Marketing Letters 8:3 (1997): 273–286

Modeling Methods for Discrete Choice Analysis

MOSHE BEN-AKIVA
Massachusetts Institute of Technology, Department of Civil and Environmental Engineering, 77 Massachusetts Avenue, Cambridge, MA 02139

DANIEL MCFADDEN
University of California, Department of Economics, 655 Evans Hall, Berkeley, CA 94720

MAKOTO ABE
University of Illinois at Chicago, Department of Marketing, 601 S. Morgan Street, MC 243, Chicago, IL 60607

ULF BÖCKENHOLT
University of Illinois at Urbana-Champaign, Department of Psychology, 603 East Daniel Street, Champaign, IL 61820

DENIS BOLDUC
Université Laval, Département d'économique, Pavillon J.-A. De Sève, Sainte-Foy, Quebec, G1K 7P4, Canada

DINESH GOPINATH
Mercer Management Consulting, 33 Hayden Avenue, Lexington, MA 02173

TAKAYUKI MORIKAWA
Nagoya University, Department of Civil Engineering, Chikusa-ku, Nagoya, 464-01, Japan

VENKATRAM RAMASWAMY
NBD Bancorp, Assistant Professor of Business Administration, University of Michigan, 701 Tappan Street, Ann Arbor, MI 48109

VITHALA RAO
Johnson Graduate School of Management, Cornell University, 529 Malott Hall, Ithaca, NY 14853

DAVID REVELT
University of California at Berkeley, Department of Economics

DAN STEINBERG
San Diego State University, Department of Economics, San Diego, CA 92182

Abstract

This paper introduces new forms, sampling and estimation approaches for discrete choice models. The new models include behavioral specifications of latent class choice models, multinomial probit, hybrid logit, and non-parametric methods. Recent contributions also include new specialized choice based sample designs that permit greater efficiency in data collection. Finally, the paper describes recent developments in the use of simulation methods for model estimation. These developments are designed to allow the applications of discrete choice models to a wider variety of discrete choice problems.

Key words: discrete choice models, multinomial probit, simulation estimation, sample design

1. Introduction

Much of the empirical work in discrete choice modeling has occurred in relatively simple choice contexts. Typically, a single decision maker chooses one alternative among a small well-defined set, and taste homogeneity is assumed among a well-defined population. Expansion of the choice model application domain, however, necessitates deviation from this simple and well-defined arena. In many choice situations, problems of taste heterogeneity and unobservable choice sets are expected and the choice decision itself is fairly complex, as illustrated by the following cases:

1. *Taste heterogeneity*: It is quite common for consumers' preferences to differ for the same set of product or service attributes. In the random utility models this should be expressed by different sets of taste parameters and/or the random component of utility. If there exist substantial taste variations (systematic or idiosyncratic) not adequately captured by the additive random term, then the taste parameters should be specified probabilistically. Two well-known approaches are latent class choice models and the random coefficients models.
2. *Choice from a large number of alternatives*: In many choice applications, numerous alternatives are available. In such cases it is unrealistic to expect the decision maker to compare the many attributes of all available alternatives before choosing. Most researchers agree that a two-stage choice is made: (i) simplified rules screen the many alternatives down to a manageable number, and (ii) through an elaborate process the most preferable alternative is found.
3. *Choice of a collection of multi-attributed items*: In general these collections can be assortments or systems. An assortment is a collection of relatively homogeneous items, such as the set of magazines a consumer subscribes to in a given year. A system is a collection of heterogeneous items, each drawn from a relatively homogeneous group; for example, a work station organized by a work-at-home professional, consisting of a micro-computer, monitor, printer, and fax machine.
4. *Sequential decisions*: Various decisions are made by an individual in a sequential manner. Examples include choice of employment, domicile, transportation mode, and investments. Sequential decisions are dynamic by nature.
5. *Decisions involving externalities*: For example, whether or not to subscribe to an electronic communications network; such a decision may be driven by how many other individuals (including family and friends) have already subscribed or will subscribe. Social issues caused by externality effects (environmental damage, traffic congestion) may complicate the choice decision due to the social dilemma problem. "Green" product purchase, recycling, and motor vehicle use decisions typically involve social externality effects.

Development of a choice model for these types of decisions requires significant adaptation of the standard choice modeling framework and often a new way of thinking about the decision problem. The remainder of this paper introduces approaches for tackling these problems, in particular referencing model forms, sample design, and model estimation.

2. Model forms

2.1. Latent class choice models[1]

In general the Latent Class Choice Model (LCCM) can be described by the following equation:

$$P_n(i) = \sum_{s=1}^{S} P_n(i|s)Q_n(s) \tag{1}$$

where

$P_n(i)$	=	probability of individual n choosing alternative i;
$Q_n(s)$	=	probability of individual n belonging to latent class s;
$P_n(i\|s)$	=	probability of individual n choosing alternative i given n belonging to class s; and
S	=	number of latent classes.

The latent classes may represent:

(a) Different decision protocols adopted by consumers;
(b) Variable choice sets considered by the consumer; and
(c) Segments of the population with varying tastes or sensitivity to product and service attributes.

Note that the latent class choice model assumes that the available data include the choice indicator, attributes of alternatives, and the consumer's socioeconomic and demographic characteristics. Specifically, no data is available on attitudinal and perceptual indicators[2]. We now turn our attention to recent developments in LCCM.

(1) *Form of the class membership probability model*: Gopinath (1994) presents modeling approaches for cases wherein the underlying latent construct could be: (a) categorical, as in the case of latent class characterizing decision protocols (*categorical criterion model*); (b) binary latent class, as in the case of latent class characterizing choice set (*binary criteria model*); and (c) latent class with ordered levels in each dimension, as in the case of latent class characterizing taste variations through ordered levels of

consumer's sensitivity to attributes (*ordinal criteria model*). The presentation emphasizes the various types of class membership models which assign consumers to classes, and which are formulated via a behavioral theory of unobserved criterion functions. The criterion functions may represent unobserved attitudes, consumer's constraints, and decision rules. The basic elements of the binary criteria model were embedded in the works of Swait and Ben-Akiva (1987a, 1987b) and Ben-Akiva and Boccara (1995).

(2) *Dynamic LCCM*: Until now the discussion of LCCM has been limited to a static setting. Dynamic LCCMs have been developed that capture: (1) preference changes over time, and (2) structural dependencies between past and present latent states. Böckenholt and Langeheine (1996) recently demonstrated the relevance of these extensions for predicting choice behavior. Böckenholt and Dillon (1997) present a parsimonious latent class framework for analyzing competitive changes before and after a new product is introduced. Their latent class models are dynamic in nature, allowing the classes' product preferences to change after the new product's introduction, as well as considering the possibility that consumers may switch to a different latent class after product introduction. In a similar vein, Ramaswamy (1996) uses a latent Markov model in capturing the evolution of preference segments before and after new product introductions.

(3) *Latent classes on multiple dimensions*: Ramaswamy et al. (1996) discuss a latent class framework for modeling portfolio choice on two bases that are conceptually distinct, but interdependent. Their latent class model explicitly considers potential interdependence between the bases at the latent class level by specifying the joint distribution of the latent classes on each basis.

(4) *Multistage choice models*: Consumers often use multiple decision making rules. This is especially relevant to choice contexts with many possible alternatives. In such a case, consumers may use a simpler rule to reduce the number of alternatives considered, at which point a more elaborate rule may be employed to choose from the reduced set of alternatives. A number of researchers have investigated this multistage choice decision protocol, and it is well known that misspecified choice sets yield biased parameters (see Stopher (1980), Williams and Ortuzar (1982), Swait and Ben-Akiva (1986)).

The choice set formation process requires a probabilistic model since the consumer's choice set is unobservable. In this sense, the two-stage choice model can be viewed as a latent class model where each class corresponds to a possible choice set. A practical problem of the two-stage model arises from the number of potential choice sets. To alleviate the computational issue Morikawa (1995) derives choice probability for the independent availability model of Swait and Ben-Akiva (1987b) by a pairwise comparison process.

Ben-Akiva and Boccara (1995) propose a methodology for incorporating indicators of availability of alternatives. Using attitudinal data as indicators of latent constructs is a promising method for modeling consumer behavior and suggests a future direction of this research topic.

(5) *LCCM with latent class indicators*: As a first step in incorporating latent class indicators which may be viewed as *attributes* characterizing the class, Gopinath (1994)

elaborates on the different types of specification of the measurement model depending on the characterization of the latent class. These measurement models can be linked to the different class membership models to obtain an extended latent class choice model that is more likely to be empirically identifiable and to yield more efficient estimates.

2.2. Model forms for choice probabilities

(1) *Multinomial Probit Model*: Discrete choice empiricists face a well-known dilemma involving the trade-off between model and behavioral complexity, and model simplicity and ease of estimation. Relaxed assumptions concerning the utility functions' error structure limit the model application to choice contexts with a small number of alternatives. On the other hand, restrictive assumptions concerning the error structure (e.g., MNL model), lead to quick calculation of the choice probabilities for even large choice sets. The most practical method for resolving this trade-off involves incorporating simulation techniques into estimation (see Hajivassiliou and Ruud, 1994, for a discussion of simulation methods). Most approaches exploit a conventional estimation technique, such as the method of moments and the Maximum Likelihood (ML) frameworks, where easy to compute simulators replace the choice probabilities. In situations with very large choice sets, the method of moments often loses some interest because it requires the computation of all choice probabilities instead of only the probability associated with the observed choice, which is the case within the ML framework. However, we note that in recent years, no significant advances have been made in the MNP literature. Some incremental contributions include the application of existing simulation techniques to situations with large choice sets and improved simulators.

(2) *Hybrid logit/Random coefficients Model*: The MNP estimation approach which has received recent attention is the *hybrid logit* (or logit kernel). This approach is characterized by specification of the error term, of which a portion is assumed to be a Gumbel variate. Clearly it has a great potential for application because of its speed and its ability to apply to a variety of choice situations. Bolduc and Ben-Akiva (1991) suggested a hybrid logit specification to estimate MNP models with very large choice sets.

The hybrid logit approach partitions the $J_n \times 1$ utility vector into three independent components[3]:

$$U_n = X_n\beta + Z_n S\eta_n + v_n, n = 1,\ldots, N$$

The first term is the systematic portion of utility. X_n is a $J_n \times K$ matrix of explanatory variables, and β is a $K \times 1$ vector of coefficients. The second term is a random vector with mean zero. Specifically, S is a matrix of coefficients, Z_n is a matrix of individual characteristics or explanatory variables and η_n is a vector of independent random variates. Finally, v_n is assumed to be a Gumbel variate, independent and identically distributed across both alternatives and observations. In addition, v_n and η_n are independent. Since the Gumbel and normal distributions are similar in shape, the logit kernel is a close approximation to the MNP model when η_n is standard normal. Ben-Akiva and Bolduc (1996) consider many different sub-models subsumed in this model form. For our purposes, two specific forms of the utility functions are notable. First, if Z_n is equal to X_n, S is a $K \times K$ lower triangular matrix of coefficients (with possibly some of its elements constrained to zero), and η_n is a $K \times 1$ vector of random variates, then this equation represents a random coefficients model. A degenerate form of the random coefficients model—that nonetheless has an interesting interpretation—is if Z_n is a $J_n \times J_n$ identity matrix, S is a $J_n \times M$ matrix of factor loadings, and η_n is an $M \times 1$ vector of random variates. Under these conditions, the representation of utility corresponds to a factor analytic structure. The logit kernel simulator can be derived easily. By conditioning on η, the likelihood that observation n chooses alternative i is written as:

$$
\begin{aligned}
P_{in} &= \int \frac{e^{X_{in}\beta + Z_{in}S\eta_n}}{\sum_{j \in J_n} e^{X_{jn}\beta + Z_{jn}S\eta_n}} f_\eta d\eta \\
&\approx \frac{1}{R}\sum_{r=1}^{R} \frac{e^{X_{in}\beta + Z_{in}S\eta_{rn}}}{\sum_{j \in J_n} e^{X_{jn}\beta + Z_{jn}S\eta_{rn}}} \\
&= \hat{P}_{in}
\end{aligned}
$$

where R equals the number of draws from the distribution of η. By construction, $\hat{P}_{in}$ (known as the *logit kernel* simulator) is an unbiased estimate of P_{in}. Given that P_{in} is a smooth (twice-differentiable) function of the parameters, it can be used within standard quasi-Newton estimation techniques. For a finite value of R, however, the $\log(\hat{P}_{in})$ is a biased estimator of the $\log(P_{in})$ (see Revelt and Train, 1996). The section on estimation discusses the value of R for which this bias is negligible.

Table 1. Classification of nonparametric methods

Classification	Systematic utility	Random component
1. Nonparametric	nonparametric	distribution-free
2. Semiparametric I	parametric (usually linear-in-parameters)	distribution-free
3. Semiparametric II	nonparametric	parametric distribution

(3) *Nonparametric methods*: Here, the term "nonparametric" refers to issues outside heterogeneity characterized by the latent class or random coefficients models. In the context of random utility models, nonparametric methods can be classified as in Table 1, where the utility is expressed as the sum of systematic utility and a random component.
   (a) Nonparametric: Matzkin (1993) discusses identification conditions. Her work however is purely theoretical and no applications exist on this topic.
   (b) Semiparametric I: This is studied extensively in the econometrics literature. For examples, see the special issue of the *Journal of Econometrics* (1993, v. 58, July, p. 1–274) and *Marketing Letters* (1994, v. 5:4, p. 335–350). However, all empirical studies are limited to a binary response variable, and even many theoretical developments tend to focus on binary response models, as described in Horowitz et al. (1994).
   (c) Semiparametric II: While correct specification of a random component is important for consistent estimation of systematic utility, response shape of systematic utility could suggest interesting behavioral implications. For a binary response, the generalized additive model (GAM) allows estimation of an additive nonparametric utility function when the random component is assumed to be logistically distributed (Hastie and Tibshirani, 1990). Abe (1995) extends the GAM to a multinomial response variable in the framework of a multinomial logit model. He found that the algorithm was simple, estimation was quite robust under violation of the logit distributional assumption through a simulation study, and the computations were quick for typical scanner-panel databases.

In the estimation of a nonparametric systematic utility, note that the bottleneck is typically the number of covariates. When the utility function has many covariates, the data points are sparse over the high dimensional space and the estimation becomes unstable. This is known as the "curse of dimensionality" (Silverman 1986, Stone 1985): as the number of dimensions increase, an exponential increase in sample size is needed to maintain reliable estimation. Therefore, practical implementation necessitates the implementation of some form of constraints (or structure) to the nonparametric specification, such as additivity in GAM.

Future research on nonparametric methods should pursue the following:

- Demonstration of the advantages and benefits of pursuing nonparametric methods through useful and interesting applications in choice contexts;
- Development of convenient estimation algorithms (GAM is one example); and
- Empirical studies to gain better understanding of the operational characteristics of estimators—such as robustness—in practical settings.

3. Sampling and data collection

3.1. Sample design and analysis for discrete panel data

A common survey design in empirical analysis of choice behavior is a panel in which a sequence of discrete responses are observed. Examples are consumers enrolled in supermarket scanner panels, travelers who either provide diaries or retrospectively recall timing and destination of trips, and workers periodically polled for employment status. The classical protocol for enrolling a panel is simple random selection from a target population, and virtually all of the statistical theory for analysis of discrete panel data assumes this protocol. However, in some applications, the recruitment protocol involves stratification by choice. Examples of choice-based recruitment protocols would be enrolling scanner panel members from customers at a specific supermarket chain, selecting travelers by intercepting them at destination sites, and selecting workers based on current employment status.

Choice-based samples can be quite economical in collecting relevant information at low cost, but statistical analysis must be adjusted to correct for sampling effects. Also, choice-based samples are quite susceptible to bias, since minor variations in survey procedure by choice can produce differences that the statistical analysis attributes to differences in the measured attributes of the alternatives. Model coefficients that are specific to alternatives are particularly vulnerable to this bias. In biostatistics, choice-based samples, also called case/control or retrospective designs, are used mostly as a preliminary screening device, and are viewed as insufficiently reliable to be definitive. Only exogenous sample designs, also called prospective designs, are considered reliable enough to provide a firm basis for medical policy (see Feinstein 1979). While there is less accumulated experience with choice-based sampling strategies in consumer behavior applications, the lessons from biostatistics suggest that at the very least, a choice-based design user assumes a heavy burden of ensuring that conclusions drawn from the data are not contaminated by the sample design.

A sample design in which members are recruited by interception at chosen sites, and then followed up as a panel with diaries and/or periodic surveys, is an Intercept and Follow (I&F) design. Such a designed survey was conducted for the State of Montana: anglers were intercepted at selected fishing sites on selected days, and then followed up with diaries over a fishing season (see Morey et al. 1995). The purpose of the study was to estimate a model of angler participation level and site choice as a function of individual and site attributes. In the following discussion, choices are termed “sites,” and recruitment is interpreted as interception by screening at specified sites.

Recruitment by interception at selected sites has several important statistical implications. First, individuals who visit a site frequently have a greater chance of being captured and hence are oversampled relative to those who visit the site infrequently. Further, characteristics associated with more frequent visits, such as male gender or high avidity in the case of fishing, are oversampled relative to characteristics associated with infrequent visits. Second, visits to intercept sites, compared with nonintercept sites and nonparticipation, are more frequent in the sample than they are in the population. Third, if the capture rates vary across sites, then the effects of site attributes on behavior are confounded by variations in apparent site popularity-an artifact of the sample design.

Techniques for adjusting the statistical analysis of choice-based cross-section samples to correct for sampling effects have been investigated extensively in biostatistics and econometrics; see Manski and Lerman (1977), Manski and McFadden (1981), Hsieh, Manski, and McFadden (1985), and Imbens (1992). A critical requirement for these methods is that all alternatives appearing in the behavioral model of choice must be sampled at positive rates.

The literature on choice-based sampling has not treated the problem of analyzing panel data with choice-based recruitment. McFadden (1996) extends the statistical analysis of choice-based cross-section samples to the case of panels with choice-based recruitment, and finds that the simple WESML and CML weighting schemes that work in the cross-section case are no longer valid. McFadden (1996) provides a readily implemented statistical procedure for the case of panels where there is no unobserved heterogeneity that persists through time, and provides solutions for some cases. The results are clearly useful for applications with explicit choice-based recruitment, and are relevant for a broader class of panels where screening protocols, refusals, and attrition create de facto stratification by choice.

Suppose there is *no* persistent heterogeneity, so that choices on different occasions are *statistically independent*, conditioned on observed exogenous and predetermined variables. Then the intercept choice behaves like a cross-section choice-based observation, and can be treated using a WESML or CML criterion. All subsequent choices behave like observations from a random sample. Thus, for example, one can discard the intercept observation and analyze the remaining data as if the sampling were random. However, even in this case the distribution of explanatory variables is not representative of the population, but rather biased toward configurations of explanatory variables that induce more frequent visits to the intercept sites. Then, estimation of any population quantity requires re-weighing of the sample to correct for the oversampling of frequent visitors.

Alternately, suppose there is persistent heterogeneity, perhaps due to taste variations across individuals. Then the simple expedient of discarding the intercept observation and analyzing the panel as if it were randomly recruited does not work. Further, quasi-maximum likelihood methods that are a standard computationally practical tool for analyzing randomly recruited panels with persistent heterogeneity break down under choice-based recruitment. McFadden (1996) shows that an extension of the CML criterion is applicable; implementation requires use of computer-intensive simulation methods.

3.2. Incomplete choice based samples

A common shortcoming of revealed preference (RP) data is incompleteness: information on the choice set from which choices are made may be missing or incomplete, and the sample may lack data on some of the alternatives. In an extreme but familiar case, data may be available for only one of the choice alternatives, as when a firm keeps files on only its own customers. For such data it would appear to be impossible to estimate a choice model. However, Steinberg and Cardell (1992) show that even in such unfavorable circumstances it may be possible to estimate binary discrete choice models by pooling the RP data with commonly available public use data. Thus, for example, an insurance company pooling data on its own customers with census data lacking information on insurance choices could still estimate a binary model of insurance choice. In companion work, Cardell and Steinberg (1993) extend their results to incomplete multinomial data. Here the core argument for the binary logit case is summarized.

Consider three types of data. The first, called a choice-restricted sample, is a random sample available from a stratum defined by the response variable Y. The second, dubbed a supplementary sample, is a random sample with observations drawn from all strata of the choice variable Y, but which contains data only on the covariates X and does *not* record any information on Y. The third is a random sample from the entire population with data on both Y and X. In the latter case, for a sample of size N, the log likelihood function can be written as:

$$\sum_{n=1}^{N} [Y_n \, log \, P_n + (1 - Y_n)\log(1 - P_n)] \tag{2}$$

where P_n is a probability model and n indexes observations.

For model identification, equation (2) requires that data contain observations for both $Y = 1$ and $Y = 0$. The model cannot be estimated when the data is restricted to observations with $Y = 1$ (choice-restricted samples). However, when a supplementary random sample containing information on X but not on the response variable Y is pooled with a choice-restricted sample containing data on both X and Y, it may be possible to estimate a discrete response model.

Consider the artificial case where the sampling rates for the supplementary and choice-restricted samples are both 1; that is, where the entire population is surveyed in the random (supplementary) sample, and all persons with the response variable equal to 1 are surveyed in the choice-restricted sample. Applied to the entire population, equation (2) can be rewritten as

$$\sum_{n=1}^{N} \log(1 - P_n) + \sum_{n:Y_n=1} \log P_n - \sum_{n:Y_n=1} \log(1 - P_n) \tag{3}$$

where N now equals the population size.

If the sampling rates are 1, then equation (3) can be computed from the combined supplementary and choice-restricted samples as follows: (a) The first sum is computed from the supplementary sample alone, accumulating a $\log(1 - P)$ term for each observation. This is equivalent to erroneously treating every observation in this sample as if it had a value of 0 for the response variable. (b) The second sum is computed from the choice-restricted sample only. It accumulates a correct $\log P$ term for observations having a value of 1 for the response variable. (c) The third sum is also computed from the choice-restricted sample only and functions as a correction term.

When the sampling rates are less than 1, a related pseudo-likelihood can be applied to a pooled sample of supplementary and choice-restricted samples. It can be written as

$$\sum_{n=1}^{N} \log(1 - P_n) + \frac{r_0}{r_1} \sum_{n=N+1}^{N+M} \log(P_n/(1 - P_n)) \tag{4}$$

where r_0 = the sampling rate in the supplementary sample; r_1 = the sampling rate in the choice-restricted sample; N = the size of the supplementary sample; and M = the size of the choice-restricted sample. The data are arranged so that the first N observations are from the supplementary sample and the following M observations are from the choice-restricted sample. Aside from some simplification, equation (4) differs from equation (3) only in that contributions to the pseudo-likelihood from the choice-restricted sample are weighted by the ratio of the sampling weights r_0 and r_1. Thus, to implement equation (4), we need to know only either r_0/r_1, or the share of the population with $Y = 1$.

When r_0 and/or r_1 are less than 1, Cardell and Steinberg (1993) show that for a broad class of functions P, the maximization of equation (4) when applied to a pooled sample of supplementary and choice-restricted samples yields consistent estimates.

4. Estimation

We now turn our attention to developments in the estimation of MNP, for which three general approaches exist. A non-classical approach is Gibbs Sampling, or more generally, the method of Monte Carlo Markov Chains. Overall experience with this technique has been mixed. While some find it to be a viable alternative to more classical simulation methods (see Bolduc, Fortin, and Gordon, 1996), others note that it often requires a large number of random draws and, thus, is computationally inefficient. The Gibbs sampling approach has great potential but needs more work before the technique can be generally adopted for discrete choice analysis.

A second approach involves simulating the choice probabilities using GHK (see Börsch-Supan and Hajivassiliou, 1993) or Stern's method (Stern, 1992). Such simulators assume that the random component of utility is multivariate normal, but by partitioning the error term in a convenient manner, the probabilities can be estimated quickly. Many authors have compared the plethora of estimators that fall into the class (e.g., Hajivassiliou, McFadden, and Ruud, 1996), and most find the GHK approach to dominate others. Bolduc (1996) extends Stern's original approach in order to produce a simulator that performs as well as GHK in terms of precision, but computes faster.

A third approach, discussed earlier, is the hybrid logit. This section focuses on some new insights not covered in the several survey papers referenced above, and exclusively concerns the hybrid logit model estimation.

(1) *Identification*: Ben-Akiva and Bolduc (1996) demonstrate that the hybrid logit model is subject to identification problems that do not arise in the conventional MNP frame-

work. In fact, the authors show that this problem appears in any MNP formulation based on an error component decomposition.

(2) *Simulation variance and local maxima*: Simulation variance is simply the additional variation in the parameter estimates due to approximation of the likelihood function by a finite number of random draws. The R variates used in simulating the integral are a function of the initial seed to the random number generator. Two easy solutions for reducing this variance are: increase the value of R or use antithetic variates. From a programming perspective, the first option is easier, but it increases the computation time. The second option requires a trade-off as well, since computing the antithetic variates is more costly than simply drawing pseudo-random numbers.
An interesting feature of the logit kernel estimator is that the variation in the likelihood value seems to be small for even small values of R (e.g. 20), but the variation in the parameter estimates can be as much as 10% of the sampling variance for values of R as large as 500. In other words, large changes in the parameter estimates result in only small changes in the likelihood value. Needless to say, the optimal number of random draws needed to minimize both simulation variance and computation time largely remains an open question.
Local maxima are a common problem among many MNP estimators that use simulation techniques, and as of yet, little research has focused on this issue. The method most often prescribed for overcoming this problem is to estimate the function at different starting values, but since maximization often requires hours (or days!) of computation time, this solution is clearly impractical. Inevitably, researchers must develop new algorithms that are robust to local roots, and one candidate may be simulated annealing (SA) (see Aarts, 1989, for a description). SA has already established a position in computer science and engineering literature as one of the best methods for solving combinatorial optimization problems which often have many local roots, but it has not yet found its way into the statistical packages commonly used by discrete choice empiricists. Research should begin to focus on modifying these algorithms to facilitate efficient estimation of the MNP model.

5. Conclusion

Increasing the complexity of discrete choice models often requires complex computations and data collection procedures to estimate the choice probabilities, and as in the MNP case, issues of computational tractability quickly become important. Simulated maximum likelihood is a practical method that often works well in resolving this problem, but the breakdown points of these estimators are ill understood. Clearly future work must identify when problems such as simulation variance and local maxima are important, and develop methods to deal with them. While the speed of the average computer has increased, so also have the ambitions of modern discrete choice empiricists.

Notes

1. DeSarbo et al. (1997) discusses latent class choice models with reference to capturing heterogeneity in consumer response models. This section provides an overview of recent developments not addressed in DeSarbo et al. (1997).
2. In this respect the term "latent class choice model" is somewhat of a misnomer since latent class models utilize the existence of indicators of latent classes. It must be noted that the choice itself can be construed as an indicator of both the underlying preference and the associated latent class. The ideas developed within latent class choice models are extended in Gopinath (1994) wherein models incorporating attitudinal and perceptual indicators, referred to as latent structure choice models, are formulated.
3. Throughout this section, let J_n, denote the total number of alternatives faced by individual n, K denote the total number of explanatory variables in the systematic portion of utility, and N the sample size.

References

Aarts, E. H. L. (1989). *Simulated Annealing and Boltzmann Machines.* New York: Wiley.

Abe, M. (1995). "A Generalized Additive Model for Discrete Choice Data," Working paper, The University of Illinois at Chicago.

Ben-Akiva, M., and B. Boccara. (1995). "Discrete Choice Models with Latent Choice Sets," *International Journal of Research in Marketing* 12, 9–24.

Ben-Akiva, M., and D. Bolduc. (1996). "Multinomial Probit with a Logit Kernel and General Parametric Specification of the Covariance Structure," Working paper, Département d'économique, Université Laval.

Böckenholt, U., and W. R. Dillon. (1996). "Some New Methods for an Old Problem: Modeling Preference Changes and Competitive Market Structures in Pre-Test Market Data," *Journal of Marketing Research* 34, 130-142.

Böckenholt, U., and R. Langeheine. (1996). "Latent Change in Recurrent Choice Data," *Psychometrika* 61, 285–302.

Bolduc, D. (1996). "A Fast Maximum Simulated Likelihood Estimation Technique for MNP Models," Working paper, Département d'économique, Université Laval.

Bolduc, D., and M. Ben-Akiva. (1991). "A Multinomial Probit Formulation for Large Choice Sets," *Proceedings of the 6th IATBR Conference*, Québec, May.

Bolduc, D., B. Fortin, and S. Gordon. (1996). "Multinomial Probit Estimation of Spatially Interdependent Choices", forthcoming *International Regional Science Review*.

Börsch-Supan, A., and V. Hajivassiliou. (1993). "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models," *Journal of Econometrics* 58, 347–368.

Cardell, N. S., and D. Steinberg. (1993). "Optimal Design of Stated preference Choice Experiments," Working Paper, Department of Economics, San Diego State University.

DeSarbo, W. et al. (1997). "Representing Heterogeneity in Consumer Response Models," *Marketing Letters* (this issue).

Feinstein, A. (1979). "Methodological Problems and Standards in Case-Control Research," *Journal of Chronic Disease* 32, 35–41.

Gopinath, D. (1994). "Modeling Heterogeneity in Discrete Choice Processes: Application to Travel Demand," Ph.D. thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Hajivassiliou, V., D. McFadden, and P. Ruud. (1996). "Simulation of Multivariate Normal Rectangle Probabilities and their Derivatives: Theoretical and Computational Results," forthcoming, *Journal of Econometrics*.

Hajivassiliou, V., and P. Ruud. (1994). "Classical Estimation Methods for LDV Models using Simulation," *Handbook of Econometrics* IV, R. Engle and D. McFadden, eds., Elsevier Science B.V.

Hastie, T., and R. Tibshirani. (1990). *Generalized Additive Models*, Monographs on Statistics and Applied Probability. New York: Chapman and Hall.

Horowitz, J. L. et al. (1994). "Advances in Random Utility Models: Report of the Workshop on Advances in Random Utility Models," *Marketing Letters* 5, 311–322.

Hsieh, D., C. Manski, and D. McFadden. (1985). "Estimation of Response Probabilities from Augmented Retrospective Observations," *Journal of the American Statistical Association* 80, 651–662.

Imbens, G. (1992). "A Method of Moments Estimator for Discrete Choice Models with Choice Based Sampling," *Econometrica* 60, 1187–1214.

Manski, C., and S. Lerman. (1977). "Estimation of Choice Probabilities from Choice Based Samples," *Econometrica* 45, 1977–88.

Manski, C., and D. McFadden. (1981). "Alternative Estimators and Sample Designs for Discrete Choice Analysis," in C. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data*, MIT Press.

McFadden, D. (1996). "On the Analysis of 'Intercept and Follow' Surveys," Working Paper, University of California, Berkeley.

Matzkin, R. L. (1993). "Nonparametric Identification and Estimation of Polychotomous Choice Models," *Journal of Econometrics* 58, 137–168.

Morey, E., R. Rowe, and D. Waldman. (1995). "Assessment of Damages to Anglers and Other Recreators from Injuries to the Upper Clark Fork River Basin," Technical Report, State of Montana Natural Resource Damage Litigation Program.

Morikawa, T. (1995). "A Hybrid Probabilistic Choice Set Model with Compensatory and Noncompensatory Choice Rules," *Proceedings of the 7th World Conference on Transport Research*, Sydney, in press.

Ramaswamy, V., A. Chatterjee, and S. Cohen. (1996). "Joint Segmentation on Distinct Interdependent Bases with Categorical Data," *Journal of Marketing Research* 33 (August), 337–350.

Ramaswamy, V. (1996). "Evolutionary Preference Segmentation with Panel Survey Data: An Application to New Products," forthcoming *International Journal of Research in Marketing*.

Revelt, D. and K. Train. (1996). "Incentives for Appliance Efficiency: Random Parameters Logit Model of Households' Choices," Working paper, Department of Economics, University of California, Berkeley.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability. New York: Chapman and Hall.

Steinberg, D., and N. S. Cardell. (1992). "Estimating Logistic Regression Models When the Dependent Variable Has No Variance," *Communications in Statistics* 21, 423–450.

Stern, S. (1992). "A Method of Smoothing Simulated Moments of Discrete Probabilities in Multinomial Probit Models," *Econometrica* 60, 943–952.

Stone, C. J. (1985). "Additive Regression and Other Nonparametric Models," *Annals of Statistics* 13, 689–705.

Stopher, P. (1980). "Captivity and Choice in Travel Behavior Models," *Transportation Journal of ASCE* 106, 427–435.

Swait, J., and M. Ben-Akiva. (1986). "Analysis of the Effects of Captivity on Travel Time and Cost Elasticities," *Behavioural Research for Transport Policy*: *The 1985 Internat'l Conference on Travel Behavior*. The Netherlands: VNU Science Press.

Swait, J., and M. Ben-Akiva. (1987a). "Incorporating Random Constraints in Discrete Models of Choice Set Generation," *Transportation Research* 21B, 91–102.

Swait, J., and M. Ben-Akiva. (1987b). "Empirical Test of a Constrained Discrete Choice Model: Mode Choice in Sao-Paulo, Brazil," *Transportation Research* 21B, 103–115.

Williams, H. and J. Ortuzar. (1982). "Behavioral Theories of Dispersion and Misspecification of Travel Demand Models," *Transportation Research B* 16B, 167–219.