Approximation Approaches to Probabilistic Choice Set Models for Large Choice Set Data

Koichi Kuriyama
University of California, Berkeley

W. M. Hanemann
University of California, Berkeley

Linwood Pendleton
University of Wyoming

DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS AND POLICY
DIVISION OF AGRICULTURE AND NATURAL RESOURCES
UNIVERSITY OF CALIFORNIA AT BERKELEY

WORKING PAPER NO. 967

APPROXIMATION APPROACHES TO PROBABILISTIC CHOICE SET MODELS FOR LARGE CHOICE SET DATA

by

Koichi Kuriyama, W. Michael Hanemann, and Linwood Pendleton

California Agricultural Experiment Station
Giannini Foundation of Agricultural Economics
May 2003

Name and institutional affiliation of paper presenter and co-author(s);
Presenter:
Koichi Kuriyama (Department of Agricultural and Resource Economics, UC Berkeley)
Co-author:
W. Michael Hanemann (Department of Agricultural and Resource Economics, UC Berkeley)
Linwood Pendleton (Department of Economics and Finance, University of Wyoming)

Title of paper;
Approximation Approaches to Probabilistic Choice Set Models for Large Choice Set Data

Address and e-mail address of paper presenter;
Department of Agricultural & Resource Economics, 207 Giannini Hall #3310,
University of California, Berkeley, CA 94720-3310, USA
e-mail: kkuri@waseda.jp

Up to six key words;
random utility model, travel cost method, recreation, environmental valuation

JEL codes; Q26

Approximation Approaches to Probabilistic Choice Set Models for Large Choice Set Data

Koichi Kuriyama, W. Michael Hanemann, and Linwood Pendleton

1 INTRODUCTION

In standard discrete choice models, it is assumed that the choice set can be exogenously specified by analyst. In general, however, individual subjective choice sets are unobserved, therefore, they cannot be imputed with certainty on the basis of observational data. This implies that individual choice behavior would treat choice set generation as a probabilistic process. Haab and Hicks (1997) propose the probabilistic choice set model for recreation data, however, it is difficult to use this model for large choice set data, because it requires calculating the probability of the all subset of the choice set due to is the restriction of the size of the choice set.

This paper investigates two approaches of the probabilistic choice set model for large number of alternatives: randomly drawn choice set and smoothing approximation. McFadden (1978) has shown the estimated parameters by conditional logit model using random draws can give unbiased estimates of the model with the full set. Haab and Hicks (1997) suggest the investigation of the probabilistic choice set approach using randomly drawn choice set, however, our results of theoretical analysis show that this approach may yield biased estimates.

We propose alternative approach, smoothing approximation, which approximates the summation of discrete term in the probabilistic choice set model by smoothing probability function. This approach does not need to calculate of all possible combination of alternatives, therefore, it can analyze the large choice set data. As we show in the theoretical analysis, the smoothing approximation approach can yield unbiased estimates and the error of approximation converges to zero when the probability of alternative availability is close to one or standard deviation of the probability of alternative availability is close to zero.

This paper provides theoretical and empirical analysis of these two approaches. Section 2 provides theoretical analysis of these approaches to the probabilistic choice set model. Section 3 presents Monte Carlo experiments which compare the probabilistic choice set model using randomly drawn choice set and smoothing approximation. Section 4 provides an empirical analysis of this approach using the data of beach recreation trips in Southern California. Finally, section 5 offers concluding comments.

2 THE MODEL

The probabilistic choice set models are based on the following expression:

$$P(i|C,x,\theta) = \sum_{B\subseteq C} P(i|B,x,\theta)P(B|C,x,\theta) \tag{1}$$

where $P(i|C,x,\theta)$ is the probability of choosing alternative i. C is the full set which is the set of all possible alternatives, x is a vector of attributes and θ is a vector of parameters. B is all non-empty subsets of the full set C. $P(i|B,x,\theta)$ is the probability of choosing alternative i given that the choice set is B, and it is assumed to be the conditional logit. $P(B|C,x,\theta)$ is the probability that individual consider the choice set B given C. Following Swait and Ben-Akiva (1987), assume that an alternative is available if a set of relevant constraints specific to that alternative are met and the random components of the elimination criteria across alternatives is independent. Then the choice probabilities is given by

$$P(i|C,x,\theta) = \sum_{B\subseteq C} \frac{1(i\in B)\cdot \exp\left(V_i(x,\theta)\right)}{\sum\limits_{j\in B}\exp\left(V_j(x,\theta)\right)} \cdot \frac{\prod\limits_{j\in B} p_j \prod\limits_{j\in C-B}\left(1-p_j\right)}{1-\prod\limits_{j\in C}\left(1-p_j\right)}. \tag{2}$$

where $1(i\in B)=1$ if $i\in B$ and $1(i\in B)=0$ if $i\notin B$, $V_i(x,\theta)$ is observable component of the utility function of the alternative i, p_j is the probability that alternative j is available to individual.

One of empirical problems in the probabilistic choice set models is the restriction of the size of the choice set. When the full set has j alternatives, there are 2^j-1 non-empty subsets. For example, when the number of alternatives in the full set is 10, the number of possible subset is 1,023. For this restriction, most empirical studies of the probabilistic choice set models use small size of the choice set. Haab and Hicks (1997) propose the investigation of the probabilistic choice set model with randomly drawn choice set, however, we find that this approach may yield biased parameters unless p_j is close to one.

We propose alternative approach which approximates the summation of discrete term $1[i \in B]$ by smoothing function. The approximated probability is as follows:

$$P(i|C, x, \theta) = \frac{p_i \cdot \exp(V_i(x, \theta))}{\sum_{j \in C} p_j \cdot \exp\left(V_j(x, \theta)\right)}. \tag{3}$$

Our theoretical analysis shows that the error of approximation converges to zero if the alternative probabilities p_j are close to one or standard deviation of the distribution function is close to zero.

3 MONTE CARLO EXPERIMENTS

This section provides Monte Carlo experiments to analyze a precision of the probabilistic choice set model using randomly drawn choice set and the smoothing approximation. Our theoretical analysis demonstrates that a precision of these approaches is related to the alternative probabilitiesp_j and standard deviation σ. We compare three cases: (1) p_j is relatively high and σ is relatively low, (2) p_j is relatively low and σ is relatively low, and (3) p_j is relatively low and σ is relatively high. A prediction of our theoretical analysis is as follows: (1) The bias of both approaches should be small, (2) While the bias of the random draw approach should be large, the bias of the smoothing approximation approach should be small, and (3) The bias of both approaches should be large.

Table 1 shows the compensating variation estimated by the Monte Carlo experiments. The number of observations is assumed to be 500. The number of alternatives in full set is assumed to be ten and the number of possible subset is 1,023, therefore, it is difficult to use exact probabilistic choice set model with all possible subset from the full set. We compared three models: the conditional logit, the probabilistic choice set model using randomly drawn choice set, and smoothing approximation model. While the conditional logit and the smoothing approximation model uses full set, the randomly drawn choice set model uses small set (3, 4, and 6 sets).

While the compensating variation estimated by the randomly drawn set model tends to be underestimated, the compensating variation estimated by the smoothing approximation is similar to the assumed value. Even in the case 3, which predicts large error in the approximation, the difference from assumed value is only 6%.

Table 1: Estimated Compensating Variation

The Compensating Variation (CV)						
	Assumed Values	Conditional Logit	Probabilistic Choice Set Model with Randomly Drawn Set			Smoothing Approximation
		10 sets	3 sets	4 sets	6 sets	10 sets
CASE 1	5.8575	6.1338 (0.462)	5.0638 (1.447)	5.5138 (1.376)	5.7090 (0.889)	5.9189 (0.530)
CASE 2	5.7345	6.3855 (0.565)	5.0124 (1.746)	5.2156 (1.510)	5.4485 (1.264)	5.7177 (0.687)
CASE 3	5.5582	6.3289 (0.633)	5.1273 (1.515)	5.3347 (1.389)	5.6088 (1.142)	5.8904 (0.553)
Percentage of Estimated CV / Assumed CV						
	Assumed Values	Conditional Logit	Probabilistic Choice Set Model with Randomly Drawn Set			Smoothing Approximation
		10 sets	3 sets	4 sets	6 sets	10 sets
CASE 1	100%	104.7%	86.5%	94.1%	97.5%	101.0%
CASE 2	100%	111.4%	87.4%	91.0%	95.0%	99.7%
CASE 3	100%	113.9%	92.2%	96.0%	100.9%	106.0%

Note: The values in parentheses are standard errors.

4 APPLICATION

The probabilistic choice set models using randomly drawn choice set and smoothing approximation are applied beach recreation data in Southern California. A panel of Southern California respondents was recruited between November 1999 and January 2000. The number of beach sites of our data is 52 and the number of possible combination of subset is $2^{52}-1=4.5*10^{15}$, therefore, it is difficult to compute exact choice probability of the probabilistic choice set model. Preliminary results of our empirical study support the theoretical and experimental results.

5 CONCLUDING COMMENTS

We consider the probabilistic choice set model for large choice set data. The probabilistic choice set model need to compute the probability of all possible combination of alternatives, therefore, it is necessary to use the approximation approaches when the number of alternative is large. We investigate two alternatives: the randomly drawn choice set and the smoothing approximation approaches. Our Monte Carlo experiments lead the superiority of our proposed smoothing approximation model.

References

[1] Haab, Timothy C. and Hicks, Robert L. (1997) Accounting for Choice Set Endogeneity in Random Utility Models of Recreation Demand, Journal of Environmental Economics and Management, 34(2), (October), 127-147.

[2] McFadden, Daniel. L. (1978) Chapter 5: Modelling the Choice of Residential Location. Karlqvist, L. Lundqvist; Snickars, F. and Weibull, J. (eds.), Spatial Interaction Theory and Planning Models. North-Holland Press, Amsterdam, 75-96.