Combining attitudinal and choice data to estimate preferences and preference heterogeneity: a discrete-choice, random-utility, latent-class model

William Breffle

Stratus Consulting Inc., 1881 9th St # 201, Boulder CO 80302, USA

Edward Morey *

Department of Economics, University of Colorado-Boulder, 256 UCB, Boulder CO 80309, USA

Jennifer Thacher †

Department of Economics, University of New Mexico, MSC05 3060, Albuquerque NM 87131, USA

Abstract

Many preference surveys collect both choice data and attitudinal data, but the majority of econometric models of preferences rely solely on choice data. We suggest how choice data and answers to Likert-scale attitudinal questions can be jointly modeled to more efficiently estimate preferences and preference heterogeneity. This is done in the context of a discrete-choice, random-utility latent-class model, simultaneously estimating: (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.

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

JEL code: C51,Q51,D12

Preference surveys often include attitudinal questions. A common type of attitudinal question assesses the relative importance the individual places on different attributes of a good, indicating how the individual "feels" about those attributes. Consider the following example Likert-scale attitudinal questions. The first is from a survey of Green Bay anglers. The second is from a survey of depressed individuals about the possible side effects of treatment alternatives.

*All are equal authors.

†Corresponding author. Tel.: 505 277 1965; fax: 505 277 9445. *Email address:* jthacher@unm.edu

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" ("Not at all Bothersome .... Very Bothersome")

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

If attitudinal data is an expression of preferences, it can provide significant information about the existence of different preference classes and how preferences vary. We proceed assuming preferences are latent (unobservable) and both choices (actual and hypothetical) and answers to attitudinal questions are manifestations of those unobserved preferences. If correct, there is a strong argument for incorporating attitudinal data into the estimation of preferences and preference heterogeneity (Ben-Akiva et al., 2002; McFadden, 1986; Boxall and Adamowicz, 2002).

A discrete-choice, latent-class model, random-utility model is assumed. Using the E-M (expectation-maximization) algorithm (Dempster et al., 1977), we implement full-information maximum likelihood (FIML) estimation to find the model parameters that maximize the likelihood of observing both the attitudinal and choice data. The application identifies and estimates preference heterogeneity for an environmental amenity. Class membership and the preferences of each class are latent. What is observed are the choices made, the attributes of the alternatives in the choice sets, and the answers to attitudinal questions about those attributes.

1 Background

Latent-class models can model preference heterogeneity among discrete groups without assuming some observable deterministic explanation for that heterogeneity. Denote a discrete-choice latent-class random utility model estimated with only choice data a LC_C model. LC denotes "latent-class" and the subscript(s) denote what type or types of data are used to estimate the model. Examples of LC_C models include Provencher et al. (2002) and Scarpa and Thiene (2005). 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.

Alternatively, one can estimate a latent-class model of preferences using only the answers to attitudinal questions - a LC_A model. Examples of LC_A models include McCutcheon (1987) and Morey et al. (2006). An LC_A model identifies preferences classes, the probabilities of class membership, and probability of answering level s to attitudinal question q, but does not estimate preference parameters, by class, in utility functions.

Here, we build on previous work (Morey et al., 2006) and combine the LC_A model and LC_C model to estimate the LC_{AC} model.

2 A latent-class model of choice and attitudinal data

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

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

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

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

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

Given the independence across responses,

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

where

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

and

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

$\pi_{qs|c}$ is the probability that an individual in class c answers level s to 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 discrete-choice set k, conditional on being a member of class c. Each $\pi_{qs|c}$is estimated as a separate parameter subject to the constraint that $\prod_{s=1}^{S}(\pi_{qs|c}) = 1$.

The $P_{jk|c}$ are functions of the parameters in the class-specific conditional-indirect utility functions, the β_c parameters. That is, $P_{jk|c}$ can be a probit or logit probability of choosing alternative j from discrete-choice set k, conditional on being a member of class c.

Estimation is with the E-M algorithm. Further implementation details are available upon request.

3 Application: preferences of Green Bay anglers

To demonstrate the approach, 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 for PCBs. Anglers answered 15 Likert-scale attitudinal questions and eight stated preference questions of the type: Would you rather fish Green Bay under conditions A or B? 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.

To keep the example simple, we restrictively assume only two classes: $\Pr(class\ 2) = 1 - \Pr(class\ 1)$. More generally, one would estimate the number of classes.

The estimated $\pi_{qs|c}$ can be used to calculate the average Likert-scale response to each of the attitudinal questions, by class. These averages are reported in Table 1. 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. 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 were the catch and FCA levels for Perch and Walleye.

The estimated probability of class membership for the FCA class is 0.404; that is, 40.4% of Green Bay anglers are predicted to be in the FCA class. 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 86% of the sample and effectively 100% for 59% of the sample.

The deterministic part of the conditional-indirect utility function for a Green Bay fishing day is assumed a linear function of the catch times (average time to catch a fish) for the different species, FCA levels, and the cost of a trip to Green Bay. 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. We assumed a logistic probability.

Table 2 reports the estimated preference parameters. All the estimated preference parameters are significant. They also tell the same story as the estimated responses to the attitudinal questions: 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. In terms of willingness-to-pay, the utility parameter estimates indicate that those in the FCA class have a greater WTP for the absence of PCBs than those in the Perch/Walleye class.

3.1 Possible extensions

The model is easily 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.

References

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, ed., 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, 421–446.

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

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, 256–275.

McFadden, D., 1986, The choice theory approach to market research, Marketing Science, 5, 275–297.

Morey, E., J. Thacher, and B. Breffle, 2006, Using angler characteristics and attitudinal data to identify environmental preference classes: A latent-class model, Environmental and Resource Economics, 34, 91–115.

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, 1066–1075.

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

Table 1: Mean Response to Attitudinal Questions by Class

	FCA class	Perch/Walleye class
Importance of:[a]		
Smallmouth bass catch rate	3.36	2.92
Yellow perch catch rate	3.64	3.48
Trout/salmon catch rate	3.13	2.52
Walleye catch rate	3.75	3.34
Smallmouth bass FCA	4.05	2.19
Yellow perch FCA	4.47	3.20
Trout/salmon FCA	4.08	2.46
Walleye FCA	4.56	3.10
Boat launch fee/access fee	3.05	3.09
Amount bothered if PCBs resulted in an FCA of:[b]		
Eat no more than 1 meal a week	3.81	2.46
Eat no more than 1 meal a month	4.34	3.19
Do not eat	4.72	4.00
Willing to pay higher boat launch fees if: [c]		
Catch rate were higher	2.92	2.78
Fish had no PCBs	3.75	3.04
Comparison to Other Sites[d]		
How does the quality of fishing in Green Bay compare to other sites?	3.68	3.84

[a] 1=Not at all Important and 5=Very Important

[b] 1=Not at all Bothersome and 5=Very Bothersome

[c] 1=Strongly Disagree and 5=Strongly Agree

[d] 1=Much Worse and 5=Much Better

Table 2: Preference Parameter Estimates (Est/SE)

Parameters	**FCA Class**	**Perch/Walleye Class**
Pr(class)	0.404	0.596
Time required to catch Yellow Perch	−0.378	−0.623
	(−5.029)	(−10.884)
Time required to catch Walleye	−0.031	−0.04
	(−5.261)	(−9.401)
FCA2	−0.484	−0.03
	(−4.802)	(−0.404)
FCA3	−0.602	−0.084
	(−6.008)	(−1.168)
FCA4	−1.051	−0.178
	(−10.448)	(−2.468)
FCA5	−1.325	−0.345
	(−12.424)	(−4.687)
FCA6	−1.035	−0.178
	(−10.128)	(−2.545)
FCA7	−1.456	−0.451
	(−14.124)	(−6.463)
FCA8	−1.949	−0.639
	(−17.439)	(−8.706)
FCA9	−2.205	−0.699
	(−18.891)	(−9.428)
	Same Parameters for Both Classes	
Fee	−0.48	
	(−15.366)	
Time required to catch Salmon/Trout	−0.028	
	(−7.891)	
Time required to catch Smallmouth Bass	−0.032	
	(−9.341)	