Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables

by

Joan Leslie Walker

Bachelor of Science in Civil Engineering, University of California at Berkeley (1991)
Master of Science in Transportation, Massachusetts Institute of Technology (1994)

Submitted to the Department of Civil and Environmental Engineering
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in

Transportation Systems

at the

Massachusetts Institute of Technology

February 2001

Author ____
Department of Civil and Environmental Engineering
February 14, 2001

Certified by ____
Moshe E. Ben-Akiva
Edmund K. Turner Professor of Civil and Environmental Engineering
Thesis Supervisor

Accepted by ____
Oral Buyukozturk
Chair, Departmental Committee on Graduate Studies

Extended Discrete Choice Models:
Integrated Framework, Flexible Error Structures, and Latent Variables

by

Joan Leslie Walker

Submitted to the Department of Civil and Environmental Engineering
on February 14, 2001 in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy in Transportation Systems

Abstract

Discrete choice methods model a decision-maker's choice among a set of mutually exclusive and collectively exhaustive alternatives. They are used in a variety of disciplines (transportation, economics, psychology, public policy, etc.) in order to inform policy and marketing decisions and to better understand and test hypotheses of behavior. This dissertation is concerned with the enhancement of discrete choice methods.

The workhorses of discrete choice are the multinomial and nested logit models. These models rely on simplistic assumptions, and there has been much debate regarding their validity. Behavioral researchers have emphasized the importance of amorphous influences on behavior such as context, knowledge, and attitudes. Cognitive scientists have uncovered anomalies that appear to violate the microeconomic underpinnings that are the basis of discrete choice analysis. To address these criticisms, researchers have for some time been working on enhancing discrete choice models. While there have been numerous advances, typically these extensions are examined and applied in isolation. In this dissertation, we present, empirically demonstrate, and test a generalized methodological framework that integrates the extensions of discrete choice.

The basic technique for integrating the methods is to start with the multinomial logit formulation, and then add extensions that relax simplifying assumptions and enrich the capabilities of the basic model. The extensions include:

– *Specifying factor analytic (probit-like) disturbances* in order to provide a flexible covariance structure, thereby relaxing the IIA condition and enabling estimation of unobserved heterogeneity through techniques such as random parameters.

– *Combining revealed and stated* preferences in order to draw on the advantages of both types of data, thereby reducing bias and improving efficiency of the parameter estimates.

– *Incorporating latent variables* in order to provide a richer explanation of behavior by explicitly representing the formation and effects of latent constructs such as attitudes and perceptions.

– *Stipulating latent classes* in order to capture latent segmentation, for example, in terms of taste parameters, choice sets, and decision protocols.

The guiding philosophy is that the generalized framework allows for a more realistic representation of the behavior inherent in the choice process, and consequently a better understanding of behavior, improvements in forecasts, and valuable information regarding the validity of simpler model structures.

These generalized models often result in functional forms composed of complex multidimensional integrals. Therefore a key aspect of the framework is its 'logit kernel' formulation in which the disturbance of the choice model includes an additive i.i.d. Gumbel term. This formulation can replicate all known error structures (as we show here) and it leads to a straightforward probability simulator (of a multinomial logit form) for use in maximum simulated likelihood estimation. The proposed framework and suggested implementation leads to a flexible, tractable, theoretically grounded, empirically verifiable, and intuitive method for incorporating and integrating complex behavioral processes in the choice model.

In addition to the generalized framework, contributions are also made to two of the key methodologies that make up the framework. First, we present new results regarding identification and normalization of the disturbance parameters of a logit kernel model. In particular, we show that identification is not always intuitive, it is not always analogous to the systematic portion, and it is not necessarily like probit. Second, we present a general framework and methodology for incorporating latent variables into choice models via the integration of choice and latent variable models and the use of psychometric data (for example, responses to attitudinal survey questions).

Throughout the dissertation, empirical results are presented to highlight findings and to empirically demonstrate and test the generalized framework. The impact of the extensions cannot be known a priori, and the only way to test their value (as well as the validity of a simpler model structure) is to estimate the complex models. Sometimes the extensions result in large improvements in fit as well as in more satisfying behavioral representations. Conversely, sometimes the extensions have marginal impact, thereby showing that the more parsimonious structures are robust. All methods are often not necessary, and the generalized framework provides an approach for developing the best model specification that makes use of available data and is reflective of behavioral hypotheses.

Thesis Supervisor: Moshe E. Ben-Akiva
Title: Edmund K. Turner Professor of
Civil and Environmental Engineering

Acknowledgments

I am indebted to a great number of people who generously offered advise, encouragement, inspiration, and friendship throughout my time at MIT.

I offer my sincere gratitude to my advisor and mentor, Professor Moshe Ben-Akiva, with whom it has been an honor and a pleasure to work. I thank him for a great number of things: for sharing his knowledge, for treating me as a colleague, for the opportunities he has provided me, for forcing me to dig deeper into my research and then helping me find the solutions, for his patience, his accessibility, his encouragement, and his invaluable ideas, which form the backbone of this thesis.

Thank you to the members of my doctoral thesis committee. Professor Nigel Wilson, who provided input on my research as well as truly valued advice, and to whom I am grateful for looking out for me throughout my tenure at MIT. Professor Denis Bolduc, who provided vital assistance with my estimation programs and the empirical case studies, as well as instrumental guidance for Chapter 2. Dr. Dinesh Gopinath, whose thorough reading of the thesis draft greatly improved the final product, and whose command of econometrics proved over and over again to be invaluable. Fred Salvucci, whose wealth of fascinating insight and stories never ceases to amaze me, and who generously offered nothing but encouragement when my research headed away from his primary interests.

Thank you to the MIT faculty and staff who have enriched my experience. Professor Cynthia Barnhart who is an incredible role model and a tremendous asset to the program. Professor Ismail Chabini for his enthusiastic support and for bringing such energy to the department. Professor Joseph Sussman for his continued support that started the first day I arrived at MIT. Dr. Joseph Coughlin for his cheerful encouragement. My honorary committee members, Dr. Shlomo Bekhor and Professor Ikki Kim, for their interest in my research and useful comments. To the staff of CTS and CEE, especially Cynthia, Ginny, Jan and Sydney, who provide immeasurable assistance to the students in their care. Thanks to Nora, who cheerfully checked in on me on many a late nights. I also gratefully acknowledge the numerous sources from which my education has been funded, including Tren Urbano, the National Science Foundation, the University Transportation Centers Program, the United Parcel Service, the US Department of Transportation, and the Center for Transportation Studies at MIT.

Thank you to all the friends and colleagues who have made the process a whole lot more enjoyable. Leanne for providing endless entertainment and support. Sarah for her empathy, wonderful conversations, and inspiration. Denise for her encouragement and the fun. Jon for making the long hours in the office much more bearable and often enjoyable. Fani for her laughter and stories. My fellow modelers Scott and John, for the conversations about statistics and, thankfully, other topics as well. Salal for doing his best to teach me not to sweat the small stuff, and for listening so well when I forgot this lesson. Julie for her friendship and kindness. Neal, Paul, Emmanuelle, and Benoit for providing great escapes from MIT. Kristie and Rob, who were with me when I entered the doctoral program and touched me by returning for the conclusion. Mary, who has been influencing and supporting my academic pursuits for as long as I can remember. Alberta and Linda for their continuous support and interest in my endeavors. And Raj, for his great wisdom and unending encouragement.

Finally, thanks to my amazing family, who provide me with more love, support, and friendship then I would otherwise think possible. To my siblings, Kathy, Jeff, and Sue, who have done such a good job looking out for their little sister. They have each had a tremendous influence on who I am. To the recent additions to the family: kind and fun Arne; sweet and generous Judy; and beautiful, smart, and perfect little Julia. And to my parents, who have taught me countless philosophies of life, many of which I found to be invaluable during this process! To my dad, who has inspired me through his passion and dedication to both his family and his work. And to my mom, my biggest fan, to whom I hope I have given back a fraction of what I have received.

Table of Contents

List of Figures

List of Tables

Chapter 1: ***Introduction***

This dissertation is concerned with the enhancement of *discrete choice models*, which are methods used to model a decision-maker's choice among a set of mutually exclusive and collectively exhaustive alternatives. The guiding philosophy is that such enhancements allow for more behaviorally realistic representations of the choice process, and consequently a better understanding of behavior, improvements in forecasts, and valuable information regarding the validity of simpler model structures.

Motivation

There are 4 major factors that motivate the work described in this dissertation:

- The desire to model discrete choice behavior in a broad array of disciplines (transportation, economics, psychology, public policy, etc.) for a variety of reasons, including:
  - to provide forecasts to inform policy and marketing decisions, and
  - to better understand and test hypotheses of behavior.
- The complexity of the behavioral processes by which people make choices, which is influenced by latent concepts such as context, knowledge, and attitudes. (As advanced by behavioral theorists.)
- Conversely, the simplistic behavioral representation of the standard quantitative models of behavior, which, in practice, are dominated by the multinomial and nested logit formulations. (As developed by discrete choice modelers.)
- Continuing advances in the areas of computational power, estimation methodologies, and the availability of different types of behavioral data.

The work presented here aims to develop, demonstrate, and test a methodological framework to close the gap between the simplistic behavioral representation in today's models (*discrete choice models*) and the complexity of the actual behavioral process (*behavioral theory*), thereby improving the specification and explanatory power of discrete choice models.

The Foundation of Quantitative Models of Discrete Choice Behavior

The standard tool for modeling individual choice behavior is the choice model based on the random utility hypothesis. These models have their foundations in classic economic consumer theory, which is the source of many of the important assumptions of the models. Therefore, it is also the source of much debate surrounding the models as well as the fuel for extensions. In this section we briefly overview economic consumer theory, discuss how it extends to discrete choice theory, and present the basics of the random utility choice model.

Economic consumer theory states that consumers are rational decision makers. That is, when faced with a set of possible consumption bundles of goods, they assign preferences to each of the various bundles and then choose the most preferred bundle from the set of affordable alternatives. Given the properties of *completeness* (any two bundles can be compared, i.e., either *a* is preferred to *b*, or *b* is preferred to *a*, or they are equally preferred), *transitivity* (if *a* is preferred to *b* and *b* is preferred to *c*, then *a* is preferred to *c*) and *continuity* (if *a* is preferred to *b* and *c* is arbitrarily 'close' to *a*, then *c* is preferred to *b*), it can be shown that there exists a continuous function (the *utility function*) that associates a real number with each possible bundle, such that it summarizes the preference orderings of the consumer. Consumer behavior can then be expressed as an optimization problem in which the consumer selects the consumption bundle such that their utility is maximized subject to their budget constraint. This optimization function can be solved to obtain the demand function. The demand function can be substituted back into the utility equation to derive the indirect utility function, which is the maximum utility that is achievable under the given prices and income. The indirect utility function is what is used in discrete choice models, and we refer to this simply as 'utility' throughout the dissertation. (See, for example, Varian, 1992, for further information on consumer theory.)

There are several extensions to classic consumer theory that are important to discrete choice models. First, consumer theory assumes homogeneous goods (a car is a car), and therefore the utility is a function of quantities only and not attributes. Lancaster (1966) proposed that it is *the attributes* of the goods that determine the utility they provide, and therefore utility can be expressed as a function of the attributes of the commodities.

Second is the concept of random utility theory originated by Thurstone (1927) and further developed by Marschak (1960) and Luce (1959). Whereas classic consumer theory assumes deterministic behavior, random utility theory introduces the concept that individual choice behavior is intrinsically probabilistic. The idea behind random utility theory is that while the decision maker may have perfect discrimination capability, the analyst has incomplete information and therefore uncertainty must be taken into account. Therefore, utility is modeled as a random variable, consisting of an observable (i.e., measurable component) and an unobservable (i.e., random) component. Manski (1977) identified four sources of uncertainty: unobserved alternative attributes, unobserved individual attributes (or taste variations), measurement errors, and proxy (or instrumental) variables.

Finally, consumer theory deals with continuous (i.e., infinitely divisible) products. Calculus is used to derive many of the key results, and so a continuous space of alternatives is required. Discrete choice theory deals with a choice among a set of finite, mutually exclusive alternatives and so different techniques need to be used. However, the underlying hypotheses of random utility remain intact.

The standard technique for modeling individual choice behavior is the discrete choice model derived from random utility theory. As in consumer theory, the model is based on the notion that an individual derives utility by buying or choosing an alternative. Usually, the models assume that the individual selects the alternative that has the maximum utility, but other decision protocols can be used. The (indirect) utilities are latent variables, and the actual choice, which is what can be observed, is a manifestation of the underlying utilities. The utilities are specified as proposed by Lancaster (1966) and McFadden (1974), in which they are assumed to be a function of (i.e., caused by) the attributes of the alternatives and the characteristics of the decision maker (introduced to capture heterogeneity across individuals). The final component of the utility is a random disturbance term. Assumptions on the distributions of the disturbances lead to various choice models (for example, probit and logit). The outputs of the models are the probabilities of an individual selecting each alternative. These individual probabilities can then be aggregated to produce forecasts for the population.

Simplifying assumptions are made in discrete choice models in order to maintain a parsimonious and tractable structure. Such assumptions include utility maximizing behavior, deterministic choice sets, straightforward explanatory variables (for example, easily measurable characteristics of the decision-maker and attributes of the alternatives), and simple error structures such as GEV disturbances (multinomial logit, nested logit, cross-nested logit). There is a more extensive discussion of discrete choice models later in this chapter, and these models and their variants will be described in detail throughout the dissertation. (For a general discussion of discrete choice theory, see Ben-Akiva and Lerman, 1985, or McFadden, 1984.)

Qualitative Concepts of Behavioral Theory

Due to the strong assumptions and simplifications in quantitative discrete choice models, there has been much debate in the behavioral science and economics communities on the validity of such models. For example, one well-publicized issue with multinomial logit models is the property of Independence from Irrelevant Alternatives (or IIA), which will be discussed later.

Behavioral researchers have stressed the importance of the cognitive processes on choice behavior. Far from the concept of innate, stable preferences that are the basis of traditional discrete choice models, they emphasize the importance of things such as experience and circumstances and a whole host of amorphous concepts, some of which are listed in Table 1-1. These behavioral constructs are pervasive throughout consumer behavior textbooks (for example, Engel, Blackwell and Miniard, 1995; Hawkins, Best and Coney, 1989; and Olson, 1993) and research journals (for example, *Journal of Applied Social Psychology*, *Journal of Marketing Research*, *Journal of Consumer Psychology*, etc.). Many detailed and comprehensive representations of the consumer choice process have been proposed by behavioral researchers, the most widely cited being those by Engel et al. (EKB) (1968, 1982, 1995); Howard and Sheth (1969) and Howard (1977 and 1989); and Nicosia (1966) and Nicosia and Wind (1977). These models are described in many consumer behavior textbooks including Engel and Blackwell (1982), Onkvisit and Shaw (1994), and Rice (1993). These researchers take a systems dynamics approach in which equations (often linear) are associated with connectivity as represented in a flow diagram. The behavioral process that is represented is complex, with extensive connectivity and feedback between the behavioral states and constructs. For example, the Howard and EKB frameworks are presented in Figure 1-1 and Figure 1-2. As would be expected, mathematically capturing this process is difficult. Some of the issues with the estimation techniques used for these models are that they are not grounded in economic consumer theory, and they depend on the use of psychometric indicators (for example, responses to survey questions regarding attitudes, perceptions, and memory) as causal variables in the process (see Chapter 3 for a discussion). Nonetheless, such representations are extremely valuable in conceptualizing and studying the behavioral process.

In addition to the grand behavioral frameworks discussed above, there has been a lot of research on specific aspects of the behavioral process, including every concept shown in Table 1-1, Figure 1-1, and Figure 1-2. It is a huge literature, which we cannot hope to give justice here. Ajzen (2001), Olson and Zanna (1993), and Wood (2000) provide a summary of research on attitudes, which is a major emphasis in the literature. Jacoby et al. (1998) and Simonson et al. (2001) provide a broader review of consumer behavior research.

Furthermore, a great deal of research has been conducted to uncover cognitive anomalies that appear to violate the basic axioms of the utility maximization foundations of discrete choice theory. The fundamental work in this area was performed by Kahneman and Tversky (for example, Kahneman and Tversky, 1979, Tversky, 1977, and Tversky and Kahneman, 1974), who accumulated experimental evidence of circumstances in which individuals exhibit surprising departures from rationality. They found that decision makers are sensitive to context and process, they are inconsistent at forming perceptions and processing information, and they use decision-making heuristics. Some of the issues emphasized by cognitive psychologists are the degree of complexity, familiarity, and risk of the choice at hand (see, for example, Ajzen, 1987, and Gärling, 1998); the use of non-utility maximizing decision protocols such as problem-solving, reason-based, and rule-driven processes (see, for example, Payne et al., 1992, and Prelec, 1991); and the concept of 'framing effects', which is that people often accept and use information in the form in which they receive it (see, for example, Slovic, 1972, and Schweitzer, 1995); and a whole host of other perceived biases and errors associated with rational theory.

Table 1-1: Influences on the Choice Process

Context	***Knowledge***	***Point of View***	***Choice***
Experience	Awareness	Perceptions	Problem Recognition
Involvement	Search	Attitudes	Constraints
Motivation	Exposure	Beliefs	Compliance
Attention	Memory	Lifestyle	Evaluation Criteria
Stimuli	Learning	Behavior modification	Decision protocol
Intention	Comprehension	Cultural Norm/Values	
	Recall	Satisfaction	
	Information		
	Reference Groups		

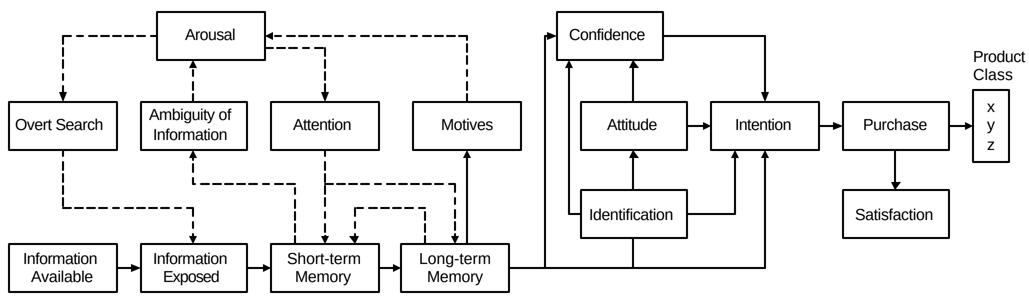


Figure 1-1: The Howard Model of Consumer Behavior
(Figure taken from Engel and Blackwell, 1982)

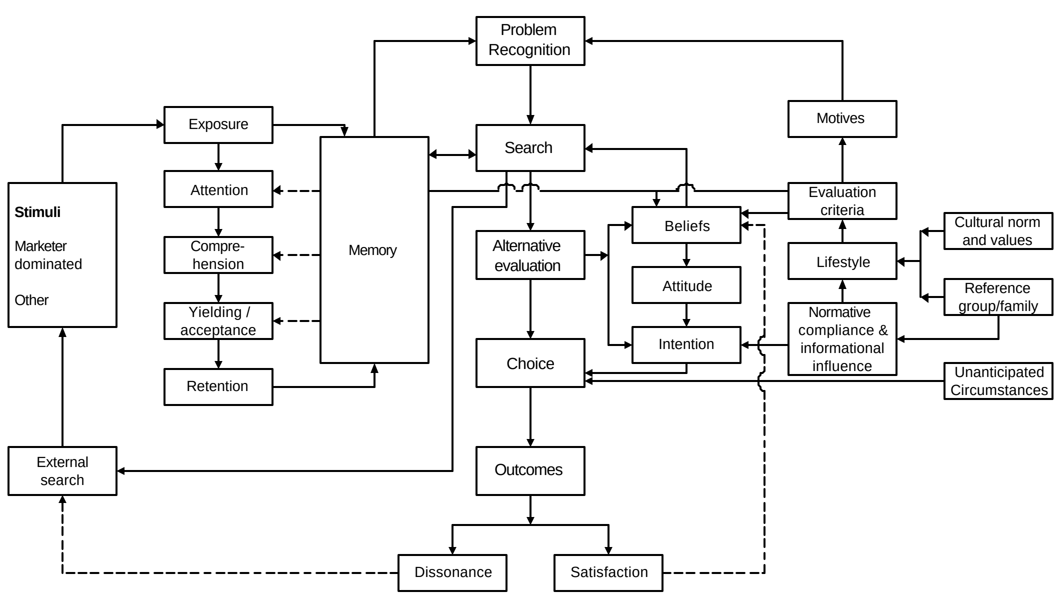


Figure 1-2: The EKG Model of Consumer Behavior
(Figure taken from Engel and Blackwell, 1982)

Camerer (1987), Mellers et al. (1998), Rabin (1998), and Thaler (1991) provide surveys of the research in cognitive anomalies from a behavioral scientists reference point. McFadden (1997) provides a summary of the work from a discrete choice modelers view. He argues that "most cognitive anomalies operate through errors in perception that arise from the way information is stored, retrieved, and processed" and that "empirical study of economic behavior would benefit from closer attention to how perceptions are formed and how they influence decision-making."

The Gap Between Behavioral Theory and Discrete Choice Models

As implied by the discussion above, there is a large gap between behavioral theory and discrete choice models. The gap exists because of the driving forces behind the two disciplines: while discrete choice modelers are focused on mapping inputs to the decision, behavioral researchers aim to understand the nature of how decisions come about, or the decision-process itself. The graphic in Figure 1-3 highlights this difference. This figure, as well as the remaining figures in the dissertation, follows the convention that unobservable variables are shown in ovals, observable variables in rectangles, causal relationships by solid arrows, and measurement relationships by dashed arrows.

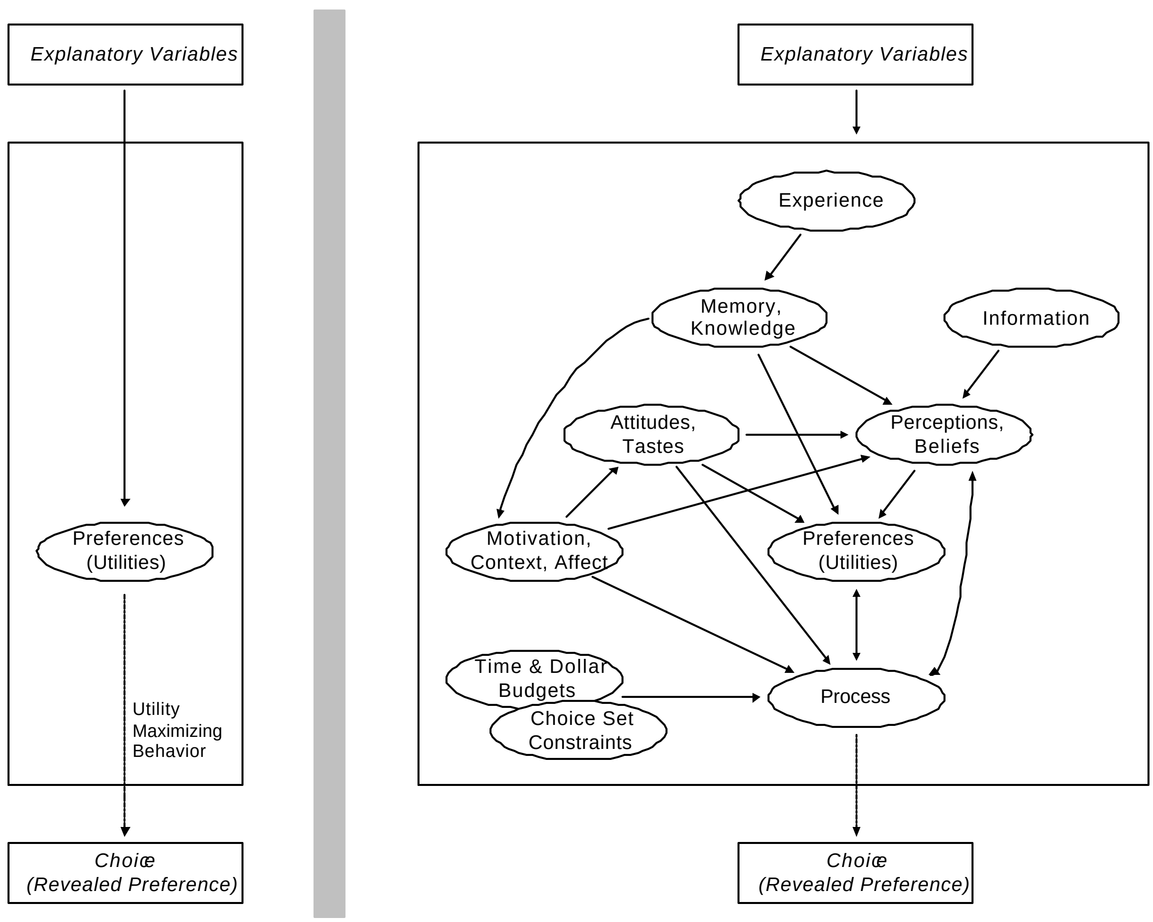


Figure 1-3: The gap between basic Discrete Choice Models (left) and the Complexity of Behavior (right)

The framework for the basic discrete choice model is shown on the left. The preferences (represented by utilities) are unobservable, but they are assumed to be a function of explanatory variables as well as unknown parameters (or weights) and a disturbance term. The choice is a manifestation of the preferences, and the typical assumption is that the alternative with the maximum utility is chosen. This model is often described as an "optimizing black box", because the model directly links the observed inputs to the observed output, and thereby assumes that the model implicitly captures the behavioral choice process.

The right side of Figure 1-3 is an attempt to show the inherent complexity of the behavioral process itself.[1] While one could argue about the specific terminology, components, and connectivity, the objective of the figure is to provide an example of a more realistic representation of the underlying choice process.

The question is, does the gap matter? Or, is the optimizing black box an adequate representation? In terms of applying the models, clearly the most desirable model is the one that is as parsimonious as possible, and yet serves the purpose at hand. We have found in many instances that the multinomial logit formulation is quite robust. However, there are instances in which a more complex model structure could be of use, for example:

- To provide confidence that a parsimonious specification is adequate.
- To improve forecasts.
- To test a particular behavioral theory or hypothesis having to do with a construct in the black box.
- To correct for biases and so-called cognitive anomalies in responses.
- To introduce different types of measurement relationships (beyond just the revealed choice preference indicator) that are hypothesized to provide information on the choice process.

What specifically can we do to enhance the choice model? Researchers have been working on this for some time, and this is the topic of this dissertation.

The State of the Practice in Discrete Choice Modeling and Directions of Research

The background of the random utility model was presented above, and a framework shown in Figure 1-3. The general model is written mathematically as follows:

$$U_{in} = V(X_{in};q) + e_{in} \ , \qquad \text{"Structural Equation"}$$

$$y_n = f(U_n) \ , \qquad \text{"Measurement Equation"}$$

where: n denotes an individual, $n = 1,...,N$;

i, j denote alternatives, $i, j = 1,..., J$;

[1] The figure is adapted from Ben-Akiva, McFadden et al. (1999) and McFadden (2000).

J_n is the number of alternatives considered by individual n;

U_{in} is the utility of alternative i as perceived by individual n; U_n is the $(J_n \times 1)$ vector of utilities;

y_{in} is the choice indicator (equal to 1 if alternative i is chosen, and 0 otherwise), and y_n is the $(J_n \times 1)$ vector of choice indicators;

V is a function that expresses the systematic utility in terms of explanatory variables;

f is a function that represents the decision protocol as a function of the utility vector;

q are a set of unknown parameters;

e_{in} are random disturbance terms; and

X_{in} is a $(1 \times K)$ vector describing n and i; X_n is the $(J_n \times K)$ matrix of stacked X_{in}.

The most common discrete choice model is the linear in parameters, utility maximizing, multinomial logit model (MNL), developed by McFadden (1974), which is specified as:

$$U_{in} = X_{in} b + n_{in}, \quad n_{in} \text{ are i.i.d. Gumbel random variates with scale parameter } m, \qquad [1\text{-}1]$$

$$y_{in} = \begin{cases} 1, & if\ U_{in} = \max_j \{U_{jn}\} \\ 0, & otherwise \end{cases} . \qquad [1\text{-}2]$$

Equations [1-1] and [1-2] lead to the following individual choice probability:

$$P(y_{in} = 1 \mid X_n; b) = \frac{e^{m(X_{in} b)}}{\sum_{j \in C_n} e^{m(X_{jn} b)}} ,$$

where: C_n is the choice set faced by individual n, comprised of J_n alternatives; and

b is a $(K \times 1)$ vector of unknown parameters.

One of the most noteworthy aspects of the multinomial logit model is its property known as Independence from Irrelevant Alternatives (or IIA), which is a result of the i.i.d. disturbances. The IIA property states that, for a given individual, the ratio of the choice probabilities of any two alternatives is unaffected by other alternatives. This property was first stated by Luce (1959) as the foundation for his probabilistic choice model, and was a catalyst for McFadden's development of the tractable multinomial logit model. There are some key advantages to IIA, for example, the ability to estimate a choice model using a sample of alternatives, developed by McFadden (1978). However, as Debreu (1960) pointed out, IIA also has the distinct disadvantage that the model will perform poorly when there are some alternatives that are very similar to others (for example, the now famous red bus – blue bus problem).

There are many ways to relax the IIA assumption, and many variations of discrete choice models aim at doing just that. Nested logit (NL), introduced by Ben-Akiva (1973) and derived as a random utility model as a special case of GEV by McFadden (1978, 1981), partially addresses this issue by explicitly allowing correlation within sets of mutually exclusive groups of alternatives. The beauty of nested logit is that it retains an extremely tractable closed form solution, and therefore is widely used (second only in popularity to multinomial logit).

Multinomial and nested logit are the workhorses of discrete choice modeling, and form the foundation of models in areas such as travel demand modeling and marketing. This is because they are extremely tractable and fairly robust models that are widely described in textbooks (for example, Ben-Akiva and Lerman, 1985; Greene, 2000; Louviere et al., 2000; Ortuzar and Willumsen, 1994) and can be easily estimated by numerous estimation software packages (for example, *HieLow*[2] and *Alogit*[3]). Nested logit models have been used to estimate extremely complex decision processes, for example, detailed representations of individual activity and travel patterns (see Ben-Akiva and Bowman, 1998).

Beyond MNL and NL, there are many directions for enhancements that are pursued by discrete choice modelers. These directions are loosely categorized (with admitted overlap across categories) and discussed below, and the chapters that follow contain more detailed literature reviews on many of these topics. For further information, McFadden (2000) provides an excellent review of the history and future directions of discrete choice theory.

Specification of the Disturbances

There has been a lot of research focused on introducing more flexibility to the covariance structure of MNL in order to relax IIA and improve the performance of the model. Nested logit is one example of this area. In addition, there are a numerous other variations on the logit theme, albeit none that comes close to the popularity of MNL and NL. Cross-nested logit (CNL), relaxes the error structure of nested logit by allowing groups to overlap. CNL was first mentioned by McFadden (1978) and further investigated and applied by Small (1987) for departure time choice, Vovsha (1997) for mode choice, and Vovsha and Bekhor (1998) for route choice. MNL, NL, and CNL are all members of the General Extreme Value, or GEV, class of models, developed by McFadden (1978, 1981), a general and elegant model in which the choice probabilities still have tractable logit form but do not necessarily hold to the IIA condition. There is also the heteroscedastic extreme value logit model, which allows the variance of the disturbance to vary across alternatives. This was developed and applied by Bhat, 1995, for travel mode choice and tested against other GEV and probit models using synthetic data by Munizaga et al. (2000).

---

[2] Distributed by Stratec.

[3] Distributed by Hague Consulting Group.

The other major family of discrete choice models is the probit family, which has a multivariate *normal* distributed disturbance. The early investigations of probabilistic choice models (Aitchison and Bennett, 1970; Bock and Jones, 1968; Marschak, 1960) were of probit form, because it is natural to make normality assumptions. Probit is extremely flexible, because it allows for an unrestricted covariance matrix, but is less popular than the GEV forms primarily due to the difficulty in estimation (i.e., lack of a closed form solution). Much of the research on probit is in the areas of estimation (for example, Clark, 1961, developed an early used approximation; Lerman and Manski, 1981, pioneered the use of simulation for econometric models; and Geweke, Hajivassiliou, and Keane developed the now common GHK simulator[4], which made great strides in increasing the tractability of probit) and in simplifying the error structure (for example, McFadden, 1984, proposed using a factor analytic form to reduce the dimensionality of the integral). Daganzo (1979) provides a thorough examination of probit, and the model is widely described in Econometrics textbooks, for example, Amemiya (1985), Ben-Akiva and Lerman (1985), and Greene (2000).

Logit kernel (or continuous mixed logit model) is a model that attempts to combine the relative advantages of probit and GEV forms, and this is the subject of Chapter 3. It is a powerful and practical model that has recently exploded in the applied literature (see Chapter 3 for references) and is making its way into econometric textbooks, for example, Greene, 2000, and Louviere et al., 2000. The disturbance of the logit kernel model is composed of two parts: a probit-like term, which allows for flexibility, and an i.i.d. Gumbel (or GEV) term, which aids in estimation. The technique was used as early as Boyd and Mellman (1980) and Cardell and Dunbar (1980) for the specific application of random parameter logit. The more general form of the model came about through researchers quest for smooth probability simulators for use in estimating probit models. McFadden's 1989 paper on the Method of Simulated Moments, includes a description of numerous smooth simulators, one of which involved probit with an additive i.i.d. Gumbel term. Stern (1992) described a similar simulator, which has an additive i.i.d. normal term instead of the Gumbel. At the time of these papers, there was a strong desire to retain the pure probit form of the model. Hence, the algorithms and specifications were designed to eventually remove the additive "contamination" element from the model (for example, McFadden, 1989) or ensure that it did not interfere with the pure probit specification (for example, Stern, 1992). Bolduc and Ben-Akiva (1991)[5] did not see the need to remove the added noise, and began experimenting with models that left the Gumbel term in tact, and found that the method performed well. There have been numerous relatively recent applications and investigations into the model (see Chapter 3). A particularly important contribution is McFadden and Train's (2000) paper on mixed logit, which both *(i)* proves that any well-behaved RUM-consistent behavior can be represented as closely as desired with a mixed logit specification and *(ii)* presents easy to implement specification tests for these models.

[4] See Hajivassiliou and Ruud, 1994, for a description of GHK.

[5] Later generalized in Ben-Akiva and Bolduc (1996).

Incorporating Methods from Related Fields

There has been a growing effort to incorporate the findings and techniques from related fields into applied discrete choice models. We highlighted above the contributions of psychologists and behavioral theorists, who have studied how decisions are made and have researched cognitive anomalies that appear to violate the axioms of the discrete choice model. There have also been key influences from two other groups of researchers.

Psychometricians

Psychometricians, in their quest to understand behavioral constructs, have pioneered the use of psychometric data, for example, answers to direct survey questions regarding attitudes, perceptions, motivations, affect, etc. A general approach to synthesizing models with latent variables and psychometric-type measurement models has been advanced by a number of researchers including Keesling (1972), Jöreskog (1973), Wiley (1973), and Bentler (1980), who developed the structural and measurement equation framework and methodology for specifying and estimating latent variable models. Such models are widely used to define and measure unobservable factors, including many of the constructs shown in Figure 1-3. The incorporation of these latent variable techniques (for example, factor analysis) into choice models is the topic of Chapter 3.

Market Researchers

Whereas psychometricians tend to focus on behavioral constructs such as attitudes and perceptions, market researchers tend to focus on preferences. They have long used stated preference (conjoint) data to provide insight on preferences. The analysis of stated preference data originated in mathematical psychology with the seminal paper by Luce and Tukey (1964). The basic idea is to obtain a rich form of data on behavior by studying the choice process under hypothetical scenarios designed by the researcher. There are many advantages to these data including the ability to: capture responses to products not yet on the market, design explanatory variables such that they are not collinear and have wide variability, control the choice set, easily obtain numerous responses per respondent, and employ various response formats that are more informative than a single choice (for example, ranking, rating, or matching). Areas of research include experimental design, design of choice experiments, developing the choice model, and validity and biases. See Carroll and Green (1995) for a discussion of the methods and Louviere et al. (2000) for a general review of all issues. The primary drawback to stated preference data is that they may not be congruent with actual behavior. For this reason, techniques to combine stated and revealed preferences (developed by Ben-Akiva and Morikawa, 1990, and described in Chapter 4), which draw on the relative advantages of each type of data, are becoming increasingly popular (see Chapter 4 for references).

Preference and Behavior Indicators

We highlight the different type of indicators because a major emphasis in this thesis is making use of the various types of information we have to provide insight on the behavioral process. First, there are many different types of choice indicators, and variations of the logit model have been developed for the various types, for example ordinal logit when responses are in the form of an ordinal scale or dynamic choice models for panel data (see, for example, Golob et al., 1997). Second, there has been a lot of research on techniques specific to stated preference responses, as mentioned above. Finally, there can also be indicators for the behavioral process itself (e.g., survey questions regarding attitudes, memory, or decision protocol), and the latent variable techniques described above aim to make use of such psychometric indicators; the use of such data in choice models is the topic of Chapter 3.

Choice Process Heterogeneity

A key area of enhancements to discrete choice models is related to the idea that there is heterogeneity in behavior across individuals, and ignoring this heterogeneity can result in forecasting errors. For example, Ben-Akiva, Bolduc, and Bradley (1994) demonstrated the significance of unobserved heterogeneity on the demand curve for toll facilities. The most straightforward way to address this issue is to capture so-called "observed heterogeneity" by introducing socio-economic and demographic characteristics in the systematic portion of the utility function (i.e., in $V(\bullet)$). This has been an emphasis in forecasting models since the early applications, for example in the urban travel demand models developed by Domencich and McFadden (1975) and Ruiter and Ben-Akiva (1978). Alternatively, there are numerous techniques aimed at capturing *unobserved* heterogeneity. Quandt (1970) and Hausman and Wise (1978) introduced the concept of random parameters to the probit model, and Boyd and Mellman (1980) and Cardell and Dunbar (1980) estimated random parameter logit models. There are numerous recent applications of this technique, see, for example, Hensher and Reyes, 2000, and Mehndiratta and Hansen, 1997. This will be discussed in Chapter 3 within the context of the logit kernel framework. Another technique is latent class models, which can be used to capture unobservable segmentation regarding tastes, choice sets, and decision protocols. The concept of discrete mixing of functions (termed finite mixture models) has been around a long time (at least since Pearson, 1894), and McLachlan and Basford (1988) offer a review of these methods. The technique entered the choice behavior context with work by Manski (1977) in the context of choice set generation and Kamakura and Russell (1987) in the context of taste variation. Gopinath (1995) developed a general and rigorous treatment of the problem within a choice context. Latent class models are further discussed in Chapter 4.

Data, Estimation Techniques, and Computational Power

Fueling all of the extensions discussed above are the advances being made in data collection (for example, information technology, the collection of more refined data, stated preferences, and psychometric data), estimation techniques (in particular, the use of the simulation techniques pioneered by Lerman and Manski, 1981, McFadden, 1989, and Pakes and Pollard, 1989, and excellently reviewed in Stern, 2000), and computational power. These improvements make the estimation of behaviorally realistic models more attainable.

Objectives

While there have been numerous advances in discrete choice modeling, typically each of these extensions is examined and applied in isolation and there does not exist an integrated methodological framework. The objective of this research is to develop a generalized discrete choice model to guide the progress of models towards more behaviorally realistic representations with improved explanatory power. The resulting framework must be mathematically tractable, empirically verifiable, theoretically grounded, and have the ability to incorporate key aspects of the behavioral decision making process.

To achieve this objective, we develop, demonstrate, and test an overall framework that meets the stated specifications, including the synthesis of the various extensions discussed in the preceding section. We also provide in-depth analysis regarding specification, estimation, and identification of two of the key components of the framework:

1. The specification of flexible error structures and the logit kernel model.
2. The incorporation of latent variables into discrete choice models.

Overview of the Generalized Framework

The proposed generalized framework is shown in Figure 1-4. The framework draws on ideas from a great number of researchers, including Ben-Akiva and Morikawa (1990) who developed the methods for combining revealed and stated preferences; Cambridge Systematics (1986) and McFadden (1986) who laid out the original ideas for incorporating latent variables and psychometric data into choice models; Ben-Akiva and Boccara (1987) and Morikawa, Ben-Akiva, and McFadden (1996) who continued the development for including psychometric data into choice models; Gopinath (1995) who developed rigorous and flexible methods for capturing latent class segmentation in choice models; and Ben-Akiva and Bolduc (1996) who introduced an additive factor analytic parameterized disturbance to the multinomial logit i.i.d Gumbel.

As shown in Figure 1-4, the core of the model is a standard multinomial logit model (highlighted in bold), and then extensions are added to relax simplifying assumptions and enrich the capabilities of the basic model. The extensions include:

- *Factor analytic (probit-like) disturbances* in order to provide a flexible covariance structure, thereby relaxing the IIA condition and enabling estimation of unobserved heterogeneity through, for example, random parameters.
- *Combining revealed and stated preferences* in order to draw on the advantages of the two types of data, thereby reducing bias and improving efficiency of the parameter estimates.
- *Incorporating latent variables* in order to provide a richer explanation of behavior by explicitly representing the formation and effects of latent constructs such as attitudes and perceptions.
- *Stipulating latent classes* in order to capture latent segmentation in terms of, for example, taste parameters, choice sets, and decision protocols.

The framework has its foundation in the random utility theory described above, makes use of different types of data that provide insight into the choice process, allows for any desirable disturbance structure (including random parameters and nesting structures) through the factor analytic disturbance, and provides means for capturing latent heterogeneity and behavioral constructs through the latent variable and latent class modeling structures. Through these extensions, the choice model can capture more behaviorally realistic choice processes. Furthermore, the framework can be practically implemented via use a the logit kernel smooth simulator (as a result of the additive i.i.d. Gumbel) and a maximum simulated likelihood estimator.

The dissertation includes both an in-depth presentation and application of this framework, as well as extended investigations into two key aspects of the framework: the specification and identification of the disturbances and the incorporation of latent variables.

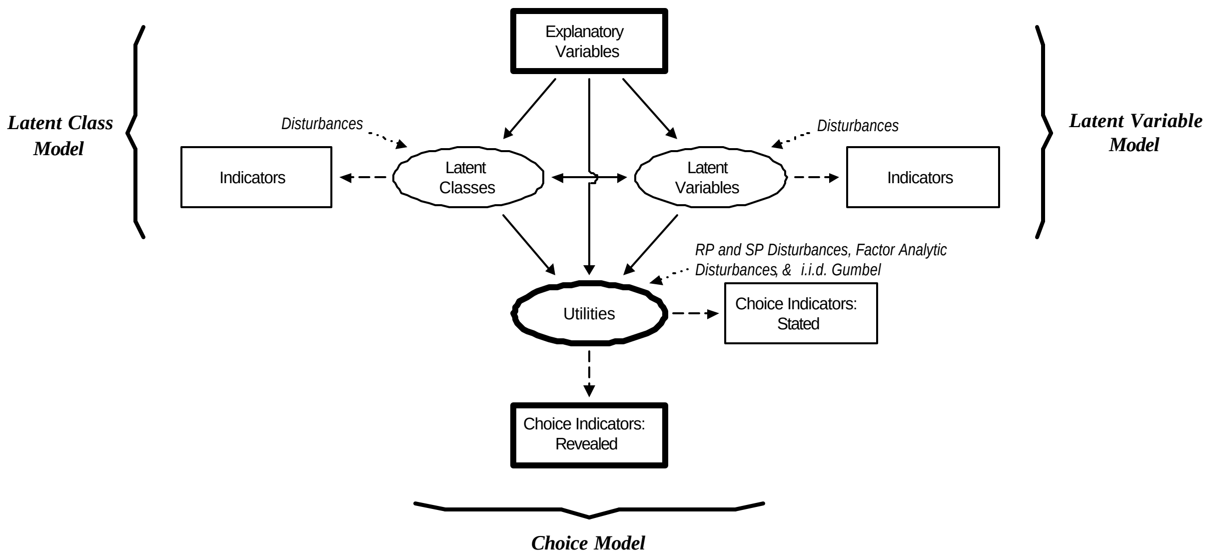


Figure 1-4: Generalized Discrete Choice Framework

Outline of the Dissertation

The dissertation is organized as follows:

- Chapter 2 focuses on the specification of the random component of the utility function. The basic idea behind the extension that is discussed is to develop general and tractable models with flexible error structures. Such structures aim to relax the IIA property of the logit model and are able to capture a variety of sources of heterogeneity among individuals. The model discussed is a hybrid between logit and probit, called logit kernel, which is a model that is becoming wildly popular in the discrete choice model literature. We specify the model using a factor analytic structure, and we show that this specification can represent any desirable error structure. In addition, we establish specific rules for identification, which has thus far been largely ignored in the literature. Empirical results are presented using both synthetic and real data to highlight issues of specification and identification.

- Chapter 3 focuses on the specification of the systematic part of the utility function. The motivation for the methodology we investigate is that there are often causal variables and behavioral constructs that are important to the choice process, but which are not directly observable. The method discussed in this chapter is the explicit incorporation of latent constructs such as attitudes and perceptions (or, more generally, any of the concepts in Table 1-1 or Figure 1-3) in the choice model. The objective is to develop models that more accurately represent the behavioral process, and therefore provide more accurate forecasts of demand. This method makes use of what are called psychometric indicators, for example, responses to survey questions about attitudes or perceptions, which are hypothesized to be manifestations of the underlying latent behavioral constructs. The chapter presents a general framework and methodology for incorporating latent variables into choice models. Empirical results from prior dissertations are reviewed to provide examples of the method and to demonstrate its practicality and potential benefits.

- Chapter 4 provides the generalized framework that aims to incorporate all extensions to the discrete choice model. The framework includes as important components the latent variable techniques described in Chapter 2 and the flexible error structures discussed in Chapter 3. These methods are summarized along with other techniques that are incorporated in the framework. Empirical results are presented to demonstrate and test the use and practicality of the generalized discrete choice model.

- Chapter 5 provides a summary and directions for further research.

Contributions

This dissertation represents a combination of summary, synthesis, and development. The specific contributions presented in this document are as follows:

Flexible Error Structures and the Logit Kernel Model

The Logit Kernel Model, which is the focus of Chapter 2, is a very flexible and powerful method for introducing flexible error structures in discrete choice models. It is a relatively new and extremely *en vogue* model form – even deemed 'the model of the future' by some. There are two important contributions in this chapter. The first is the use of a factor analytic form for the error structure, which we show is able to represent any desirable (additive) error structure. (This contribution was originally presented by Ben-Akiva and Bolduc, 1996, of which the chapter presented here represents a major revision). The second contribution is that it turns out that there are numerous specification and identification issues that are vital to practical application of these models and yet, to our knowledge, are not recognized in the existing literature. This chapter presents new results in this area, including the development of specific rules for identification and normalization of the parameters in the error structure.

Empirical results using both synthetic and real data are provided to highlight the specification and identification issues raised in the chapter.

The research presented here has important implications on the logit kernel model, which is the focus of the chapter. Furthermore, there are results that are applicable to any kernel specification (for example, probit kernel) and to any random parameter discrete choice specifications (for example, random parameter probit).

Integrating Choice and Latent Variable Models

While the ideas of combining choice and latent variables have been around for some time (for example, Cambridge Systematics, 1986; McFadden, 1986; and Ben-Akiva and Boccara, 1987), the literature contains only empirical applications to specific problems (for example, the case studies reviewed here) or restricted model formulations (for example, the elegant formulation for a binary probit and MIMC model presented in McFadden, 2000, and Morikawa et al., 1996). The contribution in this dissertation is the development of a general framework and methodology (including specification, identification, and estimation) for incorporating latent variables in discrete choice models. The described method provides complete flexibility in terms of the formulation of both the choice model and the latent variable model. In addition, the method is placed within a larger framework of alternative approaches, and a theoretical comparison of the various methods is provided. The case studies reviewed in Chapter 3 were developed earlier by others and are reviewed here to provide examples of the methodology. The empirical results for the choice and latent variable model presented in Chapter 4 are new to this dissertation.

Generalized Discrete Choice Model

The final chapter summarizes and synthesizes a variety of extensions to the discrete choice model. While the existing literature focuses on developing and applying the methods independently, the key contribution of this chapter is the integration of methods and presentation of a generalized discrete choice model and estimation method that incorporates all extensions. The basic technique for integrating the methods is to start with the multinomial logit formulation, and then add extensions that relax simplifying assumptions and enrich the capabilities of the basic model. These models often result in functional forms composed of complex multidimensional integrals. The core multinomial logit formulation allows for relatively straightforward estimation via the maximum simulated likelihood techniques and the logit kernel simulator. The proposed framework and suggested implementation leads to a flexible, tractable, practical, and intuitive method for incorporating and integrating complex behavioral processes in the choice model. This chapter provides empirical results that demonstrate and test the practicality of the generalized framework.

Chapter 2: **Flexible Error Structures and the Logit Kernel Model**

The extension presented in this chapter focuses on the specification of the error portion of the utility function. The basic idea is the development of general and tractable models with flexible error structures that relax the independence from irrelevant alternatives (IIA) property of the Logit model, and are able to capture a variety of sources of heterogeneity among individuals.

The model discussed in this chapter (called the Logit Kernel Model) is a very flexible and powerful method for introducing flexible error structures in discrete choice models. In this chapter we show how a factor analytic form of the error structure can be used to replicate all known error structures. We also present new results regarding normalization and identification of the disturbance parameters of the logit kernel model.

Introduction

The logit kernel model is a straightforward concept: it is a discrete choice model in which the disturbances (of the utilities) consist of both a probit-like portion and an additive i.i.d. Gumbel portion (i.e., a multinomial logit disturbance).

Multinomial logit (MNL) has its well-known blessing of tractability and its equally well-known curse of a rigid error structure leading to the IIA property. The nested logit model relaxes the rigidity of the MNL error structure and has the advantage of retaining a probability function in closed form. Nonetheless, nested logit is still limited and cannot capture many forms of unobserved heterogeneity, including, for example, random parameters. The logit kernel model with its probit-like disturbances completely opens up the specification of the disturbances so that almost any desirable error structure can be represented in the model. As with probit, however, this flexibility comes at a cost, namely that the probability functions consist of multi-dimensional integrals that do not have closed form solutions. Standard practice is to estimate such models by replacing the choice probabilities with easy to compute and unbiased simulators. The beauty of the additive i.i.d. Gumbel term is that it leads to a particularly convenient and attractive probability simulator, which is simply the average of a set of logit probabilities. The logit kernel probability simulator has all of the desirable properties of a simulator including being convenient, unbiased, and smooth.

Terminology

There are numerous terms floating around the literature that are related to the logit kernel model that we present here. McFadden, Train, and others use the term "mixed logit" to refer to models that are comprised of a mixture of logit models. This is a broad class that encompasses any type of mixing distribution, including discrete distributions (for example, latent class) as well as continuous distributions. Within this reference, logit kernel is a special case of mixed logit in which the mixing distribution is continuous. There are also numerous terms that are used to describe various error specifications in discrete choice models, including error components, taste variation, random parameters (coefficients), random effects, unobserved heterogeneity, etc. When such models are specified in a form that includes an additive i.i.d. Gumbel term, then they fall within the logit kernel (as well as mixed logit) class of models. Many of these special cases are described later in the chapter.

We choose to use the term *logit kernel*, because conceptually these models start with a logit model at the core and then are extended by adding a host of different error terms. In addition, the term is descriptive of the form of the likelihood function and the resulting logit kernel simulator.

Organization of the Chapter

The chapter is organized as follows. First, we introduce the logit kernel model and present a general discussion of identification. Then we discuss specification and identification of several important special cases, which are all based on a factor analytic representation of the error covariance structure. Next, we focus on the estimation of logit kernel via maximum (simulated) likelihood. In the final section, we present empirical results that highlight some of the specification and identification issues.

Related Literature

There have been many previous efforts to extend the logit model to allow more flexible covariance structures. The most widely used extension is nested logit. The advantage of nested logit is that it relaxes the classic IIA assumption and yet has a closed form. Nonetheless it is still a fairly rigid model. Nested logit is not a logit kernel model, although it can be approximated in the logit kernel structure. In terms of logit kernel models, the earliest applications were in random parameter logit specifications, which appeared 20 years ago in the papers by Boyd and Mellman (1980) and Cardell and Dunbar (1980). The more general form of the model came about through researchers quest for smooth probability simulators for use in estimating probit models. McFadden's 1989 paper on the Method of Simulated Moments, includes a description of numerous smooth simulators, one of which involved probit with an additive i.i.d. Gumbel term. Stern (1992) described a similar simulator, which has an additive i.i.d. normal term instead of the

Gumbel. At the time of these papers, there was a strong desire to retain the pure probit form of the model. Hence, the algorithms and specifications were designed to eventually remove the additive "contamination" element from the model (for example, McFadden, 1989) or ensure that it did not interfere with the pure probit specification (for example, Stern, 1992). Bolduc and Ben-Akiva (1991)[6] did not see the need to remove the added noise, and began experimenting with models that left the Gumbel term in tact, and found that the models performed well. There have been numerous relatively recent applications and investigations into the model, including Bhat (1997 & 1998), Bolduc, Fortin and Fournier (1996), Brownstone, Bunch and Train (2000), Brownstone and Train (1999), Goett, Hudson, and Train (2000), Gönül and Srinivasan (1993), Greene (2000), Mehndiratta and Hansen (1997), Revelt and Train (1998 & 1999), Srinivasan and Mahmassani (2000), and Train (1998). A very important recent contribution is McFadden and Train's (2000) paper on mixed logit, which both (i) proves that any well-behaved random utility consistent behavior can be represented as closely as desired with a mixed logit specification, and (ii) presents easy to implement specification tests for these models.

While logit kernel has strong computational advantages, it, like probit, does not have a closed form solution and can easily lead to high dimensional integrals. The well-known Gaussian Quadrature method of numerical integration is not computationally feasible for dimensionalities above 3 or so, and therefore estimation via simulation is a key aspect to applications of the logit kernel model. The basic idea behind simulation is to replace the multifold integral (the probability equations) with easy to compute probability simulators. Lerman and Manski (1981) introduced this concept and proposed the use of a frequency simulator to simulate probit probabilities. The frequency simulator was found to have poor computational properties primarily because it is not smooth (i.e., not continuous and not differentiable). Basically the frequency simulator maps each draw to a value of either 0 or 1, whereas a smooth simulator would map each draw to a value somewhere between 0 and 1 (and therefore retains more information). The result is that discontinuous simulators require a prohibitively large number of simulation draws to obtain acceptable accuracy. In addition, a theoretical advantage of smoothness is that it greatly simplifies asymptotic theory. For these reasons, there has been a lot of research on various smooth simulators (see, for example, Börsch-Supan and Hajivassiliou, 1993; McFadden, 1989; Pakes and Pollard, 1989; and Stern, 1992). The discovery of the GHK simulator provided a smooth simulator for probit, which quickly became the standard for estimating probit models (see Hajivassiliou and Ruud, 1994). Now there is great interest in the logit kernel smooth simulator because it is conceptually intuitive, flexible, and relatively easy to program.

With simulation, the types and number of draws that are made from the underlying distribution to calculate the simulated probabilities are always important issues. Traditionally, simple pseudo-random draws (for example, Monte Carlo) have been used. Bhat (2000) and Train (1999) present an interesting addition to the econometric simulation literature, which is the use of intelligent drawing mechanisms (in many cases non-random draws known as Halton sequences). These draws are designed to cover the integration space in a more uniform way, and therefore can significantly reduce the number of draws required. We employ this approach for the empirical results presented later in this chapter.

[6] Later generalized to Ben-Akiva and Bolduc (1996).

A final point is that we use Maximum Likelihood Estimation (ML) or Maximum Simulated Likelihood (MSL). An alternative to this is the Method of Simulated Moments (MSM) proposed by McFadden (1989) and Pakes and Pollard (1989). MSM is often favored over MSL because a given level of accuracy in model parameter estimation can be obtained with a fairly small number of replication draws. The accuracy of the MSL methodology critically depends on using a large number of simulation draws because the log-likelihood function is simulated with a non-negligible downward bias. For several reasons, we still stick to the MSL approach. First, MSL requires the computation of the probability of only the chosen alternative, while MSM needs all choice probabilities. With large choice sets this factor can be quite important. Second, the objective function associated with MSL is numerically better behaved than the MSM objective function. Third, with the increase in computational power and the implementation of intelligent drawing mechanisms, the number of draws issue is not as critical as it once was.

The Logit Kernel Model

The Discrete Choice Model

Consider the following discrete choice model. For a given individual n, $n = 1,...,N$ where N is the sample size, and an alternative i, $i = 1,...,J_n$ where J_n is the number of alternatives in the choice set C_n of individual n, the model is written as:

$$y_{in} = \begin{cases} 1 & if\ U_{in} \geq U_{jn},\ for\ j = 1,...,J_n \\ 0 & otherwise \end{cases} ,$$

$$U_{in} = X_{in}b + e_{in} ,$$

where y_{in} indicates the observed choice, and U_{in} is the utility of alternative i as perceived by individual n. X_{in} is a $(1 \times K)$ vector of explanatory variables describing individual n and alternative i, including alternative-specific dummy variables as well as generic and alternative-specific attributes and their interactions with the characteristics of individual n. b is a $(K \times 1)$ vector of coefficients and e_{in} is a random disturbance. The assumption that the disturbances are i.i.d. Gumbel leads to the tractable, yet restrictive logit model. The assumption that the disturbances are multivariate normal distributed leads to the flexible, but computationally demanding probit model. The logit kernel model presented in this chapter is a hybrid between logit and probit and represents an effort to incorporate the advantages of each.

In a more compact vector form, the discrete choice model can be written as follows:

$$y_n = [y_{1n},...,y_{J_n n}]' ,$$

$$U_n = X_n b + e_n , \qquad [2\text{-}1]$$

where y_n, U_n, and e_n are $(J_n \times 1)$ vectors and X_n is a $(J_n \times K)$ matrix.

The Logit Kernel Model with Factor Analytic Form

Model Specification

In the logit kernel model, the e_{in} random utility term is made up of two components: a probit-like component with a multivariate distribution, and an i.i.d. Gumbel random variate. The probit-like term captures the interdependencies among the alternatives. We specify these interdependencies using a factor analytic structure[7], which is a flexible specification that includes all known error structures, as we will show below. It also has the ability of capturing complex covariance structures with relatively few parameters. This formulation of the logit kernel was originally presented in the working paper by Ben-Akiva and Bolduc (1996), and this chapter represents a major revision of that paper.

Using the factor analytic form, the disturbance vector e_n is specified as follows:

$$e_n = F_n x_n + n_n \ , \qquad [2\text{-}2]$$

where x_n is an $(M \times 1)$ vector of M multivariate distributed latent factors, F_n is a $(J_n \times M)$ matrix of the factor loadings that map the factors to the error vector (F_n includes fixed and/or unknown parameters and may also be a function of covariates), and v_n is a $(J_n \times 1)$ vector of i.i.d. Gumbel random variates. For estimation, it is desirable to specify the factors such that they are independent, and we therefore decompose x_n as follows:

$$x_n = T z_n \ , \qquad [2\text{-}3]$$

where z_n are a set of standard independent factors (often normally distributed), TT' is the covariance matrix of x_n, and T is the Cholesky factorization of it. The number of factors, M, can be less than, equal to, or greater than the number of alternatives. To simplify the presentation, we assume that the factors have standard normal distributions, however, they can follow any number of different distributions, such as lognormal, uniform, etc.

Substituting Equations [2-2] and [2-3] into Equation [2-1], yields:

<u>The Factor Analytic Logit Kernel Specification</u>

$$U_n = X_n b + F_n T z_n + n_n \ , \qquad [2\text{-}4]$$

$$\text{cov}(U_n) = F_n T T' F_n' + (g / m^2) I_{J_n} \qquad [2\text{-}5]$$

(which we denote as $\Omega_n = \Sigma_n + \Gamma_n$),

where: U_n is a $(J_n \times 1)$ vector of utilities;

X_n is a $(J_n \times K)$ matrix of explanatory variables;

---

[7] The Factor Analytic Structure was proposed for probit by McFadden (1984) as a means of reducing the dimensionality of the integral.

b is a $(K \times 1)$ vector of unknown parameters;

F_n is a $(J_n \times M)$ matrix of factor loadings, including fixed and/or unknown parameters;

T is a $(M \times M)$ lower triangular matrix of unknown parameters, where $TT' = Cov(x_n = Tz_n)$;

z_n is a $(M \times 1)$ vector of i.i.d. random variables with zero mean and unit variance; and

n_n is a $(J_n \times 1)$ vector of i.i.d. Gumbel random variables with zero location parameter and scale equal to $m > 0$. The variance is g/m^2, where g is the variance of a standard Gumbel ($p^2/6$).

The unknown parameters in this model are m, b, those in F_n, and those in T. X_n are observed, whereas z_n and n_n are unobserved.

It is important to note that we specify the model in level form (i.e., U_{jn}, $j = 1,..., J_n$) rather than in difference form (i.e., $(U_{jn} - U_{J_n n})$, $j = 1,\ldots,(J_n - 1)$). We do this for interpretation purposes, because it enables us to parameterize the covariance structure in ways that capture specific (and conceptual) correlation effects. Nonetheless, it is the difference form that is estimable, and there are multiple level structures that can represent any unique difference covariance structure. We return to this issue later in the chapter.

Response Probabilities

As will become apparent later, a key aspect of the logit kernel model is that if the factors z_n are known, the model corresponds to a multinomial logit formulation:

$$\Lambda(i \mid z_n) = \frac{e^{m\left(X_{in}b + F_{in}Tz_n\right)}}{\sum_{j \in C_n} e^{m\left(X_{jn}b + F_{jn}Tz_n\right)}} \quad , \qquad [2\text{-}6]$$

where $\Lambda(i \mid z_n)$ is the probability that the choice is i given z_n, and F_{jn} is j^{th} row of the matrix F_n, $j = 1,..., J_n$.

Since the z_n is in fact not known, the unconditional choice probability of interest is:

$$P(i) = \int_z \Lambda(i \mid z)n(z, I_M)dz \quad , \qquad [2\text{-}7]$$

where $n(z, I_M)$ is the joint density function of z, which, by construction, is a product of standard univariate normals:

$$n(z, I_M) = \prod_{m=1}^{M} f(z_m) \ .$$

The advantage of the logit kernel model is that we can naturally estimate $P(i)$ with an unbiased, smooth, tractable simulator, which we compute as:

$$\hat{P}(i) = \frac{1}{\mathbb{D}} \sum_{d=1}^{\mathbb{D}} \Lambda(i \mid z_n^d) \ ,$$

where z_n^d denotes draw d from the distribution of z, thus enabling us to estimate high dimensional integrals with relative ease.

Finally, note that if $T = 0$ then the model reduces to logit.

Identification and Normalization

It is not surprising that the estimation of such models raises identification and normalization issues. There are two sets of relevant parameters that need to be considered: the vector b and the unrestricted parameters of the distribution of the disturbance vector e_n, which include F_n, T, and m. For the vector b, identification is identical to that for a multinomial logit model. Such issues are well understood, and the reader is referred to Ben-Akiva and Lerman (1985) for details.

The identification of the parameters in error structure is more complex, and will be discussed in detail in this chapter.

Comments on Identification of Pure Probit versus Logit Kernel

Recall that the error structure of the logit kernel model consists of a probit-like component and an additive i.i.d. extreme value term (the Gumbel). Bolduc (1992), Bunch (1991), Dansie (1985) and others address identification issues for disturbance parameters in the multinomial probit model. Bunch (1991) presents clear guidelines for identification (consisting of Order and Rank conditions, which are described below) and provides examples of identified and unidentified error structures. He also provides a good literature review of the investigations into probit identification issues. For the most part, the identification guidelines for pure probit are applicable to the probit-like component of the logit kernel model. However, there are some differences, which are touched on here, and will be expanded on in the detailed discussion that follows.

We will see below that by applying the mechanics that are used to determine identification of a Probit model (Order and Rank) to the logit kernel model, effectively what happens is that the number of identifying restrictions that were necessary for a pure probit model are also required for the probit-like portion of the logit kernel model. However, there are some subtle, yet important, differences. Recall that one constraint is always necessary to set the scale of the model. In a pure probit model, this is done by setting at least one of the elements of the covariance structure[8] to some positive value (usually 1). Call this element that is constrained s_p. With logit kernel, on the other hand, the scale of the model is set as in a standard logit model by constraining the m parameter of the i.i.d. Gumbel term. Since the scale of the logit kernel model is set by m, the normalization of s_p is now a regular identifying restriction in the logit kernel model. One issue with the normalization of s_p for the logit kernel model is that in order to be able to trivially test the hypothesis that a logit kernel model is statistically different from a pure logit model, it is desirable to set s_p equal to zero so that pure logit is a special case of a logit kernel specification. A second difference is that while the specific element of the covariance matrix that is used to set the scale in a probit model is arbitrary, the selection of s_p is not necessarily arbitrary in the equivalent logit kernel model. This is due to the structure of the logit kernel model, and will be explained further below (in the discussion of the 'positive definiteness' condition.)

Finally, it turns out that the fact that s_p must be constrained in a logit kernel model is not exactly correct. In a *probit* kernel model (i.e., with an i.i.d. normal term), it is true that s_p must be constrained. In this case, there is a perfect trade-off between the multivariate normal term and the i.i.d. normal term. However, in the logit kernel model, this perfect trade-off does not exist because of the slight difference between the Gumbel and Normal distributions. Therefore, there will be an optimal combination of the Gumbel and Normal distribution, and this effectively allows another parameter to be estimated. This leads to somewhat surprising results. For example, in a heteroscedastic logit kernel model a variance term can be estimated for *each* of the alternatives, whereas probit, probit kernel, or extreme value logit requires that one of the variances be constrained. The same holds true for an unrestricted covariance structure. Nonetheless, the reality is that without the constraint, the model is nearly singular (i.e., the objective function is very flat at the optimum), as will be demonstrated in the estimation results that follow. Due to the near singularity, it is advisable to impose the additional constraint, and we proceed using this approach throughout the rest of the discussion.

Overview of Identification

The first step of identification is to determine the model of interest, that is, the disturbance structure that is a priori assumed to exist. For example, an unrestricted covariance matrix (of utility differences) or various restricted covariance matrices such as heteroscedasticity or nesting. Once that is determined, there are three steps to determining the identification and normalization of the hypothesized model. The first two have to do with identification. For the model to be identified, both the order condition (necessary) and the rank condition (sufficient) must hold. The order condition establishes the maximum number of parameters that can be estimated, which is based on the number of alternatives in the choice set. The rank condition establishes the actual number of parameters that can be estimated, which is based on the number of independent equations available. In cases in which the conclusion from the order and rank conditions is that additional restrictions are in order, then a third condition (which we refer to as the positive definiteness condition) is necessary to verify that the chosen normalization is valid. Recall that the reason that an identifying restriction is necessary is that there are an infinite number of solutions (i.e., parameter

[8] Technically, the constraint is on the covariance matrix of utility *differences*.

estimates) to match the given model structure. The point of an identifying restriction is to establish the existence of a single unique solution, but not change the underlying model in any way. The positive definiteness condition asks the question of whether the models true structure (i.e., the one on which the rank and order conditions were applied) is maintained given the chosen identifying restriction. This is not an important issue for probit, but, as we will see, it has important implications for logit kernel. Each of the conditions is expanded on below, and we use the heteroscedastic logit kernel model to illustrate each condition.

The Specification of the Heteroscedastic Logit Kernel Model

The heteroscedastic model, assuming a universal choice set $(C_n = C \ \ \forall n)$, is written as:[9]

Vector notation: $U_n = X_n b + T z_n + n_n$, $(M = J$ and F_n equals the identity matrix $I_J)$,

$$T = \begin{bmatrix} s_1 & & & \\ 0 & s_2 & & \\ 0 & 0 & \ddots & \\ 0 & 0 & 0 & s_J \end{bmatrix} (J \times J), \quad z_n \ (J \times 1),$$

and, defining $s_{ii} = (s_i)^2$, the $Cov(U_n)$ is:

$$\Omega = \begin{bmatrix} s_{11} + g/m^2 & & & \\ 0 & s_{22} + g/m^2 & & \\ 0 & 0 & \ddots & \\ 0 & 0 & 0 & s_{JJ} + g/m^2 \end{bmatrix} (J \times J).$$

Scalar notation: $U_{in} = X_{in} b + s_i z_{in} + n_{in}$, $i \in C$.

Note that for a heteroscedastic model with a universal choice set, the covariance matrix does not vary across the sample, and so we can drop the subscript n from Ω_n.

We carry the identification conditions through for a binary heteroscedastic model, a three alternative heteroscedastic model, and a four alternative heteroscedastic model, because the three models serve well to highlight various aspects of identification and normalization. The covariance structures for these three models are as follows:

$$J = 2: \ \Omega = \begin{bmatrix} s_{11} + g/m^2 & \\ 0 & s_{22} + g/m^2 \end{bmatrix},$$

[9] Note that our notation for symmetric matrices is to show only the lower triangular portion.

$$J=3:\ \Omega=\begin{bmatrix} s_{11}+g/m^2 & & \\ 0 & s_{22}+g/m^2 & \\ 0 & 0 & s_{33}+g/m^2 \end{bmatrix},$$

$$J=4:\ \Omega=\begin{bmatrix} s_{11}+g/m^2 & & & \\ 0 & s_{22}+g/m^2 & & \\ 0 & 0 & s_{33}+g/m^2 & \\ 0 & 0 & 0 & s_{44}+g/m^2 \end{bmatrix}.$$

Setting the Location

The general approach to identification of the error structure is to examine the covariance matrix of utility differences, denoted in the general case as Ω_{n,Δ_j}. Taking the differences sets the "location" of the model, a necessity for random utility models. The covariance matrix of utility differences for any individual is:

$$\Omega_{n,\Delta_j}=Cov(\Delta_j U_n)=\Delta_j F_n T T' F_n' \Delta_j' + \Delta_j (g/m^2) I_J \Delta_j',$$

where Δ_j is the linear operator that transforms the J utilities into $(J-1)$ utility differences taken with respect to the j^{th} alternative. Δ_j is a $(J-1)\times J$ matrix that consists of a $(J-1)\times(J-1)$ identity matrix with a column vector of -1's inserted as the j^{th} column. We use the notation $\Omega_{n,\Delta}$ to denote the covariance matrix of utility differences taken with respect to the J^{th} alternative.

Setting the Location for the Heteroscedastic Model

For the example heteroscedastic models using J as the base, the covariance matrices of utility differences are as follows:

$$J=2:\ \Delta_J=\begin{bmatrix}1 & -1\end{bmatrix}, \qquad \Omega_\Delta=\begin{bmatrix} s_{11}+s_{22}+2g/m^2 \end{bmatrix},$$

$$J=3:\ \Delta_J=\begin{bmatrix}1 & 0 & -1\\ 0 & 1 & -1\end{bmatrix}, \qquad \Omega_\Delta=\begin{bmatrix} s_{11}+s_{33}+2g/m^2 & \\ s_{33}+g/m^2 & s_{22}+s_{33}+2g/m^2 \end{bmatrix},$$

$$J=4:\ \Delta_J=\begin{bmatrix}1 & 0 & 0 & -1\\ 0 & 1 & 0 & -1\\ 0 & 0 & 1 & -1\end{bmatrix},$$

$$\Omega_\Delta=\begin{bmatrix} s_{11}+s_{44}+2g/m^2 & & \\ s_{44}+g/m^2 & s_{22}+s_{44}+2g/m^2 & \\ s_{44}+g/m^2 & s_{44}+g/m^2 & s_{33}+s_{44}+2g/m^2 \end{bmatrix}.$$

Order Condition

The first condition is the order condition, which is necessary for identification. When discussing the Order Condition, it is useful to separate the covariance matrix into that which is constant across the sample (called the 'alternative-specific' portion) and that which varies across the sample (for example, in the case of random parameters). The order condition only applies to the alternative-specific portion of the covariance matrix. It states that a maximum of $s = J(J-1)/2 - 1$ alternative-specific parameters are estimable in Ω, which is equal to the number of distinct cells in Ω_Δ (symmetric) minus 1 to set the scale (another necessity of random utility models). Therefore:

with 2 alternatives, no alternative-specific covariance terms can be identified;
with 3 alternatives, up to 2 terms can be identified;
with 4 alternatives, up to 5 terms can be identified;
with 5 alternatives, up to 9 terms can be identified;
etc.

When the error structure has parameters that are not alternative-specific, for example, random parameters, it is possible to estimate more than s parameters, because there is additional information derived from the variations of the covariance matrix across individuals. Technically, there still is an order condition, but the limit is large (related to the size of the sample) and is therefore never a limiting condition.

The Order Condition and the Heteroscedastic Model

The disturbance parameters in the heteroscedastic model are alternative-specific, so the order condition must hold. Each heteroscedastic model has $J+1$ unknown parameters: J s_{ii}'s and one m. The order condition then provides the following information regarding identification:

$J = 2$: $unknowns = \{s_{11}, s_{22}, m\}$; $s = 0$ → 0 variances are identified

$J = 3$: $unknowns = \{s_{11}, s_{22}, s_{33}, m\}$; $s = 2$ → up to 2 variances are identified

$J = 4$: $unknowns = \{s_{11}, s_{22}, s_{33}, s_{44}, m\}$; $s = 5$ → potentially *all* variances are identified

Note that there are published probit and logit kernel models in the literature that do not meet the order condition, see, for example, Greene (2000) Table 19.15 and Louviere et al. (2000) Table B.6. While the logit kernel models in Greene and Louviere do not meet the order condition, these models are nonetheless barely identified due to the slight difference between the normal and Gumbel distributions (as discussed earlier). However, the probit model does not have this luxury, and therefore the probit model reported in Greene is not identified (as will be demonstrated in the mode choice application).

While the order condition provides a quick check for identification, it is clearly shown in Bunch (1991) that the number of parameters that can be estimated is often less than s, depending on the covariance structure postulated. Therefore, the rank condition must also be checked, which is described next.

Rank Condition

The rank condition is more restrictive than the order condition, and it is a sufficient condition for identification. The order condition simply counts cells, and ignores the internal structure of Ω_Δ. The rank condition, however, counts the number of linearly independent equations available in Ω_Δ that can be used to estimate the parameters of the error structure. Bolduc (1992) and Bunch (1991) describe the mechanics of programming the rank condition. The basic idea behind determining this count is to examine the Jacobian matrix, which is equal to the derivatives of the elements in Ω_Δ with respect to the unknown parameters. The number of parameters that can be estimated is equal to the Rank of the Jacobian matrix minus 1 (to set the scale). These mechanics are demonstrated below with the heteroscedastic example.

The Rank Condition and the Heteroscedastic Model

The first step is to vectorize the unique elements of Ω_Δ into a column vector (we call this operator *vecu*):[10]

$$J=3:\ vecu(\Omega_\Delta)=\begin{bmatrix} s_{11}+s_{33}+2g/m^2 \\ s_{22}+s_{33}+2g/m^2 \\ s_{33}+g/m^2 \end{bmatrix},$$

$$J=4:\ vecu(\Omega_\Delta)=\begin{bmatrix} s_{11}+s_{44}+2g/m^2 \\ s_{22}+s_{44}+2g/m^2 \\ s_{33}+s_{44}+2g/m^2 \\ s_{44}+g/m^2 \end{bmatrix}.$$

By examination, it is clear that we are short an equation in both cases. This is formally determined by examining the Rank of the Jacobian matrix of vecu(Ω_Δ) with respect to each of the unknown parameters $(s_{11},...,s_{JJ},g/m^2)$:

$$J=3:\ \begin{matrix}Jacobian\\ matrix\ \ of\\ vecu(\Omega_\Delta)\end{matrix}=\begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix},\ Rank=3 \quad \rightarrow \quad \begin{matrix}can\ estimate\ 2\ of\ the\ parameters; \\ must\ normalize\ m\ and\ one\ s_{ii}.\end{matrix}$$

$$J=4:\ \begin{matrix}Jacobian\\ matrix\ \ of\\ vecu(\Omega_\Delta)\end{matrix}=\begin{bmatrix} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix},\ Rank=4 \quad \rightarrow \quad \begin{matrix}can\ estimate\ 3\ of\ the\ parameters; \\ must\ normalize\ m\ and\ one\ s_{ii}.\end{matrix}$$

So for both of these cases, the scale term m as well as one of the s_{ii}'s must be normalized.

---

[10] Note that there's no need to continue with identification for the binary heteroscedastic case, since the order condition resolved that none of the error parameters are identified.

Which s_{ii} should be fixed? And to what value? This is where the positive definiteness condition comes into play, and it turns out that the normalizations for logit kernel models are not always arbitrary or intuitive.

Positive Definiteness

When the conclusion from the order and rank conditions is that further identifying restrictions (normalizations) are required, the positive definiteness condition is used to determine the set of acceptable normalizations. Conceptually, the need for the positive definiteness condition is as follows. First note that the reason for the additional normalization is that there are infinite possible solutions that result in the hypothesized covariance structure. The normalization is necessary to establish the existence of a unique solution, but it does not change the underlying model structure (i.e., the covariance matrix of utility differences) in any way. The positive definiteness condition is necessary to verify that the chosen normalization is valid, i.e., that the remaining parameters that are estimated are able to replicate the underlying model structure. It turns out that with logit kernel models, there can be seemingly obvious normalizations that are not valid, because the structure of the model prevents the underlying covariance matrix of utility differences from being recovered.

To work through the details of the positive definiteness condition, we rephrase the above discussion as follows. There are two overriding issues behind the positive definiteness condition:

Statement 1: There are infinite possible normalizations that can be imposed to identify the model. However, note that all valid normalizations for a particular specification will result in identical $W_{n,D}$, that is, { $W_{n,D}^{N1}$ from normalization 1} = { $W_{n,D}^{N2}$ from normalization 2}. For example, with this relationship, one can convert the estimated parameters from a particular normalization (say $s_{11} = 0$) to the parameters that will be estimated if a different normalization (say $s_{11} = 1$) is imposed (as long as both normalizations are valid).

Statement 2: The logit kernel covariance matrix is $\Omega_n = \Sigma_n + \Gamma_n$, where $\Sigma_n = (F_n T)(F_n T)'$ (Equation [2-5]). Therefore, by construction, Σ_n is necessarily positive semi-definite ('semi' because $F_n T$ can equal zero).

Given these two issues, any valid normalization must be such that both of the following conditions hold for all observations:

I. $\Omega_{n,\Delta}^{N} = \Omega_{n,\Delta} \quad \rightarrow \quad \Sigma_{n,\Delta}^{N} + \Gamma_{n,\Delta}^{N} = \Omega_{n,\Delta}$ (by definition of a normalization).

The covariance matrix of utility differences of the normalized model (denoted by N) equals the covariance matrix of utility differences of the non-normalized (theoretical) model.

II. Σ_n^N *is positive semi-definite* (by construction).

If the normalization is such that both Conditions I and II cannot be met, the parameter estimates will be inconsistent and result in a loss of fit. It turns out that for logit kernel, these conditions can impose restrictions on the feasible set of normalizations, as we describe below.

We have already stated that Condition II necessarily holds due to the construction of the model. Therefore, the issue is whether the imposed normalization is such that Condition I can be met, given the restriction that Σ_n^N is positive semi-definite. Problems can arise with logit kernel models due to the additive i.i.d. Gumbel portion of the covariance structure, Γ_n. Because of Γ_n, there can be normalizations for which satisfying Condition I requires a negative definite Σ_n^N. However, this conflicts with Condition II, and so any such normalization is not valid. Note that this issue actually arises with any model structure that includes an i.i.d. disturbance term along with a parameterized disturbance, for example, a probit kernel model.

Positive Definiteness and the Heteroscedastic Model

Looking at the heteroscedastic case, we will use the three alternative model as an example. It is useful in the analysis to deal directly with the estimated (i.e., scaled) parameters, so we introduce the notation $\dot{s}_{ii} = (m s_i)^2$. Say we impose the normalization that the third heteroscedastic term, $\dot{s}_{33}$, is constrained to some fixed value we denote as $\dot{s}_{ff}^N$. Condition I can then be written as:

$$\begin{bmatrix} (\dot{s}_{11}^N + \dot{s}_{ff}^N + 2g)/m_N^2 & \\ (\dot{s}_{ff}^N + g)/m_N^2 & (\dot{s}_{22}^N + \dot{s}_{ff}^N + 2g)/m_N^2 \end{bmatrix} = \begin{bmatrix} (\dot{s}_{11} + \dot{s}_{33} + 2g)/m^2 & \\ (\dot{s}_{33} + g)/m^2 & (\dot{s}_{22} + \dot{s}_{33} + 2g)/m^2 \end{bmatrix},$$

where the matrix on the left represents the normalized model $\left(\dot{s}_{ii}^N = (m_N s_i^N)^2\right)$ and the matrix on the right represents the theoretical (non-normalized) model. This relationship states that when the normalization is imposed, the remaining parameters in the normalized model will adjust such that the theoretical (or true) covariance matrix of utility differences is recovered. It also provides us with three equations:

$$(\dot{s}_{ff}^N + g)/m_N^2 = (\dot{s}_{33} + g)/m^2 \quad , \qquad [2\text{-}8]$$

$$(\dot{s}_{11}^N + \dot{s}_{ff}^N + 2g)/m_N^2 = (\dot{s}_{11} + \dot{s}_{33} + 2g)/m^2 \quad \text{, and} \qquad [2\text{-}9]$$

$$(\dot{s}_{22}^N + \dot{s}_{ff}^N + 2g)/m_N^2 = (\dot{s}_{22} + \dot{s}_{33} + 2g)/m^2 \quad . \qquad [2\text{-}10]$$

Condition II states that Σ^N must be positive semi-definite, where:

$$\Sigma^N = \begin{bmatrix} \dot{s}_{11}^N & & \\ 0 & \dot{s}_{22}^N & \\ 0 & 0 & \dot{s}_{ff}^N \end{bmatrix} * \frac{1}{m_N^2} \quad .$$

This matrix is positive semi-definite if and only if the diagonal entries are non-negative and m_N^2 is strictly positive, or:

$$m_N^2 > 0 \ , \qquad [2\text{-}11]$$
$$\dot{s}_{11}^{\,N} \geq 0 \ , \qquad [2\text{-}12]$$
$$\dot{s}_{22}^{\,N} \geq 0 \ \text{, and} \qquad [2\text{-}13]$$
$$\dot{s}_{ff}^{\,N} \geq 0 \ . \qquad [2\text{-}14]$$

The positive definiteness condition requires that all valid normalizations satisfy the restrictions stated by Equations [2-8] to [2-14]. The question is, what values of $\dot{s}_{ff}^{\,N}$ guarantee that these relationships hold?

To derive the restrictions on $\dot{s}_{ff}^{\,N}$, we first use Condition I (Equations [2-8] to [2-10]) to develop equations for the unknown parameters of the normalized model $(m_N^2,\ \dot{s}_{11}^{\,N},\ and\ \dot{s}_{22}^{\,N})$ as functions of the normalized parameter $\dot{s}_{ff}^{\,N}$ and the theoretical parameters $(m^2,\ \dot{s}_{11},\ \dot{s}_{22},\ and\ \dot{s}_{33})$, which leads to:

$$m_N^2 = m^2(\dot{s}_{ff}^{\,N} + g)\big/(\dot{s}_{33} + g) \ , \qquad [2\text{-}15]$$
$$\dot{s}_{11}^{\,N} = \left((\dot{s}_{11} + g)\dot{s}_{ff}^{\,N} + (\dot{s}_{11} - \dot{s}_{33})g\right)\big/\left(\dot{s}_{33} + g\right) \text{, and} \qquad [2\text{-}16]$$
$$\dot{s}_{22}^{\,N} = \left((\dot{s}_{22} + g)\dot{s}_{ff}^{\,N} + (\dot{s}_{22} - \dot{s}_{33})g\right)\big/\left(\dot{s}_{33} + g\right) \ . \qquad [2\text{-}17]$$

Equations [2-11] to [2-14] impose restrictions on the parameters of the normalized model, and so we can combine them with Equations [2-15] to [2-17], which results in the following set of restrictions:

$$\dot{s}_{ff}^{\,N} \geq 0 \ , \qquad \text{(Eq. [2-14])} \qquad [2\text{-}18]$$
$$m^2(\dot{s}_{ff}^{\,N} + g)\big/(\dot{s}_{33} + g) > 0 \ , \qquad \text{(Eqs. [2-11] \& [2-15])} \qquad [2\text{-}19]$$
$$\left((\dot{s}_{11} + g)\dot{s}_{ff}^{\,N} + (\dot{s}_{11} - \dot{s}_{33})g\right)\big/\left(\dot{s}_{33} + g\right) \geq 0 \ \text{, and} \qquad \text{(Eqs. [2-12] \& [2-16])} \qquad [2\text{-}20]$$
$$\left((\dot{s}_{22} + g)\dot{s}_{ff}^{\,N} + (\dot{s}_{22} - \dot{s}_{33})g\right)\big/\left(\dot{s}_{33} + g\right) \geq 0 \ . \qquad \text{(Eqs. [2-13] \& [2-17])} \qquad [2\text{-}21]$$

The other information we have is that Σ is positive semi-definite (by construction), and therefore:

$$m^2 > 0 ,\ \dot{s}_{11} \geq 0 ,\ \dot{s}_{22} \geq 0 \text{, and } \dot{s}_{33} \geq 0 . \qquad [2\text{-}22]$$

So going back to restrictions [2-18]-[2-21], the first two restrictions are trivial: Equation [2-18] just states that the normalization has to be non-negative; and given Equations [2-18] and [2-22] , Equation [2-19] will always be satisfied. Equations [2-20] and [2-21] are where it gets interesting, because solving for $\dot{s}_{ff}^{\,N}$ leads to the following restrictions on the normalization:

$$\dot{s}_{ff}^{N} \geq \left(\dot{s}_{33} - \dot{s}_{ii}\right) g \Big/ (g + \dot{s}_{ii}) \quad , \ i = 1,2 \ . \qquad [2\text{-}23]$$

($\dot{s}_{33}$ is the heteroscedastic term that is fixed.)

What does this mean? Note that as long as alternative 3 is the minimum variance alternative, the right hand side of Equation [2-23] is negative, and so the restriction is satisfied for any $\dot{s}_{ff}^{N} \geq 0$. However, when alternative 3 is not the minimum variance alternative, $\dot{s}_{ff}^{N}$ must be set "large enough" (and certainly above zero) such that Equation [2-23] is satisfied. This latter approach to normalization is not particularly practical since the $\dot{s}_{ii}$ are unknown (how large is large enough?), and it has the drawback that MNL is not a case nested within the logit kernel specification. Therefore, the following normalization is recommended:

The preferred normalization for the heteroscedastic logit kernel model is to constrain the heteroscedastic term of the minimum variance alternative to zero.

A method for implementing this normalization is described later in the section on heteroscedastic logit kernel models.

Positive Definiteness and a Probit Model

What about the positive definiteness condition for pure probit? Pure probit models also must satisfy a positive definiteness condition, but it turns out that these do not impose any problematic restrictions on the normalization. With pure probit, there is obviously no Gumbel term, so Condition I can be written as $\Sigma_{n,\Delta}^{N} = \Sigma_{n,\Delta}$. Condition II is similar to that for logit kernel, except that Σ_{n}^{N} must now be positive definite (since it cannot equal zero). Since $\Sigma_{n,\Delta}$ is well-behaved (by construction), Condition I states that $\Sigma_{n,\Delta}^{N}$ will also be well-behaved, and, therefore, so will Σ_{n}^{N}. The result is that the positive definiteness condition automatically holds for normalizations that are intuitively applied to probit.

Positive Definiteness and a *Probit* Heteroscedastic Model

This can be demonstrated for the heteroscedastic pure probit case, Condition I is:

$$\begin{bmatrix} (\dot{s}_{11}^{N} + \dot{s}_{ff}^{N}) / \tilde{m}_{N}^{2} & \\ (\dot{s}_{ff}^{N}) / \tilde{m}_{N}^{2} & (\dot{s}_{22}^{N} + \dot{s}_{ff}^{N}) / \tilde{m}_{N}^{2} \end{bmatrix} = \begin{bmatrix} (\dot{s}_{11} + \dot{s}_{33}) / \tilde{m}^{2} & \\ (\dot{s}_{33}) / \tilde{m}^{2} & (\dot{s}_{22} + \dot{s}_{33}) / \tilde{m}^{2} \end{bmatrix},$$

where $\tilde{m}$ is the scale of the probit model (i.e., not the traditional Gumbel m).

Solving for the unknown parameters from the normalized model:

$$\tilde{m}_{N}^{2} = \tilde{m}^{2} \dot{s}_{ff}^{N} / \dot{s}_{33} \ ,$$
$$\dot{s}_{11}^{N} = \dot{s}_{11} \dot{s}_{ff}^{N} / \dot{s}_{33} \ , \text{ and}$$
$$\dot{s}_{22}^{N} = \dot{s}_{22} \dot{s}_{ff}^{N} / \dot{s}_{33} \ .$$

Condition II requires:

$\tilde{m}_N^2 > 0$,
$\dot{s}_{11}^{N} > 0$,
$\dot{s}_{22}^{N} > 0$, and
$\dot{s}_{ff}^{N} > 0$.

Given that the theoretical Σ_{Δ} is well behaved (i.e., all theoretical variances and scale are strictly positive), it is clear that any $\dot{s}_{ff}^{N} > 0$ will result in Conditions I and II being satisfied. So, the normalization is arbitrary, and the standard practice of normalizing any one of the terms to 1 is valid.

Examination of the normalization unrestricted probit and logit kernel models are provided in Appendix A. The heteroscedastic and unrestricted covariance matrix examples illustrate the nature of the problem. The issue arises due to the manner in which the normalized parameter estimates adjust to replicate the true covariance structure. With probit, the parameters shift in a simple multiplicative manner. However, with logit kernel, the parameters shift in an additive manner, and this can lead to infeasible 'negative' variances and a factor analytic term that is not positive definite.

The brief summary of identification is that the order and rank conditions need to be applied to verify that any estimated model is identified, and the positive definiteness condition needs to be applied to verify that a particular normalization is valid. It is critical to examine identification on a case-by-case basis, which is how we will proceed in the remainder of the chapter. There is also an empirical issue concerning identification, which is whether or not the data provide enough information to estimate any given theoretically identified structure. This is the usual multicollinearity problem, and it arises when there are too many parameters in the error structure and therefore the Hessian is nearly singular.

Special Cases

Many interesting cases can be embedded in the general factor analytic logit kernel specification presented in Equation [2-4]. We will cover the following special cases in this section:

- *Heteroscedastic* – a summary and generalization of the discussion above.
- *Nested and Cross-nested* – analogous to nested and cross-nested logit.
- *Error Components* – a generalization of heteroscedastic and nested structures.
- *Factor Analytic* – a further generalization in which parameters in F_n are also estimated.
- *General Auto-Regressive* – particularly useful for large choice sets.
- *Random parameters* – where most of the current applications of logit kernel in the literature are focused.

This is not meant to be an exhaustive list. There are certainly other special cases of the logit kernel model, some of which are presented in papers listed in the references. The objective of this section is to show the flexibility of logit kernel, to provide specific examples of specification and identification, and to establish rules for identification and normalization for some of the most common special cases.

Heteroscedastic

The heteroscedastic model was presented above. The scalar notation form of the model is repeated here for convenience:

$$U_{in} = X_{in} b + s_i z_{in} + n_{in} , \qquad i \in C_n .$$

Identification

Identification was already discussed above for $J = 2, 3$, and 4. These results can be straightforwardly generalized to the following:

Identification

$J = 2$ none of the heteroscedastic variances can be identified.

$J \geq 3$ $J - 1$ of the heteroscedastic variances can be identified.

Normalization

For $J \geq 3$, a normalization must be imposed on one of the variance terms, denote this as $\dot{s}_{jj} = \dot{s}_{ff}^{N}$ where $\dot{s}_{jj}$ is the true, albeit unknown, variance term that is fixed to the value $\dot{s}_{ff}^{N}$.

This normalization is not arbitrary, and must meet the following restriction:

$$\dot{s}_{ff}^{N} \geq \left(\dot{s}_{jj} - \dot{s}_{ii} \right) g \big/ (g + \dot{s}_{ii}) \quad , \; i = 1,..., J \; .$$

This restriction shows that the natural tendency to normalize an arbitrary heteroscedastic term to zero is incorrect. If the alternative does not happen to be the minimum variance alternative, the parameter estimates will be inconsistent, there can be a significant loss of fit (as demonstrated in the application section), and it can lead to the incorrect conclusion that the model is homoscedastic. This is an important issue, which, as far as we can tell, is ignored in the literature. It appears that arbitrary normalizations are being made for models of this form (see, for example Gönül and Srinivasan, 1993, and Greene, 2000, Table 19.15). Therefore, there is a chance that a non-minimum variance was normalized to zero, which would mean that the model is misspecified. It is important to note that it is the addition of the i.i.d. disturbance that causes the identification problem. Therefore, heteroscedastic pure probit models as well as the heteroscedastic extreme value models (see, for example, Bhat, 1995, and Steckel and Vanhonacker, 1988) do not exhibit this property.

Ideally, we would like to impose a normalization such that MNL is a special case of the model. Therefore, the best normalization is to fix the minimum variance alternative to zero. However, there is in practice no prior knowledge of the minimum variance alternative. A brute force solution is to estimate J versions of the model, each with a different heteroscedastic term normalized; the model with the best fit is the one with the correct normalization. This is obviously cumbersome as well as time consuming. Alternatively, one can estimate the unidentified model with all J heteroscedastic terms. Although this model is not identified, it will pseudo-converge to a point that reflects the true covariance structure of the model. The heteroscedastic term with minimum estimated variance in the unidentified model is the minimum variance alternative, thus eliminating the need to estimate J different models. Examples of this method are provided in the applications section.

Nesting & Cross-Nesting Error Structures

Nesting and cross-nesting logit kernel is another important special case, and is analogous to nested and cross-nested logit. The nested logit kernel model is specified as follows:

$$U_n = X_n b + F_n T z_n + n_n \ ,$$

where: z_n is $(M \times 1)$, M is the number of nests, and one factor is defined for each nest.

F_n is $(J_n \times M)$, $f_{jm} = \begin{cases} 1 & \text{if alternative } j \text{ is a member of nest } m \\ 0 & \text{otherwise} \end{cases}$

T is $(M \times M)$ diagonal, which contains the standard deviation of each factor.

In a strictly hierarchical nesting structure, the nests do not overlap, and $F_n F_n'$ is block diagonal. In a cross-nested structure, the alternatives can belong to more than one group.

Identification

As usual, the order and rank conditions are checked for identification. The order condition states that at most $J(J-1)/2-1$ nesting parameters can be identified. However, the rank condition leads to further restrictions as described below.

Models with 2 Nests

The summary of identification for a 2 nest structure is that only 1 of the nesting parameters is identified. Furthermore, the normalization of the nesting parameter is arbitrary. This is best shown by example. Take a 5 alternative case (with universal choice set) in which the first 2 alternatives belong to one nest, and the last 3 alternatives belong to a different nest. The model is written as:

$$\begin{aligned} U_{1n} &= \ldots + s_1 z_{1n} + n_{1n} \\ U_{2n} &= \ldots + s_1 z_{1n} + n_{2n} \\ U_{3n} &= \ldots + s_2 z_{2n} + n_{3n} \\ U_{4n} &= \ldots + s_2 z_{2n} + n_{4n} \\ U_{5n} &= \ldots + s_2 z_{2n} + n_{5n} \end{aligned} \ , \quad \text{where: } F = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \text{ and } T = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} .$$

We denote this specification as 1, 1, 2, 2, 2 (a shorthand notation of the matrix F). The covariance matrix of utility differences (with alternative 5 as the base) is as follows:

$$\Omega_{\Delta} = \begin{bmatrix} s_{11} + s_{22} + 2g/m^2 & & & \\ s_{11} + s_{22} + g/m^2 & s_{11} + s_{22} + 2g/m^2 & & \\ g/m^2 & g/m^2 & 2g/m^2 & \\ g/m^2 & g/m^2 & g/m^2 & 2g/m^2 \end{bmatrix}.$$

It can be seen from this matrix that only the sum $(s_{11} + s_{22})$ can be identified. This is verified by the rank condition as follows:

$$vecu(\Omega_{\Delta}) = \begin{bmatrix} s_{11} + s_{22} + 2g/m^2 \\ s_{11} + s_{22} + g/m^2 \\ g/m^2 \\ 2g/m^2 \end{bmatrix} \quad \rightarrow \quad \begin{matrix} Jacobian \\ matrix \end{matrix} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 2 \end{bmatrix} \quad \rightarrow \quad \text{RANK=2}$$

$\rightarrow$ can estimate 1 of the parameters; must normalize m and one s_{ii} .

Furthermore, unlike the heteroscedastic logit kernel model, either one of the variance terms can be normalized to zero (i.e., the normalization is arbitrary). This can be seen intuitively by noticing that only the sum $(s_{11} + s_{22})$ appears in Ω_{Δ} , and so it is always this sum that is estimated regardless of which term is set to zero. This can also be verified via the positive definiteness condition, as follows. Say we impose the normalization $\dot{s}_{22}^{N} = 0$. Condition I leads to the relationships $m_N^2 = m^2$ and $\dot{s}_{11}^{N} = (\dot{s}_{11} + \dot{s}_{22})$. Condition II states that Σ^N must be positive semi-definite, where:

$$\Sigma^N = \begin{bmatrix} \dot{s}_{11}^{N} & & & & \\ \dot{s}_{11}^{N} & \dot{s}_{11}^{N} & & & \\ 0 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} * \frac{1}{m_N^2}.$$

A matrix is positive semi-definite if all of its eigenvalues are non-negative. The eigenvalues for Σ^N shown above are: $2\dot{s}_{11}^{N} / m_N^2,\ 0,\ 0,\ 0,\ 0$. We know from Condition I that $m_N^2 > 0$ and $\dot{s}_{11}^{N} \geq 0$, which means $2s_{11}^{N} / m_N^2 \geq 0$, Σ^N is positive semi-definite, and the normalization $\dot{s}_{22}^{N} = 0$ is valid. Similarly, it can be shown that the normalization $\dot{s}_{11}^{N} = 0$ is also valid.

While it is not possible to estimate both variance parameters of the 1, 1, 2, 2, 2 structure, the following structures are all identified and result in *identical* covariance structures (i.e., identical models):

{ 1, 1, 0, 0, 0 } = { 0, 0, 2, 2, 2 } = { 1, 1, 2, 2, 2 with $s_1 = s_2$ } .

These results straightforwardly extend to all two nest structures regardless of the number of alternatives (as long as at least one of the nests has 2 or more alternatives).

Models with Three or More Nests

The summary of identification for models with 3 or more nests is that *all* of the nesting parameters are identified. To show this, we will again look at a 5 alternative model, this time imposing a 3 nest structure (1, 1, 2, 3, 3):

$$\begin{aligned} U_{1n} &= ... + s_1 z_{1n} + n_{1n} \\ U_{2n} &= ... + s_1 z_{1n} + n_{2n} \\ U_{3n} &= ... + s_2 z_{2n} + n_{3n} \\ U_{4n} &= ... + s_3 z_{3n} + n_{4n} \\ U_{5n} &= ... + s_3 z_{3n} + n_{5n} \end{aligned} \quad , \quad \text{where: } F = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } T = \begin{bmatrix} s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & s_3 \end{bmatrix}.$$

The covariance matrix of utility differences is:

$$\Omega_\Delta = \begin{bmatrix} s_{11} + s_{33} + 2g/m^2 & & & \\ s_{11} + s_{33} + g/m^2 & s_{11} + s_{33} + 2g/m^2 & & \\ s_{33} + g/m^2 & s_{33} + g/m^2 & s_{22} + s_{33} + 2g/m^2 & \\ g/m^2 & g/m^2 & g/m^2 & 2g/m^2 \end{bmatrix}.$$

A check of the rank condition verifies that all three variance parameters are identified:

$$vecu(\Omega_\Delta) = \begin{bmatrix} s_{11} + s_{33} + 2g/m^2 \\ s_{11} + s_{33} + g/m^2 \\ s_{33} + g/m^2 \\ s_{22} + s_{33} + 2g/m^2 \\ g/m^2 \\ 2g/m^2 \end{bmatrix} \rightarrow \begin{matrix} Jacobian \\ matrix \end{matrix} = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix} \rightarrow \text{RANK=4}$$

$\rightarrow$ can estimate 3 of the parameters; only need to normalize m.

It is an interesting result that 1, 1, 0, 2, 2 structure results in both variance parameters being identified (by virtue of having a 3 nest structure) whereas only one parameter of the 1, 1, 2, 2, 2 structure is identified.

Conceptually, the number of estimable parameters can be thought of in terms of the number of differences and number of covariances that are left in the utility differences. In a two nest structure, only one difference remains and no covariances and therefore one parameter is estimable. Whereas in a three nest structure, there are two differences, plus the covariance between these two differences, and so three parameters are estimable.

This finding can be extended to any model with 3 or more nests (where 'nests' can have only 1 alternative, as long as at least one nest has 2 or more alternatives) as follows. Without loss of generality, assume that the base alternative is a member of a nest with 2 or more alternatives (as in the example above). Define m_b as the group to which the base alternative belongs, and s_{bb} as the variance associated with this base. Recall that M is the number of nests. The covariance matrix of utility differences has the following elements:

On the diagonal:

$s_{ii} + s_{bb} + 2g/m^2 \quad \forall\, i \notin m_b$, *M-1* equations, [2-24]

$2g/m^2$, *1* equation. [2-25]

On the off-diagonal:

$s_{bb} + g/m^2$, *1* equation, [2-26]

g/m^2 , irrelevant: a dependent equation,

$s_{ii} + s_{bb} + g/m^2$ for some $i \notin m_b$, irrelevant: a dependent equation.

Equations [2-24] through [2-26] provide identification for all nesting parameters, and the remaining equations are dependent. In the two-nest case, Equation [2-26] does not exist, and thus is an equation short of identification.

Cross-Nested Models

There are no general rules for identification and normalization of cross-nested structures, and one has to check the rank condition on a case-by-case basis. For example, in the five alternative case in which the third alternative belongs to both nests (1, 1, 1-2, 2, 2), the (non-differenced) covariance matrix is:

$$\Omega = \begin{bmatrix} s_{11} + g/m^2 & & & & \\ s_{11} & s_{11} + g/m^2 & & & \\ s_{11} & s_{11} & s_{11} + s_{22} + g/m^2 & & \\ 0 & 0 & s_{22} & s_{22} + g/m^2 & \\ 0 & 0 & s_{22} & s_{22} & s_{22} + g/m^2 \end{bmatrix}.$$

A check of the order and rank conditions would find that both of the parameters in this cross-nested structure are identified. However, note that the cross-nesting specification can have unintended consequences on the covariance matrix. For example, in the (1, 1, 1-2, 2, 2) specification shown above, the third alternative is forced to have the highest variance. There are numerous possible solutions. One is to add a set of heteroscedastic terms, another is to add factors such that all the alternative-specific variances are identical as with the following specification:

$$F = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \text{ and } T = \begin{bmatrix} s_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & s_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & s_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & s_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & s_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & s_2 \end{bmatrix}.$$

The covariance matrix of utility differences for this structure is as follows:

$$\Omega_\Delta = \begin{bmatrix} 2s_{11} + 2s_{22} + 2g/m^2 & & & \\ 2s_{11} + s_{22} + g/m^2 & 2s_{11} + 2s_{22} + 2g/m^2 & & \\ 2s_{11} + g/m^2 & 2s_{11} + g/m^2 & 2s_{11} + 2g/m^2 & \\ s_{11} + g/m^2 & s_{11} + g/m^2 & s_{11} + g/m^2 & 2s_{11} + 2g/m^2 \end{bmatrix}.$$

A check of the rank condition verifies that both variance parameters are identified for this specification.

$$vecu(\Omega_\Delta) = \begin{bmatrix} 2s_{11} + 2s_{22} + 2g/m^2 \\ 2s_{11} + s_{22} + g/m^2 \\ 2s_{11} + g/m^2 \\ 2s_{11} + 2g/m^2 \\ s_{11} + g/m^2 \\ 2s_{11} + 2g/m^2 \end{bmatrix} \rightarrow \begin{matrix} Jacobian \\ matrix \end{matrix} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 1 & 1 \\ 2 & 0 & 1 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \\ 2 & 0 & 2 \end{bmatrix} \rightarrow \text{RANK=3}$$

$\rightarrow$ can estimate 2 of the parameters, only need to normalize m.

Extensions to Nested Models

There are various complexities that can be introduced to the nesting structure, including multi-level nests, cross-nested structures with multiple dimensions, and unknown parameters in the loading matrix (F). While we have investigated various special cases of these extended models, we have not yet derived general rules for identification. We recommend that identification be performed automatically on a case-by-case basis by programming the rank and order conditions into the estimation program.

Error Components

The error component formulation is a generalization that includes the heteroscedastic, nested, and cross-nested structures. The model is specified as follows:

$$U_n = X_n b + F_n T z_n + n_n ,$$

where F_n, z_n, and T are defined as in the general case, and F_n is a matrix of fixed factor loadings equal to 0 or 1. If T is diagonal (as it often is), then the disturbances in scalar form are:

$$e_{in} = \sum_{m=1}^{M} f_{imn} s_m z_{mn} + n_{in}, \qquad i \in C_n \ ,$$

where:

$$f_{imn} = \begin{cases} 1 & \textit{if the } m^{th} \textit{ element of } z_n \textit{ applies to alternative } i \textit{ for individual } n, \\ 0 & \textit{otherwise}. \end{cases}$$

The number of factors can be less than, equal to, or greater than the number of alternatives.

Identification

The order condition states that up to $J(J-1)/2-1$ parameters in T are identified. However, it is always necessary to check the rank condition for the particular specification and the positive definiteness condition for valid normalizations. Examples were provided above for the special cases of heteroscedastic, nesting, and cross-nesting specifications. Note that the rank condition should always be checked when any combination of nesting, cross-nesting, and heteroscedasticity are applied. That is, the identification rules cannot be independently applied for combinations.

Factor Analytic

The Factor Analytic specification is a further generalization in which the F_n matrix contains unknown parameters. The model is written as in the general case:

$$U_n = X_n b + F_n T z_n + n_n .$$

If T is diagonal, the disturbances can be written in scalar form as follows:

$$e_{in} = \sum_{m=1}^{M} f_{imn} s_m z_{mn} + n_{in}, \qquad i \in C_n \ ,$$

where both the f_{imn}'s and s_m's are unknown parameters.

Identification

This is a very broad class of models. Therefore, it is difficult to go beyond the rank and order generalizations of identification. However, note that some constraints must be imposed on F_n and T in order to achieve identification. For alternative-specific error structures, the minimum number of necessary constraints can be determined from the order condition: a maximum of $J(J-1)/2-1$ parameters can be estimated and there are up to $M(J+1)+1$ unknown parameters (M in T diagonal, JM in F_n, plus the scale term m). Once the order condition is met, the rank condition needs to be checked on a case-by-case basis. Finally, it must be verified that any imposed normalization satisfies the positive definiteness condition.

General Autoregressive Process

A fully unrestricted error correlation structure in models with large choice sets is problematic as the dimension of the integral is on the order of the number of alternatives and the number of parameters grows quadratically with the number of alternatives. A generalized autoregressive framework is attractive in these situations, because it allows one to capture fairly general error correlation structures using parsimonious parametric specifications. The key advantage of the method is that the number of parameters in the error structure grows linearly with the size of the choice set.

The disturbances $\dot{x}_n = (\dot{x}_{1n},...,\dot{x}_{J_n n})'$ [11] of a first-order generalized autoregressive process [GAR(1)] is defined as follows:

$$\dot{x}_n = r W_n \dot{x}_n + T_n z_n \,, \qquad z_n \sim N(0, I_{J_n}) \,, \qquad [2\text{-}27]$$

where W_n is a $(J \times J)$ matrix of weights $w_{i\,j,n}$ describing the influence of each $\dot{x}_{jn}$ error upon the others, r is an unknown parameter, and $T_n z_n$ allows for heteroscedastic disturbances, where T_n is $(J_n \times J_n)$ diagonal (the subscript n is included to allow for different sized choice sets). Using a general notation, we write $w_{i\,j,n}$ as:

$$w_{i\,j,n} = \frac{w^*_{i\,j,n}}{\sum_{k=1}^{J_n} w^*_{ik,n}} \,, \qquad \forall \;\; j \neq i \;\; and \;\; w_{i\,j,n} = 0 \;\; \forall \; i{=}j \,, \qquad [2\text{-}28]$$

where $w^*_{i\,j,n}$ is a function of unknown parameters and observable explanatory variables, which describe the correlation structure in effect. Solving for $\dot{x}_n$ in Equation [2-27] and incorporating it into Equation [2-4] , leads to a logit kernel form of the GAR[1] specification:

$$U_n = X_n b + F_n T_n z_n + n_n \,, \qquad \text{where } F_n = (I - r W_n)^{-1} \,.$$

The normalization applied in Equation [2-28] ensures that the process is stable for values of r in the $(-1,1)$ interval. The interpretation and the sign of r , usually referred to as the correlation coefficient, depend on the definition of proximity embodied in w^*_{ij} .

In practice, the parameters in $w^*_{i\,j,n}$ could be estimated. However, there are important special cases in which they are fixed. For example, spatial studies often use spatial autoregressive of order 1 [SAR(1)] error processes, which define the contiguity structure through a Boolean contiguity matrix. In this case, $w^*_{ij} = 1$ if i and j are contiguous and $w^*_{ij} = 0$ otherwise. For this specification, a $r > 0$ implies that errors of the same sign are grouped together. A slightly more complex specification, which requires

[11] $\dot{x}_n$ has a slightly different interpretation than the x_n used elsewhere in the paper.

estimation of a single parameter q , is to set $w_{ij}^{*} = (d_{ij})^{-q}$, in which the distance d_{ij} plays the role of a contiguity or proximity measure between pairs of alternatives. For examples of SAR(1) see Anselin (1989), and Cliff and Ord (1981). For an application of SAR(1) processes in economics, see Case (1991). Bolduc, Fortin, and Fournier (1996) use an SAR(1) process to estimate a logit kernel model with 18 alternatives.

For more details on GAR(1), including a discussion on identification issues, see Bolduc (1992).

Random Parameters

The MNL formulation with normally distributed random taste parameters can be written as:

$$U_n = X_n b_n + n_n \ , \text{ where } b_n \sim N(b, \Sigma_b) .$$

b_n is a K-dimensional random normal vector with mean vector b and covariance matrix Σ_b . Replacing b_n with the equivalent relationship: $b_n = b + Tz_n$, where T is the lower triangular Cholesky matrix such that $TT' = \Sigma_b$, leads to a general factor analytic logit kernel specification where $F_n = X_n$:

$$U_n = X_n b + X_n T z_n + n_n \ .$$

The parameters that need to be estimated in this model are b and those present in T . T is usually specified as diagonal, but it does not have to be (see, for example, Train, 1998, and the application presented in Chapter 4). Independently distributed parameters are probably a questionable assumption when variables are closely related, for example in-vehicle and out-of-vehicle travel time.[12] Also, note that the distribution does not have to be normal. For example, parameters with sign constraints should be specified with a lognormal distribution. See the telephone case study presented later for an example of a model with a lognormally distributed b_n parameter.

Identification

For identification of random parameter models, it is useful to separate the random parameters into two groups: those that are applied to alternative-specific constants and those applied to variables that vary across the sample.

<u>*Alternative-specific constants*</u>

> When alternative-specific zero/one dummy variables have randomly distributed parameters, this is identical to the heteroscedastic, nested, and error component structures. In such cases, the order and rank conditions as discussed earlier hold.

<u>*Variables that include variation across the sample*</u>

---

[12] Note that if a subset of the covariances are estimated, then one has to be careful about the way the structural zeros are imposed on the Cholesky. In order for the structure of the Cholesky *T* (i.e., the location of the structural zeros) to be transferred to the covariance structure *TT'*, the structural zeros must be in the left-most cells of each row in the Cholesky. See Appendix B for more discussion.

As pointed out in the general discussion on identification, the order condition does not hold for the portion of the covariance matrix that varies across the sample. Rather, as many parameters as the data will support (without running into multicollinearity problems) can be estimated.

Continuous Attributes of the Alternatives

When random parameters are specified for continuous attributes of the alternatives, there are no identification issues per se. Data willing, the full covariance structure (i.e., variances for each parameter as well as covariances across parameters) can be estimated.

Categorical Attributes of the Alternatives

An interesting and unintuitive identification issue arises when categorical variables[13] are specified with *independently* distributed random parameters. Say there are M categories for a variable. Then there is theoretically a b_m and s_m for each category m, $m = 1,...,M$. It is well known that for the systematic terms (the b_m's), only $(M-1)$ b_m's can be identified and therefore a base must be arbitrarily selected. However, this is not necessarily true for the disturbance terms. To do the analysis, the rank condition comes into play. Identification of the s_m's can be thought of as identification for a nested structure (think of it as examining the covariance structure for a particular individual). Therefore, if there are only 2 categories, then only one random parameter is identified and the normalization is arbitrary; if there are 3 or more categories, then a random parameter for each of the categories is identified. The key here being that, unlike the systematic portion of the utility function, it is incorrect to set one of the s_m's as a base when there are 3 or more categories. Unlike the identification analysis for a nested structure, the number of alternatives J does not impact the number of s_m's that can be estimated, because of the variation across observations. Note that this analysis applies for a single categorical variable, and it is not immediately apparent that the conclusion translates to the case when random parameters are specified for multiple categorical variables in the model. The issue of identification for categorical variables is not addressed in the literature, see, for example, Goett, Hudson, and Train (2000), who include random parameters on several categorical variables in their empirical results.

When covariances are estimated (as they probably should be), then a full set of variances and covariances can be estimated for the $M-1$ b_m's estimated in the systematic utility.

Characteristics of the Decision-maker

If a random parameter is placed on a variable that is a characteristic of the decision-maker (for example, years employed), it necessarily must be interacted with an alternative-specific variable (otherwise it will cancel out when the differences are taken). The normalization or such parameters then depends on the type of variable with which it interacts. If it interacts with alternative-specific dummy variables, then the heteroscedastic rules apply (i.e., $J-1$ variance terms can be estimated,

[13] An example of a categorical variable in a housing choice context is *X*={street parking only, reserved parking space in a lot, private garage}, where each alternative has exactly one of the possible *X*'s associated with it.

and the minimum variance term must be constrained to zero). If it interacts with nest-specific constants, then the rules for nested error structures apply, etc. Furthermore, we suspect that if the characteristic is a categorical variable (for example, low income, medium income, high income), then the rules we presented for categorical attributes also apply (although this hasn't been verified).

Identification of Lognormally Distributed Parameters

Our application of the Order and Rank conditions for identification assume that the disturbance component of the utility can be separated from the systematic portion of the utility. With lognormally distributed parameters, the mean and variance of the distribution are a function of both of the disturbance parameters and therefore this separability does not exist. While the identification rules described above cannot be strictly applied, they provide guidelines for identification. And, as always, empirical tests such as examining the Hessian should also be applied.

As long as the identification restrictions described above are imposed, the number of random parameters that can be identified is dependent on the data itself in terms of the variation and the collinearity present in the explanatory variables. Therefore, empirical methods are used to verify identification of random parameter models, for example, verifying that the Hessian is non-singular at the convergence point. An issue with simulation is that identification issues often do not present themselves empirically unless a large number of draws are used. Therefore, other useful methods are to constrain one or more parameters and observe whether the likelihood changes, or to test the impact of different starting values. Also, it is particularly important in random parameter models to verify stability of parameter estimates as the number of draws increases.

McFadden and Train (2000) note the inherent difficulty of identifying the factor structure for random parameter models, because many different factor combinations will fit the data approximately as well.

Parameter Estimation

We now describe the method that we use to estimate the joint vector of parameters $d = (b', y')'$, where b is the vector of unknown parameters in the systematic portion of the utility and y is the vector of unknown parameters in the error structure. For example, in the heteroscedastic model, only the alternative-specific standard deviations are included in y. In the GAR(1) version based on a Boolean contiguity matrix, the same standard deviations are estimated in addition to r (the correlation coefficient). The factor analytic and the random parameter structures can potentially have a very large number of unknown parameters.

The approach is to employ probability simulators within a maximum likelihood framework, which leads to Maximum Simulated Likelihood (MSL). The application of this method is straightforward and provides great flexibility in terms of the structure of the covariance matrix.

Maximum Likelihood

The log-likelihood of the sample is:

$$L(d) = \sum_{n=1}^{N} \ln P(i_n \mid d) \ ,$$

where $P(i_n \mid d)$ is the probability associated with the choice made by individual n. The score vector is:

$$\frac{\partial L(d)}{\partial d} = \sum_{n=1}^{N} \frac{1}{P(i_n \mid d)} \frac{\partial P(i_n \mid d)}{\partial d} .$$

Inserting the probability equations for the logit kernel model (Equations [2-6] and [2-7]) leads to the score for the logit kernel model:

$$\frac{\partial L(d)}{\partial d} = \sum_{n=1}^{N} \frac{1}{P(i_n \mid d)} \int_z \Lambda(i_n \mid d, z) \frac{\partial \ln \Lambda(i_n \mid d, z)}{\partial d} n(z, I_M) dz \ . \qquad [2\text{-}29]$$

Note that we also use the relationship $\partial X / \partial q = X \left(\partial \ln(X) / \partial q \right)$ in Equation [2-29] in order to make the derivative tractable: $\ln \Lambda(i_n \mid d, C_n) = X_{in} b + F_{in} T z_n - \ln \sum_{j \in C_n} e^{X_{jn} b + F_{jn} T z_n}$, which is easy to differentiate.

Each factor z introduces a dimension to the integral. Unless the dimension of z is small (≤ 3), the Maximum Likelihood (ML) estimator just described cannot be computed in a reasonable amount of time. For models with z of larger dimension, we use the Maximum Simulated Likelihood (MSL) methodology, described next.

Maximum Simulated Likelihood

The response probability for alternative i is replaced with the unbiased, smooth, tractable simulator:

$$\hat{P}(i \mid d) = \frac{1}{\mathbb{D}} \sum_{d=1}^{\mathbb{D}} \Lambda(i \mid d, z_n^d) \ , \qquad [2\text{-}30]$$

where z_n^d denotes draw d from the distribution of z_n (each draw consists of M elements). Thus, the integral is replaced with an average of values of the function computed at discrete points. There has been a lot of research concerning how best to generate the set of discrete points (see Bhat, 2000, for a summary and references). The most straightforward approach is to use pseudo-random sequences. However, variance reduction techniques (for example, antithetic draws) and quasi-random approaches (for example, the Halton draws, which are used in the empirical results in this chapter) have been found to cover the dimension space more evenly and thus are more efficient.

Incorporating the simulated probability, the simulated log-likelihood is then:

$$\hat{L}(d) = \sum_{n=1}^{N} \ln \hat{P}(i_n \mid d) \ , \qquad [2\text{-}31]$$

and the simulated score is:

$$\frac{\partial \hat{L}(d)}{\partial d} = \sum_{n=1}^{N} \frac{1}{\hat{P}(i_n \mid d)} \frac{1}{\mathbb{D}} \sum_{d=1}^{\mathbb{D}} \Lambda(i_n \mid d, z_n^d) \frac{\partial \ln \Lambda(i_n \mid d, z_n^d)}{\partial d} . \qquad [2\text{-}32]$$

A well-known result previously obtained in Börsch-Supan and Hajivassiliou (1993), among others, indicates that the log-likelihood function, although consistent, is simulated with a downward bias for finite number of draws. The issue is that while the probability simulator [2-30] is unbiased, the log-simulated-likelihood [2-31] is biased due to the log transformation. This can be seen by Jensen's inequality and the concavity of the log function. It can also be seen by taking a second degree Taylor's expansion of $\ln(\hat{P}(i))$ around $P(i)$, which gives:

$$\ln(\hat{P}(i)) \approx \ln(P(i)) + \frac{1}{P(i)}(\hat{P}(i) - P(i)) - \frac{1}{2P(i)^2}(\hat{P}(i) - P(i))^2 .$$

Taking the expected value of this relationship implies that:

$$\hat{L}(d) - L(d) \approx -\frac{\mathrm{var}(\hat{P}(i \mid d))}{2P(i \mid d)^2} \leq 0 . \qquad [2\text{-}33]$$

This suggests that in order to minimize the bias in simulating the log-likelihood function, it is important to simulate the probabilities with good precision. The precision increases with the number of draws, as well as with the use of efficient methods to generate the draws. The number of draws necessary to sufficiently remove the bias cannot be determined a priori; it depends on the type of draws, the model specification, and the data.

Applications

In this section, we consider four applications: two based on synthetic data and two on real data. The first sample concerns a hypothetical choice situation among three alternatives; the focus is on the parameter identification issues of heteroscedastic models. The second sample, also using synthetic data, has 5 alternatives and focuses on identification issues of categorical variables with random parameter. The third application uses a mode choice dataset that is used for logit kernel models that appear in two recent textbooks (Greene, 2000, and Louviere, Hensher, and Swait, 2000). We replicate the models presented in the texts, and use them to highlight practical issues that arise in estimating logit kernel models. The fourth application is based on a survey collected to predict residential telephone demand. We estimate several error structures for the telephone data, including heteroscedasticity, nesting, cross-nesting, and random parameter, and highlight many of the important identification and estimation issues of logit kernel models.

Estimation Notes & Practical Issues

Optimization Algorithm

While the likelihood function for linear in the parameters logit models is strictly concave, this is not true for logit kernel models (note that it is also not true for the nested logit model). Furthermore, the simple Newton methods that are used for MNL estimation tend to lose their robustness when the optimization function is not concave. Therefore, modified Newton methods, which address non-concavity with techniques such as trust regions, should be used for logit kernel models. For details on these methods, see Dennis and Schnabel (1983). In the applications presented in this chapter, we use the `DUMIAH` routine provided in Fortran's IMSL Libraries. The `maxlik` routine provided in Gauss could also be used.[14]

Direction Matrix

To decrease estimation time, we analytically program the derivatives and approximate the matrix of second derivatives (the Hessian) with first order information. The most straightforward approximation of the Hessian is the BHHH technique (Berndt et al. 1974), which is computed as:

$$\mathbf{R} = \sum_{n=1}^{N} \left(\frac{\partial L_n(d)}{\partial d} \right) \left(\frac{\partial L_n(d)}{\partial d} \right)', \qquad [2\text{-}34]$$

where the score is defined as in Equation [2-29] (evaluated per sample observation). For Maximum Simulated Likelihood, it is computed with the simulated scores [2-32].

Under certain regularity conditions, BHHH can be shown to be a consistent estimator of the covariance matrix of parameters at the maximum likelihood estimate. There are also numerous other approximations that can be used, see Dennis and Schnabel (1983) for further discussion.

Standard Errors at Convergence

For a finite number of simulation draws, BHHH may substantially underestimate the covariance of the estimator due to simulation error (see McFadden and Train, 2000, for a discussion). BHHH (or some other approximation) is still preferred for the direction matrix due to the low cost of estimating the matrix as well as the robustness of estimation with regards to the direction matrix. However, it is advisable to use robust standard errors to generate the test statistics at convergence. A robust asymptotic covariance matrix estimator is $\mathbf{H}^{-1}\mathbf{R}\mathbf{H}^{-1}$ (Newey and McFadden, 1994), where $\mathbf{H}$ is the Hessian, calculated numerically or analytically, and $\mathbf{R}$ is defined as in Equation [2-34]. When simulation is used, the simulated Hessian and Score are used. We report robust t-statistics (calculated using a numerical Hessian) for all estimation results.

---

[14] Note that Kenneth Train of UC Berkeley provides Gauss-based estimation code for logit kernel (a.k.a. mixed logit) models from his website: http://emlab.berkeley.edu/users/train/index.html

Simulation Draws

We primarily use Halton draws for the simulation; however, some of the specifications are also estimated using pseudo-random draws for comparison. (See Bhat, 2000, and Train, 1999, for more information on Halton draws.) We have found the Halton draws to be more efficient than pseudo-random draws. For each observation, we draw $\mathbb{D}$ random vectors ($z_n^1, \ldots, z_n^{\mathbb{D}}$, each $(M \times 1)$) from the given multivariate distribution of the factors, and these draws are kept constant across iterations so that the simulator does not "chatter" as d changes (see McFadden and Train, 2000, for more information). The probability is then simulated using Equation [2-30], the log-likelihood using Equation [2-31], and the derivatives using Equation [2-32].

Simulation Bias and Identification

Two issues critical to estimating logit kernel models are simulation bias and identification.

As noted above, the number of draws, $\mathbb{D}$, must be large enough to sufficiently reduce the bias shown in Equation [2-33]. The problem is that there is no way to know a priori how large is large enough, because this depends on the particular model structure and data. Therefore it is always necessary, as we do in these applications, to verify that the estimated parameters remain stable as the number of draws is increased.

The number of draws also plays an important role in testing for identification. Note that there are two forms of unidentification: structural, as indicated by the order and rank conditions, and informational, which is when the data do not provide enough information to support the given structure (i.e., multicollinearity). It turns out that identification problems often do not appear (via a singular Hessian) when a small number of draws is used. For example, in the most extreme case, any specification (whether identified or not) will always appear identified when only 1 draw is used, because this is equivalent to adding explanatory variables to the systematic portion of the utility. This issue also emphasizes the importance of checking the rank condition prior to estimation, and of verifying robustness of estimates using different starting values.

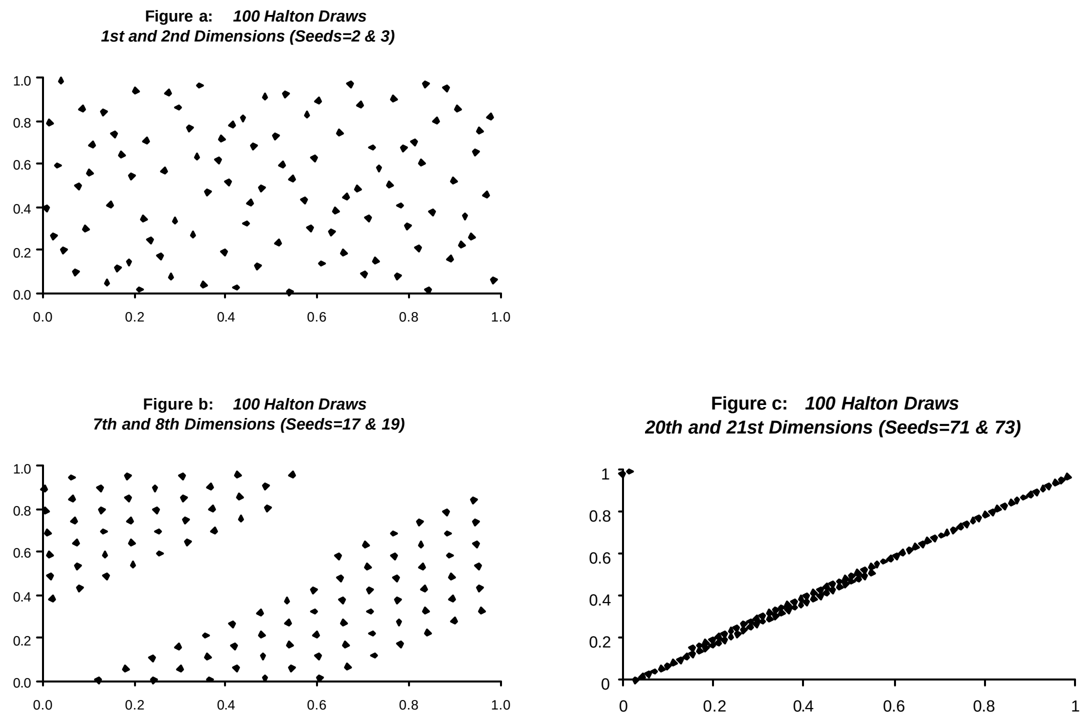


Figure 2-1: 100 Halton Draws for Different Dimensions of the Integral

Another issue with the number of draws is that as the dimension of the problem increases the number of draws necessary to estimate the model also increases. Conceptually, the issue is that it takes more draws to adequately cover the dimension space; this applies to all methods used to integrate non-closed form functions (for example, Gaussian quadrature or simulation via pseudo-random or quasi-random methods). It is interesting to note that with Halton draws, planes develop when small numbers of draws are used for high dimensional integrals. The generation of Halton draws is presented very clearly in Train (1999). Briefly, to implement Halton draws, a non-random series is developed for each dimension, each series is seeded with a prime number, and the seeds are implemented in order (2, 3, 5, 7, etc.). As an example of the problem with planes developing, take an extreme case: 100 draws are often sufficient to estimate a two dimensional model. As shown in Figure 2-1a, examination of a sample of Halton draws for a particular observation shows that the draws cover the 1st and 2nd dimensions of the sample space quite well. However, Figure 2-1b indicates that 100 draws for the 7th and 8th dimensions do not cover the space well, and Figure 2-1c shows that the 100 draws for the 20th and 21st dimensions are even worse.

To summarize, due to the issues of bias and identification, it is critical to empirically verify on a case-by-case basis that a sufficient number of draws are being used to estimate the model.

Synthetic Data I: Heteroscedasticity

The first application concerns a hypothetical choice situation among three alternatives. The model specification is as follows.

$$
\begin{aligned}
U_{1n} &= a_1 + X_{1n} b + s_1 z_{1n} + n_{1n} , \\
U_{2n} &= a_2 + X_{2n} b + s_2 z_{2n} + n_{2n} , \\
U_{3n} &= X_{3n} b + s_3 z_{3n} + n_{3n} .
\end{aligned}
$$

The true parameter values used to generate the synthetic data are:

$$a_1 = 1.5, \quad a_2 = 0.5, \quad b = -1, \quad s_1 = 3, \quad s_2 = 2, \quad s_3 = 1, \; and \; m = 1.$$

The explanatory variable, X , is simulated as a normal variable with a standard deviation of 3, independent across alternatives and observations. The utilities for each observation are generated by drawing a single random draw for each z_{jn} from independent standard normal distributions and each n_{jn} from independent standard Gumbel distributions. The utilities are calculated, and the alternative with the highest utility is then the chosen alternative.

Estimation results using the synthetic data are provided in Table 2-1. Table 2-1a presents estimation results regarding selecting and setting the base heteroscedastic term. Recall that only $J-1$ heteroscedastic terms are identified, and that it is necessary to either set the minimum variance term to zero, or set any of the other variance terms high enough according to the equation derived earlier (Equation [2-23]):

$$\dot{s}_{ff}^{N} \geq \left(\dot{s}_{jj} - \dot{s}_{ii} \right) g \big/ (g + \dot{s}_{ii}) \quad , \; i = 1,\ldots, J \;,$$

where $\dot{s}_{jj}$ is the theoretical (true) variance that is fixed to the value $\dot{s}_{ff}^{N}$.

All of the models in Table 2-1a are estimated with 10,000 observations and 500 Halton draws. The first model shows estimation results for an unidentified model; this model is used to determine the minimum variance alternative, and it correctly identifies the third alternative as having minimum variance.[15] Models 2 through 4 show identified models in which the minimum variance alternative is constrained to different values (0, 1, and 2); as expected, the log-likelihoods of these models are basically equivalent and all of these represent correct specifications. Models 5 through 10 show identified models in which the maximum variance alternative is constrained to different values (0, 1, 1.5, 2.25, 3, and 4). Applying Equation [2-23] (repeated above), the model specification will be correct as long as s_1 is constrained to a value above 2.2. The empirical results verify this. First, there is a severe loss of fit when the s_1 is constrained below 2.2. Second, the parameter estimates for the mis-specified models are biased. This can be seen by examining the ratio of the systematic parameters (for example, b / a_1) across models. While the scale shifts for various normalizations (and therefore the parameter estimates also shift), the ratio of systematic parameters should remain constant across normalizations. A cursory examination of the estimation results shows that these ratios begin to drift with successively invalid normalizations. Finally, note that these results indicate a slight loss of fit when the base alternative is constrained to a high value (s_3 =2 and

[15] We were able to calculate t-statistics for the unidentified model here (and elsewhere) for two reasons. First, simulation has the tendency to mask identification issues, and therefore does not always result in a singular Hessian for a finite number of draws. Second, the slight difference between the Gumbel and Normal distributions makes the unidentified model only 'nearly' singular, and not perfectly singular.

s_1 =4), and this is due to the issue addressed earlier regarding the slight difference between the Gumbel and normal distributions. It must be emphasized that the normalization in heteroscedastic logit kernel models is not arbitrary.

Table 2-1: Synthetic Data I - Heteroscedastic Models (3 Alternatives)

Table a: Selecting and Setting the Base Heteroscedastic Term (10,000 Observations & 500 Halton Draws)

	True	Unidentified		Identified: Minimum Variance Base						Identified: Maximum Variance Base											
Parameter	Value	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
α_1	1.5	1.27	*(3.4)*	1.24	*(15.7)*	1.51	*(15.9)*	2.18	*(15.9)*	0.97	*(29.1)*	1.02	*(27.9)*	1.08	*(23.4)*	1.24	*(5.8)*	1.57	*(17.2)*	2.03	*(17.4)*
α_2	0.5	0.43	*(2.6)*	0.42	*(8.9)*	0.53	*(9.2)*	0.76	*(9.2)*	0.37	*(11.1)*	0.40	*(11.5)*	0.41	*(10.4)*	0.42	*(2.2)*	0.54	*(6.8)*	0.70	*(7.0)*
β	-1.0	-0.80	*(3.8)*	-0.78	*(14.6)*	-0.94	*(14.1)*	-1.36	*(13.7)*	-0.51	*(55.5)*	-0.57	*(65.0)*	-0.64	*(39.1)*	-0.78	*(16.0)*	-0.98	*(37.1)*	-1.27	*(37.1)*
σ_1	3.0	2.32	*(2.9)*	2.24	*(9.7)*	2.84	*(10.3)*	4.30	*(11.0)*	*0.00*	---	*1.00*	---	*1.50*	---	*2.25*	–	*3.00*	---	*4.00*	---
σ_2	2.0	1.27	*(1.9)*	1.21	*(4.7)*	1.69	*(5.9)*	2.80	*(7.7)*	0.06	*(0.1)*	0.03	*(0.3)*	0.50	*(1.8)*	1.22	*(6.6)*	1.82	*(11.7)*	2.58	*(14.5)*
σ_3	1.0	0.35	*(0.2)*	0.00	–	1.00	–	2.00	---	0.00	*(0.9)*	0.00	*(1.6)*	0.01	*-(0.5)*	0.16	*(0.0)*	1.07	*(4.4)*	1.78	*(7.6)*
(Simul.) Log-Likelihood:		-6837		-6837		-6837		-6838		*-6907*		*-6865*		-6845		-6837		-6837		-6838	
	Model:	1		2		3		4		5		6		7		8		9		10	

Table b: Varying the Numbers and Types of Draws (10,000 Observations)

			Halton Draws								Pseudo-Random Draws							
	True	True with	200 Halton		1000 Halton		2000 Halton		4000 Halton		500 'Random'		1000 'Random'		5000 'Random'		10000 'Random'	
Parameter	Value	σ_3=0	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
α_1	1.5	1.18	1.22	*(16.5)*	1.24	*(15.4)*	1.24	*(15.5)*	1.24	*(14.5)*	1.20	*(16.5)*	1.21	*(16.2)*	1.23	*(15.6)*	1.24	*(15.7)*
α_2	0.5	0.39	0.42	*(9.1)*	0.42	*(8.8)*	0.42	*(8.8)*	0.42	*(8.9)*	0.42	*(9.3)*	0.42	*(9.1)*	0.42	*(8.9)*	0.42	*(8.8)*
β	-1.0	-0.79	-0.77	*(15.6)*	-0.78	*(14.2)*	-0.78	*(14.3)*	-0.78	*(13.0)*	-0.75	*(15.6)*	-0.76	*(15.3)*	-0.78	*(14.4)*	-0.78	*(14.6)*
σ_1	3.0	2.23	2.19	*(10.2)*	2.25	*(9.5)*	2.26	*(9.5)*	2.25	*(8.7)*	2.14	*(10.2)*	2.15	*(10.0)*	2.23	*(9.5)*	2.26	*(9.7)*
σ_2	2.0	1.37	1.14	*(4.6)*	1.22	*(4.5)*	1.23	*(4.6)*	1.23	*(4.2)*	1.06	*(4.0)*	1.10	*(4.2)*	1.19	*(4.4)*	1.22	*(4.7)*
σ_3	1.0	*0.00*	*0.00*	---	*0.00*	---	*0.00*	---	*0.00*	---	*0.00*	---	*0.00*	---	*0.00*	–	*0.00*	---
(Simul.) Log-Likelihood:			-6837		-6837		-6837		-6836		-6835		-6839		-6838		-6836	

Table c: Varying the Number of Observations (500 Halton Draws)

	True	1000 Obs		5000 Obs		10000 Obs		40000 Obs		80000 Obs	
Parameter	Value	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
α_1	1.5	2.27	*(2.1)*	1.64	*(9.6)*	1.51	*(15.9)*	1.45	*(32.1)*	1.54	*(38.4)*
α_2	0.5	0.91	*(2.4)*	0.68	*(8.4)*	0.53	*(9.2)*	0.53	*(18.4)*	0.52	*(23.7)*
β	-1.0	-1.69	*(1.9)*	-0.99	*(8.3)*	-0.94	*(14.1)*	-0.95	*(29.2)*	-1.02	*(33.2)*
σ_1	3.0	5.64	*(1.7)*	3.13	*(6.5)*	2.84	*(10.3)*	2.85	*(21.3)*	3.05	*(24.8)*
σ_2	2.0	3.58	*(1.5)*	1.62	*(3.2)*	1.69	*(5.9)*	1.72	*(12.3)*	2.08	*(17.4)*
σ_3	1.0	*1.00*	---	*1.00*	---	*1.00*	---	*1.00*	---	*1.00*	---
(Simul.) Log-Likelihood:		-655		-3369		-6837		-27499		-54944	

The models shown in Table 2-1b were estimated to investigate the impact of the number and types of draws. All of these models are estimating using the normalization s_3 =0, and so we report the true parameters as calculated given this normalization (using Equations [2-15] to [2-17]). The model estimates verify that the 500 Halton draws used for the models in Table 2-1a are sufficient. The results also show that the Halton draws are more efficient then pseudo-random draws, as the parameter estimates stabilize for a lower number of Halton draws. Table 2-1c is provided to show that as the number of observations increases, the estimated parameters converge on their true values. Note that a potentially large number of observations is required to accurately reproduce the parameters of the population. However, the required number of observations is highly dependent on the model specification and data, and generalizations cannot be drawn.

Synthetic Data II: Random parameters on Categorical Variables

The second application, which also involves synthetic data, concerns the issue of identification of random parameters for categorical variables. Recall that if the variable has two categories (i.e., a 0/1 dummy) then one systematic parameter and one random parameter are identified, and the normalization of each is arbitrary. For variables with 3 (or more) categories, two systematic parameters are identified but all 3 random parameters (one per category) are identified. Empirical results are shown in Table 2-2. Table 2-2a, b, and c all use slightly different datasets and model specifications. The general specification is as follows:

$$U_{in} = a_i + \begin{bmatrix} X_{1in} & X_{2in} & X_{3in} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} X_{1in} & X_{2in} & X_{3in} \end{bmatrix} \begin{bmatrix} s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & s_3 \end{bmatrix} \begin{bmatrix} z_{1n} \\ z_{2n} \\ z_{3n} \end{bmatrix} + n_{in}$$

$$, \forall\ i = 1,\ldots,5\ ;\ n,$$

where $a_5 = 0$ (the base alternative-specific constant) and X is a categorical variable, that is $X_{kin} = \{0,1\}$ & $X_{1in} + X_{2in} + X_{3in} = 1,\ \forall\ i;\ k = 1,\ldots,3; n$. The data are generated using the same approach as described in the synthetic data above, i.e., a X, z, $and\, n$ are sampled for each person, the utilities are calculated according to the model and parameters above, and the alternative with the highest utility is the chosen alternative. 10,000 observations are used for all of the models.

The dataset for the models in 2a includes a categorical variable with 2 categories ($X_{3in} = 0\ \forall\ i, n$). While the covariance structure varies across individuals, identification is analogous to a nested structure with two nests, for example, 1, 1, 2, 2, 2 or 1, 2, 2, 2, 2 or 1, 2, 1, 2, 1, etc. depending on the values of X for observation n.[16] Therefore, 1 systematic parameter (b) and 1 random parameter (s) can be estimated. Furthermore, the normalization of the random parameter is arbitrary. These statements are supported by the estimation results. The first two models show that the model with

[16] This concept of a categorical variable being analogous to a 2-nest nesting structure is denoted as *"~1, 1, 2, 2, 2"* in Table 2-2.

Table 2-2: Synthetic Data II – Categorical Variables with Random Parameters (5 Alternatives; 10,000 Observations)

Table a: Categorical variables with 2 categories, each enters all 5 utilities (~1, 1, 2, 2, 2)

Parameter	True Value	Unidentified 500 Halton Est	t-stat	Unidentified 500 Halton Est	t-stat	Identified: Base 1 500 Halton Est	t-stat	Identified: Base 2 500 Halton Est	t-stat	Identified: Base 2 1000 Halton Est	t-stat
α_1	0.5	0.48	*(11.2)*	0.48	*(11.2)*	0.48	*(11.2)*	0.48	*(11.2)*	0.48	*(11.2)*
α_2	0.5	0.44	*(10.2)*	0.44	*(10.2)*	0.44	*(10.2)*	0.44	*(10.2)*	0.44	*(10.2)*
α_3	1.0	0.92	*(22.7)*	0.92	*(22.7)*	0.92	*(22.7)*	0.92	*(22.7)*	0.92	*(22.7)*
α_4	1.0	0.98	*(24.2)*	0.98	*(24.2)*	0.98	*(24.2)*	0.98	*(24.2)*	0.98	*(24.2)*
β_1	0.5	0.50	*(7.9)*	0.50	*(7.9)*	0.50	*(7.9)*	0.50	*(7.9)*	0.50	*(7.9)*
σ_1	2.0	0.84	*(2.3)*	3.91	*(13.9)*			3.94	*(14.4)*	3.94	*(14.4)*
σ_2	4.0	3.85	*(13.6)*	0.47	*(0.7)*	3.94	*(14.4)*				
$(\sigma_1^2+\sigma_2^2)^{1/2}$	4.5	3.94		3.94		3.94		3.94		3.94	
(Simul.) Log-Likelihood:		-15310		-15310		-15310		-15310		-15310	
	Model:	*1*		*2*		*3*		*4*		*5*	

Table b: Categorical variables with 2 categories, each enters 4 of 5 utilities (~1, 1, 2, 2, 0)

Parameter	True Value	Misspecified 1 500 Halton Est	t-stat	Misspecified 2 500 Halton Est	t-stat	Identified 500 Halton Est	t-stat	Identified 1000 Halton Est	t-stat
α_1	0.5	0.10	*(1.5)*	0.41	*(9.6)*	0.47	*(5.1)*	0.47	*(5.1)*
α_2	0.5	0.04	*(0.6)*	0.35	*(8.2)*	0.41	*(4.4)*	0.41	*(4.5)*
α_3	1.0	0.52	*(7.8)*	0.80	*(19.5)*	0.90	*(9.7)*	0.90	*(9.8)*
α_4	1.0	0.57	*(8.7)*	0.86	*(21.0)*	0.95	*(10.3)*	0.96	*(10.4)*
β_1	0.5	0.53	*(8.7)*	0.11	*(2.8)*	0.50	*(7.3)*	0.50	*(7.3)*
σ_1	2.0			2.29	*(16.0)*	1.73	*(8.4)*	1.73	*(8.5)*
σ_2	4.0	3.45	*(15.1)*			3.55	*(13.2)*	3.55	*(13.2)*
(Simul.) Log-Likelihood:		-15398		-15537		-15378		-15378	

Table c: Categorical variables with 3 categories, each enters all utilities (~1, 1, 2, 2, 3)

Parameter	True Value	Misspecified 500 Halton Est	t-stat	Identified 500 Halton Est	t-stat	Identified 1000 Halton Est	t-stat
α_1	0.5	0.36	*(7.7)*	0.36	*(7.7)*	0.36	*(7.7)*
α_2	0.5	0.40	*(8.5)*	0.40	*(8.5)*	0.40	*(8.5)*
α_3	1.0	0.93	*(20.5)*	0.93	*(20.6)*	0.93	*(20.6)*
α_4	1.0	0.92	*(20.2)*	0.92	*(20.3)*	0.92	*(20.3)*
β_1	1.0	1.06	*(6.4)*	1.06	*(6.4)*	1.06	*(6.7)*
β_2	0.5	1.06	*(7.0)*	0.69	*(4.4)*	0.70	*(4.4)*
σ_1	2.0	3.47	*(12.2)*	2.75	*(7.5)*	2.77	*(8.1)*
σ_2	3.0			2.52	*(6.8)*	2.49	*(6.7)*
σ_3	4.0	4.74	*(11.1)*	4.37	*(10.7)*	4.38	*(10.9)*
(Simul.) Log-Likelihood:		-15376		-15368		-15368	

both random parameters is unidentified, as the fit is identical for very different estimates of the random parameters. The third and fourth models show that the normalization is arbitrary: the parameter and fit are the same for either normalization. The fifth model verifies that enough draws are being used for estimation.

The dataset used for the models in Table 2-2b is similar to that used in Table 2-2a, with the exception that the categorical variable only applies to the first four alternatives ($X_{k5n} = 0 \; \forall \; k, n$). In this case, identification is related to a nested structure with three nests (for example, 1, 1, 2, 2, 0); therefore, 1 systematic parameter is estimable and *both* of the random parameters are estimable. This is shown in the estimation results, where the models with either of the systematic terms fixed to 0 results in a significant loss of fit.

In Table 2-2c, the categorical variable contains three categories. Identification here is also related to a nested model with 3 nests (for example, 1, 1, 2, 2, 3), and therefore 2 systematic parameters are identified and all 3 random parameters are identified. This is supported by the estimation results, in which constraining one of the random terms to zero results in a significant loss of fit.

Empirical Application I: Mode Choice

The logit kernel formulation is now making its way into econometric textbooks. In this section, we investigate the identification issues of logit kernel models that appear in Greene (2000, Table 19.15) and Louviere, Hensher and Swait (2000, Table B6.5). Both texts make use of the same data and present similar model specifications.

The Data

This is a revealed choice dataset containing mode choices for travel between Sydney and Melbourne, Australia. The choices available are air, train, bus, and car.[17] There are 210 observations in the sample, and the explanatory variables are[18]:

GCost: Generalized cost ($00)
= in vehicle cost + in vehicle time*value of travel time savings.

TTime: Terminal waiting time for plane, train and bus (hours). Auto terminal time is zero.

Income: Household income ($00,000), which is interacted with the 'air' alternative specific dummy variable.

---

[17] The dataset is actually a choice-based sample, and therefore the weighted exogenous sample maximum likelihood estimator (WESML, see Ben-Akiva and Lerman, 1985) should be used for the logit-based models (and the probit-equivalent for the probit models, see Imbens, 1992) to obtain consistent estimates. However, we did not use WESML in order to replicate the models as reported in the textbooks.

[18] Note: *(i)* The Louviere, Swait, and Hensher model also included a 'party size' explanatory variable. We based our models on the more parsimonious specification used in Greene. *(ii)* We scaled the data differently than that used for the models reported in the textbooks.

Models

In this section, we use the models presented in Greene and Louviere et al. to highlight various practical issues in model estimation. Greene estimated a series of models including probit as well as several logit kernel specifications (an unrestricted covariance structure, a heteroscedastic model, and a more general random parameter model). Louviere et al. present an even more general random parameter model.

Table 2-3: Mode Choice Model – Probit

Specification:	Unidentified				Identified			
Draws:	1000 'Random'		1000 'Random'		1000 'Random'		5000 'Random'	
Parameter	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Altern. Specific constants								
Air (1)	0.270	*n/a*	0.968	*n/a*	0.456	*(1.2)*	0.377	*(0.6)*
Train (2)	0.579	*n/a*	2.10	*n/a*	0.959	*(4.8)*	0.917	*(3.5)*
Bus (3)	0.486	*n/a*	1.76	*n/a*	0.805	*(4.4)*	0.768	*(3.1)*
GCost ($00)	-0.468	*n/a*	-1.70	*n/a*	-0.772	*(4.0)*	-0.747	*(4.6)*
TTime (hours)	-0.662	*n/a*	-2.39	*n/a*	-1.10	*(3.8)*	-1.03	*(2.3)*
Income ($00,000) - Air (1)	0.700	*n/a*	2.54	*n/a*	1.15	*(2.0)*	1.16	*(2.5)*
T11	0.608	*n/a*	2.20	*n/a*	1.00	----	1.00	----
T21	0.131	*n/a*	0.476	*n/a*	0.216	*(0.9)*	0.224	*(2.3)*
T31	0.0736	*n/a*	0.267	*n/a*	0.121	*(0.5)*	0.132	*(1.5)*
T22	0.246	*n/a*	0.888	*n/a*	0.407	*(3.0)*	0.381	*(2.9)*
T32	0.113	*n/a*	0.408	*n/a*	0.186	*(1.5)*	0.175	*(2.9)*
T33	0.130	*n/a*	0.471	*n/a*	0.216	*(2.7)*	0.202	*(2.4)*
Log Likelihood (simul.):	-197.727		-197.727		-197.727		-197.784	

Unrestricted Probit

The first model we present is a probit model in which the covariance matrix of utility differences (Ω_Δ) is unrestricted. In this case, the parameters of the Cholesky decomposition of Ω_Δ are estimated, or:

$$T = \begin{bmatrix} T_{11} & 0 & 0 \\ T_{21} & T_{22} & 0 \\ T_{31} & T_{32} & T_{33} \end{bmatrix}, \text{ where } TT' = \Omega_\Delta .$$

Note that even with probit, one has to be careful about identification. The Order Condition states that only five of the six parameters can be estimated. (Greene indirectly estimates all six, and therefore reports results for an unidentified model.) The need for this restriction can be verified empirically, and we present the results in Table 2-3. These were obtained using the GHK simulator with pseudo-random draws. First we report two sets of estimation results for the unidentified model. The two models have identical fits and yet different parameter estimates (note that the difference is a scale shift). The models also have a singular Hessian and therefore t-stats could not be generated. We also report estimation results for the identified model (setting $T_{11} = 1$). The model is now identified: the fit is identical to the unidentified models and the Hessian is not singular. The 5,000 draw result is provided to verify stability.

Unrestricted Logit Kernel

Greene also presents a logit kernel version of the probit model presented Table 2-3 (which he calls a 'constants random parameters logit model'). For the logit kernel version, the disturbance parameters include the six T_{ij} parameters as well as the logit scale parameter m. The identification of this model presents some interesting issues. First, an application of the order condition suggests that the m as well as one of the T_{ij}'s must be normalized for identification. However, as we will show empirically, this is not exactly the case. The reason is due to the slight difference between the Normal and Gumbel distribution. Since there is not an exact trade-off between the probit-like term and the Gumbel, there is an optimal weighting between the two distributions that make up the disturbance, and this allows an extra term to be estimated. Nonetheless, the model is nearly singular without a constraint on a T_{ij}, and so it is advisable to impose a normalization.

The second issue relates to the manner in which T_{ij} is normalized. The covariance matrix of utility differences for this model is:

$$\begin{bmatrix} T_{11}^2 + 2g/m^2 & & \\ T_{11}T_{21} + g/m^2 & T_{21}^2 + T_{22}^2 + 2g/m^2 & \\ T_{11}T_{31} + g/m^2 & T_{21}T_{31} + T_{22}T_{32} + g/m^2 & T_{31}^2 + T_{32}^2 + T_{33}^2 + 2g/m^2 \end{bmatrix}$$

We want to impose a normalization such that the model can reduce to a pure MNL. Therefore we want to normalize some $T_{ij} = 0$. Note that we cannot set $T_{11} = 0$, because this will restrict two of the covariance terms in the probit portion to be zero. We have also found empirical evidence that it is not always valid to set $T_{22} = 0$ due to the positive definiteness condition. However, it appears that the normalization $T_{33} = 0$ (or, more generally normalizing the lowest diagonal element of the cholesky matrix) is a valid normalization, and this is what we apply for this model. (See Appendix A for more information.)

The empirical results for the unrestricted logit kernel model are provided in Table 2-4. The first two columns provide estimation results for the case in which all six T_{ij}'s are estimated. The model is identified as suggested by a non-singular Hessian and stable parameter estimates as the number of draws is increased. The middle columns provide estimation results for models in which T_{33} is normalized to various values. There is marginal loss of fit due to the normalizations, but the likelihood function is fairly flat across the normalizations. The final column is provided to verify the stability of the normalized model with a high number of draws.

Table 2-4: Mode Choice Model – Unrestricted Logit Kernel

Specification:	Multinomial Logit		'Unidentified' (Nearly Singular)				Identified with Various Normalizations								Identified	
Draws:			2000 Halton		40.000 Halton		2000 Halton		2000 Halton		2000 Halton		2000 Halton		4000 Halton	
Parameter	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Altern. Specific constants																
Air (1)	5.21	*(5.3)*	4.42	*(1.4)*	4.41	*(1.5)*	4.42	*(1.4)*	4.76	*(0.8)*	8.28	*(0.3)*	25.3	*(0.9)*	4.41	*(1.4)*
Train (2)	3.87	*(7.5)*	6.09	*(1.2)*	6.02	*(1.5)*	6.09	*(1.4)*	8.28	*(2.5)*	19.0	*(2.6)*	41.4	*(5.5)*	6.09	*(1.1)*
Bus (3)	3.16	*(5.8)*	5.00	*(1.1)*	4.93	*(1.4)*	5.00	*(1.4)*	6.92	*(2.5)*	15.9	*(3.0)*	35.1	*(5.8)*	5.00	*(1.0)*
GCost ($00)	-1.55	*(3.1)*	-4.04	*(0.7)*	-3.97	*(0.8)*	-4.04	*(0.8)*	-6.22	*(1.3)*	-15.4	*(1.5)*	-33.1	*(1.7)*	-4.04	*(0.6)*
TTime (hours)	-5.77	*(6.4)*	-7.50	*(1.8)*	-7.43	*(2.3)*	-7.50	*(2.2)*	-9.73	*(3.5)*	-21.5	*(4.6)*	-48.9	*(5.6)*	-7.50	*(1.7)*
Income ($00,000) - Air (1)	1.33	*(1.4)*	5.55	*(0.5)*	5.44	*(0.6)*	5.55	*(0.6)*	8.91	*(0.8)*	23.5	*(0.5)*	40.5	*(0.7)*	5.55	*(0.5)*
T11			4.85	*(0.6)*	4.76	*(0.7)*	4.85	*(0.7)*	7.78	*(1.0)*	20.3	*(0.8)*	40.8	*(1.5)*	4.86	*(0.5)*
T21			0.934	*(0.4)*	0.904	*(0.5)*	0.933	*(0.5)*	1.59	*(0.9)*	4.35	*(0.6)*	7.83	*(1.1)*	0.928	*(0.4)*
T31			0.554	*(0.4)*	0.538	*(0.5)*	0.554	*(0.5)*	0.913	*(0.7)*	2.50	*(0.6)*	4.30	*(0.5)*	0.551	*(0.4)*
T22			1.25	*(0.3)*	1.18	*(0.3)*	1.25	*(0.3)*	2.81	*(1.2)*	7.79	*(3.5)*	17.9	*(3.1)*	1.25	*(0.2)*
T32			0.711	*(0.3)*	0.681	*(0.4)*	0.711	*(0.4)*	1.30	*(1.4)*	3.44	*(1.4)*	7.55	*(2.2)*	0.709	*(0.3)*
T33			5.12E-03	*(0.1)*	-7.88E-05	*(0.0)*	0.000	----	1.00	----	4.00	----	10.0	----	0.00	----
Log Likelihood (simul.):	-199.128		-195.466		-195.491		-195.466		-196.500		-197.713		-197.647		-195.481	

Heteroscedastic Logit Kernel

Greene also reports a heteroscedastic logit kernel model (which he calls an 'uncorrelated random parameters logit model'). As with the unrestricted logit kernel model discussed above, the rank and order conditions suggest a normalization is necessary when this is not exactly the case. Nonetheless, a normalization is advisable since the model is otherwise nearly singular. Furthermore, as we emphasized earlier, if a normalization is imposed, the selection of the base alternative to normalize is not arbitrary.

Table 2-5: Mode Choice Model – Heteroscedastic Logit Kernel

			Heteroscedastic Models									
Specification:	Multinomial Logit		'Unidentified'		Identified: Base 1		Identified: Base 3		Identified: Base 4		Identified: Base 4	
Draws:			1000 Halton		1000 Halton		1000 Halton		1000 Halton		5000 Halton	
Parameter	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Altern. Specific constants												
Air (1)	5.21	*(5.3)*	4.65	*(3.1)*	5.21	*(6.4)*	4.65	*(3.1)*	4.62	*(3.6)*	4.69	*(3.7)*
Train (2)	3.87	*(7.5)*	5.19	*(4.6)*	3.87	*(7.9)*	5.19	*(4.8)*	5.07	*(6.8)*	5.08	*(7.2)*
Bus (3)	3.16	*(5.8)*	4.20	*(3.9)*	3.16	*(6.4)*	4.21	*(4.0)*	4.11	*(5.4)*	4.12	*(5.8)*
GCost ($00)	-1.55	*(3.1)*	-3.27	*(3.2)*	-1.55	*(3.7)*	-3.27	*(3.3)*	-3.17	*(4.3)*	-3.15	*(4.6)*
TTime (hours)	-5.77	*(6.4)*	-6.90	*(5.4)*	-5.77	*(10.8)*	-6.90	*(5.7)*	-6.78	*(7.0)*	-6.78	*(7.8)*
Income ($00,000) - Air (1)	1.33	*(1.4)*	3.68	*(1.4)*	1.33	*(1.1)*	3.68	*(1.4)*	3.53	*(1.4)*	3.45	*(1.5)*
$\sigma 1$			3.38	*(3.1)*	0.00	–	3.38	*(3.2)*	3.27	*(3.4)*	3.18	*(3.6)*
$\sigma 2$			0.143	*(0.0)*	0.0414	*(0.0)*	0.143	*(0.0)*	0.128	*(0.0)*	0.029	*(0.0)*
σ_3			0.00206	*(0.0)*	0.0181	*(0.0)*	0.00	---	0.00266	*(0.0)*	0.00584	*(0.0)*
σ_4			0.432	*(0.2)*	0.0558	*(0.0)*	0.434	*(0.2)*	0.00	---	0.00	---
Log Likelihood (simul.):	-199.128		-196.751		-199.118		-196.751		-196.768		-196.255	

The empirical results for the Mode Choice dataset are provided in Table 2-5. We estimate the 'unidentified' model to determine the parameters that are candidates for normalization. The results suggest that train, bus, or car can be used as the base (Greene normalizes the car alternative). We then report several identified models with different base alternatives normalized, and show that the model in which the air heteroscedastic term is the base is a mis-specified model (as indicated by the loss of fit).

Random Parameter Logit Kernel

Greene also reports a model that expands the unrestricted logit kernel model presented in Table 2-4 by including normally distributed random parameters for the cost, time, and income variables.[19] The primary issue here is that there are only 210 observations in the sample, and it is not a rich enough dataset to support the estimation of a large number of disturbance parameters. This is demonstrated with the empirical results reported in Table 2-6, in which we present a series of random parameter models starting with more parsimonious specifications.

The first model is the multinomial logit model, provided for comparison. Model 1-2 (estimated with 2000 and 4000 Halton draws) includes independent random parameters on the cost, time, and income variables. This model appears identified, and results in a large improvement in fit over the multinomial logit model.[20] The t-stats are low here due to the correlation among the parameter estimates. Model 4 shows that allowing for a single random parameter on the time variable achieves much of the total improvement in fit. Model 5-6 (estimated with 2000 and 4000 Halton draws) allows for a full set of correlations among the random parameters, and this results in a marginal improvement in fit over the independent model. (Note that the Cholesky parameters and not the variances and covariances are reported). Model 7 is estimated with a more parsimonious correlated structure. So far, these models all appear to be identified and provide significant (and similar) explanation of the disturbances. This is not the case for the remaining models. Model 8-9 includes the three independent random parameters along with heteroscedasticity, and the model appears unidentified. Model 10 is the model reported in Greene (although we normalized T_{33}). It includes an unrestricted covariance structure as well as the three independent random parameters, and the model appears unidentified. Louviere, Hensher and Swait report estimation results for a model similar to Greene (i.e., an unrestricted covariance structure with additional random parameters), and their model, too, appears unidentified.

The important points of these random parameter results are that, first, there are often several specifications that result in a similar improvement in fit. Second, that it is important not to overdue the specification, because it is easy to end up with an unidentified model.

---

[19] Note that since the time and cost parameters have a sign constraint, they should be specified with log-normally distributed parameters.

[20] Note that we achieved a much larger improvement in fit than any of the models reported in Greene and Louviere et al., even with this more parsimonious specification.

Table 2-6: Mode Choice Model – Random Parameters

Specification:	Multinomial Logit		Independent Random Parameters						Correlated Random Parameters					
Draws:			2000 Halton		4000 Halton		4000 Halton		2000 Halton		4000 Halton		4000 Halton	
Parameter	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Altern. Specific constants														
Air (1)	5.21	*(5.3)*	12.0	*(3.6)*	11.8	*(2.9)*	9.49	*(5.7)*	17.8	*(2.5)*	17.6	*(2.6)*	10.8	*(3.8)*
Train (2)	3.87	*(7.5)*	12.9	*(3.1)*	12.7	*(2.5)*	9.65	*(5.5)*	18.4	*(2.4)*	18.3	*(2.5)*	10.7	*(3.6)*
Bus (3)	3.16	*(5.8)*	11.6	*(3.2)*	11.5	*(2.6)*	8.69	*(5.5)*	16.7	*(2.4)*	16.5	*(2.5)*	9.7	*(3.7)*
GCost ($00)	-1.55	*(3.1)*	-4.21	*(2.0)*	-4.14	*(1.6)*	-2.57	*(3.3)*	-6.71	*(1.6)*	-6.53	*(1.8)*	-4.02	*(1.9)*
TTime (hours)	-5.77	*(6.4)*	-16.7	*(3.3)*	-16.5	*(2.7)*	-12.5	*(5.8)*	-24.1	*(2.4)*	-24.1	*(2.5)*	-13.4	*(3.9)*
Income ($00,000) - Air (1)	1.33	*(1.4)*	9.61	*(1.9)*	9.48	*(1.7)*	5.93	*(2.5)*	14.4	*(1.6)*	14.3	*(1.7)*	5.5	*(2.0)*
T11 (σ_1)														
T21														
T31														
T22 (σ_2)														
T32														
T33 (σ_3)														
GCost			0.493	*(0.4)*	0.332	*(0.1)*			4.99	*(0.9)*	4.86	*(1.1)*	3.00	*(1.3)*
TTime			10.7	*(2.5)*	10.6	*(2.1)*	7.9	*(3.7)*	13.6	*(2.0)*	14.1	*(2.0)*	3.86	*(0.4)*
Income - Air			8.34	*(1.3)*	8.18	*(1.1)*			6.94	*(1.0)*	5.56	*(1.3)*		
GCost - TTime									9.21	*(1.5)*	8.13	*(1.8)*	7.70	*(2.0)*
GCost - (Income-Air)									6.57	*(0.6)*	9.03	*(0.9)*		
TTime - (Income-Air)									-13.6	*(1.3)*	-14.6	*(1.5)*		
Log Likelihood (simul.):	-199.128		-177.523		-177.640		-178.680		-174.419		-174.420		-176.816	
Model:	*1*		*2*		*3*		*4*		*5*		*6*		*7*	

Specification:	Random Parameters & Heteroscedasticity				Random Param. & Unconstrained	
Draws:	2000 Halton		4000 Halton		2000 Halton	
Parameter	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Altern. Specific constants						
Air (1)	25.7	*n/a*	28.2	*n/a*	44.1	*n/a*
Train (2)	31.3	*n/a*	34.2	*n/a*	56.0	*n/a*
Bus (3)	27.8	*n/a*	30.4	*n/a*	48.4	*n/a*
GCost ($00)	-13.4	*n/a*	-14.6	*n/a*	-23.0	*n/a*
TTime (hours)	-39.5	*n/a*	-43.3	*n/a*	-69.9	*n/a*
Income ($00,000) - Air (1)	25.5	*n/a*	28.7	*n/a*	48.6	*n/a*
T11 (σ_1)	12.4	*n/a*	11.7	*n/a*	24.3	*n/a*
T21					2.69	*n/a*
T31					-0.389	*n/a*
T22 (σ_2)	2.16	*n/a*	2.07	*n/a*	4.90	*n/a*
T32					2.68	*n/a*
T33 (σ_3)	0.57	*n/a*	1.60	*n/a*	0.00	----
GCost	0.10	*n/a*	2.16	*n/a*	-2.67	*n/a*
TTime	25.5	*n/a*	28.1	*n/a*	45.8	*n/a*
Income - Air	6.69	*n/a*	18.69	*n/a*	13.1	*n/a*
GCost - TTime						
GCost - (Income-Air)						
TTime - (Income-Air)						
Log Likelihood (simul.):	-176.072		-176.036		-175.393	
Model:	*8*		*9*		*10*	

Empirical Application II: Telephone Service

In this section, we apply these methods to residential telephone demand analysis. The model involves a choice among five residential telephone service options for local calling. A household survey was conducted in 1984 for a telephone company and was used to develop a comprehensive model system to predict residential telephone demand (Train, McFadden and Ben-Akiva 1987). Below we use part of the data to estimate a model that explicitly accounts for inter-dependencies between residential telephone service options. We first describe the data. Then we present estimation results using a variety of error structures.

The Data

Local telephone service typically involves the choice between flat (i.e., a fixed monthly charge for unlimited calls within a specified geographical area) and measured (i.e., a reduced fixed monthly charge for a limited number of calls plus usage charges for additional calls) services. In the current application, five services are involved, two measured and three flat. They can be described as follows:

- *Budget measured* - no fixed monthly charge; usage charges apply to each call made.
- *Standard measured* - a fixed monthly charge covers up to a specified dollar amount (greater that the fixed charge) of local calling, after which usage charges apply to each call made.
- *Local flat* - a greater monthly charge that may depend upon residential location; unlimited free calling within local calling area; usage charges apply to calls made outside local calling area.
- *Extended area flat* - a further increase in the fixed monthly charge to permit unlimited free calling within an extended area.
- *Metro area flat* - the greatest fixed monthly charge that permits unlimited free calling within the entire metropolitan area.

The sample concerns 434 households. The availability of the service options of a given household depends on its geographical location. Details are provided in Table 2-7. In Table 2-8, we summarize the service option availabilities over the usable sample.

Table 2-7: Telephone Data - Availability of Service Options

Service Options	Geographic Location		
	Metropolitan Areas	Perimeter Exchanges Adjacent to Metro Areas	All Other
Budget Measured	Yes	Yes	Yes
Standard Measured	Yes	Yes	Yes
Local Flat	Yes	Yes	Yes
Extended Flat	No	Yes	No
Metro Flat	Yes	Yes	No

Table 2-8: Telephone Data - Summary Statistics on Availability of Service Options

Service Options	Chosen	Percent	Total Available
Budget Measured	73	0.168	434
Standard Measured	123	0.283	434
Local Flat	178	0.410	434
Extended Flat	3	0.007	13
Metro Flat	57	0.131	280
Total :	434	1.000	1595

Models

The model that we use in the present analysis is intentionally specified to be simple. The explanatory variables used to explain the choice between the five service options are four alternative-specific constants, which correspond to the first four service options, and a generic cost variable (the natural log of the monthly cost of each service options expressed in dollars). We investigated three types of error structures: heteroscedasticity, nested and cross-nested structures, and taste heterogeneity (random parameters).

Heteroscedastic

The results for the heteroscedastic case are provided in Table 2-9 and Table 2-10. Table 2-9 displays results from the unidentified model. To explore the issue of normalization of the minimum variance alternative, we estimated the unidentified model for various numbers of Halton draws and pseudo-random draws. The results suggest that there is no strong base alternative, and it could be either alternative 1, 2, 4, or 5. Table 2-10 provides estimation results for identified heteroscedastic models. Again, to explore the issue of the minimum variance alternatives, 5 identified models were estimated, each one with a different base heteroscedastic term. (Note that this defeats the purpose of estimating the unidentified model, but was done for illustration purposes only.) As indicated by the unidentified models, the identified model estimation results support the conclusion that any of alternatives 1, 2, 4, or 5 could be set as the base. However, constraining s_3 to zero results in a significant loss of fit, whereas constraining it to 4.0 brings it in line with the correctly specified model. Comparing the correctly specified heteroscedastic models with the MNL model, there is an obvious gain in likelihood from incorporating heteroscedasticity, primarily due to capturing the high variance of alternative 3.

Table 2-9: Telephone Model - Heteroscedastic Unidentified Models to Determine Base

	100 Halton		200 Halton		400 Halton		1000 Halton		2000 Halton		5000 'Random'		10000 'Random'	
Parameter	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Altern. Specific constants														
Budget Measured (1)	-3.30	*(6.9)*	-163.39	*n/a*	-3.28	*(7.5)*	-3.28	*(7.7)*	-3.27	*(7.6)*	-3.32	*(7.2)*	-3.29	*(7.7)*
Standard Measured (2)	-2.55	*(5.5)*	-126.84	*n/a*	-2.53	*(6.3)*	-2.53	*(6.4)*	-2.52	*(6.8)*	-2.55	*(6.4)*	-2.53	*(6.5)*
Local Flat (3)	-1.38	*(3.5)*	-78.09	*n/a*	-1.37	*(3.6)*	-1.37	*(3.6)*	-1.36	*(3.6)*	-1.38	*(3.7)*	-1.37	*(3.6)*
Extended Flat (4)	-1.07	*(1.3)*	-44.31	*n/a*	-1.04	*(1.3)*	-1.04	*(1.3)*	-1.04	*(1.5)*	-1.06	*(1.5)*	-1.04	*(1.4)*
Log Cost	-2.70	*(7.2)*	-145.18	*n/a*	-2.68	*(7.9)*	-2.68	*(8.2)*	-2.67	*(8.4)*	-2.70	*(8.1)*	-2.69	*(7.6)*
σ_1	0.10	*(0.3)*	60.29	*n/a*	0.06	*(0.3)*	0.03	*(0.2)*	0.00	*(0.1)*	0.31	*(0.5)*	0.13	*(0.4)*
σ_2	0.30	*(0.3)*	61.19	*n/a*	0.21	*(0.3)*	0.14	*(0.4)*	0.06	*(0.3)*	0.20	*(0.2)*	0.08	*(0.2)*
σ_3	2.91	*(3.2)*	196.53	*n/a*	2.88	*(3.3)*	2.88	*(3.4)*	2.87	*(3.6)*	2.91	*(4.3)*	2.91	*(3.1)*
σ_4	0.39	*(0.3)*	16.18	*n/a*	0.01	*(0.0)*	0.04	*(0.1)*	0.01	*(0.0)*	0.11	*(0.2)*	0.07	*(0.3)*
σ_5	0.22	*(0.2)*	81.36	*n/a*	0.01	*(0.1)*	0.09	*(0.3)*	0.01	*(0.0)*	0.05	*(0.1)*	0.26	*(0.2)*
(Simul.) Log-Likelihood:	-471.09		-468.27		-471.16		-471.20		-471.19		-470.89		-471.38	

Table 2-10: Telephone Model - Identified Heteroscedastic Models

	MNL		Identified Heteroscedastic Model															
			1000 Halton		1000 Halton		1000 Halton		1000 Halton		1000 Halton		1000 Halton		5000 'Random'		10000 'Random'	
Parameter	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Altern. Specific constants																		
Budget Measured (1)	-2.46	*(8.4)*	-3.27	*(7.9)*	-3.27	*(7.1)*	-5.03	*(2.4)*	-3.28	*(6.0)*	-3.27	*(7.8)*	-3.91	*(2.2)*	-3.28	*(7.6)*	-3.28	*(6.5)*
Standard Measured (2)	-1.74	*(6.6)*	-2.53	*(6.6)*	-2.52	*(6.2)*	-3.85	*(2.2)*	-2.53	*(6.1)*	-2.52	*(6.5)*	-3.02	*(2.4)*	-2.53	*(6.5)*	-2.53	*(5.0)*
Local Flat (3)	-0.54	*(2.7)*	-1.37	*(3.8)*	-1.36	*(3.2)*	-1.09	*(2.1)*	-1.37	*(3.6)*	-1.36	*(3.7)*	-1.67	*(3.3)*	-1.37	*(3.8)*	-1.37	*(3.4)*
Extended Flat (4)	-0.74	*(1.1)*	-1.04	*(1.3)*	-1.04	*(1.3)*	-1.37	*(1.5)*	-1.04	*(1.4)*	-1.04	*(1.4)*	-1.10	*(1.2)*	-1.05	*(1.3)*	-1.04	*(1.4)*
Log Cost	-2.03	*(9.6)*	-2.68	*(8.2)*	-2.67	*(4.9)*	-3.24	*(3.1)*	-2.68	*(6.2)*	-2.67	*(8.2)*	-3.33	*(2.9)*	-2.68	*(8.1)*	-2.69	*(7.6)*
σ_1					0.02	*(0.1)*	2.77	*(1.8)*	0.03	*(0.0)*	0.03	*(0.3)*	0.76	*(0.4)*				
σ_2			0.13	*(0.3)*			3.27	*(1.6)*	0.14	*(0.1)*	0.14	*(0.3)*	0.70	*(0.3)*	0.11	*(0.2)*	0.10	*(0.2)*
σ_3			2.88	*(4.9)*	2.88	*(2.4)*			2.88	*(3.3)*	2.87	*(3.8)*	4.00	----	2.89	*(4.7)*	2.91	*(2.9)*
σ_4			0.04	*(0.1)*	0.04	*(0.1)*	1.14	*(0.5)*			0.04	*(0.1)*	0.11	*(0.1)*	0.12	*(0.2)*	0.07	*(0.1)*
σ_5			0.09	*(0.3)*	0.09	*(0.2)*	0.01	*(0.0)*	0.10	*(0.0)*			1.33	*(1.3)*	0.03	*(0.1)*	0.26	*(0.2)*
(Simul.) Log-Likelihood:	-477.56		-471.20		-471.20		-476.66		-471.20		-471.20		-471.42		-470.92		-471.39	

Nested & Cross-Nested Structures

In Table 2-11, the estimation results of various nested and cross-nested specifications are provided. Table 2-11a reports results for identified model structures (as can be verified by the rank condition). The best specification is model 3, in which the first two alternatives are nested, the last two alternatives are nested, and the third term has a heteroscedastic term. This provides a significant improvement in fit over the MNL specification shown in the first column, and also provides a better fit than the heteroscedastic models in Table 2-10. The poor fit for many of the nesting and cross-nesting specifications is due to the fact that the variance for alternative 3 is constrained to be in line with the other variances. The heteroscedastic models indicated that it has a much higher variance, and when this was added to the nested and cross-nested models (see Table 2-11b) the fit improved dramatically.[21]

Table 2-11c provides results for the unidentified model in which the first two alternatives are nested and the last 3 alternatives are nested, and we attempt (incorrectly) to estimate both error parameters. The first model, estimated with 1,000 Halton draws, appears to be identified. However, the second model, estimated using different starting values, shows that this is not the case; it has an identical fit, but very different estimates of the error parameters. This is as expected, because only the sum of the variances $(s_1^2 + s_2^2)$ can be identified. The remaining columns show that it can take a very large number of draws to get the telltale sign of an unidentified model, the singular Hessian – in this case, 80,000 Halton draws. (Again, the actual number depends on the specification and the data.) Table 2-11d shows that the normalization for the 2 nest model is arbitrary. The table presents three normalizations resulting in identical fits where:

$$\{ 1, 1, 0, 0, 0 \} = \{ 0, 0, 2, 2, 2 \} = \{ 1, 1, 2, 2, 2 \text{ with } s_1 = s_2 \}.$$

Table 2-11: Telephone Model - Nested & Cross-Nested Error Structures

Table a: Identified Nesting & Cross-Nesting Error Structures

	Nested Structures										Cross-Nested Structures					
Specification*:	1, 1, 2, 2, 0		1, 1, 2, 2, 3		1, 1, 2, 3, 3		1, 1, 2, 3, 3		1, 1, 2, 2, 2 ($s_1 = s_2$)		1, 1, 1-2, 2, 2		1-2, 2-3, 3-4, 4-5, 5-6 (all s equal)		1-2, 2-3, 3-4, 4-5, 5-6 (all s equal)	
Draws:	1000 Halton		1000 Halton		1000 Halton		2000 Halton		1000 Halton		1000 Halton		1000 Halton		5000 Halton	
Parameter	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Altern. Specific constants																
Budget Measured (1)	-3.63	*(5.0)*	-3.63	*(5.0)*	-3.79	*(5.4)*	-3.80	*(5.3)*	-3.80	*(5.7)*	-3.80	*(5.7)*	-2.83	*(2.4)*	-2.72	*(3.1)*
Standard Measured (2)	-2.85	*(4.3)*	-2.85	*(4.3)*	-3.00	*(4.6)*	-3.01	*(4.6)*	-3.01	*(4.9)*	-3.00	*(4.9)*	-1.90	*(3.1)*	-1.85	*(3.9)*
Local Flat (3)	-1.48	*(3.1)*	-1.48	*(3.1)*	-1.63	*(3.1)*	-1.64	*(3.1)*	-1.09	*(3.6)*	-1.09	*(3.5)*	-0.55	*(2.3)*	-0.54	*(2.4)*
Extended Flat (4)	-1.52	*(1.5)*	-1.52	*(1.5)*	-1.18	*(1.3)*	-1.18	*(1.3)*	-1.19	*(1.4)*	-1.19	*(1.4)*	-0.76	*(1.0)*	-0.75	*(1.0)*
Log Cost	-3.05	*(4.5)*	-3.05	*(4.5)*	-3.19	*(5.0)*	-3.20	*(5.0)*	-3.25	*(6.1)*	-3.25	*(6.1)*	-2.40	*(2.1)*	-2.29	*(2.6)*
σ_1	1.32	*(1.1)*	1.32	*(1.1)*	1.55	*(1.5)*	1.55	*(1.6)*	2.16	*(3.0)*	0.01	*(0.8)*	0.65	*(0.6)*	0.53	*(0.6)*
σ_2	3.02	*(2.9)*	3.02	*(2.9)*	3.34	*(2.9)*	3.37	*(2.8)*			3.04	*(3.0)*				
σ_3			0.00	*(0.0)*	0.01	*(0.1)*	0.01	*(0.2)*								
(Simul.) Log-Likelihood:	-471.26		-471.26		-470.70		-470.64		-473.04		-473.05		-477.48		-477.51	

[21] Therefore, the problem identified earlier with the cross-nested *1, 1, 1-2, 2, 2* structure does not apply to this dataset. In fact, as shown by the models in Table 2-11c, alternative 3 has an even larger relative variance than the *1, 1, 1-2, 2, 2* structure provides.

Table b: Nesting / Cross-Nesting plus Heteroscedasticity (0, 0, 1, 0, 0)

	Combined Models					
Specification*:	2, 2, 1-3, 3, 3 (s 2= s 3)		2, 2, 2-1-3, 3, 3		2-3, 3-4, 4-1-5, 5-6, 6-7 (s 2... s 7 equal)	
Draws:	1000 Halton		1000 Halton		1000 Halton	
Parameter	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Altern. Specific constants						
Budget Measured (1)	-3.81	*(5.5)*	-3.80	*(5.3)*	-3.28	*(7.3)*
Standard Measured (2)	-3.02	*(4.7)*	-3.01	*(4.6)*	-2.53	*(6.3)*
Local Flat (3)	-1.64	*(3.1)*	-1.64	*(3.1)*	-1.37	*(3.5)*
Extended Flat (4)	-1.19	*(1.3)*	-1.18	*(1.3)*	-1.04	*(1.3)*
Log Cost	-3.21	*(5.2)*	-3.20	*(5.0)*	-2.68	*(8.0)*
$\sigma 1$	3.37	*(2.8)*	3.38	*(2.8)*	2.88	*(3.3)*
$\sigma 2$	1.11	*(1.6)*	0.03	*(0.3)*	0.09	*(0.2)*
$\sigma 3$			1.55	*(1.6)*		
(Simul.) Log-Likelihood:	-470.64		-470.69		-471.22	

Table c: Unidentified Nested Error Structures

Specification*:	1, 1, 2, 2, 2 (Unidentified - can only estimate ($\sigma 1^2+\sigma 2^2$))											
Draws:	1000 Halton		1000 Halton		10000 Halton		40000 Halton		40000 'Random'		80000 Halton	
Parameter	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Altern. Specific constants												
Budget Measured (1)	-3.80	*(5.7)*	-3.80	*(5.7)*	-3.80	*(5.7)*	-3.80	*(5.8)*	-3.81	*(5.7)*	-3.80	*n/a*
Standard Measured (2)	-3.01	*(4.9)*	-3.01	*(4.9)*	-3.01	*(4.9)*	-3.01	*(4.9)*	-3.01	*(4.8)*	-3.01	*n/a*
Local Flat (3)	-1.09	*(3.6)*	-1.09	*(3.6)*	-1.09	*(3.6)*	-1.09	*(3.6)*	-1.09	*(3.5)*	-1.09	*n/a*
Extended Flat (4)	-1.19	*(1.4)*	-1.19	*(1.4)*	-1.19	*(1.4)*	-1.19	*(1.4)*	-1.19	*(1.4)*	-1.19	*n/a*
Log Cost	-3.25	*(6.1)*	-3.25	*(6.1)*	-3.25	*(6.1)*	-3.25	*(6.1)*	-3.25	*(6.0)*	-3.25	*n/a*
$\sigma 1$	2.65	*(3.1)*	0.78	*(0.5)*	2.55	*(2.5)*	2.56	*(1.5)*	1.83	*(1.1)*	1.93	*n/a*
$\sigma 2$	1.51	*(2.2)*	2.95	*(3.3)*	1.67	*(3.8)*	1.68	*(0.4)*	2.45	*(1.9)*	2.36	*n/a*
$(\sigma 1^2+\sigma 2^2)^{1/2}$	*3.05*		*3.05*		*3.05*		*3.06*		*3.06*		*3.05*	
(Simul.) Log-Likelihood:	-473.02		-472.99		-473.02		-473.02		-472.95		-473.02	

Table d: Identical (Identified) Nested Error Structures

Specification*:	1, 1, 0, 0, 0		0, 0, 2, 2, 2		1, 1, 2, 2, 2 ($\sigma_1=\sigma_2$)			
Draws:	1000 Halton		1000 Halton		1000 Halton		2000 Halton	
Parameter	*Est*	*T-stat*	*Est*	*T-stat*	*Est*	*T-stat*	*Est*	*T-stat*
Altern. Specific constants								
Budget Measured (1)	-3.80	*(5.7)*	-3.80	*(5.7)*	-3.80	*(5.7)*	-3.80	*(5.8)*
Standard Measured (2)	-3.01	*(4.9)*	-3.01	*(4.9)*	-3.01	*(4.9)*	-3.01	*(4.9)*
Local Flat (3)	-1.09	*(3.6)*	-1.09	*(3.6)*	-1.09	*(3.6)*	-1.09	*(3.6)*
Extended Flat (4)	-1.19	*(1.4)*	-1.19	*(1.4)*	-1.19	*(1.4)*	-1.19	*(1.4)*
Log Cost	-3.25	*(6.1)*	-3.25	*(6.1)*	-3.25	*(6.1)*	-3.25	*(6.1)*
$\sigma 1$	3.05	*(3.0)*			2.16	*(3.0)*	2.15	*(3.0)*
$\sigma 2$			3.05	*(3.0)*	2.16	---	2.15	---
$(\sigma 1^2+\sigma 2^2)^{1/2}$	*3.05*		*3.05*		*3.05*		*3.04*	
(Simul.) Log-Likelihood:	-473.02		-473.03		-473.04		-473.01	

* the specification lists the factors (and sigmas) that apply to each of the five alternatives

Random Parameters

We also considered unobserved taste heterogeneity for the parameter on log of cost. Since the parameter has a sign constraint, a lognormal distribution is used. (Draws from a lognormal distribution are generated by exponentiating draws taken from a normal distribution.) The results are shown in Table 2-12. The first model shows that when there are no other covariance parameters specified, the heterogeneity on log cost is insignificant. However, the second model shows that heterogeneity does add slightly to the explanatory power of the best nested model as specified in Table 2-11a. The remaining 4 models report specifications with both heterogeneity and taste variation. While the rank and order conditions suggest that a model with 4 heteroscedastic parameters and the lognormal parameter is identified, the estimation results show that there is a multicollinearity problem. Note that when only 200 pseudo-random draws are used, this model appears, incorrectly, to be identified.

Table 2-12: Telephone Model - Taste Variation, Lognormal Parameter for Log(Cost)

Specification*:	Taste Variation		1,1,2,3,3 & Taste Variation				1,2,3,4,5 & Taste Variation							
Draws:	1000 Halton		1000 Halton		2000 Halton		1000 Halton		200 'Random'		1000 Halton		1000 Halton	
Parameter	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Altern. Specific constants														
Budget Measured (1)	-2.46	*(8.2)*	-3.48	*(5.7)*	-3.50	*(4.3)*	-24.20	*n/a*	-4.06	*(2.6)*	-30.36	*n/a*	-26.84	*n/a*
Standard Measured (2)	-1.74	*(6.5)*	-2.68	*(4.7)*	-2.70	*(3.5)*	-16.75	*n/a*	-3.06	*(2.8)*	-22.03	*n/a*	-19.41	*n/a*
Local Flat (3)	-0.54	*(2.7)*	-1.44	*(3.1)*	-1.45	*(2.7)*	-7.57	*n/a*	-1.57	*(2.4)*	-10.72	*n/a*	-9.77	*n/a*
Extended Flat (4)	-0.74	*(1.0)*	-0.98	*(1.1)*	-0.98	*(1.1)*	-3.33	*n/a*	-1.07	*(1.1)*	-5.11	*n/a*	-4.75	*n/a*
Log Cost **	-2.03	*(9.6)*	-3.17	*(5.6)*	-3.18	*(5.1)*	-23.30	*n/a*	-3.69	*(2.7)*	-28.38	*n/a*	-26.02	*n/a*
σ Log Cost **	0.00	*(0.1)*	1.18	*(1.1)*	1.16	*(1.0)*	18.39	*n/a*	1.65	*(1.4)*	18.85	*n/a*	18.54	*n/a*
$\sigma 1$			0.40	*(0.1)*	0.50	*(0.1)*	12.38	*n/a*	1.00	*(0.6)*	13.72	*n/a*	12.19	*n/a*
$\sigma 2$			3.56	*(3.0)*	3.58	*(3.0)*	9.06	*n/a*	0.72	*(0.5)*	11.34	*n/a*	9.02	*n/a*
$\sigma 3$			0.05	*(0.8)*	0.01	*(0.1)*	24.50	*n/a*	4.13	*(2.3)*	30.45	*n/a*	28.96	*n/a*
$\sigma 4$							0.49	*n/a*						
$\sigma 5$							0.88	*n/a*	0.24	*(0.6)*	1.26	*n/a*		
Log Likelihood (simul.):	-477.56		-470.36		-470.28		-469.15		-470.74		-468.69		-469.47	

** the mean and standard deviation of the lognormal are reported

Summary of Telephone Data Models

By far the most important part of the error structure for the telephone dataset is that the Local Flat Alternative (3) has a significantly higher variance than the other alternatives. Note that a simple heteroscedastic model outperforms the most obvious nested structure in which the measured alternatives are nested together and the flat alternatives are nested together. Marginal improvements can be achieved by incorporating nesting, cross-nesting or taste variation as long as alternative 3 is allowed a free variance. While this dataset served its purpose in highlighting specification and identification issues, one would ideally like to estimate such logit kernel models with larger datasets.

Conclusion

In this chapter we presented general rules for specification, identification, and estimation via maximum simulated likelihood for the logit kernel model. We presented guidelines for examining identification and normalization, which consisted of three conditions: order, rank, and positive definiteness. The positive definiteness condition is not an issue for probit models. However, as the heteroscedastic case highlights, it can have important consequences for logit kernel. We emphasized that identification must be examined on a case-by-case basis, and that it is not necessarily intuitive. Furthermore, given the fact that simulation has a tendency to mask identification problems, it becomes even more critical that identification is well understood.

We discussed in detail the specification and identification of many of the special cases, all within a general factor analytical framework, including:

Heteroscedasticity:	F_n diagonal (fixed) ; T diagonal.
Nesting (Cross-Nesting):	$F_n F_n'$ block-diagonal (fixed) ; T diagonal.
Error Components:	F_n fixed to 0/1 ; T (usually) diagonal.
Factor Analytic:	F_n unknown ; T triangular.
Autoregressive Process:	F_n moving average form of a GAR(1) process ; T diagonal.
Random parameters:	F_n a function of explanatory variables (fixed) ; T triangular.

Just as there are well-known standard rules for identification for the systematic parameters in a multinomial logit, we aimed to develop identification rules for the disturbance parameters of the logit kernel model. There are critical differences between the identification of these parameters and the identification of their counterparts in both the systematic portion of the utility as well as their counterparts in a probit model. The following summarizes these identification rules:

Heteroscedasticity

$J = 2$ alternatives:	0 parameters identified.
$J \geq 3$ alternatives:	$J-1$ parameters identified & must constrain the minimum variance term to 0.

Nesting

$M = 2$ nests:	$M-1$ parameters identified & normalization is arbitrary.
$M \geq 3$ nests:	M parameters identified.

Random parameters

Beyond the specific rules listed below, can estimate as many random parameters as the data will support.

Alternate-specific variables

Rules for heteroscedasticity, nesting, and error components apply.

Categorical variables with independently distributed parameters

$M = 2$ categories:	$M - 1$ parameters identified & normalization is arbitrary.
$M \geq 3$ or more categories:	M parameters identified. (Includes a binary categorical variable that does not enter all utilities.)

Characteristics of the Decision-maker with independently distributed parameters

Interacts with alternative-specific constants:	Analogous to the heteroscedastic case: $J - 1$ parameters identified & must constrain the minimum variance term to 0.
Interacts with nest-specific constants:	Analogous to nested case:
$M = 2$ nests:	$M - 1$ parameters identified.
$M \geq 3$ nests:	M parameters identified.

Our objectives were that through examination of the special cases we would be able to establish some identification and specification rules, and also highlight some of the broad themes and provide tools for uncovering other potential issues pertaining to logit kernel models. Clearly there are numerous identification issues that are not covered by the above list. Therefore, models have to be examined on a case-by-case basis. For the alternative-specific portion of the disturbance, it is recommended that the rank and order conditions be programmed into the estimation program. When the positive definiteness condition comes into play, it is recommended to examine the problem analytically, where possible, or empirically (by investigating various normalizations). For random parameter models, it is recommended to use the above identification rules as guidelines, and then empirically establish identification by (1) verifying that the parameter estimates are stable as the number of draws are increased and (2) checking that the Hessian is non-singular at the convergence point.

One of the most important points of the chapter is that there are critical aspects to the logit kernel specification that are often overlooked in the literature. It must be remembered that this is a relatively new methodology, and there are numerous aspects that warrant further research, including:

- More testing and experience with applications,
- Further exploration of identification and normalization issues,
- Continued compilation and analysis of special cases and rules of identification,
- Better understanding of the impact on analysis of different factor specifications (particularly since often several factor specification will provide similar fit to the data),
- Investigation of analogous specifications estimated via different methods (for example, logit kernel versus probit, nested logit, cross-nested logit, heteroscedastic extreme value, etc.)
- Additional comparisons with GHK and other smooth simulators, and

- Further examination of Halton draws as well as other pseudo- and quasi-random drawing methods.

Finally, we also may need to look at modifying the specification of the logit kernel model to alleviate some of the complications. One of the issues with the logit kernel specification is that while pure logit is a special case of the model, pure probit is not. Our analysis assumes that it is acceptable to include the Gumbel term in the model. However, the Gumbel term may, in fact, have no business being in the model. For this reason, we would ideally want to specify and estimate the model in a way that allows the Gumbel term to disappear. Conceptually, such a model could be specified as a linear combination of the two error terms, so Equation [2-4] (assuming a universal choice set) would become:

$$U_n = X_n b + \sqrt{(g/m^2)(1-l^2)} F_n T z_n + l n_n \ ,$$

where l is an unknown parameter. The covariance of the model is then a linear combination of the two covariance matrices:

$$\text{cov}(U_n) = \left((1-l^2)F_n T T' F_n' + l^2 I_J\right)\left(g/m^2\right) \ .$$

Conceptually this Combined Logit-Probit (CLP) specification is an appealing model. Note that a strict application of the order and rank conditions lead to the conclusion that the model is not identified. However, as we described in the section on identification, the slight difference between the Gumbel and Normal distributions makes the model identified (albeit, nearly singular).

To summarize, the logit kernel formulation has a tremendous amount of potential, because it can replicate any desirable error structure and is straightforward to estimate via maximum simulated likelihood. However, it also has some issues that must be understood for proper specification. As increased computational power and readily available software open up these techniques for widespread use, it is a critical time to understand and address the nuances of the logit kernel model.

Chapter 3:
Integration of Choice and Latent Variable Models[22]

Chapter 2 focused on the random portion of the utility function. The extension described in this chapter focuses on the causal structure and the specification of the systematic part of the utility function. The methodology we investigate can be used when important causal variables are not directly observable. The idea is to explicitly incorporate latent constructs such as attitudes and perceptions, or any amorphous concept affecting choice, in an effort to produce more behaviorally realistic models. This method makes use of what are called psychometric indicators (for example, responses to survey questions about attitudes, perceptions, or decision-making protocols), which are manifestations of the underlying latent variable. The objective of the work presented here is to develop a general framework and methodology for incorporating latent variables into choice models.

Introduction

Recent work in discrete choice models has emphasized the importance of the explicit treatment of psychological factors affecting decision-making. (See, for example, Koppelman and Hauser, 1979; McFadden, 1986; Ben-Akiva and Boccara, 1987; Ben-Akiva, 1992; Ben-Akiva et al., 1994; Morikawa et al., 1996.) A guiding philosophy in these developments is that the incorporation of psychological factors leads to a more behaviorally realistic representation of the choice process, and consequently, better explanatory power.

This chapter presents conceptual and methodological frameworks for the incorporation of latent factors as explanatory variables in choice models. The method described provides for explicit treatment of the psychological factors affecting the decision-making process by modeling them as latent variables. Psychometric data, such as responses to attitudinal and perceptual survey questions, are used as indicators

[22] This chapter is based on Ben-Akiva, Walker, et al. (1999).

of the latent psychological factors. The resulting approach integrates choice models with latent variable models, in which the system of equations is estimated simultaneously. The simultaneous estimation of the model structure represents an improvement over sequential methods, because it produces consistent and efficient estimates of the parameters. (See Everitt, 1984 and Bollen, 1989 for an introduction to latent variable models and Ben-Akiva and Lerman, 1985 for a textbook on discrete choice models.)

Three prototypical applications from the literature are reviewed to provide conceptual examples as well as sample equations and estimation results. The applications illustrate how psychometric data can be used in choice models to improve the definition of attributes and to better capture taste heterogeneity. They also demonstrate the flexibility and practicality of the methodology, as well as the potential gain in explanatory power and improved specifications of discrete choice models.

Related Literature

As described in the Chapter 1, discrete choice models have traditionally presented an individual's choice process as a black box, in which the inputs are the attributes of available alternatives and individual characteristics, and the output is the observed choice. The resulting models directly link the observed inputs to the observed output, thereby assuming that the inner workings of the black box are *implicitly* captured by the model. For example, discrete choice models derived from random utility theory do not model explicitly the formation of attitudes and perceptions. The framework for the random utility discrete choice model shown in Chapter 1 is repeated in Figure 3-1.[23]

There has been much debate in the behavioral science and economics communities on the validity of the assumptions of utility theory. Behavioral researchers have stressed the importance of the cognitive workings inside the black box on choice behavior (see, for example, Abelson and Levy, 1985 and Olson and Zanna, 1993), and a great deal of research has been conducted to uncover cognitive anomalies that appear to violate the basic axioms of utility theory (see, for example, Gärling, 1998, and Rabin, 1998). McFadden (1997) summarizes these anomalies and argues that "most cognitive anomalies operate through errors in perception that arise from the way information is stored, retrieved, and processed" and that "empirical study of economic behavior would benefit from closer attention to how perceptions are formed and how they influence decision-making." To address such issues, researchers have worked to enrich choice models by modeling the cognitive workings inside the black box, including the explicit incorporation of factors such as attitudes and perceptions.

A general approach to synthesizing models with latent variables and psychometric measurement models has been advanced by a number of researchers including Keesling (1972), Jöreskog (1973), Wiley (1973), and Bentler (1980), who developed the structural and measurement equation framework and methodology for specifying and estimating latent variable models. Such models are widely used to define and measure unobservable factors. Estimation is performed by minimizing the discrepancy between (a) the covariance

---

[23] Note that the terms in ellipses represent *unobservable* (i.e., latent) constructs, while those in rectangles represent *observable* variables. Solid arrows represent *structural equations* (cause-and-effect relationships) and dashed arrows represent *measurement equations* (relationships between the underlying latent variables and their observable indicators).

matrix of observed variables and (b) the theoretical covariance matrix predicted by the model structure, which is a function of the unknown parameters. Much of this work focuses on continuous latent constructs and continuous indicators. When discrete indicators are involved, direct application of the approach used for continuous indicators results in inconsistent estimates. For the case of discrete indicators, various corrective procedures can be applied. Olsson (1979), Muthén (1979, 1983, and 1984), and others developed procedures based on the application of polychoric correlations (rather than the Pearson correlations used for continuous indicators) to estimate the covariance matrix of the latent continuous indicators from the discrete indicators. Consistent estimates of the parameters can then be obtained by minimizing the discrepancy between this estimated covariance matrix and the theoretical covariance matrix. (See Bollen, 1989, for more discussion of discrete indicators.) Estimation methods for the situation of discrete latent variables and discrete indicators was developed by Goodman (1974)—see McCutcheon (1987) for a discussion.

In the area of choice modeling, researchers have used various techniques in an effort to explicitly capture psychological factors in choice models. One approach applied is to include *indicators* of psychological factors (such as responses to survey questions regarding individuals' attitudes or perceptions) directly in the utility function as depicted in Figure 3-2 (see, for example, Koppelman and Hauser, 1979; Green, 1984; Harris and Keane, 1998).

Another frequently used approach is to first perform factor analysis on the indicators, and then use the fitted latent variables in the utility, as shown in Figure 3-3. (See, for example, Prashker, 1979a,b; and Madanat et al., 1995). Note that these fitted variables contain measurement error, and so to obtain consistent estimates, the choice probability must be integrated over the distribution of the latent variables, where the distribution of the factors is obtained from the factor analysis model. (See, for example, Morikawa, 1989.)

Other approaches have been developed in market research (in an area called *internal market analysis*), in which both latent attributes of the alternatives and consumer preferences are inferred from preference or choice data. (For a review of such methods, see Elrod, 1991; and Elrod and Keane, 1995.) For example, Elrod 1988 and 1998, Elrod and Keane 1995, and Keane 1997 develop random utility choice models (multinomial logit and probit) that contain latent attributes. In estimating these models, they do not use any indicators other than the observed choices. Therefore, the latent attributes are alternative-specific and do not vary among individuals in a market segment. (In this way, they can be described as the alternative-specific factor analytic specification presented in Chapter 2.) However they do use perceptual indicators post-estimation to aid in interpretation of the latent variables. The framework for their model is shown in Figure 3-4. Wedel and DeSarbo (1996) and Sinha and DeSarbo (1997) describe a related method based on multidimensional scaling.

This research extends the above-described methods by formulating a general treatment of the inclusion of latent variables in discrete choice models. The formulation incorporates psychometric data as indicators of the latent variables. We employ a simultaneous maximum likelihood estimation method for integrated latent variable and discrete choice models, which results in consistent and efficient estimates of the model parameters. The formulation of the integrated model and the simultaneous estimator are described in the following sections of the chapter.

Contribution of the Chapter

The work on the methodology presented here began during the mid-1980s with the objective of making the connection between econometric choice models and the extensive market research literature on the study of consumer preferences (Cambridge Systematics, 1986; McFadden, 1986; and Ben-Akiva and Boccara, 1987). A number of empirical case studies, a sampling of which is reviewed in this chapter, have been undertaken over the years. While the ideas have been around for some time, the literature contains only empirical applications to specific problems (for example, the case studies reviewed here) or restricted model formulations (for example, the elegant formulation for a binary probit and MIMC model presented in McFadden, 2000, and Morikawa et al., 1996). The contribution of this chapter is the presentation of a general specification and estimation method for the integrated model, which provides complete flexibility in terms of the formulation of both the choice model and the latent variable model. In addition, the proposed method is reviewed within the context of other potential approaches, and its advantages discussed.

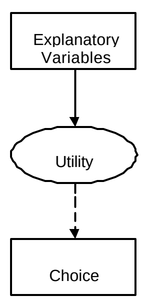


Figure 3-1:
Random Utility Discrete Choice Model

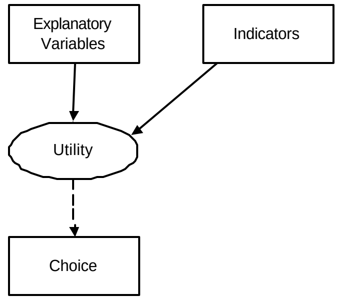


Figure 3-2:
Choice Model with Indicators Directly Included in Utility

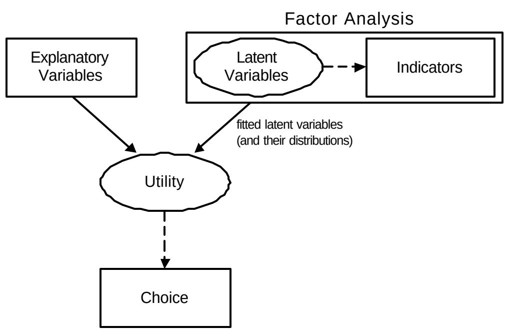


Figure 3-3:
Sequential Estimation: Factor Analysis followed by a Choice Model

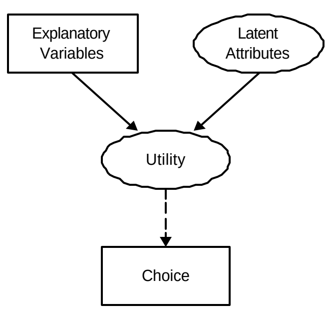


Figure 3-4:
Choice Model with Latent Attributes

Behavioral Framework for Choice Models with Latent Variables

Before presenting the methodological framework and specification, it is useful to discuss the behavioral framework behind joint choice and latent variable models. The framework is presented in Figure 3-5 (Ben-Akiva and Boccara, 1987), and the notation will be explained in the next section. The objective is to explicitly analyze latent psychological factors in order to gain information on aspects of individual behavior that cannot be inferred from market behavior or revealed preferences. In this behavioral framework, three types of latent factors are identified: attitudes, perceptions, and preferences.

Cause-Effect Behavioral Relationships

Attitudes and perceptions of individuals are hypothesized to be key factors that characterize the underlying behavior. The observable explanatory variables, including characteristics of the individual (for example, socio-economics, demographics, experience, expertise, etc.) and the attributes of alternatives (for example, price) are linked to the individual's attitudes and perceptions through a causal mapping. Since attitudes and perceptions are unobservable to the analyst, they are represented by latent constructs. These latent attitudes and perceptions, as well as the observable explanatory variables, affect individuals' preferences toward different alternatives and their decision-making process.

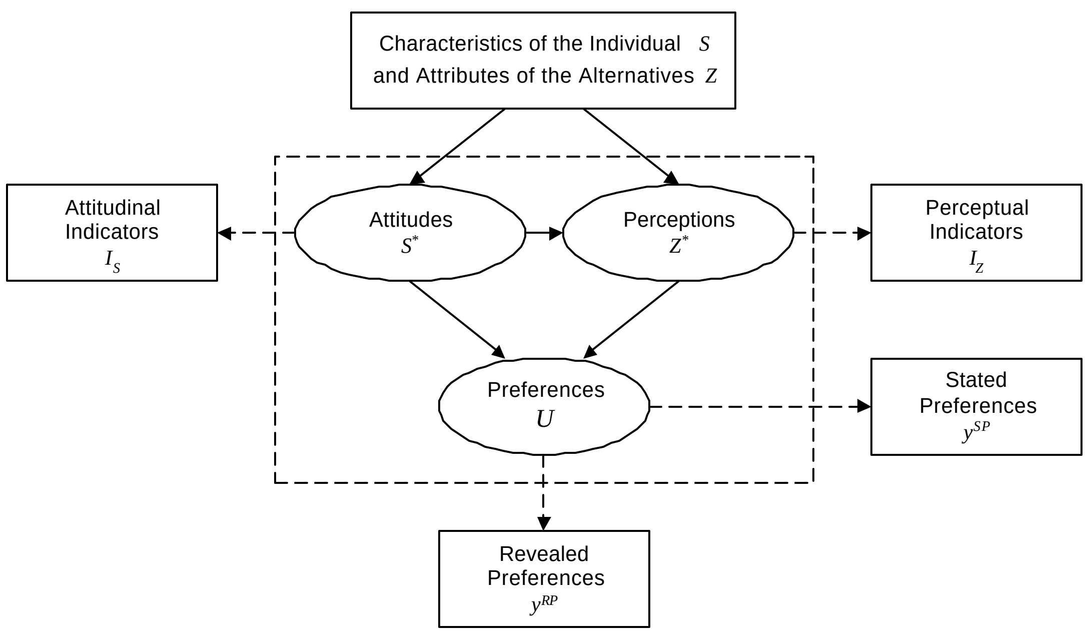


Figure 3-5: Behavioral Framework for Choice Models with Latent Variables

Perceptions are the individuals' beliefs or estimates of the levels of attributes of the alternatives. The choice process is expected to be based on perceived levels of attributes. Perceptions explain part of the random component of the utility function through individual-specific unobserved attributes. Examples of perceptions in a travel mode choice context for the transit alternative are *safety*, *convenience*, *reliability*, and *environmental friendliness*. Examples of perceptions for toothpaste are *health benefit* and *cosmetic benefit* (Elrod, 1998).

Attitudes are latent variables corresponding to the characteristics of the decision-maker. Attitudes reflect individuals' needs, values, tastes, and capabilities. They are formed over time and are affected by experience and external factors that include socioeconomic characteristics. Attitudes explain unobserved individual heterogeneity, such as taste variations, choice set heterogeneity and decision protocol heterogeneity. Examples of attitudes in a travel mode choice context are *the importance of reliability*, *preferences for a specific mode,* and *sensitivities to time and cost*. Examples of attitudes about toothpaste are the *importance of health benefits*, *cosmetic benefits*, and *price*.

In this framework, as in traditional random utility models, the individual's *preferences* are assumed to be latent variables. Preferences represent the desirability of alternative choices. These preferences are translated to decisions via a decision-making *process*. The process by which one makes a decision may vary across different decision problems or tasks, and is impacted by type of task, context, and socioeconomic factors (Gärling and Friman, 1998). Frequently, choice models assume a utility maximization decision process (as in the case studies reviewed later). However, numerous other decision processes may be appropriate given the context, for example habitual, dominant attribute, or a series of decisions each with a different decision-making process. Various types of decision processes can be incorporated into this framework.

The Measurement Relationships

The actual market behavior or revealed preference (RP) and the preferences elicited in stated preference (SP) experiments are *manifestations* of the underlying preferences, and therefore serve as indicators.[24] Similarly, there may also be available indicators for attitudes and perceptions such as responses to attitudinal and perceptual questions in surveys. For example, one could use rankings of the importance of attributes or levels of satisfaction on a semantic scale. As stated earlier, indicators are helpful in model identification and increase the efficiency of the estimated choice model parameters.

---

[24] A method for combining revealed and stated preferences is covered in Chapter 4.

Benefits of the Framework

The integrated choice and latent variable modeling framework allows us to explicitly model the cognitive processes enclosed by the dashed lines in Figure 3-5. Incorporating such latent constructs in choice models requires a hypothesis of the type and the role of the latent variables, as well as indicators of the latent variables (i.e., data).

The simple framework shown in Figure 3-5 is a bit deceiving. *Attitudes* can in fact be any latent characteristic of a decision-maker and thus incorporate concepts such as memory, awareness, tastes, goals, etc. Attitudes can be specified to have a causal relationship with other attitudes and perceptions, and vice-versa. Temporal variables can also be introduced in the specification, and different *processes* by which people make decisions could be included, such as those described in the section above. There is still a tremendous gap between descriptive behavioral theory and the ability of statistical models to reflect these behavioral hypotheses. Examining the choice process within this framework of latent characteristics and perceptions opens the door in terms of the types of behavioral complexities we can hope to capture, and can work to close the gap between these fields.

As with all statistical models, the consequences of mis-specification can be severe. Measurement error and/or exclusion of important explanatory variables in a choice model may result in inconsistent estimates of all parameters. As with an observable explanatory variable, excluding an important attitude or perception will also result in inconsistent estimates. The severity depends highly on the model at hand and the particular specification error, and it is not possible to make generalizations. Before applying the integrated choice and latent variable methodology, the decision process of the choice of interest must also be considered. For more information on behavioral decision theory, see Engel, Blackwell and Miniard (1995) and Olson (1993) for general reference, Gärling, Laitila and Westin (1998) for discussion of behavior regarding activity and transportation decisions, as well as the other references listed in the "Supporting Research" section of this chapter.

Methodology

Herein we develop a general methodology for the incorporation of latent variables as explanatory factors in discrete choice models, so that we can capture the behavioral framework represented by Figure 3-5. The resulting methodology is an integration of latent variable models, which aim to operationalize and quantify unobservable concepts, with discrete choice methods. The methodology incorporates indicators of the latent variables provided by responses to survey questions to aid in estimating the model. A simultaneous estimator is used, which results in latent variables that provide the best fit to both the choice and the latent variables indicators.

Notation

The following notation, corresponding to choice model notation, is used:

X_n observed variables, including:

S_n characteristics of individual n,
Z_{in} attributes of alternative i and individual n.

X_n^* latent (unobservable) variables, including:
S_n^* latent characteristics of individual n,
Z_{in}^* latent attributes of alternative i as perceived by individual n.

I_n indicators of X_n^*.
(For example, responses to survey questions related to attitudes, perceptions, etc.)
I_{S_n} indicators of S_n^*,
$I_{Z_{in}}$ indicators of Z_{in}^*.

U_{in} utility of alternative i for individual n.
U_n vector of utilities.

y_{in} choice indicator; equal to 1 if alternative i is chosen by individual n and 0 otherwise
y_n vector of choice indicators.

a, b, l unknown parameters.

w, e, u random disturbance terms.

Σ, s covariances of random disturbance terms.

D distribution function.

f standard normal probability density function.

Φ standard normal cumulative distribution function.

Framework and Definitions

The integrated modeling framework, shown in Figure 3-6, consists of two components, a choice model and a latent variable model.

As with any random utility choice model, the individual's utility U_n for each alternative is assumed to be a latent variable, and the observable choices y_n are *manifestations* of the underlying utility. Such observable variables that are manifestations of latent constructs are called *indicators*. A dashed arrow representing a *measurement equation* links the unobservable U_n to its observable indicator y_n. Solid arrows representing *structural equations* (i.e., the cause-and-effect relationships that govern the decision making process) link the observable and latent variables (X_n, X_n^*) to the utility U_n.

It is possible to identify a choice model with limited latent variables using only observed choices and no additional indicators (see, for example, Elrod, 1998). However, it is quite likely that the information content from the choice indicators will not be sufficient to empirically identify the effects of individual-specific latent variables. Therefore, indicators of the latent variables are used for identification, and are introduced in the form of a latent variable model.

The top portion of Figure 3-6 is a latent variable model. Latent variable models are used when we have available indicators for the latent variables X_n^*. Indicators could be responses to survey questions regarding, for example, the level of satisfaction with, or importance of, attributes. The figure depicts such indicators I_n as manifestations of the underlying latent variable X_n^*, and the associated measurement equation is represented by a dashed arrow. A structural relationship links the observable causal variables X_n (and potentially other latent causal variables X_n^*) to the latent variable X_n^*.

The integrated choice and latent variable model explicitly models the latent variables that influence the choice process. Structural equations relating the observable explanatory variables X_n to the latent variables X_n^* model the behavioral process by which the latent variables are formed. While the latent constructs are not observable, their effects on indicators are observable. The indicators allow identification of the latent constructs. They also contain information and thus potentially provide for increased efficiency in model estimation. Note that the indicators do not have a causal relationship that influences the behavior. That is, the arrow goes *from* the latent variable *to* the indicator, and the indicators are only used to aid in measuring the underlying causal relationships (the solid arrows). Because the indicators are not part of the causal relationships, they are typically used only in the model estimation stage and not in model application.

General Specification of the Model

As described above, the integrated model is composed of two parts: a discrete choice model and a latent variable model. Each part consists of one or more structural equations and one or more measurement equations. Specification of these equations and the likelihood function follow.

Structural Equations

For the latent variable model, we need the distribution of the latent variables given the observed variables, $f_1(X_n^* \mid X_n; l, \Sigma_w)$. For example:

$$X_n^* = h(X_n; l) + w_n \quad \text{and} \quad w_n \sim D(0, \Sigma_w). \qquad [3\text{-}1]$$

This results in one equation for each latent variable.

For the choice model, we need the distribution of the utilities, $f_2(U_n \mid X_n, X_n^*; b, \Sigma_e)$. For example:

$$U_n = V(X_n, X_n^*; b) + e_n \quad \text{and} \quad e_n \sim D(0, \Sigma_e). \qquad [3\text{-}2]$$

Note that the random utility is decomposed into systematic utility and a random disturbance, and the systematic utility is a function of both observable and latent variables.

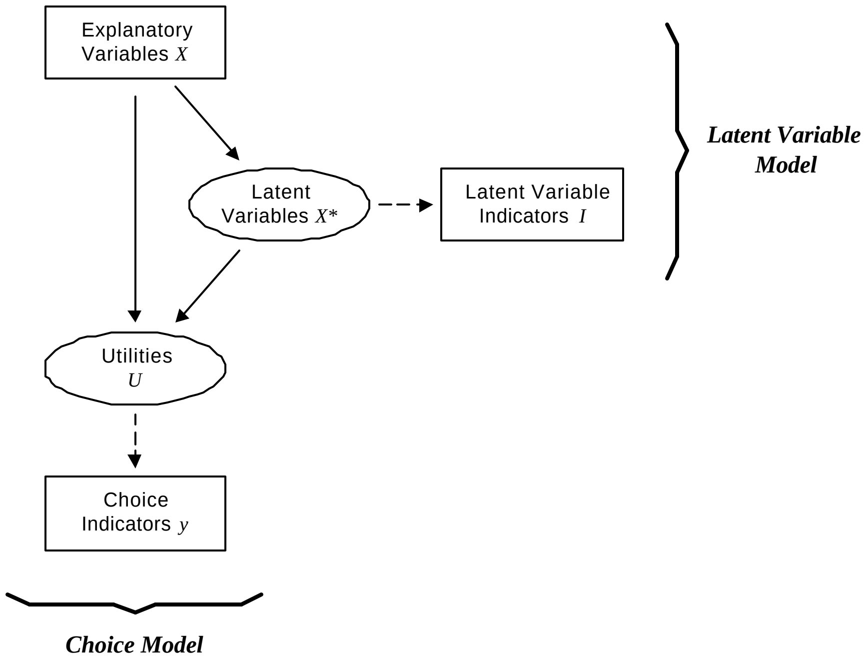


Figure 3-6: Integrated Choice and Latent Variable Model

Measurement Equations

For the latent variable model, we need the distribution of the indicators conditional on the values of the latent variables, $f_3(I_n \mid X_n, X_n^*; \alpha, \Sigma_u)$. For example:

$$I_n = m(X_n, X_n^*; \alpha) + u_n \quad \text{and} \quad u_n \sim D(0, \Sigma_u). \qquad [3\text{-}3]$$

This results in one equation for each indicator (i.e., each survey question). These measurement equations usually contain only the latent variables on the right-hand-side. However, they may also contain individual characteristics or any other variable determined within the model system such as the choice indicator. In principle, such parameterizations can be allowed to capture systematic response biases when the individual is providing indicators. For example, in a brand choice model with latent product quality (Z_n^*), one may include the indicator y_{in} for the chosen brand, for example, $I_{rn} = \alpha_{1r} Z_{in}^* + \alpha_{2r} y_{in} + u_{rn}$, where I_{rn} is an indicator of the perceived quality of alternative i. This would capture any exaggerated responses in reporting the perceived quality of the chosen brand, perhaps caused by justification bias.

For the choice model, we need to express the choice as a function of the utilities. For example, assuming utility maximization:

$$y_{in} = \begin{cases} 1, & if\ U_{in} = \max_j \{U_{jn}\} \\ 0, & otherwise \end{cases}.$$

[3-4]

Note that $h(\cdot),\ V(\cdot),\ and\ m(\cdot)$ are functions, which are currently not defined. Typically, as in the case studies reviewed later, the functions are specified to be linear in the parameters, but this is not necessary. Also note that the distribution of the error terms must be specified, leading to additional unknown parameters (the covariances, Σ). The covariances often include numerous restrictions and normalizations for model simplification and identification.

Integrated Model

The integrated model consists of Equations [3-1] to [3-4]. Equations [3-1] and [3-3] comprise the latent variable model, and Equations [3-2] and [3-4] comprise the choice model. From Equations [3-2] and [3-4] and an assumption about the distribution of the disturbance, e_n we derive $P(y_n | X_n, X_n^*; b, \Sigma_e)$, the choice probability conditional on both observable and latent explanatory variables.

Likelihood Function

We use maximum likelihood techniques to estimate the unknown parameters. The most intuitive way to create the likelihood function for the integrated model is to start with the likelihood of a choice model without latent variables:

$$P(y_n | X_n; b, \Sigma_e)\ . \qquad [3\text{-}5]$$

The choice model can be any number of forms, for example, logit, nested logit, probit, ordinal probit, logit kernel, etc., and can include the combination of different choice indicators such as stated and revealed preferences.

Now we add the latent variables to the choice model. Once we hypothesize an unknown latent construct, X_n^*, its associated distribution, and independent error components $(w_n,\ e_n)$, the likelihood function is then the integral of the choice model over the distribution of the latent constructs:

$$P(y_n | X_n; b, l, \Sigma_w, \Sigma_h) = \int_{X^*} P(y_n | X_n, X_n^*; b, \Sigma_e) f_1(X^* | X_n; l, \Sigma_w) dX^*\ . \qquad [3\text{-}6]$$

We introduce indicators to both improve the accuracy of estimates of the structural parameters as well as to allow for their identification. Assuming the error components (w_n, e_n, u_n) are independent, the joint probability of the observable variables y_n and I_n, conditional on the exogenous variables X_n, is:

$$f_4(y_n, I_n | X_n; a, b, l, \Sigma_e, \Sigma_u, \Sigma_w) = \qquad [3\text{-}7]$$

$$\int_{X^*} P(y_n | X_n, X^*; b, \Sigma_e) f_3(I_n | X_n, X^*; a, \Sigma_u) f_1(X^* | X_n; l, \Sigma_w) dX^*\ .$$

Note that the first term of the integrand corresponds to the choice model, the second term corresponds to the measurement equation from the latent variable model, and the third term corresponds to the structural equation from the latent variable model. The latent variable is only known to its distribution, and so the joint probability of y_n, I_n, and X_n^* is integrated over the vector of latent constructs X_n^*.

Functional Forms

The forms of the variables (for example, discrete or continuous) and assumptions about the disturbances of the measurement and structural equations determine the functional forms in the likelihood equation. Frequently we assume linear in the parameter functional forms, and disturbances that have normal (or extreme value for the choice model) distributions.

The choice model portion of the likelihood function is a standard choice model, except that the utility is a function of latent constructs. The form of the probability function is derived from Equations [3-2] and [3-4] and an assumption about the distribution of the disturbance, e_n. For example, for a choice of alternative i:

$U_{in} = V_{in} + e_{in}$ and $V_{in} = V_{in}(X_n, X_n^*; b)$, $i \in C_n$, C_n is the choice set for individual.

$$P(y_{in} = 1 | X_n, X_n^*; b, \Sigma_e) = P(U_{in} \geq U_{jn}, \forall j \in C_n)$$

$$= P(V_{in} + e_{in} \geq V_{jn} + e_{jn}, \forall j \in C_n)$$

$$= P(e_{jn} - e_{in} \leq V_{in} - V_{jn}, \forall j \in C_n) .$$

If the disturbances, e_n, are i.i.d standard Gumbel, then:

$$P(y_{in} = 1 | X_n, X_n^*; b) = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}} . \qquad \text{[Logit Model]}$$

Or, in a binary choice situation with normally distributed disturbances:

$$P(y_{in} = 1 | X_n, X_n^*; b) = \Phi(V_{in} - V_{jn}) , \qquad \text{[Binary Probit Model]}$$

where Φ is the standard normal cumulative distribution function

The choice model can take on other forms. For example, ordinal categorical choice indicators would result in either ordinal probit or ordinal logistic form (for example, see Case Study 3), or the logit kernel model presented in Chapter 2 can be used.

The form of the distribution of the latent variables is derived from Equation [3-1]; the form of the distribution of the indicators is derived from Equation [3-3]. The disturbances of the structural and measurement equations of the latent variable model are often assumed to be normally and independently distributed. Thus the latent variables are assumed to be orthogonal, i.e., the indicators are assumed to be conditionally (on X_n^* and X_n) independent. In this case, the resulting densities are:

$$f_1(X_n^* | X_n; l, s_w) = \prod_{l=1}^{L} \frac{1}{s_{w_l}} f\left(\frac{X_{ln}^* - h(X_n; l_l)}{s_{w_l}} \right) ,$$

$$f_3(I_n \mid X_n, X_n^*; a, s_u) = \prod_{r=1}^{R} \frac{1}{s_{u_r}} f\left(\frac{I_{rn} - m(X_n, X_n^*; a_r)}{s_{u_r}}\right),$$

where: f is the standard normal density function;

s_{u_r}, s_{w_l} are the standard deviations of the error terms of u_{rn} and w_{ln}, respectively;

R is the number of indicators; and

L is the number of latent variables.

It is trivial to remove the orthogonality assumption for the latent variables by specifying a full covariance structure for w_n (and by estimating the Cholesky decomposition of this matrix).

Both the indicators and the latent variables may be either discrete or continuous. See Gopinath (1995) and Ben-Akiva and Boccara (1995) for details on the specification and estimation of models with various combinations of discrete and continuous indicators and latent constructs. The case of discrete latent variables (i.e., latent class models) is covered in Chapter 4.

Theoretical Analysis

The methodology presented here improves upon the techniques described by Figure 3-1 through Figure 3-4.

Figure 3-1 - Omitting important latent variables may lead to mis-specification and inconsistent estimates of all parameters.

Figure 3-2 - A priori, we reject the use of the indicators directly in the choice model – they are not causal, they are highly dependent on the phrasing of the survey question, there can be multicollinearity issues, and they are not available for forecasting.

Figure 3-3a - The two-stage sequential approach without integration leads to measurement errors and results in inconsistent estimates.

Figure 3-3b - The two-stage sequential approach with integration results in consistent, but inefficient estimates. Furthermore, note that since the choice model involves an integral over the latent variable, a canned estimation procedure cannot be used. Therefore, there is no significant advantage to estimating the model sequentially.

Figure 3-4 - The choice and latent variable model without indicators is restrictive in that the latent variables are alternative-specific and cannot vary among individuals.

In summary, the approach we present is theoretically superior: it is a generalization of Figure 3-1 and Figure 3-4 (so cannot be inferior) and it is statistically superior to sequential methods represented by Figure 3-3. How much better is the methodology in a practical sense? The answer will vary based on the model and application at hand: in some cases it will not make a difference and, presumably, there are cases in which the difference will be substantial.

Identification

As with all latent variable models, identification is certainly an issue in these integrated choice and latent variable models. While identification has been thoroughly examined for special cases of the integrated framework presented here (see, e.g, Elrod 1988 and Keane 1997), necessary and sufficient conditions for the general integrated model have not been developed. Identification of the integrated models needs to be analyzed on a case-by-case basis.

In general, all of the identification rules that apply to a traditional latent variable model are applicable to the latent variable model portion of the integrated model. See Bollen (1989) for a detailed discussion of these rules. Similarly, the normalizations and restrictions that apply to a standard choice model would also apply here. See Ben-Akiva and Lerman (1985) for further information.

For the integrated model, a sufficient, but not necessary, condition for identification can be obtained by extending the *Two-step Rule* used for latent variable models to a *Three-step Rule* for the integrated model:

1. Confirm that the measurement equations for the latent variable model are identified (using, for example, standard identification rules for factor analysis models).
2. Confirm that, given the latent variables, the structural equations of the latent variable model are identified (using, for example, standard rules for a system of simultaneous equations).
3. Confirm that, given the latent variables, the choice model is identified (using, for example, standard rules for a discrete choice model).

An ad-hoc method for checking identification is to conduct Monte Carlo experiments by generating synthetic data from the specified model structure (with given parameter values), and then attempt to reproduce the parameters using the maximum likelihood estimator. If the parameters cannot be reproduced to some degree of accuracy, then this is an indication that the model is not identified.

Another useful heuristic is to use the Hessian of the log-likelihood function to check for local identification. If the model is locally identified at a particular point, then the Hessian will be positive definite at this point. The inverse Hessian is usually computed at the solution point of the maximum likelihood estimator to generate estimates of the standard errors of estimated parameters, and so in this case the test is performed automatically. (See Chapter 4 for more discussion.)

Estimation

Maximum likelihood techniques are used to estimate the unknown parameters of the integrated model. The model estimation process maximizes the logarithm of the sample likelihood function over the unknown parameters:

$$\max_{a,b,l,\Sigma} \sum_{n=1}^{N} \ln f_4\left(y_n, I_n \mid X_n; a, b, l, \Sigma\right). \qquad [3\text{-}8]$$

The likelihood function includes complex multi-dimensional integrals, with dimensionality equal to that of the integral of the underlying choice model plus the number of latent variables. There are three basic ways of estimating the model: a sequential numerical approach, a simultaneous numerical approach, and a simulation approach.

The sequential estimation method involves first estimating the latent variable model (Equations [3-1] and [3-3]) using standard latent variable estimators. The second step is to use fitted latent variables *and their distributions* to estimate the choice model, in which the choice probability is integrated over the distribution of the latent variables.[25] The two step estimation method results in consistent, but inefficient estimates. See McFadden (1986), Train et al. (1986), and Morikawa et al. (1996) for more details on the sequential approach.

An important point is that a sequential estimation procedure that treats the fitted latent variables as non-stochastic variables in the utility function introduces measurement error and results in inconsistent estimates of the parameters. If the variance of the latent variable's random error (w) is small, then increasing the sample size may sufficiently reduce the measurement error and result in acceptable parameter estimates. Increasing the sample size results in a more precise estimate of the expected value of the latent variable, and a small variance means that an individual's true value of the latent variable will not be too far off from the expected value. Train et al. (1986) found that for a particular model (choice of electricity rate schedule) the impact of the inconsistency on parameter estimates was negligible using a sample of 3,000 observations. However, this result cannot be generalized; the required size of the dataset is highly dependent on the model specification, and it requires that the variance of the latent variable's error (w) be sufficiently small. Note that the sample size has no effect on the variance of w. In other words, the measurement errors in the fitted latent variables do not vanish as the sample size becomes very large. Therefore, without running tests on the degree of inconsistency, it is a questionable practice to estimate these integrated choice and latent variable models by chaining a canned latent variable model software package with a canned choice model package. Performing these tests requires integration of the choice model.

The inconsistency issue already makes application of the sequential estimation approach quite complex, and it produces inefficient estimates. Alternatively, a fully efficient estimator can be obtained by jointly estimating Equations [3-1] through [3-4]. This involves programming the joint likelihood function (Equation [3-8]) directly in a flexible estimation package (for example, *Gauss*), which, ideally, has built in numerical integration procedures. This is the method that is used in the second and third case studies reviewed in this chapter.

The dimensionalities of the likelihoods in all three of the reviewed case studies are such that numerical integration is feasible and preferred. However, as the number of latent variables increases (and therefore the dimension of the integral increases), numerical integration methods quickly become infeasible and simulation methods must be employed. Typical estimation approaches used are Method of Simulated

[25] Note that technically this distribution should also include the estimation error from the parameter estimates.

Moments or Maximum Simulated Likelihood Estimation, which employ random draws of the latent variables from their probability distributions. For illustration purposes, consider the use of maximum simulated likelihood for the model that we later review as Case Study 1. This is a binary choice (probit) model with 2 latent variables (assumed to be orthogonal) and six indicators (see the Case Study for further details). The likelihood function is as follows:

$$f_4(y_n, I_n \mid X_n; a, b, g, \Sigma) = \int\int_{Z^*} \Phi\{y_n(X_n b_1 + Z^* b_2)\} *$$

$$\prod_{r=1}^{6} \frac{1}{s_{u_r}} f\left[\frac{I_{rn} - Z^* a_r}{s_{u_r}}\right] * \prod_{l=1}^{2} \frac{1}{s_{w_l}} f\left[\frac{Z_l^* - X_n l_l}{s_{w_l}}\right] dZ^* .$$

Note that since this is only a double integral, it is actually more efficient to estimate the model using numerical integration (as in the case studies that are reviewed later). However, the model serves well for illustration purposes.

Typically, the random draws are taken from a standard multivariate normal distribution (i.e., $\sim N(0, I)$) distribution, so we re-write the likelihood with standard normal disturbance terms for the latent variable structural equation as follows:

$$Z_{ln}^* = X_n l_l + w_{ln} \ , \ \ l = 1,2 \ , \ \ w_n \sim N(0, \Sigma_w \ \ diagonal) \ ,$$

$$w_{ln} = s_{w_l} \tilde{w}_{ln} \text{, where } \tilde{w}_{ln} \sim N(0,1) .$$

The likelihood is then written as:

$$f_4(y_n, I_n \mid X_n; a, b, \ l, \Sigma) = \int\int_{\tilde{h}} \Phi\{y_n(X_n b_1 + (X_n l_1 + s_{w_1}\tilde{w}_1) b_{12} + (X_n l_2 + s_{w_2}\tilde{w}_2) b_{22})\} *$$

$$\prod_{r=1}^{6} \frac{1}{s_{u_r}} f\left[\frac{I_{rn} - (X_n l_1 + s_{w_1}\tilde{w}_1) a_{1r} - (X_n l_2 + s_{w_2}\tilde{w}_2) a_{2r}}{s_{u_r}}\right] * \prod_{l=1}^{2} f\left(\tilde{w}_l\right) d\tilde{w} \ .$$

To simulate the likelihood, we take $\mathbb{D}$ random draws from the distributions of $\tilde{w}_1$ and $\tilde{w}_2$ for each observation in the sample, denoted $\tilde{w}_{1n}^d$ and $\tilde{w}_{2n}^d$, *d=1,…,* $\mathbb{D}$. The following is then an unbiased simulator for $f_4(y_n, I_n \mid X_n; a, b, l, \Sigma)$:

$$\hat{f}_4(y_n, I_n \mid X_n; a, b, \ l, \Sigma) =$$

$$\frac{1}{\mathbb{D}} \sum_{d=1}^{\mathbb{D}} \Bigg\{ \ \Phi\{y_n(X_n b_1 + (X_n l_1 + s_{w_1}\tilde{w}_{1n}^d) b_{12} + (X_n l_2 + s_{w_2}\tilde{w}_{2n}^d) b_{22})\}$$

$$* \prod_{r=1}^{6} \frac{1}{s_{u_r}} f\left[\frac{I_{rn} - (X_n l_1 + s_{w_1}\tilde{w}_{1n}^d) a_{1r} - (X_n l_2 + s_{w_2}\tilde{w}_{2n}^d) a_{2r}}{s_{u_r}}\right] \Bigg\} .$$

The parameters are estimated by maximizing the simulated likelihood over the unknown parameters:

$$\max_{a,b,l,\Sigma} \sum_{n=1}^{N} \ln \hat{f}_4\left(y_n, I_n \mid X_n; a, b, l, \Sigma\right) .$$

Note that, by Jensen's Inequality, $(\ln \hat{f}_4)$ is a biased estimator of $(\ln f_4)$ though consistent by the Slutsky theorem. When a small number of draws is employed, this results in a non-negligible bias in the parameter estimates. Therefore, one has to verify that a sufficient number of draws is used to reduce this bias. This is usually done by estimating the model using various number of draws, and showing empirically that the parameter estimates are stable over a certain number of draws. This issue was discussed more thoroughly in Chapter 2.

For more information on simulation methods for estimating discrete choice models, see McFadden (1989) and Gourieroux and Monfort (1996).

Model Application

The measurement equations are used in estimation to provide identification of the latent constructs and further precision in the parameters estimates for the structural equations. For forecasting, we are interested in predicting the probability of the choice indicator, $P(y_n \mid X_n; a, b, l, \Sigma)$. Furthermore, we do not have forecasts of the indicators, I. Therefore, the likelihood (Equation [3-7]) must be integrated over the indicators. This integration trivially leads to the following model structure, which is what is used for *application*:

$$P(y_n \mid X_n; a, b, l, \Sigma) = \int_{X^*} P(y_n \mid X_n, X_n^*; b, \Sigma_e) f_1(X^* \mid X_n; l, \Sigma_w) dX^* . \qquad [3\text{-}9]$$

Once the model is estimated, Equation [3-9] can be used for forecasting and there is no need for latent variable measurement models or the indicators. Typically, the latent variable structural model is substituted into Equation [3-9], and the function is then simply a choice model integrated over the distribution of the latent variable disturbances, w.

Reviewed Case Studies

The unique features of the integrated choice modeling framework are demonstrated by reviewing three case studies from the literature. For each case study, the original source, the problem context, a problem-specific modeling framework, survey questions, model equations, and results are presented. The models from the original sources were re-framed (and in some cases simplified) using the terminology, notation, and diagram conventions (including the creation of the full-path diagrams) used in this chapter.

The Role of the Case Studies

These case studies have been assembled from a decade of research investigating the incorporation of attitudes and perceptions in choice modeling. The review of the case studies provide conceptual examples of model frameworks, along with some specific equations, estimation results, and comparison of these models with standard choice models. The aim is to show that the methodology is practical, and to provide concrete examples. The reviewed case studies emphasize the general nature of the approach by providing likelihood functions for a variety of model structures, including the use of both SP and RP data, the introduction of an agent effect, and the use of logit, probit, and ordinal probit.

Model Estimation

The dimensionalities of the likelihoods in each of the three case studies were small enough such that numerical integration was feasible and preferred over simultaneous estimation techniques. Therefore, numerical integration was used in all three studies. The first reviewed case study was estimated sequentially (accounting for the distribution of the latent variable), resulting in consistent, inefficient estimates of the parameters. In the second and third reviewed case studies, the latent variable and choice models were estimated jointly, resulting in consistent, efficient estimates. Identification was determined via application of the *Three-step Rule* as described earlier, as well as using the inverse Hessian to check for local identification at the solution point.

Additional References

Applications of the integrated approach can be found in Boccara (1989), Morikawa (1989), Gopinath (1995), Bernardino (1996), Börsch-Supan et al. (1996), Morikawa et al. (1996), and Polydoropoulou (1997). A joint choice and latent variable is also presented in Chapter 4.

Case Study 1: Mode Choice with Latent Attributes

The first case study (Morikawa, Ben-Akiva, and McFadden, 1996) presents the incorporation of the latent constructs of convenience and comfort in a mode choice model. The model uses data collected in 1987 for the Netherlands Railways to assess factors that influence the choice between rail and car for intercity travel. The data contain revealed choices between rail and auto for intercity trips. In addition to revealed choices, the data also include subjective evaluation of trip attributes for both the chosen and unchosen modes, which were obtained by asking questions such as those shown in Table 3-1. The resulting subjective ratings are used as indicators for latent attributes. It is presumed that relatively few latent variables may underlie the resulting ratings data, and two latent variables, *ride comfort* and *convenience*, were identified through exploratory factor analysis.

Figure 3-7 presents the framework for the mode choice model. The revealed choice is used as an indicator of utility, and the attribute ratings are used as indicators for the two latent variables. Characteristics of the individual and observed attributes of the alternative modes are exogenous explanatory variables. Figure 3-8 provides a full path diagram of the model, noting the relationships between each variable.

Table 3-1: Indicators for Ride Comfort and Convenience

Please rate the following aspects for the auto trip:

	very poor	………………		…very good	
Relaxation during the trip	1	2	3	4	5
Reliability of the arrival time	1	2	3	4	5
Flexibility of choosing departure time	1	2	3	4	5
Ease of traveling with children and/or heavy baggage	1	2	3	4	5
Safety during the trip	1	2	3	4	5
Overall rating of the mode	1		. . .		10

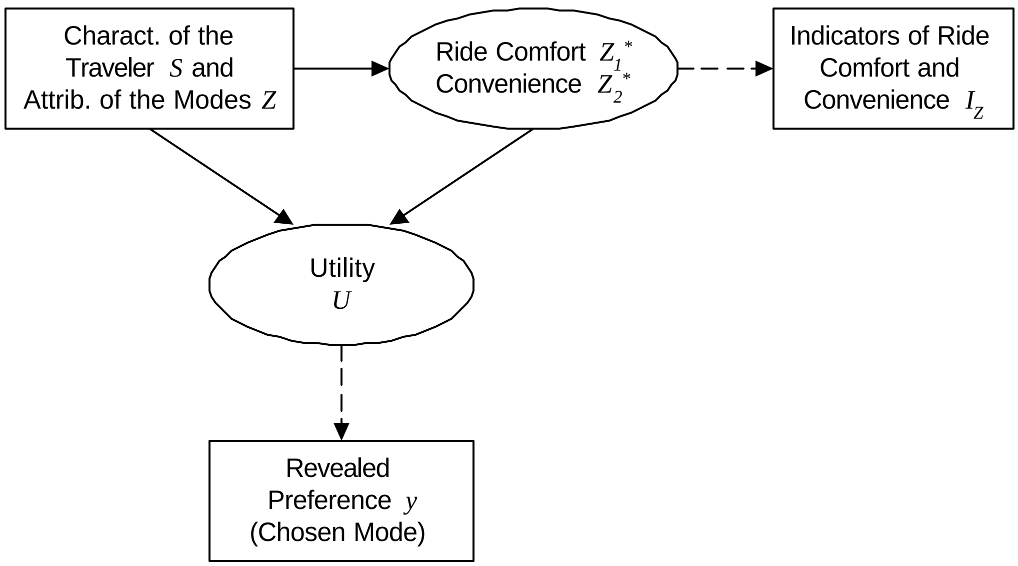


Figure 3-7: Modeling Framework for Mode Choice with Latent Attributes

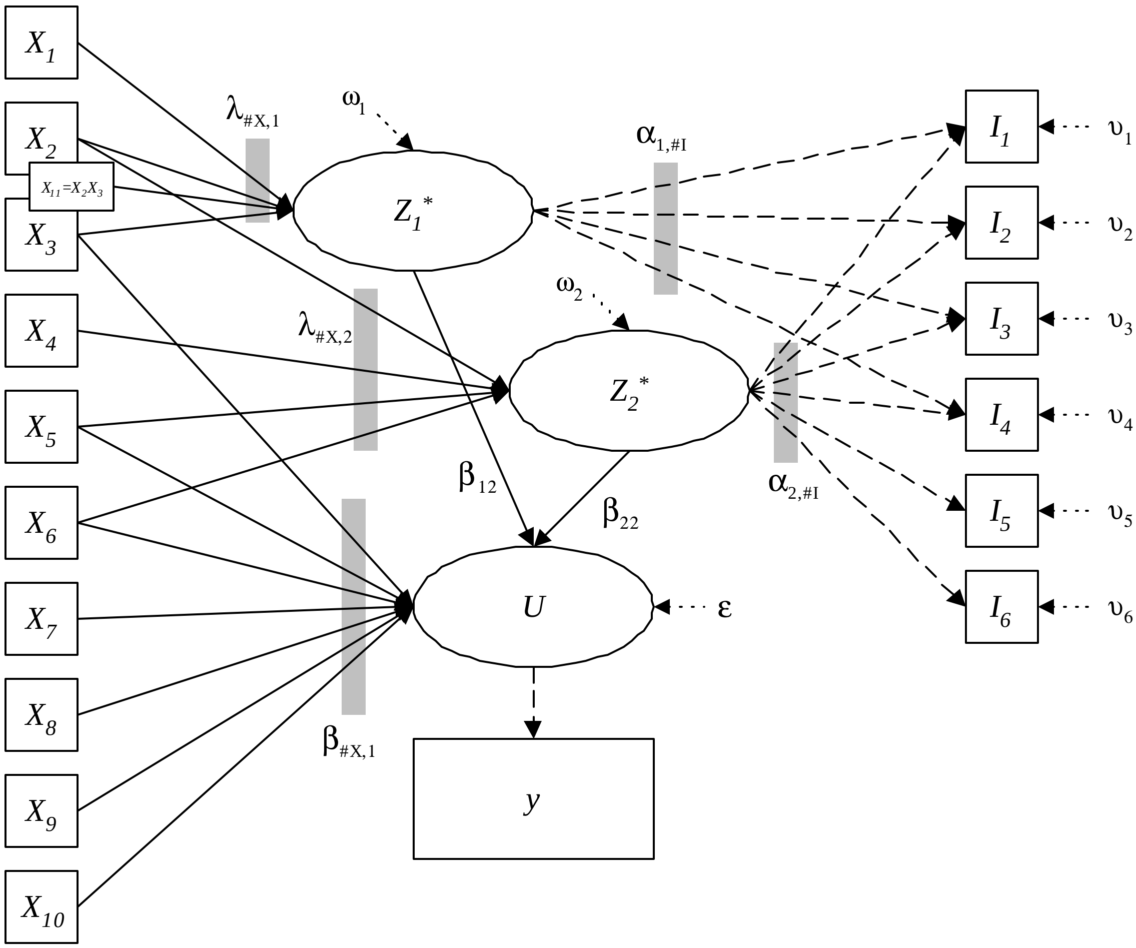


Figure 3-8: Full Path Diagram for Mode Choice Model with Latent Attributes
(SeeTable **3-2** and the model equations for notation.)

The mode choice model with latent attributes is specified by the following equations. All variables, including the latent variables, are measured in terms of the difference between rail and auto. This was done to reduce the dimensionality of the integral (from 4 to 2), and was not necessary for identification of the joint choice/latent variable model.

Structural Model

$$\underset{(1X1)}{Z^*_{ln}} = \underset{(1X10)(10X1)}{X_n\, l_l} + \underset{(1X1)}{w_{ln}}\ ,\quad l = 1,2\ ,\quad w_n \sim N(0, \Sigma_w\ \ diagonal)\,,$$ {2 equations}

$$\underset{(1X1)}{U_n} = \underset{(1X10)(10X1)}{X_n\, b_1} + \underset{(1X2)(2X1)}{Z^*_n\, b_2} + \underset{(1X1)}{e_n}\ ,\ e_n \sim N(0,1)\ .$$ {1 equation}

Measurement Model

$$I_{rn} = Z_n^* a_r + u_{rn} \ , \ \ r = 1,...,6 \ , \ \ u_n \sim N(0, \Sigma_u \ \ diagonal) \ , \qquad \text{\{6 equations\}}$$

(1X1) (1X2)(2X1) (1X1)

$$y_n = \begin{cases} 1, \ if \ U_n > 0 \\ -1, \ if \ U_n \leq 0 \end{cases} \ . \qquad \text{\{1 equation\}}$$

(1X1) (1X1)

Note that the covariances of the error terms in the latent variable structural and measurement model are constrained to be equal to zero (denoted by the "S *diagonal*" notation).

Likelihood function

$$f(y_n, I_n \mid X_n; a, b, l, \Sigma) = \int\int_{Z^*} \Phi\{y_n(X_n b_1 + Z_n^* b_2)\} *$$

$$\prod_{r=1}^{6} \frac{1}{s_{u_r}} f\left[\frac{I_{rn} - Z_n^* a_r}{s_{u_r}}\right] * \prod_{l=1}^{2} \frac{1}{s_{w_l}} f\left[\frac{Z_l^* - X_n l_l}{s_{w_l}}\right] dZ^* \ .$$

Results

The parameters to be estimated include: b (9 parameters estimated), a (8 parameters estimated, 2 parameters constrained to one for identification, 2 parameters constrained to zero based on exploratory factor analysis), $u_n \sim N(0, \Sigma_u \ \ diagonal)$ (8 parameters estimated), and the standard deviations s_u (6 parameters) and s_w (2 parameters), where the covariances of the latent variable equations are restricted to zero. Unless otherwise noted, parameters were set to zero based on statistical tests and a priori hypotheses regarding the behavior. All parameters except the variances are reported.

The results are shown in Table 3-2. Estimation was performed via a sequential estimation procedure that is described in Morikawa et al. (1996). The dataset included 219 observations. The top panel displays the estimation results of two different choice models: the second column is the choice model *without* the latent variables, and the first column is the choice model *with* the latent variables. The integrated choice and latent variable model consists of the choice model with latent variables (the first column of the upper panel) and the latent variable model (displayed in the lower panel of Table 3-2). The table for the latent variable model displays the estimation results of both the structural and measurement equations for each of the two latent variables *comfort* (the first column) and *convenience* (the second column). The latent variable model is made up of many equations: one structural equation for comfort, one structural equation for convenience, and six measurement equations for comfort and convenience.

Both of the latent attributes have significant parameter estimates. Inclusion of the latent attributes identified by the linear structural equation resulted in a large improvement in the goodness-of-fit of

Table 3-2: Estimation Results of Mode Choice Model with Latent Attributes

CHOICE MODEL

		WITH Latent Attributes		WITHOUT Latent Attributes	
	Explanatory Variables	Est. β	t-stat	Est. β	t-stat
X10	Rail constant	0.32	1.00	0.58	2.00
X9	Cost per person	-0.03	-4.10	-0.03	-4.20
X3	Line-haul time	0.08	0.20	-0.41	-1.60
X6	Terminal time	-1.18	-2.60	-1.57	-4.20
X5	Number of transfers	-0.32	-1.70	-0.20	-1.30
X8	Business trip dummy	1.33	3.60	0.94	3.60
X7	Female dummy	0.65	2.60	0.47	2.30
*Z1**	Ride comfort (latent)	0.88	2.70	------	------
*Z2**	Convenience (latent)	1.39	4.10	------	------
	Rho-bar-Squared	0.352		0.242	

LATENT VARIABLE MODEL

	Structural Model (2 equations total, 1 per column)	Comfort *Z1**		Convenience *Z2**	
		Est. $\lambda 1$	t-stat	Est. $\lambda 2$	t-stat
X2	Age >40	-0.23	-1.40	0.41	3.30
X1	First class rail rider	0.29	1.00	------	------
X3	Line haul travel time (rail-auto)	-0.29	-1.30	------	------
X6	Terminal time (rail-auto)	------	------	-0.52	-2.10
X5	Number of transfers by rail	------	------	-0.05	-0.60
X4	Availability of free parking for auto	------	------	0.16	1.60
X11	(Age >40) * (Line haul travel time)	-0.04	-0.10	------	------

	Measurement Model (6 equations total, one per row)	Comfort *Z1**		Convenience *Z2**	
		Est. $\alpha 1$	t-stat	Est. $\alpha 2$	t-stat
I1	Relaxation during trip	1.00	------	0.17	0.80
I2	Reliability of the arrival time	0.77	1.80	1.00	------
I5	Flexibility of choosing departure time	------	------	1.49	4.30
I6	Ease of traveling with children/baggage	------	------	1.16	1.16
I3	Safety during the trip	0.69	3.10	0.33	2.00
I4	Overall rating of the mode	1.64	2.60	2.43	5.90

the discrete choice model. The rho-bar-squared for the model with latent attributes uses a degree-of-freedom correction involving two variables beyond those used in the model without latent variables, and thus this degree of freedom adjustment only accounts for the estimated parameters of the choice model. Note that some of this improvement in fit would probably be captured in the choice model by including in the base choice model the additional variables that are included in latent variable structural model.

While the indicators used for comfort and convenience in this case study are adequate, the structural equations are not particularly strong because of the limited explanatory variables available for comfort and convenience. In general, it can be difficult to find causes for the latent variables. This issue needs to be thoroughly addressed in the data collection phase.

Note that numerous variations on this model are presented in Chapter 4.

Case Study 2: Employees' Adoption of Telecommuting

The second case study (Bernardino, 1996) assesses the potential for the adoption of telecommuting by employees. Figure 3-9 presents the modeling framework. The behavioral hypothesis is that an employee faced with a telecommuting arrangement will assess the impact of the arrangement on lifestyle, work-related costs and income, and then decide whether to adopt telecommuting. Telecommuting is expected to influence lifestyle quality by providing the employee with the benefit of increased flexibility to adjust work schedule, workload, personal needs, and commuting patterns. The perceived impact is expected to vary according to the characteristics of the individual and of the program. Telecommuting is also expected to impact household expenditures, such as utilities, equipment, day care, and transportation. Figure 3-10 provides a full path diagram of the model, noting the relationships between each variable.

The employee's decision to adopt a telecommuting program in a simulated choice experiment is modeled as a function of her/his motivations and constraints, as well as the impacts of the available program on lifestyle quality, work-related costs, and income. Changes in income are included in the telecommuting scenarios, while latent constructs of benefit (i.e., enhancement to lifestyle quality) and cost are estimated. To obtain indicators for benefit, respondents are asked to rate the potential benefits of the telecommuting program on a scale from 1 to 9 as shown in Table 3-3. These responses provide indicators for the latent variable model. The latent cost variable is manifested by the employees' responses to questions about the expected change in home office costs, child and elder care costs, and overall work-related costs as shown in Table 3-4. The employee is assumed to have a utility maximization behavior, and thus will choose to adopt a particular telecommuting option if the expected change in utility is positive. This decision is influenced by the characteristics of the arrangement, the individual's characteristics and situational constraints, and the perceived benefits and costs of the arrangement.

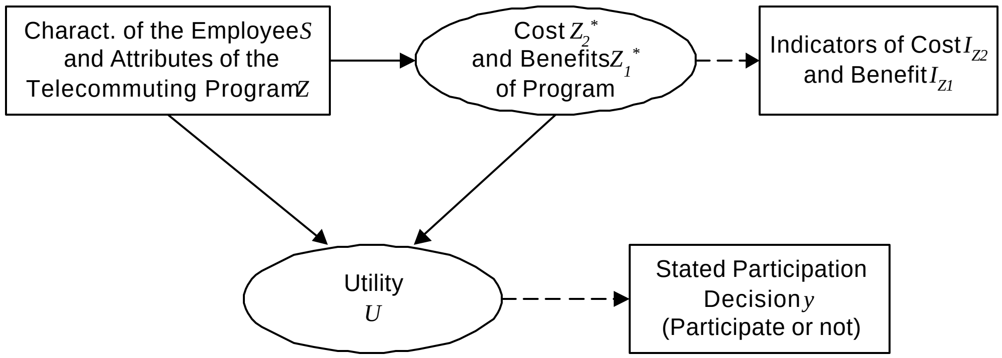


Figure 3-9:
Modeling Framework for Employee's Adoption of Telecommuting

The adoption of telecommuting model is specified by the following equations.

Structural Model

$$Z_{ln}^{*} = X_n l_l + w_{ln} \ , \ \ l = 1,2 \ , \ \ w_n \sim N(0, \Sigma_w \ \ diagonal) \ , \qquad \{2 \text{ equations}\}$$

(1X1) (1X14)(14X1) (1X1)

$$U_n = X_n b_1 + Z_n^{*} b_2 + e_n \ , \ e_n \sim standard\ logistic \ . \qquad \{1 \text{ equation}\}$$

(1X1) (1X14)(14X1) (1X2) (2X1) (1X1)

Measurement Model[26]

$$I_{rn} = Z_n^{*} a_r + u_{rn} \ , \ \ r = 1,...,14 \ , \ \ u_n \sim N(0, \Sigma_u \ \ diagonal) \ , \qquad \{14 \text{ equations}\}$$

(1X1) (1X2)(2X1) (1X1)

$$y_n^{RP} = t, \ \ if \, t_{t\text{-}1}^{RP} < U_n^{RP} \le t_t^{RP} \ \ . \qquad \{1 \text{ equation}\}$$

(1X1) (1X1)

---

[26] Note that the indicators for the cost latent variable were on a 3-point scale and therefore the specified measurement equations are actually discrete equations. We write them as linear here to simplify the presentation. See Bernardino (1996) for the actual discrete equations.

Likelihood Function

$$f(y_n, I_n \mid X_n; a, b, l, \Sigma) = \int\int_{Z^*} \left(\frac{1}{1+\exp^{-(X_n b_1 + Z^* b_2) y_n}} \right) * \prod_{r=1}^{R} \frac{1}{s_{u_r}} f\left[\frac{I_m - Z^* a_r}{s_{u_r}} \right] *$$

$$\prod_{l=1}^{2} \frac{1}{s_{w_l}} f\left[\frac{Z_l^* - X_n l_l}{s_{w_l}} \right] dZ^* \ .$$

Table 3-3: Indicators of Benefit

What type of impact would you expect the telecommuting arrangement to have on:

	extremely negative								extremely positive
Your schedule flexibility	1	2	3	4	5	6	7	8	9
Your productivity	1	2	3	4	5	6	7	8	9
Your autonomy in your job	1	2	3	4	5	6	7	8	9
The productivity of the group you work with	1	2	3	4	5	6	7	8	9
Your family life	1	2	3	4	5	6	7	8	9
Your social life	1	2	3	4	5	6	7	8	9
Your job security	1	2	3	4	5	6	7	8	9
Your opportunity for promotion	1	2	3	4	5	6	7	8	9
Your sense of well being	1	2	3	4	5	6	7	8	9
Your job satisfaction	1	2	3	4	5	6	7	8	9
Your life. overall	1	2	3	4	5	6	7	8	9

Table 3-4: Indicators of Cost

How would you expect the telecommuting arrangement to impact your expenditures on:

home utilities:	( ) decrease	( ) remain the same	( ) increase
child care:	( ) decrease	( ) remain the same	( ) increase
elder care:	( ) decrease	( ) remain the same	( ) increase
overall working costs:	( ) decrease	( ) remain the same	( ) increase

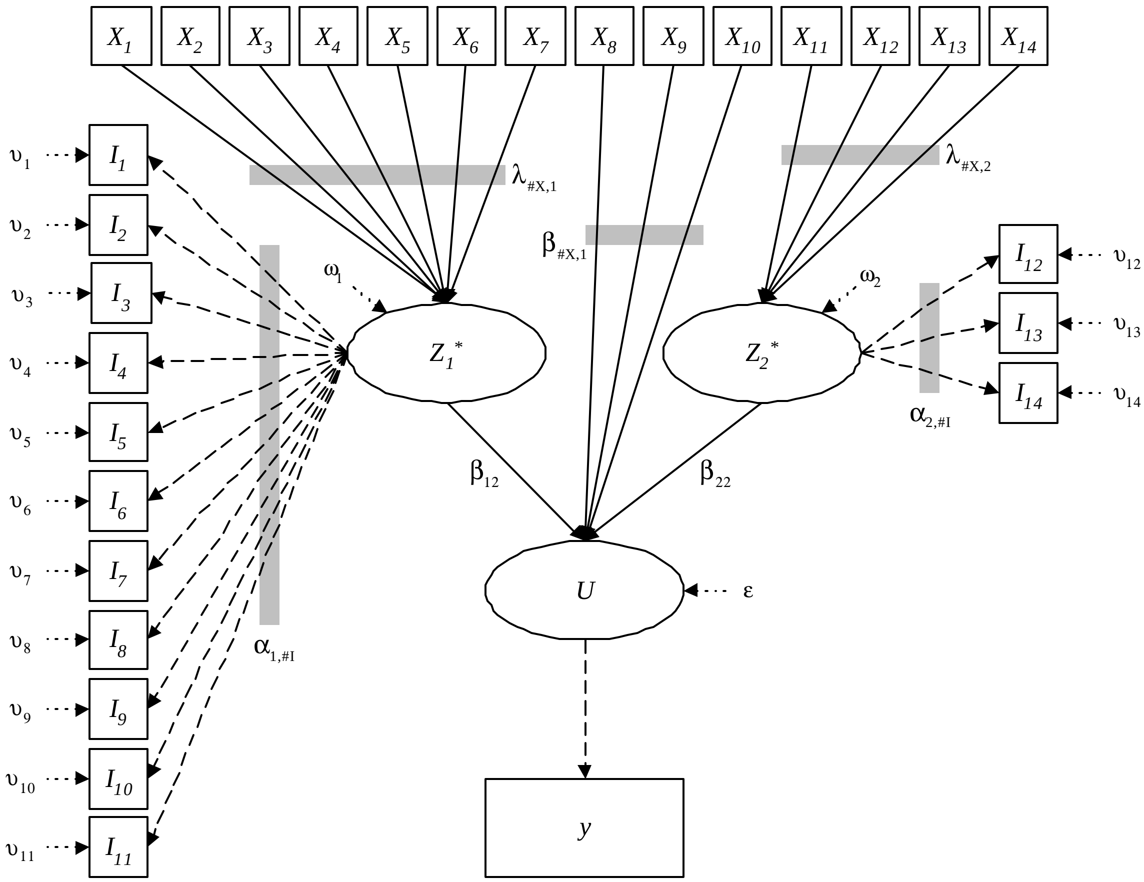


Figure 3-10:
Full Path Diagram for Model of Employee's Adoption of Telecommuting
(See Table 3-5 and the model equations for notation.)

Results

The parameters to be estimated include: b (5 parameters estimated), a (13 parameters estimated, 1 parameter constrained to one for identification), l (11 parameters estimated), and the standard deviations s_u (14 parameters estimated) and s_w (2 parameters: the benefit parameter is estimated, the cost parameter is constrained for identification rather than constraining an a), where the covariances of the latent variable equations are restricted to zero. Unless otherwise noted, parameters were set to zero based on statistical tests and a priori hypotheses about the behavior.

The estimation results are shown in Table 3-5 (estimated variances of the disturbance terms are not reported). The model was estimated using observations from 440 individuals and employed a simultaneous numerical integration estimation procedure. The top panel displays the results of the choice model, which includes the latent explanatory variables benefit and cost. The lower panel displays the results for the latent variable model. The latent variable model consists of many equations: a structural equation for benefit, a structural equation for cost, 11 measurement equations for benefit (one equation per row), and 3 measurement equations for cost (again, one equation per row).

This model of the employee's adoption decision contains more information and allows for a clearer behavioral interpretation than standard choice models. It demonstrates the impact of different telecommuting arrangements on the employee's lifestyle and work-related costs, as a function of the employee's characteristics and situational constraints. The results indicate that females and employees with young children perceive a higher beneficial impact from telecommuting on lifestyle quality than their counterparts. Note that unlike the other two case studies reviewed in this chapter, a survey was conducted that was designed specifically for this model, and, as a result, the structural models are quite strong with solid causal variables. For more information on these models and other models for telecommuting behavior, see Bernardino (1996).

Table 3-5: Estimation Results of a Telecommuting Choice Model with Latent Attributes

CHOICE MODEL

	Explanatory Variables	Est. β	t-stat
X8	Telecommuting specific constant	2.02	8.94
X9	Higher salary to telecommuters (relative to 'same')	0.50	1.12
X10	Lower salary to telecommuters (relative to 'same')	-2.36	-5.78
*Z1**	Benefit (latent variable)	0.99	7.01
*Z2**	Cost (latent variable)	-0.37	-3.12
	Rho-bar-Squared	0.35	

LATENT VARIABLE MODEL

	*Structural Model for Benefits Z1** (1 equation)	Est. $\gamma 1$	t-stat
X1	Min # of telecommuting days/week	-0.15	-6.65
X2	Max # of telecommuting days/week * team structure dummy	0.10	3.02
X3	Max # telecommuting days/week * individual structure dummy	-0.04	-1.99
X4	Telework center telecommuting dummy	-1.02	-14.75
X5	Travel time * female dummy	0.69	7.47
X6	Travel time * male dummy	0.27	3.21
X7	Child under 6 years old in household dummy	0.55	7.46
	Squared multiple correlation for structural equation	0.28	

	*Measurement Model for Benefits Z1** (11 equations)	Est. $\alpha 1$	t-stat
I1	Social life	0.59	11.61
I2	Family life	0.80	18.37
I3	Opportunity for job promotion	0.32	6.19
I4	Job security	0.41	8.15
I5	Schedule flexibility	0.76	14.40
I6	Job autonomy	0.60	12.51
I7	Your Productivity	0.92	20.87
I8	Group productivity	0.61	12.43
I9	Sense of well being	1.04	24.86
I10	Job satisfaction	1.07	24.84
I11	Life overall	1.00	-----

	*Structural Model for Cost Z2** (1 equation)	Est. $\lambda 2$	t-stat
X11	Day care costs proxy	0.39	2.00
X12	Home office utilities proxy	-0.36	-2.70
X13	Equipment costs	0.76	2.50
X14	Weekly transportation costs	0.65	2.91
	Squared multiple correlation for structural equation	0.21	

	*Measurement Model for Cost Z2** (3 equations)	Est. $\alpha 2$	t-stat
I12	Day care costs	0.37	4.78
I13	Home office utilities costs	-0.11	-3.07
I14	Overall working costs	0.50	3.63

Case Study 3: Usage of Traffic Information Systems

The objective of the third case study (Polydoropoulou, 1997) is to estimate the willingness to pay for Advanced Traveler Information Systems. The model uses data collected for the SmarTraveler test market in the Boston area. SmarTraveler is a service that provides real-time, location-specific, multi-modal information to travelers via telephone.

Figure 3-11 shows the framework for the model, which includes a latent variable of satisfaction as an explanatory variable in the usage decision. Travelers' satisfaction ratings of SmarTraveler are used as indicators of the satisfaction latent construct. Table 3-6 shows the survey questions used to obtain ratings of satisfaction. The model assumes that each traveler has an underlying utility for the SmarTraveler service. The utility is a function of the service attributes such as cost and method of payment, as well as the overall satisfaction with the service. Since utility is not directly observable, it is a latent variable, and the responses to the alternate pricing scenarios serve as indicators of utility. Respondents were presented with several pricing scenarios, and then asked what their usage rate (in terms of number of calls per week) or likelihood of subscribing to the service would be under each scenario. Two types of scenarios were presented: a 'measured' pricing structure in which travelers are charged on a per call basis (corresponds to SP1 responses) and a 'flat rate' pricing structure in which travelers pay a monthly subscription fee (corresponds to SP2 responses). Travelers' revealed preference for free service is reflected by the actual usage rate, which serves as an additional indicator of utility. Figure 3-12 provides a full path diagram of the model, noting the relationships between each variable in the model.

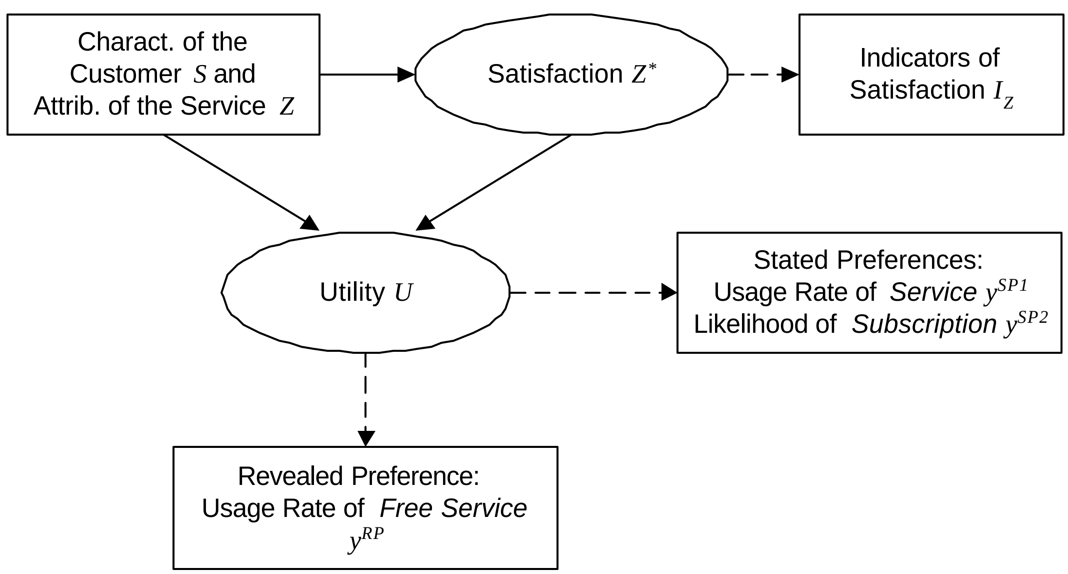


Figure 3-11:
Modeling Framework for Usage of SmarTraveler

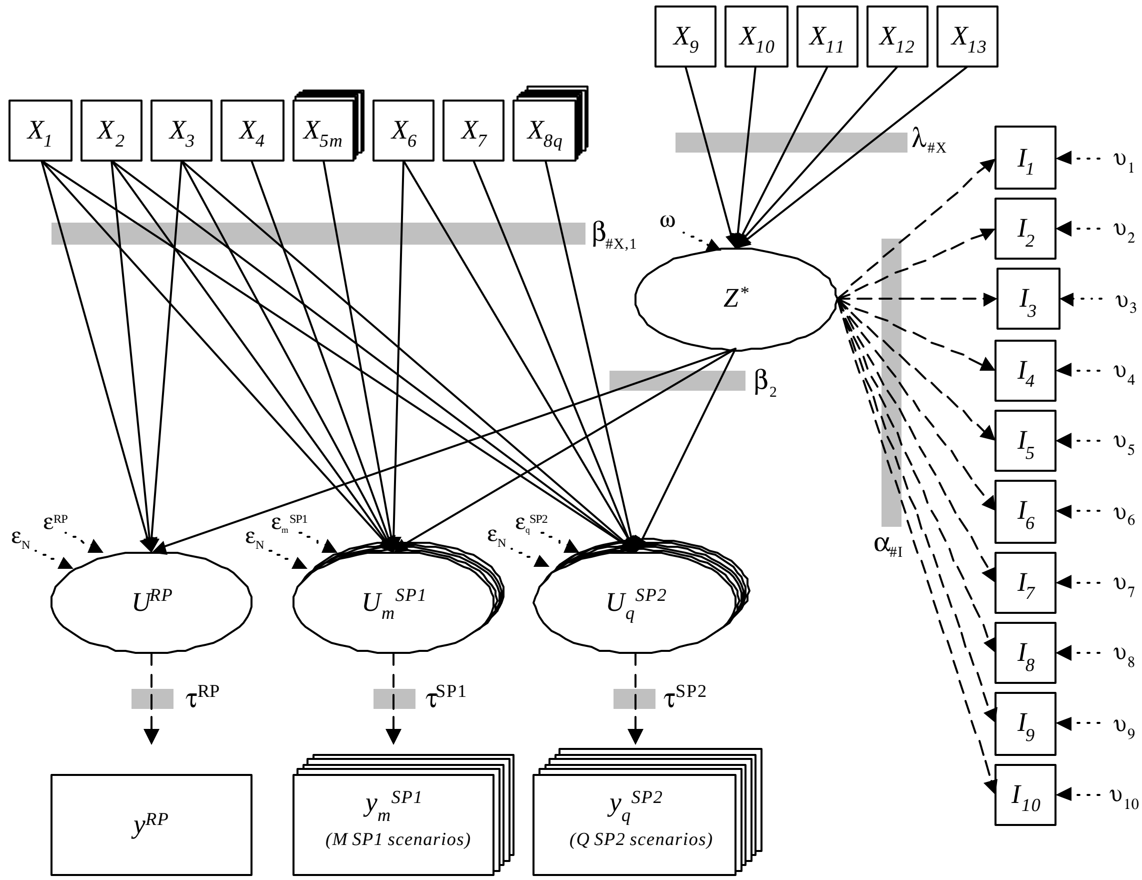


Figure 3-12:
Full Path Diagram for Model of Usage of SmarTraveler
(See Table 3-7 and model equations for notation.)

Table 3-6: Indicators of Satisfaction with SmarTraveler Service

Please rate your level of satisfaction with the following aspects of the existing SmarTraveler service.

	extremely……………………extremely dissatisfied……………..…….satisfied								
Ease of use	1	2	3	4	5	6	7	8	9
Up to the minute information	1	2	3	4	5	6	7	8	9
Availability on demand	1	2	3	4	5	6	7	8	9
Accuracy of information	1	2	3	4	5	6	7	8	9
Level of detail of information	1	2	3	4	5	6	7	8	9
Provision of alternate routes	1	2	3	4	5	6	7	8	9
Hours of operation	1	2	3	4	5	6	7	8	9
Coverage of major routes	1	2	3	4	5	6	7	8	9
Cost of service	1	2	3	4	5	6	7	8	9
Overall satisfaction with service	1	2	3	4	5	6	7	8	9

All of the choice variables are ordinal categorical, and therefore ordinal probit choice models are used. The revealed preference choice (y_n^{RP}) and the stated usage rate (y_n^{SP1}) can take on the following values:

$$y_n = \begin{cases} 1, \text{ if less than 1 call per week} \\ 2, \text{ if 1 to 4 calls per week} \\ 3, \text{ if 5 to 9 calls per week} \\ 4, \text{ if more than 9 calls per week} \end{cases} .$$

The stated likelihood of subscription (y_n^{SP2}) can take on the following values:

$$y_n = \begin{cases} 1, \text{ if very unlikely to subscribe} \\ 2, \text{ if somewhat unlikely to subscribe} \\ 3, \text{ if somewhat likely to subscribe} \\ 4, \text{ if very likely to subscribe} \end{cases} .$$

The following equations specify the model of SmarTraveler usage.

Structural Model

$$\underset{(1X1)}{Z_n^*} = \underset{(1X13)}{X_n^{RP}} \underset{(13X1)}{l} + \underset{(1X1)}{w_n} \ , \ w_n \sim N(0, s_w^2) \ ,$$ {1 equation}

Utility equations: {1+M+Q equations}

$$U_n^{RP} = V_n^{RP} + \tilde{e}_n^{RP} = X_n^{RP} b_1 + Z_n^* b_2 + e_n^N + e_n^{RP}, \qquad e_n^{RP} \sim N(0,1),$$

$$U_{mn}^{SP1} = V_{mn}^{SP1} + \tilde{e}_{mn}^{SP1} = X_{mn}^{SP1} b_1 + Z_n^* b_2 + e_n^N + e_{mn}^{SP1}, \qquad e_{mn}^{SP1} \sim N(0, s_{SP1}^2), \qquad m = 1,..., M,$$

$$U_{qn}^{SP2} = V_{qn}^{SP2} + \tilde{e}_{qn}^{SP2} = X_{qn}^{SP2} b_1 + Z_n^* b_2 + e_n^N + e_{qn}^{SP2}, \qquad e_{qn}^{SP2} \sim N(0, s_{SP2}^2), \qquad q = 1,..., Q,$$

(1X1) (1X13)(13X1) (1X1)(1X1) (1X1) (1X1)

where: m denotes a particular measured rate scenario, and

q denotes a particular flat rate scenario.

The disturbance in the utility equations, $\tilde{e}_n$, are made up of 2 components: a respondent-specific component and a dataset/scenario specific component. The random disturbance characterizing each respondent, e_n^N, is constant for any respondent across pricing scenarios, and captures the correlation among responses from the same individual (an "agent effect"). The assumed distribution for the agent effect is $e_n^N \sim N(0, s_N^2)$.

Measurement Model

$$I_{rn} = Z_n^* a_r + u_{rn}, \quad r = 1,...,10, \quad u_n \sim N(0, \Sigma_u \ diagonal), \qquad \{10 \text{ equations}\}$$

(1X1) (1X1)(1X1) (1X1)

$$y_n^{RP} = t, \ if\, t_{t-1}^{RP} < U_n^{RP} \le t_t^{RP}, \qquad t = 1,\ldots,4,$$

$$y_{mn}^{SP1} = t, \ if\, t_{t-1}^{SP1} < U_{mn}^{SP1} \le t_t^{SP1}, \qquad t = 1,\ldots,4, \qquad m = 1,..., M,$$

$$y_{qn}^{SP2} = t, \ if\, t_{t-1}^{SP2} < U_{qn}^{SP2} \le t_t^{SP2}, \qquad t = 1,\ldots,4, \qquad q = 1,..., Q,$$

t are unknown threshold parameters, with $t_0 = -\infty$, $t_1 = 0$ (for identification), $t_4 = \infty$.

Additional Notation

$$y_{tn}^{RP} = \begin{cases} 1, & if\, y_n^{RP} = t \\ 0, & otherwise \end{cases},$$

$$y_{mtn}^{SP1} = \begin{cases} 1, & if\, y_{mn}^{SP1} = t \\ 0, & otherwise \end{cases}, \text{ and}$$

$$y_{qtn}^{SP2} = \begin{cases} 1, & if\, y_{qn}^{SP2} = t \\ 0, & otherwise \end{cases}.$$

Likelihood Function

$$f(y_n, I_n | \; a, b, l, t, \Sigma) = \int\int \left[\sum_{t=1}^{4} y_{tn}^{RP} \left[\Phi\left(\frac{t_t^{RP} - V_n^{RP} - e^N)}{1} \right) - \Phi\left(\frac{t_{t-1}^{RP} - V_n^{RP} - e^N}{1} \right) \right] \right] *$$

$$\left[\prod_{m=1}^{M} \left(\sum_{t=1}^{4} y_{mtn}^{SP1} \left[\Phi\left(\frac{t_t^{SP1} - V_{mn}^{SP1} - e^N}{s_{SP1}} \right) - \Phi\left(\frac{t_{t-1}^{SP1} - V_{mn}^{SP1} - e^N}{s_{SP1}} \right) \right] \right) \right] *$$

$$\left[\prod_{q=1}^{Q} \left(\sum_{t=1}^{4} y_{qtn}^{SP2} \left[\Phi\left(\frac{t_t^{SP2} - V_{qn}^{SP2} - e^N}{s_{SP2}} \right) - \Phi\left(\frac{t_{t-1}^{SP2} - V_{qn}^{SP2} - e^N}{s_{SP2}} \right) \right] \right) \right] *$$

$$\left[\prod_{r=1}^{10} \frac{1}{s_{u_r}} f\left(\frac{I_{rn} - Z^* a_r}{s_{u_r}} \right) \right] * \left[\frac{1}{s_N} f\left(\frac{e^N}{s_N} \right) \right] \left[\frac{1}{s_w} f\left(\frac{Z^* - X_n^{RP} l}{s_w} \right) \right] dZ^* de_N \; .$$

Results

The parameters to be estimated include: b (9 parameters estimated), a (9 parameters estimated, 1 parameter constrained to 1 for identification), l (5 parameters estimated), the threshold parameters t , and the standard deviations s_u (10 parameters), s_w (1 parameter), s_{SP1} (1 parameter), s_{SP2} (1 parameter), s_N (1 parameter), where s_{RP} was constrained to 1 for identification and the covariances of the latent variable equations are restricted to zero. Unless otherwise noted, parameters were set to zero based on statistical tests and a priori hypotheses about the behavior.

Table 3-7 shows the estimation results for this model (estimated threshold parameters, t , and variances of the error terms are not reported). The model was estimated using observations from 442 individuals, all of whom are SmarTraveler users, and a simultaneous numerical integration estimation procedure. Results of two choice models are presented: one without the satisfaction latent variable (the right column of the top panel) and one that includes the satisfaction latent variable (the left column of the top panel). The integrated choice and latent variable model consists of the choice model with the satisfaction variable and the latent variable model (one structural equation and 10 measurement equations).

The incorporation of satisfaction in the utility of SmarTraveler model significantly improved the goodness of fit of the choice model. Note that some of this improvement in fit would probably be captured in the choice model by including in the base choice model the additional variables that are included in latent variable structural model. The rho-bar-squared for the model with latent attributes uses a degree-of-freedom correction involving one variable (for the satisfaction latent variable) beyond those used in the model without the latent variable, and thus this degree of freedom adjustment only accounts for the estimated parameters of the choice model. See Polydoropoulou (1997) for additional model estimation results for this model, and for additional models of behavior regarding SmarTraveler.

Table 3-7: Estimation Results of ATIS Usage Model with Latent Satisfaction

CHOICE MODEL

	Utility of SmarTraveler Service	WITH the Satisfaction Latent Variable		WITHOUT the Satisfaction Latent Variable	
	Explanatory Variables	Est. β	t-stat	Est. β	t-stat
X6	Constant for actual market behavior	0.94	5.20	0.97	5.90
X4	Constant for measured service	0.56	3.90	0.59	4.30
X7	Constant for flat rate service	0.10	0.70	0.11	0.80
X5	Price per call (cents/10)	-0.31	-15.90	-0.31	-15.80
X8	Subscription fee ($/10)	-1.29	-15.50	-1.27	-16.30
X1	Income: $30,000-$50,000	0.02	0.10	0.15	1.00
X2	Income: $50.001-$75.000	0.32	2.10	0.37	2.60
X3	Income: >$75.000	0.35	2.40	0.22	1.60
*Z**	Satisfaction Latent Variable	0.16	4.50	------	------
	Rho-bar-Squared	0.65		0.49	

LATENT VARIABLE MODEL

	Structural Model (1 equation)	Est. λ	t-stat
X9	Gender (male dummy)	-0.19	-2.40
X10	NYNEX user	-0.86	-10.50
X11	Cellular One user	-1.08	-8.20
X12	Age: 25-45 years	-0.26	-1.60
X13	Age: >45 years	-0.24	-1.40
	Squared multiple correlation for structural model	0.104	

	Measurement Model (10 equations)	Est. α	t-stat	R_{Ir}^2
I1	Ease of use	0.46	7.80	0.15
I2	Up to the minute information	1.26	21.60	0.64
I3	Availability on demand	0.47	8.2	0.18
I4	Accuracy of information	1.19	23.10	0.69
I5	Level of Detail of information	1.10	22.60	0.63
I6	Suggestions of alternative routes	0.75	7.80	0.16
I7	Hours of operation	0.57	7.40	0.13
I8	Coverage of major routes	0.59	12.60	0.25
I9	Cost of service	0.19	5.30	0.06
I10	Overall satisfaction with service	1.00	-----	0.82

Practical Findings from the Case Studies

In the case studies reviewed here, and in our other applications of the methodology, the general findings are that implementation of the integrated choice and latent variable model framework results in:

- Improvements in goodness of fit over choice models without latent variables, or, alternatively, confidence that the simple choice model is adequately specified;
- Latent variables that are statistically significant in the choice model, with correct parameter signs; and
- A more satisfying behavioral representation.

Several practical lessons were learned from our application of the methodology. First, in terms of the measurement equations [3-3], a sufficient number of indicators relevant to the latent variable under consideration, as well as variability among the indicators, are critical success factors. Second, for the structural equations [3-1], it can be difficult to find solid causal variables (X) for the latent variables. In some cases, it is difficult to even conceptually define good causal variables, that is, cases in which there are no good socioeconomic characteristics or observable attributes of the alternatives that sufficiently explain the latent attitudes and/or perceptions. However, frequently it happens that even if one can define good causal variables, these types of data have not been collected and are not included in the dataset. To address both of these issues, it is critical for the successful application of this methodology to first think clearly about the behavioral hypotheses behind the choices, then develop the framework, and *then* design a survey to support the model. The final major lesson is that these integrated models require both customized programs and fast computers for estimation. The estimation programs and models tend to be complex, and therefore synthetic data should be used to confirm the program's ability to reproduce the parameters as a matter of routine. Such a test provides assurance that the model is identified and that the likelihood is programmed correctly, but does not otherwise validate the model specification.

Conclusion

In this chapter, we presented a general methodology and framework for including latent variables—in particular, attitudes and perceptions—in choice models. The methodology provides a framework for the use of psychometric data to explicitly model attitudes and perceptions and their influences on choices.

The methodology requires the estimation of an integrated multi-equation model consisting of a discrete choice model and the latent variable model's structural and measurement equations. The approach uses maximum likelihood techniques to estimate the integrated model, in which the likelihood function for the integrated model includes complex multi-dimensional integrals (one integral per latent construct). Estimation is performed either by numerical integration or simulation (MSM or MSL), and requires customized programs and fast computers.

Three applications of the methodology are presented. The findings from the reviewed case studies are that implementation of the integrated choice and latent variable model framework results in: improvements in goodness of fit over choice models without latent variables, latent variables that are statistically significant in the choice model, and a more satisfying behavioral representation. Application of these methods requires careful consideration of the behavioral framework, and then design of the data collection phase to generate good indicators and causal variables that support the framework.

To conclude, we note that the methodology presented here and the empirical case studies that were reviewed have merely brought to the surface the potential for the integrated modeling framework. Further work is needed including investigation in the following areas:

Behavioral Framework: By integrating latent variable models and choice models, we can begin to reflect behavioral theory that has here-to-for primarily existed in descriptive, flow-type models. The behavioral framework and the methodology we present needs to be extended to continue bridging the gap between behavioral theory and statistical models. For example, including memory, awareness, process, feedback, temporal variables, tastes, goals, context, etc. in the framework.

Validation: The early signs indicate that the methodology is promising: the goodness of fit improves, the latent variables are significant, and the behavioral representation is more satisfying. For specific applications it would also be useful to conduct validation tests, including tests of forecasting ability, consequences of misspecifications (for example, excluding latent variables that should be present), and performance comparisons with models of simpler formulations.

Identification: Other than the methods we present for identification (the *Three-step Rule*, the use of synthetic data, and the evaluation of the Hessian), there are no additional rules for identification of the general formulation of the integrated choice and latent variable models. Similar to the way that necessary and sufficient rules were developed for LISREL, the knowledge base of identification issues for the integrated model must be expanded.

Computation: Application of this method is computationally intensive due to the evaluation of the integral. Estimation time varies significantly with the particular application, but is usually on the order of a few hours to several days using, for example, a 500 plus MHz Pentium processor. Investigation of techniques such as parallel computing, particularly for estimation by simulation, would greatly ease the application of such models.

The approach presented in this chapter is a flexible, powerful, and theoretically grounded methodology that will allow the modeling of complex behavioral processes.

Chapter 4:
Generalized Discrete Choice Model

In this chapter, we present a generalized discrete choice model that synthesizes a wide variety of enhancements that have been made to the basic discrete choice paradigm. The model has the ability to incorporate key aspects of behavioral realism and yet remains mathematically tractable. The chapter begins by summarizing a variety of extensions, including those described in Chapters 2 and 3 as well as others, and then presents a generalized framework and specification. The basic technique for integrating the methods is to start with the multinomial logit formulation, and then add extensions that relax simplifying assumptions and enrich the capabilities of the basic model. The extended models often result in functional forms composed of complex multidimensional integrals, and so a key part of the generalized model is the implementation of the logit kernel smooth simulator described in Chapter 2. This chapter also provides empirical results that demonstrate and test the generalized discrete choice modeling framework.

Introduction

As described in the introductory chapter, researchers have long been focused on improving the specification of the discrete choice model. A guiding philosophy in these developments is that such enhancements lead to a more behaviorally realistic representation of the choice process, and consequently a better understanding of behavior, improvements in forecasts, and valuable information regarding the validity of simpler model structures. The objective of this chapter is to extend the basic discrete choice model by integrating with it a number of extensions, including:

- *Factor Analytic Probit-like Disturbances*
- *Combining Revealed Preferences and Stated Preferences*
- *Latent Variables*
- *Latent Classes*

We present a generalized framework that encompasses these extensions, describe each enhancement and associated equations, and show relationships between methods including how they can be integrated. Note that we summarize the material presented in Chapters 2 and 3 in order to provide a complete picture of the generalized framework and to allow this chapter to stand on its own.

The extended models often result in functional forms composed of complex multidimensional integrals. Therefore, we also describe an estimation method consisting of Maximum Simulated Likelihood Estimation with a Logit Kernel smooth simulator, which provides for practical estimation of such models.

The Discrete Choice Model

The framework for the standard discrete choice model is again shown in Figure 4-1. The model is based on the notion that individual derives utility by choosing an alternative. The utilities U are latent variables, and the observable choices y are manifestations of the underlying utilities. The utilities are assumed to be a function of a set of explanatory variables X, which describe the decision-maker n and the alternative i, i.e.:

$$U_{in} = V(X_{in};q) + e_{in} \ ,$$

where: V is a function of the explanatory variables,

q is a vector of unknown parameters, and

e_{in} is a random disturbance.

This formulation is grounded in classic microeconomic consumer theory; brings in the random utility paradigm pioneered by Thurstone (1927), Marshak (1960), and Luce (1959); and incorporates the manner of specifying utilities developed by Lancaster (1966) and McFadden (1974).

Starting from this general equation, assumptions on the decision protocol and on the distributions of the disturbances lead to various choice models, most commonly the utility maximizing GEV forms (multinomial logit, nested logit, cross-nested logit) or probit.

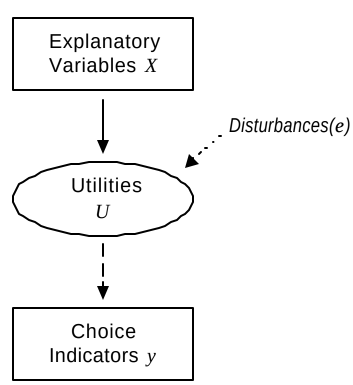


Figure 4-1: Discrete Choice Model

Simplifying Assumptions and the Base Model

In this chapter, we present the generalized discrete choice model as a set of methods that extend the multinomial logit model (MNL). For each of the described methods, MNL forms the core, and all extensions are built upon it. As will become apparent, this formulation offers complete flexibility (for example, probit-like disturbances and nested structures can easily be implemented), enables straightforward combination of methods, and has computational advantages.

In order to clarify the presentation of the generalized framework, we also make several simplifying assumptions: we assume utility maximizing behavior, linear in the parameters systematic utilities, and a universal choice set across respondents. It is straightforward to relax these assumptions, and we will do so where a deviation is useful for the discussion.

Given this, the base discrete choice model is specified as follows: [4-1]

$U_{in} = X_{in} b + n_{in}$, or, in vector notation $U_n = X_n b + n_n$, "Structural Equation"

$$y_{in} = \begin{cases} 1, & if\, U_{in} = \max_j \{U_{jn}\} \\ 0, & otherwise \end{cases}.$$ "Measurement Equation"

where:

- n denotes individuals, $n = 1,...,N$, where N is the size of the sample;
- i, j denote alternatives;
- C is the choice set, which is comprised of J alternatives;
- U_{in} is the utility of alternative i as perceived by n; U_n is the $(J \times 1)$ vector of utilities;
- X_{in} is a $(1 \times K)$ vector describing n and i; X_n is the $(J \times K)$ matrix of stacked X_{in};
- b is a $(K \times 1)$ vector of unknown parameters;
- y_{in} is the choice indicator, and y_n is the $(J \times 1)$ vector of choice indicators; and

Finally, making the assumption that the disturbance (n_{in}) is i.i.d. Extreme Value or Gumbel$(0, m)$, the structural and measurement equations lead to the MNL formulation:

$$P(i \mid X_n) = \frac{e^{m(X_{in} b)}}{\sum_{j \in C} e^{m(X_{jn} b)}} \quad \left(\text{and the likelihood is } P(y_n \mid X_n) = \prod_{i \in C} P(i \mid X_n)^{y_{in}} \right) \qquad [4\text{-}2]$$

where $P(i \mid X_n)$ is the probability that $y_{in} = 1$, given X_n (and parameters b). We denote the logit probability as $\Lambda(i \mid X_n)$. The variance of n_{in} is g / m^2, where g is the variance of a standard Gumbel $(p^2 / 6)$.

Overview of the Components of the Generalized Framework

In this section, we provide background material and a brief presentation of each of the four extensions. (Appendix D provides further detail on each of the extensions.) We will end with a summary of the generalized discrete choice model.

Factor Analytic Disturbances and Logit Kernel

This first extension deals with both the disturbances of the choice model and computational issues. The primary limitations with multinomial logit models, or Generalized Extreme Value models in general, derive from the rigidity of the error structure. One relatively new solution to this problem is the *logit kernel model* presented in Chapter 2, and which we briefly summarize here. This is a discrete choice model that has both probit-like disturbances, which provide flexibility, as well as an additive i.i.d. Extreme Value (or Gumbel) disturbance, which aids in computation.

Framework

The framework for the model is shown in Figure 4-2, which is just like the framework of a standard discrete choice model except it has a parameterized disturbance. We parameterize the error structure using a factor analytic form because this provides great flexibility and also enables one to represent complex covariance structures with relatively few parameters and factors. This is a general formulation that can be used to specify all known (additive) error structures, including, heteroscedasticity, nested, cross-nested, random parameters, and auto-regressive processes.

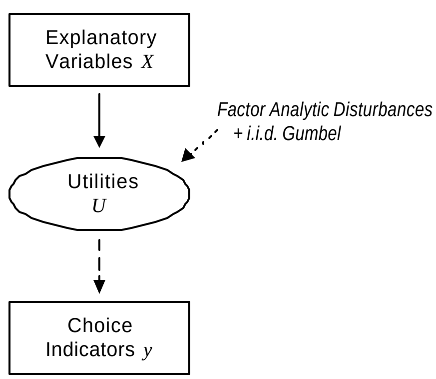


Figure 4-2: Discrete Choice Model
with Factor Analytic Disturbances and a Logit Kernel

Specification

The structure of the model (expanded on in both Chapter 2 and Appendix D) is:

$$U_n = X_n b + F_n T z_n + n_n ,$$

where: $F_n T z_n$ are the factor analytic disturbances

F_n is a $(J \times M)$ matrix of factor loadings, including fixed and/or unknown parameters,

T is an $(M \times M)$ lower triangular cholesky matrix of unknown parameters, where $TT' = Cov(Tz_n)$,

z_n is an $(M \times 1)$ vector of unknown factors with independent standard distributions, and

U, X, b, n are as in the base MNL model (Equation [4-1]).

While the factor analytic disturbances provide for flexibility, the i.i.d. Gumbel term aids in computation. Namely, if the factors z_n are known, the model corresponds to a multinomial logit formulation:

$$\Lambda(i \mid X_n, z_n) = \frac{e^{m\left(X_{in} b + F_{in} T z_n\right)}}{\sum_{j \in C} e^{m\left(X_{jn} b + F_{jn} T z_n\right)}} ,$$

Since the z_n is in fact not known, the unconditional choice probability of interest is:

$$P(i \mid X_n) = \int_z \Lambda(i \mid X_n, z) n(z, I_M) dz \quad ,$$

where $n(z, I_M)$ is the joint density function of z . We can naturally estimate $P(i \mid X_n; d)$ with an unbiased, smooth, tractable simulator, which we compute as:

$$\hat{P}(i \mid X_n) = \frac{1}{\mathbb{D}} \sum_{d=1}^{\mathbb{D}} \Lambda(i \mid X_n, z_n^d) ,$$

where z_n^d denotes draw d from the distribution of z_n , thus enabling us to estimate high dimensional integrals with relative ease.

Applications

The earliest applications of logit kernel were in random parameter logit specifications, which appeared 20 years ago in the papers by Boyd and Mellman (1980) and Cardell and Dunbar (1980). Since then, there have been numerous applications and investigations into various aspects of the model, including Ben-Akiva and Bolduc (1996), Bhat (1997, 1998), Bolduc and Ben-Akiva (1991), Bolduc, Fortin and Fournier (1996), Brownstone, Bunch and Train (2000), Brownstone and Train (1999), Goett, Hudson, and Train (2000), Gönül and Srinivasan (1993), Greene (2000), Mehndiratta and Hansen (1997), Revelt and Train (1998 & 1999), Srinivasan and Mahmassani (2000), and Train (1998). Most of the applications in the literature are in the area of random parameters, but there are also applications of heteroscedasticity (Ben-Akiva and Bolduc, 1996, and Greene, 2000), nesting (Ben-Akiva and Bolduc, 1996), cross-nesting (Bhat, 1997), dynamics (Srinivasan and Mahmassani, 2000), and auto-regressive applications (Bolduc, Fortin and Fournier, 1996). A very important recent contribution is McFadden and Train's (2000) paper on mixed logit (a generalization of logit kernel in which the mixing function does not have to be continuous), which both *(i)* proves that any well-behaved random utility consistent behavior can be represented as closely as desired with a mixed logit specification and *(ii)* presents easy to implement specification tests for these models.

Combining Stated and Revealed Preferences[27]

The second extension deals with the issue of combining choice data from different sources. There are two broad classes of choice or preference data that are used to estimate discrete choice models: revealed preferences, which are based on actual market behavior, and stated preferences, which are expressed responses to hypothetical scenarios. Each type of data has its advantages and disadvantages, including:

Choices: Revealed preferences are cognitively congruent with actual behavior, whereas stated preferences may be subject to various response biases.

Alternatives: Revealed preferences can only be gathered for existing alternatives, whereas stated preferences can be elicited for new (i.e., non-existing) alternatives.

Attributes: The attributes of the alternatives in a revealed preference setting often have limited ranges, include measurement errors, and are correlated. Stated preference surveys can address all of these issues through appropriate experimental designs.

Choice set: The actual choice sets are often ambiguous for revealed preferences, whereas for stated preferences they are well defined (albeit the respondent may not consider all alternatives).

Number of responses: It is difficult to obtain multiple revealed preferences from an individual (for example, it requires a panel setting), whereas repetitive questioning using hypothetical scenarios is easily implemented in stated preference surveys.

Response format: Revealed preferences only provide information on the actual choice, whereas stated preferences can employ various response formats such as ranking, rating, or matching data that provide more information.

Given these strengths and weaknesses, the two types of data are highly complementary, and combined estimators can be used to draw on the advantages of each. A fundamental assumption in conducting SP surveys is that the trade-off relationship among major attributes is common to both revealed and stated

[27] This SP/RP discussion in this chapter is based on Ben-Akiva and Morikawa 1990, Morikawa 1989, and Morikawa, Ben-Akiva, and McFadden 1996.

preferences. When there is such an overlap between the RP model and the SP model, there are advantages to jointly estimating the models.

Framework

Ben-Akiva and Morikawa (1990) developed techniques for combining the two types of data. (See also the review in Ben-Akiva et al., 1994.) The framework for the combined estimator is shown in Figure 4-3, in which both stated preferences and revealed preferences are indicators of the unobservable utilities. The benefits of the combined model include correcting bias that may exist in the SP responses, identifying the effect of new services, identifying the effects of attributes that have either limited range or are highly correlated in the RP data, and improving efficiency of the parameter estimates. In order to combine the preference data, there are two important issues involving the RP and SP disturbances that need to be considered. First, they are most likely correlated across multiple responses for a given individual. Second, the scale (i.e., the variances of the disturbances) may vary across the two models. Methods for addressing these issues are discussed in Appendix D.

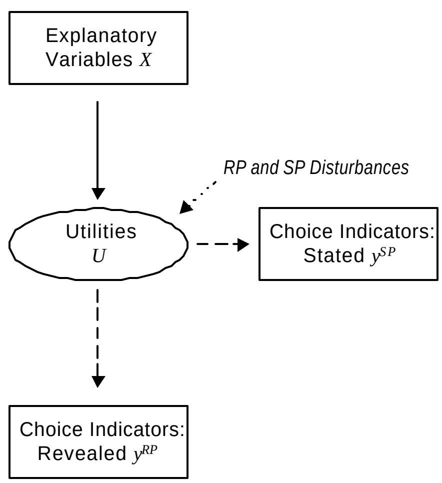


Figure 4-3: Joint Revealed and Stated Preference Model

Applications

These techniques are becoming fairly common in the literature. For example, joint SP/RP models have been used to model recreational site choice (Adamowicz et al., 1994), intercity mode choice (Ben-Akiva and Morikawa, 1990), choices among gasoline and alternative fueled vehicles (Brownstone et al., 2000), and pre-trip decisions as influenced by traveler information systems (Khattak et al., 1996).

Choice and Latent Variables

This extension deals with the causal structure of the model, and the ideas include capturing latent causal variables and also making use of different types of behavioral data. Often in behavioral sciences, there are concepts of interest that are not well defined and cannot be directly measured, for example knowledge, ambition, or personality. These concepts are referred to as latent constructs. While there exists no operational methods to directly measure these constructs, latent variable modeling techniques are often applied to infer information about latent variables. These techniques are based on the hypothesis that although the construct itself cannot be observed, its effects on measurable variables (called 'indicators') are observable and such relationships provide information on the underlying latent variable. We consider first the incorporation of continuous latent factors as explanatory variables in discrete choice models (a summary of what was presented in Chapter 3), and in the subsequent extension we also incorporate discrete latent constructs.

The behavioral framework for integrated choice and latent variable models is presented in Figure 4-4. The aim is to explicitly treat the psychological factors, such as attitudes and perceptions, affecting the utility by modeling them as latent variables. Psychometric data, such as responses to attitudinal and perceptual survey questions, are used as indicators of the latent psychological factors.

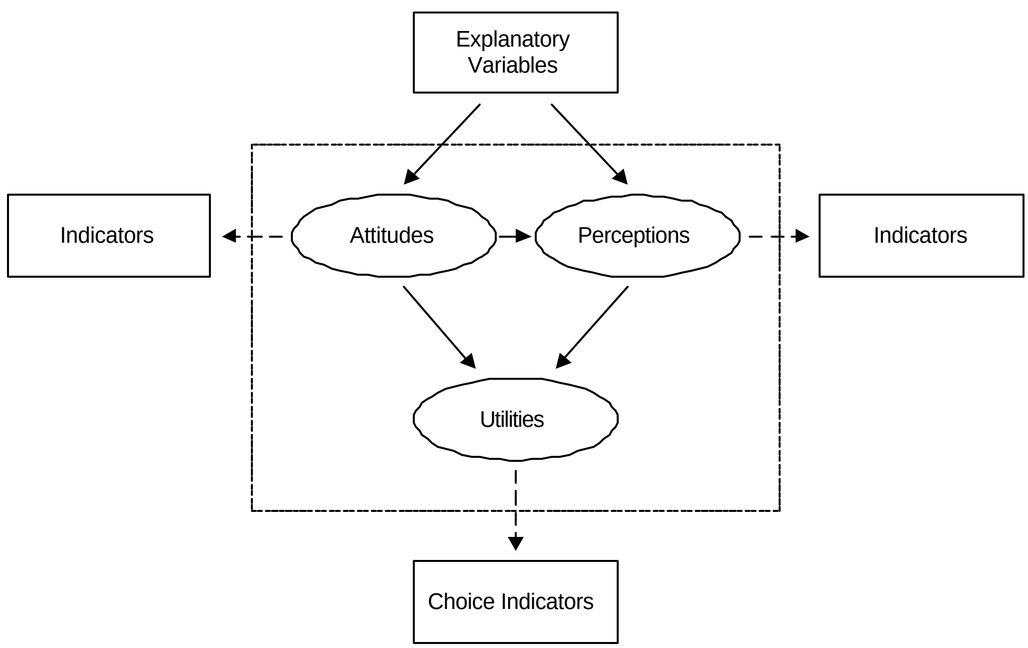


Figure 4-4: Behavioral Framework for Including Attitudes and Perceptions in Discrete choice Models

A general approach to synthesizing models with latent variables and psychometric-type measurement models has been advanced by a number of researchers including Keesling (1972), Jöreskog (1973), Wiley (1973), and Bentler (1980), who developed the structural and measurement equation framework and methodology for specifying and estimating latent variable models. Such models are widely used to define and measure unobservable factors. Much of this work focuses on continuous latent constructs and continuous indicators and is not described in relation to discrete choice models. When discrete indicators are involved, direct application of the continuous indicator approach results in inconsistent estimates. Various corrective procedures have been developed for discrete indicators (see, for example, Olsson 1979, Muthén 1979, 1983, and 1984), and methods have been developed when both the latent variables and indicators are discrete (see, for example, Goodman, 1974, and McCutcheon, 1987).

In the area of discrete choice models, researchers have used various techniques in an effort to explicitly capture latent psychological factors in choice models. Alternative approaches include directly introducing the indicators as explanatory variables, or sequentially estimating a latent variable model and then a choice model (see Chapter 3 for a discussion). The method presented here is a general treatment of the inclusion of latent variables and psychometric data in discrete choice models. The methodology integrates latent variable models with discrete choice models, resulting in a rigorous methodology for explicitly including psychological factors in choice models. A simultaneous maximum likelihood estimation method is employed, which results in consistent and efficient estimates of the model parameters.

The work on the methodology presented here began during the mid-1980s with the objective of making the connection between econometric choice models and the extensive market research literature on the study of consumer preferences (Cambridge Systematics, 1986; McFadden, 1986; and Ben-Akiva and Boccara, 1987). Since then, a number of researchers have continued developing and testing the techniques as evidenced by the variety of applications discussed below.

Framework

The integrated modeling framework, shown in Figure 4-5, consists of two components, a choice model and a latent variable model.

The choice model is as before, except that now some of the explanatory variables are not directly observable. It is possible to identify a choice model with limited latent variables using only observed choices and no additional indicators (see, for example, Elrod, 1998). However, it is quite likely that the information content from the choice indicators will not be sufficient to empirically identify the effects of individual-specific latent variables. Therefore, indicators of the latent variables are used for identification, and are introduced in the form of a latent variable model.

The top portion of Figure 4-5 is a latent variable model. Latent variable models are used when we have available indicators for the latent variables. Indicators could be responses to survey questions regarding, for example, the level of agreement, satisfaction with, or importance of attributes or an attitudinal statement. The figure depicts such indicators as manifestations of the underlying latent variable, and the associated measurement equation is represented by a dashed arrow. A structural relationship links the observable causal variables (and potentially other latent causal variables) to the latent variable, and these are shown as solid arrows.

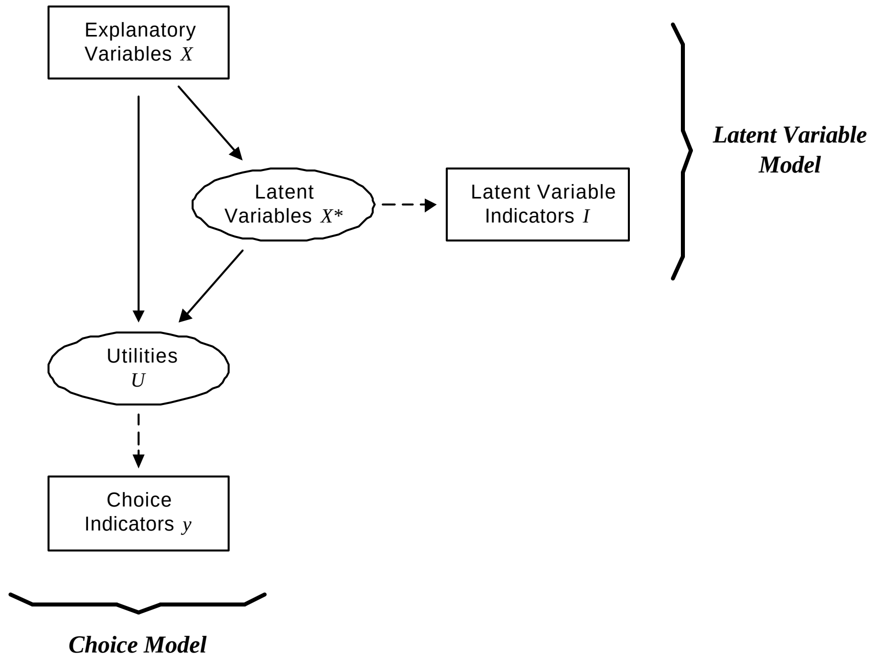


Figure 4-5: Integrated Choice & Latent Variable Model

The integrated choice and latent variable model explicitly models the latent variables that influence the choice process. Structural equations relating the observable explanatory variables to the latent variables model the behavioral process by which the latent variables are formed. While the latent constructs are not observable, their effects on indicators are observable. Note that the indicators do not have a causal relationship that influences the behavior. That is, the arrow goes from the latent variable to the indicator, and the indicators are only used to aid in measuring the underlying causal relationships (the solid arrows). Because the indicators are not part of the causal relationships, they are typically used only in the model estimation stage and not in model application.

Applications

The following are examples of how latent variables have been incorporated into choice models (some of which were described in detail in Chapter 3):

- Bernardino (1996) modeled telecommuting behavior and included latent attributes such as the costs and benefits of a program,
- Börsch-Supan et al. (1996) modeled the choice of living arrangements of the elderly and included a latent health characteristic,
- Hosoda (1999) modeled shoppers' mode choices and included latent sensitivities to time, cost, comfort, and convenience.

- Morikawa et al (1996) modeled intercity mode choices and included the latent attributes of comfort and convenience,
- Polydoropoulou (1997) modeled responses to advanced traveler information systems and included latent variables such as knowledge and satisfaction,
- Ramming (2000) modeled commuters' choice of route to work and included a latent characteristic that represents knowledge of the transportation system, and
- Train et al. (1987) modeled consumers' choices of public utility rate schedules and included latent characteristics such as the importance of energy consumption and the importance of finding new energy sources.

Choice and Latent Classes

This extension focuses on capturing latent segmentation in the population. As with random parameter models and latent variable models, latent class models also capture unobserved heterogeneity, but are employed when the latent variables are discrete constructs. The idea is that there may be discrete segments of decision-makers that are not immediately identifiable from the data. Furthermore, these segments (or classes) may exhibit different choice behavior in terms of choice sets, decision protocols, tastes, or model structure (for example, nesting). While we cannot deterministically identify the classes from the observable variables, we presume that class membership probabilities can be estimated.

Framework

The framework for a latent class model is shown in Figure 4-6, in which the latent classes are shown to either impact the formulation of the utilities in terms of, for example, taste variation, decision protocols, or choice sets. The basic form of the latent class model is:

$$P(i \mid X_n) = \sum_{s=1}^{S} P(i \mid X_n; s) P(s \mid X_n) .$$

In this equation, the choice model, $P(i \mid X_n; s)$, is class-specific and may be specified differently for different classes of individuals, s. The class membership model, $P(s \mid X_n)$, is the probability of belonging to class s, and may depend on explanatory variables X_n.

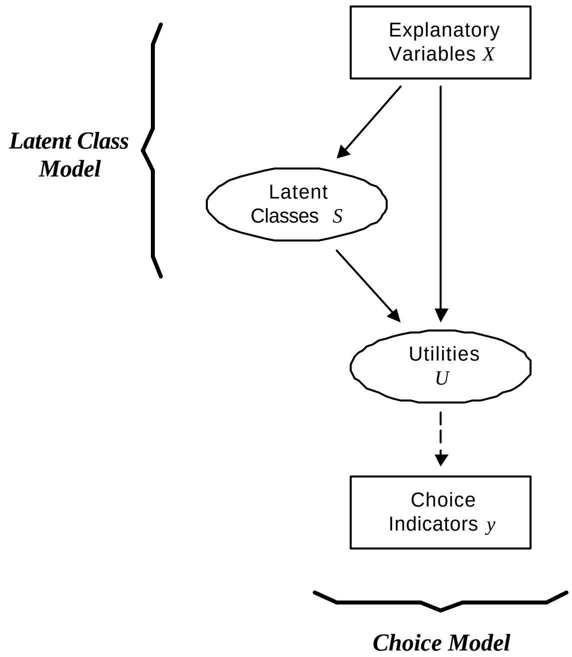


Figure 4-6: Discrete Choice Model with Latent Classes

Applications

The following are examples of how latent classes have been used to improve the behavioral representation and explanatory power of choice models:

- Ben-Akiva and Boccara (1996) modeled commuters' mode choices allowing for different choice sets among travelers,
- Gopinath (1995) modeled intercity travelers' mode choices allowing for different decision protocols among classes (for example, utility maximizers versus habitual choosers),
- Gopinath (1995) modeled shippers' choices between train and truck allowing for different sensitivities to time and cost, and
- Hosoda (1999) modeled shopper's mode choice allowing for different sensitivities of time and cost, for example, distinguishing between patient and impatient travelers.

The Generalized Discrete Choice Model

Integrating the extensions described above leads to the generalized discrete choice model as shown in Figure 4-7. The framework draws on ideas from a great number of researchers, including Ben-Akiva and Morikawa (1990) who developed the methods for combining revealed and stated preferences; Cambridge

Systematics (1986) and McFadden (1986) who laid out the original ideas for incorporating latent variables and psychometric data into choice models; Ben-Akiva and Boccara (1987) and Morikawa, Ben-Akiva, and McFadden (1996) who continued the development for including psychometric data in choice models; Gopinath (1995) who developed rigorous and flexible methods for capturing latent class segmentation in choice models; and Ben-Akiva and Bolduc (1996) who introduced an additive factor analytic parameterized disturbance to MNL's i.i.d Gumbel.

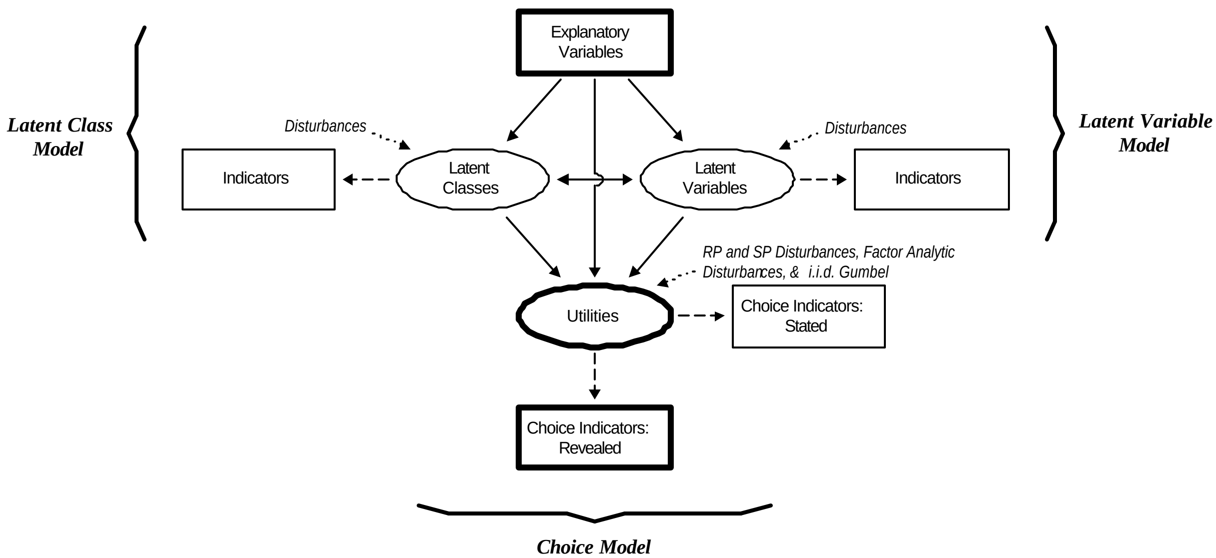
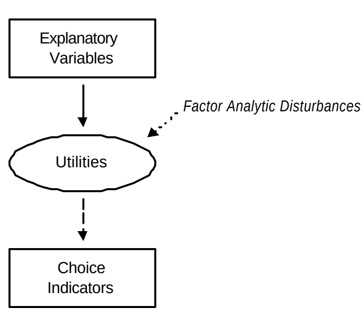


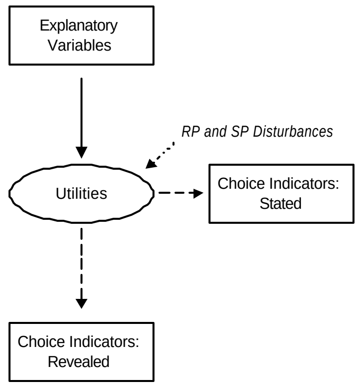
Figure 4-7: Generalized Discrete Choice Framework

As shown in Figure 4-7, the core of the model is the standard multinomial logit model (highlighted in bold), and then the extensions are built upon it:

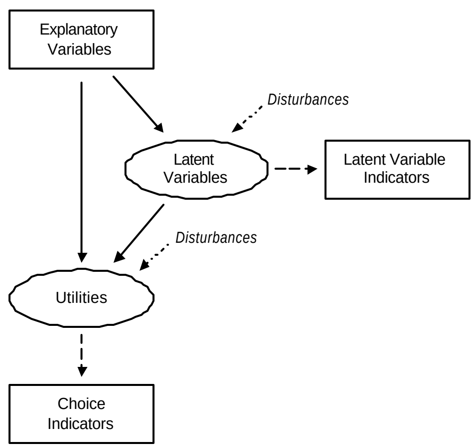
- *Factor Analytic (probit-like) disturbances* in order to provide a flexible covariance structure, thereby relaxing the independence from irrelevant alternatives (IIA) condition of MNL and enabling estimation of unobserved heterogeneity through, for example, random parameters.



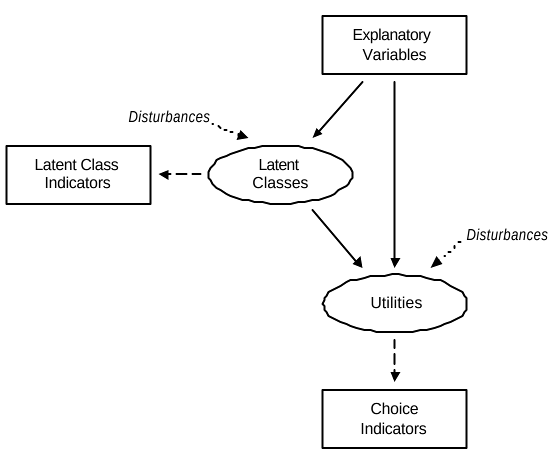
- *Combining revealed preferences* (what people actually do) *and stated preferences (*what people say that they would do) in order to draw on the advantages of the two types of data.



- *Incorporating latent variables* in order to integrate behavioral indicators and to provide a richer explanation of behavior by explicitly representing the formation and effects of latent constructs such as attitudes and perceptions.



- *Stipulating latent classes* in order to incorporate yet another type of behavioral indicator and to capture latent segmentation that may influence various aspects of the choice process including taste parameters, choice sets, and decision protocols.



Most of the methodological developments and applications found in the literature apply a single one of the extensions we describe in this chapter. Exceptions that we have found are Gönül and Srinivasan (1993) who developed a model with random parameters and latent classes and Hosoda (1999) who included continuous latent variables as explanatory variables in a latent class model. The generalized framework proposed here integrates the various extensions available for discrete choice models.

The framework has its foundation in the random utility model, makes use of different types of data that provide insight into the choice process, allows for any desirable disturbance structure (including random parameters and nesting structures) through the factor analytic disturbances, and provides means for capturing latent heterogeneity and behavioral constructs through the latent variable and latent class modeling structures. Through these extensions, the choice model can capture more behaviorally realistic choice processes and enable the validity of more parsimonious structures to be tested. Furthermore, the framework can be practically implemented via use of the logit kernel smooth simulator (as a result of the additive i.i.d. Gumbel) and a maximum simulated likelihood estimator.

Generalized Discrete Choice Model

In this section, we discuss the specification, estimation, and identification for the generalized model.

Framework

The framework for the generalized discrete choice model is in Figure 4-7, which shows how the extensions (factor analytic disturbances, joint SP/RP, latent variables, and latent classes) are conceptually integrated into a single framework.

Specification

In specifying the generalized discrete choice model, it is useful to think of two different aspects to the process. The first is specifying the behavioral model of interest, i.e., a model that explains market behavior (revealed preferences) and the causal relationships behind this behavior. Typically, a model with a rich behavioral structure cannot be estimated by drawing on revealed preferences alone. So, the second aspect of the specification has to do with incorporating additional behavioral indicators to aid in estimating and identifying the parameters in the model of interest. Each of these aspects is addressed below.

The Generalized Choice Model

The generalized model that explains the market behavior consists of several components. The core of the model is the multinomial logit probability, which we denote as $\Lambda(y_n^{RP} \mid X_n)$. As discussed above, adding features such as factor analytic disturbances (FTz), latent variables (X_n^*), and latent classes (s) can be used to relax the limiting restrictions of the multinomial logit formulation and enrich the behavioral representation of the model. While these additional elements are all unknown factors, we can write the multinomial logit probability *given* the latent variables, latent classes, and factors, which we denote as $\Lambda(y_n^{RP} \mid X_n, X_n^*, s, z)$.

However, because the latent variables, classes, and factors are, in fact, unobservable, there are additional components to the model that are necessary in order to specify their distributions. These include:

- The distribution of the factor analytic disturbances, $f(z)$;
- The distribution of the latent variables, as defined by the latent variables structural (i.e., causal) model, $f(X_n^* \mid X_n)$; and
- The class membership model, $P(s \mid X_n)$, which is the probability of belonging to class s given explanatory variables X_n.

These components are integrated together to form the generalized choice model:

$$P(y_n^{RP} \mid X_n) = \iint \sum_{s=1}^{S} \left(\Lambda(y_n^{RP} \mid X_n, X^*, s, z) P(s \mid X_n) \right) f(X^* \mid X_n) f(z) dX^* dz \qquad [4\text{-}3]$$

The conditional logit probabilities, $\Lambda(y_n^{RP} \mid X_n, X_n^*, s, z)$, are first summed over the latent classes, and then integrated over the unknown latent variables and factor analytic disturbances. The resulting function is the probability of the revealed behavior as a function of observable explanatory variables. This is the model of interest in that it explains market behavior. It also allows for a rich causal specification through incorporation of flexible disturbances, latent variables, and latent classes. This generalized choice model includes the parameters of the systematic utilities from the basic logit model (b), the parameters of the factor analytic disturbances, the parameters of the class membership model, and those of the structural equations of the latent variables. This is a lot to estimate using only the revealed choices, and this is where the other sources of data come into play.

The Likelihood Function

While the revealed preferences are the only behavior that we are interested in explaining and predicting, there also exists a host of other behavioral indicators that can provide assistance in estimating the parameters of the behavioral model presented above. These include:

- Stated preferences, y_n^{SP}, which aid in estimating the parameters of the choice model (b).
- Psychometric indicators, I_n, which help with the estimation of the class membership model, $P(s \mid X_n)$, and the latent variable structural model, $f(X_n^* \mid X_n)$.

To make use of this information, we introduce two more elements to the model. The first is the SP model, which is analogous to the RP model as written above:

$\Lambda(y_{nq}^{SP} \mid X_n, X_n^*, s, z)$, and $q = 1,...,Q_n$, denoting multiple responses per individual.

The SP model will share some parameters with the RP model. Thus by using appropriate experimental designs for the SP experiment, the inclusion of SP data can improve the estimation of the RP choice model parameters. Note that there is often correlation among SP responses and between SP and RP responses that should be captured in the joint model. (See Appendix D for further discussion.)

The second element is the measurement model for the latent constructs. This is written as the distribution of the indicators given the latent variables and the class, s, as follows:

$$f(I_n \mid X_n^*, s)$$

Note that the addition of the SP model and the measurement model will add nuisance parameters, which are not a part of the behavioral model of interest (i.e., Equation [4-3]), but also must be estimated.

Incorporating these additional elements into Equation [4-3], the likelihood function is then:

$$P(y_n^{RP}, y_n^{SP}, I_n \mid X_n) = \quad [4\text{-}4]$$

$$\iint \sum_{s=1}^{S} \left(\Lambda(y_n^{RP} \mid X_n, X^*, s, z) \prod_{q=1}^{Q_n} \Lambda(y_{nq}^{SP} \mid X_n, X^*, s, z) f(I_n \mid X^*, s) P(s \mid X_n) \right) f(X^* \mid X_n) f(z) dX^* dz$$

Alternatively,

$$P(y_n^{RP}, y_n^{SP}, I_n \mid X_n) = \iint \sum_{s=1}^{S} \left(P(y_n^{RP}, y_n^{SP} \mid X_n, X_n^*, s, z) f(I_n \mid X^*, s) P(s \mid X_n) \right) f(X^* \mid X_n) f(z) dX^* dz,$$

where $P(y_n^{RP}, y_n^{SP} \mid X_n, X_n^*, s, z) = \Lambda(y_n^{RP} \mid X_n, X^*, s, z) \prod_{q=1}^{Q_n} \Lambda(y_{nq}^{SP} \mid X_n, X^*, s, z)$.

Application

While the full specification shown in Equation [4-4] is used to estimate the model, the aim of including the additional behavioral indicators is simply to improve the specification of the parameters in Equation [4-3]. This latter equation is the model of interest, and it is the one used for model application.

Estimation

We use maximum simulated likelihood (MSL) techniques for estimation, although clearly other methods (for example, Method of Moments) could be implemented. We choose MSL because of its straightforward interpretation and implementation, as well as its performance and asymptotic properties.

As described above, Equation [4-4] is used for estimation. One key in estimation is to write the equation such that the distribution over which the integral is taken is independent multivariate standard normal, because this allows the application of general estimation code. For example, making the assumption that the latent variable structural model is of the form:

$$X_n^* = h(X_n) + w \ ,$$

where $h(\cdot)$ is a vector function of the explanatory variables and w is a vector of random disturbances. Given these relationships, X_n^* can then be replaced in the likelihood by X_n and w. Thus, $f(I_n | X_n^*, s)$ becomes $f(I_n | X_n, w, s)$; $f(X_n^* | X_n)$ becomes $f(w)$; and $P(y_n^{RP}, y_n^{SP} | X_n, X_n^*, s, z)$ becomes $P(y_n^{RP}, y_n^{SP} | X_n, w, s, z)$, which leads to the following likelihood function:

$$P(y_n^{RP}, y_n^{SP}, I_n | X_n) =$$

$$\iint \sum_{s=1}^{S} \left(P(y_n^{RP}, y_n^{SP} | X_n, w, s, z) f(I_n | X_n, w, s) P(s | X_n) \right) f(w) f(z) \, dw \, dz$$

By construction, the factors w (from the latent variables) and z (from the factor analytic disturbances and correlation among RP and SP disturbances) are i.i.d. normally distributed (via the Cholesky decomposition, if necessary). A second key to the estimation is to keep the dimensionality of the integral down. The dimension is determined by the factor analytic parameters (in z), the RP/SP correlation terms (also in z), and the latent variables (w). It is also desirable to keep the number of classes small. When the dimension of the integral is above 3, simulation techniques are required in order to evaluate the integral. The basic idea behind simulation is to replace the multifold integral (the likelihood function) with easy to compute probability simulators. The advantage of the logit kernel formulation is that it provides a tractable, unbiased, and smooth simulator for the likelihood, namely:

$$\hat{P}(y_n^{RP}, y_n^{SP}, I_n | X_n) = \qquad [4\text{-}5]$$

$$\frac{1}{\mathbb{D}} \sum_{d=1}^{\mathbb{D}} \sum_{s=1}^{S} P(y_n^{RP}, y_n^{SP} | X_n, w_n^d, s, z_n^d) f(I_n | X_n, w_n^d, s) P(s | X_n),$$

where z_n^d *and* w_n^d are particular realizations (or draws) from a standard normal distribution. Thus, the integral is replaced with an average of values of the function computed at discrete points. There has been a lot of research concerning how best to generate the set of discrete points. The most straightforward approach is to use pseudo-random sequences (for example, Monte Carlo). However, variance reduction techniques (for example, antithetic draws) and quasi-random approaches (for example, Halton draws, which are used in this chapter) have been found to cover the dimension space more evenly and thus are more efficient. See Bhat (2000) for further discussion.

Using the probability simulator, the simulated log-likelihood of the sample is:

$$\hat{L}(d) = \sum_{n=1}^{N} \ln \hat{P}(y_n^{RP}, y_n^{SP}, I_n \mid X_n) \ , \qquad [4\text{-}6]$$

where d is the vector of all the unknown parameters. The parameters are then estimated by maximizing Equation [4-6] over the unknown parameters.[28]

A well-known issue is that the simulated log-likelihood function, although consistent, is simulated with a downward bias for finite number of draws. The issue is that while the probability simulator (Equation [4-5]) is unbiased, the log-simulated-likelihood (Equation [4-6]) is biased due to the log transformation. In order to minimize the bias in simulating the log-likelihood function, it is important to simulate the probabilities with good precision. The precision increases with the number of draws, as well as with the use of intelligent methods to generate the draws. The number of draws necessary to sufficiently remove the bias cannot be determined a priori; it depends on the type of draws, the model specification, and the data. Therefore, when estimating these models, it is necessary to verify stability in the parameter estimates as the number of draws is increased. In Appendix E, we provide results verifying that the models we present in the case study (next) have 'stabilized', which we somewhat arbitrarily define as when the estimation results converge to within one standard error. Note that as the dimensionality of the integral increases, so too does the required number of draws. Also note that some of our models (particularly the high dimensional random parameter models) required 20,000 Halton draws, and they are still not perfectly stable. This suggests that the model may need to be simplified in order to make estimation feasible.

Identification

Identification can be difficult, particularly as the model gets more complex. While specific identification rules have been developed for special cases of the generalized framework, there are no general necessary and sufficient conditions for identification. The best we can do is to apply the sufficient, but not necessary technique of conditionally identifying each sub-module (as in the two- and three-step approaches). However, in many cases there remains uncertainty regarding identification and, furthermore, even models that are theoretically identified often have multicollinearity issues that impede estimation of the parameters. Therefore, the use of empirical identification tests is highly recommended. There are several possible techniques in this category, including:

- Conducting Monte Carlo experiments by generating synthetic data from the specified model structure (with given parameter values), and then attempting to reproduce the parameters using the maximum likelihood estimator. If the parameters cannot be reproduced to some degree of accuracy, then this is an indication that the model is not identified.

[28] In some cases, sequential estimation methods could be used (see, for example, Ben-Akiva et al., 1999, Morikawa 1989, and Morikawa et al., 1996), which produce consistent but inefficient estimates.

- Verifying that the parameters converge to the same point and likelihood given different starting values.

- Verifying that the Hessian of the log-likelihood function is non-singular (a test of local identification). This test is usually performed automatically in order to generate estimates of the standard errors of estimated parameters.

- Constraining one or more parameters to different values, and verifying that the log-likelihood shifts as a result. (This test is particularly useful when there is one or more suspect parameters.)

- Verifying that the parameters are stable as the number of simulation draws is increased. This is critical, as an unidentified model will usually appear identified with a small number of draws.

Case Study

To demonstrate and test the generalized discrete choice model, we applied the technique to a single model application. Appendix D provides the general equations for each of the methods, and here we provide them for a particular application.

Data

The models presented use data collected in 1987 for the Netherlands Railway. (A subset of these data was used for Case Study 1 in Chapter 3.) The purpose in collecting the data was to assess the factors that influence the choice between rail and auto for intercity travel. The data were collected by telephone, and consist of people who had traveled between Nijmegen and Randstad (approximately a two-hour trip) in the 3 months prior to the survey. The following information was collected for each of 228 respondents:

- *Demographic data:*
  Characteristics of the respondent, for example, age and gender.

- *Psychometric data:*
  Subjective ratings of latent attributes of rail and auto, for example, relaxation and reliability.

- *Revealed Preference data (RP):*
  Characteristics of the Nijmegen to Randstad trip made by the respondent, including:
  - the chosen mode (rail or auto);
  - characteristics of the trip, such as trip purpose (business or other), number of persons traveling, and whether or not there was a fixed arrival time requirement; and
  - attributes of the alternatives, including cost, in-vehicle and out-of-vehicle travel times, number of transfers (rail only).

- *Stated Preference data 1 (SP1 – rail versus rail):*
  Responses to a stated preference experiment of a choice between two hypothetical rail services.

  For each experiment, the respondent was presented with two hypothetical rail alternatives for the particular intercity trip reported in the RP experiment. Each alternative was described by travel cost, travel time, number of transfers, and level of amenities. Level of amenities is a package of different aspects such as seating room and availability, quietness, smoothness of ride, heating/ventilation, and food service, but is presented at only three levels (0, 1, and 2, the lower the better). Given the two alternatives, the respondent was asked to state his or her preference on the basis of a five point scale:

1 - definitely choose alternative 1,
2 - probably choose alternative 1,
3 - not sure,
4 - probably choose alternative 2, and
5 - definitely choose alternative 2.

Each respondent was presented with multiple pairs of choices, and a total of 2,875 responses were collected (an average of about 13 per person).

- *Stated Preference data 2 (SP2 – rail versus auto):*
  Responses to a stated preference experiment of a choice between hypothetical rail and auto services.

  For each experiment, the respondent was presented with a hypothetical rail alternative and a hypothetical auto alternative for the particular intercity trip reported in the RP experiment. Each alternative was described by travel cost, travel time, number of transfers (rail only), and level of amenities (rail only). Given the two alternatives, the respondent was asked to state his or her preference on the basis of a five point scale:

  1 - definitely choose auto,
  2 - probably choose auto,
  3 - not sure,
  4 - probably choose rail, and
  5 - definitely choose rail.

  Each respondent was presented with multiple pairs of choices, and a total of 1,577 responses were collected (an average of about 7 per person).

For additional information on the data, see Bradley, Grosvenor, and Bouma (1988).

Base Models for the Case Study

For binary choice models, it is convenient to introduce a slightly different notation than in the general case. There are 2 utilities, only the difference between the utilities matters, and so we express one utility equation, which is the difference between the two utilities:

$$U_n = U_{1n} - U_{2n} = X_n b + n_n \ ,$$

where U_n is (1×1), X_n is $(1 \times K)$ and is equal to $(X_{1n} - X_{2n})$, b $(K \times 1)$ is as before, and n_n is the difference between two independent Gumbel distributed random variables (and is therefore logistically distributed).

For the revealed preference data, the choice indicator is a standard $0/1$ binary choice indicator, and we redefine the choice indicator as:

$$y_n^{RP} = \begin{cases} 1 & if\ U_n^{RP} \geq 0 \quad (person\ n\ chose\ rail) \\ -1 & if\ U_n^{RP} < 0 \quad (person\ n\ chose\ auto) \end{cases} .$$

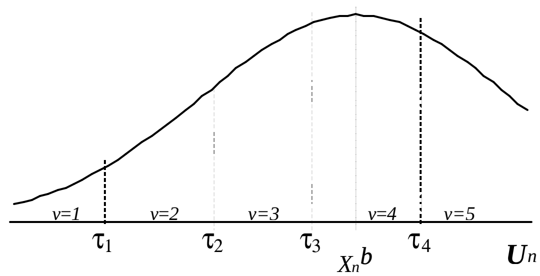
The likelihood for an RP response is then:

$$\Lambda(y_n^{RP} \mid X_n^{RP}; b, m) = \frac{1}{1 + e^{-m(X_n^{RP} b) y_n^{RP}}} .$$

[4-7]

The stated preference choice indicators consist of a five-point preference rating, and so an ordinal logit model is used. The utility is specified as above (in differenced form), and threshold values (t) are specified in the utility scale such that:

$$P_n(1) = P(t_0 < U_n^{SP} \leq t_1),$$
$$P_n(2) = P(t_1 < U_n^{SP} \leq t_2),$$
$$P_n(3) = P(t_2 < U_n^{SP} \leq t_3),$$
$$P_n(4) = P(t_3 < U_n^{SP} \leq t_4),$$
$$P_n(5) = P(t_4 < U_n^{SP} \leq t_5),$$



where $t_0 = -\infty$ and $t_5 = \infty$.

We define the ordinal choice indicator as:

$$y_{in}^{SP} = \begin{cases} 1 & if\ t_{i\text{-}1} < U_n^{SP} \leq t_i \\ 0 & otherwise \end{cases}, \qquad i = 1,\ldots,5.$$

and the vector of these indicators is $y_n^{SP} = (y_{1n}^{SP}, \ldots, y_{5n}^{SP})'$.

The likelihood for each ordinal preference rating is then:

$$P(y_n^{SP} \mid X_n^{SP}; b, t, m) = \prod_{i=1}^{5} \left(\frac{1}{1 + e^{-m(X_n^{SP} b - t_i)}} - \frac{1}{1 + e^{-m(X_n^{SP} b - t_{i-1})}} \right)^{y_{in}^{SP}},$$

where there is a different specification for each SP dataset (SP1 and SP2).

One final detail on the ordinal model is the normalization of the threshold parameters. For the Rail versus Rail stated preference data, the order of the alternatives is irrelevant (i.e., they can be swapped without affecting the model), therefore, the threshold parameters must be symmetric, i.e., $t_1 = -t_4$ and $t_2 = -t_3$. We verified that the data support this constraint (via a likelihood ratio test), and all models presented here impose the constraint. For the Rail versus Auto stated preference data, the symmetry condition is not necessary (and statistical tests on the data verified that it does not hold). However, since we estimate a constant in the model, we must impose one constraint on the threshold parameters to identify the model. The constraint we impose is $t_2 = -t_3$, because this maintains zero as the center point of the threshold parameters.

Table 4-1: Revealed Preference Binary Logit Mode Choice Model

Mode Choice Model: Rail versus Auto

Parameter	U_{rail}	U_{auto}	Est.	Std Er.	*t-stat*
Rail constant	✓		0.637	0.425	*(1.5)*
Work trip dummy	✓		1.21	0.48	*(2.5)*
Fixed arrival time dummy	✓		0.736	0.368	*(2.0)*
Female dummy	✓		0.949	0.352	*(2.7)*
Cost per person in Guilders	✓	✓	-0.0477	0.0122	*(3.9)*
Out-of-vehicle time in hours	✓	✓	-2.90	0.80	*(3.6)*
In-vehicle time in hours	✓	✓	-0.554	0.462	*(1.2)*
Number of transfers	✓		-0.255	0.255	*(1.0)*
Number of observations:			228		
Log-likelihood:			-109.89		
Rho-bar-squared:			0.254		

Revealed Preference Model

The first model we present using the mode choice data is a binary logit model using the revealed preference data. This is equivalent to a classic mode choice model. The likelihood for this model is as written in Equation [4-7]. The estimation results are shown in Table 4-1. We report robust standard errors[29] and/or t-statistics for all models. The check marks in the U_{rail} and U_{auto} columns signify whether the parameter is included in the rail and/or auto utility. The signs of the parameters are as expected. With the exception of in-vehicle time and number of transfers, the parameters are significantly different from zero. The monetary value of in-vehicle time is 11.6 Guilders per hour or about $5.60 per hour[30], and for out-of-vehicle time it jumps to 60 Guilders or $29 per hour.

Joint Stated and Revealed Preference Model

We first apply the joint RP/SP technique, because this model then forms the basis for all other models that we estimate. For each respondent, we have the following choice indicators available:

[29] Using the robust asymptotic covariance matrix estimator $H^{-1}BH^{-1}$, where H is the Hessian (calculated numerically, in our case) and B is the cross product of the gradient. (Newey and McFadden, 1994)

[30] In 1985 dollars, using 1985 exchange rate.

Type of Indicator	*# Per Person*
Revealed preference	1
Stated preferences from rail versus rail hypothetical scenarios	Q_n
Stated preferences from rail versus auto hypothetical scenarios	R_n

The utilities are:

$$U_n^{RP} = X_n^{RP} b + y^{RP} h_n + n_n^{RP},$$
$$U_{nq}^{SP1} = X_{nq}^{SP1} b + n_{nq}^{SP1}, \qquad q = 1,..., Q_n, \qquad \textit{(rail versus rail)},$$
$$U_{nr}^{SP2} = X_{nr}^{SP2} b + y^{SP2} h_n + n_{nr}^{SP2}, \qquad r = 1,..., R_n, \qquad \textit{(rail versus auto)},$$

Where an agent effect (h) is included to capture correlation among the SP responses and between the SP and RP responses for a given individual. It does not enter the SP1 model, because it does not have defined alternatives (i.e., it is rail versus rail).

The likelihood function for the joint model is:

$$P(y_n^{RP}, y_n^{SP1}, y_n^{SP2} \mid X_n) = \qquad [4\text{-}8]$$
$$\int_h \Lambda(y_n^{RP} \mid X_n^{RP}, h) P(y_n^{SP1} \mid X_n^{SP1}, h) P(y_n^{SP2} \mid X_n^{SP2}, h) f(h) dh,$$

where: h is a scalar parameter,

$f(\cdot)$ denotes the standard normal distribution,

d includes b, m, y,

$$\Lambda(y_n^{RP} \mid X_n^{RP}, h_n) = \frac{1}{1 + e^{-(X_n^{RP} b + y^{RP} h_n) y_n^{RP}}}, \qquad [4\text{-}9]$$

$$P(y_n^{SP1} \mid X_n^{SP1})$$
$$= \prod_{q=1}^{Q_n} \prod_{j=1}^{5} \left(\frac{1}{1 + e^{-m_{SP1}(X_{nq}^{SP1} b - t_j^{SP2})}} - \frac{1}{1 + e^{-m_{SP1}(X_{nq}^{SP1} b - t_{j-1}^{SP1})}} \right)^{y_{jnq}^{SP}}, \qquad [4\text{-}10]$$

$$P(y_n^{SP2} \mid X_n^{SP2}, h_n)$$
$$= \prod_{r=1}^{R_n} \prod_{i=1}^{5} \left(\frac{1}{1 + e^{-m_{SP2}(X_{nr}^{SP2} b + y^{SP2} h_n - t_i^{SP2})}} - \frac{1}{1 + e^{-m_{SP2}(X_{nr}^{SP2} b + y^{SP2} h_n - t_{i-1}^{SP2})}} \right)^{y_{inr}^{SP}}. \qquad [4\text{-}11]$$

Table 4-2: Joint Stated Preference & Revealed Preference Mode Choice Model

			Joint RP/SP1/SP2			RP Only (Rail vs. Auto)			SP1 Only (Rail vs. Rail)			SP2 Only (Rail vs. Auto)		
Parameter	U_{rail}	U_{auto}	Est.	*Std Er.*	*t-stat*	Est.	*Std Er.*	*t-stat*	Est.	*Std Er.*	*t-stat*	Est.	*Std Er.*	*t-stat*
Rail constant RP	✓		0.444	0.493	*(0.9)*	0.637	0.425	*(1.5)*						
Rail constant SP2	✓		-0.466	0.777	*(0.6)*							-2.10	0.63	*(3.3)*
Work trip dummy	✓		1.17	0.51	*(2.3)*	1.21	0.48	*(2.5)*						
Fixed arrival time dummy	✓		0.723	0.381	*(1.9)*	0.736	0.368	*(2.0)*						
Female dummy	✓		0.990	0.381	*(2.6)*	0.949	0.352	*(2.7)*						
Cost per person in Guilders	✓	✓	-0.0608	0.0132	*(4.6)*	-0.0477	0.0122	*(3.9)*	-0.141	0.012	*(11.8)*	-0.0703	0.0180	*(3.9)*
Out-of-vehicle time in hours	✓	✓	-2.23	0.83	*(2.7)*	-2.90	0.80	*(3.6)*				-0.841	0.935	*(0.9)*
In-vehicle time in hours	✓	✓	-0.710	0.158	*(4.5)*	-0.554	0.462	*(1.2)*	-1.64	0.16	*(10.2)*	-1.23	0.41	*(3.0)*
Number of transfers	✓		-0.100	0.036	*(2.8)*	-0.255	0.255	*(1.0)*	-0.238	0.066	*(3.6)*	0.0798	0.1995	*(0.4)*
Amenities	✓		-0.361	0.080	*(4.5)*				-0.821	0.073	*(11.2)*	-0.925	0.237	*(3.9)*
Inertia dummy (RP Choice)	✓		2.97	1.02	*(2.9)*							5.92	0.68	*(8.7)*
Agent effect RP			0.686	0.490	*(1.4)*									
Agent effect SP2			2.44	0.50	*(4.9)*							3.11	0.29	*(10.8)*
Scale (mu) SP1			2.31	0.50	*(4.6)*									
Scale (mu) SP2			1.31	0.30	*(4.4)*									
Tau1 SP1 (=-Tau4 SP1)			*-0.195*	----	----				*-0.450*	----	----			
Tau2 SP1 (=-Tau3 SP1)			*-0.0127*	----	----				*-0.0292*	----	----			
Tau3 SP1			0.0127	0.0036	*(3.5)*				0.0292	0.0060	*(4.9)*			
Tau4 SP1			0.195	0.049	*(4.0)*				0.450	0.038	*(11.7)*			
Tau1 SP2			-0.986	0.219	*(4.5)*							-1.30	0.13	*(10.2)*
Tau2 SP2 (=-Tau3 SP2)			*-0.180*	----	----							*-0.238*	----	----
Tau3 SP2			0.180	0.053	*(3.4)*							0.238	0.055	*(4.3)*
Tau4 SP2			1.32	0.32	*(4.1)*							1.75	0.18	*(9.6)*
Number of observations:			4680			228			2875			1577		
Number of draws (Halton):			1000			1000			1000			1000		
Log-likelihood:			-4517.43			-109.89			-3131.10			-1271.29		
Rho-bar-squared:			0.380			0.254			0.322			0.495		

The estimation results are presented in Table 4-2.[31] The joint model is presented along with models estimated individually on each of the three datasets. A likelihood ratio test was performed to verify that the restrictions imposed by the joint model are supported by the data: the 8 restrictions result in a reduction of under 6 log-likelihood points and therefore the restrictions are not rejected at a 10% significance level.

One clear benefit of the joint model is that the parameters for in-vehicle travel time and number of transfers are now statistically significant. The monetary value of in-vehicle time remains consistent with the RP model at about \$5.60/hour, whereas the value of out-of-vehicle time falls from around \$29 to under \$18/hour. Another benefit is that the concept of ‘amenities’ is now captured in the model. Both the inertia

[31] All models are estimated using Maximum Simulated Likelihood Estimation techniques. The method and related issues (for example, number of draws and Halton draws) will be covered when estimation is discussed for the integrated model.

and agent effect are highly significant, and therefore estimating this model with the inertia effect and without the agent effect would result in biased estimates of the parameters.

Random Parameter (Factor Analytic) Model

As an example of a logit kernel model, we have taken the joint SP/RP mode choice model presented in the previous section and allowed some of the parameters to be randomly distributed. Separating out the parameters that are fixed across the population (b) from those that are allowed to vary across the population (g_n), the model is now specified as follows:

$$U_n^{RP} = X_n^{RP} b + W_n^{RP} g_n + y^{RP} h_n + n_n^{RP} ,$$

$$U_{nq}^{SP1} = X_{nq}^{SP1} b + W_{nq}^{SP1} g_n + n_{nq}^{SP1} , \qquad q = 1,..., Q_n ,$$

$$U_{nr}^{SP2} = X_{nr}^{SP2} b + W_{nr}^{SP2} g_n + y^{SP2} h_n + n_{nr}^{SP2} , \qquad r = 1,..., R_n .$$

where X and W are the explanatory variables (formerly all included in X).

In the random parameter model presented, we allow the parameters associated with attributes of the alternatives to be distributed, i.e., W includes the following five variables:

- Cost per person
- Out-of-vehicle travel time
- In-vehicle travel time
- Number of transfers
- Amenities

All of these parameters have sign restrictions, and therefore we specify the parameters with a multivariate lognormal distribution. Replacing g_n with the equivalent lognormal relationship $g_n = -mexp(g + Tz_n)$, (where the minus constrains the signs to be negative and *mexp()* is defined below) the model is then written as follows:

$$U_n^{RP} = X_n^{RP} b + W_n^{RP} \left(\text{-}mexp(g + Tz_n)\right) + y^{RP} h_n + n_n^{RP} ,$$

$$U_{nq}^{SP1} = X_{nq}^{SP1} b + W_{nq}^{SP1} \left(\text{-}mexp(g + Tz_n)\right) + n_{nq}^{SP1} ,$$

$$U_{nr}^{SP2} = X_{nr}^{SP2} b + W_{nr}^{SP2} \left(\text{-}mexp(g + Tz_n)\right) + y^{SP2} h_n + n_{nr}^{SP2} ,$$

where:

- g is a (5×1) vector of unknown parameters ,
- z is a (5×1) vector of independent standard normals ,
- T is a (5×5) lower triangular matrix of unknown parameters , and
- $mexp(x)$ is an operator that exponentiates each element in the vector x .

The likelihood is then:

$$P(y_n^{RP}, y_n^{SP1}, y_n^{SP2} \mid X_n) = \iint_{z,h} \Lambda(y_n^{RP} \mid X_n^{RP}, z, h) \prod_{q=1}^{Q_n} P(y_{nq}^{SP1} \mid X_{nq}^{SP1}, z)$$

$$* \prod_{r=1}^{R_n} P(y_{nr}^{SP2} \mid X_{nr}^{SP2}, z, h) \prod_{k=1}^{5} f(z_k) f(h) dz\, dh \ ,$$

where the unknown parameters include $b, m, \Psi, g,$ and T (using the notation defined earlier).

The results for the random parameter mode choice model are shown in Table 4-3.[32] The first model is the joint SP/RP model shown in Table 4-2, and is repeated here for comparison purposes. The second model provides estimation results for a random parameter model in which the parameters are independently distributed (i.e., T is diagonal). We find that there is a large improvement in fit over the model with fixed parameters. The third model allows for correlations among the random parameters (i.e., T is lower triangular), which provides a marginal improvement in fit.

Note that because of the structure of the lognormally distributed parameters, the t-stats do not have their normal interpretation. The parameter estimates and standard errors reported in Table 4-3 for the distributed parameters are g and the elements of T. However, these parameters are related to the mean and variance of the distributed parameters as follows:

$$mean(g_{kn}) = e^{g_k + 0.5(TT')_{kk}} \ ,$$

$$variance(g_{kn}) = e^{2g_k} \left(e^{2(TT')_{kk}} - e^{(TT')_{kk}} \right) \ ,$$

where $(TT')_{kk}$ is the k^{th} diagonal element of TT'.

[32] See Random Parameter section of Chapter 2 for a discussion of identification of lognormally distributed random parameters.

Table 4-3: Random Parameter Mode Choice Model

Group	Parameter	Base RP/SP Model: Not Distributed Est.	Std. Er.	t-stat	Distributed Model 1: Independent Distributions Est.	Std. Er.	t-stat	Distributed Model 2: Multivariate Distributions Est.	Std. Er.	t-stat
	Rail constant RP	0.444	0.493	*(0.9)*	2.80	0.97	*(2.9)*	1.67	0.81	*(2.1)*
	Rail constant SP2	-0.466	0.777	*(0.6)*	4.05	1.20	*(3.4)*	2.19	0.79	*(2.8)*
	Work trip dummy	1.17	0.51	*(2.3)*	0.891	0.762	*(1.2)*	1.16	0.65	*(1.8)*
	Fixed arrival time dummy	0.723	0.381	*(1.9)*	0.513	0.647	*(0.8)*	0.850	0.522	*(1.6)*
	Female dummy	0.990	0.381	*(2.6)*	1.61	0.61	*(2.7)*	1.51	0.51	*(2.9)*
Distributed	1 Cost per person in Guilders	-0.0608	0.0132	*(4.6)*	-2.19	0.26	*	-2.33	0.26	*
	2 Out-of-vehicle time in hours	-2.23	0.83	*(2.7)*	1.56	0.34	*	0.97	0.36	*
	3 In-vehicle time in hours	-0.710	0.158	*(4.5)*	0.284	0.279	*	0.149	0.271	*
	4 Number of transfers	-0.100	0.036	*(2.8)*	-2.29	0.33	*	-2.25	0.31	*
	5 Amenities	-0.361	0.080	*(4.5)*	-0.644	0.265	*	-0.722	0.274	*
Parameters in the T Matrix	T11				0.993	0.129	*(7.7)*	1.29	0.06	*(21.5)*
	T21							-0.479	0.043	*(11.2)*
	T31							0.470	0.045	*(10.4)*
	T41							0.645	0.055	*(11.7)*
	T51							0.404	0.043	*(9.4)*
	T22				0.723	0.166	*(4.4)*	0.658	0.060	*(10.9)*
	T32							0.281	0.063	*(4.5)*
	T42							0.287	0.021	*(13.8)*
	T52							0.035	0.048	*(0.7)*
	T33				0.818	0.057	*(14.3)*	0.894	0.042	*(21.3)*
	T43							0.106	0.036	*(2.9)*
	T53							0.136	0.033	*(4.1)*
	T44				1.96	0.21	*(9.3)*	1.83	0.11	*(17.4)*
	T54							0.344	0.024	*(14.1)*
	T55				1.06	0.05	*(20.9)*	1.11	0.07	*(15.6)*
	Inertia dummy (RP Choice)	2.97	1.02	*(2.9)*	-0.245	0.680	*(0.4)*	1.097	0.481	*(2.3)*
	Agent effect RP	0.686	0.490	*(1.4)*	3.19	1.28	*(2.5)*	2.07	0.65	*(3.2)*
	Agent effect SP2	2.44	0.50	*(4.9)*	4.14	1.14	*(3.6)*	3.74	1.05	*(3.6)*
	Scale (mu) SP1	2.31	0.50	*(4.6)*	4.07	1.11	*(3.7)*	5.21	1.44	*(3.6)*
	Scale (mu) SP2	1.31	0.30	*(4.4)*	1.79	0.48	*(3.8)*	1.88	0.54	*(3.5)*
	Tau1 SP1 (=-Tau4 SP1)	*-0.195*	----		*-0.241*			*-0.196*	----	----
	Tau2 SP1 (=-Tau3 SP1)	*-0.0127*	----		*-0.0159*			*-0.0128*	----	----
	Tau3 SP1	0.0127	0.0036	*(3.5)*	0.0159	0.0052	*(3.0)*	0.0128	0.0043	*(2.9)*
	Tau4 SP1	0.195	0.049	*(4.0)*	0.241	0.081	*(3.0)*	0.196	0.071	*(2.7)*
	Tau1 SP2	-0.986	0.219	*(4.5)*	-0.904	0.241	*(3.8)*	-0.856	0.241	*(3.6)*
	Tau2 SP2 (=-Tau3 SP2)	*-0.180*	----		*-0.160*			*-0.150*	----	----
	Tau3 SP2	0.180	0.053	*(3.4)*	0.160	0.055	*(2.9)*	0.150	0.053	*(2.8)*
	Tau4 SP2	1.32	0.32	*(4.1)*	1.15	0.31	*(3.8)*	1.08	0.31	*(3.5)*
	Number of observations:	4680			4680			4680		
	Number of draws (Halton):	1000			20000			20000		
	Log-likelihood:	-4517.43			-3931.20			-3911.72		
	Rho-bar-squared:	0.380			0.460			0.461		

* Testing that the lognormal location parameter is different from 0 is meaningless.

First, note that it is meaningless to test that the location parameter, g_k is different from zero. What we want to test is that the $mean(g_{kn})$ is different from zero. Second, testing that the parameters in T are significantly different from zero provides some information, but again it would be better to test the (co)variances directly. We did generate t-stats for the mean and standard deviations of the population parameters for the independently distributed parameters (not shown), and the t-stats for the standard deviations ranged between 1.8 and 4.8, which are more in line with the t-stats for the other parameters in the model. Regardless of the t-stats, we can tell by the increase in fit that the additional parameters improved the model.

Table 4-4 provides the estimated mean and standard deviation for each of the distributed parameters.

Table 4-4: Mean and Standard Deviations of the Distributed Parameters

	Base RP/SP Model: Not Distributed		Distributed Model 1: Independent Distributions		Distributed Model 2: Multivariate Distributions	
Parameter	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Cost per person in Guilders	-0.0608	0.000	-0.183	0.237	-0.223	0.459
Out-of-vehicle time in hours	-2.23	0.000	-6.19	5.12	-3.68	3.57
In-vehicle time in hours	-0.710	0.000	-1.86	1.81	-2.01	2.84
Number of transfers	-0.100	0.000	-0.689	4.629	-0.728	4.971
Amenities	-0.361	0.000	-0.922	1.331	-1.04	1.99

Choice and Latent Variable Model

Our mode choice dataset includes information pertaining to the respondents' subjective ratings of various latent attributes. Following the RP portion of the survey, the respondents were asked to rate the following aspects for both rail and auto:

- Relaxation during the trip
- Reliability of arrival time
- Flexibility of choosing departure time
- Ease of traveling with children and/or heavy baggage
- Safety during the trip
- Overall rating of the mode

Responses for the first 5 attributes were in the form of a 5-point scale (from very bad to very good), and the overall rating was on a 10-point scale (again, from very bad to very good).

Clearly these responses provide information on the behavior. The question is how do we use this information? Frequently, such data are directly inserted as explanatory variables in the choice model, resulting in highly significant parameter estimates and large improvements in model fit. However, there are several issues with such an approach. First, the data are not available for forecasting, so if forecasting is desired then such a specification is problematic. Second, the multicollinearity inherent in responses to such a string of questions often makes it difficult to estimate the full set of parameters. The third and most fundamental issue is that it is not clear that such data are causal. For these reasons, we use the latent variable modeling approach, which assumes that these responses are indicators for a smaller number of underlying causal latent attributes. Furthermore, these latent attributes can be explained by observable attributes of the alternatives and characteristics of the respondent.

The equations for the RP/SP mode choice and latent variable model follow. First, some notes on the model:

- All variables, including the latent variables and their indicators, are measured in terms of the difference between rail and auto. This was done to simplify the specification: it reduces the dimensionality of the integral by 2, it cuts down on the number of parameters, and it lowers the potential for multicollinearity among the latent variable structural equations.
- The indicators in differenced form have a 9-point scale for the first 5 attribute ratings, and a 19-point scale for the 'overall' attribute rating, and therefore are treated as continuous variables.
- We performed a combination of exploratory and confirmatory analysis to arrive at the final structure of the latent variable model, which consists of 2 latent variables labeled comfort and convenience.
- The indicators pertain to the RP choice, and therefore the latent variables are specified using only RP data in the structural equation. However, we hypothesize that these latent perceptions also impact the stated preference rail versus auto experiment (SP2), and so we include the latent variables in the SP2 model, but allow them to have different weights (i.e., b 's).

To specify the joint choice and latent variable model, we need to write the structural and measurement equations for both the latent variable component and the choice component. The equations are as follows:

Latent variable structural equations:

$$X^*_{ln} = X^{LV}_n l + w_{ln} \; ; \quad l = 1,2 \; ; \quad w_n \sim N(0, I) \, .$$

The variances of the disturbance w_n are set equal to 1 to set the scale of the latent variables (necessary for identification). We experimented with models that allowed a covariance term (i.e., non-orthogonal latent variables), but it was consistently insignificant.

Choice model structural equations (as before, but with the addition of the latent variable):

$$U^{RP}_n = X^{RP}_n b_1 + X^*_n b^{RP}_2 + y^{RP} h_n + n^{RP}_n \; ,$$

$$U^{SP1}_{nq} = X^{SP1}_{nq} b_1 + n^{SP1}_{nq} \; , \qquad q = 1, ..., Q_n \; ,$$

$$U^{SP2}_{nr} = X^{SP2}_{nr} b_1 + X^*_n b^{SP2}_2 + y^{SP2} h_n + n^{SP2}_{nr} \; , \qquad r = 1, ..., R_n \; .$$

Latent variable measurement equations:

$$I_{bn} = X^*_n a_b + u_{bn} \; ; \quad b = 1, \ldots, 6 \; ; \quad u_n \sim N(0, \Sigma_u \; diagonal) \, .$$

Choice model measurement equations (as before):

$$y_n^{RP} = \begin{cases} 1 & if\, U_n^{RP} \geq 0 \\ -1 & if\, U_n^{RP} < 0 \end{cases},$$

$$y_{inq}^{SP1} = \begin{cases} 1 & if\, t_{i-1}^{SP1} < U_{nq}^{SP1} \leq t_i^{SP1} \\ 0 & otherwise \end{cases}, \qquad i = 1,\ldots,5\,,\;\; q = 1,\ldots, Q_n\,,$$

$$y_{inr}^{SP2} = \begin{cases} 1 & if\, t_{i-1}^{SP2} < U_{nr}^{SP2} \leq t_i^{SP2} \\ 0 & otherwise \end{cases}, \qquad i = 1,\ldots,5\,,\;\; r = 1,\ldots, R_n\,.$$

The likelihood function for the joint model is:

$$P(y_n^{RP}, y_n^{SP1}, y_n^{SP2}, I_n \mid X_n) =$$
$$\iint_{h,X^*} \Lambda(y_n^{RP} \mid X_n^{RP}, X^*, h) \prod_{q=1}^{Q_n} P(y_{nq}^{SP1} \mid X_{nq}^{SP1}, h) \prod_{r=1}^{R_n} P(y_{nr}^{SP2} \mid X_{nr}^{SP2}, X^*, h)$$
$$* f_2(I_n \mid X^*)\, f_1(X^* \mid X_n^{RP}) f(h) dh dX^*\,,$$

where:

$\Lambda(y_n^{RP} \mid X_n^{RP}, X_n^*, h_n)$, $P(y_n^{SP1} \mid X_n^{SP1})$, and $P(y_n^{SP2} \mid X_n^{SP2}, X_n^*, h_n)$ are as in Equations [4-9], [4-10], and [4-11], but with the latent explanatory variables (i.e., the utilities as written above) ;

$$f_2(I_n \mid X^*) = \prod_{b=1}^{6} \frac{1}{s_{u_b}} f\left(\frac{I_{bn} - X^* a_b}{s_{u_b}}\right), \qquad s_{u_b}^2 = \text{var}(u_b)\;;$$

$$f_1(X^* \mid X_n^{RP}) = \prod_{l=1}^{2} f\left(X_l^* - X_n^{LV} l_l\right)\;; \text{ and}$$

The unknown parameters (using the notation defined earlier) include $b, m, y, a, l,$ and s_u.

Table 4-5: Choice and Latent Variable Mode Choice Model

CHOICE MODEL

	Base RP/SP Choice Model			Choice and Latent Variable RP/SP Model (latent variable portion below)		
Parameter	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat
Rail constant RP	0.444	0.493	*(0.9)*	-0.442	0.750	*(0.6)*
Rail constant SP2	-0.466	0.777	*(0.6)*	-0.890	0.837	*(1.1)*
Work trip dummy	1.17	0.51	*(2.3)*	1.67	0.64	*(2.6)*
Fixed arrival time dummy	0.723	0.381	*(1.9)*	0.692	0.532	*(1.3)*
Female dummy	0.99	0.38	*(2.6)*	1.13	0.45	*(2.5)*
Cost per person in Guilders	-0.0608	0.0132	*(4.6)*	-0.0605	0.0163	*(3.7)*
Out-of-vehicle time in hours	-2.23	0.83	*(2.7)*	-0.983	0.936	*(1.1)*
In-vehicle time in hours	-0.710	0.158	*(4.5)*	-0.691	0.186	*(3.7)*
Number of transfers	-0.100	0.036	*(2.8)*	-0.0982	0.0384	*(2.6)*
Amenities	-0.361	0.080	*(4.5)*	-0.358	0.097	*(3.7)*
Latent Comfort - RP				1.16	1.17	*(1.0)*
Latent Comfort - SP2				1.16	0.55	*(2.1)*
Latent Convenience - RP				1.30	0.76	*(1.7)*
Latent Convenience - SP2				0.764	0.331	*(2.3)*
Inertia dummy (RP Choice)	2.97	1.02	*(2.9)*	2.52	1.24	*(2.0)*
Agent effect RP	0.686	0.490	*(1.4)*	0.210	0.611	*(0.3)*
Agent effect SP2	2.44	0.50	*(4.9)*	2.08	0.64	*(3.3)*
Scale (mu) SP1	2.31	0.50	*(4.6)*	2.32	0.63	*(3.7)*
Scale (mu) SP2	1.31	0.30	*(4.4)*	1.31	0.42	*(3.1)*
Tau1 SP1 (=-Tau4 SP1)	-0.195	----	----	-0.194	----	----
Tau2 SP1 (=-Tau3 SP1)	-0.0127	----	----	-0.0126	----	----
Tau3 SP1	0.0127	0.0036	*(3.5)*	0.0126	0.0041	*(3.0)*
Tau4 SP1	0.195	0.049	*(4.0)*	0.194	0.058	*(3.3)*
Tau1 SP2	-0.986	0.219	*(4.5)*	-0.988	0.313	*(3.2)*
Tau2 SP2 (=-Tau3 SP2)	-0.180	----	----	-0.181	----	----
Tau3 SP2	0.180	0.053	*(3.4)*	0.181	0.065	*(2.8)*
Tau4 SP2	1.32	0.32	*(4.1)*	1.33	0.44	*(3.0)*
Number of observations:	4680			4680		
Number of draws (Halton):	1000			5000		
Log-likelihood (Choice&Latent):				-6656.12		
Log-likelihood (Choice):	-4517.43			-4517.97		
Rho-bar-squared (Choice):	0.380			0.380		

LATENT VARIABLE MODEL

Structural Equations (2 equations, 1 per column)

	Comfort Equation			Convenience Equation		
Parameter	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat
Constant - Comfort	0.106	0.219	*(0.5)*			
Constant - Convenience				0.489	0.303	*(1.6)*
Age dummy - over 40	-0.449	0.622	*(0.7)*	0.871	0.287	*(3.0)*
First class rail rider	0.431	0.567	*(0.8)*			
In-vehicle time in hours	-0.481	0.331	*(1.5)*			
Out-of-vehicle time in hours				-1.18	0.71	*(1.6)*
Number of transfers				-0.122	0.199	*(0.6)*
Free parking dummy (auto)				0.222	0.242	*(0.9)*
Variance(ω)	1.00	----	----	1.00	----	----
Squared Multiple Correlation (SMC)	0.092			0.230		

Measurement Equations (6 equations, 1 per row)

	Comfort Parameters			Convenience Parameters			Disturbance Params. (StdDev(υ))			Fit (SMC)
Equation	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat	
Relaxation	0.522	0.240	*(2.2)*	0.131	0.135	*(1.0)*	1.17	0.13	*(9.3)*	0.172
Reliability	0.331	0.105	*(3.1)*	0.446	0.089	*(5.0)*	0.899	0.055	*(16.3)*	0.263
Flexibility				0.731	0.288	*(2.5)*	0.877	0.242	*(3.6)*	0.366
Ease				0.571	0.168	*(3.4)*	1.15	0.09	*(12.1)*	0.188
Safety	0.381	0.135	*(2.8)*	0.132	0.117	*(1.1)*	0.803	0.081	*(10.0)*	0.197
Overall Rating	1.25	0.82	*(1.5)*	1.39	0.51	*(2.7)*	1.28	0.26	*(5.0)*	0.616

The likelihood is a 3 dimensional integral: 1 for the agent effect and 1 for each latent variable. To estimate the model, we substitute the structural equation throughout, and the likelihood function is then an integral over 3 independent standard normal distributions.

The results of the model are provided in Table 4-5, and again we provide the base RP/SP model (shaded) for comparison.[33] In this case, the latent variables of comfort and convenience are borderline significant in the choice model (t-stats of 1.0 to 2.3). The latent variable model appears to reasonably capture the latent constructs, and it does add richness to the behavioral process represented by the model. However, the impact is certainly not overwhelming. We also report the log-likelihood for just the choice model portion of the joint model. Note that there are various ways to calculate this log-likelihood. What we report is the case in which the latent variable score and distribution are extracted using the structural equation only (a partial information extraction), and then the log-likelihood of the choice model is calculated given this information. This method is representative of the forecasting process, in which the measurement equation is not used (since the indicators are not known). The log-likelihood actually increases slightly over the base choice model. The decrease in fit for the choice model portion does not necessarily mean that the joint model is inferior. First, a full information extraction method (using both the structural and measurement equations) would improve the fit of the choice model portion (particularly since, in this case, the structural model is relatively weak.) Second, it is not surprising that the likelihood increases slightly, because we compare a value that is already optimized to the choice data (the base choice model) versus a value that is optimized to both the choice and indicator data, i.e., the comparison is made across different metrics. As long as the parameters for the latent variables in the choice model are significant, then the latent variable portion is bringing some explanation to the model. The best method to determine the magnitude of the benefits of the joint choice and latent variable model is to perform forecasting tests using either a hold out sample or real data.

Latent Class Model

For the latent class mode choice model, we estimate a model that is analogous to the random parameter model presented in Table 4-3. However, instead of representing the unobserved heterogeneity with random parameters, we specify that there are two distinct classes of people, each with its own set of parameters for the 5 attributes of the alternatives. Parameters other than those for the 5 attributes are common across the classes. The likelihood is as follows:

$$P(y_n^{RP}, y_n^{SP1}, y_n^{SP2} \mid X_n) = \sum_{s=1}^{2} P(y_n^{RP}, y_n^{SP1}, y_n^{SP2} \mid X_n, s)\Lambda(s \mid X_n) ,$$

where $P(y_n^{RP}, y_n^{SP1}, y_n^{SP2} \mid X_n)$ is as in Equation [4-8], with the exception that there are a different set of parameters for each class, and $\Lambda(s \mid X_n)$ is a binary logit model.

---

[33] There are 3 additional explanatory variables in the choice and latent variable model (age dummy, first class rail rider, and free parking), which enter the latent variable structural equations. These variables were tested in the base RP/SP model and are not significant (t-stats of 0.9, 0.4, and 0.2, respectively).

The estimation results are presented in Table 4-6. The model suggests that there are at least two classes. Class one is defined by younger travelers, recreational travelers, and people traveling in groups who are more sensitive to cost, in-vehicle time, and transfers. Class two is defined by business travelers and older travelers who are more sensitive to out-of-vehicle time and amenities. The sample is skewed towards class 2 as the class membership statistics show at the bottom of the table. The 2 latent classes do provide a significant improvement in fit over the base model, but the fit of the model falls well below that captured by the random parameter model. This is not surprising since we did not have strong behavioral justification for two distinct segments of the travelers, and therefore a continuous distribution provides more explanatory power.

Combination Models

The estimation results thus far have provided examples of integrating joint RP/SP models with random parameters, latent variable, and latent class, individually. Here we provide examples of further combinations. Ideally, one would like to have strong behavioral justification or motivation to introduce more complexity. In the case of our mode choice example, we really do not. Our objectives of further integrating the model are to both improve the overall fit and behavioral representation of the model, as well as to strengthen the relationship between the latent variables and the choice model. Several models are presented below.

- *Choice and Latent Variable Model with Latent Class Heterogeneity of Mode Attributes*

  provides estimation results for a model that is a direct combination of the choice and latent variable model presented in Table 4-5 and the latent class model presented in Table 4-6. The generalized model now captures the latent concepts of comfort and convenience, as well as the unobserved heterogeneity represented by the latent class structure

- *Choice and Latent Variable Model with Random Parameters*

  Table 4-8 provides results for the latent and choice variable model in which we have added random parameters to both the choice model portion and the structural equations of the latent variable model. To keep the dimension of the integral down and to avoid potential multicollinearity issues, it is important to be selective in terms of the parameters that are distributed. We selected 4 parameters in the choice model (those with the most significant distributions from the random parameter model presented in Table 4-3) and 3 parameters in the structural equations (those with highest significance in the fixed parameter model presented in Table 4-5). There is a significant improvement in the overall fit of the model. However, again, the latent variables have only a marginal impact on the choice model.[34]

---

[34] Note that the original estimate of this model was empirically unidentified (the parameters trended away from zero), and so the parameter corresponding to the RP agent effect is constrained to be equal to 1.

- *Choice and Latent Variable Model with Unobserved Heterogeneity of Latent Variable Parameters*

  In an effort to strengthen the relationship between the latent variable constructs of comfort and convenience with the choice model, we experimented with unobserved heterogeneity of the parameters for the latent variables in the choice model. First we specified the parameters to be lognormally distributed. Next we specified the parameters as having latent heterogeneity defined by a two-class structure, in which the latent variables only impact one of the classes. We used a naïve class membership model, because a richer specification proved to have identification problems. The results for the choice model portion of both of these models are shown in Table 4-9. We did not report the estimation results for the latent variable model, because they are very close to the results reported in Table 4-5. Note that neither approach significantly impacted the choice model. Therefore to improve the specification, the latent variable model probably needs major reworking. One possibility is to not specify the latent variables in their differenced form (rail-auto), and therefore specify the measurement equations as having discrete indicators. Another possibility is to specify different latent variable models for different latent classes. Early experimentation with this latter approach showed some promise.

Conclusion

We presented a flexible, powerful framework that incorporates key extensions to discrete choice models. The experimental results we have provided using the mode choice dataset explored various specifications and demonstrated the practicality of the generalized model. The conclusions from the application of the generalized model to the mode choice case study are that introducing stated preferences and random taste variation greatly improves the specification of the model, whereas latent variables and latent classes had less significant impacts. It is important to note that we cannot draw conclusions on the various methods from the series of estimation results presented in this chapter. The results will vary based on the application and data. For example, in contrast to the results we presented here, we have had cases in which a latent class model outperforms a random parameter specification, and also have had cases in which the latent variable model has a large impact on the choice model.

Table 4-6: Latent Class Mode Choice Model

MODE CHOICE MODEL

Parameter	Base RP/SP Model Est.	Base RP/SP Model Std. Er.	Base RP/SP Model t-stat	Latent Class Model: Parameters Common Across Classes Est.	Std. Er.	t-stat	Parameters Unique to Class 1 Est.	Std. Er.	t-stat	Parameters Unique to Class 2 Est.	Std. Er.	t-stat
Rail constant RP	0.444	0.493	*(0.9)*	1.26	0.756	*(1.7)*						
Rail constant SP2	-0.466	0.777	*(0.6)*	1.42	0.772	*(1.8)*						
Work trip dummy	1.17	0.51	*(2.3)*	1.10	0.620	*(1.8)*						
Fixed arrival time dummy	0.723	0.381	*(1.9)*	0.641	0.497	*(1.3)*						
Female dummy	0.990	0.381	*(2.6)*	1.03	0.432	*(2.4)*						
Cost per person in Guilders	-0.0608	0.0132	*(4.6)*				-0.231	0.063	*(3.7)*	-0.0408	0.0115	*(3.5)*
Out-of-vehicle time in hours	-2.23	0.83	*(2.7)*				-1.31	1.21	*(1.1)*	-3.47	1.34	*(2.6)*
In-vehicle time in hours	-0.710	0.158	*(4.5)*				-1.69	0.48	*(3.5)*	-0.876	0.244	*(3.6)*
Number of transfers	-0.100	0.036	*(2.8)*				-0.216	0.092	*(2.3)*	-0.149	0.055	*(2.7)*
Amenities	-0.361	0.080	*(4.5)*				-0.408	0.114	*(3.6)*	-0.540	0.146	*(3.7)*
Inertia dummy (RP Choice)	2.97	1.02	*(2.9)*	0.99	0.696	*(1.4)*						
Agent effect RP	0.686	0.490	*(1.4)*	2.09	0.76	*(2.8)*						
Agent effect SP2	2.44	0.50	*(4.9)*	2.87	0.73	*(3.9)*						
Scale (mu) SP1	2.31	0.50	*(4.6)*	2.25	0.59	*(3.8)*						
Scale (mu) SP2	1.31	0.30	*(4.4)*	1.56	0.35	*(4.5)*						
Tau1 SP1 (=-Tau4 SP1)	*-0.195*	----	----	-0.236	----	----						
Tau2 SP1 (=-Tau3 SP1)	*-0.0127*	----	----	-0.0154	----	----						
Tau3 SP1	0.0127	0.0036	*(3.5)*	0.0154	0.0050	*(3.1)*						
Tau4 SP1	0.195	0.049	*(4.0)*	0.236	0.070	*(3.4)*						
Tau1 SP2	-0.986	0.219	*(4.5)*	-0.895	0.210	*(4.3)*						
Tau2 SP1 (=-Tau3 SP2)	*-0.180*	----	----	-0.161	----	----						
Tau3 SP2	0.180	0.053	*(3.4)*	0.161	0.051	*(3.1)*						
Tau4 SP2	1.32	0.32	*(4.1)*	1.17	0.28	*(4.2)*						
Number of observations:	4680			4680								
Number of draws (Halton):	1000			1000								
Log-likelihood:	-4517.43			-4283.04								
Rho-bar-squared:	0.380			0.411								

CLASS MEMBERSHIP MODEL

Parameter	Est.	Std. Er.	t-stat	*Class Membership Statistics*
Constant	-0.455	0.395	*(1.2)*	Probability(Class 1) < 0.2 for 16% of the sample
Female dummy	-0.0832	0.3625	*(0.2)*	0.2 <= Probability(Class 1) < 0.4 for 19% of the sample
Number of persons in party	0.174	0.121	*(1.4)*	0.4 <= Probability(Class 1) < 0.6 for 62% of the sample
Work trip dummy	-1.94	0.73	*(2.7)*	0.6 <= Probability(Class 1) < 0.8 for 3% of the sample
Age over 40 dummy	-0.472	0.371	*(1.3)*	Probability(Class 1) >= 0.6 for 0% of the sample

Table 4-7: Choice and Latent Variable Mode Choice Model with Latent Classes

MODE CHOICE MODEL

	Parameters Common Across Classes			Parameters Unique to Class 1			Parameters Unique to Class 2		
Parameter	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat
Rail constant RP	0.293	0.905	*(0.3)*						
Rail constant SP2	0.940	1.143	*(0.8)*						
Work trip dummy	1.96	1.26	*(1.6)*						
Fixed arrival time dummy	0.590	0.609	*(1.0)*						
Female dummy	1.04	0.53	*(2.0)*						
Cost per person in Guilders				-0.220	0.104	*(2.1)*	-0.0406	0.0196	*(2.1)*
Out-of-vehicle time in hours				0.0541	1.5606	*(0.0)*	-2.27	1.64	*(1.4)*
In-vehicle time in hours				-1.61	0.76	*(2.1)*	-0.909	0.375	*(2.4)*
Number of transfers				-0.180	0.127	*(1.4)*	-0.167	0.079	*(2.1)*
Amenities				-0.415	0.166	*(2.5)*	-0.566	0.241	*(2.3)*
Latent Comfort - RP	1.32	0.69	*(1.9)*						
Latent Comfort - SP2	1.62	0.53	*(3.0)*						
Latent Convenience - RP	1.90	1.04	*(1.8)*						
Latent Convenience - SP2	1.32	0.61	*(2.2)*						
Inertia dummy (RP Choice)	0.0277	1.0414	*(0.0)*						
Agent effect RP	2.24	1.61	*(1.4)*						
Agent effect SP2	2.73	1.13	*(2.4)*						
Scale (mu) SP1	2.21	0.92	*(2.4)*						
Scale (mu) SP2	1.38	0.43	*(3.2)*						
Tau1 SP1 (=-Tau4 SP1)	-0.242	----	----						
Tau2 SP1 (=-Tau3 SP1)	-0.0157	----	----						
Tau3 SP1	0.0157	0.0070	*(2.2)*						
Tau4 SP1	0.242	0.111	*(2.2)*						
Tau1 SP2	-1.00	0.32	*(3.1)*						
Tau2 SP1 (=-Tau3 SP2)	-0.181	----	----						
Tau3 SP2	0.181	0.071	*(2.5)*						
Tau4 SP2	1.31	0.43	*(3.0)*						
Number of observations:	4680								
Number of draws (Halton):	5000								
Log-likelihood (Choice&Latent):	-6423.09								
Log-likelihood (Choice):	-4284.96								
Rho-bar-squared (Choice):	0.412								

LATENT VARIABLE MODEL

Structural Equations (2 equations, 1 per column)

	Comfort Equation			Convenience Equation		
Parameter	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat
Constant - Comfort	0.132	0.158	*(0.8)*			
Constant - Convenience				0.497	0.245	*(2.0)*
Age dummy - over 40	-0.540	0.400	*(1.4)*	0.876	0.246	*(3.6)*
First class rail rider	0.454	0.402	*(1.1)*			
In-vehicle time in hours	-0.519	0.324	*(1.6)*			
Out-of-vehicle time in hours				-1.23	0.54	*(2.3)*
Number of transfers				-0.107	0.156	*(0.7)*
Free parking dummy (auto)				0.218	0.259	*(0.8)*
Variance(ω)	1.00	----	----	1.00	----	----
Squared Multiple Correlation (SMC)	0.115			0.236		

Measurement Equations (6 equations, 1 per row)

	Comfort Parameters			Convenience Parameters			Disturbance Params. (StdDev(υ))			Fit (SMC)
Equation	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat	
Relaxation	0.551	0.183	*(3.0)*	0.156	0.134	*(1.2)*	1.15	0.10	*(11.5)*	0.194
Reliability	0.343	0.106	*(3.2)*	0.462	0.090	*(5.1)*	0.887	0.055	*(16.0)*	0.282
Flexibility				0.716	0.171	*(4.2)*	0.892	0.139	*(6.4)*	0.352
Ease				0.570	0.128	*(4.4)*	1.15	0.09	*(13.5)*	0.187
Safety	0.377	0.092	*(4.1)*	0.153	0.103	*(1.5)*	0.800	0.051	*(15.6)*	0.201
Overall Rating	1.10	0.38	*(2.9)*	1.44	0.26	*(5.5)*	1.37	0.18	*(7.7)*	0.579

CLASS MEMBERSHIP MODEL

Parameter	Est.	Std. Er.	t-stat	*Class Membership Statistics*
Constant	-0.375	0.467	*(0.8)*	Probability(Class 1) < 0.2 for 16% of the sample
Female dummy	0.0489	0.4128	*(0.1)*	0.2 <= Probability(Class 1) < 0.4 for 18% of the sample
Number of persons in party	0.165	0.125	*(1.3)*	0.4 <= Probability(Class 1) < 0.6 for 60% of the sample
Work trip dummy	-1.85	0.74	*(2.5)*	0.6 <= Probability(Class 1) < 0.8 for 5% of the sample
Age over 40 dummy	-0.496	0.384	*(1.3)*	Probability(Class 1) >= 0.8 for 0% of the sample

Table 4-8: Choice and Latent Variable Mode Choice Model with Random Parameters

CHOICE MODEL

Parameter	Location Parameters Est.	Std. Er.	t-stat	Distribution Parameters Est.	Std. Er.	t-stat	
Rail constant RP	0.100	0.796	*(0.1)*				
Rail constant SP2	1.53	0.67	*(2.3)*				
Work trip dummy	1.07	0.83	*(1.3)*				
Fixed arrival time dummy	0.397	0.651	*(0.6)*				
Female dummy	1.48	0.63	*(2.4)*				
Cost per person in Guilders	-2.18	0.29	*	1.02	0.05	*(22.1)*	*lognormal*
Out-of-vehicle time in hours	0.06	0.94	*(0.1)*				
In-vehicle time in hours	0.228	0.305	*	0.864	0.040	*(21.5)*	*lognormal*
Number of transfers	-2.14	0.38	*	1.76	0.15	*(12.0)*	*lognormal*
Amenities	-0.609	0.271	*	1.13	0.05	*(22.3)*	*lognormal*
Latent Comfort - RP	2.98	0.84	*(3.5)*				
Latent Comfort - SP2	3.08	0.87	*(3.5)*				
Latent Convenience - RP	1.54	0.37	*(4.2)*				
Latent Convenience - SP2	1.18	0.37	*(3.2)*				
Inertia dummy (RP Choice)	-1.05	0.57	*(1.8)*				
Agent effect RP	1.00	----	----				
Agent effect SP2	1.84	0.53	*(3.5)*				
Scale (mu) SP1	4.28	1.24	*(3.5)*				
Scale (mu) SP2	2.03	0.55	*(3.7)*				
Tau1 SP1 (=-Tau4 SP1)	-0.229	----	----				
Tau2 SP1 (=-Tau3 SP1)	-0.0152	----	----				
Tau3 SP1	0.0152	0.0053	*(2.9)*				
Tau4 SP1	0.229	0.083	*(2.8)*				
Tau1 SP2	-0.812	0.220	*(3.7)*				
Tau2 SP1 (=-Tau3 SP2)	-0.143	----	----				
Tau3 SP2	0.143	0.049	*(2.9)*				
Tau4 SP2	1.03	0.28	*(3.7)*				
Number of observations:	4680						
Number of draws (Halton):	20000						
Log-likelihood (Choice&Latent):	-6066.08						
Log-likelihood (Choice):	-3935.04						
Rho-bar-squared (Choice):	0.458						

LATENT VARIABLE MODEL

Structural Equations (2 equations, 1 per column)

	Comfort Equation						Convenience Equation						
	Location Parameters			Distribution Parameters			Location Parameters			Distribution Parameters			
Parameter	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat	
Constant - Comfort	0.0688	0.1362	*(0.5)*										
Constant - Convenience							0.649	0.239	*(2.7)*				
Age dummy - over 40	-0.435	0.145	*(3.0)*				0.961	0.286	*(3.4)*	-0.281	0.072	*(3.9)*	*normal*
First class rail rider	-0.434	0.211	*(2.1)*										
In-vehicle time in hours	-3.03	0.43	*	1.964	0.154	*(12.7)*							*lognormal*
Out-of-vehicle time in hours							0.246	0.386	*	-0.674	0.133	*(5.1)*	*lognormal*
Number of transfers							-0.294	0.126	*(2.3)*				
Free parking dummy (auto)							0.147	0.180	*(0.8)*				
Variance(ω)	1.00	----	----				1.00	----	----				

Measurement Equations (6 equations, 1 per row)

	Comfort Parameters			Convenience Parameters			Disturbance Params. (StdDev(υ))		
Equation	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat
Relaxation	0.408	0.138	*(3.0)*	0.136	0.084	*(1.6)*	1.20	0.07	*(16.4)*
Reliability	0.220	0.100	*(2.2)*	0.402	0.072	*(5.6)*	0.896	0.052	*(17.1)*
Flexibility				0.603	0.109	*(5.6)*	0.870	0.087	*(10.0)*
Ease				0.453	0.085	*(5.3)*	1.16	0.07	*(15.8)*
Safety	0.242	0.095	*(2.5)*	0.152	0.069	*(2.2)*	0.838	0.044	*(19.1)*
Overall Rating	1.05	0.13	*(8.1)*	1.12	0.12	*(9.0)*	1.39	0.14	*(10.0)*

Table 4-9: Choice and Latent Variable Models with Heterogeneity of Latent Variable Parameters

CHOICE MODEL (Latent Variable Portion not Shown)

	Choice and Latent Variable RP/SP Model with Randomly Distributed Parameters (Lognormal)						Choice and Latent Variable RP/SP Model with Latent Class Heterogeneity					
Parameter	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat
Rail constant RP	-0.390	0.707	*(0.6)*				-0.391	0.722	*(0.5)*			
Rail constant SP2	-0.856	0.748	*(1.1)*				-0.908	0.778	*(1.2)*			
Work trip dummy	1.76	0.74	*(2.4)*				1.72	0.66	*(2.6)*			
Fixed arrival time dummy	0.707	0.504	*(1.4)*				0.702	0.520	*(1.4)*			
Female dummy	1.16	0.48	*(2.4)*				1.17	0.48	*(2.4)*			
Cost per person in Guilders	-0.0637	0.0165	*(3.9)*				-0.0635	0.0174	*(3.7)*			
Out-of-vehicle time in hours	-1.09	0.88	*(1.2)*				-1.14	0.99	*(1.2)*			
In-vehicle time in hours	-0.728	0.192	*(3.8)*				-0.726	0.198	*(3.7)*			
Number of transfers	-0.103	0.040	*(2.6)*				-0.103	0.041	*(2.5)*			
Amenities	-0.377	0.100	*(3.8)*				-0.376	0.104	*(3.6)*			
	Location Parameters			Distribution Parameters			Class 1 Parameters			Class 2 Parameters		
Latent Comfort - RP	0.161	0.699	*	0.187	0.787	*(0.2)*	1.34	0.94	*(1.4)*	0.000	-----	-----
Latent Comfort - SP2	0.186	0.391	*	0.340	0.079	*(4.3)*	1.42	0.63	*(2.3)*	0.000	-----	-----
Latent Convenience - RP	0.267	0.467	*	0.314	0.511	*(0.6)*	1.48	0.61	*(2.4)*	0.000	-----	-----
Latent Convenience - SP2	-0.252	0.359	*	0.214	0.115	*(1.9)*	0.834	0.366	*(2.3)*	0.000	-----	-----
Inertia dummy (RP Choice)	2.56	1.07	*(2.4)*				2.62	1.21	*(2.2)*			
Agent effect RP	0.256	0.566	*(0.5)*				0.125	0.571	*(0.2)*			
Agent effect SP2	2.10	0.61	*(3.5)*				2.12	0.66	*(3.2)*			
Scale (mu) SP1	2.20	0.58	*(3.8)*				2.21	0.61	*(3.6)*			
Scale (mu) SP2	1.26	0.38	*(3.3)*				1.24	0.41	*(3.0)*			
Tau1 SP1 (=-Tau4 SP1)	-0.204	----	----				-0.204	----	----			
Tau2 SP1 (=-Tau3 SP1)	-0.0133	----	----				-0.0132	----	----			
Tau3 SP1	0.0133	0.0043	*(3.1)*				0.0132	0.0044	*(3.0)*			
Tau4 SP1	0.204	0.060	*(3.4)*				0.204	0.062	*(3.3)*			
Tau1 SP2	-1.03	0.31	*(3.4)*				-1.05	0.34	*(3.1)*			
Tau2 SP1 (=-Tau3 SP2)	-0.189	----	----				-0.192	----	----			
Tau3 SP2	0.189	0.064	*(3.0)*				0.192	0.070	*(2.7)*			
Tau4 SP2	1.39	0.43	*(3.2)*				1.41	0.49	*(2.9)*			
Number of observations:	4680						4680					
Number of draws (Halton):	10000						10000					
Log-likelihood (Choice&Latent):	-6655.79						-6655.96					
Log-likelihood (Choice):	-4518.08						-4518.19					
Rho-bar-squared (Choice):	0.379						0.380					

CLASS MEMBERSHIP MODEL

Parameter	Est.	Std. Er.	t-stat
Constant	2.50	1.39	*(1.8)*
	Probability (Class 1) = 92%		

Summary of Latent Variable Parameters from the Different Models

Model:	Base Model	Random Parameter Model		Latent Class Model	
Parameter	Fixed	Mean	Std. Dev.	Class 1	Class 2
Latent Comfort - RP	1.16	1.20	0.23	1.34	0.000
Latent Comfort - SP2	1.16	1.28	0.45	1.42	0.000
Latent Convenience - RP	1.30	1.37	0.44	1.48	0.000
Latent Convenience - SP2	0.764	0.795	0.173	0.834	0.000

Chapter 5:
Conclusion

Summary

We started the discussion by pointing out the gap between traditional discrete choice model and behavioral theory, which is depicted in Figure 5-1. Researchers have long been working on a host of different enhancements to improve the performance of discrete choice model. These new techniques are mostly explored and applied in isolation from one another. In order to develop models that are behaviorally realistic, reflecting anything close to the complexity depicted in Figure 5-1, we must draw on a toolbox of methodologies. To meet this end, we proposed an generalized discrete choice modeling framework (Figure 5-2) that incorporates key extensions to the discrete choice model, including:

- The ability to represent any desirable (additive) error structure via the parameterized disturbance with factor analytic form, enabling us to relax the IIA restriction as well as represent unobserved heterogeneity, for example, in the form of random parameters;
- The use of different behavioral indicators, including revealed preferences, stated preferences, and psychometric data, all of which provide insight on the choice process;
- The capability of explicitly modeling the formation of important latent behavioral constructs, such as attitudes and perceptions, and their effect on the choices; and
- The capacity to represent latent segmentation of the population (or multimodal behavior, for example, leisure or rushed time) as well as the respective tastes, decision protocols, and choice sets of each segment.

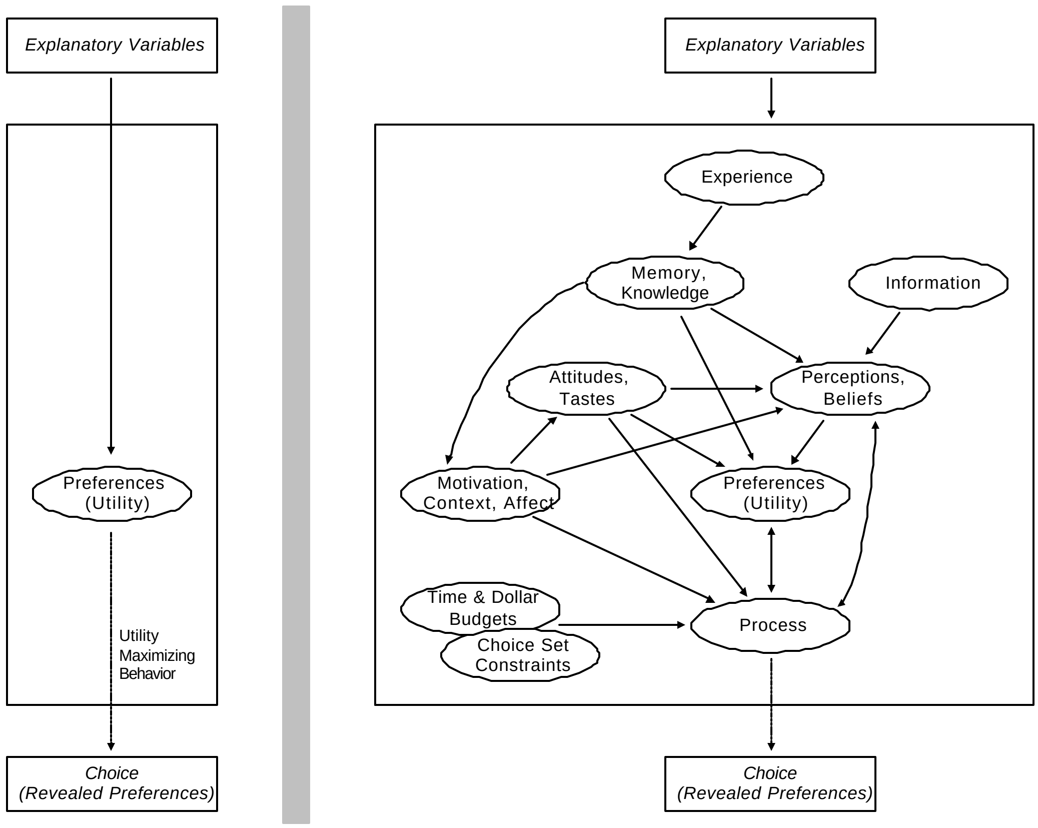


Figure 5-1: The Gap Between Traditional Discrete Choice Theory and Behavior

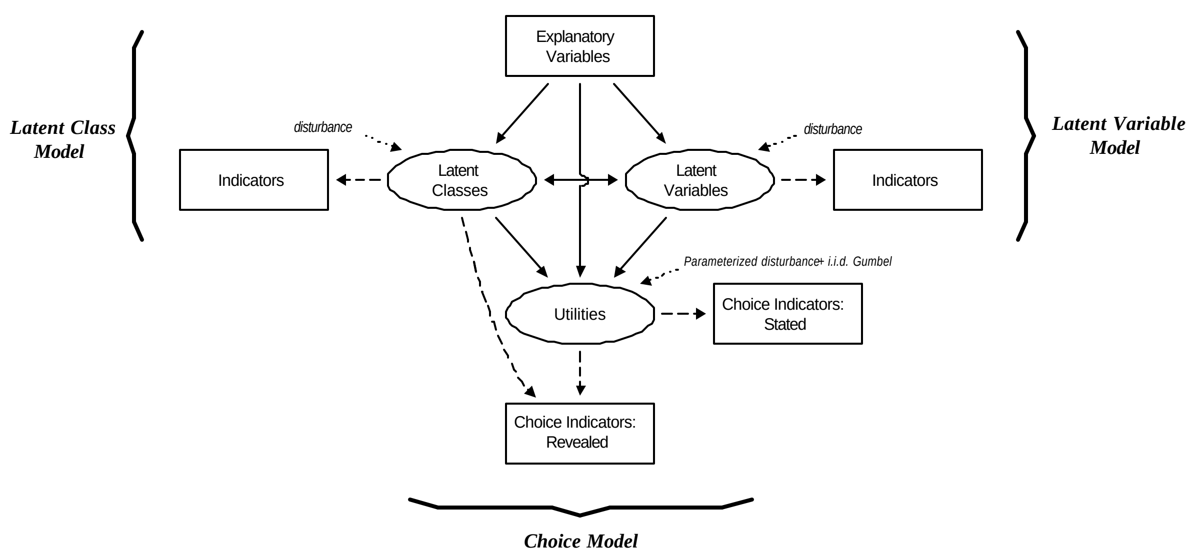


Figure 5-2: The Generalized Discrete Choice Framework

The basic integration technique that is recommended is to start with multinomial logit formulation, and then add extensions that relax simplifying assumptions and enrich the capabilities of the basic model. This technique results in a 'logit kernel' formulation of the model, and leads to a straightforward probability simulator for use in maximum simulated likelihood estimation.

We provided estimation results using a mode choice application to demonstrate and test the use and practicality of the generalized model. Some of our applications result in large improvements in fit as well as a more satisfying behavioral representation. However, in some cases the extensions have no impact on the choice model. These latter cases provide the valuable information that the parsimonious structures are robust.

In addition to the overall generalized model, we also provided expanded coverage of two of the key methodologies that make up the generalized framework. The first methodology that we emphasized was the logit kernel model (Figure 5-3), which is a discrete choice model in which the disturbance is composed of a probit-like multivariate normal (or other) distributed term and an i.i.d Gumbel term. We showed that a factor analytic specification of the disturbances can be used to specify all known (additive) error structures, including heteroscedasticity, nested and cross-nested structures, and random parameters. The inclusion of the i.i.d Gumbel term leads to a convenient smooth probability simulator, which allows for straightforward estimation via maximum simulated likelihood. A key contribution is our investigation of the normalization and identification of logit kernel models. We found that it is not necessarily intuitive, and the rules can differ from those for the systematic portion of the utility as well as those for analogous probit models. We established specific rules of normalization and identification for many of the most common forms of the logit kernel model. We also presented empirical results that highlighted various specification and identification issues.

The second emphasized methodology was the development of a general framework and methodology for incorporating latent variables into choice models. The framework is shown in Figure 5-4; it is essentially the integration of the latent variable methodologies developed by psychometricians and a classic discrete choice model. This method is critical for developing behaviorally realistic models, because so many of the constructs that cognitive researchers emphasize as being essential to the choice process (for example, the ovals in Figure 5-1) cannot be directly measured. However, we can build surveys that gather psychometric data on all aspects of the choice process, and then use these data to aid in specifying the structural equations of the choice model.

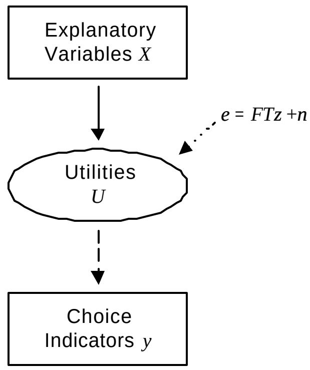


Figure 5-3: Emphasized Methodology I –
Factor Analytic Parameterized Disturbance with Logit Kernel

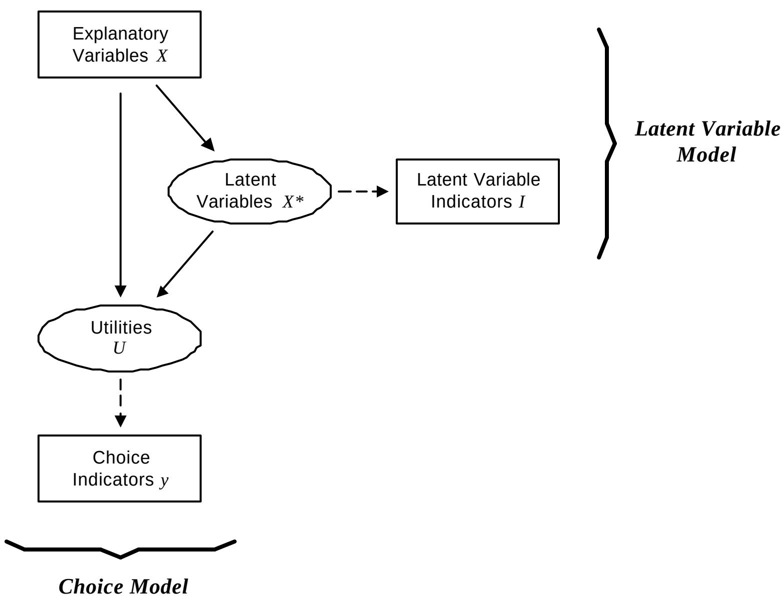


Figure 5-4: Emphasized Methodology II –
Integration of Choice and Latent Variable Models

Research Directions

The methodology presented here and the empirical case studies have merely brought to the surface the potential for the generalized modeling framework. These are relatively new, untested methods, and they require further investigation into numerous issues, including:

Applications: The first issue is simply that more testing and experience with applications are necessary to uncover related issues and to better understand the potential of the generalized framework.

Validation: Thus far for validation we have looked at aspects such as the goodness of fit, significance of the parameters that are part of the extensions, and simply examining the behavioral process represented by the model structure. The findings so far suggest that the advanced methodologies provide promise. Now more work needs to be done in conducting validation tests, including tests of forecasting ability, consequences of misspecifications (for example, excluding latent variables or heterogeneity that should be present), and performance comparisons with models of simpler formulations.

Identification: There is a need for further exploration of identification and normalization issues, including pursuit of general necessary and sufficient rules for identification as well as continued compilation and analysis of special cases and rules of identification. Also more fundamental identification issues of identification need to be explored related to, for example, the shape of the objective function.

Comparison of Various Approaches for Estimation, Simulation, and Specification: One of the things we do in this dissertation is suggest a particular modeling approach in terms of estimation (maximum simulated likelihood), simulation (Halton draws and logit kernel), and specification (the use of the factor analytic disturbance to reflect the covariance structure). We suggest these approaches because it leads to a flexible, tractable, practical, and intuitive method for incorporating complex behavioral processes in the choice model. However, there are alternative approaches in each of these directions, method of simulated moments; other types of pseudo- and quasi-random draws; semi-parametric approaches; empirical Bayes estimation (versus classic techniques); probit and the GHK simulator; classic nested, cross-nested, and heteroscedastic logit formulations; and many more. We need a better understanding of the relationships among various techniques, and the implications of various specifications. In addition, investigations into new techniques such as the Combined Logit Probit model described in Chapter 2 would be valuable.

Dynamics: We have not directly addressed the issue of dynamics in this dissertation, although dynamics is clearly a critical aspect of behavior. With the existing generalized framework, the choice indicators could be of panel data form, and it is then relatively straightforward to introduce standard dynamic choice modeling techniques into the framework. However, a more elusive issue is that of feedback, which is very prevalent in behavioral theory, and more thought needs to be put into this area.

Computation: Applying these methods are computationally intensive. Estimation time varies significantly with the particular application. The models presented in Chapters 2 and 4 were estimated using 550-733 MHz Pentium II processors. Depending on the specification, they took on the order of either hours or days to estimate. For example, the telephone service models in Chapter 2, which involve only 434 observations, took less than an hour. The mode choice models presented in Chapter 4 that did not involve random parameters took on the order of several hours (more observations than the telephone dataset and also a more complex logit kernel with its RP/SP specification). The models using synthetic datasets in Chapter 2 (which have 10,000 observations) and the random parameter mode choice models in Chapter 4 took on the order of a day (24 hours) to estimate. Furthermore, all of the models presented in this dissertation are relatively small in terms of the number of observations and number of alternatives, and therefore the estimation time for real applications could easily extend to over a week. Due to the long estimation times, investigation into techniques such as parallel computing, for which simulation is a perfect application, would greatly ease the application of such models.

Data: One of the key ideas of the generalized framework is to make use of various types of indicators that can provide insight on the choice process. These include the revealed preferences, stated preferences, and attitudinal and perceptual indicators that are dealt with in some detail in this document. More generally, it includes any type of verbal or other indicator for the behavioral process depicted in Figure 5-1, including, for example, verbal descriptions of decision protocols. Cognitive researchers as well as others have long investigated data collection and surveys, and this research needs to be synthesized in conjunction with the behavioral framework and generalized methodological framework.

Behavioral Framework: Our focus throughout the dissertation has been on methodological tools and not on the substantive issues in psychology and behavioral sciences. The generalized framework provides potential to reflect behavioral theory that has here-to-for primarily existed in descriptive, flow-type models. Clearly, application requires careful consideration of the behavioral framework, strong behavioral justification for the added complexity and, ideally, design of a data collection effort that generates good indicators and causal variables to support the framework.

Conclusion

Behavior is clearly complex, and the basic discrete choice model is a simplistic representation of this behavior. We have the tools available to improve the behavioral representation of models by integrating methods that exploit the use of different types of data, capture unobserved heterogeneity for all aspects of the choice process, and explicitly model behavioral constructs such as attitudes and perceptions. With increasing computational power and increasingly rich datasets, techniques such as those described in this dissertation can be practically applied and offer great potential to better understand behavior and test behavioral hypothesis, instill confidence in parsimonious specifications, and improve forecasts. The approach presented in this dissertation is a flexible, powerful, practical, intuitive, and theoretically grounded methodology that allows the modeling of complex behavioral processes.

There are still practical questions. How much of a difference do these techniques make? It is highly dependent on the question being asked, the behavior being modeled, the strength of the behavioral framework, and the quality of the data. Do we really need to capture the inner workings of the black box if we are only interested in the bottom line choices? It is certainly debatable. However, the best way, and perhaps the only way, to answer the question is to explore more behaviorally realistic models and then compare their performance against parsimonious specifications.

The best test of this framework would be to start with a situation in which there are strong behavioral hypotheses and objectives for the modeling; then develop a methodological framework that represents the assumed behavior (making use of the various methodologies and potential data sources); then develop a data collection plan to gather data that supports the framework, and then estimate a series of models to test the impact of various levels of complexity. The problem is, that each of these four stages is difficult: we are dealing with behavior like in Figure 5-1 and complex equations.

The bottom line is that we need to continue to explore, and the answers lie in bringing together the techniques and expertise of econometricians, psychometricians, cognitive researchers, and market analysts.

Appendix A: Normalization of Unrestricted Probit and Logit Kernel Covariance Structures

This appendix examines the normalization of unrestricted probit and logit kernel models. The important point is that while the normalization of pure probit leads to straightforward scale shifts of all of the parameter estimates, this is not the case for logit kernel.

Case 1: Probit with 4 Alternatives

Using the notation from Chapter 2, the unrestricted four alternative probit model written in differenced form has the error structure Tz_n, where:

$$T = \begin{bmatrix} a_{11}/\tilde{m} & 0 & 0 \\ a_{21}/\tilde{m} & a_{22}/\tilde{m} & 0 \\ a_{31}/\tilde{m} & a_{32}/\tilde{m} & a_{33}/\tilde{m} \end{bmatrix}$$

Note that we use a 's instead of s 's since these aren't variance terms. Also $\tilde{m}$ is the scale of the probit model (i.e., not the traditional Gumbel m).

The covariance structure is then (using new notation):

$$TT' \; theoretical: \begin{bmatrix} (a_{11}^2)/\tilde{m}^2 & & \\ (a_{11}a_{21})/\tilde{m}^2 & (a_{21}^2 + a_{22}^2)/\tilde{m}^2 & \\ (a_{11}a_{31})/\tilde{m}^2 & (a_{21}a_{31} + a_{22}a_{32})/\tilde{m}^2 & (a_{31}^2 + a_{32}^2 + a_{33}^2)/\tilde{m}^2 \end{bmatrix}$$

A normalization must be made in order to achieve identification. Normalizing $a_{33} = a_{ff}^N$, and noting the unknown parameters as a and m, then the normalized covariance structure is:

$$TT'\ normalized: \begin{bmatrix} \left(a_{11}^N\right)^2/\tilde{m}_N^2 & & \\ \left(a_{11}^N a_{21}^N\right)/\tilde{m}_N^2 & \left(\left(a_{21}^N\right)^2+\left(a_{22}^N\right)^2\right)/\tilde{m}_N^2 & \\ \left(a_{11}^N a_{31}^N\right)/\tilde{m}_N^2 & \left(a_{21}^N a_{31}^N + a_{22}^N a_{32}^N\right)/\tilde{m}_N^2 & \left(\left(a_{31}^N\right)^2+\left(a_{32}^N\right)^2+\left(a_{ff}^N\right)^2\right)/\tilde{m}_N^2 \end{bmatrix}$$

Setting $TT'\ normalized = TT'\ theoretical$, leads to the following equations:

$$\left(a_{11}^N\right)^2/\tilde{m}_N^2 = \left(a_{11}\right)^2/\tilde{m}^2$$

$$\left(a_{11}^N a_{21}^N\right)/\tilde{m}_N^2 = \left(a_{11}a_{21}\right)/\tilde{m}^2$$

$$\left(a_{11}^N a_{31}^N\right)/\tilde{m}_N^2 = \left(a_{11}a_{31}\right)/\tilde{m}^2$$

$$\left(\left(a_{21}^N\right)^2+\left(a_{22}^N\right)^2\right)/\tilde{m}_N^2 = \left(\left(a_{21}\right)^2+\left(a_{22}\right)^2\right)/\tilde{m}^2$$

$$\left(a_{21}^N a_{31}^N + a_{22}^N a_{32}^N\right)/\tilde{m}_N^2 = \left(a_{21}a_{31}+a_{22}a_{32}\right)/\tilde{m}^2$$

$$\left(\left(a_{31}^N\right)^2+\left(a_{32}^N\right)^2+\left(a_{ff}^N\right)^2\right)/\tilde{m}_N^2 = \left(\left(a_{31}\right)^2+\left(a_{32}\right)^2+\left(a_{33}\right)^2\right)/\tilde{m}^2$$

And solving for each of the unknown parameters in the normalized model leads to:

Solution: $\left(a_{11}^N\right)^2 = \left(a_{11}\right)^2\dfrac{\tilde{m}_N^2}{\tilde{m}^2} \rightarrow a_{11}^N = a_{11}\dfrac{\tilde{m}_N}{\tilde{m}}$

$$a_{21}^N = \frac{a_{11}a_{21}}{a_{11}^N}\frac{\tilde{m}_N^2}{\tilde{m}^2} \rightarrow a_{21}^N = a_{21}\frac{\tilde{m}_N}{\tilde{m}}$$

$$a_{31}^N = \frac{a_{11}a_{31}}{a_{11}^N}\frac{\tilde{m}_N^2}{\tilde{m}^2} \rightarrow a_{31}^N = a_{31}\frac{\tilde{m}_N}{\tilde{m}}$$

$$\left(a_{22}^N\right)^2 = \left(\left(a_{21}\right)^2+\left(a_{22}\right)^2\right)\frac{\tilde{m}_N^2}{\tilde{m}^2} - \left(a_{21}^N\right)^2 \rightarrow a_{22}^N = a_{22}\frac{\tilde{m}_N}{\tilde{m}}$$

$$a_{32}^N = \frac{1}{a_{22}^N}\left(\left(a_{21}a_{31}+a_{22}a_{32}\right)\frac{\tilde{m}_N^2}{\tilde{m}^2} - a_{21}^N a_{31}^N\right) \rightarrow a_{32}^N = a_{32}\frac{\tilde{m}_N}{\tilde{m}}$$

$$\frac{\left(a_{31}^N\right)^2+\left(a_{32}^N\right)^2+\left(a_{ff}^N\right)^2}{\tilde{m}_N^2} = \frac{\left(a_{31}\right)^2+\left(a_{32}\right)^2+\left(a_{33}\right)^2}{\tilde{m}^2} \rightarrow \tilde{m}_N = \frac{a_{ff}^N}{a_{33}}\tilde{m}$$

Therefore, for probit, the normalization just scales all of the parameters, and any positive normalization is acceptable.

Case 2: Logit Kernel with 4 Alternatives

Now, we will show that the equivalent logit kernel case is not so straightforward. Following the same process, the covariance matrix of utility differences for the four alternative unrestricted logit kernel model is:

$$\begin{array}{l} TT'+G \\ theoretical: \end{array} \quad \begin{bmatrix} \left(a_{11}^2+2g\right)/m^2 & & \\ \left(a_{11}a_{21}+g\right)/m^2 & \left(a_{21}^2+a_{22}^2+2g\right)/m^2 & \\ \left(a_{11}a_{31}+g\right)/m^2 & \left(a_{21}a_{31}+a_{22}a_{32}+g\right)/m^2 & \left(a_{31}^2+a_{32}^2+a_{33}^2+2g\right)/m^2 \end{bmatrix}$$

Imposing the normalization $a_{33}=a_{ff}$ leads to:

$$\begin{array}{l} TT'+G \\ normalized: \end{array}$$

$$\begin{bmatrix} \left(\left(a_{11}^N\right)^2+2g\right)/m_N^2 & & \\ \left(a_{11}^N a_{21}^N+g\right)/m_N^2 & \left(\left(a_{21}^N\right)^2+\left(a_{22}^N\right)^2+2g\right)/m_N^2 & \\ \left(a_{11}^N a_{31}^N+g\right)/m_N^2 & \left(a_{21}^N a_{31}^N+a_{22}^N a_{32}^N+g\right)/m_N^2 & \left(\left(a_{31}^N\right)^2+\left(a_{32}^N\right)^2+\left(a_{ff}^N\right)^2+2g\right)/m_N^2 \end{bmatrix}$$

Setting the normalized covariance structure to the normalized structure leads to the following equations (the C notation is just to clean up the math later on):

$$\left(\left(a_{11}^N\right)^2+2g\right)/m_N^2=\left(a_{11}^2+2g\right)/m^2\equiv C_1$$

$$\left(a_{11}^N a_{21}^N+g\right)/m_N^2=\left(a_{11}a_{21}+g\right)/m^2\equiv C_2$$

$$\left(a_{11}^N a_{31}^N+g\right)/m_N^2=\left(a_{11}a_{31}+g\right)/m^2\equiv C_3$$

$$\left(\left(a_{21}^N\right)^2+\left(a_{22}^N\right)^2+2g\right)/m_N^2=\left(a_{21}^2+a_{22}^2+2g\right)/m^2\equiv C_4$$

$$\left(a_{21}^N a_{31}^N+a_{22}^N a_{32}^N+g\right)/m_N^2=\left(a_{21}a_{31}+a_{22}a_{32}+g\right)/m^2\equiv C_5$$

$$\left(\left(a_{31}^N\right)^2+\left(a_{32}^N\right)^2+\left(a_{ff}^N\right)^2+2g\right)/m_N^2=\left(a_{31}^2+a_{32}^2+a_{33}^2+2g\right)/m^2\equiv C_6$$

And solving for the estimated parameters in the normalized model leads to:

$$\left(a_{11}^{N}\right)^2 = C_1 m_N^2 - 2g$$

$$a_{21}^{N} = \frac{C_2 m_N^2 - g}{\sqrt{C_1 m_N^2 - 2g}}$$

$$a_{31}^{N} = \frac{C_3 m_N^2 - g}{\sqrt{C_1 m_N^2 - 2g}}$$

$$\left(a_{22}^{N}\right)^2 = C_4 m_N^2 - 2g - \frac{\left(C_2 m_N^2 - g\right)^2}{C_1 m_N^2 - 2g}$$

$$a_{32}^{N} = \frac{C_5 m_N^2 - g - \frac{\left(C_2 m_N^2 - g\right)\left(C_3 m_N^2 - g\right)}{\left(C_1 m_N^2 - 2g\right)}}{\sqrt{C_4 m_N^2 - 2g - \frac{\left(C_2 m_N^2 - g\right)^2}{\left(C_1 m_N^2 - 2g\right)}}}$$

$$m_N^2 = \frac{1}{C_6}\left(\frac{\left(C_3 m_N^2 - g\right)^2}{C_1 m_N^2 - 2g} + \frac{\left(C_5 m_N^2 - g - \frac{\left(C_2 m_N^2 - g\right)\left(C_3 m_N^2 - g\right)}{\left(C_1 m_N^2 - 2g\right)} \right)^2}{C_4 m_N^2 - 2g - \frac{\left(C_2 m_N^2 - g\right)^2}{\left(C_1 m_N^2 - 2g\right)}} + \left(a_{ff}^{N}\right)^2 + 2g \right)$$

Unlike probit, this is not a simple scale shift, i.e., the model must adjust to the normalization in complex, non-linear ways. Furthermore, it is not clear from these equations what the potential restrictions are on the normalization.

Empirical results exploring the normalization issue for a 4 alternative unrestricted logit kernel model are shown in Table A-1. The table includes estimation results using two different synthetic datasets (the true parameters vary across the datasets). There are 4 alternatives, and the model is specified with three alternative specific dummy parameters, one explanatory variable, and then an unrestricted covariance structure. The final column in the first table shows that, under some circumstances, restricting a_{22} to zero is an invalid normalization. The remaining estimation results suggest that restricting a_{33} to zero is a valid normalization regardless of the true parameter estimates. However, these results are not conclusive.

Table A-1: Normalization of Unrestricted Logit Kernel Model

(2 Synthetic Datasets; 4 Alternatives; 10,000 Observations; 1,000 Halton draws)

	Parameter	True	Unidentified		Valid Normalizations						Invalid Normalization	
			Est	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Systematic	Alt. 1 dummy	1.0	1.38	*(2.8)*	0.93	*(11.5)*	1.02	*(11.8)*	1.31	*(12.1)*	0.76	*(12.4)*
	Alt. 2 dummy	1.0	1.28	*(2.8)*	0.85	*(10.2)*	0.94	*(10.3)*	1.21	*(10.5)*	0.67	*(11.0)*
	Alt. 3 dummy	0.0	0.03	*(0.3)*	0.04	*(0.5)*	0.04	*(0.5)*	0.03	*(0.3)*	0.02	*(0.3)*
	Variable 1	-1.0	-1.37	*(2.9)*	-0.93	*(23.5)*	-1.02	*(25.6)*	-1.30	*(28.8)*	-0.76	*(38.5)*
Disturbance	$\alpha 11$	**2.0**	3.16	*(2.1)*	1.60	*(9.1)*	1.96	*(11.3)*	2.94	*(15.7)*	-0.34	*(3.1)*
	$\alpha 21$	1.0	1.75	*(2.1)*	0.86	*(3.7)*	1.09	*(4.7)*	1.63	*(6.2)*	-2.39	*(15.1)*
	$\alpha 31$	2.0	2.86	*(2.7)*	2.01	*(9.1)*	2.13	*(9.4)*	2.70	*(10.9)*	-1.12	*(8.9)*
	$\alpha 22$	**3.0**	4.62	*(2.6)*	2.89	*(14.6)*	3.25	*(16.2)*	4.35	*(19.1)*	**0.00**	---
	$\alpha 32$	1.0	1.79	*(2.5)*	1.16	*(6.9)*	1.27	*(7.8)*	1.69	*(9.3)*	-0.01	*(0.0)*
	$\alpha 33$	**1.0**	2.20	*(1.7)*	**0.00**	---	**1.00**	---	**2.00**	---	0.00	*(0.0)*
	(Simul.) Log-Likelihood:		-7973.176		-7974.867		-7973.843		-7973.187		-7998.768	

	Parameter	True Value	Unidentified		Valid Normalizations							
			Est	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*	*Est*	*t-stat*
Systematic	Alt. 1 dummy	1.0	0.94	*(8.5)*	0.92	*(9.4)*	0.92	*(9.4)*	0.92	*(9.4)*	0.94	*(8.9)*
	Alt. 2 dummy	1.0	0.95	*(8.2)*	0.93	*(9.1)*	0.92	*(9.1)*	0.93	*(9.1)*	0.96	*(8.4)*
	Alt. 3 dummy	0.0	0.18	*(1.5)*	0.17	*(1.5)*	0.17	*(1.5)*	0.17	*(1.5)*	0.18	*(1.5)*
	Variable 1	-1.0	-0.86	*(17.1)*	-0.85	*(31.8)*	-0.85	*(31.8)*	-0.85	*(31.6)*	-0.87	*(27.7)*
Disturbance	$\alpha 11$	**2.0**	1.43	*(5.3)*	1.37	*(6.9)*	1.37	*(6.9)*	-1.38	*(7.0)*	1.45	*(7.2)*
	$\alpha 21$	1.0	0.79	*(4.6)*	0.76	*(5.0)*	0.76	*(5.0)*	-0.76	*(5.0)*	0.80	*(5.3)*
	$\alpha 31$	2.0	2.53	*(3.9)*	2.50	*(3.8)*	2.48	*(3.8)*	-2.50	*(3.9)*	2.56	*(3.9)*
	$\alpha 22$	**1.0**	0.39	*(0.9)*	-0.22	*(1.6)*	-0.22	*(1.6)*	-0.25	*(1.6)*	0.43	*(1.9)*
	$\alpha 32$	1.0	3.19	*(1.2)*	-4.87	*(14.2)*	-4.78	*(13.8)*	-4.46	*(12.0)*	3.03	*(5.4)*
	$\alpha 33$	**6.0**	3.83	*(1.5)*	**0.00**	---	**1.00**	---	**2.00**	---	**4.00**	---
	(Simul.) Log-Likelihood:		-8983.725		-8984.556		-8984.62		-8984.222		-8983.735	

Case 3: Logit Kernel with 3 Alternatives

The three alternative logit kernel case is a bit easier to compute. Following the same process as above:

$$T: \quad \begin{bmatrix} a_{11}/m & 0 \\ a_{21}/m & a_{22}/m \end{bmatrix}$$

$$TT'+G \ \text{theoretical}: \quad \begin{bmatrix} \left(\left(a_{11}\right)^2 + 2g\right)/m^2 & \\ \left(a_{11}a_{21} + g\right)/m^2 & \left(\left(a_{21}\right)^2 + \left(a_{33}\right)^2 + 2g\right)/m^2 \end{bmatrix}$$

$$TT'+G \quad normalized: \quad \begin{bmatrix} \left(\left(a_{11}^N\right)^2+2g\right)/m_N^2 & \\ \left(a_{11}^N a_{21}^N+g\right)/m_N^2 & \left(\left(a_{21}^N\right)^2+\left(a_{ff}^N\right)^2+2g\right)/m_N^2 \end{bmatrix}$$

$$\rightarrow \quad \left(\left(a_{11}^N\right)^2+2g\right)/m_N^2=\left(\left(a_{11}\right)^2+2g\right)/m^2 \equiv C_1$$

$$\left(a_{11}^N a_{21}^N+g\right)/m_N^2=\left(a_{11}a_{21}+g\right)/m^2 \equiv C_2$$

$$\left(\left(a_{21}^N\right)^2+\left(a_{ff}^N\right)^2+2g\right)/m_N^2=\left(\left(a_{21}\right)^2+\left(a_{33}\right)^2+2g\right)/m^2 \equiv C_3$$

Solution

$$a_{11}^N=\sqrt{C_1 m_N^2-2g} \qquad \text{… or …} \qquad a_{11}^N=\frac{C_2 m_N^2-g}{\sqrt{C_3 m_N^2-a_{33}^2-2g}}$$

$$a_{21}^N=\frac{C_2 m_N^2-g}{\sqrt{C_1 m_N^2-2g}} \qquad \text{… or …} \qquad a_{21}^N=\sqrt{C_3 m_N^2-\left(a_{ff}^N\right)^2-2g}$$

$$m_N^2=\frac{\left(-\left(2g(C_1-C_2+C_3)+\left(a_{ff}^N\right)^2 C_1\right) \pm\sqrt{\left(2g(C_1-C_2+C_3)+\left(a_{ff}^N\right)^2 C_1\right)-4\left(C_2^2-C_1C_3\right)\left(-2g\left(a_{ff}^N\right)^2-3g^2\right)}\right)}{2\left(C_2^2-C_1C_3\right)}$$

Here, the restrictions are

$$\left(2g(C_1-C_2+C_3)+\left(a_{ff}^N\right)^2 C_1\right)^2-4\left(C_2^2-C_1C_3\right)\left(-2g\left(a_{ff}^N\right)^2-3g^2\right)\geq 0 \ ,$$

$$m^2>0 \ ,$$

$$C_1 m^2-2g>0 \ \text{… or …} \ C_3 m^2-\left(a_{ff}^N\right)^2-2g>0 \ ,$$

$$a_{11}^2\left(a_{21}^2+\left(a_{ff}^N\right)^2\right)-\left(a_{11}^2 a_{21}^2\right)\geq 0 \ , \text{ where } a_{11}=f(a_{ff}^N) \text{ and } a_{21}=f(a_{ff}^N) \ ,$$

and only 1 of the two possible m^2 satisfies the conditions.

Again, it’s not clear in which cases these restrictions become limiting. Our empirical tests suggests that the normalization of the lowest diagonal element in the cholesky matrix is, in fact, a valid normalization regardless of the true parameters (unlike, for example, the heteroscedastic case).

Appendix B: **Structural Zeros in Random Parameter Models**

For random parameter models in which a subset of possible covariances are estimated, there is an issue as to how to impose the constraints in order to obtain the desired covariance structure. For example, in the random parameter model presented in Chapter 4, say we want to include covariances among the travel time parameters and not among all 5 random parameters.

Recall that the random parameter logit kernel model is specified as:

$$U_n = X_n b + X_n T z_n + n_n \ ,$$

where the notation is as in Chapter 2.

The issue arises because the constraints are placed on the Cholesky Matrix, T , and not the covariance structure TT'. The key is that introducing the constraint $T_{ij} = 0$ does not necessarily lead to the equivalent cell of the covariance matrix TT' to be zero.

Guideline for Imposing Structural Zeros

The solution is to place the structural zeros in the left-most cells of each row in the Cholesky. If this is done, then TT' will have the same structure as T . To implement this may require reorganizing the data (i.e., specification). We first provide an example below, and then prove the result using the general case.

Example

Say we have 3 variables with random parameters, and we desire the following covariance structure (i.e., 2 of the 3 covariance terms estimated):

$$TT' = \begin{bmatrix} s_{11} & s_{21} & s_{31} \\ s_{21} & s_{22} & 0 \\ s_{31} & 0 & s_{33} \end{bmatrix}.$$

The following restriction on the Cholesky does not retain the structural zero in the covariance matrix:

$$T = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & 0 & a_{33} \end{bmatrix} \rightarrow \qquad TT' = \begin{bmatrix} a_{11}^2 & a_{21}a_{11} & a_{31}a_{11} \\ a_{21}a_{11} & a_{21}^2 + a_{22}^2 & a_{31}a_{21} \\ a_{31}a_{11} & a_{31}a_{21} & a_{31}^2 + a_{33}^2 \end{bmatrix}.$$

But by reorganizing the variables (variable 2, variable 1, variable 3), we get the correct two of three covariances estimated:

$$T = \begin{bmatrix} a_{22} & 0 & 0 \\ a_{21} & a_{11} & 0 \\ 0 & a_{31} & a_{33} \end{bmatrix} \rightarrow \qquad TT' = \begin{bmatrix} a_{22}^2 & a_{21}a_{22} & 0 \\ a_{21}a_{22} & a_{21}^2 + a_{11}^2 & a_{31}a_{11} \\ 0 & a_{31}a_{11} & a_{31}^2 + a_{33}^2 \end{bmatrix}.$$

General Case

A general cholesky matrix can be written as follows:

$$T = \begin{bmatrix} a_{11} & & & & & \\ a_{21} & a_{22} & & & & \\ a_{31} & a_{32} & a_{33} & & & \\ a_{41} & a_{42} & a_{43} & a_{44} & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \\ a_{K1} & a_{K2} & a_{K3} & a_{K4} & \cdots & a_{KK} \end{bmatrix}_{lower\ triangular}.$$

The covariance matrix is then:

$V = TT'$, where

$V_{ij} = V_{ji} = \sum_{k=1}^{K} a_{ik} a_{jk}$ and $i < j$, which simply takes advantage of the symmetry.

The conditions of interest are those under which $V_{ij} \neq 0$ and $V_{ij} = 0$.

(I) $V_{ij} \neq 0$ as long as $a_{ij} \neq 0$ and $a_{jj} \neq 0$.

Explanation: $V_{ij} = ... + a_{ij} a_{jj} + ... \neq 0$.

(II) $V_{ij} = 0$ if $a_{i1},...,a_{ij} = 0$.

Explanation: $V_{ij} = T_i T_j$ where T_i is the i^{th} row of T and T_j is the j^{th} row of T .

The first j elements of T_i are zero (due to the restriction),

the last $K - j$ elements of T_j are zero (due to the lower diagonal structure of the Cholesky),

which leads to $V_{ij} = T_i T_j = 0$.

Therefore, as long as the data are reorganized such that the structural zeros are entered at the beginning of each row of the Cholesky matrix, then the structure of the covariance matrix (TT') will match the structure of the Cholesky matrix (T).

Appendix C: Identification of Agent Effect Parameters

This appendix examines the identification of agent effect parameters as described for the joint revealed and stated preference models described in Chapter 4.

General Specification

The general utility equation for alternative j, person n, and response q

$$U_{jnq} = X_{jnq} b + y_j h_{jn} + n_{jnq} .$$

The utility consists of:

a systematic portion $X_{jnq} b$,
the agent effect $y_j h_{jn}$, and
a Gumbel white noise n_{jnq}, which has variance g / m^2 .

3 Alternative Model, 2 Responses per person:

In levels form, the utilities are as follows:

U for response *1* from person *n*:

$$U_{1n1} = X_{1n1} b + y_1 h_{1n} + n_{1n1} ,$$
$$U_{2n1} = X_{2n1} b + y_2 h_{2n} + n_{2n1} , \text{ and}$$
$$U_{3n1} = X_{3n1} b + y_3 h_{3n} + n_{3n1} .$$

U for response *2* from person *n*:

$$U_{1n2} = X_{1n2} b + y_1 h_{1n} + n_{1n2} ,$$
$$U_{2n2} = X_{2n2} b + y_2 h_{2n} + n_{2n2} , \text{ and}$$
$$U_{3n2} = X_{3n2} b + y_3 h_{3n} + n_{3n2} .$$

Assuming the h 's are independent, the covariance matrix is:

$$Cov(U_{jn1}, U_{jn2}) =$$
$$\begin{bmatrix} y_{11} + g/m^2 & & & & & \\ 0 & y_{22} + g/m^2 & & & & \\ 0 & 0 & y_{33} + g/m^2 & & & \\ y_{11} & 0 & 0 & y_{11} + g/m^2 & & \\ 0 & y_{22} & 0 & 0 & y_{22} + g/m^2 & \\ 0 & 0 & y_{33} & 0 & 0 & y_{33} + g/m^2 \end{bmatrix},$$

where $y_{ii} = (y_i)^2$.

The Utility Differences are as follows:

Response 1: $U_{1n1} - U_{3n1} = ... + y_1 h_{1n} - y_3 h_{3n} + n_{1n1} - n_{3n1}$, and

$U_{2n1} - U_{3n1} = ... + y_2 h_{2n} - y_3 h_{3n} + n_{2n1} - n_{3n1}$.

Response 2: $U_{1n2} - U_{3n2} = ... + y_1 h_{1n} - y_3 h_{3n} + n_{1n2} - n_{3n2}$, and

$U_{2n2} - U_{3n2} = ... + y_2 h_{2n} - y_3 h_{3n} + n_{2n2} - n_{3n2}$.

The covariance matrix of utility differences is:

$$Cov(\Delta U_{jn1}, \Delta U_{jn2}) =$$
$$\begin{bmatrix} y_{11} + y_{33} + 2g/m^2 & & & \\ y_{33} + g/m^2 & y_{22} + y_{33} + 2g/m^2 & & \\ y_{11} + y_{33} & y_{33} & y_{11} + y_{33} + 2g/m^2 & \\ y_{33} & y_{22} + y_{33} & y_{33} + g/m^2 & y_{22} + y_{33} + 2g/m^2 \end{bmatrix}.$$

Applying the rank condition:

$$vecu\left(Cov(\Delta U_{jn1}, \Delta U_{jn2})\right) = \begin{bmatrix} y_{11} + y_{33} + 2g/m^2 \\ y_{22} + y_{33} + 2g/m^2 \\ y_{33} + g/m^2 \\ y_{11} + y_{33} \\ y_{33} \\ y_{22} + y_{33} \end{bmatrix} \rightarrow \text{Jacobian: } \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix} \rightarrow \text{Rank} = 4 .$$

Therefore, we can estimate all 3 of the agent effect parameters and the only required normalization is to m . Empirical verification of this result using synthetic data is provided in Table C-2.

Table C-2: Empirical Tests of Agent Effect Normalization

Synthetic Data

3 Alternatives, multiple responses per respondent.

Beta is a generic parameter for an attribute.

Alphas are the alternative specific constants (Alpha_3 is the base).

Psis are the agent effect parameters.

Base tests (1000 records, 500 Halton draws)

True	Parameter	Est	StdErr	t-stat	Est	StdErr	t-stat	Est	StdErr	t-stat	Est	StdErr	t-stat
1.00	Beta	1.04	0.15	*(6.9)*	1.03	0.15	*(6.9)*	1.03	0.15	*(7.0)*	1.06	0.15	*(7.2)*
1.50	Alpha_1	1.23	0.53	*(2.3)*	1.26	0.52	*(2.4)*	1.17	0.59	*(2.0)*	0.24	0.46	*(0.5)*
1.50	Alpha_2	1.23	0.56	*(2.2)*	1.19	0.62	*(1.9)*	1.40	0.52	*(2.7)*	0.28	0.43	*(0.6)*
1.00	Psi_1	1.32	0.62	*(2.1)*	**0.00**	-----	-----	2.31	0.27	*(8.5)*	3.48	0.51	*(6.9)*
2.00	Psi_2	1.83	0.50	*(3.6)*	2.21	0.27	*(8.3)*	**0.00**	-----	-----	3.32	0.47	*(7.0)*
4.00	Psi_3	4.05	0.60	*(6.8)*	4.31	0.62	*(6.9)*	4.08	0.55	*(7.3)*	**0.00**	-----	-----
	Log-likelihood	-706.76			-708.26			-707.61			-737.32		
	# respondents	100			100			100			100		
	# responses/respondent	10			10			10			10		
	# of records	1000			1000			1000			1000		
	# of draws (H)	500			500			500			500		

Doubling the number of draws, and everything is within a standard error.

True	Parameter	Est	StdErr	t-stat	Est	StdErr	t-stat
1.00	Beta	1.04	0.15	*(7.0)*	1.03	0.15	*(6.9)*
1.50	Alpha_1	1.11	0.54	*(2.1)*	1.25	0.52	*(2.4)*
1.50	Alpha_2	1.24	0.54	*(2.3)*	1.19	0.62	*(1.9)*
1.00	Psi_1	1.67	0.56	*(3.0)*	**0.00**	-----	-----
2.00	Psi_2	1.56	0.58	*(2.7)*	2.23	0.27	*(8.3)*
4.00	Psi_3	3.98	0.59	*(6.8)*	4.27	0.61	*(6.9)*
	Log-likelihood	-706.49			-708.23		
	# respondents	100			100		
	# responses/respondent	10			10		
	# of records	1000			1000		
	# of draws (H)	**1000**			**1000**		

Using 10 times the number of respondents, and the parameters get closer to true.

True	Parameter	Est	StdErr	t-stat
1.00	Beta	0.98	0.05	*(20.8)*
1.50	Alpha_1	1.46	0.16	*(9.1)*
1.50	Alpha_2	1.50	0.16	*(9.1)*
1.00	Psi_1	1.25	0.20	*(6.3)*
2.00	Psi_2	1.77	0.15	*(11.9)*
4.00	Psi_3	3.76	0.18	*(21.4)*
	Log-likelihood	-7257.34		
	# respondents	**1000**		
	# responses/respondent	10		
	# of records	10000		
	# of draws (H)	500		

2 Alternative Model, 2 Responses per person:

In levels form, the utilities are as follows:

U for response *1* from person *n*:

$$U_{1n1} = X_{1n1} b + y_1 h_{1n} + n_{1n1}$$

$$U_{2n1} = X_{2n1} b + y_2 h_{2n} + n_{2n1}$$

U for response *2* from person *n*:

$$U_{1n2} = X_{1n2} b + y_1 h_{1n} + n_{1n2}$$

$$U_{2n2} = X_{2n2} b + y_2 h_{2n} + n_{2n2}$$

Assuming the h's are independent, the covariance matrix is:

$$Cov(U_{jn1}, U_{jn2}) = \begin{bmatrix} y_{11} + g/m^2 & & & \\ 0 & y_{22} + g/m^2 & & \\ y_{11} & 0 & y_{11} + g/m^2 & \\ 0 & y_{22} & 0 & y_{22} + g/m^2 \end{bmatrix}$$

The Utility Differences are as follows:

Response 1: $U_{1n1} - U_{2n1} = \ldots + y_1 h_{1n} - y_2 h_{2n} + n_{1n1} - n_{2n1}$

Response 2: $U_{1n2} - U_{2n2} = \ldots + y_1 h_{1n} - y_2 h_{2n} + n_{1n2} - n_{2n2}$

The covariance matrix of utility differences is:

$$Cov(\Delta U_{jn1}, \Delta U_{jn2}) = \begin{bmatrix} y_{11} + y_{22} + 2g/m^2 & \\ y_{11} + y_{22} & y_{11} + y_{22} + 2g/m^2 \end{bmatrix}$$

By inspection, we can only estimate one agent effect parameter, and the normalization is arbitrary.

Appendix D: Specification and Estimation of the Components of the Generalized Discrete Choice Model

In this appendix, we provide more detail regarding the specification, estimation, and identification of each of the components included in the generalized discrete choice model presented in Chapter 4.

Factor Analytic Disturbances and Logit Kernel

Specification

The disturbance of the logit kernel model has a probit-like portion as well as an i.i.d. Gumbel portion, and is specified as follows:

$$e_n = F_n x_n + n_n \ , \qquad \text{[D-1]}$$

where x_n is an $(M \times 1)$ vector of M multivariate distributed latent factors, F_n is a $(J_n \times M)$ matrix of the factor loadings that map the factors to the error vector (F_n includes fixed and/or unknown parameters and may also be a function of covariates), and n_n is an i.i.d. Gumbel term. For computational reasons, it is desirable to specify the factors such that they are independent, and we therefore decompose x_n as follows:

$$x_n = Tz_n \ , \qquad \text{[D-2]}$$

where z_n are a set of standard independent factors (often normally distributed), TT' is the covariance matrix of x_n, and T is the Cholesky factorization of it. To simplify the presentation, we assume that the factors have standard normal distributions, however, they can follow any number of different distributions, such as lognormal, uniform, etc.

Substituting Equations [D-1] and [D-2] into the standard random utility equation, yields the Factor Analytic Logit Kernel Specification (the framework for which was shown in Figure 4-2):

$$U_n = X_n b + F_n T z_n + n_n , \qquad \text{[D-3]}$$

where: F_n is a $(J \times M)$ matrix of factor loadings, including fixed and/or unknown parameters,

T is an $(M \times M)$ lower triangular cholesky matrix of unknown parameters, where $TT' = Cov(Tz_n)$,

z_n is an $(M \times 1)$ vector of unknown factors with independent standard distributions, and

U, X, b, n are as in the base MNL model.

The covariance structure of the model is:

$$\text{cov}(U_n) = F_n T T' F_n' + (g / m^2) I_J ,$$

where I_J is a $(J \times J)$ identity matrix, and g and m are as in the base MNL model.

$F_n T z_n$ provides for flexibility, as highlighted by the special cases presented below, and n_n aids in computation, as will be explained in the section on estimation.

Special Cases

The logit kernel model with its probit-like component completely opens up the specification of the disturbances so that any desirable error structure can be represented in the model. In particular, several useful special cases of the model are:

Heteroscedastic

The heteroscedastic model relaxes MNL’s i.i.d. Gumbel error structure by allowing the variances to vary across alternatives. The model is specified as:

$$U_n = X_n b + T z_n + n_n ,$$

where: F_n from the general logit kernel equation [D-3] equals the identity matrix,

T is $(J \times J)$ diagonal, which contains the standard deviation of each alternative,

z_n is $(J \times 1)$.

Nested & Cross-Nested Error Structures

Models that are analogous to nested and cross-nested logit can also be specified. The nested logit kernel model is as follows:

$$U_n = X_n b + FTz_n + n_n \ ,$$

where: z_n is $(M \times 1)$, M is the number of nests, and one factor is defined for each nest,

F is $(J \times M)$, $F_{jm} = \begin{cases} 1 & \text{if alternative } j \text{ is a member of nest } m \\ 0 & \text{otherwise} \end{cases}$, and

T is $(M \times M)$ diagonal, which contains the standard deviation of each factor.

In a strictly hierarchical nesting structure, the nests do not overlap, and FF' is block diagonal. In a cross-nested structure, the alternatives can belong to more than one group.

Error Components

The error component formulation is a generalization that includes the heteroscedastic, nested, and cross-nested structures. The model is specified as follows:

$$U_n = X_n b + FTz_n + n_n \ ,$$

where: F is a $(J \times M)$ matrix of fixed factor loadings equal to 0 or 1,

$$f_{jm} = \begin{cases} 1 & \text{if the } m^{th} \text{ element of } z_n \text{ applies to alternative } j \\ 0 & \text{otherwise} \end{cases},$$

z_n, T are defined as in the general case (Equation [D-3]).

Factor Analytic Errors

The Factor Analytic specification is a further generalization in which the F_n matrix contains unknown parameters. The model is written as in the general case:

$$U_n = X_n b + F_n T z_n + n_n \ .$$

If T is diagonal, the disturbances can be written in scalar form as follows:

$$e_{in} = \sum_{m=1}^{M} f_{imn} s_m z_{mn} + n_{in} \ ,$$

where both the f_{imn}'s and s_m's (the diagonal elements of T) are unknown parameters.

Random Parameters

The MNL formulation with normally distributed random taste parameters can be written as:

$$U_n = X_n b_n + n_n \ ,$$

where $b_n \sim N(b, \Sigma_b)$.

Replacing b_n with the equivalent relationship: $b_n = b + Tz_n$, where T is the lower triangular Cholesky matrix such that $TT' = \Sigma_b$, leads to a general factor analytic logit kernel specification where $F_n = X_n$:

$$U_n = X_n b + X_n T z_n + n_n \ .$$

The unknown parameters are the vector b and those present in T. Note that T is often diagonal, but does not have to be. Also, the distribution does not have to be normal. For example, it is often specified as lognormal for parameters that have sign constraints.

Autoregressive Process

The disturbances $x_n = (x_{1n}, ..., x_{Jn})'$ of a first-order generalized autoregressive process [GAR(1)] is defined as follows:

$$x_n = r A_n x_n + T z_n \ ,$$

where: A_n is a $(J \times J)$ matrix of weights $a_{i\,j,n}$ describing the influence of each x_{jn} error upon the others. A_n can either be fixed or a function of unknown parameters;

r is an unknown parameter; and

Tz_n allows for heteroscedastic disturbances, T is $(J \times J)$ diagonal and z_n is $(J \times 1)$.

Solving for x_n and incorporating it into the logit kernel general form, leads to a logit kernel GAR[1] specification:

$$U_n = X_n b + F_n T z_n + n_n \ ,$$

where $F_n = (I - r A_n)^{-1}$.

Estimation

As with probit, the flexibility in specifying the error terms comes at a cost, namely the probability functions consist of multi-dimensional integrals that do not have closed form solutions. Standard practice is to estimate such models by replacing the choice probabilities with easy to compute and unbiased simulators. A key aspect of the logit kernel model is that if the factors z_n are known, the model corresponds to a multinomial logit formulation:

$$\Lambda(i \mid X_n, z_n) = \frac{e^{m\left(X_{in}b + F_{in}Tz_n\right)}}{\sum_{j \in C} e^{m\left(X_{jn}b + F_{jn}Tz_n\right)}} ,$$

where $\Lambda(i \mid X_n, z_n; d)$ is the probability that the choice is i given X_n and z_n. The unknown parameters include $m, b,$ and those in F and T.

Since the z_n is in fact not known, the unconditional choice probability of interest is:

$$P(i \mid X_n) = \int_z \Lambda(i \mid X_n, z)n(z, I_M)dz \quad , \qquad \text{[D-4]}$$

where $n(z, I_M)$ is the joint density function of z, which, by construction, is composed of i.i.d. standard normal components. The advantage of the logit kernel model is that we can naturally estimate $P(i \mid X_n; d)$ with an unbiased, smooth, tractable simulator, which we compute as:

$$\hat{P}(i \mid X_n) = \frac{1}{\mathbb{D}} \sum_{d=1}^{\mathbb{D}} \Lambda(i \mid X_n, z_n^d) ,$$

where z_n^d denotes draw d from the distribution of z_n, thus enabling us to estimate high dimensional integrals with relative ease. The logit kernel probability simulator has all of the desirable properties of a simulator including being convenient, unbiased, and smooth, and can straightforwardly be applied in maximum simulated likelihood estimation.

Identification

It is not surprising that the estimation of such models brings identification and normalization issues. There are two sets of relevant parameters that need to be considered: the vector b and the unrestricted parameters in the disturbance term, which include F_n, T, and m. For the vector b, identification is identical to that for a multinomial logit model. Such issues are well understood, and the reader is referred to Ben-Akiva and Lerman (1985) for details.

The identification of the parameters in the disturbances is much more complex. Identification and normalization, including the order, rank, and positive definiteness conditions were covered in detail in Chapter 2. The summary is that one has to be careful with identification of the logit kernel model. In particular:

- *Identification is not necessarily intuitive.* For example, only one disturbance parameter is identified with a two nest structure, whereas three disturbance parameters are identified with a three nest structure.

- *Identification is not necessarily analogous to the systematic portion.* For example, if random parameters are estimated for a categorical variable with 3 categories, only two systematic parameters are identified, whereas three disturbance parameters are identified.

- *Normalization is not necessarily like probit.* For example, the normalization for a probit heteroscedastic model is arbitrary, whereas the normalization for a logit kernel heteroscedastic model is not (the minimum variance alternative must be normalized).

- Using a small number of draws in simulation will mask identification problems, which makes analytical verification of identification even more critical.

Combining Stated and Revealed Preferences

Specification

The framework for combining stated and revealed preferences is shown in Figure 4-3. The choice models for the RP and SP models can be written individually as follows:

Revealed: $U_n^{RP} = X_n^{RP} b + e_n^{RP}$,

Stated: $U_n^{SP} = X_n^{SP} b \quad + e_n^{SP}$,

where X_n^{RP} and X_n^{SP} are the explanatory variables for the RP and SP experiments, respectively, and b are the unknown parameters where at least a subset of the parameters are common across the two models. In order to combine the models, there are two important issues involving the disturbances e_n^{RP} and e_n^{SP} that need to be considered. First, they are most likely correlated across multiple responses for a given individual. Second, the scale (i.e., their variances) may vary across the two models.

Issue 1: Correlation Across Responses from the Same Individual

It is highly likely that multiple responses from a given individual will exhibit correlated disturbance terms. In the best possible scenario, ignoring potential correlation will result in consistent, but inefficient estimates. However, in the worst case, it can lead to inconsistent estimates. This occurs, for example, when the revealed choice (y_n^{RP}) is included in the SP model (often done to capture response bias); if e_n^{SP} and e_n^{RP} are correlated then y_n^{RP} is endogenous to the SP model, and therefore the resulting estimates are inconsistent.

To deal with the issue of correlation, the model should be specified in a way that allows for correlation among the SP responses as well as correlation between the SP and RP responses from a given individual. To achieve this, Morikawa et al. (1996) suggest decomposing the error component into two portions[35]:

$$e_n^{RP} = \Psi^{RP} h_n + n_n^{RP} \ ,$$

$$e_n^{SP} = \Psi^{SP} h_n + n_n^{SP} \ ,$$

[35] Again, we're assuming that the MNL specification is appropriate for both the SP and RP model, although clearly any choice model can be substituted.

where: h_n is a $(J \times 1)$ vector of i.i.d. standard normal disturbances. These are assumed independent across alternatives, but identical across responses for a given individual (also called an 'agent effect').

Ψ^{RP}, Ψ^{SP} are $(J \times J)$ diagonal matrices, which contain unknown parameters that capture the correlation across responses.

n_n^{RP}, n_n^{SP} are $(J \times 1)$ vectors of disturbances (white noise), and each vector is i.i.d. Gumbel.

h_n, n_n^{RP}, n_n^{SP} are independent.

Thus, this structure allows for correlations between RP and SP responses for the same individual:

$$Cov(U_{in}^{RP}, U_{in}^{SP}) = y_i^{RP} y_i^{SP} \ ,$$

where y_i denotes the i^{th} diagonal element of the matrix Ψ.

If there are multiple SP responses per individual, for example, SP_a and SP_b, then:

$$Cov(U_{in}^{SP_a}, U_{in}^{SP_b}) = (y_i^{SP})^2 \ .$$

Given this structure, the likelihood for, say, 1 RP response (y_n^{RP}) and Q SP responses ($y_n^{SP} = \{ y_{n1}^{SP}, ..., y_{nQ}^{SP} \}$) observed for the respondent n is:

$$P(y_n^{RP}, y_n^{SP} \mid X_n) = \int_h \Lambda(y_n^{RP} \mid X_n^{RP}, h) \prod_{q=1}^{Q} \Lambda(y_{nq}^{SP} \mid X_{nq}^{SP}, h) f(h) \, dh \ , \qquad \text{[D-5]}$$

where the unknown parameters include $b, m,$ and Ψ, and it is necessary to integrate out over the unknown correlation factor h.

In calculating the probabilities within the integral, the scale issue becomes important, which is described next.

Issue 2: Different Scales for Different Datasets

Since the effect of unobserved factors will be different between revealed and stated preference surveys, there is good reason to suspect that n_n^{RP} and n_n^{SP} have different variances, which leads to different scales m_{RP} and m_{SP}. The conditional probabilities are then:

$$\Lambda(i^{RP} \mid X_n^{RP}, h_n) = \frac{e^{m_{RP}(X_{in}^{RP} b + y_i^{RP} h_{in})}}{\sum_{j \in C} e^{m_{RP}(X_{jn}^{RP} b + y_j^{RP} h_{jn})}}, \qquad \text{[D-6]}$$

$$\Lambda(i^{SP} \mid X_{nq}^{SP}, h_n) = \frac{e^{m_{SP}(X_{inq}^{SP} b + y_i^{SP} h_{in})}}{\sum_{j \in C} e^{m_{SP}(X_{jnq}^{SP} b + y_j^{SP} h_{jn})}}. \qquad \text{[D-7]}$$

Estimation

The likelihood for the sample can then be built from Equations [D-5], [D-6], and [D-7]. Joint estimators are obtained by maximizing the log-likelihood of the sample over the unknown parameters (b, m, Ψ).

This model requires numerical integration with respect to h to evaluate the likelihood, and therefore requires customization of the likelihood in a flexible programming package. However, if serial correlation is not considered, the model simplifies considerably as the integration over the agent effect (h) is no longer necessary, or:

$$P(y_n^{RP}, y_n^{SP} \mid X_n) = \Lambda(y_n^{RP} \mid X_n^{RP}) \prod_{q=1}^{Q} \Lambda(y_{nq}^{SP} \mid X_{nq}^{SP}),$$

where the unknown parameters are b and m.

In this case the log-likelihood can be decomposed into the standard log-likelihood for the RP data plus the log-likelihood for the SP data. The independent model can be estimated either sequentially or simultaneously. (See Morikawa, 1989, for a discussion.) Bradley and Daly (1997) developed a method for estimating this model (no agent effect) simultaneously by creating an artificial tree structure and using a standard Nested Logit software package.

Identification

The standard identification rules for discrete choice apply to the specifications of both the RP and SP portions (see Ben-Akiva and Lerman, 1985). The only unique issues here are with the agent effect parameters, Ψ, and the scale terms, m. The required normalizations for Ψ are determined by a rank condition (see Appendix C), and the resulting identification restrictions depend on the specification, for example[36]:

[36] The mirror image of Cases I and II also hold in which there are multiple RP responses and a single SP response.

	Number of RP Responses	*Number of SP Responses*	*Number of Alternatives*	*Number of Identification Restrictions*
Case I:	1	1 or more	3 or more	none
Case II:	1	2 or more	2	one (either RP or SP)
Case III:	1	1	2	one for RP & one for SP

For the scale normalization, recall that in order to identify the coefficients of a discrete choice model, the scale must be set by arbitrarily fixing the variance of the disturbance term, i.e., by fixing m. For the joint SP/RP model, it is only necessary to fix one of the two scale terms. Therefore, we arbitrarily set the scale of the model to be that of the RP data (i.e., $m_{RP} = 1$) and we estimate one parameter m, which equals the ratio of standard deviations between n_n^{RP} and n_n^{SP}, or $m = m_{SP} / m_{RP}$. The conditional probabilities are then:

$$\Lambda(i^{RP} \mid X_n^{RP}, h_n) = \frac{e^{(X_{in}^{RP} b + y_i^{RP} h_{in})}}{\sum_{j \in C} e^{(X_{jn}^{RP} b + y_j^{RP} h_{jn})}} ,$$

$$\Lambda(i^{SP} \mid X_{nq}^{SP}, h_n) = \frac{e^{m(X_{inq}^{SP} b \;\; + y_i^{SP} h_{in})}}{\sum_{j \in C} e^{m(X_{jnq}^{SP} b \;\; + y_j^{SP} h_{jn})}} .$$

Choice and Latent Variables

Specification

The framework for the integrated choice and latent variable model is shown in Figure 4-5. The integrated model is composed of two parts: a discrete choice model and a latent variable model. Each part consists of one or more structural equations and one or more measurement equations. Specification of these equations and the likelihood function follow.[37]

Structural Equations

For the latent variable model, we need the distribution of the latent variables, denoted as X_n^*, given the observed variables, for example:

$$X_n^* = X_n l + w_n \quad \text{and} \quad w_n \sim N(0, \Sigma_w) . \qquad \text{[D-8]}$$

This results in one equation for each latent variable. Equation [D-8] can also be generalized to include latent variables as explanatory variables.

The structural equation for the choice model is as before, except now contains latent explanatory variables:

[37] Here, as elsewhere in the chapter, we make simplifying assumptions to clarify the explanation, for example, we assume linear in the parameters and normally distributed disturbances (except for the Gumbel term for the choice model).

$$U_n = X_n b_1 + X_n^* b_2 + e_n \ . \qquad \text{[D-9]}$$

Measurement Equations

For the latent variable model, we need the distribution of the indicators (I) conditional on the values of the latent variables, for example:

$$I_n = X_n^* a + u_n \quad \text{and} \quad u_n \sim N(0, \Sigma_u) \ . \qquad \text{[D-10]}$$

This results in one equation for each indicator (for example, each survey question). These measurement equations usually contain only the latent variables on the right-hand-side. However, they may also contain individual characteristics or any other variable determined within the model system such as the choice indicator. In principle, such parameterizations can be allowed to capture systematic response biases when the individual is providing indicators.

The measurement equation for the choice model is exactly as before:

$$y_{in} = \begin{cases} 1, \ if\ U_{in} = \max_j \{U_{jn}\} \\ 0, \ otherwise \end{cases} \ . \qquad \text{[D-11]}$$

Integrated Choice and Latent Variable Model

The integrated model consists of Equations [D-8] through [D-11]. Equations [D-8] and [D-10] comprise the latent variable model, and equations [D-9] and [D-11] comprise the choice model.

Estimation

Likelihood Function

The most intuitive way to create the likelihood function for the integrated model is to start with the likelihood of a choice model without latent variables:

$$P(y_n \mid X_n) \ .$$

The choice model can be any number of forms, for example, logit, nested logit, random parameter logit, probit, ordinal probit, and can include the combination of different choice indicators such as stated and revealed preferences.

Now we add the latent variables to the choice model. Once we hypothesize an unknown latent construct, X^*, its associated distribution, and independent error components (w, e), the likelihood function is then the integral of the choice model over the distribution of the latent constructs:

$$P(y_n \mid X_n) = \int_{X^*} P(y_n \mid X_n, X_n^*) f_1(X_n^* \mid X_n) dX^* \ ,$$

where the unknown parameters include the b from the choice model (as well as any estimated disturbance terms), and the l and parameters in Σ_w from the latent variable structural model.

We introduce indicators to improve the accuracy of estimates of the structural parameters. Assuming the error components (w, e, u) are independent, the joint probability of the observable variables y_n and I_n, conditional on the exogenous variables X_n, is:

$$f_4(y_n, I_n \mid X_n) = \qquad \text{[D-12]}$$

$$\int_{X^*} P(y_n \mid X_n, X_n^*) f_3(I_n \mid X_n^*) f_1(X_n^* \mid X_n) \, dX^* \ ,$$

which now includes the unknown parameters from the measurement model: a and those in Σ_w.

Note that the first term of the integrand corresponds to the choice model, the second term corresponds to the measurement equation from the latent variable model, and the third term corresponds to the structural equation from the latent variable model. The latent variable is only known to its distribution, and so the joint probability of y, I, and X^* is integrated over the vector of latent constructs X^*.

Identification

As with all latent variable models, identification is certainly an issue in these integrated choice and latent variable models. While identification has been thoroughly examined for special cases of the integrated framework presented here (see, for example, Elrod 1988 and Keane 1997), necessary and sufficient conditions for the general integrated model have not been developed. Therefore, identification of the integrated models needs to be analyzed on a case-by-case basis.

In general, all of the identification rules that apply to a traditional latent variable model are applicable to the latent variable model portion of the integrated model. See Bollen (1989) for a detailed discussion of these rules. Similarly, the normalizations and restrictions that apply to a standard choice model would also apply here. See Ben-Akiva and Lerman (1985) for further information.

For the integrated model, a sufficient, but not necessary, condition for identification can be obtained by extending the Two-step Rule used for latent variable models to a Three-step Rule for the integrated model:

1. Confirm that the measurement equations for the latent variable model are identified (using, for example, standard identification rules for factor analysis models).

2. Confirm that, given the latent variables, the structural equations of the latent variable model are identified (using, for example, standard rules for a system of simultaneous equations).

3. Confirm that, given the distribution of the latent variables, the choice model is identified (using, for example, standard rules for a discrete choice model).

Because identification is not always straightforward, empirical tests of identification can be extremely useful for these models, which are discussed later.

Choice and Latent Classes

Specification

The framework for the latent class model is shown in Figure 4-6. The model is written as:

$$P(i \mid X_n) = \sum_{s=1}^{S} P(i \mid X_n, s) P(s \mid X_n) \quad , \qquad \text{[D-13]}$$

$P(i \mid X_n; s)$ is the class-specific choice model, and can include variation across classes in terms of all aspects of the choice process, for example taste parameters, choice sets, decision protocol, or covariance structure (for example, nesting). $P(s \mid X_n)$ is the class membership model, i.e., the probability of belonging to class s given X_n.

The primary issue in latent class models is how to specify the class membership model. Gopinath (1995) provides extensive detail on this issue, and a summary is provided here. Many applications of latent class choice models in the literature employ a naïve class membership model in which $P(s \mid X_n; q)$ is a logit model and the class-specific constants are the only parameters (see, for example, Kamakura and Russell, 1987). Such models are more commonly called 'finite mixture models' (see McLachlan and Basford, 1988, for a review).

The most straightforward extension of the naïve model is to include descriptive information about the decision-makers as explanatory variables in $P(s \mid X_n; q)$ to improve the prediction of the class probabilities. If $P(s \mid X_n; q)$ is an MNL (or analogous) model, then this is called a 'categorical criterion model' (see, for example, Dillon et al., 1993, or Gupta and Ghintagunta, 1994). If it is ordinal MNL (i.e., the classes represent varying degrees along a single dimension) than it is called an 'ordinal criteria model'. Gopinath (1995) developed a flexible and rigorous methodology for specifying latent class membership models. His methodology includes those methods described above as well as models in which the class membership specification has ordinal criteria in more than one dimension. For example, if the latent classes represent taste variations, the class membership could be based on an individual's sensitivity to two or more attributes (for example, time and cost).

As with the continuous latent variables, it is helpful if indicators of the latent classes are available. This is similar to the continuous latent variable case in which another component (the measurement equations for the indicators) is added to the likelihood. See Ben-Akiva and Boccara (1995) and Gopinath (1995) for more information.

Estimation

Like the other models described in this chapter, estimation can be performed using maximum likelihood techniques. However, one key difference is that as long as the conditional choice model does not require integration, then the latent class model does not require integration.

One important issue with latent class models is that there can be numerous local maxima. Therefore, it is necessary to explore different starting values. In our empirical tests, we have found that it works well to start the model at a point with very distinct class-specific behavior, and allow the classes to move together.

Identification

Identification of latent class models follows the general rules for latent variable models. A sufficient but not necessary two-step rule can be used in which the first step is to verify that the class membership model is identified, and the second step is to verify that the conditional choice model is identified given the class.

Appendix E:
Stability of Parameter Estimates

The results presented in this appendix are used to test the stability of the models presented in Chapter 4. With the exception of the random parameter models, the final parameter estimates are within one standard deviation of a model estimated with fewer draws, and therefore are very stable. The random parameter models tend to be more unstable, particularly with lognormal (as opposed to normal) distributions. Nonetheless, with a couple of exceptions, all parameter estimates are within 2 standard deviations, and therefore fairly stable. The parameters that are outside of 2 standard deviations (2 in Table E-4 and 3 in Table E-8, shown in bold) are in all cases lognormal distribution parameters that have extremely small standard deviations.

Table E-3: Stability of Joint SP/RP Model (Table 4-2)

Draws:	500	1000	
Parameter	Est.	Est.	*Std Er.*
Rail constant RP	0.439	0.444	0.493
Rail constant SP2	-0.480	-0.466	0.777
Work trip dummy	1.17	1.17	0.51
Fixed arrival time dummy	0.724	0.723	0.381
Female dummy	0.989	0.990	0.381
Cost per person in Guilders	-0.0607	-0.0608	0.0132
Out-of-vehicle time in hours	-2.22	-2.23	0.83
In-vehicle time in hours	-0.709	-0.710	0.158
Number of transfers	-0.100	-0.100	0.036
Amenities	-0.361	-0.361	0.080
Inertia dummy (RP Choice)	2.98	2.97	1.02
Agent effect RP	0.680	0.686	0.490
Agent effect SP2	2.44	2.44	0.50
Scale (mu) SP1	2.31	2.31	0.50
Scale (mu) SP2	1.31	1.31	0.30
Tau1 SP1 (=-Tau4 SP1)	*-0.195*	*-0.195*	----
Tau2 SP1 (=-Tau3 SP1)	*-0.0126*	*-0.0127*	----
Tau3 SP1	0.0126	0.0127	0.0036
Tau4 SP1	0.195	0.195	0.049
Tau1 SP2	-0.987	-0.986	0.219
Tau2 SP2 (=-Tau3 SP2)	*-0.180*	*-0.180*	----
Tau3 SP2	0.180	0.180	0.053
Tau4 SP2	1.32	1.32	0.32
Log-likelihood:	-4517.50	-4517.43	

Table E-4: Stability of Random Parameter Model (Table 4-3)

	Distributed Model 1: Independent Distributions					Distributed Model 2: Multivariate Distributions				
Draws:	1000	5000	10000	20000		1000	5000	10000	20000	
Parameter	Est.	Est.	Est.	Est.	Std. Er.	Est.	Est.	Est.	Est.	Std. Er.
Rail constant RP	1.94	1.84	2.78	2.80	0.97	1.96	2.94	2.04	1.67	0.81
Rail constant SP2	2.53	2.63	3.82	4.05	1.20	2.65	3.73	2.59	2.19	0.79
Work trip dummy	0.820	0.902	0.814	0.891	0.762	1.179	1.181	1.234	1.16	0.65
Fixed arrival time dummy	0.569	0.620	0.559	0.513	0.647	0.526	0.702	0.710	0.850	0.522
Female dummy	1.48	1.49	1.56	1.61	0.61	1.55	1.50	1.55	1.51	0.51
1 Cost per person in Guilders	-2.24	-2.38	-2.21	-2.19	0.26	-2.24	-2.14	-2.14	-2.33	0.26
2 Out-of-vehicle time in hours	1.14	1.10	1.60	1.56	0.34	1.33	1.41	1.19	0.97	0.36
3 In-vehicle time in hours	0.224	0.0852	0.295	0.284	0.279	0.142	0.226	0.247	0.149	0.271
4 Number of transfers	-2.09	-2.37	-2.05	-2.29	0.33	-2.32	-2.14	-1.93	-2.25	0.31
5 Amenities	-0.567	-0.765	-0.600	-0.644	0.265	-0.642	-0.656	-0.650	-0.722	0.274
T11	0.903	1.00	0.917	0.993	0.129	1.283	1.201	1.194	1.29	0.06
T21						-0.293	-0.434	-0.414	-0.479	0.043
T31						0.422	0.512	0.478	0.470	0.045
T41						0.711	0.408	0.437	**0.645**	**0.055**
T51						0.352	0.329	0.356	0.404	0.043
T22	0.827	0.873	0.700	0.723	0.166	0.497	0.559	0.550	0.658	0.060
T32						-0.073	0.160	0.355	0.281	0.063
T42						0.533	0.262	0.308	0.287	0.021
T52						0.229	0.163	0.127	0.035	0.048
T33	0.779	0.871	0.885	0.818	0.057	0.867	0.856	0.890	0.894	0.042
T43						-0.210	-0.283	-0.190	**0.106**	**0.036**
T53						0.223	0.150	0.125	0.136	0.033
T44	1.78	1.84	1.77	1.96	0.21	1.93	1.97	1.77	1.83	0.11
T54						0.221	0.350	0.351	0.344	0.024
T55	0.94	1.09	1.07	1.06	0.05	0.95	1.04	1.01	1.11	0.07
Inertia dummy (RP Choice)	0.675	0.667	0.300	-0.245	0.680	0.387	0.990	1.068	1.097	0.481
Agent effect RP	2.55	2.21	2.47	3.19	1.28	2.58	2.10	2.02	2.07	0.65
Agent effect SP2	4.08	3.26	3.59	4.14	1.14	3.97	3.51	3.47	3.74	1.05
Scale (mu) SP1	4.07	4.82	3.99	4.07	1.11	5.02	4.75	4.88	5.21	1.44
Scale (mu) SP2	1.54	2.08	1.86	1.79	0.48	1.92	1.91	1.91	1.88	0.54
Tau1 SP1 (=-Tau4 SP1)	*-0.231*	*-0.202*	*-0.244*	*-0.241*	----	*-0.194*	*-0.213*	*-0.209*	*-0.196*	----
Tau2 SP1 (=-Tau3 SP1)	*-0.0153*	*-0.0134*	*-0.0161*	*-0.0159*	----	*-0.0127*	*-0.0140*	*-0.0137*	*-0.0128*	----
Tau3 SP1	0.0153	0.0134	0.0161	0.0159	0.0052	0.0127	0.0140	0.0137	0.0128	0.0043
Tau4 SP1	0.231	0.202	0.244	0.241	0.081	0.194	0.213	0.209	0.196	0.071
Tau1 SP2	-1.004	-0.777	-0.865	-0.904	0.241	-0.841	-0.856	-0.859	-0.856	0.241
Tau2 SP2 (=-Tau3 SP2)	*-0.178*	*-0.137*	*-0.153*	*-0.160*	----	*-0.147*	*-0.151*	*-0.151*	*-0.150*	----
Tau3 SP2	0.178	0.137	0.153	0.160	0.055	0.147	0.151	0.151	0.150	0.053
Tau4 SP2	1.291	0.988	1.11	1.15	0.31	1.03	1.09	1.08	1.08	0.31
Log-likelihood:	-3934.21	-3933.36	-3932.50	-3931.20		-3916.15	-3909.88	-3908.71	-3911.72	

Table E-5: Stability of Choice and Latent Variable Model (Table 4-5)

CHOICE MODEL

Draws:	1000	5000	
Parameter	Est.	Est.	Std. Er.
Rail constant RP	-0.525	-0.442	0.750
Rail constant SP2	-1.193	-0.890	0.837
Work trip dummy	1.70	1.67	0.64
Fixed arrival time dummy	0.748	0.692	0.532
Female dummy	1.15	1.13	0.45
Cost per person in Guilders	-0.0593	-0.0605	0.0163
Out-of-vehicle time in hours	-0.946	-0.983	0.936
In-vehicle time in hours	-0.679	-0.691	0.186
Number of transfers	-0.097	-0.0982	0.0384
Amenities	-0.351	-0.358	0.097
Latent Comfort - RP	1.10	1.16	1.17
Latent Comfort - SP2	1.14	1.16	0.55
Latent Convenience - RP	1.38	1.30	0.76
Latent Convenience - SP2	0.746	0.764	0.331
Inertia dummy (RP Choice)	3.04	2.52	1.24
Agent effect RP	-0.087	0.210	0.611
Agent effect SP2	2.02	2.08	0.64
Scale (mu) SP1	2.37	2.32	0.63
Scale (mu) SP2	1.27	1.31	0.42
Tau1 SP1 (=-Tau4 SP1)	-0.190	-0.194	----
Tau2 SP1 (=-Tau3 SP1)	-0.0124	-0.0126	----
Tau3 SP1	0.0124	0.0126	0.0041
Tau4 SP1	0.190	0.194	0.058
Tau1 SP2	-1.014	-0.988	0.313
Tau2 SP2 (=-Tau3 SP2)	-0.185	-0.181	----
Tau3 SP2	0.185	0.181	0.065
Tau4 SP2	1.36	1.33	0.44
Log-likelihood (Choice&Latent):	-6656.87	-6656.12	
Log-likelihood (Choice):	-4518.72	-4517.97	

LATENT VARIABLE MODEL

Structural Equations (2 equations, 1 per column)

	Comfort Equation			Convenience Equation		
Draws:	1000	5000		1000	5000	
Parameter	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.
Constant - Comfort	0.087	0.106	0.219			
Constant - Convenience				0.529	0.489	0.303
Age dummy - over 40	-0.444	-0.449	0.622	0.885	0.871	0.287
First class rail rider	0.441	0.431	0.567			
In-vehicle time in hours	-0.505	-0.481	0.331			
Out-of-vehicle time in hours				-1.22	-1.18	0.71
Number of transfers				-0.151	-0.122	0.199
Free parking dummy (auto)				0.242	0.222	0.242
Variance(ω)	1.00	1.00	----	1.00	1.00	----

Measurement Equations (6 equations, 1 per row)

	Comfort Parameters			Convenience Parameters			Disturbance Params. (StdDev(υ))		
Draws:	1000	5000		1000	5000		1000	5000	
Equation	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.
Relaxation	0.540	0.522	0.240	0.126	0.131	0.135	1.16	1.17	0.13
Reliability	0.329	0.331	0.105	0.443	0.446	0.089	0.903	0.899	0.055
Flexibility				0.714	0.731	0.288	0.894	0.877	0.242
Ease				0.570	0.571	0.168	1.15	1.15	0.09
Safety	0.394	0.381	0.135	0.134	0.132	0.117	0.796	0.803	0.081
Overall Rating	1.21	1.25	0.82	1.42	1.39	0.51	1.29	1.28	0.26

Table E-6: Stability of Latent Class Model (Table 4-6)

MODE CHOICE MODEL

	Parameters Common Across Classes			Parameters Unique to Class 1			Parameters Unique to Class 2		
Draws:	500	1000		500	1000		500	1000	
Parameter	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.
Rail constant RP	1.28	1.26	0.756						
Rail constant SP2	1.45	1.42	0.772						
Work trip dummy	1.11	1.10	0.620						
Fixed arrival time dummy	0.637	0.641	0.497						
Female dummy	1.03	1.03	0.432						
Cost per person in Guilders				-0.227	-0.231	0.063	-0.0405	-0.0408	0.0115
Out-of-vehicle time in hours				-1.38	-1.31	1.21	-3.51	-3.47	1.34
In-vehicle time in hours				-1.66	-1.69	0.48	-0.871	-0.876	0.244
Number of transfers				-0.211	-0.216	0.092	-0.149	-0.149	0.055
Amenities				-0.402	-0.408	0.114	-0.537	-0.540	0.146
Inertia dummy (RP Choice)	0.97	0.99	0.696						
Agent effect RP	2.12	2.09	0.76						
Agent effect SP2	2.91	2.87	0.73						
Scale (mu) SP1	2.27	2.25	0.59						
Scale (mu) SP2	1.56	1.56	0.35						
Tau1 SP1 (=-Tau4 SP1)	-0.235	-0.236	----						
Tau2 SP1 (=-Tau3 SP1)	-0.0153	-0.0154	----						
Tau3 SP1	0.0153	0.0154	0.0050						
Tau4 SP1	0.235	0.236	0.070						
Tau1 SP2	-0.895	-0.895	0.210						
Tau2 SP1 (=-Tau3 SP2)	-0.161	-0.161	----						
Tau3 SP2	0.161	0.161	0.051						
Tau4 SP2	1.17	1.17	0.28						
Log-likelihood:	-4283.14	-4283.04							

CLASS MEMBERSHIP MODEL

Draws:	500	1000	
Parameter	Est.	Est.	Std. Er.
Constant	-0.450	-0.455	0.395
Female dummy	-0.0776	-0.0832	0.3625
Number of persons in party	0.175	0.174	0.121
Work trip dummy	-1.93	-1.94	0.73
Age over 40 dummy	-0.476	-0.472	0.371

Table E-7: Stability of Choice and Latent Variable with Latent Classes Model (Table 4-7)

MODE CHOICE MODEL

	Parameters Common Across Classes			Parameters Unique to Class 1			Parameters Unique to Class 2		
Draws:	1000	5000		1000	5000		1000	5000	
Parameter	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.
Rail constant RP	0.119	0.293	0.905						
Rail constant SP2	0.834	0.940	1.143						
Work trip dummy	1.94	1.96	1.26						
Fixed arrival time dummy	0.619	0.590	0.609						
Female dummy	1.04	1.04	0.529						
Cost per person in Guilders				-0.232	-0.220	0.104	-0.0412	-0.0406	0.0196
Out-of-vehicle time in hours				0.251	0.0541	1.5606	-1.44	-2.27	1.64
In-vehicle time in hours				-1.71	-1.61	0.76	-0.893	-0.909	0.375
Number of transfers				-0.207	-0.180	0.127	-0.158	-0.167	0.079
Amenities				-0.419	-0.415	0.166	-0.557	-0.566	0.241
Latent Comfort - RP	1.53	1.32	0.69						
Latent Comfort - SP2	1.75	1.62	0.53						
Latent Convenience - RP	2.08	1.90	1.04						
Latent Convenience - SP2	1.42	1.32	0.61						
Inertia dummy (RP Choice)	0.124	0.0277	1.0414						
Agent effect RP	2.09	2.24	1.61						
Agent effect SP2	2.45	2.73	1.13						
Scale (mu) SP1	2.20	2.21	0.92						
Scale (mu) SP2	1.48	1.38	0.43						
Tau1 SP1 (=-Tau4 SP1)	-0.242	-0.242	----						
Tau2 SP1 (=-Tau3 SP1)	-0.0158	-0.0157	----						
Tau3 SP1	0.0158	0.0157	0.0070						
Tau4 SP1	0.242	0.242	0.111						
Tau1 SP2	-0.937	-1.00	0.32						
Tau2 SP1 (=-Tau3 SP2)	-0.169	-0.181	----						
Tau3 SP2	0.169	0.181	0.071						
Tau4 SP2	1.23	1.31	0.43						
Log-likelihood (Choice&Latent):	-6419.63	-6423.09							
Log-likelihood (Choice):	-4282.48	-4284.96							

LATENT VARIABLE MODEL

Structural Equations (2 equations, 1 per column)

	Comfort Equation			Convenience Equation		
Draws:	1000	5000		1000	5000	
Parameter	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.
Constant - Comfort	0.149	0.132	0.158			
Constant - Convenience				0.497	0.497	0.245
Age dummy - over 40	-0.565	-0.540	0.400	0.851	0.876	0.246
First class rail rider	0.366	0.454	0.402			
In-vehicle time in hours	-0.417	-0.519	0.324			
Out-of-vehicle time in hours				-1.09	-1.23	0.54
Number of transfers				-0.139	-0.107	0.156
Free parking dummy (auto)				0.264	0.218	0.259
Variance(ω)	1.00	1.00	----	1.00	1.00	----

Measurement Equations (6 equations, 1 per row)

	Comfort Parameters			Convenience Parameters			Disturbance Params. (StdDev(υ))		
Draws:	1000	5000		1000	5000		1000	5000	
Equation	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.
Relaxation	0.559	0.551	0.183	0.160	0.156	0.134	1.15	1.15	0.10
Reliability	0.340	0.343	0.106	0.468	0.462	0.090	0.888	0.887	0.055
Flexibility				0.736	0.716	0.171	0.875	0.892	0.139
Ease				0.578	0.570	0.128	1.14	1.15	0.09
Safety	0.398	0.377	0.092	0.153	0.153	0.103	0.789	0.800	0.051
Overall Rating	1.09	1.10	0.38	1.43	1.44	0.26	1.40	1.37	0.18

CLASS MEMBERSHIP MODEL

Draws:	1000	5000	
Parameter	Est.	Est.	Std. Er.
Constant	-0.442	-0.375	0.467
Female dummy	-0.00192	0.0489	0.4128
Number of persons in party	0.169	0.165	0.125
Work trip dummy	-1.93	-1.85	0.74
Age over 40 dummy	-0.472	-0.496	0.384

Table E-8: Stability of Choice & Latent Variable Model with Random Parameters (Table 4-8)

CHOICE MODEL

	Choice and Latent Variable RP/SP Model with Randomly Distributed Parameters (latent variable portion below)						
	Location Parameters			Distribution Parameters			
Draws:	10000	20000		10000	20000		
Parameter	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.	
Rail constant RP	0.596	0.100	0.796				
Rail constant SP2	2.23	1.53	0.67				
Work trip dummy	1.28	1.07	0.83				
Fixed arrival time dummy	0.195	0.397	0.651				
Female dummy	1.55	1.48	0.63				
Cost per person in Guilders	-2.09	-2.18	0.29	1.18	**1.02**	**0.05**	*lognormal*
Out-of-vehicle time in hours	-0.699	0.0579	0.9423				
In-vehicle time in hours	0.462	0.228	0.305	0.938	0.864	0.040	*lognormal*
Number of transfers	-1.92	-2.14	0.38	1.60	1.76	0.15	*lognormal*
Amenities	-0.460	-0.609	0.271	1.24	**1.13**	**0.05**	*lognormal*
Latent Comfort - RP	3.29	2.98	0.84				
Latent Comfort - SP2	3.35	3.08	0.87				
Latent Convenience - RP	2.17	1.54	0.37				
Latent Convenience - SP2	1.64	1.18	0.37				
Inertia dummy (RP Choice)	-1.32	-1.05	0.57				
Agent effect RP	1.00	1.00	----				
Agent effect SP2	1.64	1.84	0.53				
Scale (mu) SP1	3.79	4.28	1.24				
Scale (mu) SP2	2.12	2.03	0.55				
Tau1 SP1 (=-Tau4 SP1)	-0.258	-0.229	----				
Tau2 SP1 (=-Tau3 SP1)	-0.0170	-0.0152	----				
Tau3 SP1	0.0170	0.0152	0.0053				
Tau4 SP1	0.258	0.229	0.083				
Tau1 SP2	-0.805	-0.812	0.220				
Tau2 SP1 (=-Tau3 SP2)	-0.142	-0.143	----				
Tau3 SP2	0.142	0.143	0.049				
Tau4 SP2	1.01	1.03	0.28				
Log-likelihood (Choice&Latent):	-6074.22	-6066.08					
Log-likelihood (Choice):	-3937.94	-3935.04					

LATENT VARIABLE MODEL

Structural Equations (2 equations, 1 per column)

	Comfort Equation						Convenience Equation						
	Location Parameters			Distribution Parameters			Location Parameters			Distribution Parameters			
Draws:	10000	20000		10000	20000		10000	20000		10000	20000		
Parameter	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.	
Constant - Comfort	0.0916	0.0688	0.1362										
Constant - Convenience							0.467	0.649	0.239				
Age dummy - over 40	-0.485	-0.435	0.145				0.963	0.961	0.286	-0.101	**-0.281**	**0.072**	*normal*
First class rail rider	-0.358	-0.434	0.211										
In-vehicle time in hours	-2.87	-3.03	0.43	1.75	1.96	0.15							*lognormal*
Out-of-vehicle time in hours							0.379	0.246	0.386	-0.536	-0.674	0.133	*lognormal*
Number of transfers							-0.162	-0.294	0.126				
Free parking dummy (auto)							-0.188	0.147	0.180				
Variance(ω)	1.00	1.00	----				1.00	1.00	----				

Measurement Equations (6 equations, 1 per row)

	Comfort Parameters			Convenience Parameters			Disturbance Params. (StdDev(υ))		
Draws:	10000	20000		10000	20000		10000	20000	
Equation	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.
Relaxation	0.384	0.408	0.138	0.201	0.136	0.084	1.20	1.20	0.07
Reliability	0.217	0.220	0.100	0.424	0.402	0.072	0.903	0.896	0.052
Flexibility				0.597	0.603	0.109	0.909	0.870	0.087
Ease				0.468	0.453	0.085	1.16	1.16	0.07
Safety	0.211	0.242	0.095	0.182	0.152	0.069	0.845	0.838	0.044
Overall Rating	0.949	1.05	0.13	1.24	1.12	0.12	1.45	1.39	0.14

Table E-9: Stability of Choice and Latent Variable Models with Heterogeneity of Latent Variable Parameters (Table 4-9)

CHOICE MODEL (Latent Variable Portion not Shown)

	Choice and Latent Variable RP/SP Model with Randomly Distributed Parameters (Lognormal)						Choice and Latent Variable RP/SP Model with Latent Class Heterogeneity					
Draws:	2000	10000		2000	10000		5000	10000		5000	10000	
Parameter	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.	Est.	Est.	Std. Er.
Rail constant RP	-0.383	-0.390	0.707				-0.389	-0.391	0.722			
Rail constant SP2	-0.923	-0.856	0.748				-0.797	-0.908	0.778			
Work trip dummy	1.80	1.76	0.74				1.70	1.72	0.66			
Fixed arrival time dummy	0.706	0.707	0.504				0.688	0.702	0.520			
Female dummy	1.21	1.16	0.48				1.17	1.17	0.48			
Cost per person in Guilders	-0.0648	-0.0637	0.0165				-0.0637	-0.0635	0.0174			
Out-of-vehicle time in hours	-1.18	-1.09	0.88				-1.12	-1.14	0.99			
In-vehicle time in hours	-0.742	-0.728	0.192				-0.727	-0.726	0.198			
Number of transfers	-0.105	-0.103	0.040				-0.103	-0.103	0.041			
Amenities	-0.384	-0.377	0.100				-0.377	-0.376	0.104			
	Location Parameters			Distribution Parameters			Class 1 Parameters			Class 2 Parameters		
Latent Comfort - RP	0.267	0.161	0.699	0.320	0.187	0.787	1.37	1.34	0.94	0.000	0.000	-----
Latent Comfort - SP2	0.280	0.186	0.391	0.291	0.340	0.079	1.42	1.42	0.63	0.000	0.000	-----
Latent Convenience - RP	0.233	0.267	0.467	0.362	0.314	0.511	1.48	1.48	0.61	0.000	0.000	-----
Latent Convenience - SP2	-0.430	-0.252	0.359	0.379	0.214	0.115	0.894	0.834	0.366	0.000	0.000	-----
Inertia dummy (RP Choice)	2.69	2.56	1.07				2.38	2.62	1.21			
Agent effect RP	0.225	0.256	0.566				0.245	0.125	0.571			
Agent effect SP2	2.21	2.10	0.61				2.11	2.12	0.66			
Scale (mu) SP1	2.16	2.20	0.58				2.20	2.21	0.61			
Scale (mu) SP2	1.20	1.26	0.38				1.26	1.24	0.41			
Tau1 SP1 (=-Tau4 SP1)	-0.208	-0.204	----				-0.204	-0.204	----			
Tau2 SP1 (=-Tau3 SP1)	-0.0135	-0.0133	----				-0.0133	-0.0132	----			
Tau3 SP1	0.0135	0.0133	0.0043				0.0133	0.0132	0.0044			
Tau4 SP1	0.208	0.204	0.060				0.204	0.204	0.062			
Tau1 SP2	-1.08	-1.03	0.31				-1.03	-1.05	0.34			
Tau2 SP1 (=-Tau3 SP2)	-0.198	-0.189	----				-0.188	-0.192	----			
Tau3 SP2	0.198	0.189	0.064				0.188	0.192	0.070			
Tau4 SP2	1.45	1.39	0.43				1.38	1.41	0.49			
Log-likelihood (Choice&Latent):	-6655.46	-6655.79					-6656.10	-6655.96				
Log-likelihood (Choice):	-4518.59	-4518.08					-4518.06	-4518.19				

CLASS MEMBERSHIP MODEL

Draws:	5000	10000	
Parameter	Est.	Est.	Std. Er.
Constant	2.46	2.50	1.39

Bibliography

Abelson, R.P. and A. Levy (1985) "Decision Making and Decision Theory", *Handbook of Social Psychology* **1**, G. Lindzey and E. Aronsom, Eds., Random House, New York.

Adamowicz, W., J. Louviere, and M. Williams (1994) "Combining Stated and Revealed Preference Methods for Valuing Environmental Amenities", *Journal of Environmental Economics and Management* **26**, 271-292.

Aitchison, J. and J. Bennett (1970) "Polychotomous Quantal Response by Maximum Indicant" *Biometrika* **57**, 253-262.

Ajzen, I. (1987) "Attitudes, Traits, and Actions: Dispositional Prediction of Behavior in Personality and Social Psychology", *Advances in Experimental Social Psychology* **20**, L. Berkowitz, Ed., Academic Press, 1-63.

Ajzen, I. (2001) "Nature and Operation of Attitudes", *Annual Reviews of Psychology* **52**, 27-58.

Amemiya, T. (1985) *Advanced Econometrics*, Harvard University Press, Cambridge, Massachusetts.

Anselin, L. (1989) *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers.

Ben-Akiva, M. (1973) *Stucture of Travel Passenger Demand Models*, Ph.D. Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Ben-Akiva, M. (1992) "Incorporation of Psychometric Data in Individual Choice Models", presented at the American Marketing Association Advanced Research Techniques Forum, Lake Tahoe, Nevada.

Ben-Akiva, M. and B. Boccara (1987) "Integrated Framework for Travel Behavior Analysis", presented at the International Association of Travel Behavior Research (IATBR) Conference, Aix-en-Provence, France.

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 a General Parametric Specification of the Covariance Structure", working paper, Massachusetts Institute of Technology.

Ben-Akiva, M., D. Bolduc, and M. Bradley (1994) "Estimation of Travel Choice Models with a Distributed Value of Time", *Transportation Research Record* **1413**.

Ben-Akiva, M. and J.L. Bowman (1998) "Activity Based Travel Demand Model Systems", *Equilibrium and Advanced Transportation Models*, P. Marcotte and S. Nguyen, Eds., Kluwer Academic Publishers.

Ben-Akiva, M., M. Bradley, T. Morikawa, J. Benjamin, T. Novak, H. Oppewal and V. Rao (1994) "Combining Revealed and Stated Preferences Data", *Marketing Letters* **5(4)**, 335-350.

Ben-Akiva, M. and S. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, The MIT Press, Cambridge, MA.

Ben-Akiva, M., D. McFadden, T. Gärling, D. Gopinath, J. Walker, D. Bolduc, A. Börsch-Supan, P. Delquié, O. Larichev, T. Morikawa, A. Polydoropoulou, and V. Rao (1999) "Extended Framework for Modeling Choice Behavior", *Marketing Letters* **10(3)**, 187-203.

Ben-Akiva, M. and T. Morikawa (1990) "Estimation of Travel Demand Models from Multiple Data Sources", *Transportation and Traffic Theory*, 461-476.

Ben-Akiva, M., J. Walker, A.T. Bernardino, D.A. Gopinath, T. Morikawa, A. Polydoropoulou (1999) "Integration of Choice and Latent Variable Models", forthcoming, International Association of Traveler Behavior Research (IATBR) book from the 1997 Conference in Austin, Texas.

Bentler, P.M. (1980) "Multivariate Analysis with Latent Variables", *Annual Review of Psychology* **31**, 419-456.

Bernardino, A.T. (1996) *Telecommuting: Modeling the Employer's and the Employee's Decision-Making Process*, Garland Publishing, New York. (also 1995 MIT Ph.D. dissertation, Department of Civil and Environmental Engineering.)

Berndt, E.R., B.H. Hall, R.E. Hall, and J.A. Hausman (1974) "Estimation and Inference in Nonlinear Structural Models", *Annals of Economic & Social Measurement* **3**, 653-665.

Bhat, C.R. (1995) "A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice", *Transportation Research B* **29(6)**, 471-483.

Bhat, C.R. (1997) "Accommodating Flexible Substitution Patterns in Multi-dimensional Choice Modeling: Formulation and Application to Travel Mode and Departure Time Choice", *Transportation Research B* **32(7)**, 455-466.

Bhat, C.R. (1998) "Accommodating Variations in Responsiveness to Level-of-Service Measures in Travel Mode Choice Modeling", *Transportation Research A* **32(7)**, 495-507.

Bhat, C.R. (2000) "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model", forthcoming, *Transportation Research*.

Boccara, B. (1989) *Modeling Choice Set Formation in Discrete Choice Models*, Ph.D. Dissertation, Department of Civil Engineering, Massachusetts Institute of Technology.

Bock, R.D. and L. Jones (1968) *The Measurement and Prediction of Judgement and Choice*, Holden-Day, San Francisco.

Bolduc, D. (1992) "Generalized Autoregressive Errors in the Multinomial Probit Model", *Transportation Research B* **26(2)**, 155-170.

Bolduc, D. and M. Ben-Akiva (1991) "A Multinominal Probit Formulation for Large Choice Sets", Proceedings of the 6th International Conference on Travel Behaviour **2**, 243-258.

Bolduc, D., B. Fortin and M.A. Fournier (1996) "The Impact of Incentive Policies to Influence Practice Location of General Practioners: A Multinomial Probit Analysis", *Journal of Labor Economics* **14**, 703-732.

Bollen, K. A. (1989) *Structural Equations with Latent Variables*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons.

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.

Börsch-Supan, A., D.L. McFadden, and R. Schnabel (1996) "Living Arrangements: Health and Wealth Effects", *Advances in the Economics of Aging*, D.A. Wise, Ed., The University of Chicago Press.

Boyd, J.H. and R.E. Mellman (1980) "The Effect of Fuel Economy Standards on the U.S. Automotive Market: An Hedonic Demand Analysis", *Transportation Research A* **14**, 367-378.

Bradley, M.A. and A.J. Daly (1997), "Estimation of Logit Choice Models Using Mixed Stated Preference and Revealed Preference Information", *Understanding Travel Behaviour in an Era of Change*, P.R. Stopher and M. Lee-Gosselin, Eds., Oxford: Pergamon, 209-232.

Bradley, M., T. Grosvenor, and A. Bouma (1988) "An Application of Computer-based Stated Preference to Study Mode-switching in the Netherlands", *Planning and Transport Research and Computation*, 307-319.

Brownstone, D., D.S. Bunch and K. Train (2000) "Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-fuel Vehicles", *Transportation Research B* **34**, 315-338.

Brownstone, D. and K. Train (1999) "Forecasting New Product Penetration with Flexible Substitution Patterns", *Journal of Econometrics* **89**, 109-129.

Bunch, D.A. (1991) "Estimability in the Multinomial Probit Model", *Transportation Research B* **25**, 1-12.

Cambridge Systematics, Inc. (1986) *Customer Preference and Behavior Project Report*, prepared for the Electric Power Research Institute.

Camerer, C. (1987) "Do Biases in Probability Judgment Matter in Markets? Experimental Evidence", *American Economic Review* **77**, 981-997.

Cardell, N.S. and F.C. Dunbar (1980) "Measuring the Societal Impacts of Automobile Downsizing" *Transportation Research A* **14**, 423-434.

Carroll, J.D. and P.E. Green (1995) "Psychometric Methods in Marketing Research. 1. Conjoint Analysis", *Journal of Marketing Research* **32(4)**, 385-391.

Case, A. (1991) "Spatial Correlation in Household Demand", *Econometrica* **59(4)**, 953-965.

Clark, C. (1961) "The Greatest of a Finite Set of Random Variables", *Operations Research* **9**, 145-162.

Cliff, A.D, and J.K. Ord (1981) *Spatial Processes, Models and Application*, Pion, London.

Daganzo, C. (1979) *Multinomial Probit: the Theory and its Application to Demand Forecasting*, Academic Press, New York.

Dansie, B.R. (1985) "Parameter Estimability in the Multinomial Probit Model", *Transportation Research B* **19(6)**, 526-528.

Debreu, G. (1960) Review of D. Luce *Individual Choice Behavior: A Theoretical Analysis*, *American Economic Review* **50**, 186-188.

Dennis, J.E. and R.B. Schnabel (1983) *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs.

Domencich, T. and D. McFadden (1975) *Urban Travel Demand*, North Holland, Amsterdam.

Dillon, W., A. Kumar and M.S. de Borrero (1993) "Capturing Individual Differences in Paired Comparisons: an Extended BTL Model Incorporating Descriptor Variables", *Journal of Marketing Research* **30**, 42-51.

Elrod, T. (1988) "Choice Map: Inferring a Product-Market Map from Panel Data", *Marketing Science* **7(1)**, 21-40.

Elrod, T. (1991) "Internal Analysis of Market Structure: Recent Developments and Future Prospects", *Marketing Letters* **2(3)**, 253-266.

Elrod, T. (1998) "Obtaining Product-Market Maps from Preference Data", American Marketing Association Advanced Research Techniques Forum, Keystone, Colorado.

Elrod, T. and M.P. Keane (1995) "A Factor-Analytic Probit Model for Representing the Market Structure in Panel Data", *Journal of Marketing Research* **32(1)**, 1-16.

Engel, J.F. and R.D. Blackwell (1982) *Consumer Behavior: Fourth Edition*, The Dryden Press

Engel, J.F., R.D. Blackwell and P.W. Miniard (1995) *Consumer Behavior: Eighth Edition*, The Dryden Press, Harcourt Brace College Publishers.

Engel, J.F., D.T. Kollat, and R.D. Blackwell (1968) *Consumer Behavior (1st Edition)*, New York, Holt, Rinehart, and Winston.

Everitt, B.S. (1984) *An Introduction to Latent Variable Models, Monographs on Statistical and Applied Probability*, Chapman and Hall.

Gärling, T. (1998) "Theoretical Framework", Working paper, Göteborg University.

Gärling, T. and M. Friman (1998) "Psychological Principles of Residential Choice", Draft chapter prepared for *Residential Environments: Choice, Satisfaction and Behavior*, J. Aragones, G. Francescato and T. Gärling, Eds.

Gärling, T., T. Laitila and K. Westin, Eds. (1998) *Theoretical Foundations of Travel Choice Modeling*, Elseveier Science, New York.

Goett, A., K. Hudson and K. Train (2000) "Customers' Choice Among Retail Energy Suppliers: The Willingness-to-Pay for Service Attributes", working paper, AAG Associates and University of California at Berkeley.

Golob, T.F., R. Kitamura and L Long, Eds. (1997) *Panels for Transportation Planning: Methods and Applications*, Transportation Research, Economics and Policy **5**, Kluwer Academic Publishers.

Gönül, F. and K. Srinivasan (1993) "Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues", *Marketing Science* **12(3)**, 213-229.

Goodman, L.A. (1974) "The Analysis of Systems of Qualitative Variables When Some of the Variables are Unobservable Part 1-A: Modified Latent Structure Approach", *American Journal of Sociology* **79**, 1179-1259.

Gopinath, D.A. (1995) *Modeling Heterogeneity in Discrete Choice Processes: Application to Travel Demand*, Ph.D. Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Gourieroux, C. and A. Monfort (1996) *Simulation-Based Econometric Methods*, Oxford University Press.

Green, P. (1984) "Hybrid Models for Conjoint Analysis: An Expository Review", *Journal of Marketing Research* **21**, 155-169.

Greene, W.H. (2000) *Econometric Analysis Fourth Edition*, Prentice Hall, Upper Saddle River, New Jersey.

Gupta, S. and P.K. Ghintagunta (1994) "On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models", *Journal of Marketing Research* **XXXI(1)**, 128.

Hajivassiliou, V. and P. Ruud (1994) "Classical Estimation Methods for LDV Models using Simulation", *Handbook of Econometrics* **IV**, R. Engle and D. McFadden, Eds., 2384-2441.

Harris, K.M. and M.P. Keane (1998) "A Model of Health Plan Choice: Inferring Preferences and Perceptions from a Combination of Revealed Preference and Attitudinal Data", Forthcoming in *Journal of Econometrics*.

Hausman, J.A. and D.A. Wise (1978) "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences", *Econometrica* **46**, 403-426.

Hawkins, D.I., R.J. Best and K.A. Coney (1989) *Consumer Behavior, Implications for Marketing Strategy: Fourth Edition*, BPI Irwin.

Hensher, D.A. and M. Bradley (1993) "Using Stated Response Choice Data to Enrich Revealed Preference Discrete Choice Models", *Marketing Letters* **4(2)**, 139-151.

Hensher, D.A. and A.J. Reyes (2000) "Trip Chaining as a Barrier to the Propensity to Use Public Transport", *Transportation* **27(4)**, 341-361.

Hosoda, T. (1999) *Incorporating Unobservable Heterogeneity in Discrete Choice Model: Mode Choice Model for Shopping trips*, Masters Thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Howard, J.A. (1977) *Consumer Behavior: Application of Theory*, McGraw-Hill, New York.

Howard, J.A. (1989) *Consumer Behavior in Marketing Strategy*, Englewood Cliffs, NJ: Prentice Hall.

Howard, J.A. and J.N. Sheth (1969) *The Theory of Buyer Behavior*, John Wiley and Sons, New York.

Imbens, G.W. (1992) "An Efficient Method of Moment Estimator for Discrete Choice Models with Choice Based Sampling", *Econometrica* **60**, 1187-1214.

Jacoby, J., G.V. Johar and M. Morrin (1998) "Consumer Behavior: A Quadrennium", *Annual Review of Psychology* **49**, 319-344.

Jöreskog, K.G. (1973) "A General Method for Estimating a Linear Structural Equation System", *Structural Models in the Social Sciences*. A.S. Goldberger and O.D. Duncan, Eds., Academic Press, New York.

Kahneman, D. and A. Tversky (1979) "Intuitive Prediction: Biases and Corrective Procedures", *Studies in Management Science* **12**, 313-327.

Kamakura, W.A. and G.J. Russell (1989) "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure", *Journal of Marketing Research* **25**, 379-390.

Keane, M.P. (1997) "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior", *Journal of Business and Economic Statistics* **15(3)**, 310-327.

Keesling, J.W. (1972) *Maximum Likelihood Approaches to Causal Analysis*, Ph.D. Dissertation, University of Chicago.

Khattak, A., A. Polydoropoulou and M. Ben-Akiva (1996) "Modeling Revealed and Stated Pretrip Travel Response to Advanced Traveler Information Systems", *Transportation Research Record* **1537**, 46-54.

Koppelman, F. and J. Hauser (1979) "Destination Choice for Non-Grocery-Shopping Trips", *Transportation Research Record* **673**, 157-165.

Lancaster, K. (1966) "A New Approach to Consumer Theory", *Journal of Political Economy* **74**, 132-157.

Lerman and Manski (1981) "On the Use of Simulated Frequencies to Approximate Choice Probabilities", *Structural Analysis of Discrete Data with Econometric Applications*, C.F. Manski and D. McFadden, Eds., The MIT Press, Cambridge, Massachusetts, 305-319.

Louviere, J.J., D.A. Hensher and J.D. Swait (2000) *Stated Choice Methods: Analysis and Application*, Cambridge University Press.

Luce, R.D. (1959) *Individual Choice Behavior*, Wiley, New York.

Luce, R.D. and J.W. Tukey (1964) "Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement", *Journal of Mathematical Psychology* **1**, 1-27.

Luce, R.D. and P. Suppes (1965) "Preference, Utility, and Subjective Probability", *Handbook of Mathematical Psychology* **3**, R.D. Luce, R.R. Bush and E. Galanter, Eds., John Wiley & Sons.

Manski, C. (1977) "The Structure of Random Utility Models", *Theory and Decision* **8**, 229-254.

Madanat, S.M., C.Y.D. Yang, Y-M Yen (1995) "Analysis of Stated Route Diversion Intentions Under Advanced Traveler Information Systems Using Latent Variable Modeling", *Transportation Research Record 1485*, 10-17.

Marschak, J. (1960) "Binary Choice constraints on Random Utility Indicators" *Stanford Symposium on Mathematical Methods in the Social Sciences*, K. Arrow, Ed., Stanford University Press, Stanford.

McCutcheon, A.L. (1987) *Latent Class Analysis*, Sage Publications, Newbury Park.

McFadden, D. (1974) "Conditional Logit Analysis of Qualitative Choice Behavior", *Frontiers of Econometrics*, P. Zarembka, Ed., Academic Press.

McFadden, D. (1978) "Modelling the Choice of Residential Location", *Spatial Interaction Theory and Planning Models*, F. Snickars and J. Weibull, Eds., North-Holland.

McFadden, D. (1981) "Econometric Models for Probabilistic Choice", *Structural Analysis of Discrete Data with Econometric Applications*, C. Manski and D. McFadden, Eds., Harvard University Press.

McFadden, D. (1984) "Econometric Analysis of Qualitative Response Models", *Handbook of Econometrics* **II**, Z. Friliches and M.D. Intriligator, Eds., Elsevier Science Publishers.

McFadden, D. (1986) "The Choice Theory Approach to Marketing Research", *Marketing Science* **5(4)**, 275-297.

McFadden, D. (1989) "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration", *Econometrica* **57(5)**, 995-1026.

McFadden, D. (1997) "Rationality for Economists", Presented at the NSF Symposium on Eliciting Preferences. Berkeley, California, July.

McFadden, D. (2000) "Disaggregate Behavioral Travel Demand's RUM Side: A 30-Year Retrospective", presented at the International Association of Travel Behavior Research (IATBR) Conference, Gold Coast, Queensland, Australia.

McFadden, D. and K. Train (2000) "Mixed MNL Models for Discrete Response", forthcoming, *Journal of Applied Econometrics*.

McLachlan, G.J. and K.E. Basford (1988) *Mixture Models: Inference and Applications to Clustering*, Marcel Dekker, New York.

Mehndiratta, R.M. and M. Hansen (1997) "Analysis of Discrete Choice Data with Repeated Observations: Comparison of Three Techniques in Intercity Travel Case", *Transportation Research Record* **1607**, 69-73.

Mellers, B.A., A. Schwartz an dA.D.J. Cooke (1998) "Judgment and Decision Making", *Annual Reviews of Psychology* **49**, 447-477.

Morikawa, T. (1989) *Incorporating Stated Preference Data in Travel Demand Analysis*, Ph.D. Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Morikawa, T., M. Ben-Akiva and D. McFadden (1996) "Incorporating Psychometric Data in Econometric Choice Models", Working paper, Massachusetts Institute of Technology.

Munizaga, M.A., B.G. Heydecker and J. de D. Ortuzar (2000) "Representation of Heteroskedasticity in Discrete Choice Models", *Transportation Research B* **34**, 219-240.

Muthén, B. (1979) "A Structural Probit Model with Latent Variables", *Journal of the American Statistical Association* **74**, 807-811.

Muthén, B. (1983) "Latent Variable Structural Equation Modeling with Categorical Data", *Journal of Econometrics* **22**, 43-65.

Muthén B. (1984) "A General Structural Equation Model with Dichotomous, Ordered Categorical and Continuous Latent Variable Indicators", *Psychometrika* **49**, 115-132.

Newey, W. and D. McFadden (1994) "Large Sample Estimation and Hypothesis Testing", *Handbook of Econometrics* **IV**, R. Engle and D. Mcfadden, Eds., 2111-2245.

Nicosia, F.M. (1966) *Consumer Decision Processes; Marketing and Advertising Implications*, Englewood Cliffs, N.J., Prentice-Hall.

Nicosia, F.M. and Y. Wind (1977) *Behavioral Models for Market Analysis: Foundations for Marketing Action"*, Dryden Press.

Olson, J.M., and M.P. Zanna (1993) "Attitudes and Attitude Change", *Annual Review of Psychology* **44**, 117-154.

Olson, P. (1993) *Consumer Behavior and Marketing Strategy: Third Edition*, Irwin, Inc.

Olsson, U. (1979) "Maximum Likelihood Estimation of the Polychoric Correlation Coefficient", *Psychometrika* **44**, 443-460.

Onkvisit, S. and J.J. Shaw (1994) *Consumer Behavior: Strategy for Analysis*, Macmillan College Publishing Company.

Ortuzar, J. de. D. and L.G. Willumsen (1994) *Modelling Transport, Second Edition*, John Wiley and Sons Ltd.

Pakes, A. and D. Pollard (1989) "Simulation and the Asymptotics of Optimization Estimators", *Econometrica* **57(5)**, 1027-1057.

Payne, J., J. Bettman, and E. Johnson (1992) "Behavioral Decision Rsearch: A Constructive Process Perspective", *Annual Review of Psychology* **43**, 87-131.

Pearson, K. (1894) "Contributions to the Theory of Mathematical Evolution", *Philosophical Transactions of the Royal Society of London A* 185**,** 71-110.

Polydoropoulou, A. (1997) *Modeling User Response to Advanced Traveler Information Systems (ATIS)*, Ph.D. Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Prashker, J.A. (1979a) "Mode Choice Models with Perceived Reliability Measures", *Transportation Engineering Journal* **105 TE3**, 251-262.

Prashker, J.A. (1979b) "Scaling Perceptions of Reliability of Urban Travel Modes Using Indscal and Factor Analysis Methods", *Transportation Research A* **13**, 203-212.

Prelec, D. (1991) "Values and Principles: Some Limitations on Traditional Economic Analysis", *Perspectives on Socioeconomics*, A. Etzioni and P. Lawrence, Eds., M.E. Sharpe, London.

Quandt, R.E. (1970) *The Demand for Travel*, D.C. Heath, London.

Rabin, M. (1998) "Psychology and Economics", *Journal of Economic Literature* **XXXVI(1)**, 11-46.

Ramming, S. (2000) personal communication, Massachusetts Institute of Technology.

Rice, C. (1993) *Consumer Behaviour: Behavioural Aspects of Marketing*, Butterworth Heinemann.

Revelt, D. and K. Train (1998) "Mixed Logit with Repeated Choices: Households' Choice of Appliance Efficiency Level", *Review of Economics and Statistics* **80(4)**, 647-657.

Revelt, D. and K. Train (1999) "Customer-Specific Taste Parameters and Mixed Logit", working paper, University of California at Berkeley.

Ruiter, E.R. and M.E. Ben-Akiva (1978) "Disaggregate Travel Demand Models fo rthe San Francisco Bay Area", *Transportation Research Record* **673**, 121-128.

Schweitzer, M. (1995) "Multiple Reference Points, Framing, and the Status Quo Bias in Health Care Financing Decisions, *Organizational Behavior and Human Decision Processes* **63**, 69-72.

Simonson, I., Z. Carmon, R. Dhar, A. Drolet, and S.M. Nowlis (2001) "Consumer Research: In Search of Identity", *Annual Reviews of Psychology* **52**, 249-275.

Sinha, I. and W.S. DeSarbo (1997) "An Integrated Approach Toward the Spatial Modeling of Perceived Customer Value", The American Marketing Association Advanced Research Techniques Forum, Monterey, California.

Small, K. (1987) "A Discrete Choie Model for Ordered Alternatives", *Econometrica* **55**, 409-424.

Slovic (1972) "From Shakespeare to Simon: Speculations – and Some Evidence – About Man's Ability to Process Information", *Research Bulletin: Oregon Research Institute* **12**, 1-29.

Srinivasan, K.K. and H.S. Mahmassani (2000) "Dynamic Kernel Logit Model for the Analysis of Longitudinal Discrete Choice Data: Properties and Computational Assessment", presented at the International Association of Travel Behavior Research (IATBR) Conference, Gold Coast, Queensland, Australia.

Steckel, J.H. and W.R. Vanhonacker (1988) "A Heterogeneous Conditional Logit Model of Choice", *Journal of Business & Economic Statistics* **6(3)**, 391-398.

Stern, S. (1992) "A Method for Smoothing Simulated Moments of Discrete Probabilities in Multinomial Probit Models", *Econometrica* **60(4)**, 943-952.

Stern, S. (2000) "Simulation Based Inference in Econometrics: Motivation and Methods", *Simulation-based Inference in Econometrics: Methods and Applications*, R. Mariano, T. Schuermann and M.J. Weeks, Eds., Cambridge University Press.

Svenson, O. (1998) "The Perspective from Behavioral Decision Theory: On Models of Travel Choice", *Theoretical Foundations of Travel Choice Modeling*, T. Gärling, T. Laitila and K. Westin, Eds., New York: Pergamon Press, 141-172.

Swait, J. and M. Ben-Akiva (1987) "Incorporating Random Constraints in Discrete Models of Choice Set Generation", *Transportation Research B* **21(2)**, 91-102.

Thaler, R. (1991) *Quasi-Rational Economics*, Russell Sage Foundation.

Thurstone, L. (1927) "A Law of Comparative Judgement", *Psychological Review* **34**, 273-286.

Train, K.E. (1998) "Recreational Demand Models with Taste Differences Over People", *Land Economics* **74(2)**, 230-239.

Train, K. (1999) "Halton Sequences for Mixed Logit", working paper, University of California at Berkeley.

Train, K., D. McFadden and M. Ben-Akiva (1987) "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices", *Rand Journal of Economics* **18(1)**, 109-123.

Train, K., D. McFadden and A. Goett (1986) "The Incorporation of Attitudes in Econometric Models of Consumer Choice", Working paper, Cambridge Systematics.

Train, K.E., D.L. McFadden and A.A. Goett (1987) "Consumer Attitudes and Voluntary Rate Schedules for Public Utilities", *The Review of Economics and Statistics* **LXIX(3)**, 383-391.

Tversky, A. (1977) "On the Elicitation of Preferences: Descriptive and Prescriptive Considerations", *Iconflicting Objectives in Decisions*, D. Bell, R. Kenney and H. Raiffa, Eds., Wiley, New York.

Tversky, A. and D. Kahneman (1974) "Judgment Under Uncertainty: Heuristics and Biases", *Science* **185**, 1124-1131.

Varian, H.R. (1992) *Microeconomic Analysis, Third Edition*, W.W. Norton & Company, New York.

Vovsha, P. (1997) "The Cross-Nested Logit Model: Application to Mode Choice in the Tel-Aviv Metropolitan Area", *Transportation Research Record* **1607**, 6-15.

Vovsha, P. and S. Bekhor (1998) "The Link-Nested Logit Model of Route Choice: Overcoming the Route Overlapping Problem", *Transportation Research Record* **2645**, 133.

Wedel, M. and W.S. DeSarbo (1996) "An Exponential-Family Multidimensional Scaling Mixture Methodology", *Journal of Business & Economic Statistics* **14(4)**, 447-459.

Wiley, D.E. (1973) "The Identification Problem for Structural Equation Models with Unmeasured Variables", *Structural Models in the Social Sciences*, A.S. Goldberger and O.D. Duncan, Eds., Academic Press, New York.

Wood, W. (2000) "Attitude Change: Persuasion and Social Influence", *Annual Reviews of Psychology* **51**, 539-570.