

Discrete Choice Methods with Simulation

Kenneth Train
University of California, Berkeley
National Economic Research Associates

Version dated March 8, 2002

Publisher: Cambridge University Press
Scheduled publication date: Autumn 2002.

Please contact me with any corrections, comments,
and suggestions, at train@econ.berkeley.edu
or 415-291-1023.

to
Daniel McFadden
and
in memory of
Kenneth Train, Sr.

This copy is made available for use by individuals for their personal research and study. Permission is not granted to use any part of this work for any other purpose whatsoever without the written consent of Cambridge University Press.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Choice Probabilities and Integration	3
1.3	Outline of book	7
1.4	Topics not covered	8
1.5	A couple notes	12
I	Behavioral Models	13
2	Properties	15
2.1	Overview	15
2.2	The choice set	15
2.3	Derivation of choice probabilities	19
2.4	Specific models	22
2.5	Identification of choice models	24
2.5.1	Only differences in utility matter	24
2.5.2	The overall scale of utility is irrelevant	29
2.6	Aggregation	35
2.6.1	Sample enumeration	37
2.6.2	Segmentation	38
2.7	Forecasting	39
2.8	Recalibration of constants	39
3	Logit	41
3.1	Choice probabilities	41
3.2	The scale parameter	48
3.3	Power and limitations of logit	50
3.3.1	Taste variation	50

3.3.2	Substitution patterns	53
3.3.3	Panel data	59
3.4	Non-linear representative utility	61
3.5	Consumer Surplus	64
3.6	Derivatives and Elasticities	67
3.7	Estimation	70
3.7.1	Exogenous sample	70
3.7.2	Choice-based samples	76
3.8	Goodness of Fit and Hypothesis Testing	78
3.8.1	Goodness of fit	78
3.8.2	Hypothesis testing	80
3.9	Case Study	81
3.10	Derivation of Logit Probabilities	85
4	GEV	87
4.1	Introduction	87
4.2	Nested logit	88
4.2.1	Substitution patterns	88
4.2.2	Choice probabilities	90
4.2.3	Decomposition into two logits	93
4.2.4	Estimation	96
4.2.5	Equivalence of nested logit formulas	98
4.3	Three-Level Nested Logit	98
4.4	Overlapping Nests	101
4.4.1	Paired combinatorial logit	102
4.4.2	Generalized nested logit	104
4.5	Heteroskedastic Logit	105
4.6	The GEV family	106
5	Probit	111
5.1	Choice probabilities	111
5.2	Identification	114
5.3	Taste variation	121
5.4	Substitution patterns/non-IIA	123
5.5	Panel data	126
5.6	Simulation of the choice probabilities	130
5.6.1	Accept-reject simulator	131
5.6.2	Smoothed A-R simulators	136
5.6.3	GHK simulator	139

6	Mixed Logit	153
6.1	Choice probabilities	153
6.2	Random coefficients	156
6.3	Error-components	158
6.4	Substitution patterns	160
6.5	Approximation to any random utility model	161
6.6	Simulation	163
6.7	Panel data	165
6.8	Case Study	168
7	Variations on a Theme	173
7.1	Introduction	173
7.2	SP/RP	174
7.3	Ranked Data	178
7.3.1	Standard and mixed logit	179
7.3.2	Probit	181
7.4	Ordered Responses	182
7.4.1	Multiple ordered responses	186
7.5	Contingent Valuation	188
7.6	Mixed Models	190
7.6.1	Mixed Nested Logit	191
7.6.2	Mixed Probit	192
7.7	Dynamic optimization	193
7.7.1	Two-periods, no uncertainty about future impacts	195
7.7.2	Multiple periods	199
7.7.3	Uncertainty about future impacts	203
II	Estimation	209
8	Numerical Maximization	211
8.1	Motivation	211
8.2	Notation	212
8.3	Algorithms	213
8.3.1	Newton-Raphson	213
8.3.2	BHHH	220
8.3.3	BHHH-2	223
8.3.4	Steepest Ascent	224
8.3.5	DFP and BFGS	225
8.4	Convergence criterion	226

8.5	Local versus global maximum	227
8.6	Variance of the Estimates	228
8.7	Information Identity	229
9	Drawing from Densities	233
9.1	Introduction	233
9.2	Random Draws	234
9.2.1	Standard normal and uniform	234
9.2.2	Transformations of standard normal	234
9.2.3	Inverse cumulative for univariate densities	235
9.2.4	Truncated univariate densities	236
9.2.5	Choleski transformation for multivariate normals	236
9.2.6	Accept-reject for truncated multivariate densities	238
9.2.7	Importance sampling	239
9.2.8	Gibbs sampling	241
9.2.9	Metropolis-Hastings algorithm	242
9.3	Variance Reduction	244
9.3.1	Antithetics	245
9.3.2	Systematic sampling	248
9.3.3	Halton sequences	252
9.3.4	Randomized Halton draws	263
9.3.5	Scrambled Halton draws	265
9.3.6	Other procedures	269
10	Simulation-Assisted Estimation	271
10.1	Motivation	271
10.2	Definition of estimators	273
10.2.1	Maximum simulated likelihood: MSL	273
10.2.2	Method of simulated moments: MSM	274
10.2.3	Method of simulated scores: MSS	277
10.3	The central limit theorem	280
10.4	Traditional estimators	282
10.5	Simulation-based estimators	285
10.5.1	Maximum simulated likelihood	290
10.5.2	Method of simulated moments	291
10.5.3	Method of simulated scores	292
10.6	Numerical Solution	293

11 Individual-Level Parameters	295
11.1 Introduction	295
11.2 Derivation of conditional distribution	298
11.3 Implications of estimation of θ	301
11.4 Monte Carlo illustration	304
11.5 Average conditional distribution	307
11.6 Case Study	307
11.7 Discussion	319
12 Bayesian Procedures	321
12.1 Introduction	321
12.2 Overview of Bayesian concepts	324
12.2.1 Bayesian properties of $\bar{\theta}$	326
12.2.2 Classical properties of $\bar{\theta}$: The Bernstein-von Mises theorem	327
12.3 Simulation of posterior mean	331
12.4 Drawing from the posterior	333
12.5 Posteriors for Normals	335
12.5.1 Result A: Unknown mean, known variance	335
12.5.2 Result B: Unknown variance, known mean	337
12.5.3 Unknown mean and variance	340
12.6 Hierarchical Bayes for mixed logit	340
12.6.1 Succinct restatement	345
12.7 Case Study: Choice of energy supplier	346
12.7.1 Independent normal coefficients	347
12.7.2 Multivariate normal coefficients	349
12.7.3 Fixed coefficients for some variables	350
12.7.4 Lognormals	352
12.7.5 Triangulars	353
12.7.6 Summary of results	355
12.8 Bayesian procedures for probit models	355

Chapter 1

Introduction

1.1 Motivation

When I wrote my first book, *Qualitative Choice Analysis*, in the mid 1980's, the field had reached a critical juncture. The break-through concepts that defined the field had been made. The basic models — mainly logit and nested logit — had been introduced, and the statistical and economic properties of these models had been derived. Applications had proven successful in many different areas, including transportation, energy, housing, and marketing — to name only a few.

The field is at a similar juncture today for a new generation of procedures. The first-generation models contained important limitations that inhibited their applicability and realism. These limitations were well recognized at the time, but ways to overcome them had not yet been discovered. Over the past twenty years, tremendous progress has been made, leading to what can only be called a sea-change in the approach and methods of choice analysis. The early models have now been supplemented by a variety of more powerful and more flexible methods. The new concepts have arisen gradually, with researchers building on the work of others. However, in a sense, the change has been more like a quantum leap than a gradual progression. The ways that researchers think about, specify, and estimate their models has changed. Importantly, a kind of consensus, or understanding, seems to have emerged about the new methodology. Among researchers working in the field, a definite sense of purpose and progress prevails.

My purpose in writing this new book is to bring these ideas together, in a form that exemplifies the unity of approach that I feel

has emerged, and in a format that makes the methods accessible to a wide audience. The advances have mostly centered on simulation. Essentially, simulation is the researcher's response to the inability of computers to perform integration. Stated more precisely, simulation provides a numerical approximation to integrals, with different methods offering different properties and being applicable to different kinds of integrands.

Simulation allows estimation of otherwise intractable models. Practically any model can be estimated by some form of simulation. The researcher is therefore freed from previous constraints on model specification – constraints that reflected mathematical convenience rather than the economic reality of the situation. This new flexibility is a tremendous boon to research. It allows more realistic representation of the hugely varied choice situations that arise in the world. It enables the researcher to obtain more information from any given dataset and, in many cases, allows previously unapproachable issues to be addressed.

This flexibility places a new burden on the researcher. First, the methods themselves are more complicated than earlier procedures and utilize many concepts and procedures that are not covered in standard econometrics courses. Understanding the various techniques — their advantages and limitations, and the relations among them — is important when choosing the appropriate method in any particular application and for developing new methods when none of the existing models seems right. The purpose of this book is to assist readers along this path.

Second, to implement a new method, or a variant on an old method, the researcher needs to be able to program the procedure into computer software. This means that the researcher will often need to know how maximum likelihood and other estimation methods work from a computational perspective, how to code specific models, and how to take existing code and change it to represent variations in behavior. Some models, such as mixed logit and pure probit in addition of course to standard logit, are available in commercially available statistical packages. In fact, code for these and other models, as well as manuals and sample data, are available (free) at my website <http://elsa.berkeley.edu/~train>. Whenever appropriate, researchers should use available codes rather than writing their own. However, the true value of the new approach to choice modeling is the ability to

create tailor-made models. The computational and programming steps that are needed to implement a new model are usually not difficult. An important goal of the book is to teach these skills as an integral part of the exposition of the models themselves. I personally find programming to be extremely valuable pedagogically. The process of coding a model helps me to understand how exactly the model operates, the reasons and implications of its structure, what features constitute the essential elements that cannot be changed while maintaining the basic approach, and which features are arbitrary and can easily be changed. I imagine other people learn this way too.

1.2 Choice Probabilities and Integration

To focus ideas, I will now establish the conceptual basis for discrete choice models and show where integration comes into play. An agent (i.e., person, firm, decision-maker) faces a choice, or a series of choices over time, among a set of options. For example, a customer chooses which of several competing products to buy; a firm decides which technology to use in production; a student chooses which answer to give on a multiple choice test; a survey respondent chooses an integer between 1 and 5 on a Lickert-scale question; a worker chooses whether to continue working each year or retire. Denote the outcome of the decision(s) in any given situation as y , where y indicates the chosen option or sequence of options. We assume for the purposes of this book that the outcome variable is discrete in that it takes a countable number of values. Many of the concepts that we describe are easily transferable to situations where the outcome variable is continuous. However, notation and terminology is different with continuous outcome variables than discrete ones. Also, discrete choices generally reveal less information about the choice process than continuous-outcome choices, such that the econometrics of discrete choice is usually more challenging.

Our goal is to understand the behavioral process that leads to the agent's choice. We take a causal perspective. There are factors that collectively determine, or cause, the agent's choice. Some of these factors are observed by the researcher and some are not. The observed factors are labeled x and the unobserved factors ε . The factors relate to the agent's choice through a function $y = h(x, \varepsilon)$. This function is called the behavioral process. It is deterministic in the sense that given x and ε , the choice of the agent is fully determined.

Since ε is not observed, the agent's choice is not deterministic and cannot be predicted exactly. Instead, the *probability* of any particular outcome is derived. The unobserved terms are considered random with density $f(\varepsilon)$. The probability that the agent chooses a particular outcome from the set of all possible outcomes is simply the probability that the unobserved factors are such that the behavioral process results in that outcome: $P(y|x) = \text{Prob}(\varepsilon \text{ s.t. } h(x, \varepsilon) = y)$.

We can express this probability in a more useable form. Define an indicator function $I[h(x, \varepsilon) = y]$ that takes the value of 1 when the statement in brackets is true and 0 when the statement is false. That is, $I[\cdot] = 1$ if the value of ε , combined with x , induces the agent to choose outcome y and 0 if the value of ε , combined with x , induces the agent to choose some other outcome. Then the probability that the agent chooses outcome y is simply the expected value of this indicator function, where the expectation is over all possible values of the unobserved factors:

$$\begin{aligned} P(y|x) &= \text{Prob}(I[h(x, \varepsilon) = y] = 1) \\ &= \int I[h(x, \varepsilon) = y] f(\varepsilon) d\varepsilon. \end{aligned} \quad (1.1)$$

Stated in this form, the probability is an integral — specifically an integral of an indicator for the outcome of the behavioral process over all possible values of the unobserved factors.

To calculate this probability, the integral must be evaluated. There are three possibilities.

Complete closed-form expression

For certain specifications of h and f , the integral can be expressed in closed form. In these cases, the choice probability can be calculated exactly from the closed-form formula. For example, consider a binary logit model of whether or not a person takes a given action, such as buying a new product. The behavioral model is specified as follows. The person would obtain some net benefit or utility from taking the action. This utility, which can be either positive or negative, consists of a part that is observed by the researcher, $\beta'x$ where x is a vector of variables and β is a vector of parameters, and a part that is not observed, ε : $U = \beta'x + \varepsilon$. The person takes the action only if the utility is positive, that is, only if doing so provides a net benefit. The

probability that the person takes the action, given what the researcher can observe, is therefore $P = \int I[\beta'x + \varepsilon > 0]f(\varepsilon)d\varepsilon$ where f is the density of ε . Assume that ε is distributed logistic, such that its density is $f(\varepsilon) = e^{-\varepsilon}/(1 + e^{-\varepsilon})^2$ with cumulative distribution $F(\varepsilon) = 1/(1 + e^{-\varepsilon})$. Then the probability of the person taking the action is:

$$\begin{aligned}
 P &= \int I[\beta'x + \varepsilon > 0]f(\varepsilon)d\varepsilon \\
 &= \int I[\varepsilon > -\beta'x]f(\varepsilon)d\varepsilon \\
 &= \int_{\varepsilon=-\beta'x}^{\infty} f(\varepsilon)d\varepsilon \\
 &= 1 - F(-\beta'x) = 1 - \frac{1}{1 + e^{\beta'x}} \\
 &= \frac{e^{\beta'x}}{1 + e^{\beta'x}}
 \end{aligned}$$

For any x , the probability can be calculated exactly as $P = \exp(\beta'x)/(1 + \exp(\beta'x))$.

Other models also have closed-form expressions for the probabilities. Multinomial logit (in chapter 3), nested logit (chapter 4), and ordered logit (chapter 7) are prominent examples. The methods that I described in my first book and that served as the basis for the first wave of interest in discrete choice analysis relied almost exclusively on models with closed-form expressions for the choice probabilities. In general, however, the integral for probabilities cannot be expressed in closed form. More to the point, restrictions must be placed on the behavioral model h and the distribution of random terms f in order for the integral to take a closed form. These restrictions can make the models unrealistic for many situations.

Complete simulation

Rather than solve the integral analytically, it can be approximated through simulation. Simulation is applicable in one form or another to practically any specification of h and f . Simulation relies on the fact that integration over a density is a form of averaging. Consider the integral $\bar{t} = \int t(\varepsilon)f(\varepsilon)d\varepsilon$ where $t(\varepsilon)$ is a statistic based on ε which has density $f(\varepsilon)$. This integral is the expected value of t over all possible values of ε . This average can be approximated in an intuitively

straightforward way. Take numerous draws of ε from its distribution f , calculate $t(\varepsilon)$ for each draw, and average the results. This simulated average is an unbiased estimate of the true average. It approaches the true average as more and more draws are used in the simulation.

This concept of simulating an average is the basis for all simulation methods, at least all of those that we consider in this book. As given in eq. (1.1), the probability of a particular outcome is an average of the indicator $I(\cdot)$ over all possible values of ε . The probability, when expressed in this form, can be simulated directly as follows: (1) Take a draw of ε from $f(\varepsilon)$. Label this draw ε^1 , where the superscript denotes that it is the first draw. (2) Determine whether $h(x, \varepsilon^1) = y$ with this value of ε . If so, create $I^1 = 1$, otherwise set $I^1 = 0$. (3) Repeat steps 1 and 2 many times, for a total of R draws. The indicator for each draw is labeled I^r for $r = 1, \dots, R$. (4) Calculate the average of the I^r 's. This average is the simulated probability: $\check{P}(i | x) = \frac{1}{R} \sum_{r=1}^R I^r$. It is the proportion of times that the draws of the unobserved factors, when combined with the observed variables x , result in outcome y .

As we will see in the chapters to follow, this simulator, while easy to understand, has some unfortunate properties. Choice probabilities can often be expressed as averages of other statistics, rather than the average of an indicator function. The simulators based on these other statistics are calculated analogously, by taking draws from the density, calculating the statistic and averaging the results. Probit (in chapter 5) is the most prominent example of a model estimated by complete simulation. Various methods of simulating the probit probabilities have been developed based on averages of various statistics over various (related) densities.

Partial simulation/partial closed form

So far we have provided two polar extremes: either solve the integral analytically or simulate it. In many situations, it is possible to do some of both.

Suppose the random terms can be decomposed into two parts labeled ε_1 and ε_2 . Let the joint density of ε_1 and ε_2 be $f(\varepsilon) = f(\varepsilon_1, \varepsilon_2)$. The joint density can be expressed as the product of a marginal and a conditional density: $f(\varepsilon_1, \varepsilon_2) = f(\varepsilon_2 | \varepsilon_1) \cdot f(\varepsilon_1)$. With this decomposition, the probability in eq. (1.1) can be expressed as:

$$P(y|x) = \int I[h(x, \varepsilon) = y] f(\varepsilon) d\varepsilon$$

$$= \int_{\varepsilon_1} \left[\int_{\varepsilon_2} I[h(x, \varepsilon_1, \varepsilon_2) = y] f(\varepsilon_2 | \varepsilon_1) d\varepsilon_2 \right] f(\varepsilon_1) d\varepsilon_1.$$

Now suppose that a closed form exists for the integral in large brackets. Label this formula $g(\varepsilon_1) \equiv \int_{\varepsilon_2} I[h(x, \varepsilon_1, \varepsilon_2) = y] f(\varepsilon_2 | \varepsilon_1) d\varepsilon_2$, which is conditional on the value of ε_1 . The probability then becomes $P(y|x) = \int_{\varepsilon_1} g(\varepsilon_1) f(\varepsilon_1) d\varepsilon_1$. If a closed-form solution does not exist for this integral, then it is approximated through simulation. Note that it is simply the average of g over the marginal density of ε_1 . The probability is simulated by taking draws from $f(\varepsilon_1)$, calculating $g(\varepsilon_1)$ for each draw, and averaging the results.

This procedure is called “convenient error partitioning” (Train, 1998). The integral over ε_2 given ε_1 is calculated exactly while the integral over ε_1 is simulated. There are clear advantages to this approach over complete simulation. Analytic integrals are both more accurate and easier to calculate than simulated integrals. It is useful, therefore, when possible, to decompose the random terms such that some of them can be integrated analytically, even if the rest must be simulated. Mixed logit (in chapter 6) is a prominent example of a model that uses this decomposition effectively. Other examples include Gourieroux and Monfort’s (1993) binary probit model on panel data and Bhat’s (1999) analysis of ordered responses.

1.3 Outline of book

Discrete choices analysis consists of two interrelated tasks: specification of the behavioral model and estimation of the parameters of that model. Simulation plays a part in both tasks. Simulation allows the researcher to approximate the choice probabilities that arise in the behavioral model. As we have stated, the ability to use simulation frees the researcher to specify models without the constraint that the resulting probabilities have a closed form. Simulation also enters the estimation task. The properties of an estimator, such as maximum likelihood, can change when simulated probabilities are used instead of the actual probabilities. Understanding these changes, and mitigating any ill effects, is important for a researcher. In some cases, such as with Bayesian procedures, the estimator itself is an integral over a density (as opposed to the choice probability being an integral). Simulation allows these estimators to be implemented even when the integral that defines the estimator does not take a closed form.

The book is organized around these two tasks. Part I describes behavioral models that have been proposed to describe the choice process. The chapters in this section move from the simplest model, logit, to progressively more general and consequently more complex models. A chapter is devoted to each of: logit, the family of generalized extreme value models (whose most prominent member is nested logit), probit, and mixed logit. This part of the book ends with a chapter entitled “variations on a theme,” which covers a variety of models that build upon the concepts in the previous chapters. The point of this chapter is more than simply to introduce various new models. The chapter illustrates the underlying concept of the book, namely, that researchers need not rely on the few common specifications that have been programmed into software but can design models that reflect the unique setting, data and goals of their project, writing their own software and using simulation as needed.

Part II describes estimation of the behavioral models. Numerical maximization is covered first, since most estimation procedures involve maximization of some function, such as the log-likelihood function. We then describe procedures for taking draws from various kinds of densities, which is the basis for simulation. This chapter also describes different kinds of draws, including antithetic variants and quasi-random sequences, that can provide greater simulation accuracy than independent random draws. We then turn to simulation-assisted estimation, looking first at classical procedures, including maximum simulated likelihood, method of simulated moments, and method of simulated scores. Finally, we examine Bayesian estimation procedures, which use simulation to approximate moments of the posterior distribution. The Bayesian estimator can be interpreted from either a Bayesian or classical perspective and has the advantage of avoiding some of the numerical difficulties associated with classical estimators. The power that simulation provides when coupled with Bayesian procedures makes this chapter a fitting finale for the book.

1.4 Topics not covered

I feel it is useful to say a few words about what the book does not cover. There are several topics that could logically be included but are not. One is the branch of empirical industrial organization that involves estimation of discrete choice models of consumer demand on market-

level data. Customer-level demand is specified by a discrete choice model, such as logit or mixed logit. This formula for customer-level demand is aggregated over consumers to obtain market-level demand functions that relate prices to shares. Market equilibrium prices are determined as the interaction of these demand functions with supply, based on marginal costs and the game that the firms are assumed to play. Berry (1994) and Berry, Levinsohn and Pakes (1995) developed methods for estimating the demand parameters when the customer-level model takes a flexible form such as mixed logit. The procedure has been implemented in numerous markets for differentiated goods, such as ready-to-eat cereals (Nevo, 2001).

I have decided not to cover these procedures, despite their importance, because they are very different, in motivation, data, and methodology, than all the other material in the book. They use market-level data rather than customer-level. The observed demand is a set of continuous variables, representing market shares, rather than discrete variables representing individual customers' choices. Prices are endogenous, determined by the interaction of demand and supply. In contrast, prices faced by an individual customer are not affected by that customer's choice; prices are therefore exogenous from market forces when modeling customer-level demand. Other types of endogeneity may arise with customer-level data, but the endogeneity from demand/supply interaction, which market level data embody, does not.

The focus of the market-level methods, and the insight provided by the contributors to these methods, are different from those for customer-level data. Instruments become paramount since price is endogenous. The dominant questions, and the basis for much of the discussion in the field, concern the selection of instruments: what instruments are appropriate and what identifying assumptions are necessary with any given instruments? These questions do not arise with customer-level data (unless, of course, some other form of endogeneity is present). The dominant methodological issue is also different. Since instruments are necessary, the methodological question becomes: how can market share data, which arise from demand functions in which the unobserved factors enter nonlinearly, be transformed to allow the use of instrumental variables estimation, which operates effectively when the unobserved factors enter linearly? Berry *et al.*'s contribution was in determining the appropriate transformation and how to perform it. This issue does not arise in customer-level data. For interested read-

ers, Nevo (2000) provides a useful guide to the procedures, including programming instructions and an discussion of the appropriate choice of instruments.

A second area that this book does not cover is discrete/continuous models. These models arise when a regression equation for a continuous variable is related in any of several ways to a discrete choice. The most prominent situations are the following.

1. The continuous variable depends on a discrete explanatory variable that is determined endogenously with the dependent variable. For example, consider an analysis of the impact of job-training programs on wages. A regression equation is specified with wages as the dependent variable and a dummy variable for whether the person participated in a job-training program. The coefficient of the participation dummy indicates the impact of the program on wages. The situation is complicated, however, by the fact that participation is voluntary: people choose whether to participate in job-training programs. The decision to participate is at least partially determined by factors that also affect the person's wage, such as the innate drive, or "go-for-it" attitude, of the person. Estimation of the regression by ordinary least squares is biased in this situation, since the program-participation dummy is correlated with the errors in the wage equation.
2. A regression equation is estimated on a sample of observations that are selected on the basis of a discrete choice that is determined endogenously with the dependent variable. For example, a researcher might want to estimate the impact of weather on peak energy load (that is, consumption during the highest-demand hour of the day). Data on energy loads by time of day are only available only for households that have chosen time-of-use rates. However, the households' choice of rate plan can be expected to be related to their energy consumption, with customers who have high peak loads tending not to choose time-of-use rates, since those rates charge high prices in the peak. Estimation of the regression equation on this "self-selected" sample is biased unless the endogeneity of the sample is accounted for.
3. The continuous dependent variable is truncated. For example, consumption of goods by households is necessarily positive. Stated

statistically, consumption is truncated below at zero, and for many goods (such as opera tickets) observed consumption is at this truncation point for a large share of the population. Estimation of the regression without regard to the truncation can cause bias.

The initial concepts regarding appropriate treatment of continuous/discrete models were developed by Heckman (1978, 1979) and Dubin and McFadden (1984). These early concepts are covered in my earlier book (Train, 1986, Ch. 5). Since then, the field has expanded tremendously. An adequate discussion of the issues and procedures would take a book in itself. Moreover, the field has not reached (at least in my view) the same type of juncture that discrete choice modeling has reached. Many fundamental concepts are still being hotly debated, and potentially valuable new procedures have been introduced so recently that there has not been an opportunity for researchers to test them in a variety of settings. The field is still expanding more than it is coalescing.

There are several on-going directions of research in this area. The early procedures were highly dependent on distributional assumptions that are hard to verify. Researchers have been developing semi- and non-parametric procedures that are hopefully more robust. The special 1986 issue of the *Journal of Econometrics* provides a set of important articles on the topic. Papers by Lewbel and Linton (forthcoming) and Levy (2001) describe more recent developments. Another important development concerns the representation of behavior in these settings. The relation between the discrete and continuous variables has been generalized beyond the fairly simple representation that the early methods assumed. For example, in the context of job training, it is likely that the impact of the training differs over people and that people choose to participate in the training program on the basis of the impact it will have on them. Stated in econometric terms: the coefficient of the participation dummy in the wage equation varies over people and affects the value of the dummy. The dummy is correlated with its own coefficient, as well as with the unobserved variables that enter the error of the regression. A recent discussion of approaches to this issue is provided by Carneiro *et al.* (2001).

1.5 A couple notes

Throughout the book, I refer to the researcher as "she" and the decision-maker as "he." This usage, as well as being comparatively gender neutral (or at least symmetrically non-inclusive), allows both people to be referred to in the same paragraph without confusion.

Many colleagues have provided valuable comments and suggestions on earlier drafts of the book. I am very grateful for this help. I thank: Greg Allenby, Moshe Ben-Akiva, Chandra Bhat, Denis Bolduc, David Brownstone, Siddhartha Chib, Jon Eisen-Hecht, Florian Heiss, David Hensher, Joe Herriges, Rich Johnson, Frank Koppelman, Jordan Louviere, Aviv Nevo, Juan de Dios Ortuzar, Ken Small, Joan Walker, Cliff Winston, Joachim Winter, and the students in my graduate econometrics course.

I welcome readers to contact me if you feel I have not covered material that you consider important, or if I have confused rather than enlightened any of the material that I *do* cover. Hopefully, another edition of this book will someday materialize.

Part I

Behavioral Models

Chapter 2

Properties of Discrete Choice Models

2.1 Overview

This chapter describes the features that are common to all discrete choice models. We start by discussing the choice set, which is the set of options that are available to the decision-maker. We then define choice probabilities and derive them from utility maximizing behavior. The most prominent types of discrete choice models, namely logit, GEV, probit, and mixed logit, are introduced and compared within the context of this general derivation. Utility, as a constructed measure of well-being, has no natural level or scale. This fact has important implications for the specification and normalization of discrete choice models, which we explore. We then show how individual-level models are aggregated to obtain market-level predictions, and how the models are used for forecasting over time.

2.2 The choice set

Discrete choice models describe decision-makers' choices among alternatives. The decision-makers can be people, households, firms, or any other decision-making unit, and the alternatives might represent competing products, courses of action, or any other options or items over which choices must be made. To fit within a discrete choice framework, the set of alternatives, called "the choice set," needs to exhibit three