of characteristics. In our application, a type is a set of anglers that share the same gender, retirement status, and income category. The idea is that these characteristics of the individual affect the class-membership probabilities. The characteristics of fishing Green Bay are the cost of a trip, how long it takes on average to catch each of the predominant species (perch, salmon, walleye, and bass), and the FCA level (e.g., "do not eat salmon" and "eat perch no more than once a week").

We compare a joint model estimated with both the choice data and the preference-statement data with a model estimated with only the choice data. We use MWTP estimates to make comparisons both across classes and between models.

Our estimates demonstrate five points. First, within a class, the importance anglers associate with the different Green Bay site characteristics is in accordance with their responses to the preference statements. Second, the across-class heterogeneity in the estimated $\beta_{k|c}$ is affected by the inclusion of the preference-statement data. Third, the across-class heterogeneity in the estimated $\pi_{qs|c}$ is affected by the inclusion of the choice data. Fourth, both data sets influence the number of classes and the probability of belonging to a class as a function of the individual's type. Finally, the parameters estimated with the joint data appear more efficient—in the small-sample sense of the word—than estimates based on only the choice data.

This paper makes two contributions to the literature. Our modeling contribution is linking the choice data and preference-statement data in a latent-class framework: as explained below, each half of our model is increasingly common. Our estimation contribution is showing how our joint model can be easily estimated using the E-M algorithm (Dempster et al. 1977), even though our model has hundreds of preference parameters.

Section 2 provides a brief overview of the Green Bay study site and the type of data that we are interested in modeling. We then present a general latent-class joint model in Sect. 3 and compare the joint model to two special cases. After outlining the model, we discuss how the previous literature has combined choice and preference-statement data and how the

[2] Timmins and Murdock (2007) is a recent example.

motivation for combining these two types of data differs from the joint modeling of RP and SP data. In Sect. 4, we discuss how the E-M algorithm was implemented for this application and methods for identifying the number of preference classes. Results are presented in Sect. 5. In Sect. 6, we discuss some of the benefits of the joint model. Finally, we summarize the main results of the paper in Sect. 7.

2 Green Bay Application

The target population is active Green Bay anglers who purchase fishing licenses in eight Wisconsin counties near Green Bay; most Green Bay fishing days are by these anglers. The sample consists of 640 Green Bay anglers.

Our choice data is SP questions over states of the world: "Would you rather fish Green Bay under conditions A or B?" (See Fig. 1.)

There were ten survey versions, each with eight choice pairs, so a total of 80 different choice pairs. Each pair was asked and answered approximately sixty-four times ($\frac{640}{10}$). See Breffle et al. (1999) for a full description of the data.

Each angler also answered the following fifteen preference statements:

- 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?
- 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:

  1. Eat not more than one meal a week.
  2. Eat not more than one meal a month.
  3. Do not eat.

- 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.

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

- On a scale from 1 to 5 where 1 is "Not at all important" and 5 is "Very Important", when you were making your [Green Bay] choices, how important were each of the following?

  1. The average catch rate for yellow perch
  2. The fish consumption advisory for yellow perch
  3. The average catch rate for trout/salmon
  4. The fish consumption advisory for trout/salmon
  5. The average catch rate for walleye
  6. The fish consumption advisory for walleye
  7. The average catch rate for smallmouth bass

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..	❑	❑

Fig. 1 Example choice pair

8. The fish consumption advisory for smallmouth bass
9. Your share of the boat launch fee (or daily access fee if not fishing from a boat)

3 The Latent-Class Joint Model

Assume the population consists of C different preference classes. The researcher observes, for each individual, three types of data: $\mathbf{x}_i$, $\mathbf{y}_i$ and t_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. $\mathbf{y}_i$ represents individual $i's$ choice-data, where $y_{ijh} = 1$ if individual i chose alternative j in choice pair h and 0 otherwise. The scalar t_i is the individual's type.

Latent-class models of discrete choice assume that individuals in the same class have the same preferences. The response patterns of individuals from the same class are more correlated with each other than with individuals in other classes. However, by assumption, 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; once one conditions on class, an individual's answers to all of the choice questions and preference statements are independent.

If one observes $\mathbf{x}_i$, $\mathbf{y}_i$, and t_i, and assumes each individual's class is unobserved, for C classes the likelihood function is:[3]

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

$\Pr(\mathbf{x}_i, \mathbf{y}_i | c)$ is the probability of observing the individual's responses, conditional on belonging to class c. $\Pr(c : t_i)$ is the unconditional probability of belonging to class c as a function of the individual's type or the *unconditional class-membership probability*.

Given the independence induced by conditioning on class, the likelihood function can be rewritten to account explicitly for the two types of data:

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

where

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

and

$$\Pr(\mathbf{y}_i | c) = \prod_{h=1}^{H} \prod_{j=1}^{J} (P_{jh|c})^{y_{ijh}}. \tag{4}$$

$\pi_{qs|c}$ is the probability that an individual in class c answers level s to statement q; it is a *response probability*. $P_{jh|c}$ is the probability of choosing alternative j in discrete-choice set h, conditional on being a member of class c. Each response probability $\pi_{qs|c}$ is estimated as a separate parameter subject to the constraint that $\sum_{s=1}^{S} \pi_{qs|c} = 1$.[4]

Assume the utility individual i in class c gets from alternative j in pair h follows a random utility model:

$$U_{jhi|c} = V_{jh|c}(\bullet) + \varepsilon_{jhi}. \tag{5}$$

$V_{jh|c}(\bullet)$ is the deterministic quality of the alternative, conditioned on class. It is a function of the attributes and the $\beta_{k|c}$. The $P_{jh|c}$ specification can be a probit or logit.

---

3 In what follows, the semicolon denotes *as a function of* and the symbol | denotes *conditional on*.

4 As noted in the introduction, the $\pi_{qs|c}$ are defined as direct preference parameters. Alternatively, one could, in theory, make the $\pi_{qs|c}$ a function of some smaller set of more "primitive" preference parameters.

Maximizing the likelihood function, Eq. 2, with respect to the class-membership probabilities $[\Pr(c : t_i)]$ and the response probabilities $(\pi_{qs|c})$ leads to the following useful characterizations of the maximum-likelihood estimates:

$$\Pr(c : t_i) = \frac{1}{N_{t_i}} \sum_{t_j \in t_i} \Pr(c : t_j \left| \mathbf{x}_j, \mathbf{y}_j \right.) \tag{6}$$

and

$$\pi_{qs|c} = \frac{\sum_{i=1}^{N} \Pr(c : t_i \left| \mathbf{x}_i, \mathbf{y}_i \right.) x_{iqs}}{\Pr(c) N}. \tag{7}$$

N_{t_i} is the number of sampled individuals of $i's$ type. $\sum_{t_j \in t_i}$ denotes summation over all of the individuals of $i's$ type. $\Pr(c : t_j \left| \mathbf{x}_j, \mathbf{y}_j \right.)$ is the probability that an individual belongs to class c as a function of his or her type and conditional on his/her choice data and answers to the preference-statement questions; it is a *conditional class-membership probability*.[5]

Equation 6 simply says the maximum-likelihood estimate of the unconditional probability of belonging to class c for individuals of a type is the average of the conditional probabilities for that type. Equation 7 is the estimated number of times individuals in class c answer level s to statement q divided by the estimated number of individuals in class c; so it is an estimate of the proportion of times individuals in class c answer level s to statement q. Both Eqs. 6 and 7 are intuitive and what one would expect of the maximum-likelihood estimates.

Bayes' theorem can be used to derive one more useful relationship:

$$\begin{aligned}
\Pr(c : t_i | \mathbf{x}_i, \mathbf{y}_i) &= \frac{\Pr(\mathbf{x}_i, \mathbf{y}_i | c) \Pr(c : t_i)}{\Pr(\mathbf{x}_i, \mathbf{y}_i)} \\
&= \frac{\Pr(\mathbf{x}_i, \mathbf{y}_i | c) \Pr(c : t_i)}{\sum_{c=1}^{C} \Pr(\mathbf{x}_i, \mathbf{y}_i | c) \Pr(c : t_i)} \\
&= \frac{\Pr(c : t_i) \prod_{q=1}^{Q} \prod_{s=1}^{S} (\pi_{qs|c})^{x_{iqs}} \prod_{h=1}^{H} \prod_{j=1}^{J} (P_{jhc})^{y_{ijh}}}{\sum_{c=1}^{C} \Pr(\mathbf{x}_i, \mathbf{y}_i | c) \Pr(c : t_i)}.
\end{aligned} \tag{8}$$

Equation 8 simply says that the conditional class-membership probability can be expressed as the probability of an individual's response pattern as a function of type, divided by the probability of his/her response pattern without knowledge of type.

The model is completed by specifying $V_{jh|c}$ as a function of the attributes and the $\beta_{k|c}$. In this application, we assume the following simple linear specification for $V_{jh|c}$, based on the attributes in the choice questions (see Fig. 1):

$$\begin{aligned}
V_{jh|c} = {} & \beta_{FEE|c}\text{FEE} + \beta_{FCA2|c}\text{FCA2} + \beta_{FCA3|c}\text{FCA3} + \ldots + \beta_{FCA9|c}\text{FCA9} \\
& + \beta_{BASS|c}\text{BASS} + \beta_{SALMON|c}\text{SALMON} + \beta_{PERCH|c}\text{PERCH} \\
& + \beta_{WALLEYE|c}\text{WALLEYE}.
\end{aligned} \tag{9}$$

FEE is the fee to fish and $FCAx$ is level x of the FCA levels. The remaining variables refer to the average amount of time to catch a particular fish species. $SALMON$, for example, is the average amount of time to catch a salmon (the reciprocal of the catch rate). In our application, $j = 2$: A and B. For this application, we assumed a probit specification for $P_{jh|c}$.

[5] Note that the probability of belonging to class c is the sum of the probabilities that each type belongs to class c.

Past research with this data set indicated that gender, retirement status, and income level (above or below \$50,000) are likely to influence the class-membership probabilities. Based on this, we divided the anglers in the data set into four types: females (f), retired males (rm), working males with income greater than \$50,000 ($wm > 50$), and working males with income less than \$50,000 ($wm \leq 50$). The percentage share of each of these four types are respectively: 18%, 7% 45%, and 30%.

3.1 Special Cases of the Joint Model

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

3.1.1 Latent-Class Choice-Only Model

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

$$L_{choice} = \prod_i \left[\sum_{c=1}^{C} \Pr\left(\mathbf{y}_i | c\right) \Pr(c : t) \right]. \tag{10}$$

It is a function of only the choice data. The preference statements are ignored. The estimates of the unconditional class-membership probabilities, the number of classes, and the indirect-utility parameters are those that maximize Eq. 10. The software Latent Choice (Vermunt and Magidson 2003) has a package for estimating latent-class choice-only models or one can program the likelihood function in software such as GAUSS (2000) or R (R Development Core Team 2005). Economic examples of latent-class choice-only models are multiplying and include Provencher et al. (2002); Greene and Hensher (2003); Scarpa and Thiene (2005); Scarpa et al. (2005); Kemperman and Timmermans (2006); Colombo and Hanley (2007); and Patunru et al. (2007).

3.1.2 Latent-Class Preference-Statement Model

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

$$L_{PS} = \prod_i \left[\sum_{c=1}^{C} \Pr\left(\mathbf{x}_i | c\right) \Pr(c : t) \right]. \tag{11}$$

It is a function of only the preference statements and ignores the choice data. The estimates of the unconditional class-membership probabilities, the number of classes, and the response probabilities are those that maximize Eq. 11. One can estimate latent-class preference-statement models using the *LC Cluster* package in Latent Gold (Vermunt and Magidson 2005) or software such as GAUSS (2000) or R (R Development Core Team 2005). While using choice data alone is *de rigeur* in recreation-demand modeling, preference-statement data are widely used in other fields and increasingly in environmental economics. McCutcheon (1987) is an early example of a latent-class preference-statement model. Economic examples include Thacher et al. (2005), Morey et al. (2006), Aldrich et al. (2006), Choi et al. (2007); Owen and Videras (2007), Morey et al. (2008), and Ward et al. (2008).

Note that if the number of classes is known and if each individual's class is known, the maximum-likelihood estimate of each $\pi_{qs|c}$ is simply the proportion of times individual's in class c answer level s to preference-statement q (Eq. 7).[6] So if C is known and and everyone's class membership is known, one might reasonably question of the value of a latent-class preference-statement model. But, neither is known: the number of classes and the class-membership probabilities must be estimated. Morey et al. (2008) for example, shows the richness of what can be determined about environmental preferences from a latent-class preference-statement model.

Economists identify the estimated indirect-utility parameters (the $\beta' s$) as preference parameters. A question is whether they are the only preference parameters; that is, the only measures of preference that influence choices and preference responses. If one assumes the indirect-utility parameter estimates are the only preference parameters, then by assumption, the response probabilities (the $\pi_{qs|c}$) are not preference parameters unless one makes the response probabilities a function of the $\beta_{k|c}$ in the conditional indirect utility for Green Bay. We do not ascribe to this view and do not make the $\pi_{qs|c}$ a function of the $\beta_{k|c}$. If we did ascribe to this view, the latent-class preference-statement model would be a reduced-form economic-model that could be made more structured by making the $\pi_{qs|c}$ a function of the $\beta_{k|c}$.[7]

3.1.3 Relationship Between the Three Models

All the parameters in the joint model are jointly determined by both the choice data and preference-statement data. The theory implies this and the results demonstrate it. As will be seen in Sect. 5, going from the choice-only model to the joint model changes the estimates of the indirect-utility parameters and the unconditional class-membership probabilities. Going from a latent-class preference-statement model to the joint model changes the estimates of the response probabilities, the unconditional class-membership probabilities, and the number of classes.

3.2 Comparison to the Literature

3.2.1 Approaches to Combining Choice and Preference-Statement Data

It is important to distinguish our joint model from the latent-class models of Swait and Sweeney (2000), Boxall and Adamowicz (2002), Owen and Videras (2007), and Patunru et al. (2007). These models also combine choice data and preference statements but use these data differently.

Swait and Sweeney (2000) and Boxall and Adamowicz (2002) use responses to preference statements as exogenous variables, as one might use age to explain observed choices. That is, they "regress" choice on the responses to an attitudinal questions. This is inconsistent with our prior that preferences simultaneously determine both choices and the responses to the preference statements.

Both Owen and Videras (2007) and Patunru et al. (2007) use sequential estimators. Owen and Videras (2007) estimate a latent-class preference-statement model, then impose each

[6] Imagine that one estimates a C class joint model only to determine that the data are best explained by only one class; that is, one estimates there is no preference heterogeneity. In this case, the $\pi_{qs|c}$ simplify to π_{qs} and their estimates are simply the proportion of the sample that answered level s to preference-statement q. Even in this case of one estimated class, the β_k are jointly determined by both the choice data and the preference-statement data: the result that there is only set of β_k is jointly determined by the two types of data.

[7] If one followed this path, the $\beta_{k|c}$ would appear in both parts of the model, making the two parts of the model more linked. For example, one could make $\pi_{qs|c}$ for the question about the importance of perch catch times a function of the $\beta_{|c}$ on perch catch times. Alternatively, one could make the $\beta_{|c}$ a function of the $\pi_{qs|c}$.

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. (2006)), 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. (2007) 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.

3.2.2 Motivations for Combining Multiple Sources of Data

Much of the recreation-demand literature on combining data types concerns combining SP and RP choice-data (e.g., Ben-Akiva and Morikawa 1990; Cameron 1992; Hensher and Bradley 1993; Adamowicz et al. 1994).[8]

There are two primary motivations for combining RP and SP data. The first is to introduce additional variability for an attribute, through inclusion of SP data. The second motivation is to use RP data to "stabilize" the SP data (Smith 2009). This second motivation for combining SP and RP choice-data is based on two premises: both data types contain information about preferences, but one of the types (typically the RP data) is considered a more trustworthy representation of those preferences. One then wants to use the SP data because it contains information about preferences, but one also wants to use RP data to discipline the SP data. Neither adding additional variability nor stabilizing the SP data are the motivation for the latent-class joint model. In contrast, we assume our SP data and our preference-statement data are equally reflective of the anglers' preferences.

4 Methods

4.1 Estimation Methods

It is possible, in theory, to maximize the log-likelihood function (Eq. 2) by searching *simultaneously* over all the model parameters. We do not investigate this traditional method of estimation. In practice its feasibility will depend on one's problem: the underlying preferences and the specifics and extent of the choice and preference-statement data. We chose the E-M algorithm for estimation because it has worked well for us in simpler latent-class applications, it is less familiar than the traditional method, and our adaptation of it for joint data sets should have applicability in other contexts with multiple data types. Our code is available and can be easily modified in numerous ways.[9]

The E-M algorithm also maximizes the log-likelihood function but does not do it by simultaneously searching over all of the parameters. Put loosely, the E-M algorithm divides the parameters into two groups, and we begin by specifying starting values for the parameters in group 1. Then, conditional on the the parameters in group 1, the log-likelihood is maximized with respect to parameters in group 2. Based on the newly minted parameter estimates for

[8] Note that in our application there is no RP data; we combine two types of SP data. That said, the SP choice component of the joint model could be easily replaced by a RP choice component, or one could add a RP choice component to the joint model and have a three component model.

[9] The code is available at: http://www.colorado.edu/economics/morey/dataset.html.

group 2, the parameter estimates for group 1 are revised, using Bayes' theorem. The process repeats until updating the group 1 parameters based on the most recent group 2 parameters does not change the group 1 parameters.

Given our data set and model, C classes, and assuming four types of anglers, there are $4(C-1)$ identified class-membership probabilities: $\Pr(c : f)$, $\Pr(c : rm)$, $\Pr(c : wm > 50)$, and $\Pr(c : wm \leq 50)$. There are $13C$ identified $\beta_{k|c}$ parameters and $77C$ identified $\pi_{qs|c}$ parameters.[10] So, for example, a two-class joint model has 184 parameters and a four-class model has 368 parameters.

In the context of this application, the E-M algorithm specifies initial numerical starting values for all of the conditional class-membership probabilities subject to the constraint that they sum to one: $\sum_{c=1}^{C} \Pr(c : t_i \,|\mathbf{x}_i, \mathbf{y}_i) = 1$. Note that each individual is of only one type and there is a maximum of 640 $(\mathbf{x}_i, \mathbf{y}_i, t_i)$ patterns, so there are $640C$ of these conditional class-membership probabilities.

Equation 6 is then used calculate the unconditional class-membership probabilities. These are the maximum-likelihood estimates of the unconditional class-membership probabilities, conditional on the specified conditional class-membership probabilities. Given the conditional class-membership probabilities, Eq. 7 is used to calculate the $77C$ identified response-probability parameters $(\pi_{qs|c})$.

One then uses a search algorithm to find the $\beta_{k|c}$ parameters that maximize the log-likelihood function assuming the current $\Pr(c : t)$ and $\pi_{qs|c}$ estimates.[11] Equation 8 is then used to revise the $\Pr(c : t_i \,|\mathbf{x}_i, \mathbf{y}_i)$ estimates. This is the end of the first E-M iteration. E-M iterations are repeated until the $\Pr(c : t_i \,|\mathbf{x}_i, \mathbf{y}_i)$ estimates remain stationary and the log-likelihood function does not increase in value.

Note that in the application, and compared to our experience with traditional estimation, each iteration of the E-M algorithm was slow, but total convergence time was faster: convergence tended to be smooth and continuous. Different starting values were used to make sure the global maximum was achieved. The joint model tended to converge more quickly than the choice-only models, leading us to conjecture that adding the preference-statement component and data to the model made the maximum of the likelihood function more distinct—adding the preference-statement data weakens multicollinearity, making identification of the maximum easier.

4.2 Identifying the Number of Classes

The basic estimation approach is to repeatedly estimate the model for different numbers of preference classes. Fit criteria are then used to identify the number of classes that best fits the data.

There is no classical statistical test to determine whether increasing the number of latent classes significantly improves model fit. A likelihood-ratio test does not exist because the discrete nature of adding a class violates assumptions needed to prove the statistic is chi-squared distributed (Wedel and Kamakura 2000).

In the economic latent-class literature, the number of classes is almost exclusively chosen on the basis of *information-criteria scores*. See, for example, Scarpa and Thiene (2005), Morey et al. (2006), Kemperman and Timmermans (2006), and Patunru et al. (2007). Every

[10] There are fourteen preference statements, each with six levels (including no response), and one statement with eight levels (including no response).

[11] In the E-M literature, Eq. 2 in terms of the current $\Pr(c : t)$ and $\pi_{qs|c}$ is referred to an "expected" likelihood function. One can use a search algorithm such as Optimum or Maxlik in GAUSS (2000) to maximize it in terms of the $\beta_{|c}$.

proposed information criterion is increasing in how much the likelihood function increases when a class is added, and decreasing in the number of additional parameters that result from adding a class: the improvement in the likelihood function is penalized by a function of the additional number of parameters. Numerous information criteria have been proposed. See Akaike (1974), Bozdogan (1987), Hurvich and Tsai (1989), and Schwarz (1978). Information criteria differ in terms of how the improvement in the likelihood function is expressed, the form of the penalty function, and whether the criteria is a function of the sample size. One would like the "best" number of classes to be consistent across the different information criteria, but this is often not the case. Yang and Yang (2007) examine the ability of information criteria, including those with sample size adjustments, to differentiate between latent-class models.[12]

5 Results

In this section we first identify the number of latent classes and assess the fit. We then discuss results from the best-fitting models for both the joint and choice-only models.

5.1 Number of Classes and Fit

For the Green Bay data set, we estimated models with one through five classes. For each model we calculated six standard information criteria (AIC3, CAIC, CAIC*, AICC, BIC, and BIC*). The best fit, and thus number of classes, is identified by the lowest information-criteria score. While the criteria do not all indicate the same number of classes, taken together, they suggest four classes for both the joint model and the choice-only model.[13]

Consider how well our joint model "fits" the data.[14] Because 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 4-class joint model correctly predicts 59% of the SP choice pairs answered and 31% of the preference statements. Thus, overall it correctly predicts 40% of all the responses. A random allocation would correctly predict 50% of the choice pairs, about 20% of the preference statements, and about 30% of all of the responses. There are two possible issues with this measure

[12] Information criteria are not the only way to identify the number of classes, and practice varies across fields. In education and psychology, the practice is to find the minimum number of classes that adequately explain the data: first find models with enough classes to produce "good fit" in terms of actual versus predicted responses—parsimonious models. Information criteria are then used to examine the trade-off between parsimony and fit.

[13] In contrast, estimation of a preference-statement-only model suggests three classes. The choice data suggests more preference heterogeneity than does the preference-statement data, possibly because the specific fees are explicit in the Green Bay choice pairs.

[14] In education and psychology, 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 (Formann 2003). However, as has been frequently noted (Eid et al. 2003; Yang and Yang 2007), 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, most possible response-patterns are never observed, meaning that the chi-squared approximation for the Pearson and Read-Cressie statistics will not be valid. Alternate statistical tests have been proposed for the case of binary-response data (Reiser and Lin 1999; Bartholomew and Leung 2002; Maydeu-Olivares and Joe 2005).

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 (Train 2003). 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.

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.[15] In all cases, we are unable to reject the null hypothesis (p-value $\geq$ 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. Because of the problem of sparse data, we cannot do a chi-square on all of the preference statements simultaneously.

We also examined how precisely the chosen model assigns respondents to classes. For each respondent, of the four estimated conditional class-membership probabilities, one is usually 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 72% of the sample. Thus, for a given response pattern, individuals are predicted to belong to one particular class with a very high probability. This implies that there are notable differences in classes; in a case of classes that do not vary, one would expect probabilities closer to 25%.

5.2 The Four-Class Model

Here we report and discuss results for the estimated joint-model. For comparison we also present results from the estimated choice-only model. As noted earlier, four classes were identified as providing the best fit for both models. Note that there is nothing in the theory of latent-class models that requires that the four classes estimated in the choice-only model correspond in total, or by number, to the four estimated classes in the joint model. That is, the preferences of anglers in Class x in the choice-only model do not have to correspond to the preferences of any class, including Class x, in the joint model. The fact that both of these models have four estimated classes is a result of the data; this is not a restriction imposed on the model.

5.2.1 The Joint Model

On the basis of a likelihood-ratio test, the null hypothesis that type is not a determinant of the class-membership probabilities is rejected—preference heterogeneity is explained, in part, by our four types of anglers.

Table 1 reports the estimated unconditional class-membership probabilities by class and angler type (Eq. 6). Class 3 is the largest: there is a 37% probability of belong to Class 3. The class-membership probability for the three other classes is approximately 20% each.

Women are most likely to belong to Class 3. Working men with incomes less than or equal to $50,000 are twice as likely to be in Class 2 as are working men with higher incomes. However, one gets the opposite result for Class 4, indicating income is an important determinant

[15] 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 excluded the no-response category from the Pearson test.

Table 1 Estimated unconditional class-membership probabilities (%) by class and by type

Description	Class 1	Class 2	Class 3	Class 4
Joint model				
Class average	22	19	37	22
Women	28	28	34	10
Working men with income $\leq$ \$50K	21	24	39	16
Working men with income > \$50K	19	12	38	30
Retired men	27	24	25	24
Choice-only model				
Class average	35	26	13	25
Women	44	19	27	10
Working men with income $\leq$ \$50K	34	43	5	18
Working men with income > \$50K	33	18	15	34
Retired men	36	26	3	35

Detail may not sum to 100 due to rounding

of the class-membership probabilities. Women are unlikely to be in Class 4. Retired men are fairly equally distributed across the four classes, suggesting they are a diverse type.

Tables 2 and 3 report the estimated response-probabilities. Class 1 is most bothered by FCAs while Class 4 is least bothered. Classes 2 and 3 are fairly similar in the degree to which they are bothered by FCAs.[16] Though similarly bothered by FCAs, members of Class 3 are estimated to much more strongly agree with the statement "I would be willing to pay higher boat launch fees if the fish had no PCB contamination:" 62% of Class 3 agree or strongly agree while 22% of Class 2 agree or strongly agree. The model estimates that 42% of Class 2 strongly disagrees with paying a higher fee, whereas only 4% of Class 3 strongly disagrees. Class 2 is estimated to place substantially more importance on the fee than the other classes.

These differences between Classes 2 and 3 in terms of propensity to pay is consistent with the class-membership estimate that low-income working males are twice as likely to be in Class 2 as are high-income working-males; in other words, poorer anglers are more likely to be in Class 2.[17] Members of Class 1 have the highest estimated agreement with the statement "I would be willing to pay higher boat launch fees if the fish had no PCB contamination." Summarizing, the estimated response-probabilities suggest that Class 1 will have the highest WTP for reducing PCB contamination, Class 3 lower estimates, and Class 2 even lower estimates than Class 3. We expect this even though members of Class 2 are estimated to be as bothered by FCAs as are members of Class 3.

The estimated response probabilities associated with the importance of the different Green Bay site characteristics indicate that for Classes 1 and 3, FCA levels are more important than catch times. This suggests Classes 1 and 3 will be willing to pay more for PCB removal than for increased catch. For Classes 2 and 4, catch times are generally more important than FCA levels, suggesting the opposite.

Class 4's estimated response probabilities stand out in terms of "not at all bothersome," "strongly disagree," and "not at all important"—members of Class 4 are predicted to be much

[16] Class 2 is more likely than Class 3 to find an FCA of "eat no more than one meal a week" very bothersome but less likely than Class 3 to find "do not eat" very bothersome.

[17] Class 3 is also estimated to be more likely to agree with paying higher fees for higher catch rates.

Table 2 Estimated response probabilities (%) by Class (7-level)

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

Class 1 levels							Class 2 levels							Class 3 levels							Class 4 levels						
1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	12	20	25	22	13	4	4	15	12	18	24	7	4	3	9	23	33	17	10	1	3	6	15	31	20	14	7

Numbers do not sum to 100 as no-response is not reported

[a] 1 = Much Worse, …, 7 = Much Better

Table 3 Estimated response probabilities (%) by Class (5-level)

Statement	Class 1 levels					Class 2 levels					Class 3 levels					Class 4 levels				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Extent bothered statements[a]																				
FCA of eat no more than one meal a week	12	4	10	15	60	15	19	16	12	30	22	18	27	17	17	47	11	17	6	15
FCA of eat no more than one meal a month	2	1	10	12	74	7	5	18	18	38	5	8	22	28	37	35	8	20	13	22
Do-not-eat FCA	0	0	1	1	96	9	1	3	4	72	3	2	6	6	82	25	4	4	7	54
Agreement statements[b]																				
WTP for higher catch rates	15	8	37	23	17	62	9	20	6	2	6	16	38	26	14	33	12	26	10	20
WTP for no PCB contamination	7	0	19	15	59	42	12	24	13	9	4	10	23	37	25	32	10	30	8	21
Importance statements[c]																				
Catch rate: perch	8	14	30	15	32	9	7	28	17	39	2	14	16	29	39	29	13	21	3	32
FCA: perch	2	2	6	7	82	8	8	17	19	49	3	10	20	34	34	39	10	28	11	9
Catch rate: trout/salmon	17	11	36	16	20	21	20	21	14	24	16	32	33	14	4	45	3	18	13	20
FCA: trout/salmon	9	0	5	11	75	17	17	11	18	36	14	28	34	22	3	52	4	18	14	11
Catch rate: walleye	4	16	34	19	27	3	14	21	21	41	2	12	35	32	19	21	4	30	20	24
FCA: walleye	1	3	4	7	84	2	11	10	23	54	2	15	29	36	19	31	6	34	15	9
Catch rate: bass	9	14	32	22	23	7	14	38	14	26	14	20	23	35	7	32	14	14	6	31
FCA: bass	4	6	7	7	76	17	14	21	16	32	23	22	25	27	2	60	10	17	4	4
Fee	24	17	32	10	15	6	5	16	17	56	15	19	36	21	8	22	16	29	13	19

Numbers do not sum to 100 because the no-response category 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

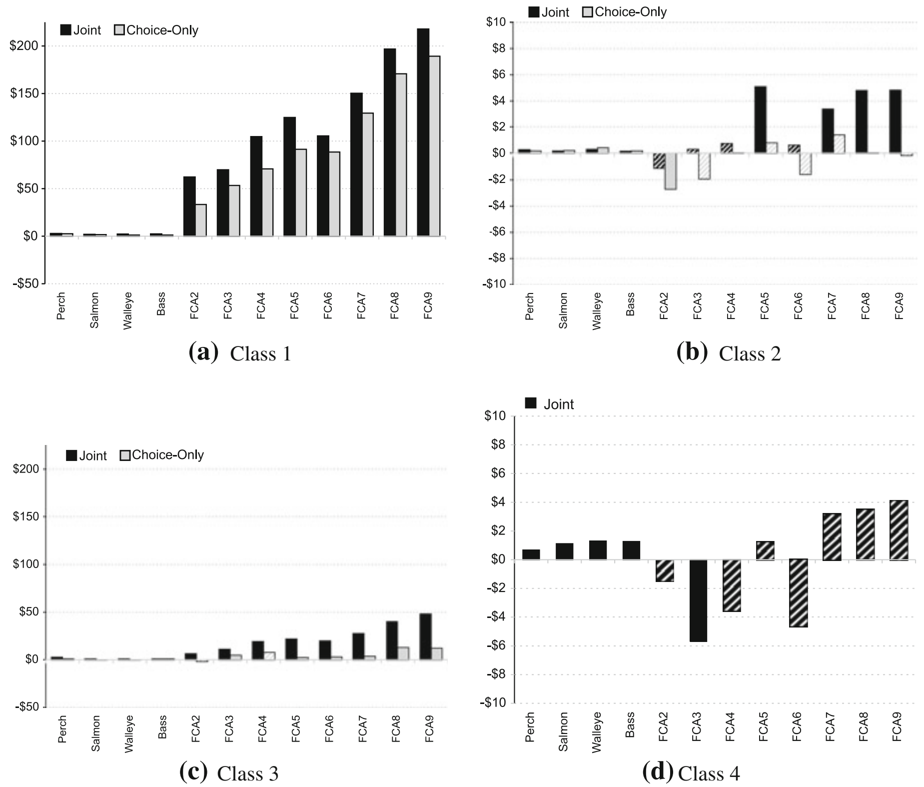

Fig. 2 Per-Trip MWTP by Class—bar is hatched if estimated attribute parameter is not significantly different from zero

more likely than the other classes to choose response level 1. "Not at all bothersome" and "not at all important," combined with a tendency to "strongly disagree" with having a willingness to pay a higher fee for better attributes, suggests an indifference to Green Bay's attributes, or at least to changes in the levels of those attributes. One might expect members of Class 4 to have little or no WTP for improved Green Bay attributes. Consistent with this, Class 4 is estimated to judge Green Bay better, relative to other sites, than do members of the other classes. Maybe they are content with Green Bay as it is.

Table 4 reports the estimated $\beta_{|c}$ for the joint model. Figure 2 reports the corresponding MWTP estimates by class for each of the twelve Green Bay site attributes.[18]

The MWTP estimates are simply the attribute parameter estimate divided by the negative of the fee parameter. MWTP usefully summarizes the β estimates because, while across models, the $\beta_{k|c}$ estimates are subject to a potential scaling effect, the MWTP estimates are not.

Note how the estimated fee parameters vary by class; the estimated fee parameter for Class 2 is many times larger (in absolute value) than the other estimated fee parameters, indicating that Class 2 is much more sensitive to money. This is consistent with the message conveyed

[18] Note that Fig. 2 reports Classes 1 and 3 using the same income scale, and a different income scale for Classes 2 and 4. A cross-hatched bar indicates the attribute parameter estimate is not significantly different from zero.

Table 4 Estimated utility parameters ($\beta_{k|c}$) from the joint model

Parameters	Class 1		Class 2		Class 3		Class 4	
	Est	t-stat	Est	t-stat	Est	t-stat	Est	t-stat
β_{PERCH}	−0.0461	−3.962***	−0.0383	−2.797***	−0.0918	−11.122***	−0.0263	−2.665***
β_{SALMON}	−0.0318	−3.131***	−0.0234	−2.310**	−0.0304	−4.738***	−0.0431	−5.612***
$\beta_{WALLEYE}$	−0.0367	−3.689***	−0.0425	−4.175***	−0.0405	−6.500***	−0.0503	−6.673***
β_{BASS}	−0.0392	−4.203***	−0.0231	−2.433**	−0.0288	−4.814***	−0.0493	−6.632***
β_{FCA2}	−0.990	−5.471***	0.158	0.968	−0.212	−1.987**	0.0567	0.473
β_{FCA3}	−1.111	−6.131***	−0.0412	−0.243	−0.356	−3.506***	0.220	1.819*
β_{FCA4}	−1.664	−8.784***	−0.105	−0.649	−0.626	−6.106***	0.137	1.119
β_{FCA5}	−1.985	−9.873***	−0.719	−4.272***	−0.700	−6.584***	−0.0481	−0.388
β_{FCA6}	−1.676	−8.648***	−0.0862	−0.536	−0.636	−6.209***	0.180	1.483
β_{FCA7}	−2.391	−12.051***	−0.476	−3.217***	−0.887	−8.709***	−0.123	−1.039
β_{FCA8}	−3.130	−13.597***	−0.674	−3.929***	−1.285	−11.880***	−0.135	−1.07
β_{FCA9}	−3.465	−14.812***	−0.680	−4.034***	−1.545	−13.152***	−0.159	−1.318
β_{FEE}	−0.0159	−1.753*	−0.141	−12.757***	−0.0322	−5.481***	−0.0387	−5.798***

Based on probit specification with 640 observations
***,**, and * indicate significance at the 1%, 5%, and 10% levels, respectively

by the estimated response probabilities and the estimate that low-income working males are twice as likely to be in this class as are high-income working males.

The FCA marginal-utility parameters for Classes 1 and 3 are all negative, significant, and strictly monotonic in the FCA levels. They are larger (in absolute value) for Class 1. Correspondingly, the MWTP estimates reported in Fig. 2 are higher for Class 1 than for Class 3 because of the larger FCA parameters and the smaller fee parameter.

For Class 2, the estimated indirect-utility parameters on FCA levels are significantly negative and monotonic for the three highest FCA levels, but smaller in absolute values than the corresponding Class 3 parameters. These estimates, combined with the large negative fee parameter for Class 2 translate into MWTP estimates that are positive but small, a few dollars at most. This result is consistent with the response-probability estimates and the estimate that lower-income males are more likely to be members of Class 2.

For Class 4, only one of the indirect-utility parameters on FCAs is significantly different from zero—the one significant estimate is positive—meaning that one cannot reject the null hypothesis that members of Class 4 have zero WTP to eliminate PCBs and their corresponding FCAs. Again, this is consistent with the Class 4 story in terms of the estimated response probabilities. Women are unlikely to belong to Class 4, implying that the vast majority of women, unlike men, have a positive WTP for reducing PCB contamination. This finding is consistent with numerous studies that have found that women have higher WTP for an environmental improvement if the improvement is a reduction in a health risk, particularly a risk to children (e.g., Dupont 2004).

The catch-time parameters are negative and significant for all four classes—catching fish faster is good.[19] In terms of catch, Class 3 cares most about perch; estimated $\beta_{PERCH|3}$ is twice the absolute value of the other perch parameters in the other classes. A big difference in catch-time parameters is not found across classes for the other three species.

As suggested by the response-probability estimates, the Class 1 and Class 3 MWTP estimates to reduce FCA levels are much larger than their MWTP estimates for reducing catch times; for Classes 2 and 4 they are much smaller.

Table 5 reports the predicted average response level by class for each of the preference statements.

Summarizing the joint model, the estimated utility parameters, response probabilities, and class-membership probabilities are in accordance. Furthermore, the model indicates an extensive amount of heterogeneity across the four estimated classes in terms of the MWTP estimates, responses to the preference-statement questions, and who belongs to the different classes.

5.2.2 Comparison with the Choice-Only Model

To distinguish results from the joint and choice-only models, we denote the choice-only classes as $1C$, $2C$, etc. and the joint classes as $1J$, $2J$, etc.

From Table 1, note that the sizes of the four classes and the class-membership probabilities by type are quite different in the choice-only and joint models: Class $1C$ is much larger than Class $1J$. Class $3C$ is much smaller than Class $3J$. In the joint model, retired men were spread approximately evenly across the four classes; in the choice-only model, they are unlikely to

[19] Because the average real perch catch time is low (0.75 h), all perch catch-time parameters reported in the paper represent the marginal utility of a change of one-tenth of an hour (i.e., 6 min). The notion of a change in perch catch times of an hour is much too large and is counter-factual. Average real catch times for the other three species average multiple hours, so the marginal utilities for these species correspond to a change in catch time of 1 h.

Table 5 Joint model: predicted average response level, by class, for preference statements

	Class 1	Class 2	Class 3	Class 4
Amount bothered statements[a]				
FCA of eat no more than one meal/week	4.08	2.97	2.89	2.16
FCA of eat no more than one meal/month	4.55	3.37	3.84	2.70
Do-not-eat FCA	4.89	3.93	4.64	3.43
Agreement statements[b]				
WTP for higher catch rates	3.19	1.76	3.26	2.72
WTP for no PCB contamination	4.19	2.37	3.66	2.75
Importance statements[c]				
Catch rate: perch	3.48	3.71	3.90	2.87
FCA: perch	4.65	3.91	3.87	2.33
Catch rate: trout/salmon	3.09	2.99	2.57	2.56
FCA: trout/salmon	4.44	3.40	2.71	2.23
Catch rate: walleye	3.49	3.83	3.55	3.15
FCA: walleye	4.68	4.17	3.56	2.50
Catch rate: bass	3.36	3.32	2.98	2.82
FCA: bass	4.44	3.32	2.60	1.66
Fee	2.69	4.13	2.89	2.85
Comparison statement[d]				
Green Bay compared to other sites	3.97	3.37	3.68	4.10

Calculated by multiplying each predicted response probability to the corresponding Likert level. The full wording of the statements can be found beginning on Page 7

[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

be in Class $3C$, and most likely to be in Classes $1C$ or $4C$. Women are much more likely to belong to Class $1C$ than in $1J$. The joint-model finding that working men with incomes of \$50,000 or lower are much more likely to be members of Class $2J$ remains in the choice-only model.

The fact that the class-membership probabilities are different raises the question of how much the preferences of members of the choice-only classes line up, if at all, with the preferences of members of the joint classes. Table 6 reports the estimated $\beta_{|c}$ for the choice-only model, and Fig. 2 shows the corresponding MWTP estimates by class for each of the twelve Green Bay site attributes. Put simply, the qualitative description of Classes 1 and 3 remains the same in the joint and choice-only models. However, the estimate parameters on the Green Bay attributes are smaller in absolute value in the choice-only model; this reduces all of the Class 1 and Class 3 MWTP estimates (see Fig. 2).

Whereas the parameter estimates indicate that reduced catch times and lower FCA levels make members of Class $4C$ better off, the estimated parameter on fee is positive and significant. Interpreted literally, this would mean that, *ceteris paribus*, a higher fee is preferred, and anglers in Class $4C$ would pay to increase FCAs. More likely, the choice data, by itself, suggests that members of Class $4C$ attach little importance to the fee magnitudes in the choice pairs (pushing their estimated fee parameter to zero). This, combined with an incorrect belief

Table 6 Estimated utility parameters ($\beta_{k|c}$) from the choice-only model

Parameters	Class 1		Class 2		Class 3		Class 4	
	Est	t-stat	Est	t-stat	Est	t-stat	Est	t-stat
β_{PERCH}	−0.0524	−4.130***	−0.0296	−1.832*	−0.255	−4.975***	−0.0812	−5.940***
β_{SALMON}	−0.0348	−3.351***	−0.0321	−2.793***	−0.0443	−1.294	−0.0514	−5.313***
$\beta_{WALLEYE}$	−0.0256	−2.472**	−0.0683	−6.272***	−0.0465	−1.955*	−0.0491	−4.925***
β_{BASS}	−0.0263	−2.822***	−0.0302	−3.182***	−0.209	−4.923***	−0.0304	−3.615***
β_{FCA2}	−0.691	−3.931***	0.440	2.420**	0.383	0.767	−0.279	−1.882*
β_{FCA3}	−1.106	−5.800***	0.312	1.585	−1.068	−2.700***	−0.100	−0.580
β_{FCA4}	−1.468	−7.169***	−0.0058	−0.030	−1.751	−3.800***	−0.023	−0.140
β_{FCA5}	−1.894	−8.636***	−0.127	−0.661	−0.540	−1.006	−0.477	−2.997***
β_{FCA6}	−1.834	−8.517***	0.258	1.346	−0.653	−2.005**	−0.162	−0.902
β_{FCA7}	−2.685	−10.677***	−0.228	−1.371	−0.833	−2.278**	−0.470	−3.009***
β_{FCA8}	−3.542	−11.275***	−0.0041	−0.021	−2.855	−4.777***	−0.411	−2.640***
β_{FCA9}	−3.925	−11.303***	0.0315	0.133	−2.670	−4.738***	−0.648	−3.231***
β_{FEE}	−0.0207	−1.977**	−0.163	−12.115***	−0.221	−5.133***	0.0204	2.560**

Based on probit specification with 640 observations

***,**, and * indicate significance at the 1%, 5%, and 10% levels, respectively that a higher fee implies better quality in terms of some unmentioned attribute, pushes the fee parameter positive, a rejection of the scenario.[20] Of interest is that the parameter on fee in Class 4 is significantly negative when the preference-statement data are added in the joint model; this indicates that the preference-statement data are "disciplining" the estimate of the fee parameter. We do not report MWTP estimates for Class $4C$; it is unlikely that the estimated attribute parameter divided by the negative of the estimated positive fee parameter is an estimate of MWTP—it is difficult to imagine that the members of Class $4C$ would pay for increased FCAs given their estimated negative and significant parameters on the different FCA levels. Similar to Class $2J$, Class $2C$ MWTP estimates for catch are less than a dollar.

Summarizing, the joint model estimates positive MWTP to eliminate high-level FCAs for Class $2J$. The choice-only model, however, basically says this class has no WTP for eliminating FCAs. In terms of the choice-only model, Classes 2 and 4 look similar, but they look quite different in terms of the joint model.

6 Benefits of the Joint Model

Historically, preference-statement data have been collected in recreation-demand surveys whose primary intent was to collect choice data. The justification for including preference-statement questions was often to focus and "warm-up" the respondent for the upcoming choice questions. We, and others, often used them only retrospectively (Lynne and Rola 1988; Breffle and Rowe 2002; Johnson et al. 2003), particularly in applications for policy or litigation. A model was estimated using only the choice data, often a model difficult for the client to comprehend, and then the response questions were used to show that the predictions of the choice model were consistent with what the respondents said was important. The underlying premise was that the same underlying preferences dictate both the individual's choices and their responses to questions about their preferences.

The joint model is consistent with this tradition but integrates the preference-statement data into the model to obtain better estimates of the individuals' preferences. Our motivation is a *use-all-the-data* approach to achieve increased small-sample efficiency.

One must be careful when discussing efficiency gains from combining data sources, even if one assumes, as we have, a single set of preferences generates both types of data. Maximum likelihood estimates are, in general, asymptotically efficient but typically not efficient in small samples, and no claim will be made that our estimates are the most efficient from a small-sample perspective.

The implications of maximum likelihood estimation for combining choice data and preference-statement data are as follows. If there is no interest in the parameters that appear only in the preference-statement part of the likelihood function, the joint model has no asymptotic-efficiency advantage over a model estimated with only choice data; one cannot asymptotically improve on asymptotic efficiency. Nevertheless, the joint estimates are small-sample more efficient (not most efficient) than the estimates based on only choice data. Paraphrasing Ben-Akiva et al. (2002), the responses to preferences statements contain information and thus potentially provide for increased efficiency in the estimation of the number of classes, the class-membership probabilities, and the utility parameters in the class-specific conditional indirect utility functions. Paraphrasing Morikawa et al. (2002), if a parameter is shared by the choice model and the preference-statement model, joint estimation of the two

[20] In latent-class models, particularly those with more than a few classes, there will often be at least one class that violates the standard assumptions of neoclassical demand theory - subsets of the sample that choose strangely are par for the course. This is valuable information, but some might find it disconcerting.

models improves statistical efficiency. Both mean small-sample efficient, but that is never made explicit. Assuming the estimates are unbiased, small-sample more efficient simply means smaller estimated standard errors.

The estimated joint model indicates that the choices in the Green Bay choice pairs and the responses to the preference statements are consistent with the same underlying preferences. This implies that a model that ignores the preference statements is ignoring useful data about preferences. It provides an argument in favor of the joint model, suggesting the joint model has a small-sample efficiency advantage (more good data is better than fewer good data). That said, we know of no statistical test of our hypothesis of a small-sample efficiency gain. However, the estimated $\beta_{|c}$ and their corresponding t statistics indicate that, with the joint model, one is more likely to reject the null hypothesis that an attribute parameter is zero.

It is true that if one only uses choice data to estimate choices, one will do a better job predicting the choices in the data set than if one adds preference-statements to the data and requires some common parameters help explain both the observed choices and responses to the preference statements. But this point, in our view, is not an argument for excluding preference-statement data: one would do a better job predicting the Green Bay choices in half of the choice pairs if one excluded the other half of the choice-pair data from the analysis, but few would argue for such an exclusion. The principle is the same.

One might argue that since the MWTP estimates for reducing PCBs estimated with only the choice data are often smaller than the corresponding estimates from the joint model, the preference-statement data must be biasing upward the MWTP estimates. We do not make this argument; the fact that adding data changes parameter estimates does not imply that the original estimates are correct, unless choice data are assumed to be unbiased and preference-statement data are assumed to be biased.

Given these benefits of the joint model, under what conditions might one report results from a choice-only model when one has preference-statement data? We believe two conditions should hold for a researcher to make this decision. First, the researcher must only be interested in predicting choice and choice heterogeneity, not preferences in a bigger sense of the word. Second, estimation of a joint model, a choice-only model, and a preference-statement-only model, indicates significant discord between the choice data and the preference-statement data. In such a circumstance, one might want to ignore the preference-statement data, but only after using it to estimate the joint model.

A final benefit of the joint model is that it has a number of useful applications in the policy arena. For example, it allows us to identify several additional probabilities. Consider some examples in the context of the Green Bay application. One can identify the probability of observing the angler's SP choices as a function of his/her type and conditional on his/her responses to the preference statements, $\Pr(\mathbf{y}_i : t_i | \mathbf{x}_i)$. In addition, one can calculate the probability of observing the angler's responses to the preference statements as a function of his/her type and conditional on his SP choices, $\Pr(\mathbf{x}_i : t_i | \mathbf{y}_i)$. In terms of prediction, these probabilities will predict behavior with more accuracy than $\Pr(\mathbf{y}_i : t_i)$ and $\Pr(\mathbf{x}_i : t_i)$.

After the joint model is estimated, one can numerically calculate $\Pr(\mathbf{y}_i : t_i | \mathbf{x}_i)$ and $\Pr(\mathbf{x}_i : t_i | \mathbf{y}_i)$ for any $\mathbf{x}_i$, $\mathbf{y}_i$ and t_i combination, and for any individual, regardless of whether that individual is in the sample. So, for example, one could have some new angler report his/her type and answer the preference-statement questions, and then use $\Pr(y_i : t_i | \mathbf{x}_i)$ to estimate the probability of him/her making a particular Green Bay choice. This will be a more accurate prediction than the one possible with no knowledge of responses to the preference statements.

In the survey that produced our data, the preference statements asked were not asked with the idea of using them to estimate our joint model–our joint model comes many years after the survey was designed and implemented. But now that we have the joint model, we can imagine preference statements designed to help identify classes based on many different criteria: classes with the same WTP but for different reasons (you worry about children eating contaminated fish versus simply worrying about the contaminated fish), protest classes, classes that are "rejecting the scenario," and individuals who vote similarly but have different motivations.[21]

7 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. Preference-statement data have strengths that complement choice data. Survey respondents are generally very familiar with preference-statement questions. Likert-scale questions, a common form for preference statements, allow individuals to indicate intensity of preference and nuances in their preferences. Preference-statement data are often ignored, however, as it has been unclear how to incorporate preference-statement data into economic decision making.[22] 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.

Our application is 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 main contribution of this paper is linking these two data types in a joint latent-class model. Our 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. Estimated parameters are the number of latent classes C, the probability that an individual of type t belongs to class c, response probabilities for preference statements, and the parameters on the fishing characteristics in the conditional indirect-utility function for fishing Green Bay. Both the response probabilities and indirect-utility parameters are preference parameters; their estimation is tied together by the estimated number of classes and the estimated probability that an individual is in class c as a function of his/her type.

Estimation is with the E-M algorithm. Its use allowed us to smoothly estimate joint models with many classes and hundreds of parameters. While our application combines preference-statement data with data from SP choice pairs, a third RP choice-component could easily be added to our joint model, and to our estimation code.

Our intent in joining choice data with preference-statement data is to enhance our picture of preference heterogeneity both in terms of its extent—why we chose a latent-class approach—and predicting the individual's preferences as a function of the characteristics of the individual. Our sense is that our joint model provides more and better estimates of preferences and their heterogeneity than does a latent-class choice-only model. Of course, views

21 Cunha-e-Sá et al. (2010), in an unpublished paper and citing an earlier version of this paper, investigate the identification of scenario rejectors in a latent-class model with attitudinal data.

22 As discussed earlier, when preference-statement data are used, they are often used, in our view, incorrectly: regressing choices on the responses to preference statements, rather than assuming the responses to the preference statements and the responses to the choice questions are jointly determined by the same underlying preferences.

can differ on what is meant by "better." We argue that they are better in a number of ways. First, adding the preference-statement data make the estimates more small-sample efficient (more good data is better than less). Second, in our application the picture of heterogeneity is clearer. Finally, the addition of the preference-statement data "discipline" the estimated $\beta_{k|c}$ in the choice component of the model. By "discipline" we mean estimates are more likely to be significant and of the correct sign. For example, in the choice-only model one of our classes has an estimated price coefficient that is positive and significant, but it is negative and significant in the joint model. Some $\beta_{|c}$ parameters on site characteristics that in the choice-only model are not significantly different from zero are significantly different from zero in the joint model.

Consider some other potential uses of our joint model. One might use it to identify a class or classes of individuals that are either "rejecting the SP choice scenario" or "protesting." For example, a protesting class would have responses to the preference statements that indicate positive WTP for an environmental improvement, but the choice data indicate WTP is zero. The preference-statement data could also help to identify classes with similar WTP but for different reasons. For example, if one had degree-of-agreement statements, "I am bothered by the PCBs because it is wrong to have unnatural chemical in the environment," and "I am bothered by the PCBs because many children and pregnant women eat the contaminated fish," one might identify two classes that make similar SP choices, but for different reasons.

References

Adamowicz W, Louviere J, Williams M (1994) Combining revealed and stated preference methods for valuing environmental amenities. J Environ Econ Manag 26:271–271

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19:716–723

Aldrich G, Grimsrud K, Thacher J, Kotchen M (2006) Relating environmental attitudes and contingent values: how robust are methods for identifying heterogeneous groups? Environ Resour Econ doi:10.1007/s10640-006-9054-7

Bartholomew D, Leung S (2002) A goodness of fit test for sparse 2p contingency tables. Br J Math Stat Psychol 55(1):1–15

Ben-Akiva M, Morikawa T (1990) Estimation of travel demand models from multiple data sources. In: Koshi M (ed) Transportation and traffic theory. Elsevier, New York

Ben-Akiva M, Walker M, Bernardino A, Gopinath D, Morikawa T, Polydoropoulos A (2002) Integration of choice and latent variable models. In: Mahmassani H (ed) Perpetual motion: travel behavior research opportunities and application challenges. Pergamon, Oxford

Boxall P, Adamowicz W (2002) Understanding heterogeneous preferences in random utility models: a latent class approach. Environ Resour Econ 23(4):421–446

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

Breffle W, Morey E, Rowe R, Waldman D, Wytinck S (1999) 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

Breffle W, Rowe R (2002) Comparing choice question formats for evaluating natural resource tradeoffs. Land Econ 78(2):65–82

Cameron T (1992) Combining contingent valuation and travel cost data for the valuation of nonmarket goods. Land Econ 68(3):302–317

Choi A, Papandrea F, Bennett J (2007) Assessing cultural values: developing an attitudinal scale. J Cult Econ 31(4):311–335

Colombo S, Hanley N (2007) Modelling preference heterogeneity in stated choice data for environmental goods: a comparison of random parameter, covariance heterogeneity and latent class logit models. EAERE Annual Conference, Thessalonica, Greece, pp 27–30

Cunha-e-Sá M, Madureira L, Nunes L, and Otrachshenko V (2010) Protesting or justifying? A latent-class model for contingent valuation with attitudinal data. Working paper

Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete observations. J R Stat Soc Ser B 39:1–38
Dupont D (2004) Do children matter? An examination of gender difference in environmental valuation. Ecol Econ 49:273–286
Eid M, Langeheine R, Diener E (2003) Comparing typological structures across cultures by multigroup latent class analysis - a primer. J Cross Cult Psychol 34(2):195–210
Formann A (2003) Latent class model diagnostics—a review and some proposals. Comput Stat Data Anal 41:549–559
GAUSS (2000) Manual. Aptech Systems Inc, Maple Valley, WA
Greene W, Hensher D (2003) A latent class model for discrete choice analysis: contrasts with mixed logit. Transp Res B Methodol 37(8):681–698
Hensher D, Bradley M (1993) Using stated response choice data to enrich revealed preference discrete choice models. Market Lett 4(2):139–151
Hurvich M, Tsai C (1989) Regression and time series model selection in small samples. Biometrika 76(2):297–307
Johnson R, Swallow S, Bauer D, Anderson C (2003) Resource economics review. Agric Resour Econ Rev 32(1):65–82
Kemperman A, Timmermans H (2006) Preferences, benefits, and park visits: a latent class segmentation analysis. Tour Anal 11(4):221–230
Kritzberg D, Morey E (2008) 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. Working Paper
Lynne G and Rola L (1988) Improving attitude-behavior prediction models with economic variables: farmer actions toward soil conservation. J Soc Psychol 24(1)
Maydeu-Olivares A, Joe H (2005) Limited-and full-information estimation and goodness-of-fit testing in [2. Sup. N] contingency tables: a unified framework. J Am Stat Assoc 100(471):1009–1021
McCutcheon A (1987) Sexual morality, pro-life values, and attitudes toward abortion - a simultaneous latent structure analysis for 1978-1983. Sociol Methods Res 16(2):256–275
McFadden D (1986) The choice theory approach to market research. Marketing Science 5(4):275–297
Morey E, Thacher J, Breffle W (2006) Using angler characteristics and attitudinal data to identify environmental preference classes: a latent-class model. Environ Resour Econ 34(1):91–115
Morey E, Thiene M, De Salvo M, Signorello G (2008) Using attitudinal data to identify latent classes that vary in their preference for landscape preservation. Ecol Econ 68(1–2):536–546
Morikawa T, Ben-Akiva M, McFadden D (2002) Discrete choice models incorporating revealed preference and psychometric data. In: PH F, Montgomery A (eds) Economic models in marketing, vol 16. Elsevier Science, Oxford
Owen A, Videras J (2007) Culture and public goods: the case of religion and the voluntary provision of environmental quality. J Environ Econ Manag 54(2):162–180
Patunru A, Braden J, and Chattopadhyay S (2007) Who cares about environmental stigmas and does it matter? A latent segmentation analysis of stated preferences for real estate. Am J Agric Econ 1–15. doi:10.1111/j.1467-8276.2007.00988
Provencher B, Baerenklau K, Bishop R (2002) A finite mixture logit model of recreational angling with serially correlated random utility. Am J Agric Econ 84(4):1066–1075
R Development Core Team (2005) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org
Reiser M and Lin Y (1999) A goodness of fit test for the latent class model when expected frequencies are small. Sociol Methodol 81–111
Scarpa R, Thiene M (2005) Destination choice models for rock-climbing in the North-East Alps: a latent-class approach investigating intensity of preferences. Land Econ 81(3):426–444
Scarpa R, Willis K, Acutt M (2005) Individual-specific welfare measures for public goods: a latent class approach to residential customers of Yorkshire water. In: Koundouri (ed) Econometrics informing natural resource management. Edward Elgar Publisher, Aldershot
Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464
Smith V (2009) Personal communication on combining data types. personal email
Swait J, Sweeney J (2000) Perceived value and its impact on choice behavior in a retail setting. J Retail Consum Services 7(2):77–88
Thacher J, Morey E, Craighead E (2005) Using patient characteristics and attitudinal data to identify treatment preference groups: a latent-class model. Depress Anxiety 21(2):47–54
Timmins C, Murdock J (2007) A revealed preference approach to the measurement of congestion in travel cost models. J Environ Econ Manag 53:230–249
Train K (2003) Discrete choice methods with simulation, 1st edn.. Cambridge University Press, Cambridge

Vermunt J, Magidson J (2003) Latent GOLD choice. Statistical Innovations, Belmont
Vermunt J, Magidson J (2005) Latent GOLD. Statistical Innovations, Belmont
Ward K, Stedman R, Luloff A, Shortle J, Finley J (2008) Categorizing deer hunters by typologies useful to game managers: A latent-class model. Soc Nat Resour 21(3):215–229
Wedel M, Kamakura W (2000) Market segmentation: conceptual and methodological foundations, second edition. Kluwer, Boston
Yang CC, Yang CC (2007) Separating latent classes by information criteria. J Classif 24:183–203