Modeling Heterogeneity in Discrete Choice Processes: Application to Travel Demand

by

Dinesh Ambat Gopinath

B.Tech., Indian Institute of Technology, Madras (1989)
M.S., Massachusetts Institute of Technology, Cambridge (1992)

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 and Decision Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1995

Author ....
Department of Civil and Environmental Engineering
December 2, 1994

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

Accepted by ....
Joseph Sussman
Chairman, Departmental Committee on Graduate Students

Modeling Heterogeneity in Discrete Choice Processes:

Application to Travel Demand

by

Dinesh Ambat Gopinath

Submitted to the Department of Civil and Environmental Engineering
on December 2, 1994, in partial fulfillment of the
requirements for the degree of
DOCTOR OF PHILOSOPHY
in
Transportation Systems and Decision Sciences

Abstract

Discrete choice models and, more recently, latent variable models have been widely applied to problems in transportation, economics, geography, marketing, public policy and psychology. The primary motivation for this work is the need for a comprehensive framework for capturing heterogeneity in choice processes, and which is consistent with existing behavioral theories, emphasizes a causal structural formulation, is mathematically tractable, and empirically verifiable. Specifically, the framework must be flexible enough to capture *unobserved heterogeneity* stemming from:

1. *decision-protocols* adopted by individuals;
2. *choice sets* considered by individuals;
3. *taste variations* among the members of the population; and
4. *psychological factors* such as attitudes and perceptions which affect the decision-making process.

To this end, this thesis significantly advances upon existing approaches to incorporate heterogeneity. The models developed in this thesis have been catalyzed by the recognition of the significance of choice process heterogeneity and the potential for incorrect forecasts if we ignore it, coupled with the advances in estimation and modeling methods. Further, the availability of computational power has engendered the development of the sophisticated models.

In this thesis, we extend the conceptual frameworks of McFadden [1986] and Ben-Akiva and Boccara [1986] of incorporating psychometric data within choice models to better reflect the underlying behavioral process. Specifically, we advance a rich class of choice models which builds on the simplicity and elegance of microeconomic theory, and incorporates the key psychological factors which endeavor to explain and quantify *seemingly irrational or inconsistent* behavior.

We develop the *latent class choice model* (LCCM) wherein the unobserved constructs are discrete or categorical, and hence are characterized through *latent classes*. LCCM is useful in situations wherein the analyst postulates that the *factors* "generating" heterogeneity can be conceptualized as discrete or categorical constructs such as choice sets considered, decision protocols adopted, market segments, etc. As part of the development of the LCCM, we formulate different class membership models which assign individuals to classes. These class membership models are derived rigourously from a behavioral theory perspective, and through a set of *criterion functions* which represent *unobserved* attitudes, individual's constraints and decision rules. We also develop the *latent structure choice model* (LSCM) which incorporates the gamut of attitudinal and perceptual indicators through latent attitudes, perceptions and classes, and discuss issues of estimation. Operationally, LSCM links latent structure models, including latent variable models and the latent class model, with choice models.

The unique features of the developed methodology are demonstrated in three domains. In the first problem, we apply the latent class choice model for taste heterogeneity in the estimation of travel choice models with distributed value of time (VOT). The modeling approach is the use of concepts such as "cost-sensitivity" and "time-sensitivity" to capture the degree to which individuals weigh travel cost and travel time in the choice process. The substantial improvement in the overall fit of the estimated models demonstrates the potential of the latent class approach for capturing taste variations, and indicates its efficacy and practicability compared to extant approaches of introducing interaction variables in the systematic utility functions, and the random coefficients model. The models also evidenced the significance of the unobserved variations in the VOT in the sample which persisted even after the systematic variations due to socio-economic and demographic variables were accounted for. Further, the models capture certain segments of the population having considerably higher VOT.

In the second problem, we apply the latent class choice model for decision protocol heterogeneity in a transportation mode choice context with data from simulated choice experiments. The estimated models have significantly better explanatory power compared to the standard probit model. Further, we observe that a significant fraction of the sample do not adopt the utility maximizing decision protocol. We also note the significance of the effects of the choices made by the individuals in the actual market environment on the decision protocol adopted in the choice experiments.

In the third problem, a class of choice models incorporating attitudinal data in choice models is formulated. The key feature of this approach is the concept of *latent attitudes*, and the respondent's responses to attitudinal questions relating to importance of attributes of alternatives (usually referred to as importance ratings) are manifestations of these attitudes towards the different attributes. The approach is demonstrated in a shipper's freight mode choice context.

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

Acknowledgments

I take this opportunity to thank Prof. Moshe Ben-Akiva for being a wonderful advisor, technical and otherwise, a constant source of encouragement since my arrival at MIT, and for taking a keen interest in my career. I have learned to become a perfectionist. I have more to learn from him, and I look forward to working with him in the years to come in our future endeavors.

I thank the other members in my doctoral committee – Professors David Bernstein and John Little – for their critical review of different aspects of this research and assistance in its presentation. I also wish to thank:

Mark Bradley from the Hague Consulting Group, Carl Martland from MIT and Taka Morikawa from Nagoya University for providing the data sets used in the case studies, in a clean and convenient format, without which the significance of this work would not have been realized.

The MIT University Transportation Center for funding the research in part through a UTC grant, and the many MIT teaching assistantships which funded my education.

My fellow students and friends including Adriana, András, Antulio, Ashok, Jai-kue, Jiang, Joan, John, Kazi, Leech, Milly, Nagi, Nicola, Prodyut, Qi, and Ted for making my long tenure at MIT enjoyable.

Denise, Rabi and Shenoi for their friendship, for the long discussions, academic and otherwise, and for buoying me up during difficult times. My special thanks to them.

Cobol, Sumanth and Tom for the fun time away from school, and for the many philosophical discussions and heated arguments.

Faculty including Professors Barnhart, Sussman and Wilson for their help in providing the perfect academic and research environment, and staff including Julie, Lisa and Paula for making my stay pleasant.

Finally, I wish to thank my parents – M. Gopinatha Menon and Shantha G. Menon – and my brothers – Suresh and Ramesh – for their encouragement and unconditional support for my intellectual pursuits.

To

Achan and Amma

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis concerns itself with enhancing existing tools to model the behavior of an individual[1] choosing an alternative[2] from a set of mutually exclusive and collectively exhaustive alternatives. Such tools, referred to as discrete choice models, have been of interest to researchers and practitioners for many years in a gamut of disciplines, such as transportation, economics, marketing, geography, psychology, public policy, and bioassay (see for example, Ben-Akiva and Lerman [1985], Train [1986], Anderson *et al.* [1992] for the theory and application of such models). These models do not attempt to predict deterministically the choice of an individual. Rather, they predict the probability of an individual picking an alternative. Further, the emphasis is on capturing the underlying process undertaken by the individual, hitherto referred to as the *discrete choice process* or the choice process for brevity, while choosing an alternative.

[1]We use the terms "individual", "consumer" and "decision-maker" interchangeably throughout the thesis.

[2]Alternative refers to an option available to an individual in a particular decision context. In a travel choice context, the option may be multi-dimensional characterized over time-of-day of travel, destination, travel mode, and route.

1.1 Motivation

To motivate the thesis objectives, we present briefly the views of two fields of research which have eyed the choice process from seemingly disparate perspectives.

1.1.1 Microeconomic Choice Theory

Microeconomic theory concentrates on the choice process by postulating that the individual is an "optimizing black box" (see, for example, McFadden [1973, 1986], Manski [1973, 1977], Samuelson [1983]). Thus the underlying decision protocol assumed is that the individual chooses an alternative which maximizes his/her *utility* or well-being. The attributes[3] of different alternatives, individual characteristics, past experiences, ambient information and situational constraints are inputs to the black box, while the observed choice is the output. The operationalization of the microeconomic choice theory relies heavily on the individual's preferences towards alternatives as elicited in the actual market environment, usually referred to as *revealed preferences*. Further, it assumes that the individual has perfect information about available alternatives and their attributes. But in reality, an individual has limited *prior information* about consumption opportunities and *information processing* capabilities. The individual's preferences to different alternatives may contain random components mainly due to the inability of the observer/analyst to measure all the inputs precisely. So the theory is operationalized by linking the random preference to the choice probabilities through a parameterized statistical model. Essentially, the model links directly the observed inputs to the observed output, and model parameters are then estimated through statistical procedures.

1.1.2 Choice Theory in Cognitive and Behavioral Science

Cognitive and behavioral scientists focus on the substantive theoretical aspects which guide the real-world choice process. For the last three decades, behavioral scientists

[3]For example, in a travel mode choice context, attributes may include travel times and travel costs of competing travel modes.

have been arguing that the microeconomic theory is not an accurate description of how individuals make decisions (e.g., Edwards [1954], Simon [1955, 1956], Kahneman and Tversky [1979], Abelson and Levy [1985]). In fact, cognitive psychologists argue that the analyst would be unjustified in believing that he/she will ever be possible to discover *quantitative laws* that apply to human behavior (Simon [1990]). In stark contrast to the basic tenets of microeconomic theory, people's preferences have been shown to be inconsistent. For example, Tversky [1969] presents evidence of intransitivity of preferences.

Kahneman and Snell [1990] note the vagueness of the concept of utility and argue that a distinction should be made between experience utility (satisfaction from use of an alternative), predicted utility (anticipated or expected satisfaction from use of an alternative) and decision utility (the weight actually given to an alternative when choosing). Further, social psychologists contend that social motives and pressures may sometimes be as important determinants of choice as individual's egoistic motives. Psychologists argue that the assumption of perfect information underlying microeconomic theory is rarely justified. Furthermore, the boundedly rational principle of Simon [1955] exemplifies the fact that the errors individuals generate in acquiring and processing information before making choices are systematic. Some economists agree with the psychologists that the microeconomic theory is incomplete (March [1978], Thaler [1992]).

The conceptualizations of cognitive psychologists and behavioral scientists are based on the following key features:

1. Individuals have limited information acquisition and processing capabilities which vary;

2. Prior experiences affect the choice process as individuals generate/update perceptions about alternatives and their attributes;

3. Individuals adopt different *heuristic* decision rules contingent on time pressure, information availability and reliability, etc. So in contrast to the *compensatory* utility maximization approach in the microeconomic choice theory, individuals

are hypothesized to adopt *non-compensatory* heuristic rules.

4. The social, economic and political environment in conjunction with individual's characteristics condition individual's attitudes, values and opinions, and orientation towards work, leisure, consumption patterns, and activities pursued, which in turn affect the choice process.

Unfortunately, these conceptualizations have not led to any operational choice models. On the other hand, economists assume that the above features are captured *implicitly* in the "black box".

1.1.3 Directions to Improve Choice Models

The "choice" of the direction pursued in this thesis, given these disparate views of choice processes, is derived from the following viewpoints:

- *Confluence of ideas from disparate fields*: The microeconomic and psychological choice theories are not antithetical, but instead can and should be utilized in conjunction with developments in psychometrics and econometrics to advance a richer class of choice models. Specifically, this class of models builds on the simplicity and elegance of microeconomic theory, and incorporates the key psychological factors which endeavor to explain and quantify *seemingly irrational or inconsistent* behavior.

- *Operational approach*: Keeping in mind that the bottom line objective of the choice modeling exercise is to have a predictive model which can be utilized for policy and operational analysis in the context of travel choice models, pricing and product portfolio decisions in the context of brand choice models, etc., we pursue an approach which leads to practical models.

- *Statistical approach:* To facilitate the testing of alternate behavioral hypotheses and microeconomic theories during the construction of choice models we adopt a statistical approach, thereby postulating that the observed choice behavior is the outcome of a probabilistic data generating process.

In recent years few researchers have acknowledged the importance of psychological factors, and have persevered to develop conceptual frameworks which integrate choice models, psychometric models, and a host of other marketing research methods (Koppelman and Hauser [1979], McFadden [1986], Cambridge Systematics [1986], Ben-Akiva and Boccara [1987], Ben-Akiva et al. [1994]). It is important to note that these conceptual frameworks and the choice models advanced in this thesis have been catalyzed on the following counts:

- *Recognition of choice process heterogeneity*: Recent literature has recognized the significance of heterogeneity, especially variations in individuals sensitivity to attributes of alternatives, in choice process, and that the choice process is affected by individual's attitudes toward and perceptions about alternatives and their attributes. Further, individuals may adopt different decision-making protocols as well as differ in the utility derived from nominally "identical" alternatives. It must be noted that such variations are different from observed variations such as inclusion of socio-economic and demographic characteristics in traditional choice models. The heterogeneity of the choice process stemming from such variations in individual's characteristics are observable. On the other hand, the focus of this thesis is on *unobservable* heterogeneity. It must be noted that models ignoring heterogeneity may produce incorrect forecasts.

- *Advances in estimation and modeling methods*: Recent methodological advances in estimation and computational methods include:

  - Choice set formation models: It is not always appropriate to impute the choice set deterministically from situational constraints. Probabilistic choice set models have addressed this issue by focusing on the existence of *random* constraints that imply the unavailability of certain alternatives (Ben-Akiva [1977], Swait and Ben-Akiva [1987a], Shocker *et al.* [1991]). More recent work by Boccara [1989] and Ben-Akiva and Boccara [1993] incorporates into a single framework of choice set formation modeling the effects of stochastic constraints and the influence of perceptions and attitudes on the choice set formation process.

– Dynamic choice models: In the literature, four major categories of intertemporal formulations of the dynamic choice process exist (Heckman [1981]:

  1. Models which assume that the multi-period choices in the individual's choice history are *independent*, and such models are as easy to estimate as the single period choice model.
  2. Models which allow the choice process at period t to depend on the *choice history* up to period t, and such models are referred to as *models with state dependence*. These models are relatively easy to estimate and one can utilize readily available software.
  3. Models which allow the random components of utilities of alternatives to be correlated over periods, and such models are referred to as *models with serial correlation*. These models are relatively harder to estimate and no stand alone software currently exists for estimation.
  4. Models wherein the choice process at period t depends on the choice history up to period t and the random components of utilities of alternatives are correlated over periods, and are referred to as *models with state dependence and serial correlation*. Such models are essentially intractable for estimation.

– Models using Revealed Preference (RP) and Stated Preference (SP) data: In order to exploit advantages of both RP data and SP data, methods have been developed (Ben-Akiva and Morikawa [1990a], Ben-Akiva et al. [1994]) which improve on the accuracy of parameter estimates in the RP model by sharing some of its parameters with the SP model, while potential biases and errors specific to SP data are explicitly considered in the SP model.

– Estimation by Simulation and Multinomial Probit Model (MNP): Recently a class of estimation methods such as the Method of Simulated Moments for choice models (McFadden [1989]), Method of Simulated Scores for limited dependent variable models (Hajivassilou and McFadden [1992]) have been developed paving the way for estimation through simulation of models which are impractical to estimate through numerical means (see also Geweke *et al.* [1992], Börsch-Supan and Hajivassilou [1993]). These developments have eased the estimation difficulty of MNP. Further, models with large choice sets where the interdependencies among alternatives are explicitly modeled have been developed (Ben-Akiva and Bolduc [1991]).

- *Availability of more refined data*: New information technologies (IT) have changed the manner in which data collection efforts are conducted by facilitating collection of large quantities of detailed data. For example, IT have eased the data collection process through "electronic questionnaires," computer logs of informational transactions and data collected through computer generated, realistic simulated experiments, usually referred to as stated preference data, by accelerating the decision-making environment. Modeling approaches were usually constrained/dictated by available data. Now, large quantities of disaggregate data are easier to collect, and are useful for the development of more refined models and "disaggregate" forecasting procedures.

- *Availability of computational power*: The availability of faster and cheaper computers, in conjunction with the availability of more refined data and analysis methods have engendered the development of both general purpose software (such as GAUSS from Aptech Systems [1993]) and specialized software (such as ALOGIT from Hague Consulting Group [1992] for the efficient estimation of Nested Logit Models) to estimate models hitherto considered computationally impractical, and thus paving the way for the development of better predictive models.

1.1.4 The Need to Model Unobserved Heterogeneity

For a systematic study of unobserved heterogeneity, we classify unobserved heterogeneity into four categories:

- *Unobserved decision protocols*: Most of the theoretical and empirical work in choice analysis is centered on the "utility maximizing" principle which assumes a rather sophisticated and cognitively demanding representation of the decision protocol. In reality individuals may adopt a variety of other decision protocols such as dominance rules, satisfaction rules, lexicographic rules, random choice, etc., of varying complexity (see for example, Slovic *et al.* [1977], Svenson [1979]).

- *Unobserved choice set*: Since the choice set actually *considered* by individuals can vary across the members of the population, this process must be explicitly treated in choice modeling to estimate the parameters of the choice model consistently. This theme of heterogeneity has been studied by Ben-Akiva [1977], Swait [1984], Swait and Ben-Akiva [1987a], and Ben-Akiva and Boccara [1990] by positing an explicit probabilistic choice set formation model.

- *Unobserved taste variations*: These refer to the variations of the parameters of the choice model across the members of the population. A significant part of this thesis is devoted to capturing such taste variations in choice models.

- *Unobserved attributes*: Some attributes are not directly observable in surveys, but which may be used by decision-makers while choosing an alternative from a choice set. Such attributes are usually individual's perceptions of alternatives and their attributes. For example, in a travel mode choice context, such attributes include "safety", "reliability", "comfort", etc.

1.2 Thesis Objectives

The primary objective for this work is to develop a general framework for modeling choice behavior which is consistent with existing behavioral theories, emphasizes a causal structural formulation, is mathematically tractable and empirically verifiable. Specifically, it must be flexible enough to:

- capture *unobserved heterogeneity* in choice behaviors stemming from:

1. *choice sets considered* by individuals;
2. *taste variations* across individuals;
3. *attitudes and perceptions* of individuals since the underlying choice process has a significant psychological component; and
4. *decision protocols* adopted by individuals in arriving at the choice.

- enable *conjunctive* use of data from different sources and in different response formats, including revealed preferences (RP) and stated preferences (SP), and psychometric data such as attitudinal and perceptual data, in the efficient estimation of choice models with explicit characterization of the different decision protocols which may be adopted in the actual market environment and the SP tasks.

1.3 Choice Modeling Framework

Before we detail the framework for incorporating the attitudinal and perceptual data, we briefly review the relevant psychology and sociology literature, to highlight how attitudes and perceptions may affect behavior, and to discuss issues of their measurement. For a more comprehensive study of these issues, the reader is directed to reviews by Fishbien and Ajzen [1972], Cooper and Croyle [1984], Chaiken and Stangor [1987], Tesser and Shaffer [1990], and Olson and Zanna [1993].

An attitude is an idea charged with emotion which predisposes a class of actions to a particular class of situations (see Rosenberg and Hovland [1960], Triandis [1971]). Thus attitude is theorized to consist of three components:

1. A *cognitive* component, that is the “idea” which is generally some category used by individuals in thinking (also referred to as “beliefs”).
2. An *affective* component, i.e., the emotion which charges the idea, such as “feels good” or “feels bad”, when the individual thinks about the category.

3. A *behavioral* component, which is a predisposition to action. This component is of immense interest to us since it is postulated to be linked to overt action.

Among the three components consistency may or may not exist. For example, an individual may have a positive affect towards luxury cars. But this does not necessarily lead to the purchase of one of the luxury cars given that the individual purchases a car. In general, overt behavior is conditioned by budget and situational constraints, acceptance of social norms, etc.

According to the functional theory of attitude and perception formation (see Smith [1947], Katz and Scotland [1959], and Katz [1960]), attitudes and perceptions are formed to understand the world around us, to adjust in this complex world, to protect our self-esteem, and to express our fundamental values. Further, attitudes and perceptions summarize the individual's complex interactions with the decision-making environment.

Conceptually, attitudes and perceptions are inferred from what an individual says about an object, from the way he feels about it, and from the way he behaves towards it. It must be noted that attitudes and perceptions involve what individuals *think* about, *feel* about, and how they *would like to behave* towards an object. Behavior is not only determined by what individuals *would like to do*, but also by what they should do (i.e., social norms, etc.), habits and expected consequences of the actions.

The general structure for incorporating attitudes and perceptions and multiple data sources such as observed market behavior and stated preference data, in choice modeling is presented in Figure 1-1. This framework builds on earlier conceptualizations of McFadden [1986], Ben-Akiva and Boccara [1987], and Morikawa [1989]. In the figure, ellipses represent *unobservable* constructs, while rectangles represent *observable* variables relevant to the problem context. Attitudes and perceptions of individuals are hypothesized to be key factors which characterize the underlying behavior. The socio-economic and demographic characteristics of the individual, attributes of alternatives, and information available to the individual 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 and other individual's socio-economic and demographic characteristics, affect the individual's preferences toward different alternatives.

In this framework, as in traditional random utility models, the individual's preference is assumed to be unobservable (hence a latent variable), and the actual market behavior and observed responses to alternate SP surveys with different response elicitation formats (if such data is also available), are only *manifestations* of the underlying preferences. Such observable variables which are manifestations of latent constructs are called *indicators* (Everitt [1984], Bollen [1989]). The responses to attitudinal and perceptual questions in surveys, form the corresponding indicators of attitudes and perceptions. Thus the observed RP and SP responses are linked to the preferences, while the attitudinal and perceptual indicators are linked to attitudes and perceptions.

Perceptions capture unobservable factors which affect the decision-making protocol. These factors are related to some problem-specific latent concepts. For example, in a travel mode choice context, such concepts for the transit alternative may include "safety", "convenience", "reliability", "environmentally friendly", etc. Perceptions are also related to the individual's "estimate" of the levels of different attributes of an alternative based on his/her available information. These perceptions are different from the "true" levels. Also, the individual's decision protocol is expected to be based on "perceived" levels of attributes and not on the "true" attributes. Perceptual differences in attribute levels arise from informational constraints or heterogeneity in information processing.

Responses to attitudinal questions in surveys represent or reflect individual's sensitivity to the different attributes of alternatives and "holistic" evaluations of alternatives. Such attitudinal data could be generic opinions or be obtained for a specific context. For example, attitudinal data collected on the importance of the attributes of alternatives from the perspective of the decision-maker reflects the individual's sensitivity to the attributes.

It is postulated that unobserved heterogeneity in the choice process may also be generated by *unobserved factors* which can be conceptualized as discrete or categorical latent constructs. For example, the sources of heterogeneity may include:

1. Segments of the population with varying tastes;

2. Choice sets considered by the individual which may vary; and

3. Different decision protocols adopted by individuals.

It must be noted that these sources of heterogeneity are not directly observable, and consequently are operationalized through the specification of unobservable concepts which we refer to as *latent classes*. Since the latent classes are discrete or categorical variables, a probabilistic latent class assignment process, referred to as the *class membership model* is postulated with the individual's attitudes and perceptions affecting the individual's class membership. Specifically, in this framework the latent classes are expected to capture latency which may be appear as:

> **Case 1:** The latent classes as well as the number of classes are well-defined. The latency is due to the analyst's inability to observe the classes. For example individuals in a particular choice situation may not consider all the deterministically available alternatives. Consequently, the choice set actually considered by an individual is not observable to the analyst and hence may be considered latent. In this situation, the possible choice sets considered are well defined as the power set[4] of the individual's deterministically available choice set.
>
> **Case 2:** The latent classes are not as well-defined and the classes are characterized through indicators of the latent classes which may be viewed as the latent class attributes. For example, market segmentation approaches seek to group consumers in terms of their sensitivities towards attributes of products. But, the actual number of latent segments and their characterization is not known until an "exploratory" data analysis.

[4] Ignoring the empty set of course.

Indicators of latent class depend on the characterization of the latent class. For example, consider a latent class which characterizes the latent choice set. Holistic evaluations of different alternatives which include responses to questions such as: "Would you consider alternative j as being available to you?" measured on a rating scale are manifestations of the perceived availability of the alternative, and hence form indicators of the latent class.

Now we turn our attention to the rationale for the utilization of both RP and SP data. As enunciated in Ben-Akiva *et al.* [1994], RP and SP data have certain advantages and certain disadvantages. The objective of combining multiple data sources is to exploit the advantages of each type of data and to overcome some of the disadvantages. Specifically RP and SP data are combined to address the validity of SP data, and to improve the accuracy of model parameter estimates. Consider a situation with RP and SP data are available. The key advantages of combining RP and SP data are:

- *Efficiency*: joint estimation of underlying preference from all the available data;
- *Bias correction*: explicit response models for SP data which include both preference parameters and bias parameters; and
- *Identification*: estimation of preference parameters not identifiable from RP data due to low variability.

The operationalization of the data combination method stems from the realization that the SP responses may not have the same relationship to latent preferences as revealed preferences do but clearly indicate some aspects of latent preferences. The linkages between SP responses, revealed preferences and the latent preference can be described by assuming different data generating processes for RP and SP data. The RP model represents actual behavior, while SP responses are represented by a different model. The task is then to estimate the unknown parameters in the preference functions of both models. The key feature of the combined RP/SP estimation method is that the preference functions have common parameters which typically represent the trade-off among the most important attributes of alternatives.

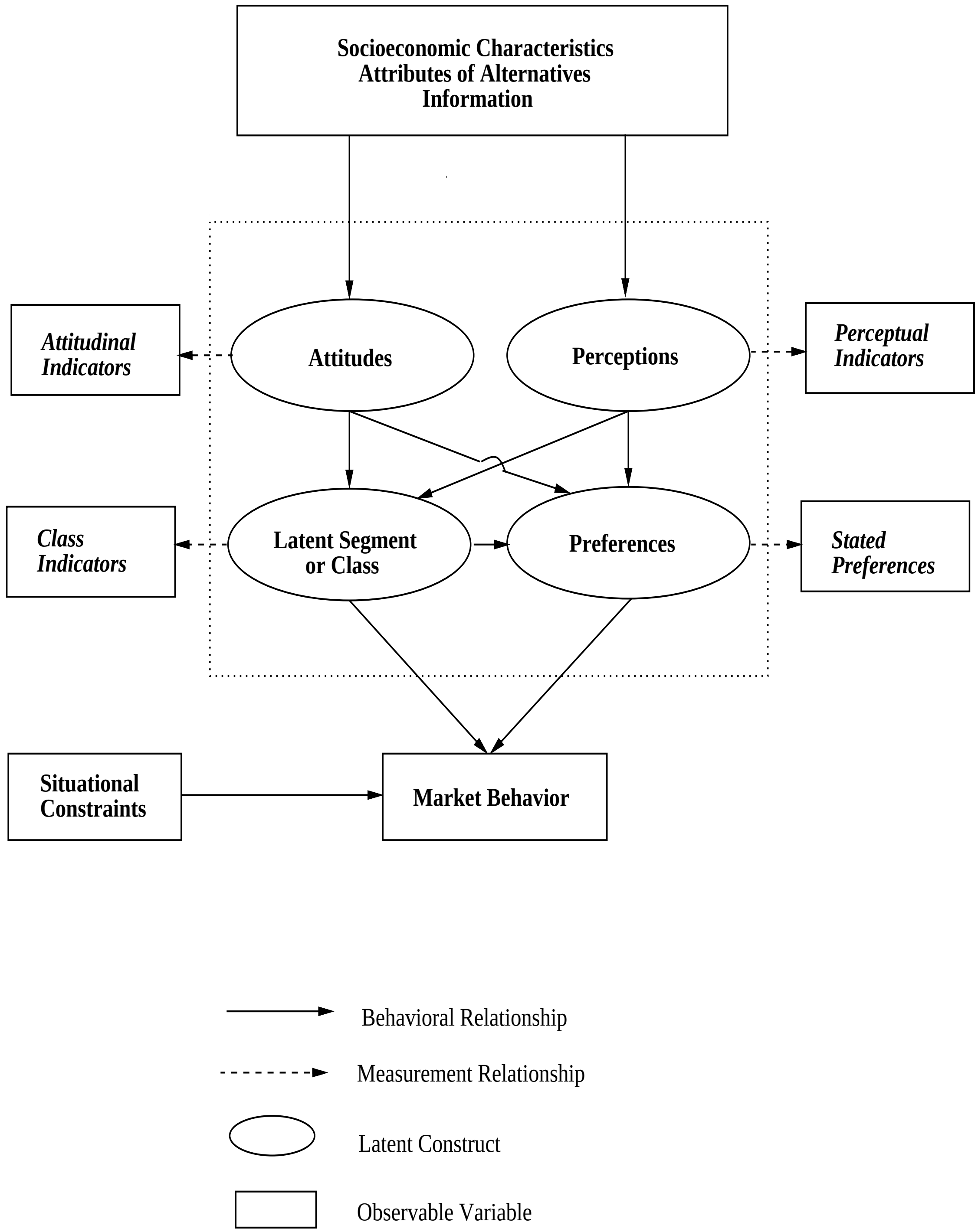


Figure 1-1: Choice Modeling Framework

1.4 Thesis Contributions

The main contribution of this thesis is methodological with an emphasis on capturing different forms of unobserved heterogeneity in choice models. The specific contributions of this thesis include:

1. We develop the *latent class choice model* (LCCM) wherein the latent constructs are discrete or categorical, and hence are characterized through *latent classes*. LCCM can be useful to capture unobserved heterogeneity in choice modeling situations wherein the analyst postulates that the *factors* "generating" heterogeneity can be conceptualized as discrete or categorical constructs such as choice sets considered, decision protocols adopted, etc.

2. As part of the development of the LCCM, we formulate different class membership models which assign individuals to classes. These class membership models are derived from a behavioral theory perspective, and through a set of *criterion functions*. These criterion functions may represent *unobserved* attitudes, individual's constraints and decision rules.

3. We extend and refine traditional latent class models (LCM) by linking the aforementioned class membership model (structural model) with the indicators of latent classes (measurement model). We also elaborate on the different types of specification of the measurement model depending on the characterization of the latent class. Further, the framework for the latent class model, presented herein, is analogous to the latent variable model.

4. We formulate and specify the *latent structure choice model* (LSCM) which incorporates the gamut of attitudinal and perceptual indicators through latent attitudes, perceptions and classes, and discuss issues of estimation. Operationally, LSCM links latent structure models, including latent variable models and the aforementioned latent class model, with choice models. Functionally, LSCM transcribes the main ideas presented in the conceptual framework for choice modeling illustrated in Figure 1-1, into an empirically testable statistical

model system wherein we postulate that the observed choice behavior or stated preferences is the outcome of a probabilistic data generating process molded by a host of psychological factors.

5. As a special case of LSCM, we develop a class of choice models which incorporates attitudinal indicators such as individual's importance ratings of attributes of alternatives. The emphasis is on "generating" unobserved taste variations from variations in attitudes.

In addition to the aforementioned methodological developments, the applications of the modeling approaches to enhance travel demand models form a significant and important part of this thesis. Specifically,

1. We apply the latent class choice model for taste heterogeneity in the estimation of travel choice models with distributed value of time (VOT). We demonstrate the efficacy and practicability of this model compared to extant approaches of introducing interaction variables in the systematic utility functions, and random coefficient models.

2. We apply the latent class choice model for decision protocol heterogeneity in a transportation mode choice context with data from simulated choice experiments. Since decision protocols in RP and SP settings may differ for the *same individual*, we also discuss the need to combine RP and SP data, and outline an approach to validate decision protocols exhibited in SP analysis with those of RP data, if both RP and SP data are available. This approach builds on previous work wherein choice models utilize both RP and SP data (Ben-Akiva and Morikawa [1990a]).

3. We apply the class of choice models incorporating attitudinal data, in a shipper's freight transportation mode choice study. In principle, we extend the work of Vieira [1992] by linking the choice model with an explicit causal model for attitude formation, and specifying responses to attitudinal questions in surveys as indicators of attitudes.

In the case study on distributed VOT, the substantial improvement in the overall fit of the estimated models indicate the potential of the latent class approach for capturing taste variations compared to extant approaches of introducing interaction variables and random coefficient models. Specifically, the models suggest the existence of two *unobserved* individual's sensitivity dimensions – cost-sensitivity and time-sensitivity. As expected, an increase in an individual's income level is associated with a decrease in his/her cost sensitivity, and consequently higher VOT. Time budget constraints arising from household and individual characteristics such as household type, employment status, age, gender, and available free time affect individual's time-sensitivity. In general, an individual in a household with children has higher VOT due to tighter time budget constraints, a part-time worker has higher VOT compared to a full-time worker, and VOT decreases for older people, especially for individuals 51 years or older. Further, a female commuter has a lower VOT compared to a male commuter, and as expected, an individual with lower available free time has higher VOT.

The models also evidenced the significance of the unobserved variations in the VOT in the sample which persisted even after the systematic variations due to socio-economic and demographic variables were accounted for. Consequently, prediction results from the estimated models reflect significant variations in the willingness to pay for travel time savings. In general, compared to the fixed coefficients model and a model wherein the implied VOT is lognormally distributed, the latent class choice model captures certain segments of the population having considerably higher willingness to pay. It is transparent that such variations in the VOT, if not properly accounted for, have substantial policy implications.

In the case study on capturing decision protocols through latent classes, the estimated latent class choice models have significantly better explanatory power compared to the standard probit model. We also note the significance of the effects of the actual choices made by the individuals in the RP context on the decision protocol adopted in the SP tasks. More specifically, we observe that the actual choice affects the choice process of the individual in the SP tasks through "inertial" effects in the systematic utility functions as well as "inertial" effects in the decision protocol adopted.

The estimated models provide only a preliminary assessment of the potential for capturing decision protocol heterogeneity. An empirical caveat in the estimation of such models is that the parameter estimates tend to be "sensitive" or "non-robust" in the sense that inclusion or exclusion of variables in the class membership model tends to change the choice model parameters appreciably. Further empirical work is needed to assess the differential impacts of including individual characteristics in the class membership model and the utility function, and their substantive significance and interpretation. In the data utilized in the case study the traveler's characteristics which potentially guide the "choice" of the decision protocol are limited. More empirical work with other surveys is necessary before such tools can be meaningfully adopted in practice.

In the case study on shipper's transportation mode choice, although the the effects of shippers attitudes such as cost-sensitivity and time-sensitivity on the choice models appear to be significant, the improvement in the overall fit over the fixed coefficient model is minimal. This is due in part to the limited variability of shipper's importance ratings of service attributes in the sample.

On the other hand, the shipper's attitude formation model provides strategic information to the marketing manager of a railroad. Considering the existence of an overall shipper's sensitivity to service attributes, we find that as the number of employees increase the shipper's sensitivity decreases, and firms with higher annual sales are more sensitive. Shippers with higher acceptable delays are less sensitive, and surprisingly, users of EDI are less sensitive as one would expect them to be more sensitive to service attributes, especially service-quality attributes such as payment terms and billing, responsiveness, etc. Further, shippers transporting high value goods over longer distances are more sensitive. Annual tonnage shipped and early acceptable delivery time do not seem to have an affect on the shipper's sensitivity.

Considering the existence of two shipper's sensitivity dimensions – time-sensitivity and cost-sensitivity – we observe that shippers with larger workforce, earlier acceptable delivery times and higher acceptable delays are less time sensitive. Shippers with higher acceptable delays are more cost sensitive, while shippers using EDI are less cost sensitive. Further, shippers with high annual tonnage are less cost sensitive, while shippers transporting high value goods over longer distances are more cost sensitive.

1.5 Outline of thesis

The remainder of this thesis is structured as follows:

- Chapter 2 reviews briefly choice models derived from the foundations of random utility theory, followed by a discussion of *latent structure models*. We also review existing methods to capture heterogeneity in choice models.

- Chapter 3 develops the latent class choice model, and formulates different class membership models.

- In chapter 4 estimation results of a special case of the latent class choice model to capture taste variations are presented. The case study conducted utilizes stated preference data wherein hypothetical travel alternatives were generated to evaluate traveler's trade-offs between travel time and travel cost, and we allow for taste variations to travel time and travel cost variables.

- In chapter 5 estimation results of a latent class choice model with explicit incorporation of the decision protocols adopted by individuals in a stated preference setting for travel mode alternatives are presented.

- In chapter 6 we advance a rich class of choice models, which incorporates attitudinal and perceptual data, and which are referred to as *latent structure choice models*. We also discuss approaches for the estimation of such models.

- In chapter 7, given the emphasis of the thesis on capturing heterogeneity, we elaborate on a class of choice models which incorporate attitudinal data. We also present estimation results for shipper's freight transportation mode choice models wherein the importance ratings of shippers of different attributes are utilized as indicators of sensitivity to attributes.

- Chapter 8 presents conclusions from this research and suggests future research directions.

Chapter 2

Literature Review

2.1 Introduction

As noted in chapter 1, we pursue a statistical approach to address heterogeneity in choice processes with an explicit characterization of psychological factors. To this end, in this chapter we review the relevant literature in three areas:

- *Random Utility Models*: To put this thesis in the context of existing probabilistic choice models, we review briefly the class of random utility models.
- *Psychometric Modeling*: To operationalize and quantify unobservable concepts such as social class, public opinion, personality, intelligence, etc., psychometricians have pioneered a class of models called *latent structure models*. We are interested in characterizing unobservable concepts such as attitudes and perceptions, and consequently the review will focus only on the methodological tools and not on the substantive issues in psychology and behavioral sciences.
- *Heterogeneity in Choice Processes*: This is the main theme of the thesis and the review will cover the extant ad hoc and model-based methods of capturing heterogeneity. Further, more recent attempts to integrate psychometric models with choice models will be discussed.

The remainder of this chapter is organized as follows: In section 2.2 an overview of random utility models is presented. In section 2.3 we review latent structure models. We categorize existing approaches to capture taste variations in choice models into: ad hoc grouping approach and model-based approach, and outline the techniques adopted in each in section 2.4 and section 2.5, respectively. Although this thesis focuses on static choice modeling, for completeness of exposition we discuss methods adopted to capture heterogeneity in discrete panel data models in section 2.6. This is followed by a review of models which address heterogeneity in choice sets in section 2.7, and models which incorporate psychometric data in section 2.8.

2.2 Random Utility Models

Extensive discussions of the micro-economic and psychological underpinnings of random utility models can be found in McFadden [1973], Manski [1977] and Ben-Akiva and Lerman [1985]. The model is based on the notion that the individual derives utility by buying or choosing an alternative, and the individual is postulated to pick that alternative which maximizes his/her utility. Since the utilities are not known to the analyst, they are treated as random variables. More specifically, the random utility of an alternative can be expressed as a sum of observable and unobservable components as:

$$U_{in} = V_{in} + \epsilon_{in}, \quad \forall i \in C_n \tag{2.1}$$

where

U_{in} = random utility of alternative i for individual n;

V_{in} = observable (systematic) component of utility;

C_n = choice set available to individual n with $|C_n| = J_n$; and

ϵ_{in} = random component of utility.

The systematic component V_{in} is written as:

$$V_{in} = V(X_{in}; \beta) \tag{2.2}$$

where

X_{in} = attributes of alternative i and characteristics of individual n; and

β = parameter vector.

Further under the maximum utility decision rule, the event that alternative i is chosen is linked to the random utilities as:

$$\{i \text{ chosen}\} \Leftrightarrow \{U_{in} \geq U_{jn} \quad \forall j\} \tag{2.3}$$

Consequently, the probability that alternative i is chosen by individual n is written as:

$$\Pr(i \text{ chosen}) = \Pr(U_{in} \geq U_{jn} \quad \forall j) \tag{2.4}$$

A class of probabilistic choice models can be constructed by appropriate specifications of the joint probability density of $(\epsilon_{1n}, \ldots, \epsilon_{J_n n})$. For example, if we assume that ϵ_{in} are independently and identically distributed Gumbel across alternatives and individuals with scale parameter set to 1 and location parameter set to 0, we obtain the choice probability in a closed form expression referred to as the Multinomial Logit Model (MNL) with

$$\Pr(y_{in} = 1 | X_n; \beta) = \frac{\exp(V_{in})}{\sum\limits_{j \in C_n} \exp(V_{jn})} \tag{2.5}$$

where

$$y_{in} = \begin{cases} 1 & \text{if alternative } i \text{ is chosen by individual } n \\ 0 & \text{otherwise} \end{cases}$$

$X_n = \{X_{in}, \forall i \in C_n\}$.

Further, if the systematic utility function $V(\cdot)$ is a *linear function in the parameters*, maximum likelihood estimation of the model parameters (if the estimates exist) is easy through gradient methods due to the global concavity of the log-likelihood function (McFadden [1973]) making the linear-in-parameters MNL the most widely used probabilistic choice model.

One of the important properties of the MNL model is termed the "independence of irrelevant alternatives" (IIA) in that for a specific individual the ratio of choice probabilities of any two alternatives is entirely unaffected by the systematic utilities of any other alternative. An important manifestation of this property is that the cross elasticities of choice probabilities of all alternatives with respect to a change in an attribute affecting only the utility of alternative j are equal for all alternatives $i \neq j$. This rather innocuous property leads to the popular *red bus/blue bus paradox* wherein the model gives counterintuitive forecasts. The fundamental cause of the paradox is the assumption of mutually independent disturbances in the MNL model. Further, this paradox is inherited by a much wider class of random utility models which rests on the assumption that the disturbances are independent.

The class of non-IIA models such as the Nested Logit (NL) model and the Multinomial Probit Model (MNP) attempt to address the problem of IIA. The NL model allows for a restrictive pattern of correlation among the random components of utilities and has the major advantage that the choice probabilities can be expressed in closed form. On the other hand, if $(\epsilon_{1n}, \ldots, \epsilon_{J_n n})$ is a multivariate normal random vector, we obtain the MNP model which allows for a more general pattern of correlation. But the estimation of MNP is computationally cumbersome for models with a large set of alternatives.

It is instructive at this point to discuss the sources of the random component of utility. Manski [1973, 1977] identifies four distinct sources of randomness. Without any loss of generality consider the case of *linear in parameters* and *linear in variables* systematic utility specification and a *pure idiosyncratic* identically distributed random component ϵ for each alternative (i.e., ϵ_i and ϵ_j are independent with var(ϵ_i)=var(ϵ_j)=σ_ϵ^2). We discuss how the different sources of the random component lead to correlation as well as heteroscedasticity among utilities of alternatives with some illustrative examples.

1. *Unobserved Variables*: Consider a situation wherein a particular variable z is unobserved (considered purely random with var(z)=σ_z^2) and this variable ought to have been in the systematic utility functions of alternatives i and j with corresponding parameters β_i and β_j. Thus the new random components for alternatives i and j are $\tilde{\epsilon_i} = \epsilon_i + \beta_i z$ and $\tilde{\epsilon_j} = \epsilon_j + \beta_j z$ respectively, and are correlated with $\text{cov}(\tilde{\epsilon_i}, \tilde{\epsilon_j}) = \beta_i \beta_j \sigma_z^2$. Further, the unobserved variable induces heteroscedasticity if $\beta_i \neq \beta_j$ since $\text{var}(\tilde{\epsilon_i}) = \sigma_\epsilon^2 + \beta_i^2 \sigma_z^2$ and $\text{var}(\tilde{\epsilon_j}) = \sigma_\epsilon^2 + \beta_j^2 \sigma_z^2$.

2. *Unobserved Taste Variations*: Assume that there exists a coefficient β which varies randomly among individuals with mean $\bar{\beta}$ and variance σ_β^2 and is independent of ϵ_i, $\forall i$. Let $\beta = \bar{\beta} + \nu$ where $\text{E}(\nu) = 0$ and $\text{var}(\nu) = \sigma_\beta^2$. If the random coefficient appears in two systematic utilities with variables x_i and x_j, then the new random components for alternatives i and j are $\tilde{\epsilon_i} = \epsilon_i + \nu x_i$ and $\tilde{\epsilon_j} = \epsilon_j + \nu x_j$ respectively, and are correlated with $\text{cov}(\tilde{\epsilon_i}, \tilde{\epsilon_j}) = x_i x_j \sigma_\beta^2$. As before heteroscedasticity is induced if $x_i \neq x_j$.

3. *Measurement Errors*: Consider a variable z which appears in the systematic utility functions of alternative i and j, with associated parameters β_i and β_j, and we only observe $\tilde{z}$ which is an imperfect measurement of z. Let $\tilde{z} = z + \nu$ where $\text{var}(\nu) = \sigma_\nu^2$. Then the new random components are $\tilde{\epsilon_i} = \epsilon_i - \nu\beta_i$ and $\tilde{\epsilon_j} = \epsilon_j - \nu\beta_j$ respectively, and are correlated with $\text{cov}(\tilde{\epsilon_i}, \tilde{\epsilon_j}) = \beta_i \beta_j \sigma_\nu^2$. In this case, the more serious problem of included variable being correlated with the random component arises because $\tilde{z}$ and ν are correlated.

4. *Instrumental Variables*: Suppose z is the variable which ought to appear in the systematic utility function, but instead we have an imperfect instrument $\tilde{z}$ with $z = g(\tilde{z}) + \nu$. The arguments in the previous case carry through leading to correlation among random components.

Therefore ignoring unobserved taste variations and unobserved attributes, if they really exist, may generate random disturbances which may not satisfy the assumptions made in the derivation of the probabilistic choice model such as the MNL model.

Methods to address taste variations have received, as we will see in this chapter, much attention with very many approaches and related empirical work. On the other hand, methods to capture unobserved factors have been developed in principle but empirical applications are rather limited.

2.3 Latent Structure Models

Certain concepts are not well defined in the behavioral and social sciences, like social class or personality. These concepts are not directly observable, and often referred to as *latent* constructs; they are hypothetical constructs conceived by an analyst with the intention of comprehending some research area of interest, and for which there exists no operational methods for direct measurement. Although latent constructs are not observable, one can hypothesize that their effects on measurable variables are observable.

Latent constructs occur in many areas; for example, in psychology *intelligence* and *verbal ability*, in sociology, *ambition and racial prejudice*, and in economics, *economic expectation*. In some cases, the observed variables, considered to be manifestations of the underlying latent construct (hence are also called manifest variables or indicators), will be *discrete* (nominal), in others *continuous* (interval or ratio) variables. Statistical models with latent constructs are in general referred to as *latent structure models*. Lazarsfeld and Henry [1968] categorize latent structure models based on the discrete and continuous nature of latent constructs and manifest variables into:

1. *Latent Class Model*: <u>Discrete</u> latent construct and <u>discrete</u> indicators (see McCutheon [1987] for a comprehensive treatment of latent class models).

2. *Latent Profile Model*: <u>Discrete</u> latent construct and <u>continuous</u> indicators.

3. *Latent Trait Model*: <u>Continuous</u> latent construct and <u>discrete</u> indicators. In this vein, we can view the random utility model as a latent trait model with continuous latent preferences and discrete choice indicator.

4. *Factor Analytic Model*: <u>Continuous</u> latent construct and <u>continuous</u> indicators.

The most well-known method for investigating the dependence of a set of manifest variables, is factor analysis (Lawley and Maxwell [1971], Johnson and Wichern [1982]). Initially, this technique was developed by psychologists, such as Spearman [1904] interested in examining ideas about the organization of mental ability suggested by a study of correlation and covariance matrices for sets of cognitive test variates.

The following description, adapted from Everitt [1984] and Bartholomew [1987], reflects the basic ingredients of the latent structure model. Let the vector $Z = [z_1, \ldots, z_P]$ denote the observed indicators, and the vector $S = [s_1, \ldots, s_M]$ denote the latent constructs. The number of latent constructs M is typically much smaller than the number of indicators P.

Let $f_{Z|S}(z|s)$ denote the joint probability density of Z given S. If the latent constructs are continuous, and letting the probability density of S be denoted by $f_S(s)$, the unconditional density of Z denoted by $f_Z(z)$ is given by

$$f_Z(z) = \int f_{Z|S}(z|s) f_S(s)\, ds \tag{2.6}$$

In general terms it is the density functions, $f_{Z|S}(z|s)$ and $f_S(s)$, that we would like to infer from the known or assumed density of $f_Z(z)$, in order to discover how the indicators depend upon the latent constructs. However, it is impossible to infer $f_{Z|S}(z|s)$ and $f_S(s)$ unless some assumption are made about their form. It is instructive to view $f_Z(z)$ as a mixture model where $f_{Z|S}(z|s)$ are the components of the mixture and $f_S(s)$ is the mixing density (McLachlan and Basford [1988]). If S is discrete or categorical, as in the latent profile and latent class models, the integral in equation (2.6) is replaced by a summation, and we have a *finite mixture* model. Else, $f_Z(z)$ is an *infinite mixture* model.

The crucial assumption of latent structure models is that of *conditional independence* which states that given the values of the latent constructs, the indicators are independent of one another. This is expressed as follows:

$$f_{Z|S}(z|s) = \prod_{i=1}^{P} f_{Z_i|S}(z_i|s) \tag{2.7}$$

The assumption of conditional independence implies that it is the latent constructs which produce the observed relationships amongst the indicators. The observed interdependence among the indicators is due to their common dependence on the latent constructs and that once these have been defined, the behavior of the indicators is essentially random.

In practice, it is assumed that the functional forms of the densities $f_{Z|S}(z|s)$ and $f_S(s)$ are known, but dependent on a set of unknown parameters. In such a case, the problem of inferring $f_{Z|S}(z|s)$ and $f_S(s)$ from $f_Z(z)$ becomes that of estimating the unknown parameters. Then using Bayes' theorem one can obtain the density of the latent constructs given indicators, denoted by $f_{S|Z}(s|z)$, from $f_{Z|S}(z|s)$, $f_S(s)$ and $f_Z(z)$.

In the foregoing discussion, the latent structure model did not have a specific *causal representation* for $f_S(s)$. Until the early 1960s, latent structure models contained only the relationships between the latent construct and indicators. The conceptual synthesis of causal models prevalent in econometrics literature and latent structure models was proposed by sociologists such as Blalock [1963] in the 1960s and early 1970s. A specific model with such a simultaneous representation was developed by Duncan *et al.* [1968]. Essentially, causal modeling is concerned with the estimation of the parameters in a system of simultaneous equations relating dependent and independent or explanatory variables. In the econometric literature, these two types of variables are termed *endogenous* and *exogenous*; the former are variables determined *within* the system, and their values are affected both by other variables in the system and by variables outside the system. In contrast, exogenous variables are those measured outside the system; they can affect the behavior of the system, but not themselves be affected by the fluctuations in the system. The endogenous variables may affect each other. A typical equation in the system attempts to explain one of the endogenous variables in terms of other endogenous variables, a number of exogenous variables, plus a disturbance term.

For ease of exposition, we categorize latent structure models based only on discrete or continuous nature of the latent constructs, i.e., latent variable models if the construct is *continuous*, and latent class models if the construct is *discrete*.

2.3.1 Latent Variable Models

A general and comprehensive approach to synthesizing econometric-type models with latent variables and psychometric-type measurement models has been primarily developed over the last two decades by a number of researchers including Keesling [1972], Jöreskog [1973], Wiley [1973], and Bentler [1980] (see, for example, the review in Cambridge Systematics [1986]). Specifically, these models assume that the indicators are continuous. Further, these models are referred to as *linear latent variable models* due to the linear specification of the relationships between the observed and latent variables[1]. Such a model consists essentially of two parts: a measurement model and a structural model. The first of these specifies how the latent variables are related to the indicators, and the second specifies the relationships among the latent variables. The structural model relates two types of latent variables – *endogenous* and *exogenous* – through *linear* structural equations of the form:

$$\eta = B\eta + \Gamma\xi + \zeta \tag{2.8}$$

In equation (2.8), η, the vector of latent endogenous random variables is $m \times 1$; ξ, the vector of latent exogenous random variables, is $n \times 1$; B is the $m \times m$ parameter matrix[2] showing the influence of the latent endogenous variables on each other; Γ is the $m \times n$ parameter matrix capturing the effects of ξ on η. The matrix $(I - B)$ is assumed to be non-singular. ζ is a $m \times 1$ disturbance vector with expected value of zero, and uncorrelated with ξ.

The measurement model can be written as:

$$y = \Lambda_y\eta + \epsilon \tag{2.9}$$

[1]Linear latent variable model is popularly referred to as LISREL model.

[2]Since a latent endogenous variable does not affect itself, the diagonal elements of B are set to zero.

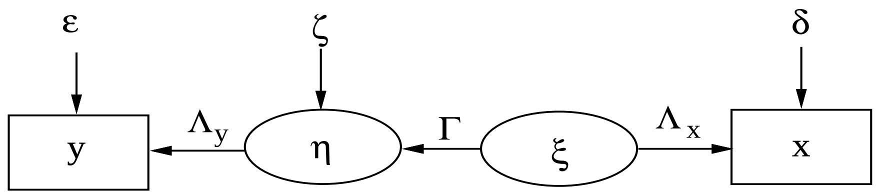


Figure 2-1: Linear Latent Variable Model: A Schematic Representation

$$x = \Lambda_x \xi + \delta \tag{2.10}$$

$y : (p \times 1)$[3] and $x : (q \times 1)$ form the indicators of η and ξ, respectively. $\Lambda_y : (p \times m)$ and $\Lambda_x : (q \times n)$ are parameter matrices that show the relation of η to y and ξ to x, respectively, and $\epsilon : (p \times 1)$ and $\delta : (q \times 1)$ are the errors of measurement for y and x, respectively. The expected values of ϵ and δ are zero, and ϵ and δ are assumed to be uncorrelated with η, ξ, ζ and with each other. To simplify matters y, x are written as deviations from their respective means (without any loss of generality). The latent variable model is schematized in Figure 2-1.

Let the unknown parameters[4] be stacked in a vector θ. S_{yy} represents the observed covariance matrix of the y variables, and S_{yx} the observed covariance between the y and x variables, and S_{xx} the covariance matrix of the x variables. Then the covariance matrix of the observed $[y', x']'$ is given by

$$S = \begin{bmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{bmatrix} \tag{2.11}$$

Let $\Sigma(\theta)$ represent the covariance matrix of the vector $[y', x']'$ implied by the model

[3]The notation $y : (p \times 1)$ denotes that y is a $p \times 1$ column vector.
[4]Parameters include those in the matrices B, Γ, Λ_y, Λ_x, and in the distributions of ξ, ζ, ϵ and δ.

system, i.e., as a function of the unknown parameter vector θ

$$\Sigma(\theta) = \begin{bmatrix} \Sigma_{yy}(\theta) & \Sigma_{yx}(\theta) \\ \Sigma_{xy}(\theta) & \Sigma_{xx}(\theta) \end{bmatrix} \tag{2.12}$$

Estimation of the parameters is based on the idea of replicating the observed covariance matrix with the implied covariance matrix. The model implies a particular structure for the population covariance matrix for the observed variables, in the sense that the elements are given by particular functions of the parameters of the model. Let Θ be the parameter space such that $\theta \in \Theta$. The parameter vector θ is obtained by minimizing a fitting function $\mathcal{F}(S, \Sigma(\theta))$ over $\theta \in \Theta$ satisfying the conditions [Everitt 1984]:

1. $\mathcal{F} \geq 0$
2. $\mathcal{F} = 0$ iff $S = \Sigma(\theta)$
3. $\mathcal{F}$ is continuous over S and $\Sigma(\theta)$

A comprehensive treatment of the theory and estimation of latent variable models is found in Everitt [1984] and Bollen [1989]. If the exogenous variables are assumed to be directly observable without any measurement error, equation (2.10) reduces to

$$x = \xi \tag{2.13}$$

Such a model system is referred to as a MIMIC (Multiple Indicators Multiple Causes) model. In a significant generalization of latent variable models, Muthén [1984] developed models which allow for dichotomous, ordered categorical and continuous indicators of latent variables[5]. Even in Muthén's work the latent variables are assumed to be continuous.

[5]It must be noted this generalization adopts the idea of the well established "threshold" crossing models such as the probit model and the ordinal probability model of McKelvey and Zavoina [1975].

2.3.2 Latent Class Models

Herein we turn our attention to models where the latent constructs are discrete. We provide only an overview of the latent class model and the reader is directed to McCutheon [1987] (see also Bartholomew [1987]) for an extensive discussion of identification and estimation of the model parameters. Assume that there exists S latent classes, P categorical indicators of the latent class with the p^{th} indicator taking on L_p levels, say $1, \ldots, L_p$. Let π_s denote the probability of an observation belonging to latent class s, where $s = 1, \ldots, S$. π_s is called the *latent class probability*. Conditional on the latent class s, the probability of the p^{th} indicator taking on level l_p is denoted by $\pi_{ps}(l_p)$. $\pi_{ps}(l_p)$ is called the *conditional response probability* of p^{th} indicator.

Under the assumption of conditional independence of the P indicators, the latent class model, which expresses the probability of observing the indicators $[l_1, \ldots, l_P]$, is written as:

$$\sum_{s=1}^{S} \prod_{p=1}^{P} \pi_{ps}(l_p)\pi_s \tag{2.14}$$

In the above representation, latent class probabilities are unaffected by any causal variables. We refer to such models as *simple latent class models*. A natural approach to address the issue of variations in the latent class probabilities is to group observations to examine group differences due to sex, race, geographic location, time of observation, etc. Consequently, a latent class model, referred to as a group-specific latent class model, is specified and estimated for each group of observations (see Clogg and Goodman [1984, 1985, 1986], Dayton and Macready [1980]). It is apparent that one may impose restrictions in the latent class probabilities and conditional response probabilities across the group-specific latent class models. To this end, two basic classes of models include: (a) models that allow for *partial homogeneity* across groups, and (b) models that allow *complete heterogeneity* across groups, where homogeneity or heterogeneity is with respect to the latent class probabilities and/or the conditional response probabilities. To clarify this categorization, assume that there exists G groups. Let π_s^g denote the probability of an observation from group g belong to latent class s. Let $\pi^g_{ps}(l_p)$ denote the probability that the p^{th} indicator of an observation from group g is in level l_p given that the observation belongs to latent class s. Then, the definition of complete homogeneity across groups would imply: $\pi^g_s = \pi_s$ and $\pi^g_{ps}(l_p) = \pi_{ps}(l_p)$ for all g. The definition of complete heterogeneity across groups would imply: $\pi^g_s \neq \pi^{g'}_s$ and $\pi^g_{ps}(l_p) \neq \pi^g_{ps}(l_p)$, for every g and g'. Consequently, partial homogeneity (or partial heterogeneity) refers to the case where some of the group-specific latent class and conditional response probabilities are equal across two or more groups.

We refer to the model which maps from the individual characteristics to the latent class probabilities as the class membership model. Dayton and Macready [1988] allowed only the latent class probabilities to depend on causal variables. Formann [1992] proposed logistic representations both for the latent class probabilities and the conditional response probabilities. Specifically,

$$\pi_{sn} = \frac{\exp(\theta'_s Z_n)}{\sum\limits_{s'=1}^{S} \exp(\theta'_{s'} Z_n)} \tag{2.15}$$

and

$$\pi_{ps}(l_p) = \frac{\exp(\gamma'_{ps;l_p} Z_n)}{\sum\limits_{l'_p=1}^{L_p} \exp(\gamma'_{ps;l'_p} Z_n)} \tag{2.16}$$

where $\theta_s \; \forall s = 1, \ldots, S$, and $\gamma_{ps;l_p} \; \forall l_p = 1, \ldots, L_p; p = 1, \ldots, P; s = 1, \ldots, S$, form unknown parameter vectors.

A similar approach for expressing latent class probability was independently and naturally developed in Gopinath and Ben-Akiva [1993] in a choice modeling setting wherein the indicators are individual's ratings of alternatives and the latent class represented *unobserved choice* (hence latent) with the latent class probability expressed by any *probabilistic choice model*. It must also be noted that a special latent class model with a structural model based on the "threshold crossing" idea was developed as part of a model to capture latent choice sets in choice models (Ben-Akiva and Boccara [1993]), although the significance of the development as a latent class model was not enunciated in their work. In a similar vein, Shyr [1993] in the context of rail fatigue analysis, adopted a probit model to express the latent class probabilities wherein the two latent classes corresponded to two "unobserved" types of rail defects.

Tests for Number of Latent Classes

To determine S heuristics may be adopted to address an analogous problem in finite mixture models (see, for example, Titterington, Smith and Makov [1985] and McLachlan and Basford [1988]). The standard generalized likelihood ratio statistic to test the null-hypothesis H_0 of S classes against the alternate hypothesis H_1 of $S+1$ classes is not asymptotically distributed chi-square since H_0 corresponds to a boundary of the parameter space for H_1, so that under H_0 the generalized likelihood test statistic is not asymptotically a full rank quadratic form (Ghosh and Sen [1985], Titterington [1990]).

Titterington *et al.* [1985], Anderson [1985], Yarmal-Vuarl and Ataman [1987] have proposed various test procedures for special types of component mixtures. Aitkin *et al.* [1981] and McLachlan [1987] apply Monte-Carlo test procedures to finite mixture problems. The basic idea behind such procedures is the comparison of likelihood ratio statistic for $S+1$ versus S latent classes from the data with a distribution of that statistic obtained from R datasets containing S classes, which are generated by replacing the unknown parameters in the component densities by their likelihood estimates from the original data. Such procedures are computationally cumbersome.

Another class of testing procedures are based on information criteria wherein a penalty is imposed on the maximized log-likelihood function. Sclove [1977] and Bozdogan and Sclove [1984] proposed the use of Akaike's Information criterion (AIC: Akaike [1974]) wherein the penalty equals the number of parameters estimated. To account for sample sizes, Bozdogan [1987] proposed the Consistent Akaike Information criterion (CAIC), while Schwartz [1987] proposed a Bayesian Information Criterion (BIC). The last two approaches are recommended when the data entail a large number of observations. It must be noted that, all these test procedures lack statistical rigor and rely on the same asymptotic properties as the likelihood ratio test (see

Sclove [1987]). Therefore, such tests are useful only in indicating the actual number of classes present.

2.4 Ad hoc Grouping Approaches to Capture Taste Variations

2.4.1 Individual-specific Models

Marketing researchers have popularized the idea of estimating individual-specific choice models to address taste variations in the population following the tradition of estimating such models in conjoint analysis. The "rationale" for using the responses from each individual to estimate an individual-specific model is highlighted in the following quote:

> "... the theoretical development of the logit model is based on utility-maximizing behavior at the individual or household level. Therefore, ideally, the parameters of the logit model should be estimated at the household level." (Chintagunta *et al.* [1991])

Unfortunately, estimating individual-specific models is not an efficient procedure as it ignores similarities in tastes which might exist across individuals. Further, the number of responses per individual is small precluding the employment of the classical large sample properties of consistency and efficiency of the maximum likelihood estimation procedure. In fact, small sample biases may be accentuated by the non-linearity of the likelihood function when compared to a linear model with the same sample size. Further, if there is no variability in the response pattern for an individual, then estimates do not exist for the individual-specific model.

2.4.2 A Priori Grouping by Observed Characteristics

This entails grouping data into subsets based on socio-economic and demographic variables such as age, income, gender, etc., and estimating separate choice models for each subset of observations. Likelihood ratio tests can be conducted to test similarities in preferences across the different subsets of observations (see chapter 7 of the textbook by Ben-Akiva and Lerman [1985] for an extensive treatment of such tests). The individual-specific choice models can also be viewed as a special case of grouping by individual.

2.4.3 Clustering-based Scheme

Cluster analysis of observable covariates (see, for example, Salomon [1980] for a travel choice situation) or latent variables (see Vieira [1992] for a shipper's freight transportation mode choice study) capture the (dis)similarity of the market segments in a multivariate space. In the latter case, latent variables are constructed using psychometric data data as indicators, and the clustering or a classification scheme is based on the latent variables. Pursuant to clustering, choice models are estimated for each cluster. The efficacy of the market segmentation in accounting for taste variations is tested by comparing these models with the performance of other segmentation schemes and with the pooled data model.

In the marketing research literature, clustering methods have been primarily utilized to develop segments of customers (see, for example, Doyle and Sanders [1985]), and to group similar/competing products (see, for example, Srivastava *et al.* [1981], Moore *et al.* [1986]) together to better understand the market structure. For more comprehensive reviews of application of clustering methods in marketing see Frank and Green [1968] and Punj and Stewart [1983].

The overall main drawbacks of ad-hoc grouping approaches are:

1. The need to separate the data into different groups and conduct separate estimation in each of the groups usually leads to imprecise parameter estimates due to (potential) small sample sizes in some groups.

2. Deterministic assignment of individuals into the different groups.

3. The grouping is typically conducted independent of the choice problem being analyzed since the grouping scheme precedes the analysis of the choice behavior.

2.5 Capturing Taste Variations through Model-based Approach

2.5.1 Random Coefficients Models

In the multiple regression model, a number of researchers have suggested that parameter heterogeneity can be assumed be randomly distributed in the population (see, for example, Hildreth and Houck [1968], Swamy [1971, 1974], Hsiao [1975]). This basic theme has also been pursued in the context of choice models. Economists usually assume the existence of a "representative" or "average" individual who is assumed to have tastes equal to the average over all the individuals in the population. The basic idea in the random coefficients model is the assumption that each individual n has his/her own taste parameter vector β_n and which differs from the average parameter vector $\bar{\beta}$ of the "representative" individual by an unknown (hence random) amount. Assuming a parametric distribution, $f(\beta;\Theta)$ for the taste parameter vector, the choice model is written as:

$$\mathrm{P}(y_{in}|X_n;\Theta) = \int \mathrm{P}(y_{in}|X_n;\beta)\, f(\beta;\Theta)\, d\beta \tag{2.17}$$

where $\mathrm{P}(y_{in}|X_n;\beta)$ is the choice model given β. The application of such an approach dates back to the work of Quandt [1968] in a binary choice situation. The Electric Power Research Institute (EPRI) implemented a form of the multinomial logit (MNL) model that allows for the parameters to be distributed across the population (EPRI [1977]). Such random taste variations can be naturally incorporated into MNP model when the random taste parameter vector is distributed multivariate normal and the systematic component of the utility function is *linear in the parameters* by employing the convenient convolution property of the normal distribution. Hausman and Wise [1978] estimated such a random coefficients probit model to analyze the travel mode choice decisions of commuters in Washington D.C. with the choice set containing three alternatives: driving alone, car pooling and public transit. Prediction tests of the fixed coefficient MNL model and the random coefficients MNP model revealed substantial differences, although the explanatory power of the MNP model is not much higher than that of the MNL model.

Fischer and Nagin [1981] present an empirical comparison of the fixed coefficients probit model and random coefficients probit model using data from an experimental setting. The respondents were faculty and staff in a university. Each respondent was requested to choose between pairs of parking spot alternatives, with each alternative being characterized by the attributes of price per year and walking distance (in minutes) from the parking lot to the building where the respondent worked. Individual-specific models using responses from each respondent, and fixed coefficient model and random coefficients model on pooled data were estimated. They conclude that after accounting for variations in individual characteristics such as income, substantial taste variations exist and may be adequately captured through the random coefficients model. Further, random coefficients model is shown to be robust to inappropriate specification compared to a fixed coefficient model.

Gönül and Srinivasan [1993] estimate a sequence of random coefficients MNL models for a choice situation with three alternatives (three brands of disposable diapers). The coefficients for price and promotion are allowed to vary randomly. They conclude that substantial unobserved taste variations exist and the explanatory power of the random coefficients MNL model is better than that of a fixed coefficients model. Further, in prediction tests the random coefficients MNL model performed better.

Some of the drawbacks of the random coefficients approach to capture unobserved taste variations are:

- In most choice situations, the analyst has prior expectations of the sign of the coefficients of important variables. For example, in a travel model choice context, the travel cost and travel time variables are expected to have negative coefficients to reflect the *disutility* associated with travel cost and travel time. If the random coefficients have unbounded support such as with the normal distribution, behaviorally implausible parameter values occur with positive probability however small it may be.

- Limited behavioral basis exists to aid in the specification of the distribution of the random coefficients, and the usual distributional assumptions are motivated by computational tractability. But, the choice model is susceptible to the distributional assumptions of the random coefficients (Heckman and Singer [1984]).

- Estimation of the MNL model with random coefficients is difficult as the choice probability calculation entails the evaluation of a multidimensional integral since a model with K random coefficients leads to a choice model which is expressed as a K-dimensional integral.

To address the difficulty in the estimation of the random coefficients model, researchers have adopted the idea of postulating a function for generating the random coefficients, whereby the randomness in K coefficients is generated by a deeper $\tilde{K}$-dimensional random variate where $\tilde{K} \ll K$. Such an approach is seen in Ben-Akiva *et al.* [1993] wherein a non-negative random variable such as a log-normal random variate is used to *scale* a subset of coefficients. It must be noted that this scaling does not change the signs of the coefficients. Further, the explanatory power of the random coefficients model is evidenced by the significant improvement in log-likelihood value compared to the fixed coefficients model. Gönül and Srinivasan [1993] adopt a similar idea in a multiplicative specification but pay no heed to maintaining the coefficient signs since a normal random variable is used as a multiplicand of the coefficients. In Appendix F we present a factor analytic representation for the generation of the random coefficients.

2.5.2 Choice Models with Latent Classes

In the marketing research arena, latent class models have become of late a popular tool to capture taste variations. The goal in segmenting the market, especially the market for consumer goods, is to group consumers into meaningful groups which have similar needs, tendencies and capabilities, and which react in a similar manner to specific marketing programs. The recognition that consumers differ in one or more respects has led to a stream of research on the theory and practice of segmentation (Frank, Massy and Wind [1972], Wind [1978]). Discriminant analysis and cluster analysis, as noted earlier, have been popular in segmentation research depending on whether the basis of segmentation is known in advance or is defined *a posteriori*. Marketing researchers take the view that there are many possible bases for segmentation including (Lehmann [1989]):

1. Grouping consumers based on similarities in a multi-dimensional variable space;
2. Grouping consumers based on similarities in the choice set considered; and
3. Grouping consumers for a particular choice problem based on the similarities in the relationships between consumer characteristics and the product category.

Grover and Srinivasan [1987] perform simultaneously market structure and segmentation by applying latent class analysis to brand switching data. But their approach does not explicitly account for the impacts of marketing mix variables such as price, promotions, features, advertisements, etc.

Kamakura and Russell [1989] propose a latent class – more popularly referred to as finite mixture in the statistics literature[6] (Titterington, Smith and Makov [1985] and McLachlan and Basford [1988]) – multinomial logit model with parameterized segment sizes, and each segment characterized by a vector of mean preferences and a single price sensitivity parameter. So the central idea is the partitioning of the market into consumer segments differing in both brand preferences and price sensitivity, and the existence of constant *prior* probabilities of an individual belonging to different consumer segments[7]. They apply this approach to study the competition between

[6]General necessary and sufficient conditions exist for the identification of finite mixture models, while the employment of these conditions to choice models with latent classes appears to be non-trivial.

[7]Rather, the prior probability of a random individual belonging to a particular consumer segment is the population share for that consumer segment.

national brands and private labels in one product category.

Zenor and Srivastava [1993] adopt a similar idea to identify market segments when only macro-level time-series data, such as market shares, are available. Estimates for segment characteristics such as size, brand preferences, and sensitivity to marketing mix variables are obtained by applying the latent segment logit model to aggregated panel data.

Dillon *et al.* [1993] adopt the ideas suggested by Dayton and Macready [1988] for incorporating causal variables such as individual characteristics in class membership model to capture individual differences in paired comparisons. They adopt an MNL-type class membership model with the individual characteristics utilized in the systematic functions. Similarly, Chintagunta and Gupta [1994] adopt an MNL-type class membership model in a multinomial choice context. Swait [1993] goes one step further where a MNL-type class membership model with latent variables such as individual's attitudes are utilized in the systematic functions.

The estimation of choice models with simple latent classes through the maximum likelihood criterion is difficult as they are plagued with the existence of many maxima. This is also due in part to the lack of a causal structure for latent class probabilities. To address this issue, researchers start from different starting values to ensure that the estimates are indeed the maximum likelihood estimates. In the presence of causal variables, such as the MNL-type class membership model, the causal variables (if relevant and properly specified) are expected to guide the algorithm to the global maximum. It must be noted here that since the characteristics of each latent class can be interpreted only *a posteriori* there is limited behavioral theory guiding in the specification of each systematic function. Further, the MNL model when used as a choice model is interpretable as it is derived from random utility theory, while such an interpretation is not feasible in the class membership model. So the initial stages of the model estimation necessitate using all the individual characteristics in each systematic function leading to a large[8] number of parameters in the latent class

[8] If one postulates the existence of S latent segments and Q individual-specific characteristics, the class membership model has $(S-1) \times Q$ parameters.

membership model.

It is instructive to view the choice model with latent classes as a special case of the random coefficients wherein the random coefficients have a non-parametric distribution. Specifically, choice models with a non-causal or simple class membership model, are the non-parametric versions of the usual random coefficients models. On the other hand, choice models with a causal class membership model, represent a random coefficients model wherein the distribution of the coefficients depend on causal variables. The choice models with class membership models are valuable in gaining insights into the extent of taste variations and the potential characterizations of the latent classes.

2.6 Heterogeneity in Discrete Panel Data Models

In marketing research, a popular method to capture inertia or brand loyalty of an individual, is the exponential smoothing model of brand loyalty used by Guadagni and Little [1983]. Fader and Lattin [1993] develop the Nonstationary Dirichlet Multinomial Model (NSDM) as an approach of capturing individual's loyalty to alternatives in discrete panel data. More specifically, the NSDM model was conceived to develop a new measure of alternative loyalty instead of the exponentially smoothed loyalty variable developed in Guadagni and Little [1983]. First, intrinsic Dirichlet heterogeneity in choice behavior across individuals is assumed. Second, the process over time is modeled as a renewal process wherein at each renewal the individual is assumed to "forget" his/her choice history. Fader and Lattin [1993] assume that the number of choice occasions since the last renewal for any individual is geometrically distributed. Given these assumptions, conditional on the choice history of an individual, one can easily calculate the expected probability of choosing each alternative. Fader and Lattin (conveniently) propose the natural logarithm of each alternative's probability so calculated as a measure of loyalty, and include this new measure as an additional variable in the systematic utility of the multinomial logit model. It must be noted that the Dirichlet-Multinomial model is an extension of the Beta-logistic

Model of Heckman and Willis [1977]. Such a model to capture cross-sectional heterogeneity is also seen in Ehrenberg [1988] and Fader [1993]. In appendix G, we present some characterizations of dynamic choice models derived from a bayesian approach of updating information from past choices.

2.7 Choice Models with Heterogeneity in Choice Sets

Most of the discrete choice literature assumes that the individual's choice set is known deterministically to the analyst, i.e., the availability of an alternative to an individual is treated as an observable binary variable – either an alternative is available to an an individual or it is not. There is both theoretical evidence (Swait and Ben-Akiva [1986]) and empirical evidence (Stopher [1980]) that misspecification in the choice set leads to choice model misspecification. Manski [1977] suggested a choice model with an explicit probabilistic choice set formation model as:

$$\mathrm{P}(i) = \sum_{C \in G} \mathrm{P}(i|C) \;\; \mathrm{Q}(C) \tag{2.18}$$

where

$\mathrm{P}(i)$ = choice probability of alternative i;

$\mathrm{P}(i|C)$ = choice probability of alternative i given that the choice set is C;

$\mathrm{Q}(C)$ = probability that C is the choice set;

M = universal choice set with J alternatives; and

G = set of all non-empty subsets of M.

The complete specification of the choice model entails the specification of:

1. a probabilistic choice set formation model, $\mathrm{Q}(C)$; and

2. a probabilistic choice model given the choice set, $P(i|C)$.

Since the number of elements in G, $2^J - 1$, will be large researchers have attempted to reduce the dimensionality of the choice set formation problem by placing *a priori restrictions* on the possible sets. For example, a latent captivity representation based on the restriction that individuals are either captive to an alternative or free to choose among all the alternatives, has been applied by Wermuth [1978], Gaudry and Wills [1979], Kitamura and Lam [1984] and Swait and Ben-Akiva [1987b].

Another approach is the independent availability model applied by Swait and Ben-Akiva [1987a] which imposes no restrictions on the possible choice sets, and which doles out probability masses to the $2^J - 1$ choice sets from J independent availability probabilities.

It is important to note that the probability model which describes the availability of each alternative to an individual is derived from a sound behavioral theory of *random constraints*. The random constraints approach is built on the theme that individuals are expected to have varying perceptions of the degree to which an operative constraint limits their access to certain alternatives. For example, in a travel mode choice context, the maximum acceptable walking distance to a subway stop is likely to vary across individuals. More recent work by Ben-Akiva and Boccara [1993] incorporates into a single framework of choice set formation modeling the effects of stochastic constraints and the influence of attitudes and perceptions on the choice set formation process.

In the marketing literature, choice set formation models have been gaining increasing attention. The problem of choice sets considered by individuals, also referred to as *consideration sets* has been studied by Roberts and Lattin [1991] and Hauser and Wernerfelt [1990] using a *compensatory process*. It can be argued that it is unlikely that an individual evaluates all the alternatives and the trade-offs among all attributes in order to eliminate a few in the first stage, and is expected to use simplifying heuristics (or a *non-compensatory* scheme) to restrict the choice set to a limited number of alternatives before choosing one. The information processing costs also usually preclude such a detailed evaluation. Gensch [1987] has provided empirical support to this argument that a non-compensatory first stage is followed by a compensatory second stage. The elimination-by-aspects model proposed by Tversky [1972a] is an example of a non-compensatory choice model, and Fader and McAlister [1990] have implemented such an approach. See Shocker *et al.* [1991] for a more extensive discussion of issues in consideration set formation.

In the area of behavioral decision research several theoretical and empirical studies (Miller [1956], Bruner [1958], Payne [1976], Wright and Barbour [1977]) support the two stages of choice process. Also, several studies (Payne [1976], Wright and Barbour [1977]) revealed that the individual's choice set reduction process is based on cutoff thresholds.

2.8 Incorporation of Psychometric Data in Choice Models

Most of the developments both in the theory and practice of discrete choice models have been in the context of revealed preference (RP) data. In recent years, however, there have been attempts to shift the focus to a more behaviorally rich paradigm of choice modeling (McFadden [1986], Ben-Akiva and Boccara [1987]). The main features of this new paradigm are:

- Explicit treatment of the psychological factors that affect the decision-making process; and
- Data sources other than revealed preferences data, such as stated preferences[9] (SP) can be effectively utilized in model development.

Ben-Akiva and Boccara [1987] identify four types of psychological factors:

1. Attitudes, needs and beliefs;
2. Perceptions;

[9]We assume any preference manifested *not* through actual market behavior fall into the category of stated preference data. Such data could be in different preference elicitation formats such as rating, ranking, indicated choice, etc.

3. Preferences; and

4. Behavioral intentions.

Behavioral intentions are usually manifested through SP data. For example, a popular method for measuring preferences in market research studies is conjoint analysis (see, for example, Green and Wind [1975]). In conjoint analysis, respondents are presented with descriptions of several hypothetical alternatives, each of which has different attributes. The respondent would be asked to indicate his or her relative preference towards each of the alternatives. The responses are used to infer the implicit weights respondents may use on each of the attributes while expressing their preferences. Random utility models have been applied to SP data to model individual choice behavior (e.g., Louviere and Hensher [1983], and Kroes and Sheldon [1986]). For the estimation of choice models, Morikawa *et al.* [1991] (see also, Morikawa [1989]) outlined the implications of the differing characteristics of SP and RP as:

- RP data are cognitively congruent with actual behavior;
- SP method form the only means of obtaining preferences toward new products and services; and
- Trade-offs among attributes are identifiable from SP data since the the attribute levels can be artificially set.

A fundamental problem associated with estimation of choice models from SP responses is the indifference of the respondent to the experimental task. Since a hypothetical scenario does not generally affect the value of the respondent (unlike actual market behavior), the respondent may be so uninterested and careless that he or she might not make a rational decision. Specific examples of such biases include: (1) *prominence hypothesis* wherein the respondent evaluates alternatives by considering the most important attribute, (2) *strategic behavior* or *policy-response bias* if the hypothetical scenario does affect the respondent's welfare, but it affects him or her in a way different from direct exposure to the "real market" situation, and the respondent believes that he or she will benefit by responding in a certain way, (3) *inertia* *bias* if the respondent prefers to maintain the status quo instead of changes posed in the SP surveys, and (4) *justification bias* wherein the respondent may want to justify past behavior and respond in that way even to a hypothetical scenario.

Recognizing the complementary characteristics of RP/SP data, Ben-Akiva and Morikawa [1990a, 1990b] have proposed a significant, albeit simple, combined RP/SP method for RP and SP data. The combined model is operationalized through the assumption of separate data generation processes for revealed preference data and the stated preference data with some commonalities.

Now we turn our attention to approaches to capturing psychological concepts such as attitudes and perceptions. It must be noted that such factors are unobserved (hence latent), and in principle one can adopt the latent structure models discussed in section 2.3 if adequate attitudinal and perceptual indicators are available. For example, Morikawa *et al.* [1990] present an intercity travel mode choice model wherein two perceptual attributes (ride comfort and convenience) are identified, with five point ratings (**(1)** very poor $\cdots$ **(5)** very good) of modal "attributes" such as: relaxation during the trip, reliability of the arrival time, flexibility of choosing departure time, ease of traveling with children and/or heavy baggage, safety during the trip, and overall rating of mode, serving as perceptual indicators. Further, the two perceptual factors are used as additional attributes in the choice model with associated coefficients.

McFadden [1986] suggests in a travel mode choice example an approach to capture individual's attitudinal factors (such as *cost consciousness*) by specifying additional variables in the systematic utility function which are interactions between the latent attitudes (e.g., cost consciousness) and relevant attributes (e.g., travel cost).

It must be noted that the conceptualizations of McFadden [1986] and Ben-Akiva and Boccara [1987] to incorporate psychometric data are significantly different from earlier paradigms such as Koppelman and Hauser [1979]. Earlier works adopted the notion that the perceptual and attitudinal indicators can be directly utilized as predictors in choice models.

2.9 Summary

In this chapter we reviewed the state of the art methods to capture taste variations in choice models, and approaches to characterize psychological factors such as attitudes and perceptions through latent structure models.

In the next chapter, we develop the latent class choice model. In principle this model builds on the theme in the choice model with latent classes reviewed in this chapter.

Chapter 3

Latent Class Choice Models

3.1 Introduction

In this chapter we present some of the key methodological developments of this thesis. We develop the *latent class choice model* (LCCM), wherein the latent constructs are discrete or categorical, and hence are characterized through *latent classes*. LCCM can be useful to capture unobserved heterogeneity in choice modeling situations wherein the analyst postulates that the *factors* "generating" the heterogeneity can be conceptualized as discrete constructs. For example, the sources of heterogeneity may include:

1. Different decision protocols adopted by individuals;
2. Choice sets considered by the individual which may vary; and
3. Segments of the population with varying tastes.

It must be noted that these constructs are not directly observable, and consequently are operationalized through the specification of latent classes. The emphasis of the presentation will be on the various types of class membership models which assign individuals to classes. These class membership models are derived through a set of *criterion functions*. The criterion functions may represent *unobserved* attitudes, individual's constraints and decision rules.

3.2 The Model

Before we present the LCCM, it is important to understand the potential forms of latent class characterization. As seen in section 2.2 most of the often used probabilistic choice models such as the MNL and MNP models assume a "utility maximizing" decision protocol wherein the individual is postulated to pick that alternative which maximizes his/her utility. But in reality individuals may adopt a variety of other decision protocols such as dominance rules, satisfaction rules, lexicographic rules, "random choice", etc. (see for example, Slovic *et al.* [1977], Svenson [1979]). Further, the decision protocol adopted by an individual is not directly observable to the analyst. Consequently, the unobserved decision protocol can be characterized by a D-dimensional latent class, where D equals the number of decision protocols postulated by the analyst. Each individual is expected to adopt *only* one of the decision protocols in a particular choice situation, and consequently belong to one of the latent classes. Therefore, the latent class is represented by a D-dimensional *binary* vector with only one of the components[1] taking the value 1, while all other components take the value 0. For example, if individual n adopts the d^{th} decision protocol, the class membership of the individual is represented by $T_n = [l_1 = 0, \ldots, l_d = 1, \ldots, l_D = 0]'$, with $\sum_{d'=1}^{D} l_{d'} = 1$. Correspondingly, the set of latent classes denoted by $\mathcal{M}_{DP}$ has D elements.

Consider a situation wherein the individual before making a choice considers only a subset of the alternatives available to him/her, and picks an alternative from this subset. The choice set actually considered is unobservable to the analyst, and consequently can be characterized by a latent class. In this case, the choice set considered can be viewed as a D-dimensional binary vector wherein the d^{th} component takes the value 1 if alternative d is considered and 0 otherwise. Therefore, the latent class

[1] The component *indicates* whether or not a particular decision protocol is adopted.

is denoted by a D-dimensional *binary* vector, $T_n = [l_1, \ldots, l_d, \ldots, l_D]'$, where

$$l_d = \begin{cases} 1 & \text{if individual } n \text{ considers alternative } d \\ 0 & \text{otherwise} \end{cases}$$

For example, if the universal set of alternatives $C = \{1, \ldots, 4\}$, and the choice set available to individual n is $C_n = \{1, 3\}$, then the class membership of the individual can be represented by $T_n = [l_1, \ldots, l_4]'$ where $l_2 = l_4 \equiv 0$, while for $d = 1, 3$ $l_d = 0$ or 1 depending on whether alternative d is considered. Herein it must be noted that since the empty choice set is neglected, the class denoted by $T_n = [0, 0, 0, 0]'$ should be eliminated. Consequently, the set of latent classes contains the 3 (i.e., $2^2 - 1$) elements, $\{1, 0, 1, 0\}$, $\{1, 0, 0, 0\}$, and $\{0, 0, 1, 0\}$. In general, if individual n has J_n alternatives in his/her choice set C_n, then there are $2^{J_n} - 1$ elements in the latent class set $\mathcal{M}_{CS;n}$, with $\mathcal{M}_{CS;n} \subseteq \mathcal{M}_{CS}$, where $\mathcal{M}_{CS}$ is the latent class set corresponding to the universal set of alternatives (i.e., $|\mathcal{M}_{CS}| = 2^J - 1$). Note that D equals J in this case. Further, the class membership denoted by $T_n = [l_1, \ldots, l_d, \ldots, l_D]'$ is such that $\sum_{d'=1}^{D} l_{d'} \geq 1$ (a value of 1 indicating captivity).

Consider a situation wherein the analyst expects unobserved taste variations to exist in the population, and that these variations can be adequately captured through the individual's sensitivity to different attributes. Further, assume there is a natural ordering of sensitivity to each attribute in levels[2] such as "low sensitivity" to "high sensitivity". Let $\tilde{K}$ denote the dimension of the taste parameter vector in a choice model. The taste variations in $\tilde{K}$ taste parameters can be "generated" by a set of D "deeper" sensitivity dimensions, with $D \leq \tilde{K}$. Specifically, each deeper sensitivity dimension captures the ordered sensitivity levels to one or more of the attributes. It must be noted that the basic idea in generating the variations in $\tilde{K}$ parameters through D deeper sensitivity dimensions is to capture interrelationships among individual's sensitivity to attributes, such as an individual having high (or

[2]It must be noted that the individual's sensitivity to an attribute is reflected in the magnitude of the corresponding coefficient in the utility function.

low) sensitivity to two or more attributes[3]. Consequently, the latent class is represented by D-dimensional vector with ordered levels in each dimension. For example, consider a simplistic travel mode choice situation wherein the relevant attributes include travel time, travel time reliability and travel cost. Assume that unobserved taste variations exist with respect to all the attributes. Further, assume that the analyst postulates that taste variations to travel time and travel time reliability can be generated through an unobserved "time sensitivity" dimension with three levels – high time sensitivity, medium time sensitivity, and low time sensitivity – with labels 1, 2 and 3, respectively, while taste variation to travel cost is generated through an unobserved "cost sensitivity" dimension with two levels – high cost sensitivity and low cost sensitivity – with labels 1 and 2, respectively. Specifically, the class membership of an individual be denoted by $T_n = [l_1, l_2]'$ where the first dimension represents time sensitivity and the second dimension represents cost sensitivity with $l_1 \in \{1, 2, 3\}$ and $l_2 \in \{1, 2\}$. Then the corresponding class-specific taste vector for the individual is written as:

$$\beta = \begin{pmatrix} \beta_{tt,l_1} \\ \beta_{ttr,l_1} \\ \beta_{tc,l_2} \end{pmatrix}$$

where β_{tt,l_1} and β_{ttr,l_1} are the travel time and the reliability of travel time coefficients when time sensitivity is in level, l_1, while β_{tc,l_2} is the travel cost coefficient when cost sensitivity is in level l_2. Therefore, in this example it must be noted that $\tilde{K} = 3$ and $D = 2$. Further, the number of elements in the latent class set $\mathcal{M}_{TV}$ equals 6 (i.e., product of the number of levels in each dimension). The idea of generating taste variations in $\tilde{K}$ taste parameters through possibly a smaller number of sensitivity dimensions is illustrated in the above example wherein the individual who is high sensitive to travel time is also expected to be high sensitive to travel time reliability, or vice versa. In general, the class membership is denoted by $T_n = [l_1, \ldots, l_d, \ldots, l_D]'$

[3]If an individual's sensitivity is associated with each alternative attribute then $D = \tilde{K}$. On the other hand, postulating $D < \tilde{K}$ allows capturing *prior* information or substantive knowledge of the choice context through the identification of a specific structure of the generation of individual's sensitivity to attributes. The "interrelationships" which we allude to are analogous to the correlations which may be imposed between the coefficients in a random coefficients choice model.

where $l_d \in \{1, \ldots, L_d\}$, L_d is the number of sensitivity levels in dimension d, and $\mathcal{M}_{TV} = \prod_{d=1}^{D} L_d$.

For notational simplicity the latent classes are indexed $s = 1, \ldots, S$ and the index s is associated with a *unique* element in the latent class set, i.e., $s \Leftrightarrow T_n = [l_1^s, \ldots, l_D^s]'$ for some $[l_1^s, \ldots, l_D^s]' \in \mathcal{M}$ where l_d may be a binary variable or an ordered categorical variable.

If the three different forms of unobserved heterogeneity coexist, then the latent class can be constructed through a superposition of each of the aforementioned latent classes. More formally, if $\mathcal{M}_{DP}$, $\mathcal{M}_{CS}$, and $\mathcal{M}_{TV}$ represent the latent class sets corresponding to decision protocols (DP), choice sets (CS), and taste variations (TV), respectively with S_{DP}, S_{CS}, and S_{TV} elements in each, then the superposed latent class may be represented by the cartesian product $\mathcal{M} = \mathcal{M}_{DP} \times \mathcal{M}_{CS} \times \mathcal{M}_{TV}$, with a total of $S = S_{DP} S_{CS} S_{TV}$ elements. Consequently, the superposed class membership for individual n, T_n, can be viewed as a concatenation of the corresponding class membership vectors $T_{DP;n}$, $T_{CS;n}$, and $T_{TV;n}$, i.e.,

$$T_n = \begin{pmatrix} T_{DP;n} \\ T_{CS;n} \\ T_{TV;n} \end{pmatrix}$$

To illustrate the superposition of different forms of latent classes consider a situation where the individual adopts one of two decision protocols – pick an alternative randomly or pick an alternative which maximizes utility – and even if the individual adopts a utility maximizing protocol, taste variation exists in two sensitivity levels – high sensitivity and low sensitivity, with labels 1 and 2 respectively. Consequently, the class membership vector for an individual who is a utility maximizer and having high sensitivity is written as:

$$T_n = \begin{pmatrix} T_{DP;n} \\ T_{TV;n} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

since

$$T_{DP;n} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

to denote that the second decision protocol is adopted, and

$$T_{TV;n} = (\,1\,)$$

to denote that the individual has high sensitivity to the attribute. In the superposed case the dimension of the class membership vector equals the sum of the dimensions of the class membership vectors corresponding to decision protocol heterogeneity and taste variations.

For notational brevity, in the following presentation we assume a generic D-dimensional latent class membership vector, where D equals

1. the number of decision protocols if the latent class characterizes unobserved decision protocols; or
2. the number of alternatives in the universal choice set if the latent class characterizes unobserved choice set; or
3. the number of sensitivity dimensions if the latent class characterizes unobserved taste variations; or
4. sum of the number of decision protocols, number of alternatives in the universal choice set, and number of sensitivity dimensions, or a partial sum thereof depending on the construction of the superposed latent class.

Notation for the Latent Class Choice Model

D = dimension of the class membership vector.

T_n = D-dimensional random vector which represents class membership, i.e., $T_n = [l_1, \ldots, l_D]'$ where l_d is the level in dimension d.[4]

[4]Note that $T_n = [T_{1n} = l_1, \ldots, T_{dn} = l_d, \ldots, T_{Dn} = l_D]'$ where T_{dn} denotes the d^{th} component of T_n. l_d may be a binary variable or an ordered categorical variable.

$\mathcal{M}$ = latent class set such that $T_n \in \mathcal{M}$.

y_{in} = choice indicator of individual n taking the value 1 if alternative i is chosen and zero otherwise.

C_n = choice set available[5] to individual n with $|C_n| = J_n$.

s = latent class index[6], $s = 1, \ldots, S$, where $S = |\mathcal{M}|$.

$$l^*_{sn} = \begin{cases} 1 & \text{if individual } n \text{ is in latent class } s \\ 0 & \text{otherwise} \end{cases}$$

β_s = choice model parameters[7] specific to class s.

C_s = choice set specific to class s.

R_s = decision protocol specific to class s.

X_n = attributes of alternatives and individual characteristics which affect the choice.

Z_n = attributes of alternatives and individual characteristics which affect the class membership.

We, must therefore, postulate the underlying mechanism for class membership and choice given the class to operationalize the model. This necessitates the formulation of two sub-models:

1. The *class membership model* assigns an individual to a latent class as a function of individual characteristics and attributes of alternatives, Z_n.[8] Since the

[5]This refers to the set of alternatives *deterministically* available.

[6]The index s is defined such that there is a one-to-one mapping between the latent class with index s and the class membership $T_n = [l_1, \ldots, l_D]'$, i.e., $(l^*_{sn} = 1) \Leftrightarrow T_n = [l^s_1, \ldots, l^s_D]'$.

[7]For simplicity we assume these parameters are fixed. In principle, we may allow random taste parameters for each class.

[8]It must be noted that the causal variables entering into the class membership model depend on the characterization of the latent class. For example, if the latent class represents unobserved choice set, then attributes of alternatives as well as individual characteristics may affect the assignment process. On the other hand, if the latent class characterizes taste variations, then only the individual characteristics may affect class membership.

analyst does not observe the class membership of an individual, a probabilistic assignment process is used. Let the *class membership model* be denoted by $Q_s(Z_n;\theta)$, i.e., the probability of individual n being assigned to latent class s can be written as:

$$P(l^*_{sn} = 1|Z_n) = Q_s(Z_n;\theta) \tag{3.1}$$

where θ is an unknown parameter vector.

2. The *class-specific choice model* predicts the choice behavior of an individual in latent class s. This sub-model is assumed to be class-specific, and therefore, depends on the choice set[9] (C_s), taste parameters (β_s), and the decision-protocol (R_s) associated with each class. Further, the class-specific choice model may be *deterministic* or *probabilistic*. For example, in a travel mode choice situation, if the individual in a particular latent class adopts the decision-protocol **"Pick the travel mode with minimum travel time"**, then the choice may be *deterministic* wherein the alternative with minimum travel time is chosen[10]. If the individual is a "utility maximizer" and considers the trade-offs among all the attributes of alternatives, a random utility model may be appropriate wherein the class-specific choice is *probabilistic*, and may be represented, for example, by a multinomial logit (MNL) model or a multinomial probit (MNP) model. The class-specific choice model expressing the choice probability of alternative i for individual n who is a member of class s can be written as:

$$P(y_{in} = 1|X_n; \beta_s, C_s, R_s). \tag{3.2}$$

[9]It must be noted that if the latent class characterizes the choice set considered, then the possible choice sets considered by an individual depends on C_n. Consequently, the class membership model doles out probability masses only to these possible choice sets, while the probability of considering any choice set containing deterministically unavailable alternatives is zero.

[10]In the event two or more travel alternatives tie for the minimum travel time, the choice may be probabilistic with *equal* probability of the individual choosing an alternative from the subset of alternatives of the class-specific choice set with minimum travel time. In general, if the individual "looks" at only one discrete or categorical attribute, then he/she picks an alternative from all the alternatives which possess this attribute with equal probability.

Using the class membership model, $\mathrm{Q}_s(Z_n;\theta)$, and the class-specific choice model, $\mathrm{P}(y_{in}=1|X_n,\beta_s,C_s,R_s)$, the latent class choice model for choosing alternative i is written as:

$$\mathrm{P}(y_{in}=1|X_n,Z_n;\theta,\beta)=\sum_{s=1}^{S}\mathrm{P}(y_{in}=1|X_n,\beta_s,C_s,R_s)\ \mathrm{Q}_s(Z_n;\theta). \tag{3.3}$$

Thus the log-likelihood function for a random sample of N individuals is given by:

$$\mathcal{L}(\beta,\theta)=\sum_{n=1}^{N}\log\left\{\prod_{i\in C_n}\left[\sum_{s=1}^{S}\mathrm{P}(y_{in}=1|X_n,\beta_s,C_s,R_s)\ \mathrm{Q}_s(Z_n;\theta)\right]^{y_{in}}\right\}. \tag{3.4}$$

The parameters $[\beta,\theta]$ can be obtained by maximizing the log-likelihood function.

It must be noted that the latent class choice model is built on the assumption that the data available to the analyst includes the choice indicator, attributes of alternatives, and socio-economic and demographic characteristics of the individual. Specifically, no data is available on attitudinal and perceptual indicators[11].

Given this overview of LCCM, we turn our attention to the class membership model, $Q_s(Z_n;\theta)$. In the following sections we present modeling approaches for cases wherein the underlying latent construct could be[12]:

- categorical as in the case of latent class characterizing decision protocols (*categorical criterion model*);
- binary latent class as in the case of latent class characterizing choice set (*binary criteria model*); and
- latent class with ordered levels in each dimension as in the case of latent class

[11] In this respect the term "latent class choice model" is somewhat of a misnomer since latent class models which we review in chapter 2 and detail in chapter 6 utilize the existence of indicators of latent classes. It must be noted that the choice itself can be construed as an indicator of both the underlying preference and the associated latent class. We extend the ideas developed within latent class choice models in chapter 6 wherein models incorporating attitudinal and perceptual indicators, referred to as *latent structure choice models*, are formulated.

[12] Although the class membership models are developed to address each of the specific forms of heterogeneity, it will be transparent to the reader how to extend the class membership model for the superposed forms of heterogeneity with some additional notational complexity.

characterizing taste variations through ordered levels of individual's sensitivity to attributes (*ordinal criteria model*).

The presentation is unique in its own right due to the causal formulation of the class membership models through a behavioral theory of unobserved criterion functions. It must be noted that the basic elements of the theory of the binary criteria model were embedded in the works of Swait and Ben-Akiva [1987a] and Ben-Akiva and Boccara [1993], although the significance of their work as a special case of a class membership model was not recognized.

3.3 Class Membership Model

First we note that since the latent classes are discrete or categorical variables, a *probabilistic* latent class assignment process is desirable. As the first step in the development of the class membership model, we postulate the existence of criterion functions which map from the variables Z_n to a vector of latent variables. Before we develop the different class membership models, it is important to discuss the behavioral interpretations of the criterion functions in different situations.

Suppose the latent class characterizes the individual's decision protocol. The decision protocol adopted by an individual can be viewed as being generated by a process wherein the different decision protocols compete with each other. The individual adopts the decision protocol which suits him/her the most. Consequently, each decision protocol is associated with a criterion function which captures the desirability of the decision protocol as a function of individual's characteristics such as time pressure, education, etc., coupled with intrinsic features of the decision protocol.

Consider a situation wherein the latent class characterizes the choice set actually considered by the individual. The formation of the choice set may be viewed as a process whereby the individual identifies feasible alternatives depending on individual's resource constraints, knowledge about competing alternatives and their attributes, and the individual's ability to process information about alternatives. Individuals are expected to have varying perceptions of the degree to which an operative constraint limits their access to certain alternatives. For example, the high price of an alternative may preclude its consideration in the individual's choice process, and the willingness-to-pay thresholds may vary across individuals depending on income levels. Consequently, the choice set considered by an individual may be generated through a set of non-compensatory (such as "satisficing") rules or constraints. Specifically, an alternative is considered if the set of non-compensatory rules associated with that alternative are satisfied. Since the rules adopted by individuals may arise from objective constraints as well as subjective constraints stemming from individual's attitudes and perceptions, and are not directly observable, they are considered *random*, and hence, may be operationalized through the criterion functions satisfying a set of inequalities.

Consider a situation wherein the latent class characterizes individual's sensitivity to attributes of alternatives in terms of the importance he/she places on each of the attributes. The individual's sensitivity to an attribute may be postulated to be a function of individual characteristics. Further, each sensitivity is unobserved, and hence operationalized through a criterion function.

Assume that there exists K_d criterion functions for each dimension d of the latent class membership vector. If the latent class characterizes the choice set considered, K_d represents the number of non-compensatory rules associated with alternative d; if the latent class characterizes individual's sensitivity to attributes K_d equals 1, and the associated criterion function represents the d^{th} sensitivity to attributes; and if the latent class characterizes the decision protocol wherein each decision protocol is associated with a single desirability concept as perceived by the individual, then K_d equals 1.

Let $H_{dkn}, \ \ \forall k = 1, \ldots, K_d; \forall d = 1, \ldots, D$ represent the k^{th} criterion function in dimension d. The criterion functions may be specified in terms of some function denoted $H(\cdot)$ such that

$$H_{dkn} = H(Z_n, \delta_{dkn}; \theta_{dk}) \tag{3.5}$$

where θ_{dk} is a parameter vector and δ_{dkn} is a random component associated with the criterion function to reflect the fact that H_{dkn} is unobserved and hence is a random variable. Further, for simplicity H_{dkn} may be separated into an additive systematic component (non-random component) denoted by $\tilde{H}_{dkn}$ and a random component, δ_{dkn}, i.e.,

$$\begin{aligned} H_{dkn} &= \tilde{H}_{dkn} + \delta_{dkn} \\ &= \tilde{H}(Z_n; \theta_{dk}) + \delta_{dkn}. \end{aligned} \tag{3.6}$$

where $\tilde{H}(\cdot)$ is some function which maps from Z_n to the systematic components. Given the specification of the criterion functions, the problem reduces to the development of the mapping from the criterion functions to the class membership probabilities. To this end, the different associations between the criterion functions and the latent classes may include:

1. If each dimension of the latent class is associated with a single criterion function (i.e., $K_d = 1, \quad \forall d = 1, \ldots, D$) as in the case of the latent class characterizing decision protocol, and a latent class is identified such that one and only one dimension of the multi-dimensional binary vector takes the value 1, then the total number of latent classes, S, equals D. Assuming that an individual is assigned to a latent class based on a *maximum criterion association rule*, the class membership model is analogous to a random utility choice model wherein the criterion function of each class is similar to a random utility associated with the class.

2. If the criterion functions $H_{dkn}, \quad \forall k = 1, \ldots, K_d$ map into a binary variable for every d, then the latent class is a D-dimensional binary vector and S equals 2^D. Specifically, the mapping process assumed is that the d^{th} dimension takes the value 1 if, and only if, every criterion function associated with that dimension satisfies a constraint, $H_{dkn} \geq 0, \quad \forall k = 1, \ldots, K_d$. For example, if the latent class represents the choice set considered, the d^{th} component takes the value 1 if alternative d is considered and zero otherwise. An alternative is considered, if and only if, the criterion functions associated with that alternative satisfy the constraints. It must be noted that even if deterministic constraints (such

as availability of an alternative) vary across individuals, such restrictions can be easily incorporated into the latent class representation by characterizing the latent class probability distribution only over possible choice sets to a particular individual.

3. If each dimension d is associated with a criterion function (i.e., $K_d = 1$), and the latent class is a D-dimensional vector with dimension d taking the value[13] $l_d \in \{1, \ldots, L_d\}$ where L_d is the number of levels in dimension d, then S equals $\prod_{d=1}^{D} L_d$. For example, in a transportation mode choice context, to capture unobserved taste variations for travel cost and travel time, a 2-dimensional latent class can be characterized along "cost-sensitivity" and "time-sensitivity" dimensions with ordered levels such as "low-sensitivity", "medium-sensitivity" and "high-sensitivity" along each dimension. The basic theme in the association of the criterion functions with the levels in each dimension, is the idea of "threshold crossing" wherein a particular level in a dimension is triggered if the corresponding criterion function falls between two thresholds[14].

Given the above associations between the criterion functions and the latent class, it must be emphasized that the *probabilistic* nature of the class membership model stems from the random components, δ_{dkn}, of the criterion functions. Denote $\tilde{Q}_s(\tilde{H}_{dkn}, \ \ \forall k = 1, \ldots, K_d, \forall d = 1, \ldots, D; \tau)$[15] as the probabilistic mapping from the systematic components of the latent criterion functions to the latent class s, such that

$$
\begin{aligned}
& \mathrm{P}(l^*_{sn} = 1 | Z_n), \\
= \quad & \mathrm{P}(T_n = [l_1^s, \ldots, l_D^s]' | Z_n) & (3.7) \\
= \quad & \mathrm{P}(T_{dn} = l_d^s, \ \forall d = 1, \ldots, D | Z_n) & (3.8) \\
= \quad & \mathrm{P}(T_{dn} = l_d^s, \ \forall d = 1, \ldots, D | \tilde{H}_{dkn}, \ \forall k = 1, \ldots, K_d, \forall d = 1, \ldots, D),
\end{aligned}
$$

[13]These integers are mere labels to identify the ordered levels in each dimension and do not serve any other purpose.

[14]For each dimension of the latent class, the assignment to a particular level is analogous to the ordinal probability model (see McKelvey and Zavoina [1975])

[15]For notational brevity we suppress parameters associated with the distribution of the random components δ. The origins of τ will be clearer in the discussion in section 3.3.2.

$$= \quad \tilde{Q}_s(\tilde{H}_{dkn},\ \forall k = 1, \ldots, K_d, \forall d = 1, \ldots, D; \tau) \quad \forall s = 1, \ldots, S. \tag{3.9}$$

It must be noted that the parameters τ which appear in the above equation in addition to the parameter vector θ cited in equation (3.5) are the thresholds referred to in the third type of class membership model (i.e., the ordinal criteria model).

Substituting $\tilde{H}_{dkn} = \tilde{H}(Z_n; \theta_{dk})$ in $\tilde{Q}_s(\cdot)$, the latent class probability can be denoted in terms of a function or model, $Q_s(Z_n; \theta, \tau)$, such that:

$$\mathrm{P}(l^*_{sn} = 1 | Z_n; \theta, \tau) = Q_s(Z_n; \theta, \tau) \quad \forall s = 1, \ldots, S. \tag{3.10}$$

The problem of specification of the class membership model reduces to the construction of $Q_s(Z_n; \theta, \tau)$. Essentially, $Q_s(Z_n; \theta, \tau)$, maps from the explanatory variables Z_n, to the latent class probabilities through the systematic components of the criterion functions, $\tilde{H}_{dkn}$. Before we proceed further, we assume for simplicity[16] *linear-in-parameters* functional form for the criterion functions and specify them as:

$$H_{dkn} = \theta'_{dk} Z_n + \delta_{dkn}, \qquad \forall k = 1, \ldots, K_d; \;\; \forall d = 1, \ldots, D. \tag{3.11}$$

It must be noted that the linear-in-parameters specification allows non-linear transformations of Z_n.

As noted earlier, the construction of the class membership model, $Q_s(\cdot)$, depends on the associations between the criterion functions and the latent class, which in turn depend on the specific problem context and characteristics of the latent class being modeled. Consequently, in the following paragraphs, we elaborate on the three types of class membership models.

[16]A linear-in-parameters specification eases the computational efforts required in the model estimation stage as the analytical derivatives/hessian of the likelihood function are relatively easy to obtain for this case.

3.3.1 Categorical Criterion Model

Here we assume that each latent class dimension d is associated with an criterion function H_{dn}, where

$$H_{dn} = \theta'_d Z_n + \delta_{dn}, \quad \forall d = 1, \ldots, D. \tag{3.12}$$

Further, assuming a maximum criterion association rule, the class membership $T_n = [l_1, \ldots, l_D]'$ is obtained as:

$$T_{dn} = \begin{cases} 1 & \text{if } \; H_{dn}(Z_n, \delta_{dn}; \theta_d) = \max\limits_{\forall d'=1,\ldots,D} \{H_{d'n}(Z_n, \delta_{d'n}; \theta_{d'})\} \\ 0 & \text{otherwise} \end{cases} \tag{3.13}$$

Note since the number of latent classes S equals D, we can replace l^*_{sn} with l^*_{dn}, and $l^*_{dn} = 1$ if $T_{dn} = 1$ and zero otherwise. By specifying a joint probability density function for $(\delta_{1n}, \ldots, \delta_{Dn})$, a class membership model can be constructed. For example, if the random variables, $\delta_{dn}, \;\; \forall d = 1, \ldots, D$, are independently and identically distributed Gumbel (0,1) random variables we obtain the MNL-type class membership model, i.e.,

$$\mathrm{P}(l^*_{dn} = 1 | Z_n; \theta) = \frac{\exp(\theta'_d Z_n)}{\sum\limits_{d'=1}^{D} \exp(\theta'_{d'} Z_n)}. \tag{3.14}$$

On the other hand, if the random vector $(\delta_{1n}, \ldots, \delta_{Dn})$ is multivariate normal, we obtain the MNP-type class membership model. It must be noted that in such an assignment only differences in the systematic components of criterion functions matter, and additional restrictions are necessary to have a unique set of parameters θ and the parameters in the distribution of $(\delta_{1n}, \ldots, \delta_{Dn})$. In the MNL-type and MNP-type membership models, we need to fix $\theta_d \equiv 0$ for some d. Further, in the MNP-type membership model identification restrictions must be placed in the covariance matrix of $(\delta_{1n}, \ldots, \delta_{Dn})$ (see Ben-Akiva and Bolduc [1991]).

3.3.2 Ordinal Criteria Model

Suppose the latent classes is characterized by a multi-dimensional vector with ordered levels along each dimension. Figure 3-1 illustrates such a latent concept which characterizes the sensitivity of an individual to travel time and travel cost in a travel choice situation, and correspondingly the two-dimensions of the latent class are: "time sensitivity" and "cost sensitivity". Three levels are postulated in the time sensitivity dimension and two levels in the cost sensitivity dimension. An individual falls in one of the six cells formed by the cartesian product of the levels in each dimension. The class membership model assigns the probability of an individual being in each of these six cells.

In general, let the latent class be represented across $d = 1, \dots, D$ with class membership vector $T_n = [l_1, \dots, l_d, \dots, l_D]'$. Let L_d represent the number of levels along dimension d with the levels taking on the integer values $1, \dots, L_d$. Let $\mathcal{L}_d$ denote the integer set $\{1, \dots, L_d\}$[17]. The modeling approach is to assume that each dimension is characterized by a criterion function H_d (i.e., $K_d = 1\ \forall d$).

Let the criterion functions H_{dn} for individual n be written as:

$$H_{dn} = \theta'_d Z_n + \delta_{dn}, \quad \forall d = 1, \dots, D. \tag{3.15}$$

We allow for the random components of H_d's, i.e., δ_d's, to be correlated to capture the unobserved interrelationships among the dimensions. For example, consider a case wherein the random components of dimension d and d', i.e., δ_{dn} and $\delta_{d'n}$ are independent. If an unobserved (hence assumed random) variable ν ought to have been included in the systematic components of criterion functions for dimensions d and d' with corresponding coefficients $\theta_{\nu d}$ and $\theta_{\nu d'}$, the new random components $\tilde{\delta}_{dn} = \delta_{dn} + \theta_{\nu d}\nu$ and $\tilde{\delta}_{d'n} = \delta_{d'n} + \theta_{\nu d'}\nu$ are correlated with $\mathrm{cov}(\tilde{\delta}_{dn}, \tilde{\delta}_{d'n}) = \theta_{\nu d}\theta_{\nu d'}\sigma^2_\nu$ where $\mathrm{var}(\nu) = \sigma^2_\nu$.

The basic theme in the association of the criterion functions and the latent class is the idea of "threshold crossing" wherein a particular level in a dimension is trig-

[17] Hence, $l_d \in \mathcal{L}_d, \quad \forall d = 1, \dots, D$.

gered if the associated criterion function falls between two thresholds. This theme is illustrated in Figure 3-2 wherein there are three levels in dimension d and

$$T_{dn} = \begin{cases} 1 & \Leftrightarrow & \{-\infty < H_{dn} \leq 0\} \\ 2 & \Leftrightarrow & \{0 < H_{dn} \leq \tau\} \\ 3 & \Leftrightarrow & \{\tau < H_{dn} < \infty\} \end{cases}$$

It must be noted that in the illustrative example, one of the thresholds is set to zero to fix the *location* of the criterion function H_{dn}, which is necessary if H_{dn} has an intercept. Further, two of the thresholds are set to $-\infty$ and ∞ which is typically the case if we assume infinite support for δ.

More generally, the levels in each dimension of the latent class are associated with the criterion functions as follows:

$$T_n = [l_1, \ldots, l_D]' \Leftrightarrow \left\{ \bigcap_{d=1}^{D} \left(\tau^d_{l_d - 1} \leq H_{dn} \leq \tau^d_{l_d} \right) \right\}, \quad \forall [l_1, \ldots, l_D] \in \{\mathcal{L}_1 \times \mathcal{L}_2 \times \cdots \mathcal{L}_D\} \tag{3.16}$$

where

$\tau^d_{l_d-1}$ - lower bound value (threshold) for criterion function for latent dimension d at level l_d

$\tau^d_{l_d}$ - upper bound value (threshold) for criterion function for latent dimension d at level l_d

Thus, the probability of the individual being in latent class $T_n = [l_1, \ldots, l_D]'$, is obtained by associating it with the probability of the event $\left\{ \bigcap_{d=1}^{D} (\tau^d_{l_d-1} \leq H_{dn} \leq \tau^d_{l_d}) \right\}$. It follows using equation (3.15) that,

$$\begin{aligned} \mathrm{P}(T_n = [l_1, \ldots, l_D]' | Z_n; \theta, \tau) &= \mathrm{P}\left\{ \bigcap_{d=1}^{D} (\tau^d_{l_d-1} \leq H_{dn} \leq \tau^d_{l_d}) \right\} \\ &= \mathrm{P}\left\{ \bigcap_{d=1}^{D} (\tau^d_{l_d-1} \leq \theta'_d Z_n + \delta_{dn} \leq \tau^d_{l_d}) \right\} \\ &= \mathrm{P}\left\{ \bigcap_{d=1}^{D} (\tau^d_{l_d-1} - \theta'_d Z_n \leq \delta_{dn} \leq \tau^d_{l_d} - \theta'_d Z_n) \right\} \end{aligned} \tag{3.17}$$

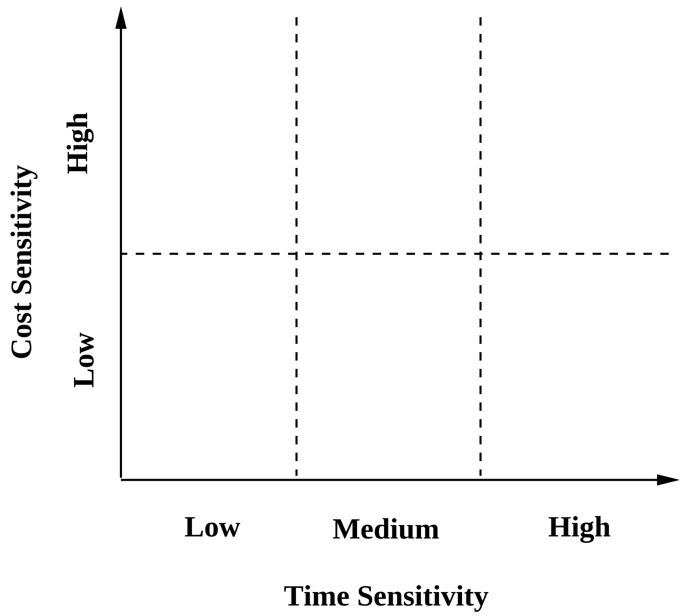


Figure 3-1: Latent Class with Ordered Levels: An example

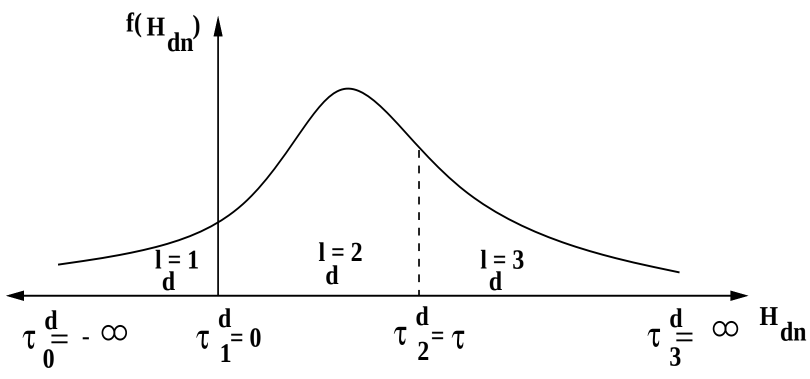


Figure 3-2: Illustration of Ordinal Criteria Model

By specifying different parametric distributions for $(\delta_{1n}, \ldots, \delta_{Dn})$, different class membership models can be constructed.

As an illustrative example, consider a latent construct represented along two dimensions, with L_1 and L_2 levels in dimensions 1 and 2, respectively. The latent class dimension levels are associated with the criterion functions as:

$$T_n = [l_1, l_2]' \Leftrightarrow \left\{ \left(\tau^1_{l_1 - 1} \leq H_{1n} \leq \tau^1_{l_1} \right) \quad \bigwedge \quad \left(\tau^2_{l_2 - 1} \leq H_{2n} \leq \tau^2_{l_2} \right) \right\}. \tag{3.18}$$

Then the probability of the individual n being in the latent class $T_n = [l_1, l_2]'$, $\mathrm{P}(T_n = [l_1, l_2]' | Z_n; \theta, \tau)$, is written as:

$$\mathrm{P} \left\{ (\tau^1_{l_1 - 1} - \theta'_1 Z_n \leq \delta_{1n} \leq \tau^1_{l_1} - \theta'_1 Z_n) \quad \bigwedge \quad (\tau^2_{l_2 - 1} - \theta'_2 Z_n \leq \delta_{2n} \leq \tau^2_{l_2} - \theta'_2 Z_n) \right\}. \tag{3.19}$$

Since the criterion functions are latent, we need to set the *scale* of the criterion function for each dimension. To this end, we set the variances of the random components to some constant, say 1. Assuming[18]

$$\begin{pmatrix} \delta_{1n} \\ \delta_{2n} \end{pmatrix} \sim \mathcal{BVN} \left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \tag{3.20}$$

the latent class probability is given by:

$$\mathrm{P}(T_n = [l_1, l_2]' | Z_n; \theta, \tau, \rho) = \int\limits_{\tau^1_{l_1 - 1} - \theta'_1 Z_n}^{\tau^1_{l_1} - \theta'_1 Z_n} \int\limits_{\tau^2_{l_2 - 1} - \theta'_2 Z_n}^{\tau^2_{l_2} - \theta'_2 Z_n} f_{\delta_1, \delta_2}(u, v) \, du \, dv \tag{3.21}$$

where $f_{\delta_1, \delta_2}(\cdot)$ is the density function of the standard bivariate normal random variate with parameters $[1, 1, \rho]$. In addition, τ^1_0 and τ^2_0 are set to $-\infty$, $\tau^1_{L_1}$ and $\tau^2_{L_2}$ to $+\infty$, and τ^1_1 and τ^2_1 are arbitrarily set to zero to fix the origin of the criterion functions.

The above correlation structure for the random components of the criterion func-

[18] $\mathcal{BVN}$=Bivariate Normal.

tions restricts the maximum number of allowable dimensions of the latent class to three to obtain tractable probabilities. Appendix A details a simulation procedure for calculating the latent class probabilities for latent classes with more criterion functions. It must be noted that if the random components are independent the class membership model reduces to the product of ordinal probit models corresponding to each dimension and hence is tractable even for large dimensions.

The ordinal criteria model is similar to the ordinal probability model of McKelvey and Zavoina [1975]. In similar vein, the class membership model can be viewed as an extension of the multivariate probit model of Ashford and Sowden [1970] which allowed for two levels in each dimension precluding the specification of unknown thresholds. The key distinguishing features are: (1) levels are unobserved in the class membership model while in the ordinal probability model and the multivariate probit model the levels are observed, and (2) class membership model generalizes the ordinal probability model to a multivariate ordinal level dependent vector, and extends the multivariate probit model to include more than two levels in each dimension. Hence, this class membership model may also be called, the *multivariate ordinal probability model*.

The specification of the criterion functions for each dimension requires special mention. The utilization of the relevant variables in the systematic component of each criterion function (i.e., specification of the fixed (zero) and the free (non-zero) θ parameters in each dimension) is based on *prior* behavioral hypotheses of the relevant latent constructs postulated to be captured by each dimension.

It is important to note a caveat of the model. By the nature of its specification, the systematic component of the criterion function is monotonic in the variables Z (monotonically increasing in Z if the associated coefficient is positive and monotonically decreasing otherwise). But the associated probabilities of an individual being in the levels of a dimension are monotonic in Z only for the first and the last level in that dimension, and not for intermediate levels. For example, consider the case of a one-dimensional latent construct with three levels. Assume that δ, the random component of the criterion function, is a standard normal random variable. Then

$\text{P}(l=1)=\Phi(-\tilde{H}_n)$, $\text{P}(l=2)=\Phi(\tau-\tilde{H}_n)-\Phi(-\tilde{H}_n)$ and $\text{P}(l=3)=1-\Phi(\tau-\tilde{H}_n)$ where $\Phi(\cdot)$ is the cumulative distribution function of δ. Then

$$\begin{aligned}
\frac{dP(l=1)}{d\tilde{H}_n} &= -\phi(-\tilde{H}_n) \\
\frac{dP(l=2)}{d\tilde{H}_n} &= -\phi(\tau-\tilde{H}_n)+\phi(-\tilde{H}_n) \\
\frac{dP(l=3)}{d\tilde{H}_n} &= \phi(\tau-\tilde{H}_n)
\end{aligned}$$

Since $\phi(\cdot)$ is the normal density function, $\text{P}(l=1)$ is monotonically decreasing in $\tilde{H}_n$, and $\text{P}(l=3)$ is monotonically increasing in $\tilde{H}_n$. On the other hand, $\text{P}(l=2)$ is monotonically increasing in $\tilde{H}_n$ for $\tilde{H}_n<\frac{\tau}{2}$, and monotonically decreasing in $\tilde{H}_n$ for $\tilde{H}_n>\frac{\tau}{2}$, with $\text{P}(l=2)$ attaining a maximum[19] at $\tilde{H}_n^*=\frac{\tau}{2}$. This argument extends to the multi-dimensional case with a symmetric joint density assumption for δ with infinite support.

3.3.3 Binary Criteria Model

In this case the latent classes are identified by a D-dimensional vector, with each dimension represented by a binary variable. A set of criterion functions $H_{dkn},\ \ \forall k=1,\ldots,K_d$ is associated with each dimension. The d^{th} component takes the value 1 if, and only if, $H_{dkn}\geq 0,\ \ \forall k=1,\ldots,K_d$.

Assume that the random components of criterion functions across dimensions of the latent class are independent (i.e., assume $\delta_d=(\delta_{d1},\ldots,\delta_{dK_d})$ and $\delta_{d'}=(\delta_{d'1},\ldots,\delta_{d'K_{d'}})$ are independent $\forall d\neq d'$). The class membership model for dimension d is written as:

$$\text{P}(T_{dn}=1|Z_n;\theta) \;=\; \text{P}\left(\bigcap_{k=1}^{K_d}(H_{dkn}\geq 0)\right) \quad \forall d=1,\ldots,D$$

[19] At the maximum, $\phi(\tau-\tilde{H}_n^*)=\phi(-\tilde{H}_n*)$. Noting the symmetry of $\phi(\cdot)$, we have $\tau-\tilde{H}_n^*=\tilde{H}_n^*$, it follows that $\tilde{H}_n^*=\frac{\tau}{2}$ (one can easily check the second order condition for maximum). Intuitively, this is obvious since the maximum is obtained when the density of the criterion function is centered on the interval $[0,\tau]$.

$$= \mathrm{P}\left(\bigcap_{k=1}^{K_d}(\theta'_{dk}Z_n + \delta_{dkn} \geq 0)\right) \quad \forall d = 1,\ldots,D$$
$$= \mathrm{P}\left(\bigcap_{k=1}^{K_d}(\delta_{dkn} \geq -\theta'_{dk}Z_n)\right) \quad \forall d = 1,\ldots,D \qquad (3.22)$$

and $\mathrm{P}(T_{dn} = 0|Z_n;\theta) = 1 - \mathrm{P}(T_{dn} = 1|Z_n;\theta)$. By specifying the joint probability density of δ_d, $f_{\delta_{d1},\ldots,\delta_{dK_d}}(u_1,\ldots,u_{K_d})$, the above probability can be obtained as:

$$\int_{v^+_{d1n}}^{\infty} \cdots \int_{v^+_{dK_dn}}^{\infty} f_{\delta_{d1},\ldots,\delta_{dK_d}}(u_1,\ldots,u_{K_d})\, du_1 \cdots du_{K_d} \qquad (3.23)$$

where $v^+_{dkn} = -\theta'_{dk}Z_n, \ \ \forall k = 1,\ldots,K_d$. We may allow for the random components of the criterion functions associated with each dimension d to be correlated.

The probability of the individual being in latent class $T_n = [l_1,\ldots,l_D]'$, equals

$$\begin{aligned}&\mathrm{P}\left(T_{dn} = l_d, \quad \forall d = 1,\ldots,D\right)\\ = \ &\mathrm{P}\left(\left\{\bigcap_{k=1}^{K_d}\left(\delta_{dkn} \geq v^+_{dkn}\right), \forall d|l_d = 1\right\} \bigwedge \right.\\ &\qquad \left.\left\{\left(\exists k \in \{1,\ldots,K_d\} : \left(\delta_{dkn} < v^+_{dkn}\right)\right), \forall d|l_d = 0\right\}\right)\end{aligned} \qquad (3.24)$$

By independence of the random components of criterion functions across dimensions of the latent class, the above equation reduces to:

$$\prod_{d=1}^{D}\left[\mathrm{P}(T_{dn} = 1|Z_n;\theta)\right]^{l_d}\left[1 - \mathrm{P}(T_{dn} = 1|Z_n;\theta)\right]^{1-l_d} \qquad (3.25)$$

It must be noted that in the foregoing model, the association between the criterion functions and the latent class is through a *conjunctive rule* wherein *all* the constraints associated with a particular dimension must be satisfied for that dimension to take the value 1. On the other hand, one may postulate a *disjunctive rule* wherein a particular dimension of the latent class vector takes the value 1 if *any* of the associated constraints is satisfied. In this case, the probability of the individual being in latent class $T_n = [l_1, \ldots, l_D]'$ equals

$$\begin{aligned}
& \mathrm{P}\left(T_{dn} = l_d, \quad \forall d = 1, \ldots, D\right) \\
= \; & \mathrm{P}\left(\left\{\bigcap_{k=1}^{K_d} \left(\delta_{dkn} < v_{dkn}^{+}\right), \forall d | l_d = 0\right\} \bigwedge \right. \\
& \left. \left\{\left(\exists k \in \{1, \ldots, K_d\} : \left(\delta_{dkn} \geq v_{dkn}^{+}\right)\right), \forall d | l_d = 1\right\}\right)
\end{aligned} \tag{3.26}$$

with the class membership model viewed as a "mirror image" of the class membership model for the conjunctive association rule.

3.4 Summary

In this chapter, we developed the latent class choice model (LCCM), and provided an overview of its adoption to capture specific forms of heterogeneity in choice processes. We also formulated different class membership models with an emphasis on their behavioral interpretations.

In chapters 4 and 5 we focus our attention on special cases of LCCM: (1) LCCM to capture taste variations through latent market segments, and (2) LCCM to capture decision protocol heterogeneity. The LCCM with the latent classes characterizing choice sets considered was reviewed in section 2.7.

Chapter 4

Latent Class Choice Model for Taste Heterogeneity: Case Study – Estimation of Distributed Value of Time

4.1 Introduction

In order to evaluate the benefits of new transportation system investment projects, one of the key input factors in the economic analysis is the value of time (VOT) concept. The principal aim in economic analysis is the comparison of the costs of the project with the potential benefits, including changes in travel time, accident rates, and operating costs, in terms of monetary units.

Individual traveler's travel time is evaluated in the marketplace by determining what price people will or do pay for travel time savings. Some of the factors that affect the value of travel time are (Winfrey [1969]):

A Individual characteristics

Age, occupation, wage earnings, whether paid during time of travel.

B Trip characteristics

Distance, purpose (business, pleasure, etc.), frequency.

C Environmental characteristics

Time of day, day of week, season of year, land use and economic conditions, rural or urban area, speed of travel, type and attributes of transportation system.

D Factors of value

Activity before and after trip, amount of available free time, productive time, utilization of travel time decrease, value of leisure time.

In the remainder of this section, we outline the operational approach to the estimation of VOT from travel choice models, and highlight certain aspects of empirical VOT studies in an effort to motivate our work. Some of the earliest studies to estimate the value of individual traveler's travel time are those by Claffey [1961], Lisco [1967], Haney [1967], and Thomas [1967]. More recent studies include: Cherlow [1981], Sharp [1988], Hague Consulting Group [1990], and Widlert [1994]. The reader is directed to Train and McFadden [1978], Bruzelius [1979], Hensher [1989], and Waters [1992] for a comprehensive treatment of the microeconomic underpinnings of the value of time concept and for the review of related empirical work.

In the value of travel time study presented here we restrict our attention to evaluation of travel times by individual travelers. In travel choice models, such as mode and route choice models, wherein travel time and travel cost are specified as attributes in the systematic utility functions, the corresponding estimated coefficients provide important information about the trade-off between travel time and travel cost. Consider a travel mode choice model with systematic utility function specified as a *linear in parameters* and *linear in attributes* function of only travel time $\boldsymbol{tt}$ and travel cost $\boldsymbol{tc}$. Specifically, the systematic utility function $V(\cdot)$ is written as:

$$V(\boldsymbol{tc}, \boldsymbol{tt}) = \beta_c \boldsymbol{tc} + \beta_t \boldsymbol{tt} \tag{4.1}$$

Setting the total differential of $V(\boldsymbol{tc}, \boldsymbol{tt})$ to zero to characterize the trade-offs between time and cost variables while maintaining the systematic utility constant, we have:

$$\begin{aligned} &\frac{\partial V}{\partial \boldsymbol{tc}}\mathrm{d}\boldsymbol{tc} + \frac{\partial V}{\partial \boldsymbol{tt}}\mathrm{d}\boldsymbol{tt} = 0 \\ \Rightarrow\ &\beta_c \mathrm{d}\boldsymbol{tc} + \beta_t \mathrm{d}\boldsymbol{tt} = 0 \\ \Rightarrow\ &-\frac{\mathrm{d}\boldsymbol{tc}}{\mathrm{d}\boldsymbol{tt}} = \frac{\beta_t}{\beta_c} \end{aligned} \tag{4.2}$$

Therefore the *implied value of time* is the ratio of the travel time coefficient to the travel cost coefficient, and is independent of the levels of attributes $\boldsymbol{tc}$ and $\boldsymbol{tt}$. Note that in a travel mode choice model, if the travel time and/or cost variables are mode-specific, then the estimated values of time differ across travel modes. To capture variations in the values of time the approaches which may be adopted include:

- *Variations along levels of time and cost variables*: Non-linearities in corresponding variables can be introduced through power series expansion of the systematic utility functions. Further, one can approximate the non-linearity by postulating a piece-wise linear function. Other transformations of variables such as the Box-Cox [1964] transformation may also be adopted.

- *Variations across individuals*: This is typically done by introducing additional interaction variables between individual characteristics and travel time and travel cost variables in the systematic utility functions.

Empirical studies to evaluate the value of time have relied primarily on two types of data: (1) *revealed preference* (RP) data such as travel mode, route, and location choices, and (2) *stated preference* (SP) data wherein travel time and travel cost trade-offs are elicited through individual's preferences towards hypothetical travel alternatives.

There is evidence in the transportation literature that the value of time varies along observed dimensions such as individual's income, age, trip purpose, etc. (see, for example, Bates and Roberts [1987], Bradley and Gunn [1990]). More recent work by Ben-Akiva *et al.* [1993] also incorporated the notion that the relative importance of time and cost changes may be influenced by individual-specific tastes and circumstances which cannot be observed by the analyst. Consequently, the value of time was allowed to be distributed randomly in the population. In a recent meta-analysis of empirical studies for estimating values of travel time, across countries and jurisdictions, Waters [1993] finds substantial variations in values of travel time with the estimates ranging between 30 to 50 percent of the wage rate. Further, the values of time varied substantially across regions within a country. Given the importance of time savings in transport project evaluation, Waters [1993] highlights the need for better methodological approaches to assess the value of travel time savings.

In the case study presented in this chapter, the emphasis is on stated preference analysis with the objective of obtaining monetary values of time that vary simultaneously along several household, personal, and situational dimensions, including the amount of free time available and income available. In contrast to RP data, the SP data collection effort was designed to distinguish between different types of time-money trading behavior: (a) for different journey purposes, (b) for different income groups, (c) for different occupation groups, (d) for different personal circumstances, and (e) for those with different amounts of leisure time.

The remainder of the chapter is organized as follows: In section 4.2 we present the model to capture unobserved heterogeneity in individual's travel time and travel cost coefficients. It must be noted that the model presented here is the latent class choice model (LCCM) for taste heterogeneity with specific reference to the classes characterized along cost sensitivity and time sensitivity dimensions. In section 4.3 we discuss the data used in the the case study to assess the potential of the model, while in section 4.4 we present estimation results. Choice models adopting existing methods to capture taste variations including the random coefficients model are estimated and compared to our modeling approach.

4.2 The Model

We assume that the data available to the analyst includes the choice indicator, attributes of travel alternatives, and socio-economic and demographic characteristics of the individual. The underlying choice process is hypothesized to vary across a finite set of groups of individuals in the population, and to be homogeneous within each such group. Since each homogeneous group of individuals is unobserved, the groups are characterized by *latent classes*.

We, must therefore, postulate the underlying mechanisms for class membership model and the choice model conditional on latent class to operationalize the model.

4.2.1 Class Membership Model

We assume that each latent class is characterized by two criteria – travel cost-sensitivity (C) and travel time-sensitivity (T) – with ordered levels along each dimension. The ordered levels are expected to reflect very high-sensitivity to very low-sensitivity or vice versa to travel time and travel cost of travel alternatives. Let the two-dimensional vector, $T_n = [l_C, l_T]'$, denote the class membership for individual n. Let L_C and L_T denote the number of levels along cost-sensitivity and time-sensitivity, respectively, with the levels taking on integer values[1] $1, \ldots, L_C$, and $1, \ldots, L_T$, respectively. Let $\mathcal{L}_C$ and $\mathcal{L}_T$ denote the integer sets $\{1, \ldots, L_C\}$ and $\{1, \ldots, L_T\}$ such that $l_C \in \mathcal{L}_C$, and $l_T \in \mathcal{L}_T$. Hence, the set of latent classes $\mathcal{M}$ is the cartesian product $\mathcal{L}_C \times \mathcal{L}_T$ and an individual belongs to *one and only one* of these latent classes. For notational convenience the latent classes are indexed $s = 1, \ldots, S$ where $S = |\mathcal{M}| = L_C L_T$, and a class membership indicator l^*_{sn} which takes the value 1 if the individual belongs to class s and zero otherwise, is defined on this index set. The index s is defined such that there is a *one-to-one* mapping between the latent class with index s and class membership, i.e. $(l^*_{sn} = 1) \Leftrightarrow T_n = [l^s_C, l^s_T]'$ for some $[l^s_C, l^s_T]' \in \mathcal{M}$. For example, if $\mathcal{L}_C = \{1, 2\}$ and $\mathcal{L}_T = \{1, 2\}$, then $\mathcal{M} = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$, $S = 4$, and the index $s = 3$ will indicate the latent class with membership $(2, 1)$.

The modeling approach is to assume that each of the cost sensitivity and time sensitivity dimensions are characterized by criterion functions, H_C and H_T, respectively. The criterion function for a particular dimension represents the individual's

[1] These integers are mere labels to identify the levels in each dimension and do not serve any other purpose.

sensitivity in that dimension. Let the criterion functions H_{Cn} and H_{Tn} for individual n be written as:

$$\begin{aligned} H_{Cn} &= \theta_C' Z_n + \delta_{Cn} \\ H_{Tn} &= \theta_T' Z_n + \delta_{Tn} \end{aligned} \tag{4.3}$$

where θ_C and θ_T are unknown parameter vectors and δ_{Cn} and δ_{Tn} are random components. We assume without loss of generality E(δ_C)=E(δ_T)=0. The specification of the criterion functions is aided by prior behavioral hypothesis as to the relevant individual characteristics affecting each dimension. For example, an individual's sensitivity to travel cost may be hypothesized to be affected by income, gender, etc. Similarly, individual characteristics such as time pressures, employment status (part-time vs. full time), household type, etc. may be postulated to affect the time sensitivity dimension. Further, correlation between the random components of H_C and H_T (i.e., δ_C and δ_T) may be allowed to capture the *unobserved* interrelationships between cost sensitivity and time sensitivity. By adopting a "threshold crossing" approach, the latent class levels in the cost and time dimension are associated with the underlying criterion functions as follows:

$$T_n = [l_C, l_T]' \;\Leftrightarrow\; \left[\left(\tau^C_{l_C-1} \leq H_{Cn} \leq \tau^C_{l_C}\right) \bigwedge \left(\tau^T_{l_T-1} \leq H_{Tn} \leq \tau^T_{l_T}\right)\right] \quad \forall [l_C, l_T]' \in \{\mathcal{L}_C \times \mathcal{L}_T\} \tag{4.4}$$

where
$\tau^C_{l_C-1}$ - lower bound value (threshold) for cost criterion function in level l_C,
$\tau^C_{l_C}$ - upper bound value (threshold) for cost criterion function in level l_C,
$\tau^T_{l_T-1}$ - lower bound value (threshold) for time criterion function in level l_T,
$\tau^T_{l_T}$ - upper bound value (threshold) for time criterion function in level l_T.

Assuming a parametric probability density function for $(\delta_{Cn}, \delta_{Tn})$, the probability of individual n belonging to latent class $[l_C, l_T]'$ (with index s such that $l_C = l^s_C$ and

$l_T = l_T^s$), denoted by $\mathrm{Q}_s(Z_n;\theta,\tau)$[2], equals

$$\begin{aligned}
& \mathrm{P}(T_n = [l_C, l_T]' | Z_n; \theta, \tau) \\
= \; & \mathrm{P}\left((\tau^C_{l_C-1} \leq H_{Cn} \leq \tau^C_{l_C}) \bigwedge (\tau^T_{l_T-1} \leq H_{Tn} \leq \tau^T_{l_T}) \right) \\
= \; & \mathrm{P}\left((\tau^C_{l_C-1} \leq \theta'_C Z_n + \delta_{Cn} \leq \tau^C_{l_C}) \bigwedge (\tau^T_{l_T-1} \leq \theta'_T Z_n + \delta_{Tn} \leq \tau^T_{l_T}) \right) \\
= \; & \mathrm{P}\Big((\tau^C_{l_C-1} - \theta'_C Z_n \leq \delta_{Cn} \leq \tau^C_{l_C} - \theta'_C Z_n) \\
& \qquad \bigwedge (\tau^T_{l_T-1} - \theta'_T Z_n \leq \delta_{Tn} \leq \tau^T_{l_T} - \theta'_T Z_n) \Big) \qquad (4.5)
\end{aligned}$$

Since the criterion functions are *latent*, it is necessary to set the scale of each criterion function, and this is operationalized by fixing the variance of each random component of the criterion function to some constant, say 1 for convenience. To see why this is necessary, $\mathrm{Q}_s(Z_n;\theta,\tau)$ equals

$$\begin{aligned}
& \mathrm{P}\Big((\tau^C_{l_C-1} - \theta'_C Z_n \leq \delta_{Cn} \leq \tau^C_{l_C} - \theta'_C Z_n) \\
& \qquad \bigwedge (\tau^T_{l_T-1} - \theta'_T Z_n \leq \delta_{Tn} \leq \tau^T_{l_T} - \theta'_T Z_n) \Big) \qquad (4.6) \\
= \; & \mathrm{P}\Big((\alpha_C \tau^C_{l_C-1} - \alpha_C \theta'_C Z_n \leq \alpha_C \delta_{Cn} \leq \alpha_C \tau^C_{l_C} - \alpha_C \theta'_C Z_n) \\
& \qquad \bigwedge (\alpha_T \tau^T_{l_T-1} - \alpha_T \theta'_T Z_n \leq \alpha_T \delta_{Tn} \leq \alpha_T \tau^T_{l_T} - \alpha_T \theta'_T Z_n) \Big) \\
= \; & \mathrm{P}\Big((\tilde{\tau}^C_{l_C-1} - \tilde{\theta}'_C Z_n \leq \tilde{\delta}_{Cn} \leq \tilde{\tau}^C_{l_C} - \tilde{\theta}'_C Z_n) \\
& \qquad \bigwedge (\tilde{\tau}^T_{l_T-1} - \tilde{\theta}'_T Z_n \leq \tilde{\delta}_{Tn} \leq \tilde{\tau}^T_{l_T} - \tilde{\theta}'_T Z_n) \Big) \qquad (4.7)
\end{aligned}$$

where α_C and α_T are arbitrary positive scalars, and any parameter $\tilde{\gamma}$ in equation (4.7) is a scaled form of the corresponding parameter γ in equation (4.5). Consequently, the latent class probability *does not* change by scaling θ and τ parameters. Hence, to identify the parameters it is necessary to fix the scale of *each* criterion function. When τ or θ parameters are shared across the cost-sensitivity and time-sensitivity dimensions, then it is necessary to fix the variance of *only* one of the random components.

[2]For notational brevity we suppress parameters associated with the distribution of the random components δ.

To illustrate the class membership model, assuming

$$\begin{pmatrix} \delta_{Cn} \\ \delta_{Tn} \end{pmatrix} \sim \mathcal{BVN}\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right), \tag{4.8}$$

$\mathrm{Q}_s(Z_n; \theta, \tau)$ reduces to

$$\int_{v_{Cn}^{+}}^{v_{Cn}^{-}} \int_{v_{Tn}^{+}}^{v_{Tn}^{-}} f_{\delta_C, \delta_T}(u_C, u_T)\, du_C\, du_T \tag{4.9}$$

where $f_{\delta_C,\delta_T}(\cdot)$ is the probability density function of the standard bivariate normal random variate with correlation parameter ρ, $v_{Cn}^{+} = \tau_{l_C-1}^{C} - \theta_C' Z_n$, $v_{Cn}^{-} = \tau_{l_C}^{C} - \theta_C' Z_n$, $v_{Tn}^{+} = \tau_{l_T-1}^{T} - \theta_T' Z_n$, and $v_{Tn}^{-} = \tau_{l_T}^{T} - \theta_T' Z_n$. In addition, τ_0^C and τ_0^T are set to $-\infty$, $\tau_{L_C}^C$ and $\tau_{L_T}^T$ to $+\infty$, and τ_1^C and τ_1^T are arbitrarily set to zero, to fix the origin of the criterion functions. By setting τ_1^C and τ_1^T to zero, we may allow for intercepts in the systematic components of the criterion functions.

4.2.2 Class-specific Choice Model

It is hypothesized that each latent class with index s has its own parameter vector β_s in the travel choice situation under consideration. The utility of travel alternative i for individual n depends on the vector of attributes of alternative i and the characteristics of the individual, X_{in}, and the latent class s to which the individual belongs. Using a linear functional form for the utility functions,

$$U_{isn} = \beta_s' X_{in} + \epsilon_{isn}, \quad \forall i \in C_n \tag{4.10}$$

where

U_{isn} = utility of travel alternative i for individual n in latent class s;

C_n = choice set available to individual n; and

ϵ_{isn} = random component of utility.

Assuming that the random component of utility is independently and identically distributed Gumbel (0,1) the class-specific choice model is an MNL model with the probability of individual n in latent class s choosing alternative i expressed as:

$$\mathrm{P}(y_{in} = 1 | l^*_{sn} = 1, X_n) = \frac{\exp(\beta'_s X_{in})}{\sum\limits_{j \in C_n} \exp(\beta'_s X_{jn})} \tag{4.11}$$

where

$$y_{in} = \begin{cases} 1 & \text{if travel alternative } i \text{ is chosen} \\ 0 & \text{otherwise} \end{cases}$$

At this time it is instructive to note the specification of β_s. Assume for notational simplicity that the cost and time variables are generic variables. The class-specific parameter vector is constructed by concatenating three elements: (1) a cost parameter, $\beta_{C,l_C}\ \forall l_C \in \mathcal{L}_C$, specific to latent classes with cost sensitivity dimension in level l_C; (2) a time parameter, $\beta_{T,l_T}\ \forall l_T \in \mathcal{L}_T$, specific to latent classes with time sensitivity dimension in level l_T; and (3) a parameter vector $\tilde{\beta}$ which is assumed be constant across classes and captures the effects of "other" attributes. Then

$$\beta_s = \begin{pmatrix} \beta_{C,l^s_C} \\ \beta_{T,l^s_T} \\ \tilde{\beta} \end{pmatrix} \tag{4.12}$$

and

$$X_{in} = \begin{pmatrix} tc_{in} \\ tt_{in} \\ \tilde{X}_{in} \end{pmatrix} \tag{4.13}$$

where tc_{in} and tt_{in} are the travel cost and travel time variables, and $\tilde{X}_{in}$ form the "other" attributes. For simplicity, we assumed the travel time and travel cost variables to be the corresponding attributes, and instead the analyst may include interaction variables between these attributes and socio-economic and demographic variables. Further, we may allow for the travel time and travel cost sensitivity levels to affect the coefficients of "other" attributes to capture interrelationships among individual's sensitivity to attributes, such as an individual having high (or low) sensitivity to two or more attributes. For example, if the "other" attributes can be separated into subsets such that $\tilde{X}_{in} = (\tilde{X}'_{1;in}, \tilde{X}'_{2;in})'$, and the analyst postulates that the individual's sensitivity to $\tilde{X}_1$ are *closely related* with cost-sensitivity[3], while *no* unobserved variations in sensitivity exists with respect to $\tilde{X}_2$, then the class-specific taste vector may be written as:

$$\beta_s = \begin{pmatrix} \beta_{C,l_C^s} \\ \beta_{T,l_T^s} \\ \tilde{\beta}_{1l_C^s} \\ \tilde{\beta}_2 \end{pmatrix} \tag{4.14}$$

and

$$X_{in} = \begin{pmatrix} tc_{in} \\ tt_{in} \\ \tilde{X}_{1;in} \\ \tilde{X}_{2;in} \end{pmatrix} \tag{4.15}$$

A more parsimonious approach to capture interrelationships between cost-sensitivity and sensitivity to X_1 is to scale the parameter vector $\tilde{\beta}_1$ such that $\tilde{\beta}_{1l_C^s} = \beta_{C,l_C^s}\tilde{\beta}_1$. Then

$$\beta_s = \begin{pmatrix} \beta_{C,l_C^s} \\ \beta_{T,l_T^s} \\ \beta_{C,l_C^s}\tilde{\beta}_1 \\ \tilde{\beta}_2 \end{pmatrix}$$

Assume that the class-specific parameter vector corresponds to equation (4.12). The implied VOT for an individual in class $[l_C, l_T]'$ equals $\beta_{T,l_T}/\beta_{C,l_C}$. If the class-specific VOT varies monotonically with the levels in a dimension[4], then monotonicity of the effects of socio-economic variables included in the criterion function associated with each dimension on the *expected* value of time can be easily established. For simplicity, consider a one-dimensional latent class with L ordered levels and characterized by a criterion function with systematic component $\tilde{H}$ and a standard normal

[3] "Closely related" refers to relationships such as an individual with high (low) sensitivity to travel cost may be expected to have high (low) sensitivity to attributes $\tilde{X}_1$, or vice versa.

[4] Monotonicity of class-specific VOT with levels is ensured if the *magnitudes* of the travel time *and* travel coefficient increase (or decrease) with the levels in each dimension.

random component δ. Let $\nu_l \; \forall l = 1, \ldots, L$ represent the value of time in level l with $\nu_1 < \nu_2 < \cdots < \nu_L$. Let $\mathrm{P}(l|\tilde{H})$ denote the latent class probability. Assuming the existence of threshold parameters $\tau_0 = 0 < \tau_1 < \cdots < \tau_{L-2}$, the latent class probabilities are written as:

$$\begin{aligned}
\mathrm{P}(1|\tilde{H}) &= \Phi(-\tilde{H}) \\
\mathrm{P}(l|\tilde{H}) &= \Phi(\tau_{l-1} - \tilde{H}) - \Phi(\tau_{l-2} - \tilde{H}), \quad \forall l = 2, \ldots, L-1 \\
\mathrm{P}(L|\tilde{H}) &= 1 - \Phi(\tau_{L-2} - \tilde{H})
\end{aligned} \tag{4.16}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable. Since $\mathrm{E}(\nu) = \sum_{l=1}^{L} \nu_l \mathrm{P}(l|\tilde{H})$,

$$\frac{d\mathrm{E}(\nu)}{d\tilde{H}} = \sum_{l=1}^{L} \nu_l \frac{d\mathrm{P}(l|\tilde{H})}{d\tilde{H}}. \tag{4.17}$$

Noting that

$$\begin{aligned}
\frac{d\mathrm{P}(1|\tilde{H})}{d\tilde{H}} &= \phi(-\tilde{H}) \\
\frac{d\mathrm{P}(l|\tilde{H})}{d\tilde{H}} &= -\phi(\tau_{l-1} - \tilde{H}) + \phi(\tau_{l-2} - \tilde{H}), \quad \forall l = 2, \ldots, L-1 \\
\frac{d\mathrm{P}(L|\tilde{H})}{d\tilde{H}} &= \phi(\tau_{L-2} - \tilde{H})
\end{aligned} \tag{4.18}$$

where $\phi(\cdot)$ is the standard normal density, we have

$$\frac{d\mathrm{E}(\nu)}{d\tilde{H}} = \phi(-\tilde{H})(\nu_2 - \nu_1) + \phi(\tau_1 - \tilde{H})(\nu_3 - \nu_2) + \cdots + \phi(\tau_{L-2} - \tilde{H})(\nu_L - \nu_{L-1}). \tag{4.19}$$

It follows that $\frac{d\mathrm{E}(\nu)}{d\tilde{H}} > 0$. It must also be noted that $\nu_1 \leq \mathrm{E}(\nu) \leq \nu_L$.

4.2.3 Unconditional Choice Model

Using equations (4.9), and (4.11), the unconditional probability of individual n choosing alternative i, $\mathrm{P}(y_{in} = 1 | X_n, Z_n)$, equals

$$\sum_{s=1}^{S} \left(\frac{\exp(\beta_s' X_{in})}{\sum\limits_{j \in C_n} \exp(\beta_s' X_{jn})} \right) \; \mathrm{Q}_s(Z_n; \theta, \tau) \tag{4.20}$$

4.2.4 Estimation

Two important issues as to the maximum likelihood estimation of any econometric model include:

1. *Existence of MLE*: This refers to whether the model parameters lie in the interior of the parameter space. Conceptually, non-existence would imply that the structural parameter estimates tend to infinity (or negative infinity), and correlation parameters tend to 1 or -1. The issue of existence is usually related to data configuration.

2. *Identification of model parameters*: This refers to whether two or more parameter vectors map into the same likelihood function. Conceptually, the model is not identified if the observed data can be generated by more than two parameterized data generating processes with identical model structure. The issue of identification is not a data configuration problem per se. Rather, it is related to the *unicity* of the underlying model structure.

If the models parameters exist and are identified, under certain regularity conditions, the parameters can be estimated by the maximum likelihood method to obtain consistent, asymptotically efficient and asymptotically normal estimates. The log-likelihood function for a random sample of N individuals is given by:

$$\mathcal{L}(\beta, \theta, \tau) = \sum_{n=1}^{N} \log \left\{ \prod_{i \in C_n} \left[\sum_{s=1}^{S} \left(\frac{\exp(\beta_s' X_{in})}{\sum\limits_{j \in C_n} \exp(\beta_s' X_{jn})} \right) \; \mathrm{Q}_s(Z_n; \theta, \tau) \right]^{y_{in}} \right\} \tag{4.21}$$

At this time, we have not investigated the conditions under which MLE exists. Assuming the existence of MLE, we do have partial results as to the unicity of a subset of the model parameters. Specifically, it can be shown that the log-likelihood function is concave in $[\beta, \theta, \tau]$ given that the correlation parameters associated with δ are *fixed*, if the class-specific choice model is an MNL model and δ is a multivariate normal random vector (see appendix E for the details).

The latent class probabilities may be obtained by numerical integration or through simulation methods. It must be noted that numerical integration to obtain the latent class probabilities might limit the maximum number of the criterion functions to 3 or 4. However, to address classes with larger number of criterion functions, recently developed estimation methods using simulators (see Hajivassiliou and McFadden [1992], Geweke *et al.* [1992], Börsch-Supan and Hajivassiliou [1993]) can be adopted. Appendix A details one such simulation procedure for approximating the latent class probabilities.

The maximization of the likelihood function can be conducted through usual gradient methods, such as Newton and quasi-newton procedures (see, for example, Luenberger [1984]). In appendix D we also outline the iterative Expectation-Maximization (EM) algorithm which may also be adopted for model estimation.

4.3 Survey Data

The data used in the case study to demonstrate the latent class choice model for taste heterogeneity was collected in The Netherlands (Hague Consulting Group [1990], Bradley and Gunn [1991]). The method of recruitment was to approach potential respondents at gas stations, parking facilities, and public transportation interchanges. The sites were selected to cover areas inside and outside the Randstad area, which includes Amsterdam, Rotterdam, and The Hague.

Travelers were asked to answer questions regarding the journey they were making at that time and whether they would be willing to participate in a mail survey. The follow-up SP questionnaire was retrospective, based as much as possible on the respondent's journeys and activities when they were intercepted.

The questionnaire contained four sections:

1. Questions about the journey they were making when intercepted, such as their frequency of making that type of journey, etc.

2. Pairwise choice questions offering different combinations of time and cost savings and losses against each other. The changes in travel time and cost were described and specified to be appropriate for the respondent's mode (car, train, bus, or streetcar) and journey distance (SP time savings or losses are limited to realistic ranges – up to 5, 10, 20, or 30 min – depending on actual journey duration). Each respondent provided 12 statements of preference regarding variations in travel times and costs for their journey. One of the 12, a "check" in which one alternative was both faster and cheaper than the other, was used to test respondents' understanding of the SP choice task.

3. Questions to gain insight into the amount of the respondent's free time and its flexibility, including the amount and rigidity of paid work hours, number of hours spent doing unpaid work, number of hours spent traveling, etc.

4. General questions about the respondents and their households, such as the income of the households and the number of workers, adults, children, and cars.

One of the unique aspects of the data is that it provides estimates of respondent's available free time from the responses to questions in section 3 of the questionnaire[5].

The data is separated into three main travel purpose groups: Commuting, Business and Other. The sample includes 485 respondents for Commuting, 469 for Business, and 1106 for Other purposes (mainly social, recreation, shopping, and educa-

[5]The individual's free time in hours per work was estimated as follows: Hours of paid work, hours of unpaid work (including work in the household) and hours spent on travel, were subtracted from the number of hours in a full week (168), further subtracted 8 hours per day for sleeping, and labelled the remaining as "free time". So a person with 35 hours/week paid work, 14 hours/week unpaid work and 3 hours per weekday travel time would have about: (7*(24*8)-35-14-(5*3)) or 48 hours/week "free time".

tion). The estimation data contains 11 SP choice observations per respondent. The variables are listed in Table 4.1[6].

4.4 Estimated Models

In this section we present travel choice models estimated on the *Commuting* dataset. Before we discuss the estimated models in detail, we provide an overview of the general theme in each of these models. The estimated models include:

1. Models which allow taste variations only along individual's sensitivity to travel cost. Specifically, two models are estimated which allow for two and three ordered cost-sensitivity levels.

2. Models wherein taste variations are allowed along individual's sensitivity to travel cost and travel time. Specifically, two models are estimated with two ordered levels in both cost-sensitivity and time-sensitivity dimensions. In the second model of this type, we allow for within-class heterogeneity in time-sensitivity by introducing interaction variables between travel time and individual characteristics in the class-specific systematic utility function.

For comparing our modeling approach with existing approaches to capture taste variations, we estimate the following models:

1. A standard choice model with fixed coefficients such as a logit model with interaction variables between cost and time attributes and individual characteristics being included in the systematic utility functions.

2. Models wherein the implied value of time (i.e., the ratio of the time coefficient to the cost coefficient) is randomly distributed. Specifically, we assume a lognormal distribution for the value of time and two variants of this type are presented.

[6] Note: Base income category is 2501-4000 Guilder/month; base age category is 21-35 years; and base free time category is > 50 hours/week.

NAME	DESCRIPTION
ID	Respondent Identification Number
CHOICE	Choice indicator = { 1 Faster alternative; 2 Cheaper alternative
tc_1, tc_2	Travel costs of alternative 1 and 2 in Dutch cents
tt_1, tt_2	Travel times of alternative 1 and 2 in minutes
INC15-	Household income dummy= { 1 0-1500 Guilder/month; 0 otherwise
INC1520	Household income dummy = { 1 1501-2000 Guilder/month; 0 otherwise
INC2025	Household income dummy = { 1 2001-2500 Guilder/month; 0 otherwise
INC4060	Household income dummy = { 1 4001-6000 Guilder/month; 0 otherwise
INC6080	Household income dummy= { 1 6001-8000 Guilder/month; 0 otherwise
INC80+	Household income dummy = { 1 >8000 Guilder/month; 0 otherwise
SOLO	Household type dummy = { 1 1 person/1 worker household; 0 otherwise
DINKS	Household type dummy = { 1 2 persons/2 workers household; 0 otherwise
KIDS	Household type dummy = { 1 1 or more children; 0 otherwise
PARTIME	Personal employment dummy = { 1 working part-time; 0 otherwise
AGE20-	Respondent's age dummy = { 1 20 or younger; 0 otherwise
AGE3650	Respondent's age dummy = { 1 36-50 years; 0 otherwise
AGE50	Respondent's age dummy = { 1 >50 years; 0 otherwise
FEMALE	Respondent's sex dummy = { 1 female; 0 male
FR-35	Respondent's free time dummy = { 1 less than 35 hours/week; 0 otherwise
FR-36-49	Respondent's free time dummy = { 1 36-49 hours/week; 0 otherwise

Table 4.1: Names and Definitions of Variables - Distributed Value of Time Study

3. Model wherein both the cost and time coefficients are randomly distributed. Specifically, we assume that these coefficients are distributed bivariate normal.

Since there are multiple responses per individual, the assumption of independence among these responses may not be entirely justified. Consequently, we also present models which attempt to capture interdependencies among multiple responses. Herein the models are based on the theme that some *unobserved individual-specific factors* may persist among the responses from the same individual. More specifically,

1. In latent class choice models, the unobserved latent class may be identical across responses, and consequently conditional on the class the multiple responses are assumed to be independent.

2. In latent class choice models, there may exist individual-specific error components in the criterion functions inducing interdependencies among responses. The specification of the criterion function is analogous to the agent-effects model in multiple regression.

3. In the lognormally distributed value of time model, the unobserved value of time may be identical across responses, thereby, inducing interdependencies among responses.

4. In the random coefficients model, the random components of cost and time coefficients may be identical across responses, thereby, inducing interdependencies among responses.

It must be noted that in none of the models do we adopt an agent-effects specification for the utility function, since the utility specification has no alternative specific constant[7].

[7] The alternatives are generic as they are described only in terms of the cost and time attributes.

4.4.1 Models Ignoring Interdependencies Among Responses

Model 0: Fixed Coefficients Model

This is the simplest of the estimated models. Noting that the individual and household characteristics are categorical dummy variables, the choice is modeled by a binary logit model expressed as a function of the utility difference between the two alternatives, using both "main" effect coefficients for travel cost and travel time, and a number of "additional" effect coefficients which apply only to certain observable segments of the sample. For respondent n for choice pair t, the systematic utility is specified as:

$$V_{tn} = (tc_{1tn} - tc_{2tn})(\alpha_0 + \sum_k \alpha_k \xi_{kn}) + (tt_{1tn} - tt_{2tn})(\beta_0 + \sum_l \beta_l \xi_{ln})$$

where

tc_{1tn}, tc_{2tn} = the travel costs of alternatives 1 and 2 for choice pair t;
tt_{1tn} , tt_{2tn} = the travel times of alternatives 1 and 2 for choice pair t;
α_0 , β_0 = the main effect cost and time coefficients, respectively;
α_k , β_l = additional effect cost and time coefficients, respectively; and
ξ_{kn} , ξ_{ln} = 0/1 variables indicating individual's membership in segments.

Membership in the k segments for additional cost-effects and in the l segments for additional time-effects is specified with regard to the respondent (e.g., age category), the household (e.g., income category), and all additional effects are estimated simultaneously. Thus each respondent may belong to a segment identified by the levels in each of the categories, keeping for each category one level as the "base".

The estimated model is presented in Table 4.2. The standard errors calculated from the estimated information matrix are incorrect in the presence of interdependencies among responses from the same individual. Consequently, to estimate the standard errors correctly, we utilize the variance-covariance matrix for extremum estimators (Amemiya [1985]), and we refer to the corrected t-statistics in conducting simple hypothesis tests. COST and TIME are cost difference and time difference variables[8], respectively. The income variables are used as additional cost-effect variables (ξ_{kn}'s), while all other variables are used as additional time-effect variables (ξ_{ln}'s). It must be noted that the coefficients of travel time and travel cost are *positive* since the cost and time difference variables are written as `attribute of alternative 1 - corresponding attribute of alternative 2`, while the estimated logit model expresses the probability of choosing alternative 2.

As noted in equation (4.2), the ratio of the travel time coefficient to the travel cost is the implied value of time. Consequently, with the coefficients that result from the model specification, a value of time for respondent n can be calculated as

$$VOT_n = \frac{(\beta_0 + \sum_l \beta_l \xi_{ln})}{(\alpha_0 + \sum_k \alpha_k \xi_{kn})}, \tag{4.22}$$

The signs and significance of the coefficients are interpreted regarding their effects on the estimated value of time. From theory we expect the value of time to increase with income levels. The coefficients of income dummies corresponding to incomes less than 2500 Guilders per month (fl/month) are positive while the coefficients of higher income dummies are negative. Further, there is a natural ordering of the estimated additional cost-effect coefficients. Noting that the income dummies enter into the denominator of equation (4.22) the value of time increases with income in tune with theory. The estimated coefficients for income dummies corresponding to incomes less than 2500 fl/month are insignificant, while the estimated coefficients for income dummies corresponding to incomes greater than 4000 fl/month are significant. Hence, the effect of income on value of time is fairly negligible for incomes less than 4000 fl/month, and it increases from thereon.

We expect individual time budgets to be influenced by individual and household characteristics. Household types are categorized into four categories: (1) households

[8]Since the alternatives do not have an identifier, such as car or transit, the marginal disutilities of travel time and travel cost are assumed to be the same for both alternatives. Consequently, the cost and time variables are *generic* in the specification.

with one or more children under working age (KIDS), (2) working adult living alone (SOLO), (3) working adult with one or more additional workers, but with no non-working adults or children (DINKS) and (4) "other" households. We interpret the coefficients corresponding to the first three categories with respect to the base "other" category. Households with children are expected to have tighter time budget constraints, and consequently higher values of time, and this is reflected in the positive and significant coefficient for variable KIDS. The coefficient for DINKS is also positive but insignificant, implying that the value of time is not significantly higher. This is due in part to possible sharing of household chores between the household members relaxing the time budget constraints of SOLO households. Consequently, the value of time is the highest for SOLO households as reflected in the magnitude and the sign of the coefficient of SOLO.

The occupation status of an individual is expected to determine his/her role in the household, and consequently influence time constraints. The coefficient corresponding to the PARTIME dummy is positive, thereby, reflecting the higher value of time of part-time workers compared to full-time workers. It can be argued that part-time workers face tighter time budget constraints, and the consequent higher value of time, or else they would have been working full-time.

The individual's age and gender are expected to affect the activities pursued and time constraints. Since the coefficient of the age dummy corresponding to age less than 21 years is positive though insignificant, it indicates that the value of time of an individual belonging to this category is higher than the base age category (21-35 years). Further, a significant decrease in the value of time is estimated for older people, especially for individuals 51 years or older. This may reflect the phenomenon that older people tend to have a less busy lifestyle, and consequently less stringent time constraints.

A female commuter has a lower value of time compared to a male commuter. Further, the available free time seems to affect the value of time, wherein individuals with lower available free time have higher values of time.

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.
Main Effect	COST	0.0064	0.0003	20.14	17.52
	TIME	0.0937	0.0067	14.05	12.91
Additional Cost Effects	INC15-	0.0016	0.0009	1.82	1.73
	INC1520	0.0008	0.0006	1.15	1.10
	INC2025	0.0007	0.0006	1.29	1.05
	INC4060	-0.0018	0.0003	-5.38	-4.74
	INC6080	-0.0019	0.0003	-5.54	-4.81
	INC80+	-0.0026	0.0004	-6.51	-5.89
Additional Time Effects	KIDS	0.0179	0.0064	2.80	2.52
	DINKS	0.0103	0.0059	1.74	1.57
	SOLO	0.0198	0.0073	2.73	2.55
	PARTIME	0.0212	0.0068	3.14	2.83
	AGE20-	0.0225	0.0114	1.98	1.80
	AGE3650	-0.0100	0.0053	-1.90	-1.77
	AGE50+	-0.0192	0.0065	-2.97	-2.61
	FEMALE	-0.0180	0.0050	-3.58	-3.33
	FR-35	0.0247	0.0086	2.89	2.59
	FR-36-49	0.0156	0.0047	3.30	2.86

Log-likelihood at zero = -3697.94
Log-likelihood at convergence = -3028.49
$\bar{\rho}^2 = 0.18$
Number of observations = 5335

Table 4.2: Model 0 – Fixed Coefficients Model

The sample average of the values of time is 12.5 fl/hr, while the sample standard deviation is 3.1 fl/hr.

Model 1: Two classes along cost-sensitivity dimension

In this model we allow unobserved taste variation to travel costs, and assume homogeneity in sensitivity to travel time. Consequently, two latent classes are postulated: "high cost-sensitive" and "low cost-sensitive" classes. The individual's sensitivity to travel cost is hypothesized to be affected by socio-economic and demographic characteristics. Hence, cost-sensitivity is captured by a criterion function which is specified as:

$$H_C = \theta_0 + \theta_1\text{INC15-} + \cdots + \theta_{16}\text{FR-36-49} + \delta_C$$

Assuming that δ_C is a standard normal random variable, and since we hypothesize two levels in the cost-sensitivity dimension, the latent class probability is given by a probit model. The choice model given the latent class s is represented by a logit model with only the cost and time difference variables specified in the systematic component of the utility function, i.e.,

$$V_{tsn} = (tc_{1tn} - tc_{2tn})\alpha_s + (tt_{1tn} - tt_{2tn})\beta, \quad s = 1, 2, \;\; t = 1, \ldots, 11$$

The estimated model is presented in Table 4.3. In the choice model, the cost coefficient specific[9] to class 1 is greater than the cost coefficient in class 2, and hence, class 1 and class 2 are interpreted as the high cost-sensitive and low cost-sensitive classes, respectively. Consequently, the value of time for an individual in class 2 is higher than that of an individual in class 1. Specifically, the class-specific VOT's are 5.7 fl/hr and 26 fl/hr for class 1 and class 2, respectively.

In the class membership model, a positive (negative) coefficient[10] for a variable would imply that the probability of being in class 2 is higher (lower) if the corresponding dummy variable takes the value 1, compared to the base case when the variable takes the value 0. Correspondingly, a positive (negative) coefficient for a variable implies that the *expected* value of time of an individual belonging to the corresponding category is higher (lower) than an individual in the base level. The estimated constant (LAT-CON) in the criterion function is negative implying that an individual in a segment wherein all the levels in the different categories are at their base levels, has a higher probability of being in class 1. The coefficients of income dummies are negative and with a decreasing trend for levels below the base level, and positive and increasing with levels above the base level[11]. This implies that the probability of

[9]For brevity we refer to the classes as class 1 and class 2, wherein class 1 and class 2 identify the latent classes with cost-sensitivity in level 1 and level 2, respectively.

[10]Note that the variables are 0/1 dummy variables.

[11]It must be noted that the coefficients of income dummies corresponding to incomes between

being in class 2 increases with income, and consequently, the expected value of time increases with increasing income. The effects of income dummies corresponding to income less than 2500 fl/month are not significant, while the effects are significant for incomes higher than 4000 fl/month. Hence, similar to Model 0 one can interpret that the effect of income on expected value of time is significant only for incomes above 4000 fl/month. The sample average of the latent class probability of being in class 1 (class 2) equals 0.42 (0.52).

The coefficients of household and employment dummies are positive implying that the expected values of time are higher for individuals in these segments compared to the base case. As age increases the expected value of time decreases. Similarly, the lower the available free time the higher is the expected value of time.

Compared to Model 0, we observe that the magnitudes of the cost and time coefficients of Model 1 are higher reflecting the better discriminatory power of Model 1. In terms of data fit, Model 1 does better than Model 0 by approximately 115 log-likelihood units, even though Model 1 has only two additional parameters. The qualitative effects of the socio-economic and demographic variables on the value of time as observed in Model 0 are retained in Model 1[12].

It must be noted that for models allowing for unobserved taste variations, the VOT for each individual is a random variable. Consequently, we can calculate individual-specific means and variances of VOT. We refer to the sample average of the individual-specific means as the "mean" VOT, the sample average of the individual-specific variances as the "mean individual" variance of VOT, and sample variance of the individual-specific means as the "variance of mean" VOT. The "total variance" of VOT refers to the sum of the variance of mean VOT and the mean individual variance of VOT.

For Model 1, the mean VOT is 17.4 fl/hr, the mean individual standard deviation

---

1501-2000 fl/month and 2001-2500 fl/month do not have follow the expected increasing trend, but they are insignificant.

[12] Although in Model 1, the value of time is class-specific, and the value of time we refer to is the expected value of time, i.e., class-specific value of time weighted by the corresponding latent class probability.

is 9.6 fl/hr, and the standard deviation of mean VOT is 3.1 fl/hr. The total standard deviation equals 10.1 fl/hr. The mean and total standard deviation of VOT are higher than those of Model 0.

Model 2: Three classes along cost-sensitivity dimension

In this model we postulate three classes: "high cost-sensitive", "medium cost-sensitive" and "low cost-sensitive" classes. The specifications of the cost-sensitivity criterion function and the utility function are as in Model 1. A class membership model with one threshold parameter (τ) captures the ordering of the levels as a function of socio-economic and demographic characteristics. Thus compared to Model 1, two additional parameters are estimated: a cost coefficient for the additional class in the choice model, and a threshold parameter in the class membership model.

The estimated model is presented in Table 4.4. The cost coefficients decrease along the classes[13], and consequently, the value of time increases from class 1 to class 3. The class-specific VOT's are 5.5 fl/hr, 22 fl/hr and 33 fl/hr for classes 1, 2, and 3 respectively.

The estimated class membership model is similar to that of Model 1. The monotonicity of the effects of socio-economic variables such as income on the expected value of time is maintained since the values of time increase monotonically along the classes. The sample averages of the latent class probabilities of being in classes 1, 2 and 3 are 0.41, 0.40 and 0.19, respectively.

This model does not improve on Model 1 if we consider the Akaike criterion since the increase is less than 2 units. The mean VOT is 17.9 fl/hr, the mean individual standard deviation is 11.1 fl/hr, and the standard deviation of mean VOT is 3.9 fl/hr. The total standard deviation equals 11.8 fl/hr.

[13]It must be noted that the coefficients of income dummies corresponding to incomes between 1501-2000 fl/month and 2001-2500 fl/month do not have follow the expected increasing trend, but they are insignificant.

Choice Model

Parameter	Estimates	Std. err.	t-stat	t-stat corr.
COST1	0.023	0.0016	14.43	13.19
COST2	0.005	0.0004	14.32	11.60
TIME	0.217	0.0122	17.73	12.89

Class Membership Model

Parameter	Estimates	Std. err.	t-stat	t-stat corr.
LAT-CON	-0.317	0.1124	-2.82	-2.73
INC15-	-0.222	0.2265	-0.98	-1.07
INC1520	-0.066	0.1632	-0.40	-0.39
INC2025	-0.158	0.1290	-1.22	-1.21
INC4060	0.407	0.0867	4.69	4.63
INC6080	0.440	0.0960	4.58	4.49
INC80+	0.781	0.1177	6.63	6.56
KIDS	0.305	0.0998	3.06	3.07
DINKS	0.177	0.0942	1.88	1.85
SOLO	0.269	0.1112	2.42	2.42
PARTIME	0.377	0.1059	3.56	3.43
AGE20-	0.545	0.1770	3.08	2.83
AGE3650	-0.109	0.0807	-1.35	-1.31
AGE50+	-0.240	0.1061	-2.26	-2.13
FEMALE	-0.319	0.0774	-4.12	-4.05
FR-35	0.379	0.1317	2.88	2.81
FR-36-49	0.401	0.0732	5.49	5.57

Log-likelihood of naive model = -3697.94
Log-likelihood at convergence = -2913.18
$\bar{\rho}^2 = 0.21$
Number of observations = 5335

Table 4.3: Model 1 – Two classes along cost-sensitivity dimension

Choice Model

Parameter	Estimates	Std. err.	t-stat	t-stat corr.
COST1	0.024	0.0017	13.92	12.77
COST2	0.006	0.0008	7.27	10.02
COST3	0.004	0.0015	2.39	4.60
TIME	0.220	0.0127	17.26	12.39

Class Membership Model

Parameter	Estimates	Std. err.	t-stat	t-stat corr.
LAT-CON	-0.268	0.1113	-2.41	-2.42
INC15-	-0.213	0.2214	-0.96	-1.05
INC1520	-0.064	0.1595	-0.40	-0.40
INC2025	-0.144	0.1255	-1.14	-1.13
INC4060	0.394	0.0832	4.74	4.71
INC6080	0.422	0.0919	4.59	4.52
INC80+	0.737	0.1116	6.61	6.55
KIDS	0.291	0.0952	3.06	3.07
DINKS	0.172	0.0900	1.91	1.88
SOLO	0.252	0.1064	2.37	2.37
PARTIME	0.347	0.1008	3.44	3.34
AGE20-	0.505	0.1636	3.09	2.89
AGE3650	-0.087	0.0769	-1.13	-1.10
AGE50+	-0.233	0.1014	-2.30	-2.20
FEMALE	-0.310	0.0738	-4.20	-4.18
FR-35	0.372	0.1233	3.01	2.93
FR-36-49	0.388	0.0696	5.58	5.69
τ	1.211	0.6910	1.75	4.87

Log-likelihood of naive model = -3697.94
Log-likelihood at convergence = -2911.73
$\bar{\rho}^2 = 0.21$
Number of observations = 5335

Table 4.4: Model 2 – Three classes along cost-sensitivity dimension

Model 3a: Two classes along cost-sensitivity and two classes along time-sensitivity dimensions

Variations in values of time could be "generated" by variations in sensitivity to both travel time and travel cost. Consequently, in Model 3a we postulate that the latent classes be identified across two dimensions: a cost-sensitivity dimension and a time-sensitivity dimension. Along each dimension we postulate the existence of two levels, leading to four classes in the population. Theory suggests that individual's income affects his/her sensitivity to travel cost, and time budget constraints arising from household and individual characteristics such as age, household type, gender, available free time, etc,. affect individual's time-sensitivity. To this end the criterion functions for the cost-sensitivity and time sensitivity dimensions are specified such that income variables are used in the criterion function for the cost-sensitivity dimension and the household type, age, gender and available free time variables are used in the time-sensitivity criterion function, i.e.,

$$
\begin{aligned}
H_C &= \theta_{C,0} + \theta_{C,1}\text{INC15-} + \cdots + \theta_{C,6}\text{INC80+} + \delta_C \\
H_T &= \theta_{T,0} + \theta_{T,1}\text{KIDS} + \cdots + \theta_{T,10}\text{FR-36-49} + \delta_T
\end{aligned}
$$

We assume that (δ_C, δ_T) is distributed as a standard bivariate normal with correlation parameter ρ.

For an individual in latent class $T_n = [l_C^s, l_T^s]'$ where $l_C^s \in \{1,2\}$ and $l_T^s \in \{1,2\}$, the utility function is specified as:

$$V_{tsn} = (tc_{1tn} - tc_{2tn})\alpha_{l_C^s} + (tt_{1tn} - tt_{2tn})\beta_{l_T^s}, \qquad t = 1, \ldots, 11$$

The estimated model is presented in Table 4.5. The cost coefficient corresponding to level 1 in the cost-sensitivity dimension is greater than the corresponding coefficient in level 1. This is also the case with the time coefficients. Consequently, along each dimension, level 1 is the high-sensitive level and level 2 is the low-sensitive level. Therefore, the value of time *increases* from level 1 to level 2 in the cost dimension, while the value of time *decreases* from level 1 to level 2 in the time dimension. Let indices 1, 2, 3 and 4 refer to the classes with membership $[1,1]'$, $[1,2]'$, $[2,1]'$, and $[2,2]'$ respectively. The class-specific VOT's are 6.9 fl/hr, 0.6 fl/hr, 32.9 fl/hr, and 2.7 fl/hr for classes 1, 2, 3 and 4 respectively.

In the cost dimension of the class membership model, we note that the constant (LAT-CON1) may also be interpreted as the coefficient corresponding to the base income category (2501-4000 fl/month). We observe an ordering of coefficients of the income dummies implying that as income increases the probability of the individual being in the latent class with the cost dimension in level 2 increases. Consequently, the expected value of time increases as income increases. We also note that the coefficients of income dummies corresponding to income less that 4000 fl/month are insignificant, leading to the same qualitative assessment made in Model 0 and Model 1.

In the time dimension, the constant (LAT-CON2) is negative and significant, implying that an individual in a segment with the dummy variables utilized in the time dimension in their base levels, has a higher probability of being in level 1 in the time dimension, and consequently higher expected value of time. The household dummies are negative and significant implying that individuals in these segments have higher expected values of time. The coefficients of age dummies reflect the lower expected value of time of older people[14]. The available free time dummies have the expected signs although one would have expected the coefficient of FR-35 to be less than FR-36-49 to reflect the higher time sensitivity of an individual with lower available free time. This may be in part due to the smaller number of observations in the sample belonging to the segment with available free time less than 35 hours/week.

Since the cost-sensitivity and time-sensitivity dimensions are expected to move in opposite directions, one would have expected the correlation between the random components of the criterion functions to be negative. But the estimated ρ is positive, and this may be due in part to the effects of omitted variables in the specification of criterion functions. The sample averages of the latent class probabilities of being in

[14]Unlike previous models wherein the coefficient corresponding to AGE50+ was significant, here the coefficient is insignificant although it reflects similar qualitative effect on value of time.

classes 1, 2, 3, and 4 are 0.4, 0.01, 0.47, and 0.12 respectively.

Compared to the fit of Model 1, fit of Model 3a is better suggesting the existence of two sensitivity dimensions – cost-sensitivity and time-sensitivity – with an increase of approximately 80 log-likelihood units for just three additional parameters (an additional time coefficient in choice model, constant for the time-sensitivity criterion function (LAT-CON2) and the correlation parameter ρ).

The mean VOT is 18.3 fl/hr, the mean individual standard deviation is 12.9 fl/hr, and the standard deviation of mean VOT is 3.1 fl/hr. The total standard deviation equals 13.3 fl/hr.

Model 3b: Two classes along cost-sensitivity and two classes along time-sensitivity dimensions with heterogeneity within each class

Even though Model 3a provides a much better fit to data, the primary drawback is the underlying assumption that there are four homogeneous groups of the population with different values of time in each group (even though by using the class membership model, one obtains the expected value of time to vary as a function of individual characteristics). To overcome this drawback, individual characteristics are included as "taste modifiers" in the choice model by interacting them with the time difference variable. This captures the systematic heterogeneity in time-sensitivity within each class.

As in Model 3a, we assume that income levels affect the individual's cost-sensitivity. Further, we hypothesize that only available free-time affects time-sensitivity, and hence the associated criterion function utilize only available free time variables, while dummy variables such as age, household type, gender, etc., are included as taste modifiers of the time coefficient in the class-specific choice model. Correspondingly, the criterion functions are specified as:

$$
\begin{aligned}
H_C &= \theta_{C,0} + \theta_{C,1}\text{INC15-} + \cdots + \theta_{C,6}\text{INC80+} + \delta_C \\
H_T &= \theta_{T,0} + \theta_{T,1}\text{FR-35} + \theta_{T,2}\text{FR-36-49} + \delta_T
\end{aligned}
$$

Choice Model

Parameter	Estimates	Std. err.	t-stat	t-stat corr.
COST1	0.086	0.0132	6.53	4.562
COST2	0.018	0.0025	7.46	5.489
TIME1	0.988	0.1546	6.39	4.429
TIME2	0.081	0.0243	3.33	4.934

Class Membership Model

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.
Cost Dimension	LAT-CON1	-0.015	0.0665	-0.22	-0.23
	INC15-	-0.132	0.2042	-0.65	-0.64
	INC1520	-0.100	0.1236	-0.81	-0.81
	INC2025	-0.063	0.1025	-0.62	-0.58
	INC4060	0.325	0.0694	4.69	4.63
	INC6080	0.391	0.0790	4.96	4.87
	INC80+	0.687	0.1049	6.54	6.04
Time Dimension	LAT-CON2	-0.837	0.1313	-6.37	-6.39
	KIDS	-0.469	0.1286	-3.65	-3.51
	DINKS	-0.299	0.1107	-2.70	-2.47
	SOLO	-0.272	0.1299	-2.09	-1.87
	PARTIME	-0.537	0.1573	-3.41	-3.40
	AGE20-	-0.581	0.2449	-2.37	-2.15
	AGE3650	0.227	0.1032	2.20	2.11
	AGE50+	0.222	0.1302	1.71	1.45
	FEMALE	0.340	0.0933	3.65	3.68
	FR-35	-0.261	0.1608	-1.62	-1.60
	FR-36-49	-0.648	0.1153	-5.61	-5.23
Noise Parameter	ρ	0.600	0.219	2.74	2.87

Log-likelihood of naive model = -3697.94
Log-likelihood at convergence = -2833.24
$\bar{\rho}^2 = 0.23$
Number of observations = 5335

Table 4.5: Model 3a – Two classes along cost-sensitivity and two classes along time-sensitivity dimensions

For an individual in latent class $T_n = [l_C^s, l_T^s]'$, where $l_C^s \in \{1,2\}$ and $l_T^s \in \{1,2\}$, the utility function is specified as:

$$V_{tsn} = (tc_{1tn} - tc_{2tn})\alpha_{l_C^s} + (tt_{1tn} - tt_{2tn})(\beta_{l_T^s} + \tilde{\beta}_1\text{KIDS} + \cdots + \tilde{\beta}_8\text{FEMALE}),$$
$$t = 1, \ldots, 11$$

where the parameters $\tilde{\beta}_1, \ldots, \tilde{\beta}_8$ represent taste modifiers.

The estimated model is presented in Table 4.6. In the choice model, the effects of age and gender on value of time are negligible as reflected in the insignificant coefficients of age and gender dummies. Further, the magnitudes of the cost and base time coefficients in the two levels are higher than those of Model 3a. This may be attributable to the inclusion of taste modifiers in the class-specific choice model which endeavor to explain part of the random component of the class-specific utility function. The class-specific VOT's are 6.3 fl/hr, 1.3 fl/hr, 92.7 fl/hr, and 19.5 fl/hr for classes 1, 2, 3 and 4 respectively.

In the class membership model, the coefficients of income dummies in the cost dimension and the coefficients of available free time in the time dimension have the expected signs[15]. The sample averages of the latent class probabilities of being in classes 1, 2, 3, and 4 are 0.38, 0.1, 0.12, and 0.4 respectively.

The overall fit of the model improved compared to Model 3a by approximately 46 log-likelihood units although both models have the same number of parameters.

The mean VOT is 22.8 fl/hr, the mean individual standard deviation is 27.1 fl/hr, and the standard deviation of mean VOT is 6.2 fl/hr. The total standard deviation equals 27.8 fl/hr. This model captures significantly higher variability in VOT.

Randomly distributed value of time

Herein we present estimation results for models with one random coefficient. The model is based on the assumption that the value of time is distributed randomly in

[15]Even here it must be noted that the coefficients of income dummies corresponding to incomes between 1501-2000 fl/month and 2001-2500 fl/month do not have follow the expected increasing trend, but they are insignificant.

Choice Model

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.
Base Parameters	COST1	0.413	0.0598	6.91	8.52
	COST2	0.028	0.0040	7.06	8.38
	TIME1	4.326	0.5955	7.26	9.04
	TIME2	0.910	0.1398	6.51	7.63
Taste Modifiers of Time Coefficient	KIDS	0.158	0.0526	3.01	3.31
	DINKS	0.175	0.0542	3.22	3.59
	SOLO	0.115	0.0565	2.03	2.13
	PARTIME	0.116	0.0500	2.33	2.39
	AGE20-	0.047	0.0644	0.73	0.78
	AGE3650	-0.012	0.0334	-0.37	-0.35
	AGE50+	0.014	0.0550	0.26	0.27
	FEMALE	-0.057	0.0407	-1.40	-1.26

Class Membership Model

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.
Cost Dimension	LAT-CON1	-0.136	0.0447	-3.05	-3.02
	INC15-	-0.283	0.1658	-1.71	-1.70
	INC1520	-0.072	0.1058	-0.68	-0.72
	INC2025	-0.128	0.0822	-1.56	-1.53
	INC4060	0.245	0.0533	4.60	4.62
	INC6080	0.304	0.0603	5.04	4.99
	INC80+	0.579	0.0697	8.30	8.35
Time Dimension	LAT-CON2	0.116	0.0585	1.98	2.05
	FR-35	-0.294	0.1482	-1.98	-1.76
	FR-36-49	-0.347	0.0887	-3.91	-3.77
Noise Parameter	ρ	0.814	0.0323	25.19	24.66

Log-likelihood of naive model = -3697.94
Log-likelihood at convergence = -2786.85
$\bar{\rho}^2 = 0.24$
Number of observations = 5335

Table 4.6: Model 3b – Two classes along cost-sensitivity and two classes along time-sensitivity dimensions (with taste modifiers)

the population *after* taking into account systematic population heterogeneity. The systematic utility function (or difference) is written as:

$$\begin{aligned} V_{tn} &= \alpha(tc_{1tn} - tc_{2tn}) + \tilde{\nu}_n(tt_{1tn} - tt_{2tn}) \\ &= \alpha\left((tc_{1tn} - tc_{2tn}) + \nu_n(tt_{1tn} - tt_{2tn})\right) \end{aligned}$$

where ν_n is the implied value of time for individual n. Suppose that ν_n is a function of characteristics of individual n:

$$\nu_n = \nu(1 + \tilde{\beta}' Z_n) \tag{4.23}$$

where ν is the base value of time which is assumed to be lognormally distributed, i.e.,

$$\ln \nu \sim \mathcal{N}(\omega, \sigma^2), \quad \nu > 0 \tag{4.24}$$

where

ω =E($\ln \nu$) = the expected value of the natural logarithm of the base value of time, and

σ^2 = the variance of the natural logarithm of the base value of time.

The probability density function of ν is given by:

$$f(\nu) = \frac{1}{\sigma\nu\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln \nu - \omega}{\sigma}\right)^2\right] \tag{4.25}$$

It can be seen[16] that ν_n is also lognormally distributed, i.e.,

$$\ln \nu_n \sim \mathcal{N}(\omega + \ln(1 + \tilde{\beta}' Z_n), \sigma^2), \quad \nu_n > 0 \tag{4.26}$$

This specification is based on the assumption that the value of time varies proportionately to some linear function of the characteristics of the individual (as in Ben-Akiva *et*

[16]The lognormal distribution for ν_n holds only if $(1 + \tilde{\beta}' Z_n) > 0$. But this is indeed the case with the estimated Model 4a.

al. 1993). Since in our case, Z_n are dummy variables, the specification reduces to the assumption of a lognormal distribution for value of time in each of the segments defined by the dummy variables. The estimated model, referred to as Model 4a, is presented in Table 4.7. The coefficients of income dummies increase with income levels suggesting that value of time increases as income increases[17]. Further, in contrast to the interpretation from the previous models, individuals with income less than 1500 fl/month have significantly lesser value of time compared to an individual in the base income category. The coefficients of other dummies have qualitatively similar effects on VOT as noted in earlier models. It must be noted that because of the assumption of lognormally distributed value of time, the expected value of time and its variance for an individual depend[18] on the parameters ω and σ^2.

The mean VOT is 20.3 fl/hr, the mean individual standard deviation is 21.0 fl/hr, and the standard deviation of mean VOT is 6.2 fl/hr. The total standard deviation equals 21.9 fl/hr.

A slight variant of the previous model is the assumption that the individual's value of time is distributed such that the mean of the logarithm of the value of time depends on characteristics of individual, i.e.,

$$\ln \nu_n \sim \mathcal{N}(\omega + \beta' Z_n, \sigma^2) \tag{4.30}$$

The estimated model, referred to as Model 4b, is presented in Table 4.8. The qualitative effects of all the individual and household characteristics on values of time remain the same as seen in previous models. The second specification provides a

[17]Even here it must be noted that the coefficients of income dummies corresponding to incomes between 1501-2000 fl/month and 2001-2500 fl/month do not have follow the expected increasing trend, but they are insignificant.

[18]If

$$\ln \nu \sim \mathcal{N}(\omega, \sigma^2), \quad \nu > 0 \tag{4.27}$$

then

$$\mathrm{E}(\nu) = \exp(\omega + \frac{\sigma^2}{2}) \tag{4.28}$$

and

$$\mathrm{Var}(\nu) = \exp(2\omega + \sigma^2)(\exp(\sigma^2) - 1) \tag{4.29}$$

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.
α	COST	0.035	0.0055	6.30	5.05
β parameters	INC15-	-0.249	0.1192	-2.09	-2.24
	INC1520	0.081	0.1119	0.73	0.78
	INC2025	-0.039	0.0874	-0.45	-0.43
	INC4060	0.378	0.0960	3.93	3.13
	INC6080	0.498	0.1316	3.78	2.72
	INC80+	1.226	0.2452	5.00	3.38
	KIDS	0.379	0.1181	3.21	2.41
	DINKS	0.247	0.1048	2.34	1.72
	SOLO	0.246	0.1147	2.15	1.52
	PARTIME	0.295	0.0897	3.29	2.75
	AGE20-	0.614	0.1892	3.25	2.81
	AGE3650	-0.093	0.0730	-1.28	-1.15
	AGE50+	-0.185	0.0747	-2.48	-2.02
	FEMALE	-0.186	0.0672	-2.77	-2.26
	FR-35	0.465	0.1318	3.53	2.92
	FR-36-49	0.523	0.1119	4.67	3.30
Lognormal dist. parameters	ω	2.479	0.0945	26.25	19.65
	σ	0.992	0.0286	34.64	31.20

Log-likelihood of naive model = -3697.94
Log-likelihood at convergence = -2894.41
$\bar{\rho}^2 = 0.22$
Number of observations = 5335

Table 4.7: Model 4a – Randomly distributed value of time model

slightly better fit to the data.

The mean VOT is 20.5 fl/hr, the mean individual standard deviation is 21.6 fl/hr, and the standard deviation of mean VOT is 6.8 fl/hr. The total standard deviation equals 22.6 fl/hr.

4.4.2 Models Allowing Interdependencies Among Responses

As noted earlier, since we have multiple responses per individual, the assumption of independence among these multiple responses may not be entirely justified. To

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.
α	COST	0.040	0.0075	5.37	3.03
β parameters	INC15-	-0.211	0.1385	-1.52	-1.40
	INC1520	0.033	0.0768	0.43	0.49
	INC2025	-0.041	0.0670	-0.62	-0.54
	INC4060	0.281	0.0501	5.61	4.09
	INC6080	0.268	0.0643	4.17	2.63
	INC80+	0.580	0.0874	6.63	3.02
	KIDS	0.230	0.0665	3.46	2.28
	DINKS	0.120	0.0607	1.97	1.25
	SOLO	0.081	0.0983	0.82	0.36
	PARTIME	0.227	0.0512	4.45	4.19
	AGE20-	0.353	0.0996	3.55	2.21
	AGE3650	-0.110	0.0562	-1.97	-1.17
	AGE50+	-0.222	0.0643	-3.45	-2.63
	FEMALE	-0.125	0.0499	-2.50	-1.44
	FR-35	0.335	0.0542	6.19	5.42
	FR-36-49	0.412	0.0661	6.23	2.63
lognormal parameters	ω	2.574	0.0708	36.33	24.07
	σ	0.999	0.0281	35.51	25.98

Log-likelihood of naive model = -3697.94
Log-likelihood at convergence = -2884.42
$\bar{\rho}^2 = 0.22$
Number of observations = 5335

Table 4.8: Model 4b – Randomly distributed value of time model

address this issue, we estimate latent class choice models which attempt to take into account these interdependencies. The lognormally distributed value of time model is re-estimated under the assumption that the unobserved value of time for the individual is identical across responses for the same individual. Further, we estimate a random coefficients model with both cost and time coefficients randomly distributed in the population.

Model 5: Latent class choice model with interdependencies among responses: Identical latent class

In this model we assume that each individual is assumed to be in one latent class, and consequently is expected to adopt the *same* class-specific choice process while making choices in the SP experiments. Therefore, *conditional* on the latent class, we assume that the responses for the same individual are independent. Then, the probability of observing the response vector $Y_n = [Y_{1n}, \dots, Y_{tn}, \dots, Y_{11n}]$, denoted by $P(Y_n|X_n, Z_n; \beta, \theta)$, can be written as

$$\sum_{s=1}^{S} \left\{ \prod_{t=1}^{11} P(Y_{tn}|X_{tn}; \beta_s) \right\} Q_s(Z_n; \theta) \tag{4.31}$$

where Y_{tn} and X_{tn} denote the choice indicator and the attributes of alternatives for the t^{th} choice pair, respectively.

The estimated model is presented in Table 4.9. The specification of the criterion and the utility functions are the same as in Model 3b. In the choice model, the base taste parameters are scaled down compared to those estimated in Model 3b, but their significance increased. On the other hand, all the taste modifiers are insignificant except for the household type dummy variable DINKS, employment status and age dummy corresponding to 51 years and older. A possible explanation for this phenomenon is that by "fixing" the class to be identical across responses, the discriminatory power of the class membership model appears to have increased since the corresponding coefficients seem to be scaled up. The class-specific VOT's are 27.9 fl/hr, 4.6 fl/hr, 59.3 fl/hr, and 9.7 fl/hr for classes 1, 2, 3 and 4 respectively.

In the class membership model, in the cost dimension the income dummies have the expected signs and trends in their magnitudes. In the time dimension, the constant term is positive and large compared to Model 3b. The sample averages of the latent class probabilities of being in classes 1, 2, 3, and 4 are 0.12, 0.23, 0.10, and 0.55 respectively.

The fit of the model increased by 253 log-likelihood units compared to that of Model 3b. This is a significant improvement given that the model structure, and the number of parameters are the same as in Model 3b. But in Model 5, we have incorporated the additional information that the latent class persists over responses from the same individual.

The mean VOT is 15.9 fl/hr, the mean individual standard deviation is 15.6 fl/hr, and the standard deviation of mean VOT is 3.4 fl/hr. The total standard deviation equals 16.0 fl/hr. These statistics are significantly lower than the corresponding statistics in Model 3b.

Model 6: Latent class choice model with interdependencies among responses: Agent-effects specification

The basic idea adopted is the agent-effects (also referred to as random-effects) model omnipresent in multiple regression models for panel data. We consider Model 3b and allow for interdependencies among responses. The error components in the criterion functions are split into two components: δ and $\tilde{\delta}$. The first component is a pure random component and is independent of the second component. The second random component is assumed to be an individual-specific random effect, and hence persists across all responses from the same individual, and is assumed to have a parametric distribution. Therefore, the criterion functions for individual n across $t = 1, \ldots, 11$ responses are specified as:

$$\begin{aligned} H_{Ctn} &= \theta'_C Z_n + \tilde{\delta}_{Cn} + \delta_{Ctn} \\ H_{Ttn} &= \theta'_T Z_n + \tilde{\delta}_{Tn} + \delta_{Ttn} \end{aligned} \tag{4.32}$$

Choice Model

	Parameter	Estimates	Std. err.	t-stat
Base Parameters	COST1	0.017	0.0012	14.70
	COST2	0.008	0.0004	21.38
	TIME1	0.791	0.0656	12.07
	TIME2	0.129	0.0093	13.86
Taste Modifiers of Time Coefficient	KIDS	0.022	0.0100	2.25
	DINKS	0.016	0.0092	1.77
	SOLO	0.006	0.0115	0.49
	PARTIME	0.028	0.0120	2.36
	AGE20-	0.015	0.0154	1.01
	AGE3650	-0.008	0.0086	-0.91
	AGE50+	-0.030	0.0095	-3.18
	FEMALE	-0.019	0.0089	-2.08

Class Membership Model

	Parameter	Estimates	Std. err.	t-stat
Cost Dimension	LAT-CON1	0.103	0.1661	0.62
	INC15-	-0.689	0.6386	-1.08
	INC1520	-0.256	0.4071	-0.63
	INC2025	-0.180	0.3112	-0.58
	INC4060	0.105	0.2103	0.47
	INC6080	0.460	0.2217	2.07
	INC80+	0.706	0.2758	2.56
Time Dimension	LAT-CON2	0.974	0.0881	11.06
	FR-35	-0.386	0.2495	-1.55
	FR-36-49	-0.544	0.1422	-3.83
Noise Parameter	ρ	0.418	0.1161	3.60

Log-likelihood of naive model = -3697.94
Log-likelihood at convergence = -2533.87
$\bar{\rho}^2 = 0.31$
Number of observations = 485

Table 4.9: Model 5 – Two classes along cost-sensitivity and two classes along time-sensitivity dimensions (with taste modifiers) and Identical latent class

Hence we allow for the criterion functions across responses for the same individual to be positively correlated. Assuming[19]

$$\begin{pmatrix} \delta_{Ctn} \\ \delta_{Ttn} \end{pmatrix} \sim \mathcal{BVN}\left(0, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \tag{4.33}$$

we have conditional on Z_n and $(\tilde{\delta}_{Cn}, \tilde{\delta}_{Tn})$ the class membership model, denoted by $Q_s(Z_n, \tilde{\delta}_{Cn}, \tilde{\delta}_{Tn}; \theta)$, is given by a bivariate probit[20] model. Then, by assuming a parametric distribution for $(\tilde{\delta}_{Cn}, \tilde{\delta}_{Tn})$, such as:

$$\begin{pmatrix} \tilde{\delta}_{Cn} \\ \tilde{\delta}_{Tn} \end{pmatrix} \sim \mathcal{BVN}\left(0, \begin{bmatrix} \sigma_C^2 & \rho\sigma_C\sigma_T \\ \rho\sigma_C\sigma_T & \sigma_T^2 \end{bmatrix}\right) \tag{4.34}$$

the probability of observing the response vector Y_n, $P(Y_n|X_n, Z_n; \beta, \theta, \sigma_C, \sigma_T, \rho)$, is written as

$$\iint \prod_{t=1}^{11} \left\{ \sum_{s=1}^{S} P(Y_{tn}|X_{tn}; \beta_s)\, Q_s(Z_n, \tilde{\delta}_C, \tilde{\delta}_T; \theta) \right\} f(\tilde{\delta}_C, \tilde{\delta}_T; \sigma_C, \sigma_T, \rho)\, d\tilde{\delta}_C\, d\tilde{\delta}_T \tag{4.35}$$

where $f(\cdot)$ is the bivariate normal density of $(\tilde{\delta}_{Cn}, \tilde{\delta}_{Tn})$. It must be noted that the agent-effects specification for the criterion functions is equivalent to the assumption of the constants associated with each criterion function to be randomly distributed in the population.

The estimated model is presented in Table 4.10. In the choice model, the base taste parameters are scaled up compared to the base parameters in Model 5, and are similar to those in Model 3b. All the taste modifiers, except for the household type dummy variable DINKS, are insignificant. Thus by allowing variability in the respondent's latent class across responses part of the random component of the class-specific utility function appears to be captured. The class-specific VOT's are 7 fl/hr,

[19]We attempted the estimation of a model wherein we allowed correlation between δ_{Ctn} and δ_{Ttn}. But the estimation procedure did not converge as the correlation parameter between δ_{Ctn} and δ_{Ttn}, and the correlation parameter between δ_{Cn} and δ_{Tn} tended towards -1 and 1.

[20]In fact since we do not allow for correlations between δ_{Ctn} and δ_{Ttn}, the class membership model is a product of probit probabilities.

1.5 fl/hr, 79.8 fl/hr, and 16.9 fl/hr for classes 1, 2, 3 and 4 respectively.

In the class membership model, in the cost dimension the income dummies have the expected signs and the coefficients are similar to those of Model 5, except for the constant which changed signs. In the time dimension, the coefficients of free time are no longer significant and the constant changed signs. The sample averages of the latent class probabilities of being in classes 1, 2, 3, and 4 are 0.35, 0.15, 0.28, and 0.22 respectively.

The fit of the model increased by 78 log-likelihood units compared to Model 5, and by 330 units compared to Model 3b[21]. This is a significant improvement given that only two additional parameters are estimated.

The mean VOT is 31.5 fl/hr, the mean individual standard deviation is 32.2 fl/hr, and the standard deviation of mean VOT is 6.4 fl/hr. The total standard deviation equals 33.3 fl/hr. These statistics are significantly greater than corresponding statistics in Model 3b.

Model 7: Randomly distributed value of time model with interdependencies among responses

In the lognormal VOT models presented earlier, we ignored interdependencies among responses from the same individual. The realization of the random value of time for an individual may persist across responses from the same individual. Thus, we assume that conditional on the value of time ν, the responses of the individual are independent, i.e.,

$$P(Y_n|X_n,\nu;\beta) = \prod_{t=1}^{11} P(Y_{tn}|X_{tn},\nu;\beta) \tag{4.36}$$

Hence the probability of observing the response vector Y_n, $P(Y_n|X_n;\beta,\omega,\sigma)$, is written as

$$P(Y_n|X_n;\beta,\omega,\sigma) = \int \prod_{t=1}^{11} P(Y_{tn}|X_{tn},\nu;\beta)\ f(\nu;\omega,\sigma)\ d\nu \tag{4.37}$$

[21]It must be noted that the log-likelihoods of Model 3b and Model 6 are comparable only to the extent that by restricting the variances of the individual-specific error components of Model 6 to zero (i.e., assuming that these terms do not exist) we obtain Model 3b.

Choice Model

	Parameter	Estimates	Std. err.	t-stat
Base Parameters	COST1	0.374	0.0560	6.69
	COST2	0.033	0.0049	6.78
	TIME1	4.390	0.7270	6.04
	TIME2	0.930	0.1439	6.47
Taste Modifiers of Time Coefficient	KIDS	0.033	0.0348	0.94
	DINKS	0.075	0.0366	2.05
	SOLO	0.022	0.0350	0.64
	PARTIME	0.039	0.0426	0.91
	AGE20-	-0.032	0.0773	-0.40
	AGE3650	0.011	0.0275	0.40
	AGE50+	0.018	0.0367	0.48
	FEMALE	-0.004	0.0266	-0.13

Class Membership Model

	Parameter	Estimates	Std. err.	t-stat
Cost Dimension	LAT-CON1	-0.257	0.1349	-1.91
	INC15-	-0.369	0.5801	-0.64
	INC1520	-0.068	0.3471	-0.20
	INC2025	-0.124	0.2766	-0.45
	INC4060	0.435	0.1787	2.43
	INC6080	0.532	0.2019	2.63
	INC80+	0.983	0.2282	4.31
Time Dimension	LAT-CON2	-0.443	0.1685	-2.63
	FR-35	-0.430	0.4786	-0.90
	FR-36-49	-0.474	0.2940	-1.61
Noise Parameters	σ_C	1.287	0.0754	17.05
	σ_T	1.573	0.2935	5.36
	ρ	0.187	0.0745	2.51

Log-likelihood of naive model = -3697.94
Log-likelihood at convergence = -2455.94
$\bar{\rho}^2 = 0.33$
Number of observations = 485

Table 4.10: Model 6 – Two classes along cost-sensitivity and two classes along time-sensitivity dimensions (with taste modifiers) and agent-effects specification for criterion functions

where $f(\cdot)$ is the lognormal density with parameters ω and σ.

We consider the specification as in Model 4a and allow for interdependencies among responses. The estimated model is presented in Table 4.11. In terms of model fit, Model 7 betters Model 4a by 306 log-likelihood units[22]. It must be noted both models have the same number of parameters. Compared to Model 4a, the cost coefficient is estimated with more precision and is scaled down. Most of the other coefficients are also scaled down. A possible explanation is that in Model 4a the unobserved (random) value of times which were independent across responses potentially captured part of the randomness in the utility function. Thus by restricting the random value of time to be same across responses, the parameters are scaled down.

The mean VOT is 15.9 fl/hr, the mean individual standard deviation is 14.7 fl/hr, and the standard deviation of mean VOT is 4.5 fl/hr. The total standard deviation equals 15.4 fl/hr. These statistics are significantly lower than the corresponding statistics in Model 4a, while these statistics are closer to those of Model 5.

Model 8: Random coefficients model

In this model we allow the coefficient of travel cost and travel time to be distributed randomly in the population. Except for the randomness in the coefficients, the systematic utility function specification is similar to the Fixed Coefficients Model (Model 0), and is written as:

$$V_{tn} = (tc_{1tn} - tc_{2tn})(\alpha_0 + \sum_k \alpha_k \xi_{nk} + \nu_{Cn}) + (tt_{1tn} - tt_{2tn})(\beta_0 + \sum_l \beta_l \xi_{nl} + \nu_{Tn})$$

where ν_{Cn} and ν_{Tn} are the random components of the travel cost and travel time coefficients. Assuming

$$\begin{pmatrix} \nu_{Cn} \\ \nu_{Tn} \end{pmatrix} \sim \mathcal{BVN}\left(0, \begin{bmatrix} \sigma_C^2 & \rho\sigma_C\sigma_T \\ \rho\sigma_C\sigma_T & \sigma_T^2 \end{bmatrix}\right) \tag{4.38}$$

[22]We note with caution that these likelihoods are not strictly comparable, although in Model 7 by utilizing additional information that the individual's unobserved value of time persists among responses, the overall fit seems to suggest improvement.

Choice Model

	Parameter	Estimates	Std. err.	t-stat
α	COST	0.010	0.0004	26.36
	INC15-	-0.586	0.1609	-3.64
	INC1520	-0.324	0.1298	-2.50
	INC2025	-0.234	0.1526	-1.54
	INC4060	0.199	0.1071	1.86
	INC6080	0.431	0.1401	3.07
	INC80+	0.528	0.1612	3.28
	KIDS	0.315	0.1248	2.53
β	DINKS	0.276	0.1225	2.26
	SOLO	0.357	0.1637	2.18
parameters	PARTIME	0.385	0.1060	3.64
	AGE20-	0.364	0.2230	1.63
	AGE3650	-0.331	0.0891	-3.71
	AGE50+	-0.220	0.1052	-2.09
	FEMALE	-0.311	0.1086	-2.86
	FR-35	0.564	0.2442	2.31
	FR-36-49	0.370	0.1192	3.10
Lognormal dist.	ω	2.552	0.1101	23.18
parameters	σ	0.881	0.0408	21.56

Log-likelihood of naive model = -3697.94
Log-likelihood at convergence = -2587.58
$\bar{\rho}^2 = 0.29$
Number of observations = 485

Table 4.11: Model 7 – Randomly distributed value of time model with interdependencies among responses

$P(Y_n|X_n;\alpha,\beta,\sigma_C,\sigma_T,\rho)$ is given by:

$$\iint \prod_{t=1}^{11} P(Y_{tn}|X_{tn},\nu_C,\nu_T;\alpha,\beta)\ f(\nu_C,\nu_T;\sigma_C,\sigma_T,\rho)\,d\nu_C\,d\nu_T \tag{4.39}$$

The estimated model is presented in Table 4.12. The estimated standard deviations of cost and time coefficients are significantly different from zero reflecting presence of unobserved taste variations along cost and time sensitivity dimensions. The correlation between the random coefficients is insignificant. Further, compared to the fixed coefficients model, the base coefficients for cost and time are higher in magnitude.

As expected by allowing for taste variations in cost coefficient in addition to random taste variations in time coefficient[23] as in Model 7, Model 8 betters Model 7 by 21 log-likelihood units.

In this model, the value of time is the ratio of two normal random variates. Consequently, the mean value of time is calculated based on a second order Taylor series expansion, while the variance of the value of time is based on a first order Taylor series expansion about the respective means (see Appendix H). If $Y = X_1/X_2$, with $\mathrm{E}(X_i) = \mu_i$, $\mathrm{var}(X_i) = \sigma_i^2$ for $i = 1,2$, and $\mathrm{cov}(X_1,X_2) = \sigma_{12}$,

$$\begin{aligned} \mathrm{E}(Y) &\approx \frac{\mu_1}{\mu_2} - \frac{\sigma_{12}}{\mu_2^2} + \frac{\mu_1\sigma_2^2}{\mu_2^3} \\ \mathrm{var}(Y) &\approx \frac{\sigma_1^2}{\mu_2^2} - \frac{2\mu_1\sigma_{12}}{\mu_2^3} + \frac{\mu_1^2\sigma_2^2}{\mu_2^4} \end{aligned}$$

The mean VOT is 18.0 fl/hr, the mean individual standard deviation is 10.4 fl/hr, and the standard deviation of mean VOT is 3.1 fl/hr. The total standard deviation equals 10.8 fl/hr.

4.4.3 Summary of Estimated models and Values of Time

Since the models cannot be specified as "nested" special cases of others, the classical likelihood ratio tests cannot be applied. While the log-likelihood values from the

[23] Rather, variations in the ratio of time to cost coefficient.

	Parameter	Estimates	Std. err.	t-stat
Cost Effects	COST	0.0130	0.0009	13.98
	INC15-	-0.0005	0.0035	-0.13
	INC1520	0.0007	0.0021	0.35
	INC2025	0.0002	0.0017	0.13
	INC4060	-0.0026	0.0011	-2.38
	INC6080	-0.0035	0.0012	-2.84
	INC80+	-0.0065	0.0014	-4.70
	KIDS	0.0379	0.0167	2.27
Time Effects	TIME	0.1735	0.0161	10.79
	DINKS	0.0262	0.0154	1.70
	SOLO	0.0302	0.0181	1.67
	PARTIME	0.0277	0.0174	1.59
	AGE20-	0.0227	0.0287	0.79
	AGE3650	-0.0336	0.0132	-2.54
	AGE50+	-0.0444	0.0172	-2.58
	FEMALE	-0.0337	0.0128	-2.62
	FR-35	0.0078	0.0216	0.36
	FR-36-49	0.0277	0.0129	2.15
Parameters of Random Coeff.	σ_C	0.0074	0.0006	12.08
	σ_T	0.0609	0.0083	7.37
	ρ	0.0561	0.1626	0.35

Log-likelihood of naive model = -3697.94
Log-likelihood at convergence = -2566.44
$\bar{\rho}^2 = 0.30$
Number of observations = 485

Table 4.12: Model 8 – Random coefficients model

different models give some indication of how closely the models fit the data, no easy distribution theory exists or can be developed to choose from among the different models. One exception, is the Hausman's test for misspecification[24]. We can conduct this test for comparing Model 3b and Model 5, where the estimates in Model 3b are consistent and inefficient, while the estimates in Model 5 are consistent and efficient under the null hypothesis of no misspecification. Since the test statistic is 49.23, and the critical value at 5% significance for 23 degrees of freedom is 35.17, we reject Model 5.

In similar vein, we can compare Model 4a and Model 7. Since the test statistic is 34.22, and the critical value at 5% significance for 19 degrees of freedom is 30.14, we reject Model 7.

In Table 4.13, we summarize the log-likelihood, the number of estimated parameters, the Akaike Information Criterion (AIC), and $\bar{\rho}^2$ for each of the estimated models. Among models which ignore interdependencies among responses, looking at AIC's, Model 3b with four latent classes and taste modifiers is the best estimated model, followed by Model 3a, Model 4b, Model 4a, Model 1, Model 2, and Model 0. Surprisingly even the best randomly distributed value of time model is worse than the worse four-latent class model (Model 4a) by 47 log-likelihood units, although it does better than the two-latent class model (Model 1) by 20 log-likelihood units.

Among models which attempt to capture interdependencies among responses, the four latent class model with taste modifiers and agent-effects specification for the criterion functions seems to fit the data the best. Further, by allowing both the time and cost coefficients to be randomly distributed led to improvement in fit vis-a-vis the lognormal value of time wherein only the ratio of the time coefficient to the cost coefficient was randomly distributed. In general, capturing taste variations using the latent class concept is effective in this case study as evidenced in the significant improvement in fit.

[24]Under the null hypothesis of no misspecification in an asymptotically efficient estimator, a χ^2 statistic is constructed as a function of asymptotically efficient parameter estimates, consistent (but inefficient) parameter estimates, and the difference of their respective covariance matrices (see Hausman [1978]).

Model type	Model	Log-lik.	# of par.	Akaike	$\bar{\rho}^2$
No interdep. among responses	Model 0	-3028.49	18	-3046.49	0.18
	Model 1	-2913.18	20	-2933.18	0.21
	Model 2	-2911.73	22	-2933.73	0.21
	Model 3a	-2833.24	23	-2856.24	0.23
	Model 3b	-2786.85	23	-2809.85	0.24
	Model 4a	-2894.41	19	-2913.41	0.22
	Model 4b	-2884.42	19	-2903.42	0.22
With interdep. among responses	Model 5	-2533.87	23	-2556.87	0.31
	Model 6	-2455.94	25	-2480.94	0.33
	Model 7	-2587.58	19	-2606.58	0.29
	Model 8	-2566.44	21	-2587.44	0.30

Table 4.13: Comparison of auxiliary statistics of estimated models

Now we turn our attention to the estimated values of time and their variability as implied by the different models. In Figures 4-1, 4-2, and 4-3 the histograms of the mean value of time are plotted for the different models. In Table 4.14 the sample mean, the average standard deviation[25], the standard deviation in the sample of the mean value of time, and the total standard deviation of the value of time in the sample are presented. The sample statistics change quite considerably for the different models.

Among models which ignore interdependencies among responses Model 3b provides the largest mean value of time and total standard deviation. The lognormal model (Model 4b) provides higher variance for the mean of VOT, while the latent class choice model captures significantly higher individual level variability. Among models capturing interdependencies among responses the four latent class model with taste modifiers and agent-effects specification for the criterion functions exhibits the largest mean and total variance in VOT. Surprisingly, the four-latent class model with taste modifiers and identical latent class among responses provides much lower

[25]Since we allow for each individual to have distributed VOT the average individual standard deviation is calculated as the square root of the sample average of the individual variances of VOT.

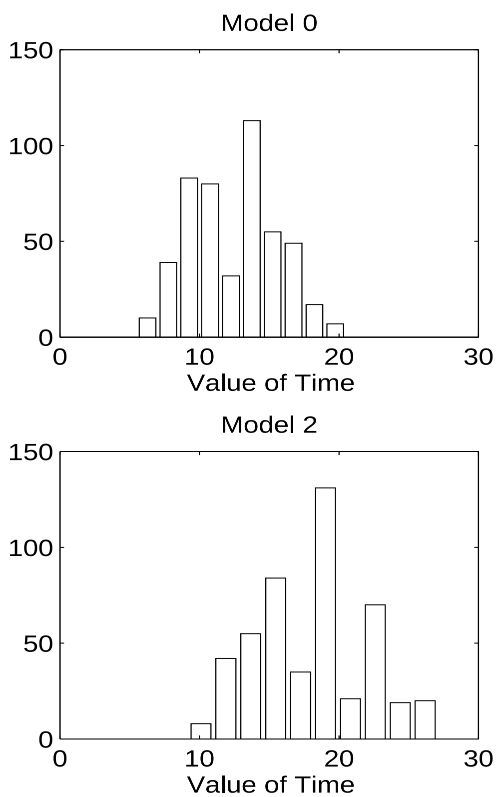


Figure 4-1: Histograms of Mean Values of Time: Model 0, Model 1 and Model 2

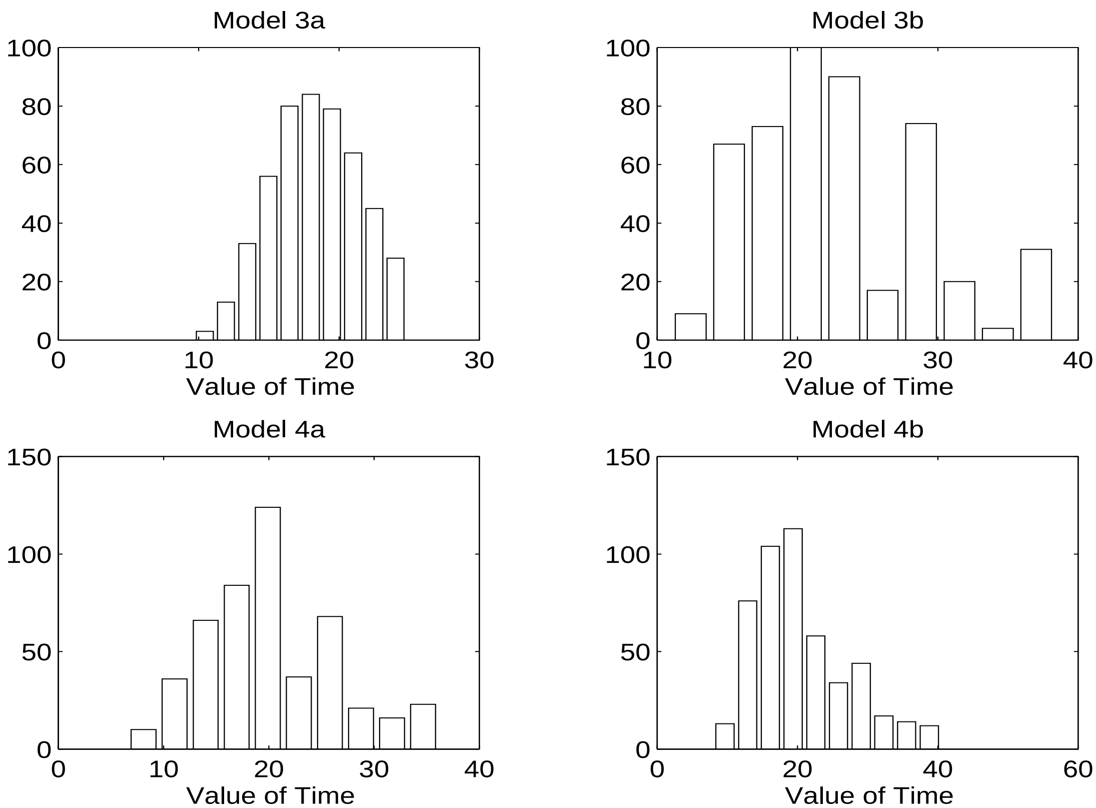


Figure 4-2: Histograms of Mean Values of Time: Model 3a, Model 3b, Model 4a and Model 4b

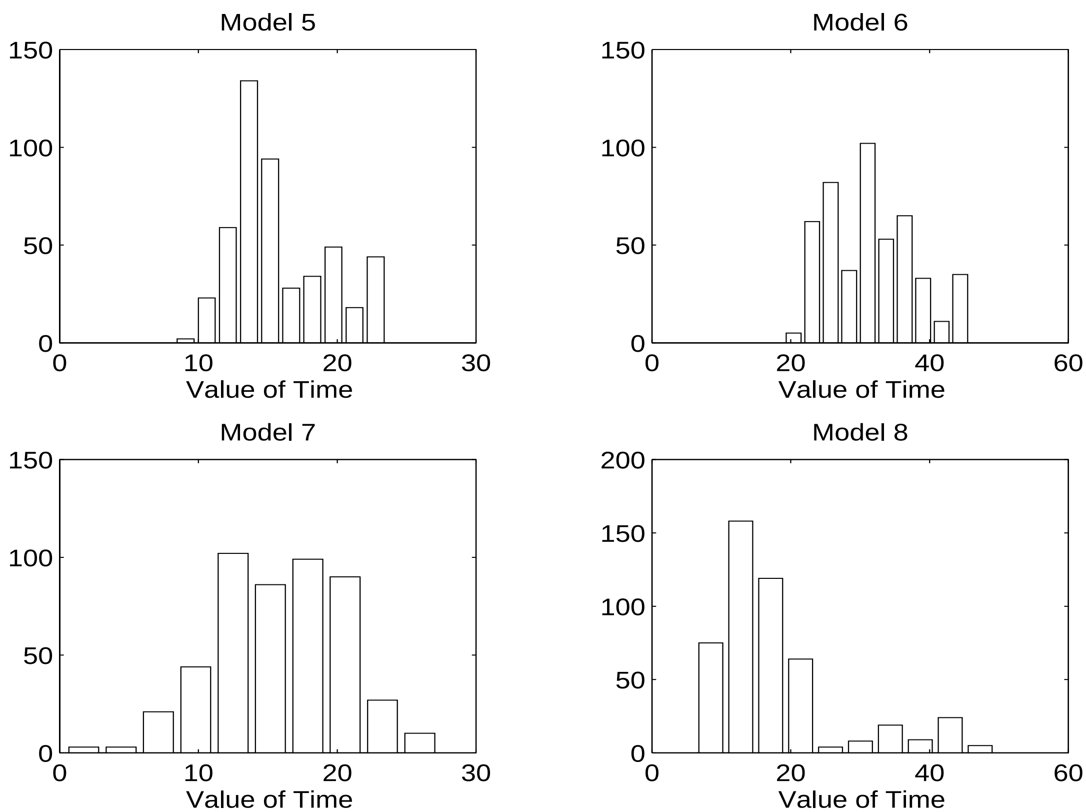


Figure 4-3: Histograms of Mean Values of Time: Model 5, Model 6, Model 7 and Model 8

Model type	Model	VOT			
		Mean	Std. dev		
		fl/hr	Mean Indivi.	of Mean	Total
No interdep. among responses	Model 0	12.5	–	3.1	3.1
	Model 1	17.4	9.6	3.1	10.1
	Model 2	17.9	11.1	3.9	11.8
	Model 3a	18.3	12.9	3.1	13.3
	Model 3b	22.8	27.1	6.2	27.8
	Model 4a	20.3	21.0	6.2	21.9
	Model 4b	20.5	21.6	6.8	22.6
With interdep. among responses	Model 5	15.9	15.6	3.4	16.0
	Model 6	31.5	32.7	6.4	33.3
	Model 7	15.6	14.7	4.5	15.4
	Model 8	18.0	10.4	3.1	10.8

Table 4.14: Estimated Values of Times from the sample

mean and variance of VOT compared to the similar model with no interdependencies among responses. This theme is manifested in the lognormal value of time model with interdependencies among responses as it provides lower mean and variance of VOT compared to the similar model with no interdependencies among responses. A partial explanation for this reduction in mean and variance of the values of times is that in the models with interdependencies we "restrict" the unobserved "factors" to be same across responses for the same individual, thereby leading to lower variability of VOT. From the table one can conclude that VOT depends significantly on the modeling approaches adopted.

Now we turn our attention to the comparison of the prediction results when the estimated models are applied. We assume that the individual is provided two travel alternatives with the travel times differing by 30 minutes. Figures 4-4, 4-5, 4-6, and 4-7 depict the average (over the sample) probability of choosing the faster alternative as the cost difference increases from 0 to 20 fl/hr.

As seen in Figure 4-4, the decline in the probability of paying for 30 minutes savings is much faster for Model 0 than those of Model 1 and Model 2. Further, for the curves of Model 1 and Model 2 there appears to be a sharper decline around

7 fl/hr. This price approximately coincides with the value of time of one of the classes. Thus at prices below this willingness to pay threshold, individuals in this class accept the faster alternative, and when price goes beyond this threshold, the individuals in the class reject the faster alternative, and prefer instead the cheaper alternative.

A similar trend is noticed in Figure 4-5 with the curves for Model 3a and Model 3b indicating segments in the population with higher willingness to pay for travel time savings. Consider, for example, the curve corresponding to Model 3b. The curve declines steeply around prices of 6 fl/hr and 20 fl/hr, and it must be noted that these points correspond to the values of times of two classes.

In Figure 4-6, the curves of Model 4a and Model 4b closely follow each other, and this similarity is also manifested in the VOT statistics and the overall goodness-of-fit.

In Figure 4-7, we graph the curves for the models incorporating interdependencies among multiple responses. Even here we notice the curves corresponding to latent class choice models to decrease at a steeper rate near prices corresponding to the values of time of classes, while the curves for the random coefficients model and the lognormal model are much smoother. The curve for Model 6 indicates much higher willingess to pay for a substantial fraction of the population, unlike all other models. In this regard, this model must be viewed with caution, and the unusual results may indicate some form of misspecification or identification problem, although the model, in general, appears to fit data the best.

4.5 Summary

The latent class choice model to capture taste variations was applied in a case study to evaluate the value of time in the transportation context. The estimated models suggest the efficacy and practicability of the modeling approach compared to extant approaches of introducing interaction variables and random coefficient models.

At this stage, the research has demonstrated the potential of our modeling approach. Further, from a practical and computational standpoint, the models were estimated without much difficulty in a 486-machine in a matter of 5-6 hours.

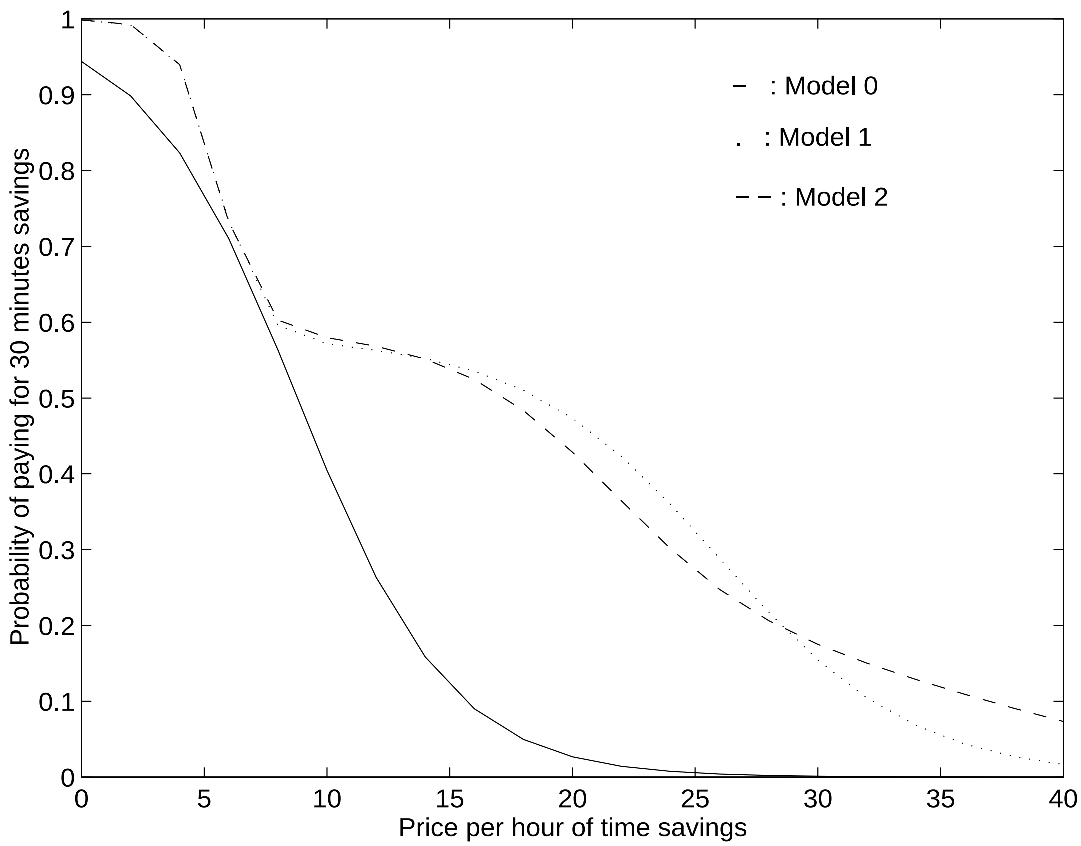


Figure 4-4: Predictions from Model 0, Model 1 and Model 2

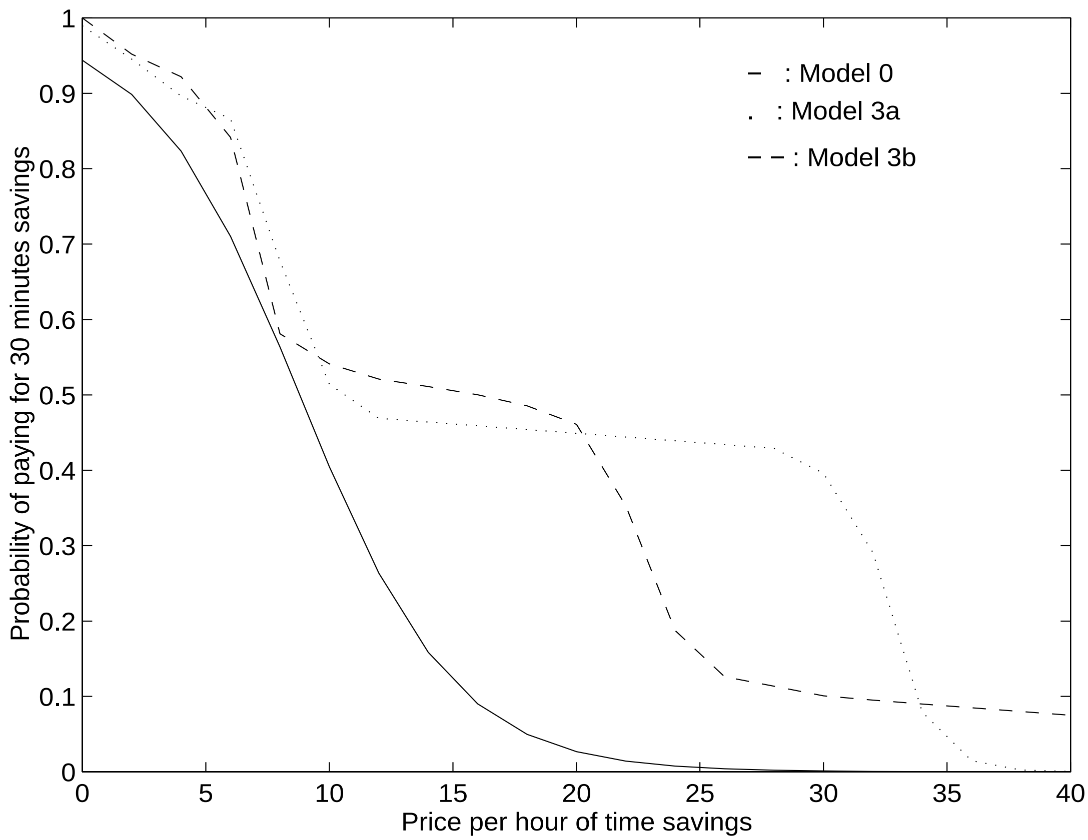


Figure 4-5: Predictions from Model 0, Model 3a and Model 3b

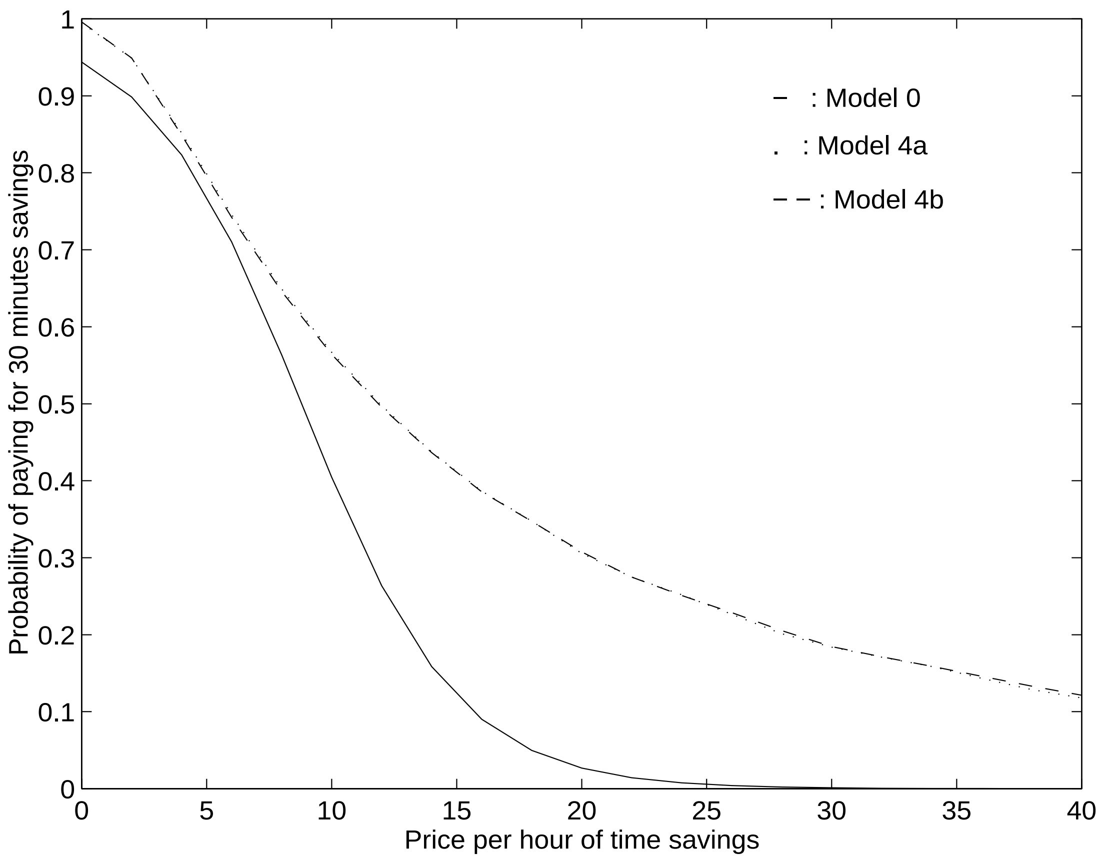


Figure 4-6: Predictions from Model 0, Model 4a and Model 4b

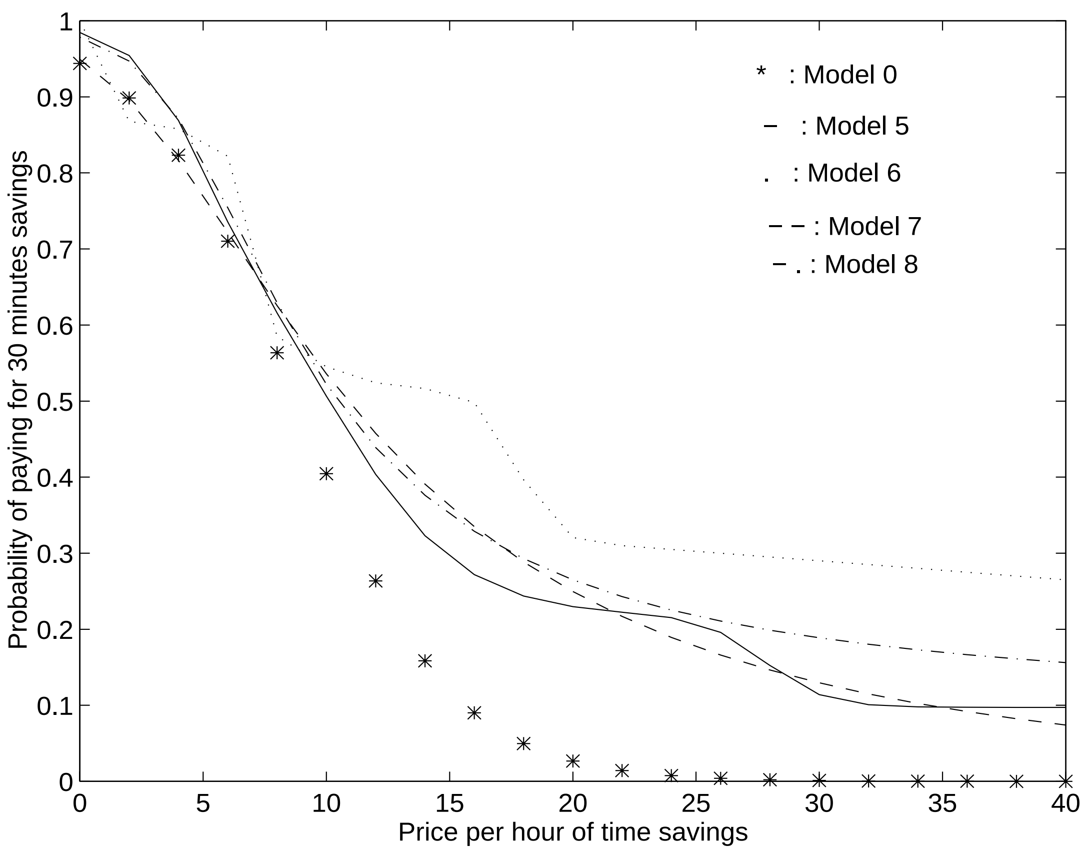


Figure 4-7: Predictions from Model 5, Model 6, Model 7 and Model 8

An (empirical) caveat of the LCCM must be noted. We attempted the estimation of LCCM's with two criterion functions, with two levels in one sensitivity dimension, and more than two levels along the other sensitivity dimension. The estimation procedure did not converge as the correlation parameter increased to 1. As noted earlier we are yet to develop general necessary and sufficient conditions for the identification of all model parameters. We conjecture that this non-convergence may be due to model identification problem, or "empirical identification" problem wherein we are attempting to capture latent classes which the *data* does not support. We also conjecture that incorporating indicators of latent class such as importance ratings of attributes may "lend character" to the latent classes since indicators may be viewed as attributes of latent classes, and thus provide adequate information for the empirical identification of the LCCM.

Chapter 5

Latent Class Choice Model for Decision Protocol Heterogeneity: Case Study – Application to Stated Preference Data

5.1 Introduction

In this chapter we focus our attention on capturing the different decision protocols adopted by individuals while making choices in the actual market environment (referred to as revealed preferences [RP]), and while indicating preferences in hypothetical preference tasks (referred to as stated preferences [SP]).

The chapter is organized as follows: In section 5.2 we provide evidence from the literature as to the significance of variations in decision protocols, and discuss possible causes for such variations in general (i.e., both in RP and SP settings). In section 5.3 we highlight the factors leading to decision protocol variations which are specific to hypothetical preference tasks. In this section, we also briefly review the RP and SP data combination technique which addresses some of the biases inherent in SP data vis-a-vis RP data. Since the analyst is unable to observe the decision protocol adopted by an individual, it may be characterized through a latent class, and to this end in section 5.4 we outline the latent class choice model for explicitly capturing unobserved decision protocols. Since decision protocols in RP and SP settings may differ for the *same individual*, we also discuss the need to combine RP and SP data, and outline an approach to validate decision protocols exhibited in SP analysis with those of RP data, if both RP and SP data are available. In section 5.5 we discuss the SP data used in the case study to assess the potential of the latent class choice model for decision protocol heterogeneity, while in section 5.6 we present estimation results. It must be noted that though the case study is in the context of simulated choice experiments the approach for capturing decision protocol heterogeneity is fairly general and may be adopted to RP data, and more judiciously to combine RP and SP data.

5.2 Decision Protocols

We are interested in a *descriptive* model that postulates how individuals behave and respond to changes in policies, prices and features of products and services, etc. It is well documented in the behavioral decision research literature that the perceptual, emotional and cognitive processes which ultimately lead to the choice of an individual differ considerably among individuals and across choice contexts (see, for example, reviews in Rapoport and Wallsten [1972], Einhorn and Hogarth [1981], Payne *et al.* [1992]).

To motivate the latent class choice model for decision protocol heterogeneity, we briefly review relevant literature highlighting the need to capture decision protocols, and its significance in different choice contexts. We focus our attention on the different decision protocols presumably adopted by individuals both in the *actual* market setting and the *simulated* market setting. This is followed by an elaboration of the causes for this heterogeneity.

Most of the theoretical and empirical work in choice analysis is centered on the "utility maximizing" principle which assumes a rather sophisticated and cognitively demanding representation of the decision protocol. It assumes that the model predicts choice regardless of the number of alternatives to be compared and the number of attributes of each alternative to be examined. It must be noted that the utility model is also often referred to as the *algebraic model* or the *information integration model* in the psychology literature (Anderson [1974a, 1974b], von Winterfeldt and Fischer [1975], Shanteau [1977]). In reality, individuals may adopt a variety of other decision protocols such as dominance rules, satisfaction rules, lexicographic rules, random choice, etc., of varying complexity (see, for example, Slovic *et al.* [1977], Svenson [1979]).

Decision protocol variations have been observed in the studies of the cognitive processes leading up to a decision, referred to as *process tracing*[1] studies. The aim of the process tracing study is to reveal the train of thought leading to the final decision, with the focus on *what* content or information is processed, and *how* it is processed. Data is collected during or after the decision process such as eye movements of the individual or information requests which represent the information search pattern, and think-aloud or verbal reports.

First, the order in which an individual seeks and evaluates the information of the choice problem may be related to the cognitive process leading to the final decision. Eye movement recordings and records of the information explicitly demanded by the individual indicate the information search pattern. Consequently, variations in these search patterns have been linked to protocol variations (see, for example, Russo and Rosen [1975], Just and Carpenter [1976], Payne [1976]). For example, complete information searches may be associated more with a compensatory decision rule, while incomplete searches eliminate compensatory decision rules.

Second, verbal protocols[2] reflect on *how* the information is processed (see, for example, Newell and Simon [1972], Russo and Rosen [1975], Nisbett and Wilson [1977]). Consequently, differences in the verbal protocol reports indicate decision protocol

[1]See Svenson [1979] for a more detailed discussion of process tracing techniques.

[2]Verbal protocols may be *retrospective* pertaining to the individual's interpretation of the decision protocol adopted, or may be *simultaneous* pertaining to data collected during the choice process.

variations.

Now we turn our attention to the possible causes for variations in decision protocols.

Information Effects

Economic theory concentrates on the individual decision-making process of what goods and how much of each good to consume or purchase, assuming perfect information about the availability of goods and their attributes. But in reality, an individual has limited *prior information* about consumption opportunities. Svenson [1979] argues that situations where information about the attribute is missing as a result of imperfect discrimination or of unreliability of available information may encourage the adoption of rules depending on minimum differences on the attribute. Specifically, differences between attribute levels across alternatives may be *perceived* only when they are greater than some attribute-specific thresholds (see, for example, the lexicographic semiorder rule of Tversky [1969]).

Deviations from Rational Behavior and the Notion of Bounded Rationality

The common assumption of utility maximizing decision protocol may not be appropriate since it assumes a rather sophisticated cost/benefit analysis paradigm, and that all individuals are *rational*. Simon [1955] argues that all behavior cannot be explained in this cost/benefit paradigm. Specifically, Simon [1955, 1957] (see also March and Simon [1958], and March [1978]) argues that individuals may exhibit rationality, but only within the constraints of the individual's perceptions of the choice context, and his/her limited information processing capabilities and ignorance of the "optimal" rules (if any). Simon [1956] suggests that individuals attempt to compensate for these limited information processing abilities by constructing a simplified representation of the choice context and behaving *rationally* within the constraints of such a representation (often referred to as the notion of *bounded rationality*).

Cognitive Representation of the Choice Problem

Individuals' limited *information processing* capabilities lead to simplification in the cognitive representation of the alternatives through techniques such as "chunking" wherein the individual interprets a complex array of attributes of alternatives by "recoding" information into larger chunks (Miller [1956], Simon [1960]). Bruner *et al.* [1962] refer to the same phenomenon as "cognitive categorization". Specifically, Park [1978] suggests that for each attribute, an individual is assumed to set up cognitive categories. These categories further takes on three different forms, depending upon how the individual codes the attributes including: (a) negative, neutral and positive affects on the categories of the attribute (e.g., price attribute may be coded into least preferred, neutral and most preferred categories depending on the price falling between certain thresholds), (b) neutral and positive affects on the categories of the attribute (e.g., price attribute may be categorized into neutral (most preferred) depending on whether the price is higher (lower) than a reservation price), and (c) indifferent to the attribute.

Svenson [1979] argues that the representation system and the sequence of rules applied may be continuously influencing each other. For example, it may be assumed that the individual changes the degree of complexity of the representation system to meet the requirements of a decision rule he/she wants to apply.

Multiple Goals or Criteria

The concept of optimality is defined with a single criterion or goal. However, judgments and choices are usually based on multiple goals or criteria. When such goals conflict there can be no optimal solution in the sense of a single criterion case (Shepard [1964]). Einhorn and Hogarth [1981] argue that even if the trade-offs or compromises between the goals are clearly defined, the single goal situation is transformed into a multiple goal case when the judgments and choices are considered over time, since conflicts between short-run and long-run strategies can exist even with a single well-defined criterion.

Time and Budget Effects

Decision protocols may vary across individual's time and budget constraints while searching for information regarding available alternatives in the actual market environment. Shugan [1980] argues that cost/benefit analysis must be expanded to include "the cost of thinking".

Montgomery and Svenson [1976] suggest that different decision rules may require different amounts of cognitive effort, so that different types of rules may be ordered on an *effort* continuum. Individuals who want to minimize the amount of cognitive effort expended may apply simpler rules before trying more complex ones. For example, if the simpler rules do not lead to a unique choice, then more complex rules may be applied.

Importance and Familiarity with the Choice Context

It must be noted that the same individual may adopt different decision protocols depending on the choice context. For example, when an individual is faced with a choice in a new situation, the more complex decision rules may be adopted. On the other hand, if similar choice situations have been faced earlier then simplifying strategies or heuristics may be used (Tversky and Kahneman [1974]). Heuristics may be developed from the individual's own experiences with similar choice situations. Hogarth [1974] argues that the choice problem may be solved on the basis of its similarity to one of a number of classes of decision situations, and each class has its own standard solution.

Further, the importance of the choice context affects the decision protocol adopted. For example, one may adopt a more complex decision protocol in the purchase of a car or house, while a simpler strategy may be adopted in the purchase of cereal.

Complexity of Choice Problem

Decision protocols may also depend on the complexity of the the choice problem where complexity is defined in terms of the product of number of alternatives and the number of attributes characterizing each alternative. Svenson [1979] in a meta-analysis of process tracing studies observes that the percent of information searched by an individual decreases with both increase in alternatives and attributes, while the rate of decrease is higher with attributes.

Even if all the individuals adopt the same decision protocol such as the "utility maximizing" principle, variations may exist in how individuals weigh the different attributes. This theme is adequately demonstrated in the case study presented in chapter 4. In similar vein, non-compensatory models such as the satisficing model which is built on attribute-specific criteria, and the lexicographic model which is built on *ranking* of attribute importance, may have these criteria and rankings individual-specific. It must be also noted that the *same individual* may adopt different decision protocols in the actual market environment and the simulated market environment.

5.3 Decision Protocol Variations Specific to SP Tasks and Modeling Approaches

Individual's actual preferences towards consumption bundles can be inferred only from observed market behavior. Consequently, most applications of discrete choice models have used as the basis of analysis RP data. However, in recent years, SP techniques, wherein preference data is collected by presenting hypothetical scenarios to the respondents and requesting for their preferences, have been increasingly used. In the following paragraphs we discuss the usefulness and the growing need for SP techniques, highlight the causes for variations of decision protocols in such data, and present empirical evidence. Such evidence motivates the heightened need to adequately address decision protocol variations in order to apply judiciously the results from SP analysis.

The analysis of SP data originated in the seminal work of Luce and Tukey [1964], wherein stated preferences and the associated analysis techniques were referred to as "conjoint measurement" and "conjoint analysis", respectively. Such techniques have been popular among market researchers since the 1970's (see, for example, Green and Rao [1971], Green and Srinivasan [1978], and Cattin and Wittink [1982], Louviere [1988a]).

In marketing research, SP techniques are particularly useful to assess the demand for new products and services. If the consumer has no access to new options, or the option is yet to be implemented since it is in the concept phase of development, then consumers may be presented with hypothetical scenarios, and asked to furnish choices/preferences. The levels of the attributes of the options, and the ambient decision-making environment of the consumer may be varied judiciously to appear plausible, relate to the consumer's experiences with similar products and ensure competitive trade-off. The consumer could be asked, for example, to choose among options (referred to as a *choice experiment*), or indicate preferences through *ranking* or *ratings* of the options presented. Further, data reflecting the cognitive processes of consumers such as how consumers learn about new products and services through information search can be collected in a simulated environment.

Although SP techniques have been used in marketing for a long time, there have been very few reports of their applications in travel demand analysis in the 1970's. Instead, applications of discrete choice models in travel demand analysis have primarily utilized revealed preferences (see, for example, Ben-Akiva and Lerman [1985]). This is due in part to the need for a travel demand model to provide forecasts which must be consistent with actual behavior. However, in the 1980's there has been an increasing interest in the adoption of such techniques to analyze travel behavior, leading to a special issue of the Journal of Transport and Economics and Policy on SP methods in transportation research (see Bates [1988], Bradley [1988], Fowkes and Wardman [1988], Hensher *et al.* [1988], Kroes and Sheldon [1988], Louviere [1988b], and Wardman [1988]). More recent applications of SP techniques in travel demand analysis include: Fowkes [1991], Bates and Terzis [1992], Copley *et al.* [1993], and Bradley and Daly [1993].

In recent years, with the new developments in transportation and information technologies and their interactions, transportation researchers have recognized the importance and the heightened need for SP techniques, and have increasingly used or proposed the adoption of such techniques to assess the transportation impacts of such new concepts (see, for example, Bernardino *et al.* [1993], Mahmassani *et al.* [1993], Sullivan *et al.* [1993], Ben-Akiva and Gopinath [1993]).

For the estimation of preference models, Ben-Akiva *et al.* [1991] outlined the implications of the differing characteristics of RP and SP as:

- RP data are cognitively congruent with actual behavior;
- SP techniques form the only means of obtaining preferences towards new products and services; and
- Trade-offs among attributes are identifiable from SP data since the attribute levels can be artificially set, the range of attribute levels extended, multicollinearity among attributes reduced, and attributes are free of measurement errors.

In the first two decades of applications of SP techniques, the econometric tools adopted were the same as those used for RP data, although researchers have recognized the issue of validity of SP responses. In the following paragraphs we highlight the important determinants of decision protocol variations specific to SP tasks.

Presentation Effects

The presentation format affects information search and evaluation patterns. For example, Bettman and Kakkar [1977] presented subjects with information about 11 alternatives, each characterized by 13 attributes (a) in matrix form with an attributes $\times$ alternatives design, (b) in an alternative centered form where each alternative was described in a booklet, and (c) in an attribute centered form in which information about each attribute was given in a booklet. The results showed that when the information was given in a alternative or an attribute centered way, the subjects adapted to the way of presentation by processing information in an intraalternative and an intraattribute manner, respectively[3].

[3]If the individual uses alternatives as reference points and investigates all the attributes for one alternative before going to the next alternative, the processing is *intraalternative*. On the other hand,

Context Effects

Tversky and Kahneman [1986] present evidence wherein descriptions of the choice problem which are normatively equivalent lead to different responses. Prospect theory (Kahneman and Tversky [1979], Thaler [1985]) has endeavored to explain such framing effects as the result of how the individual codes the outcomes of the choice problem based on some reference point or expectation level.

Further, SP studies assume that preferences are defined over alternatives and their attributes. But in order to reduce the complexity of the SP task for respondents, the number of attributes describing an alternative is usually restricted vis-a-vis the set of attributes which may be considered in the actual market setting. Consequently, the respondent may take into consideration some attributes not specified in the experiments by imputing them from his/her perceptions. This effect may be significant if the alternatives presented are "branded" (i.e., has an associated name) leading to erroneous conclusions regarding the trade-offs between the specified attributes.

Response Elicitation Effects

Decision protocol may be affected by the elicitation format in SP tasks such as ranking, rating, choice, etc. Consequently, SP responses may depend on how the questions are posed. Different response modes can lead to differential weighting of attributes and different preference assessments. If respondents are asked to value a product or service through an open bid question such as "How much are you willing to pay?", they will offer different values from the ones obtained if they are asked questions which give some starting values or ranges such as "Would you be willing to pay $X?", and "Would you be willing to pay at least $Y?". Such artificially framed SP tasks, such as trade-off exercises, and open bids tend to detract from the validity of survey responses. Tversky *et al.* [1988] citing differences in responses to choice versus matching tasks suggest the notion of strategy compatibility between the nature of

if the individual anchors on attributes and investigates the levels of attributes across alternatives before going to the next attribute, the processing is *intraattribute* (see Svenson [1979]).

the elicitation mode, such as ordinal or cardinal, and the decision process employed by the individual.

Attention and Comprehension of the Preference Tasks

Since a hypothetical scenario does not generally affect the welfare of the respondent (unlike actual market behavior), the respondent may be uninterested in the SP survey, and consequently be careless in the response as he/she might not make a rational decision.

Decision protocols for SP data may differ from those in actual market environment since situational constraints which affect the actual choice process may be ignored in the SP experiments. Further, lack of realism in the SP scenarios in creating the intended decision-making environment, and apathy and laziness on the part of the respondent while responding to SP surveys, may lead to differing decision protocols.

Other Biases

Specific examples of other forms of biases in SP responses include: (1) *prominence hypothesis* wherein the respondent evaluates alternatives by considering the most important attribute, (2) *strategic behavior* or *policy-response bias* if the hypothetical scenario does affect the respondent's welfare, but it affects him or her in a way different from direct exposure to the "real market" situation, and the respondent believes that he or she will benefit by responding in a certain way, (3) *inertia bias* if the respondent prefers to maintain the status quo instead of changes posed in the SP surveys, and (4) *justification bias* wherein the respondent may want to justify past behavior and respond in that way even to a hypothetical scenario.

Given the overview of the causes for differing decision protocols in SP tasks, we highlight its significance in the following paragraphs. The issue of validity of SP tasks has long been recognized in gambling problems. For example, Slovic *et al.* [1965] found that imaginary incentives led subjects to employ simpler decision protocols than did real payoffs. Slovic [1969] provided evidence that when choices are hypothetical, subjects maximized gain and discounted losses, but when the choices had real consequences subjects were considerably more cautious.

A recent example illustrating the issue of presentation and elicitation effects on SP responses is the work of McFadden and Leonard [1993] who conducted tests of stability of willingness to pay for saving wilderness areas from lumbering obtained using the SP approach. They compared results from alternative SP experiments that varied in response formats, question phrasing and information provided to the respondent. They found great sensitivity of the preference model to the SP elicitation format, information provided and question phrasing.

Another example, in the context of preferences towards hypothetical travel alternatives, is the work of Widlert [1994] wherein the objective was to examine how different aspects of the design of the SP experiment influence the estimated preference models. Specifically, the study assessed the differences between rating, ranking and pairwise choices, the importance of adapting the levels of the attributes to the respondent's own experiences, the effects of different number of alternatives, and a comparison between absolute and relative attribute levels. In the study, 25 different types of interviews were conducted on long distance trains in Sweden with the respondents requested to evaluate different train alternatives. Significant differences in the values of times calculated from the travel time and travel cost coefficients of preference models estimated on different SP data sets were observed. Widlert argued that such variations in values of times are mostly attributable to:

1. Respondents tend to simplify the SP task as an exploratory analysis of the data revealed that lexicographic answers based on one attribute accounted for a significant fraction of the responses. Specifically, the lexicographic rule was adopted the most when the respondent was asked to rank the alternatives, and the least when the respondent was asked to rate the alternatives or make a pairwise choice.

2. Values of times differ considerably depending on whether or not the SP tasks were adapted to be realistic to the respondent's own situation.

In the analysis of ranking data, Ben-Akiva *et al.* [1991] raised concerns about the reliability and stability of responses among the preference ranks. Also the effect of actual choice used as a reference for ranking hypothetical alternatives was significant.

To address some of the biases inherent in SP tasks, a significant, albeit simple, methodological framework for estimation of preference model from RP and SP data was developed in the late 1980's which explicitly recognizes the *complementary characteristics* of RP/SP data. Ben-Akiva and Morikawa [1990a, 1990b] proposed the combined RP/SP method for RP and SP data, the key features of which include:

- *Efficiency*: joint estimation of underlying preference from all the available data;
- *Bias correction*: explicit response models for SP data which include both preference parameters and bias parameters; and
- *Identification*: estimation of preference parameters not identifiable from RP data due to low variability.

It must be noted that in the RP/SP data combination method of Ben-Akiva and Morikawa an underlying "utility maximizing" decision protocol is assumed for the RP and SP data generating processes. Given the potential significance of the variations in decision protocols, in the next section we endeavor to capture explicitly such unobserved variations.

5.4 The Model

Although the individual has a wide array of decision protocols at his/her disposal while making a choice, the actual decision protocol adopted in a particular situation in *unobserved*. The range of decision protocols which may be used poses questions as to how one decides to "choose" (Beach and Mitchell [1978], Wallsten [1980]). The decision protocol adopted can be viewed as being generated by a process wherein the different decision protocols compete with each other. The individual adopts the decision protocol which suits him/her the most. Einhorn and Hogarth [1980] conceptualize the individual's evaluation of the decision protocol as a multidimensional object containing attributes such as speed of execution, demands on memory (e.g., storage and retrieval), computational effort, chance of making errors, etc.

Consequently, each decision protocol may be associated with an unobservable concept characterizing "desirability" of the decision protocol as a function of individual characteristics such as time pressure, education, etc., coupled with intrinsic features of the decision protocol.

<u>Notation</u>

y_{in} = choice indicator taking the value 1 if alternative i is chosen by individual n and zero otherwise.

s = latent class index, $s = 1, \ldots, S$.

R_s = decision protocol specific to class s.

C_n = choice set available[4] to individual n with $|C_n| = J_n$.

X_n = attributes of alternatives and individual characteristics which affect the class-specific choice model.

Z_n = attributes of alternatives and individual characteristics which affect the class membership.

Assume that each individual adopts one of a set of S decision protocols while making a choice. Since the decision protocols are unobserved, we assume that there are S latent classes with each latent class s characterized by its own decision protocol. A class membership model $\mathrm{Q}_s(Z_n; \theta)$ assigns an individual to a latent class, and the class-specific choice model $\mathrm{P}(y_{in} = 1|X_n, R_s; \beta)$, predicts the choice behavior of an individual with decision protocol R_s. The class-specific choice model may be *deterministic* or *probabilistic* depending on the class-specific decision protocol and the problem context. The latent class choice model expressing the probability of the

[4]This refers to the set of alternatives *deterministically* available.

individual choosing alternative i is written as:

$$\mathrm{P}(y_{in} = 1 | X_n, Z_n; \beta, \theta) = \sum_{s=1}^{S} \mathrm{P}(y_{in} = 1 | X_n, R_s; \beta)\ \mathrm{Q}_s(Z_n; \theta) \tag{5.1}$$

The assignment of individuals to the latent classes may be captured through a class membership model such as the categorical criterion model as discussed in section 3.2, with individual characteristics and alternative attributes Z_n affecting class membership. For example, consider a choice situation with two classes: (1) *Class 1*: utility maximizers who choose from the *full* deterministically available choice set, and (2) *Class 2*: utility maximizers who choose from a choice set generated through a non-compensatory screening process. The probability of an individual being in each of the two classes may be affected by characteristics such as number of alternatives available, education, sex, income, etc. Consequently, the choice model in the first class can be operationalized by an MNL model, while the choice model for the second class is a choice model with latent choice sets.

Validation of SP Decision Protocols

Even if the preference model estimated on SP data suggests the existence of more than one decision protocol, the question remains as to whether such heterogeneity may exist in the *actual market environment*. If RP data is not available, then it is left to the analyst's judgment as to how the SP model can be utilized in providing forecasts and other model applications. Specifically, if the analyst believes that a particular decision protocol identified in the SP tasks will not be exhibited in the actual market setting, then the decision protocol may be conveniently ignored in model application. For example, a decision protocol wherein the individual picks the alternative actually chosen in the market environment ignoring alternatives' attribute levels, may be eliminated in model application[5]. Else, if the analyst can assume that the psychological "laws" governing the SP data generating process and the choice

[5] It must be noted that even in this example, one may argue for the case of significant *habit persistence* or *inertia* effects.

process which could be adopted by individuals in the actual market are identical, then the latent class choice model estimated on SP data should be applied *as is*.

If RP data is also available, then we can combine it with SP data to assess the stability of decision protocols in both data sets, and more importantly, validate SP decision protocols. To this end, we outline an approach conceptually similar to the combined RP/SP estimation technique, but methodologically a significant departure from the previous approach as we allow for the decision protocols to vary across data sets.

The typical steps in the approach include:

1. *RP Model*: Estimate the latent class choice model with different decision protocols using the RP data.

2. *SP Model*: Estimate the latent class choice model with different decision protocols using the SP data.

3. *Comparison of models*: Herein the utility functions and the criterion functions for each model are "qualitatively" compared to check if the decision protocols are identical in the RP and SP data[6].

4. *Combined Model*: Depending on the similarity of coefficients of the utility functions, or the coefficients of the criterion functions, or both, subsets of the coefficients may be shared across RP and SP models. For example, consider a choice situation wherein the analyst postulates that the individual adopts one of two decision protocols with one being utility maximization. Consequently, the class membership model may be represented by a threshold crossing model with a single criterion function. The utility of alternative i in the RP context and the SP tasks for individual n who is a utility maximizer may be specified as:

$$U_{in}^{RP} = \beta' X_{in;1}^{RP} + \alpha' X_{in;2}^{RP} + \epsilon_{in}^{RP} \tag{5.2}$$

[6]It must be noted that standard likelihood ratio tests can be conducted to check for equality of the full set or subset of coefficients across the RP and SP models by estimating restricted and unrestricted models. But these statistical tests may be conducted in the next step.

$$U^{SP}_{itn} = \beta' X^{SP}_{itn;1} + \gamma' X^{SP}_{itn;3} + \epsilon^{SP}_{itn} \qquad (5.3)$$

where $t = 1, \ldots, t_n$ denotes the t^{th} SP response for individual n. In a similar manner, the criterion functions are specified as:

$$H^{RP}_{n} = \theta' Z^{RP}_{n;1} + \kappa' Z^{RP}_{n;2} + \delta^{RP}_{n} \qquad (5.4)$$

$$H^{SP}_{tn} = \theta' Z^{SP}_{tn;1} + \tau' Z^{SP}_{tn;3} + \delta^{SP}_{tn} \qquad (5.5)$$

In the above equations, β, α, γ, θ, κ, and τ are unknown parameter vectors to be estimated, and the superscripts RP and SP denote the corresponding variables from the RP and SP data, respectively. Bias factors in SP tasks are represented in the utility function through the variables $X_{in;3}$ with associated parameter vector γ, and in the criterion function through the variables $Z_{n;3}$ with associated parameter vector τ. β is a parameter vector shared by the RP and SP utility functions implying that the trade-offs among attributes in $X_{in;1}$ are the same in both the actual market setting and SP tasks. Similarly, θ is a parameter vector shared by the RP and SP criterion functions implying that the effects of $Z_{n;1}$ on the criterion function generating the decision protocols are identical in both the actual market setting and SP tasks.

It is instructive at this point to outline the variables which may be included in $Z_{n;1}$, $Z_{n;2}$ and $Z_{tn;3}$. $Z_{n;1}$ may include socio-economic and demographic characteristics of the individual such as income, gender, employment status, education, etc. Further, it is possible that $Z^{RP}_{n;1}$ and $Z^{SP}_{tn;1}$ $\forall t$ are the same set of variables. $Z_{n;2}$ may include:

1. Dummy variable(s) associated with the criterion specific constant(s) to reflect differences in intrinsic desirability of the decision protocols in the RP context; and

2. Information and situational factors which are specific to the RP context.

Similarly, $Z_{n;3}$ may include:

1. Dummy variable(s) associated with the criterion specific constant(s) to reflect

differences in intrinsic desirability of the decision protocols in the SP tasks[7];

2. Actual RP choice dummies; and

3. Actual SP choice dummies since the choices made in a sequence of SP tasks may affect the decision protocol adopted in the next task[8].

If RP and multiple SP responses are available from the same respondent, then these bits of information may not be statistically independent since unobserved individual-specific factors may be affecting both RP and SP tasks. Consequently, correlations between RP and SP responses, and between SP responses for the same individual may be captured through a variety of ways including:

1. Correlations between the error components of the utility functions of RP and SP model (i.e., ϵ^{RP} and ϵ^{SP} are correlated). Further, if there are multiple SP responses for the same respondent we can allow for correlations between ϵ^{SP} across these responses. Such correlations may be generated through an error-component structure, i.e.,

$$\epsilon_{in}^{RP} = \zeta_{in} + \nu_{in}^{RP} \tag{5.6}$$

$$\epsilon_{itn}^{SP} = \zeta_{in} + \xi_{in} + \nu_{itn}^{SP} \tag{5.7}$$

where ζ_{in}, ξ_{in}, ν_{in}^{RP}, and ν_{itn}^{SP} are assumed to be mutually independent, and independent across the alternative index i (i.e., ζ_{in} and ζ_{jn} are independent for $i \neq j$, and so on)[9]. Before we turn our attention to the correlation struc-

[7] Note that the intrinsic desirability of a decision protocol may be different across RP context and SP tasks.

[8] In principle, we may allow for the decision protocol adopted at the t^{th} SP task to depend on the decision protocols adopted in earlier tasks to capture "state dependence". This approach is tractable only if we do not allow correlations in δ_{tn}^{SP} across t.

[9] Further, if there are multiple observations $\tilde{t} = 1, \ldots, \tilde{t}_n$ of actual choices in RP as in discrete panel data, we may allow for correlations between RP choices due to individual-specific effects which are present only in RP choices. Consequently, the error-component structure may be written as:

$$\epsilon_{i\tilde{t}n}^{RP} = \zeta_{in} + \xi_{in}^{RP} + \nu_{i\tilde{t}n}^{RP} \tag{5.8}$$

$$\epsilon_{itn}^{SP} = \zeta_{in} + \xi_{in}^{SP} + \nu_{itn}^{SP} \tag{5.9}$$

ture induced by the error components, the rationale for the error-component structure needs to be outlined. As in any random utility function specification ν_{in}^{RP} and ν_{itn}^{SP} form the random components of the utilities of alternatives. The individual-specific effects are categorized into:

(a) Individual-specific unobserved intrinsic preference towards alternative i in general, i.e., assumed to be exhibited both in RP and SP contexts, and is captured by ζ_{in}.

(b) Individual-specific unobserved intrinsic preference towards alternative i exhibited only in SP responses, and is captured by ξ_{in}.

Specifically, ζ_{in} captures the correlation between the RP choice and the SP responses and part of the correlation between SP responses, while ξ_{in} captures the remaining correlation between the SP responses[10]. To see this note that $\text{cov}(\epsilon_{in}^{RP}, \epsilon_{itn}^{SP}) = \sigma_{\zeta_i}^2$, $\text{cov}(\epsilon_{itn}^{SP}, \epsilon_{it'n}^{SP}) = \sigma_{\zeta_i}^2 + \sigma_{\xi_i}^2$ for $t \neq t'$ where $\sigma_{\zeta_i}^2$=var(ζ_i), and $\sigma_{\xi_i}^2$=var(ξ_i). If the choice problem has J alternatives, the error-component structure may be allowed only for $J-1$ alternatives, i.e., for the $J-1$ alternative specific constants (see Appendix E wherein we discuss this issue in the context of agent-effects models in discrete panel data, and raise the question of which alternative to set as the base alternative[11]). Ignoring decision protocol effects for the time being, if one assumes that ν_{in}^{RP} and ν_{itn}^{SP} are independently and identically distributed Gumbel (0,1) random variables, then conditional on $\zeta_n = [\zeta_{1n}, \ldots, \zeta_{J-1;n}]$ and $\xi_n = [\xi_{1n}, \ldots, \xi_{J-1;n}]$, the probability of observing $[Y_n^{RP}, Y_n^{SP}]$, where Y_n^{RP} is the choice indicator for RP data, and $Y_n^{SP} = [Y_{1n}^{SP}, Y_{2n}^{SP}, \ldots, Y_{t_n n}^{SP}]$ is the response vector for t_n SP responses, may be

[10]An alternative error-component structure is to assume:

$$\epsilon_{in}^{RP} = \nu_{in}^{RP} \tag{5.10}$$
$$\epsilon_{itn}^{SP} = \lambda_i \nu_{in}^{RP} + \xi_{in} + \nu_{itn}^{SP} \tag{5.11}$$

In this case $\text{cov}(\epsilon_{in}^{RP}, \epsilon_{itn}^{SP}) = \lambda_i \sigma_{\nu_{in}^{RP}}^2$, $\text{cov}(\epsilon_{itn}^{SP}, \epsilon_{it'n}^{SP}) = \lambda_i^2 \sigma_{\nu_{in}^{RP}}^2 + \sigma_{\xi_i}^2$ for $t \neq t'$ where $\sigma_{\nu_{in}^{RP}}^2$=var(ν_{in}^{RP}), and $\sigma_{\xi_i}^2$=var(ξ_i).

[11]The base alternative refers to the alternative with the alternative specific constant set to 0.

written as:

$$\mathrm{P}(Y_n^{RP}, Y_n^{SP} | X_n^{RP}, X_n^{SP}) = \int\int \mathrm{P}(Y_n^{RP} | X_n^{RP}, \zeta) \mathrm{P}(Y_n^{SP} | X_n^{SP}, \zeta, \xi)\, f(\zeta) g(\xi)\, d\zeta\, d\xi \tag{5.12}$$

where conditional on (ζ, ξ), $\mathrm{P}(Y_n^{RP} | X_n^{RP}, \zeta)$ is the MNL choice probability for the RP choice, while $\mathrm{P}(Y_n^{SP} | X_n^{SP}, \zeta, \xi)$ is a product of MNL choice[12] probabilities for the t_n SP responses, and $f(\cdot)$ and $g(\cdot)$ are the density functions of ζ and ξ, respectively. Numerical approaches to the estimation of the above model is limited to a small number of alternatives since it entails a $2(J-1)$-dimensional integration. Consequently, simulation approaches may be adopted wherein the likelihood is approximated by an "average" probability for the RP choice and SP responses, with the average taken over simulation draws from $f(\zeta)$ and $g(\xi)$. It must also noted that if the analyst expects the reliability in SP tasks to differ from RP data, then the variances of ν_{in}^{RP} and ν_{itn}^{SP} may be allowed to differ, i.e., $\mathrm{var}(\nu_{in}^{RP}) = \mu^2\ \mathrm{var}(\nu_{itn}^{SP})$. This is reflected in the scaling of the taste parameters in the SP model vis-a-vis the RP model, and this scale parameter μ can also be estimated.

2. Similar to the correlations between the error components in the utility functions, we may allow for correlations in the criterion functions of RP and SP model (i.e., δ^{RP} and δ^{SP} are correlated).

   A special case emerges if the decision protocols in the RP context and SP tasks are *identical*. Then the RP and SP class-specific choice models can be specified *conditional* on the *latent class*, thereby the interdependencies between RP choice and SP responses, and between SP responses are induced by their dependence on the latent class. In our example, this is equivalent to the assumption

[12] We assumed for simplicity choice-based SP tasks such that

$$\mathrm{P}(Y_n^{SP} | X_n^{SP}, \zeta, \xi) = \prod_{t=1}^{t_n} \mathrm{P}(Y_{tn}^{SP} | X_{tn}^{SP}, \zeta, \xi) \tag{5.13}$$

where Y_{tn}^{SP} is the t^{th} SP response for individual n.

that the criterion function

$$H_n = \theta' Z_{n;1} + \kappa' Z_{n;2} + \delta_n \tag{5.14}$$

is the same for both RP choice and *all* the SP responses. Consequently, the probability of observing $[Y_n^{RP}, Y_n^{SP}]$ is written as:

$$\mathrm{P}(Y_n^{RP}, Y_n^{SP} | X_n^{RP}, X_n^{SP}, Z_n) = \sum_{s=1}^{2} \mathrm{P}(Y_n^{RP}, Y_n^{SP} | X_n^{RP}, X_n^{SP}, R_s)\ Q_s(Z_n; \theta) \tag{5.15}$$

where $Q_s(Z_n; \theta)$ is the class membership model. For example, if $\delta_n \sim \mathcal{N}(0,1)$, it is represented by a probit model.

3. In the most general case, we may allow for error-component structures for utility functions and criterion functions.

5.5 Survey Data

The survey was conducted during 1987 for the Netherlands Railways to assess factors which influence the choice between rail and car for intercity travel. Data was collected in the city of Nijmegen located in the eastern side of the Netherlands near the border with Germany. This city has typical rail connections with the major cities in the western metropolitan area called the Randstad which contains Amsterdam, Rotterdam, and The Hague. Trips from Nijmegen to the Randstad takes approximately two hours by both rail and car. The sample consisted of residents of Nijmegen who:

- made a trip in the previous three months to Amsterdam, Rotterdam or The Hague;
- did not user a yearly rail pass, or other types of pass which would eliminate the marginal cost of a rail trip;
- had the possibility of using a car, namely, possessed a driver's license and had a car available in the household; and

- had the possibility of using rail, namely, did not have any very heavy baggage, were not handicapped, and did not need to visit multiple destinations.

Qualifying residents of Nijmegen were identified in a random telephone survey and requested to participate in a home interview. 235 interviews were conducted out of the 365 people who were reached by telephone and who satisfied the above criteria. The home survey consisted of three parts:

1. the characteristics of an intercity trip to the Randstad made within the previous three months (RP data);

2. SP experiments of choice between two different rail alternatives; and

3. SP experiments of choice between rail and car.

We use Rail/Car SP data in this empirical analysis. The experiment was framed in the context of the actual trip observed in the RP data and used the full-profile pairwise comparison method. The respondent was shown a pair of hypothetical rail and car alternatives at a time, each of which was described by the following attributes: travel cost, travel time, and number of transfers (only for rail). Then, the respondent was asked which mode would be chosen for the particular intercity trip reported in the RP data in terms of a five point rating scale: (1) definitely choose car; (2) probably choose car; (3) not sure; (4) probably choose rail; and (5) definitely choose rail. Each respondent was presented with several such pairs of rail and car alternatives. The order of presentation of the alternatives to the respondent was randomized to minimize the potential response bias stemming from fixed presentation format.

The data available for our analysis included the ratings elicited in SP surveys transformed into binary choices, i.e, categories (1) and (2) into car choice and categories (4) and (5) into rail choice. The variables are listed in Table 5.1.

5.6 Estimation Results

In this section we present travel mode choice models estimated on the SP data. Choice models incorporating decision protocols are compared with a standard travel mode

NAME	DESCRIPTION
PREFER	Stated Preference choice indicator $= \begin{cases} 1 & \text{Rail} \\ 0 & \text{Car} \end{cases}$
IVTT	In-vehicle travel time in hour (Rail-Car)
OVTT	Access and egress time in hour (Rail-Car)
COST	Travel cost in Guilders (Rail-Car)
TRANS	Number of transfers for Rail
WORKDUM	Trip purpose dummy $= \begin{cases} 1 & \text{work trip} \\ 0 & \text{otherwise} \end{cases}$
FEMDUM	Sex dummy $= \begin{cases} 1 & \text{female} \\ 0 & \text{otherwise} \end{cases}$
ACTCHOICE	Actual travel mode choice dummy $= \begin{cases} 1 & \text{Rail} \\ 0 & \text{otherwise} \end{cases}$

Table 5.1: Names and Definition of Variables - Decision Protocol Study

choice model such as a probit model.

Since there are multiple responses per individual, the assumption of independence among these multiple responses may not be entirely justified. Consequently, the estimated models are categorized into:

1. Models which ignore potential interdependencies among responses; and

2. Models which attempt to capture these interdependencies.

5.6.1 Models Ignoring Interdependencies Among Responses

Model 0: Binary Probit Model

This is the simplest of the estimated models. The choice is modeled by a binary probit model expressed as a function of the utility difference between the Rail and Car alternative. There may be biases in the respondent's stated preferences due to the mode actually used in the intercity trip elicited in the revealed preference data, reflecting inertia effects, justification of actual choice, or omitted attributes that are not captured by the attributes specified in the hypothetical travel mode pairs. This bias is expected to be captured through a dummy variable which indicates the actual choice (ACTCHOICE). The utility function is specified as[13]:

$$\begin{aligned} U \;=\; & \beta_1 + \beta_2 IVTT + \beta_3 OVTT + \beta_4 COST + \beta_5 TRANS + \\ & \beta_6 WORKDUM + \beta_7 FEMDUM + \beta_8 ACTCHOICE + \epsilon \end{aligned}$$

where $\epsilon \sim \mathcal{N}(0,1)$. The estimated probit model is presented in Table 5.2. The standard errors are calculated from the estimated information matrix. Further, to estimate the standard errors correctly as multiple responses from the same individual are likely to be correlated, we utilize the variance-covariance matrix for extremum estimators (Amemiya [1985]), and consequently we refer to the corrected t-statistics in conducting simple hypothesis tests. The alternative specific constant corresponding to Rail (RAILDUM) is negative and significant, while the actual mode choice dummy has a large positive and very significant coefficient. Consequently, rail users exhibit an intrinsic preference for rail, while car users exhibit an intrinsic preference for car. Further, coefficients of in-vehicle travel time, and out-of-vehicle travel times have the expected signs, although the out-of-vehicle time coefficient is insignificant. The value of in-vehicle time is 27.3 Guilders/hour, while the value of out-of-vehicle time is 20.5 Guilders/hour. The coefficients of trip purpose and gender dummies are insignificant. Further, the number of transfers on the rail alternative apparently has no effect on the mode choice process as the corresponding coefficient is insignificant.

Model 1: "Pick Car" and "Chooser"

As a first step in the estimation of a latent class choice model incorporating decision protocols, we postulate the existence of three classes associated with three decision protocols: (1) utility maximizers (*choosers*) who consider both the rail and the car alternative and presumably analyze the trade-offs among attributes; (2) individuals who always pick the car alternative (*yea-sayers*); and (3) individuals who always pick

[13] β_1 represents the intrinsic preference for rail relative to car. Note that the alternative attributes such as travel times and travel costs are in difference form, i.e., attribute of rail profile - attribute of car profile.

Parameter	Estimates	Std. err.	t-stat	t-stat corr.
RAILDUM	-0.926	0.116	-7.98	-7.80
IVTT	-0.273	0.096	-2.85	-3.00
OVTT	-0.205	0.171	-1.20	-1.29
COST	-0.010	0.002	-4.47	-4.49
TRANS	-0.011	0.063	-0.17	-0.17
WORKDUM	-0.111	0.112	-1.00	-1.00
FEMDUM	-0.063	0.083	-0.76	-0.76
ACTCHOICE	1.650	0.102	16.23	16.32

Log-likelihood at zero = -1047.34
Log-likelihood at convergence = -655.93
$\bar{\rho}^2 = 0.37$
Number of observations = 1511

Table 5.2: Model 0: Binary Probit Model

the rail alternative (*nay-sayers*)[14]. Consequently, we postulate the existence of a latent class characterized by three ordered levels with the yea-sayers and nay-sayers pitched at the extreme levels and the choosers wedged in between. The criterion function for the class membership model is specified as:

$$\begin{aligned} H &= \tilde{H} + \delta \\ &= \theta_0 + \theta_1 WORKDUM + \theta_2 FEMDUM + \theta_3 ACTCHOICE + \delta \end{aligned}$$

Assuming $\delta \sim \mathcal{N}(0,1)$ the class membership model is represented by an ordinal probit model with an additional threshold parameter.

Conditional on the latent class, the class-specific choice model is either a deterministic or probabilistic choice model depending on the class membership. The yea-sayers

[14] It must be noted that in principle, this modeling approach is identical to the choice model with latent captivity. The only distinguishing features are: (1) applications of the captivity model are usually based on RP data while herein we utilize SP data, and (2) we allow for the influences or biases stemming from actual choices, and hence the interpretation is more meaningful as a decision protocol than that of captivity.

and the nay-sayers, choose car or rail with probability 1, while the choice model for the choosers is represented by a probit model.

The estimation of such a model was attempted. The estimation routine did not converge with the threshold parameter increasing to infinity leading to the conclusion that the probability of individuals being nay-sayers becomes negligible as it decreases with each iteration of the estimation procedure. Consequently, a model with two latent classes – yea-sayers and choosers – was estimated with the class membership captured by a probit model, i.e., P(chooser) = $\Phi(\tilde{H})$ and P(yea-sayer) = $1 - \Phi(\tilde{H})$.

The estimated model is presented in Table 5.3. In the utility function which is specified as in Model 0, the estimated rail constant is negative and significant, while the coefficient for actual choice is positive and significant. The travel cost and travel time coefficients have the expected signs. Further, compared to Model 0, these coefficients are scaled up indicating greater sensitivity of the choice model to the corresponding attributes. A possible explanation is that by taking into account decision protocols we are capturing part of the random component of the utility function, which would not have been explained otherwise. For the choosers, the value of in-vehicle time is 22.2 Guilders/hour, while the value of out-of-vehicle time is 18.1 Guilders/hour. The coefficients for trip purpose and gender dummies, and the number of transfers are insignificant.

In the criterion function, as expected the actual mode choice dummy has a positive and significant coefficient indicating that the probability of an individual being a chooser is higher if the actual choice made in the intercity trip was rail. This suggests that the decision protocol adopted by an individual while responding to SP surveys may be affected by actual market behavior. Consequently, the decision protocol adopted in the RP context potentially differs from the decision protocol adopted in the SP context. Further, by specifying a criterion function which includes the actual choice dummy, the magnitude of the coefficient of the corresponding dummy in the utility function is smaller compared to that of Model 0 as some of the "inertia" effects is captured through the criterion function. Although the coefficients of trip purpose and gender dummies are similar in magnitude to the actual choice coefficient, they

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.
Utility function	RAILDUM	-0.524	0.314	-1.68	-2.17
	IVTT	-0.334	0.131	-2.55	-2.77
	OVTT	-0.271	0.211	-1.29	-1.32
	COST	-0.015	0.004	-3.45	-2.75
	TRANS	0.009	0.082	0.11	0.11
	WORKDUM	0.751	0.730	1.03	0.57
	FEMDUM	0.334	0.252	1.33	1.07
	ACTCHOICE	1.234	0.288	4.28	3.38
Criterion function	θ_0	0.377	0.583	0.65	0.84
	WORKDUM	-1.204	0.503	-2.40	-1.80
	FEMDUM	-0.761	0.351	-2.17	-1.44
	ACTCHOICE	1.632	0.338	4.83	3.07

Log-likelihood of naive model = -1047.34
Log-likelihood at convergence = -651.24
$\bar{\rho}^2 = 0.37$
Number of observations = 1511

Table 5.3: Model 1: "Pick Car" and "Chooser"

are not estimated with precision. Further, given that the constant in the criterion function is insignificant, we conclude that a car user is equally likely to be a yea-sayer or a chooser. The sample average of the latent class probability of belonging to the class "Pick Car" ("Chooser") equals 0.38 (0.62). Thus, a significant fraction of the sample belongs to the yea-sayer class.

The log-likelihood of the model is -651.24, which betters that of Model 0 by approximately 5 units given that we have 4 additional parameters.

Model 2: "Pick Actual Mode" and "Chooser"

In this model we assume that an individual may adopt one of two decision protocols, and hence, belong to one of the latent classes: (1) *actual mode adopters* who pick the actual mode in SP experiments without considering the trade-offs among attributes, and (2) *choosers*. Individuals who adopt the first protocol are either "captive" to the actual mode, or are uninterested in the SP tasks, and consequently indicate apathy and laziness, or attempt to "justify" the actual mode choice while responding to SP tasks. The utility function for choosers is specified as in Model 0, while the criterion function is the same as in Model 1.

The estimated model is presented in Table 5.4. In the utility function, the rail constant is positive but insignificant. The coefficient of actual choice is insignificant. Thus, as expected by taking into account "inertia" at the decision protocol level, this effect is negligible at the choice level. The travel cost and travel time coefficients have the expected signs. In most empirical work in the context of travel mode choice models, the ratio of the out-of-vehicle time coefficient to the in-vehicle time coefficient is greater than 1, indicating greater marginal disutility of out-of-vehicle travel time relative to in-vehicle travel time. This relationship is reflected in this model. For the choosers, the value of in-vehicle time is 37.7 Guilders/hour, while the value of out-of-vehicle time is 47.0 Guilders/hour. Thus the values of time are higher than those of Model 1. It appears that the actual mode adopters on average have lower values of times. The coefficients for trip purpose and gender dummies, and the number of transfers are insignificant.

In the criterion function, the coefficient of work dummy is insignificant, while the coefficient of gender dummy is negative and significant. Thus, women are more likely to pick the actual mode in SP experiments compared to men. The coefficient of actual choice is positive indicating that rail users are more likely to be choosers than are car users. In other words, car users are more likely to pick the car alternative reflecting inertia effects. The sample average of the latent class probability of belonging to the class "Pick actual mode" ("Chooser") equals 0.71 (0.29).

The log-likelihood of the model is -653.25, which marginally improves on Model 0 by 3 units, and is worse compared to Model 1 by 2 units.

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.
Utility function	RAILDUM	1.552	1.020	1.52	0.88
	IVTT	-0.943	0.470	-2.00	-1.25
	OVTT	-1.176	0.853	-1.38	-0.84
	COST	-0.025	0.008	-3.25	-3.05
	TRANS	0.285	0.255	1.12	0.67
	WORKDUM	-0.300	0.352	-0.85	-0.79
	FEMDUM	0.200	0.268	0.75	0.65
	ACTCHOICE	-1.227	0.976	-1.26	-0.73
Criterion function	θ_0	-0.919	0.178	-5.16	-3.17
	WORKDUM	-0.167	0.149	-1.12	-1.04
	FEMDUM	-0.221	0.099	-2.23	-2.23
	ACTCHOICE	1.185	0.187	6.32	4.89

Log-likelihood of naive model = -1047.34
Log-likelihood at convergence = -653.25
$\bar{\rho}^2 = 0.37$
Number of observations = 1511

Table 5.4: Model 2: "Pick Actual Mode" and "Chooser"

5.6.2 Models Allowing Interdependencies Among Responses

Model 3: Model 0 with Serial Correlation

In this model, we adopt an agent-effects specification for the utility function. The utility function is similar to Model 0, i.e.,

$$\begin{aligned} U_{tn} &= \beta_1 + \beta_2 IVTT_{tn} + \beta_3 OVTT_{tn} + \beta_4 COST_{tn} + \beta_5 TRANS_{tn} + \\ &\quad \beta_6 WORKDUM_{tn} + \beta_7 FEMDUM_{tn} + \beta_8 ACTCHOICE_{tn} + \epsilon_{tn} + \tilde{\epsilon}_n \end{aligned}$$

where t denotes the t^{th} response for individual n who is presented with t_n profiles (i.e., $t \in \{1, \ldots, t_n\}$), $\tilde{\epsilon}_n$ is the individual-specific random component which persists across responses from the same individual, and ϵ_{tn} is a pure random component. Assuming $\epsilon_{tn} \sim \mathcal{N}(0,1)$, conditional on $\tilde{\epsilon}_n$, the choice model is a probit model. Noting that conditional on $\tilde{\epsilon}_n$ the multiple responses for the same individual are independent, the probability of observing the response vector $Y_n = [Y_{1n}, \ldots, Y_{t_n n}]$, $P(Y_n|X_n;\beta,\sigma)$, is written as

$$\int \left\{ \prod_{t=1}^{t_n} P(Y_{tn}|X_{tn}, \tilde{\epsilon}; \beta) \right\} f(\tilde{\epsilon};\sigma)\, d\tilde{\epsilon} \tag{5.16}$$

where Y_{tn} and X_{tn} denote the choice indicator and the attributes of alternatives for the t^{th} choice pair, respectively, and $f(\cdot)$ is the density function of the normal random variable with mean zero and variance σ^2.

The estimated model is presented in Table 5.5. As in Model 0, the coefficients of number of transfers, gender and trip purpose are insignificant, while the coefficient of actual choice is significant. The value of in-vehicle time is 20.1 Guilders/hour, while the value of out-of-vehicle time is 10.5 Guilders/hour. The significant standard deviation of the agent-effect error component indicates substantial correlation among responses.

The log-likelihood of the model is -632.88, which improves on Model 0 by 23 units.

Parameter	Estimates	Std. err.	t-stat
RAILDUM	-1.004	0.264	-3.80
IVTT	-0.523	0.196	-2.67
OVTT	-0.274	0.373	-0.73
COST	-0.026	0.005	-5.02
TRANS	-0.042	0.145	-0.29
WORKDUM	-0.345	0.260	-1.33
FEMDUM	-0.120	0.189	-0.63
ACTCHOICE	2.267	0.252	9.00
σ	1.084	0.111	9.78

Log-likelihood at zero = -1047.34
Log-likelihood at convergence = -632.88
$\bar{\rho}^2 = 0.39$
Number of observations = 226

Table 5.5: Model 3: Binary Probit Model with Serial Correlation

Model 4: "Pick Car" and "Chooser" with interdependencies among responses: Identical decision protocol

As seen in Model 1 and Model 2, the improvement in the overall model fit by incorporating variations in decision protocols is apparently limited. Recognizing that the individual may adopt the *same* decision for all the SP tasks, we estimate a model which captures such interdependencies among responses.

Herein the decision protocols are specified as in Model 1: "Pick car" and "Chooser". Assuming that conditional on class membership the responses are independent, $P(Y_n|X_n, Z_n; \beta, \theta)$ is written as

$$\sum_{s=1}^{2} \left\{ \prod_{t=1}^{t_n} P(Y_{tn}|X_{tn}; R_s) \right\} Q_s(Z_n; \theta) \tag{5.17}$$

where R_s denotes the decision protocol for an individual in class $s = 1, 2$, and Z_n form the causal variables affecting class membership (i.e., variables in $\tilde{H}_n$).

The utility function for choosers and the criterion function are same as in Model 1. Table 5.6 presents the estimation results. In the utility function the rail constant is negative and significant, while the coefficient for actual choice is positive and significant. Compared to Model 1, the coefficient of in-vehicle travel time is scaled higher. For the choosers, the value of in-vehicle time is 48.9 Guilders/hour, while the value of out-of-vehicle time is 9.3 Guilders/hour. Hence compared to Model 1, the value of in-vehicle time for choosers increased considerably, while the value of out-of-vehicle time decreased. The coefficients for trip purpose and gender dummies, and the number of transfers are insignificant.

In the criterion function, as expected the actual choice dummy has a positive and significant coefficient. The coefficients of trip purpose and gender dummies are smaller in magnitude compared to those of Model 1, and insignificant too. Further, in contrast to Model 1, the constant is insignificant. This coupled with the positive coefficient for actual choice suggests that a car user is more likely to be a yea-sayer, while a rail user is more likely to be a chooser. The sample average of the latent class probability of belonging to the class "Pick car" ("Chooser") equals 0.35 (0.65). Thus, we notice a marginal decrease in the average probability of belonging to the class 'Pick car" compared to Model 1. This is expected since in this model, for an individual to belong to the class "Pick car", he/she must pick the alternative in *all* the responses.

The log-likelihood of the model is -602.49, which improves on that of Model 1 by 49 units with no additional parameter.

Model 5: "Pick Car" and "Chooser" with interdependencies among responses: Agent-effects specification

Herein we allow the decision protocols to vary across the responses. But we capture interdependencies among the responses since the protocols are expected to be *similar*. Specifically, interdependencies is operationalized through an agent-effects specification for the criterion function, i.e., decomposing the random component of the criterion function into two components – a component which is independent across individuals and responses, and another component which is individual-specific and

	Parameter	Estimates	Std. err.	t-stat
Utility function	RAILDUM	-0.473	0.148	-3.20
	IVTT	-0.489	0.126	-3.86
	OVTT	-0.093	0.228	-0.41
	COST	-0.010	0.003	-3.37
	TRANS	0.053	0.079	0.67
	WORKDUM	0.014	0.132	0.11
	FEMDUM	0.082	0.111	0.74
	ACTCHOICE	1.259	0.129	9.76
Criterion function	θ_0	0.178	0.168	1.06
	WORKDUM	-0.204	0.320	-0.64
	FEMDUM	-0.367	0.222	-1.66
	ACTCHOICE	1.565	0.275	5.69

Log-likelihood of naive model = -1047.34
Log-likelihood at convergence = -602.49
$\bar{\rho}^2 = 0.41$
Number of observations = 226

Table 5.6: Model 4: "Pick Car" and "Chooser" with interdependencies among responses: Identical decision protocol

hence persists across responses[15]. We allow for the individual-specific component to vary randomly in the population with a parameterized distribution. Consequently, the criterion function for the class membership model is written as:

$$\begin{aligned} H_{tn} &= \tilde{H}_{tn} + \delta_{tn} + \tilde{\delta}_n \\ &= \theta_0 + \theta_1 WORKDUM + \theta_2 FEMDUM + \theta_3 ACTCHOICE + \delta_{tn} + \tilde{\delta}_n \end{aligned}$$

Conditional on $\tilde{\delta}_n$, and assuming $\delta_{tn} \sim \mathcal{N}(0,1)$ the class membership model is given by a probit model, i.e., P(chooser) = $\Phi(\tilde{H} + \tilde{\delta}_n)$ and P(yea-sayer) = $1 - \Phi(\tilde{H} + \tilde{\delta}_n)$. Then, by assuming $\tilde{\delta}_n \sim \mathcal{N}(0, \sigma^2)$ the probability of observing the response vector $Y_n = [Y_{1n}, \ldots, Y_{t_n n}]$, $P(Y_n | X_n, Z_n; \beta, \theta, \sigma)$, equals

$$\int \prod_{t=1}^{t_n} \left\{ \sum_{s=1}^{2} P(Y_{tn} | X_{tn}; R_s)\, Q_s(Z_n, \tilde{\delta}; \theta) \right\} f(\tilde{\delta}; \sigma)\, d\tilde{\delta} \tag{5.18}$$

where Y_{tn}, X_{tn}, and Z_n are as defined earlier, and $f(\cdot)$ is the density function of the normal random variable with mean zero and variance σ^2.

Table 5.7 presents the estimation results. In the utility function the rail constant is positive though insignificant, while the coefficient for actual choice is positive and significant. Compared to Model 1 and Model 4, these coefficients are scaled even higher indicating greater sensitivity of the choice model to the corresponding attributes. For the choosers, the value of in-vehicle time is 18.4 Guilders/hour, while the value of out-of-vehicle time is 14.3 Guilders/hour. Hence compared to Model 1 and Model 4, the estimated values of times for choosers decreased.

In the criterion function, as expected the actual choice dummy has a positive and significant coefficient. The estimated coefficients of trip purpose and gender dummies are smaller in magnitude compared to those of Model 1, and insignificant too. Further, in contrast to Model 1 and Model 3 the constant in the criterion

[15] In principle, we can allow agent-effects specification for the criterion function utility. We attempted such an approach in this model and in a model which discussed later. On both occasions, either the estimation procedure did not converge with some of the parameters tending to infinity or converged to a point at which the curvature of the log-likelihood function did not exist depending on the starting values.

	Parameter	Estimates	Std. err.	t-stat
Utility function	RAILDUM	1.310	0.993	1.32
	IVTT	-0.664	0.305	-2.18
	OVTT	-0.516	0.586	-0.88
	COST	-0.036	0.014	-2.56
	TRANS	6×10^{-4}	0.233	3×10^{-3}
	WORKDUM	-0.276	0.521	-0.53
	FEMDUM	-0.429	0.581	-0.74
	ACTCHOICE	1.672	0.422	3.96
Criterion function	θ_0	-1.246	0.296	-4.22
	WORKDUM	-0.050	0.437	-0.12
	FEMDUM	0.011	0.408	0.03
	ACTCHOICE	1.905	0.362	5.26
	σ	1.309	0.233	5.62

Log-likelihood of naive model = -1047.34
Log-likelihood at convergence = -571.10
$\bar{\rho}^2 = 0.44$
Number of observations = 226

Table 5.7: Model 5: "Pick Car" and "Chooser" with interdependencies among responses: Agent-effects specification

function is negative and significant. Further, the standard deviation of the individual-specific error component in the criterion function is significant. Given that this error component may be interpreted as a "random" coefficient for the intercept in the criterion function, we notice that once we take into account this randomness, the mean of the intercept is significant. The sample average of the latent class probability of belonging to the class "Pick car" ("Chooser") equals 0.63 (0.37).

The log-likelihood of the model is -571.24, which improves on that of Model 1 by 80 units, while that of Model 3 by 31 units.

Model 6: "Pick Actual Mode" and "Chooser" with interdependencies among responses: Identical decision protocol

Herein the decision protocols are specified as in Model 2: "Pick actual mode" and "Chooser". Further, as in Model 4, we assume that the individual adopts the *same* decision protocol while responding to all the SP tasks.

Table 5.8 presents the estimation results. In the utility function the rail constant is negative and significant, while the coefficient for actual mode choice dummy is positive and significant. Further, compared to Model 2, the coefficients are scaled down. For the choosers, the value of in-vehicle time is 51.6 Guilders/hour, while the value of out-of-vehicle time is 50.2 Guilders/hour. Hence compared to Model 2, values of time increased considerably. The coefficients for trip purpose and gender dummies, and the number of transfers are insignificant. Unlike Model 2, the coefficient of actual choice is positive indicating inertia effect in the utility function. This is expected since in this model, for an individual to possibly belong to the class "Pick Actual Mode", he/she must pick the actual mode in *all* the responses. Consequently, the effect of actual choice appears to be present both at the decision protocol level and the choice level.

In the criterion function, the estimated coefficients of trip purpose and gender dummies insignificant. Further, in contrast to Model 2, the constant is insignificant. Consequently, an individual tends to have a higher probability of being a chooser. This is manifested in the sample average of the latent class probability of belonging to the class "Pick actual mode" ("Chooser") which equals 0.38 (0.62), with the decrease in the share of the class "Pick actual mode" compared to that of Model 2.

The log-likelihood of the model is -606.81, which improves on that of Model 2 by 48 units with no additional parameter.

	Parameter	Estimates	Std. err.	t-stat
Utility function	RAILDUM	-0.414	0.149	-2.77
	IVTT	-0.310	0.130	-2.39
	OVTT	-0.301	0.237	-1.27
	COST	-0.006	0.003	-2.10
	TRANS	0.048	0.091	0.49
	WORKDUM	0.037	0.140	0.27
	FEMDUM	0.037	0.113	0.33
	ACTCHOICE	0.753	0.142	5.30
Criterion function	θ_0	0.111	0.155	0.72
	WORKDUM	-0.008	0.230	-0.04
	FEMDUM	-0.290	0.203	-1.43
	ACTCHOICE	1.088	0.228	4.77

Log-likelihood of naive model = -1047.34
Log-likelihood at convergence = -606.81
$\bar{\rho}^2 = 0.41$
Number of observations = 226

Table 5.8: Model 6: "Pick Actual Mode" and "Chooser" with interdependencies among responses: Identical decision protocol

Model 7: "Pick Actual Mode" and "Chooser" with interdependencies among responses: Agent-effects specification

Herein the decision protocols are specified as in Model 2. Further, as in Model 5, we allow the same individual to vary the decision protocols across responses.

Table 5.9 presents the estimation results for the model with agent-effects specification for the criterion function. In the utility function all the coefficients are insignificant except for the coefficient of travel cost and actual choice. Surprisingly, the coefficient of actual choice has a counter-intuitive sign. For the choosers, the value of in-vehicle time is 16.6 Guilders/hour, while the value of out-of-vehicle time is 33.5 Guilders/hour. Hence compared to Model 2, the value of in-vehicle time decreased. In the criterion function, only the constant and coefficient of actual choice are significant.

The log-likelihood of the model is -574.54, which improves on that of Model 2 by 84 units given that we have only 1 additional parameter. Although the model fits better than Model 6, the counter-intuitive sign for the actual choice coefficient in the utility function suggests that the model may be misspecified. Therefore, we reject this model.

5.6.3 Summary of Estimated Models

In Table 5.10, we summarize the log-likelihood value, the number of estimated parameters, the Akaike Information Criterion (AIC), $\bar{\rho}^2$, and the estimated values of time[16] (VOT) for each of the estimated models. In general, the models which incorporate interdependencies through a agent-effects specification fit the data the best. On the other hand, models which assume *same* decision protocol across responses tend to fit the data better than the binary probit (Model 0), and models which ignore these interdependencies (Model 1 and Model 2), and have more significant coefficients of important attributes such as travel time and travel cost.

[16]Comparison of values of time across models with different structures or decision protocols is not entirely appropriate since the value of time refers to the particular class of *choosers*.

	Parameter	Estimates	Std. err.	t-stat
Utility function	RAILDUM	1.399	0.768	1.82
	IVTT	-0.565	0.549	-1.03
	OVTT	-1.140	1.086	-1.05
	COST	-0.034	0.012	-2.84
	TRANS	0.579	0.403	1.44
	WORKDUM	-0.452	0.674	-0.67
	FEMDUM	-0.107	0.469	-0.23
	ACTCHOICE	-1.755	0.646	-2.72
Criterion function	θ_0	-1.225	0.213	-5.75
	WORKDUM	-0.119	0.306	-0.39
	FEMDUM	-0.299	0.231	-1.29
	ACTCHOICE	1.384	0.311	4.46
	σ	1.196	0.154	7.79

Log-likelihood of naive model = -1047.34
Log-likelihood at convergence = -574.54
$\bar{\rho}^2 = 0.44$
Number of observations = 226

Table 5.9: Model 7: "Pick Actual Mode" and "Chooser" with interdependencies among responses: Agent-effects specification

We can conduct the Hausman's specification test for comparing Model 1 and Model 4, where the estimates in Model 1 are consistent and inefficient, while the estimates in Model 4 are consistent and efficient under the null hypothesis of no mis-specification. Since the test statistic is 1.92, and the critical value at 5% significance for 12 degrees of freedom is 21.03, we accept Model 4.

In similar vein, we can compare Model 2 and Model 6. Since the test statistic is 41.17, and the critical value at 5% significance for 12 degrees of freedom is 21.03, we can reject Model 6.

It must be noted that the models presented here provide only a preliminary assessment of the potential for capturing decision protocol heterogeneity. In the data the traveler's characteristics which are postulated to guide the "choice" of the decision protocol are limited to gender and trip purpose, and consequently the criterion function does not capture the effects of the gamut of individual time and budget constraints, household constraints, etc. Further empirical work with other surveys and in problem domains is necessary before such tools can be meaningfully adopted in practice.

An empirical caveat in the estimation of such models must be noted. The parameter estimates tend to be "sensitive" or "non-robust" in the sense that inclusion or exclusion of variables in the criterion function tends to change the choice model parameters appreciably. Further empirical work is needed to assess the differential impacts of including individual characteristics in the criterion function and the utility function, and their substantive significance and interpretation.

5.7 Summary

In this chapter, we highlighted the need to capture variations in decision protocols, provided empirical evidence to substantiate their significance, and discussed possible determinants of such variations.

We outlined the latent class choice model for decision protocol heterogeneity. Since decision protocols in RP and SP settings may differ for the *same individual*, we

Model type	Model	Log-lik.	# of par.	Akaike	$\bar{\rho}^2$	VOT (Guilder/hr)	
						in-veh.	out-of-veh.
No interdep. responses	MODEL0	-655.93	8	-663.93	0.37	27.3	20.5
	MODEL1	-651.24	12	-663.24	0.37	22.2	18.1
	MODEL2	-653.25	12	-665.25	0.37	37.7	47.0
With interdep. responses	MODEL3	-632.88	9	-641.88	0.39	20.1	10.5
	MODEL4	-602.49	12	-614.49	0.41	48.9	9.3
	MODEL5	-571.10	13	-584.10	0.44	18.4	14.3
	MODEL6	-606.81	12	-618.81	0.41	51.6	50.2
	MODEL7	-574.54	13	-586.54	0.44	16.6	33.5

Table 5.10: Comparison of auxiliary statistics of estimated models: Decision Protocol Study

discussed the need to combine RP and SP data, and outlined an approach to validate the decision protocols exhibited in SP analysis. We applied the model to assess its potential in the context of simulated travel mode choice experiments.

Chapter 6

Latent Structure Choice Models

6.1 Introduction

In chapter 3 we developed the latent class choice model (LCCM) wherein the latent constructs are discrete or categorical. Unlike the latent structure model reviewed in section 2.3, wherein the latent constructs are *manifested* through a set of indicators, we did not specify explicit indicators for the classes in LCCM. Rather, only the choice indicator was utilized as an *indirect* indicator of the latent class.

To incorporate indicators of latent classes, in section 6.2 we elaborate on different specifications of the *measurement model*, which maps from the classes to the indicators, depending on the characterization of the latent class. In section 6.3 we link the measurement model with the class membership model developed in chapter 3 to obtain the latent class model (LCM). As reviewed in section 2.3, the traditional latent class model primarily links *discrete* latent constructs and *discrete* indicators. In this chapter, LCM encompasses cases wherein the indicators may be discrete and/or continuous. Specifically, the LCM maps from a set of explanatory variables to a set of indicators through intermediate constructs represented by latent classes. Subsequently, in section 6.4 we extend the LCCM to include latent class indicators.

It is instructive as this point to motivate the need for utilizing responses to attitudinal and perceptual questions as indicators of latent constructs such as attitudes, perceptions and latent classes. The reasons are primarily two-fold:

1. *Identification*: To characterize latent variables such as attitudes and perceptions it is imperative that we have attitudinal and perceptual indicators for the identification of the model parameters. It must be noted that the indicators are necessary only in the model estimation stage and not in model applications, and this notion will be apparent later in this chapter.

   An important issue in addition to the theoretical issue of model identification is that of empirical identification. The information content from only the choice indicator may not be sufficient to empirically identify the latent constructs such as choice set considered, taste variations, etc. To this end, the information from the indicators are conjectured to aid in resolving such empirical identification problems, if any.

2. *Efficiency*: We note that indicators contain important information about the latent constructs, and we highlight this theme with two examples.

   Consider a situation wherein the latent class characterizes the choice set considered and the indicators correspond to responses to alternative availability questions gathered on a Likert-type rating scale, say 1-5 where 1 represents the unavailability of the alternative while 5 represents its availability. Note that such questions should be restricted to the non-chosen alternatives since the chosen alternative is available. Thus, in principle, we can infer that an alternative with a higher availability rating is more likely to be considered than an alternative with a lower availability rating. Further, in the absence of any response biases, an individual with a higher availability rating for an alternative is more likely to consider it than another individual with a lower availability rating for the same alternative. Consequently, such ratings enhance the information content available to the analyst, thus making them legitimate candidates for the indicators of the latent class.

   Consider a situation wherein the latent class characterizes individual's sensitivity to attributes. There is clearly no doubt that the importance ratings of attributes contain information regarding the sensitivity to attributes, since we can infer that an individual with higher importance rating for an attribute is more sensitive to that attribute than an individual with lower importance rating.

Consequently, noting that the indicators of latent constructs have information content, in addition to the choice indicator, we can (potentially) gain efficiency in model estimation.

In section 6.5 we develop the latent structure choice model (LSCM) which incorporates the gamut of attitudinal, perceptual and class indicators through latent attitudes, perceptions and classes, and discuss issues of estimation. Operationally, the LSCM links latent structure models, including latent variable models and the latent class model, with choice models.

6.2 Latent Class Indicators: Measurement Model

Notation for the Latent Class Model

D = dimension of the class membership vector.

T_n = D-dimensional random vector which denotes class membership of individual n, i.e., $T_n = [l_1, \ldots, l_D]'$ where l_d is the level in dimension d.[1]

s = latent class index with $s = 1, \ldots, S$, where S number of latent classes.

$$l^*_{sn} = \begin{cases} 1 & \text{if individual } n \text{ is in latent class } s \\ 0 & \text{otherwise}^2 \end{cases}$$

A_n = $P \times 1$ vector of indicators of the latent class.

To keep the measurement models in perspective, we provide illustrative examples with special attention given to the latent class characterizations considered in chapter 3 (i.e., taste variations, choice set and decision protocols) and the indicators in

[1] l_d may be a binary variable or an ordered categorical variable.

each example. We also outline two additional examples of latent class and associated indicators. Noting that the latent class characterizes a multi-dimensional construct, we categorize the latent class indicators into:

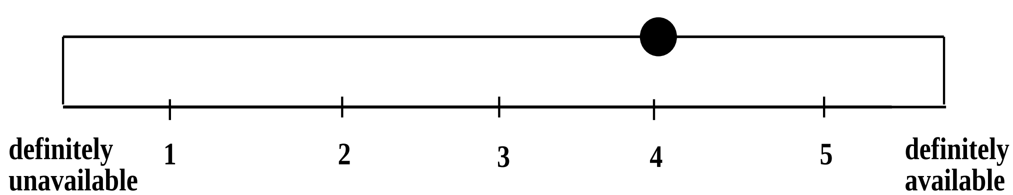
1. *Class-specific indicators*: Herein *all* the indicators are associated with each latent class as such, with the distribution of the indicators specified conditional on the latent class. This is the case when the latent class is categorical such as decision protocol.

2. *Dimension-specific indicators*: Herein one or more indicators are associated with the dimensions of the latent class vector. Further, the distribution of these dimension-specific indicators is specified conditional on the level of the corresponding dimension.

Naturally, one may allow for some indicators to be associated with a subset of the dimensions (similar to the class-specific case), while the others may be dimension-specific. The above categorization will be transparent in the following examples.

1. *Latent Choice Set Example*: Consider a situation where the latent class represents the choice set considered. The *indicators* of the latent choice set may include responses to questions such as:

   `Would you consider alternative` j `as being available to you?`

   Specifically, this alternative availability rating is an indicator of the perceived (unobserved) availability of the alternative, and thus the indicators are dimension-specific.

2. *Taste Variations Example*[3]: Consider a situation wherein a respondent rates the importance of attributes of alternatives. These importance ratings may be postulated to be *indicators* of the individual's sensitivity to attributes. Further, interrelationships among the individual's sensitivity to attributes may exist, and consequently, the individual's sensitivity to attributes may be generated by a smaller set of sensitivity dimensions. For example, consider a shipper's freight transportation mode choice situation, and the shipper's importance ratings on the following service attributes: transit time, transit time reliability, rate, payment terms and billing, loss and damage, usability of equipment, and responsiveness, are available. We may postulate that there exists three shipper's attitude dimensions – time-sensitivity, cost-sensitivity, and service-quality sensitivity – with two levels in time-sensitivity and service-quality dimensions, and three levels in the cost-sensitivity dimension. Further, the importance ratings of transit time, and reliability of transit time may be utilized as indicators of time-sensitivity, those of rate, payment terms and billing, and loss and damage as indicators of cost sensitivity, and those of usability of equipment, responsiveness, and level of effort required to deal with the carrier as indicators of service quality sensitivity. In this case, the indicators are dimension-specific, with a set of indicators associated with each dimension of the latent class.

3. *Decision Protocols*: In the case of latent class characterizing the decision protocol adopted, the class indicators are less obvious. As reviewed in section 5.2, studies in behavioral decision research (see, for example, Nisbett and Wilson [1977]) note that differences in verbal protocol reports may reflect variations in decision protocols. Consequently, these reports may be utilized as indicators if the reports can be appropriately and reliably "coded" on some measurement scale to capture the *degree of proximity* of a particular individual's verbal report to each of the decision protocols. Svenson [1974] and Thorngate and Maki [1977] report high interjudge agreement for trained coders. It must be noted though

[3]This example is taken from the work of the Vieira [1992], and is studied in more detail in chapter 7.

that such verbal reports are not collected in transportation surveys. In this case, the indicators may be class-specific, since the "proximity" ratings may be utilized as indicators of the categorical latent class.

Another potential indicator in the context of stated preference experiments is the time taken to respond to SP tasks. This indicators is intuitively appealing since an individual who takes more time to respond is more likely to adopt a more cognitively demanding decision protocol compared to an individual who takes less time.

Two additional examples of latent class and the associated indicators include:

1. *Latent Choice Example*: In stated preference experiments the respondent may be provided with a scenario wherein the attributes of the alternatives are set at different levels, and may be asked to rate each alternative on a Likert-type scale to provide preferences toward the different alternatives. A simpler preference gathering scheme, referred to as pairwise rating task, is to provide each respondent with two alternatives at a time and request a preference rating to gather the relative preferential information between the two alternatives. In such experiments the choice is unobserved, and hence may be characterized by a latent class. The ratings are postulated to be *indicators* of the latent choice (see Gopinath and Ben-Akiva [1993]). In this case, the indicators are class-specific.

2. *Latent Lifestyle Example*: In travel demand modeling, the concept of lifestyle is hypothesized to capture long term decisions of individuals and households which guide their preferred pattern of mobility, activity and travel choices, and is expected to be a key "higher" level factor substituting for traditional social class and economic status variables (Salomon [1980], Salomon and Ben-Akiva [1983]). Since the lifestyle concept is unobserved, it may be captured through latent classes. The *indicators* may include responses to questions about themes such as:

(a) orientation towards work, leisure and activities pursued;

(b) attitudes towards family formation, interactions between male and female heads of the family, etc.;

(c) (dis)like of tele-options which may substitute travel;

(d) importance of the attributes of travel and *no travel* related choices;

(e) environmental concerns, philosophical proclivities, etc.; and

(f) political affinities and attitudes.

Given an overview of the possible types of class indicators, we turn our attention to the measurement model specifications.

6.2.1 Class-specific Measurement Model

Let $a_{pn},\ \forall p = 1, \ldots, P$, denote the p^{th} indicator for individual n, where P is the number of indicators. Let $A_n = [a_{1n}, \ldots, a_{Pn}]'$ and $g_{ps}(a_{pn}; \phi_{ps})$ denote the conditional density of indicator a_p given latent class equals s with parameters ϕ_{ps}. Assuming that the indicators given the latent class are independent which is an assumption usually referred to as *conditional independence* (see section 2.3), the conditional density function of the indicators is written as

$$g(A_n | l^*_{sn} = 1; \phi_s) = \prod_{p=1}^{P} g_{ps}(a_{pn}; \phi_{ps}). \tag{6.1}$$

where $\phi_s = [\phi_{1s}, \ldots, \phi_{Ps}]$. It can be observed that if each indicator's conditional density $g_{ps}(a_{pn}; \phi_{ps})$ comes from the exponential family, such as a normal density, the complexity of the conditional density of the indicators reduces significantly.

We illustrate the measurement model for the latent choice example in Figure 6-1 for the case of pairwise stated preference rating task. Intuitively, if the individual would choose alternative i (j) then, the analyst would expect a lower (higher) individual's rating on the 1-9 point scale. Consequently, if the alternative that would be chosen is i (j), then the parametric distribution of the rating is given by the one

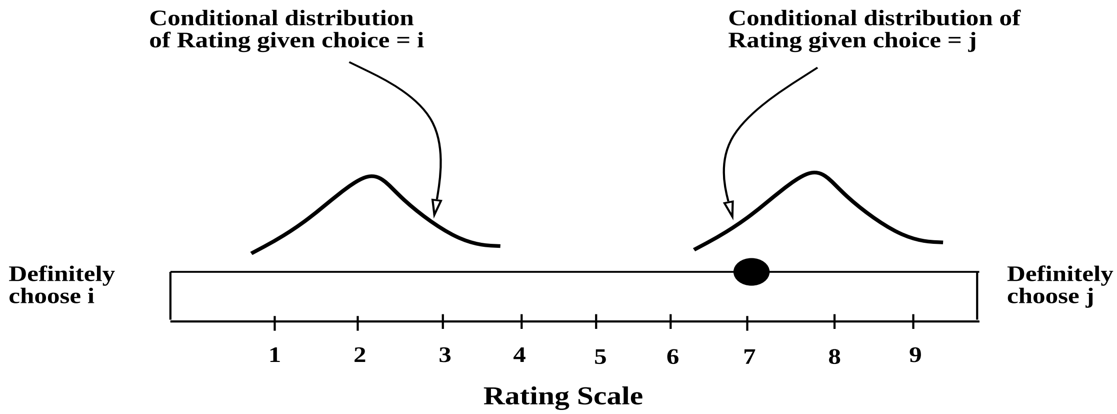


Figure 6-1: Illustration of Measurement Model for Latent Choice Example

illustrated on the left (right) side of Figure 6-1. If one assumes that the ratings are ordered categorical variables, then the conditional distributions of the ratings should be represented by probability mass functions. It must be noted that such a probability mass function may be represented through an ordinal probability model.

6.2.2 Dimension-specific Measurement Model

Let $A_n^d = [a_{1n}^d, \ldots, a_{P_d n}^d]'$ denote the P_d indicators for dimension d of the latent class. Let $g_{pdl}(a_{pn}^d; \phi_{pdl})$, $\forall p = 1, \ldots, P_d$, denote the conditional distribution of indicator a_{pn}^d given that the level of the latent class along dimension d, i.e., T_{dn}, equals l with $l \in \{1, \ldots, L_d\}$ where L_d is the number of levels in dimension d. Assuming that these indicators are independent given $T_{dn} = l$, the conditional density function of A_n^d is written as

$$g(A_n^d | T_{dn} = l; \phi_{dl}) = \prod_{p=1}^{P_d} g_{pdl}(a_{pn}^d; \phi_{pdl}). \tag{6.2}$$

where $\phi_{dl} = [\phi_{1dl}, \ldots, \phi_{P_d dl}]$. Further, assuming that the indicators A_n^d are independent across dimensions given the levels in each dimension, the conditional density of

$A_n = [A_n^{1'}, \ldots, A_n^{D'}]'$ is given as:

$$g(A_n|T_n = [l_1, \ldots, l_D]'; \phi_{1l_1}, \ldots, \phi_{Dl_D}) = \prod_{d=1}^{D} \prod_{p=1}^{P_d} g_{pdl_d}(a_{pn}^d; \phi_{pdl_d}). \quad (6.3)$$

This measurement model specification may be utilized if the analyst identifies specific indicators associated with each of the latent dimensions[4]. In most situations, at least at the stage of confirmatory analysis, the analyst has sufficient information from exploratory data analysis to "lend" character to each latent dimension through its "attributes" which are the indicators associated with that latent dimension. As noted earlier, some of the indicators may be discrete or ordered categorical with a conditional probability mass function.

To highlight this specification of the measurement model, we go back to our taste variations example schematized in Figure 6-2. The distributions of the indicators of cost-sensitivity are specified given the cost-sensitivity level. Similarly, the distributions of the indicators of time and service-quality sensitivity are specified given their corresponding levels.

Assuming that the indicators are *continuous* we can represent the above specifications of the measurement model as *linear measurement models* as in the LISREL model system, with some additional notation. Consider the measurement model wherein the indicators are class-specific, i.e., they are specified conditional on $l_{sn}^* = 1$. Denote the S-dimensional binary vector $L_n^* = [l_{1n}^* = 0, \ldots, l_{sn}^* = 1, \ldots, l_{Sn} = 0]'$ with only one of the components (here the s^{th} component) taking the value 1[5]. The measurement model is written as:

$$A_n = \Lambda L_n^* + \epsilon_n \quad (6.4)$$

[4]If an indicator a_{pn} is postulated to reflect several dimensions, say d and d', then the conditional density of a_{pn} may be specified given the corresponding levels l_d and $l_{d'}$. In this vein, the specification is analogous to the class-specific measurement model though we restrict our attention to a subset of the dimensions, which may referred to as a *sub-class*.

[5]Note that $L_n^* = T_n$, and the additional notation L_n^* is to parallel the one for the dimension-specific measurement model.

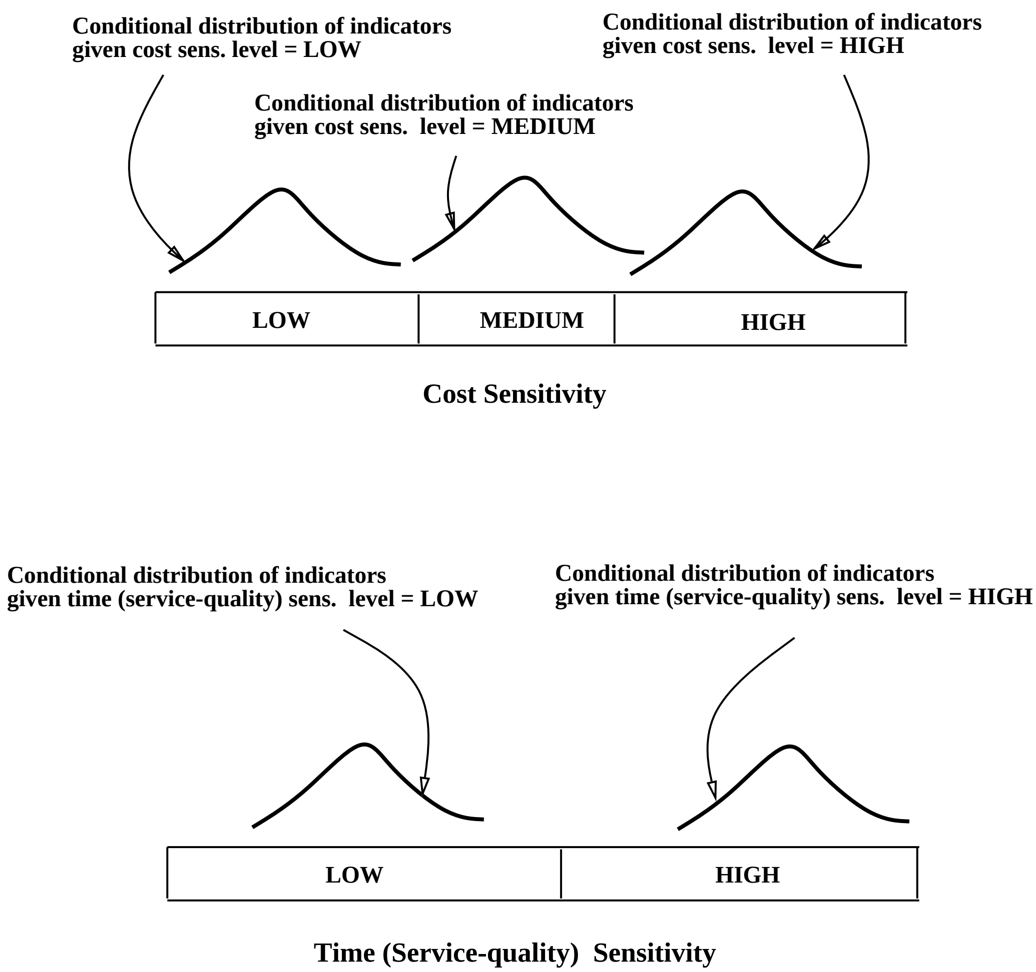


Figure 6-2: Illustration of Measurement Model for the Taste Variations Example

where Λ is a $P \times S$ parameter matrix, and ϵ_n is a $P \times 1$ random vector. Unlike the measurement model in the LISREL model wherein an element of each column of Λ is set to 1, no such scaling restrictions are necessary. This specification allows for "shifts" in the conditional distributions of the indicators, with the s^{th} column of Λ corresponding to the conditional means of the indicators given that latent class equals s. Herein, we assume that the conditional variances are invariant across classes. The assumption of conditional independence is equivalent to the assumption of independence of the components of ϵ_n.

In a similar manner, we may construct the dimension-specific measurement model. As a first step, we restrict our attention to the measurement model for the P_d indicators of dimension d. Let L^*_{dn} denote the L_d-dimensional binary vector wherein the l_d^{th} component takes the value 1 if the latent class level in dimension d equals l_d and 0 otherwise. Thus the measurement model for the P_d indicators may be written as:

$$A_n^d = \Lambda_d L^*_{dn} + \epsilon_{dn} \tag{6.5}$$

where Λ is a $P_d \times L_d$ parameter matrix. As before the l_d^{th} column of Λ_d corresponds to the conditional means of the P_d indicators given the the latent class in dimension d is in level l_d. The complete measurement model may be written in a compact form as:

$$A_n = \Lambda L^*_n + \epsilon_n \tag{6.6}$$

where

$$A_n = \begin{pmatrix} A_n^1 \\ \vdots \\ A_n^D \end{pmatrix}, \quad \epsilon_n = \begin{pmatrix} \epsilon_{1n} \\ \vdots \\ \epsilon_{Dn} \end{pmatrix}, \text{ and } \; L^*_n = \begin{pmatrix} L^*_{1n} \\ \vdots \\ L^*_{Dn} \end{pmatrix} \tag{6.7}$$

and

$$\Lambda = \begin{pmatrix} \Lambda_1 & 0 & \cdots & 0 \\ 0 & \Lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Lambda_D \end{pmatrix} \tag{6.8}$$

Even here the conditional variances are assumed be invariant across the levels in each dimension, and the assumption of conditional independence refers to independence of the components of ϵ_n.

6.3 Latent Class Model

In the traditional latent variable model we assume that the latent constructs are continuous. But this assumption may be inappropriate in many applications, particularly in the social and the behavioral sciences, wherein the latent constructs such as intelligence, character, personality, etc. are more meaningful as discrete or categorical concepts. The latent class model, which is analogous to the latent variable model, is appropriate for such situations.

The LCM presented here maps from a set of explanatory variables to a set of indicators through intermediate constructs represented by latent classes. We extend and refine traditional LCM by linking the class membership model with the indicators of latent classes.

Figure 6-3 outlines the framework for the latent class model. Z_n denotes the $Q \times 1$ vector of explanatory variables which affect the class membership of individual n.[6] The latent classes are assumed to be mutually exclusive and collectively exhaustive. The indicators, A_n, are assumed to be manifestations of the latent class. These indicators may be discrete or continuous[7].

A *class membership model*, denoted by $Q_s(Z_n;\theta)$ where θ is a parameter vector, captures the mapping from Z_n to the latent class, while the *measurement model* captures the mapping from the latent class to A_n. The measurement model specifies the distribution of the indicators given the latent class, and is denoted by $g(A_n|l^*_{sn} =$

[6]For ease of exposition, we assume a multiple indicator multiple causes analogue of the LCM such that the explanatory variables are perfectly measured. After the complete presentation of the LCM, it may be transparent to the reader how to extend the LCM for cases wherein the explanatory variables may be latent (i.e., *latent exogenous variables* in the terminology of latent variable models as discussed in section 2.3.1, page 46). Such extensions are presented within the development of LSCM.

[7]In this vein, the latent class model as presented in this chapter encompasses the latent profile model *with* causal class membership functions.

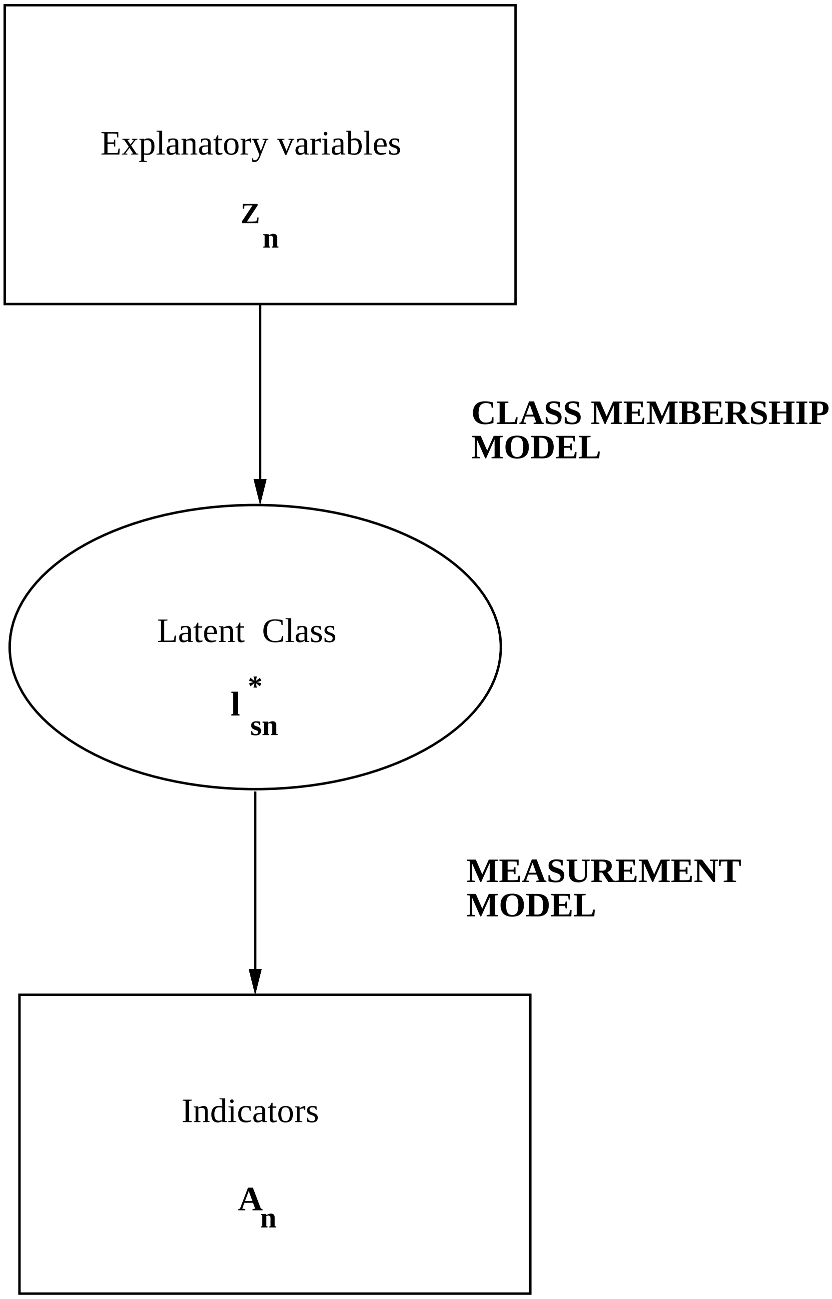


Figure 6-3: Latent Class Model: Multiple Indicator Multiple Causes

$1; \phi_s)$ where ϕ_s is a class-specific parameter vector.[8] Combining the class membership model and the measurement model, the latent class model $f(A_n|Z_n;\theta,\phi)$ is written as:

$$f(A_n|Z_n;\theta,\phi) = \sum_{s=1}^{S} g(A_n|l^*_{sn} = 1; \phi_s)\ Q_s(Z_n;\theta). \tag{6.9}$$

where $\phi = [\phi_1, \ldots, \phi_S]$.

In a similar vein, if the measurement model is dimension-specific, $f(A_n|Z_n;\theta,\phi)$ is written as:

$$\begin{aligned} f(A_n|Z_n;\theta,\phi) &= \sum_{l_1=1}^{L_1} \cdots \sum_{l_D=1}^{L_D} \Bigg[\left\{ \prod_{d=1}^{D} \prod_{p=1}^{P_d} g_{pdl_d}(a^d_{pn}; \phi_{pdl_d}) \right\} \cdot \\ & \qquad \mathrm{P}(T_n = [l_1, \ldots, l_D]|Z_n;\theta) \Bigg] \end{aligned} \tag{6.10}$$

As seen in chapter 3, the class membership model, $Q_s(Z_n;\theta)$, maps from Z_n to the latent class probabilities through a set of criterion functions. The specification of $Q_s(\cdot)$ depends on the specific problem context and characteristics of the latent class being modeled. Consequently, depending on the latent class characterization and the associated class membership model, we define three latent class models:

1. Categorical Criterion Latent Class Model;

2. Binary Criteria Latent Class Model; and

3. Ordinal Criteria Latent Class Model.

6.4 Latent Class Choice Model with Indicators

Figure 6-4 outlines the framework for the latent class choice model with class indicators. The components of the framework are self-explanatory. The model includes two sub-models: (1) latent class model, and (2) class-specific choice model.

[8] In principle, to capture response biases in the elicitation of the class indicators across gender, education, etc., we may specify the measurement model as $g(A_n|l^*_{sn} = 1, \tilde{Z}_n; \phi_s)$ where $\tilde{Z}_n$ include the corresponding socio-economic and demographic variables.

The *latent class model* consists of the class membership model, $\mathrm{Q}_s(Z_n;\theta)$, and the measurement model for the class indicators, $g(A_n|l^*_{sn}=1;\phi_s)$. The *class-specific choice model* predicts the choice behavior of an individual in latent class s, and thus depends on the choice set, (C_s), taste parameters (β_s), and the decision-protocol (R_s) associated with each class. The class-specific choice model expressing the choice probability of alternative i for individual n who is a member of class s can be written as:

$$\mathrm{P}(y_{in}=1|X_n;\beta_s,C_s,R_s). \tag{6.11}$$

Assuming that the class-specific choice model is independent of the conditional distribution of the class indicators, the probability of observing $[y_{in}, A_n]$ is written as:

$$\mathrm{P}(y_{in},A_n|X_n,Z_n;\theta,\beta,\phi)=\sum_{s=1}^{S}\mathrm{P}(y_{in}|X_n;\beta_s,C_s,R_s)\ g(A_n|l^*_{sn}=1;\phi_s)\ \mathrm{Q}_s(Z_n;\theta). \tag{6.12}$$

6.5 Latent Structure Choice Model

In the latent class choice model with indicators, we postulated that the unobserved heterogeneity is adequately captured through discrete or categorical constructs. We extend this model, to include heterogeneity stemming from individual's attitudes and perceptions, and thus incorporate the associated attitudinal and perceptual indicators. To this end, we advance a fairly general and comprehensive representation of the choice process.

The model formulated in this section is expected to transcribe the main ideas presented in the conceptual framework for choice modeling which is illustrated in Figure 1-1, into an empirically verifiable statistical model system wherein we postulate that the observed or stated choice behavior is the outcome of a probabilistic data generating process molded by a host of psychological factors.

Before we dive into the details of the model, it is instructive to view the model as

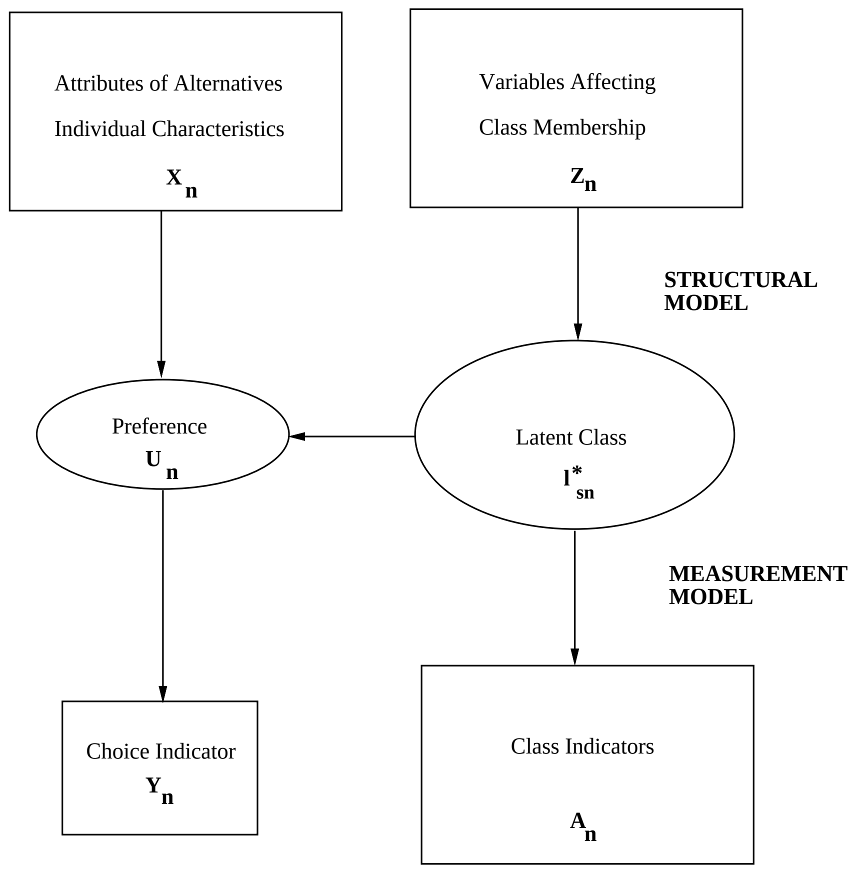


Figure 6-4: Latent Class Choice Model with Indicators

one subscribing to the following intuitive notions[9]:

- Perceptions attempt to "explain" part of the "unobserved" or random components of the alternative utility functions better through individual-specific unobserved components; and

- Attitudes attempt to "explain" unobserved individual heterogeneity, such as taste variations, choice set heterogeneity, decision protocol heterogeneity, etc, better.

The guiding philosophy in its development is that the incorporation of individual's attitudes and perceptions leads to a more behaviorally realistic representation of the decision-making process, and consequently to "better" predictive models. Further, the objective is to develop an operational modeling approach which is sufficiently general to warrant its adoption in diverse choice modeling contexts. To this end, the presentation of the modeling approach is rather abstract with illustrative examples to fathom underlying cryptic concepts.

Notation for the Latent Structure Choice Model

s = latent class index, with $s = 1, \ldots, S$, where S equals the number of latent classes.

C_n = choice set available[10] to individual n, where $|C_n| = J_n$.

β_s = choice model parameters[11] specific to class s.

C_s = choice set specific[12] to latent class s with $|C_s| = J_s$.

[9] It must be noted that perceptions may affect, in addition to alternative utilities, other stages of the choice process such as choice set formation, "choice" of adoption of decision protocol, etc. This issue will be highlighted in the illustrative example.

[10] This refers to the set of alternatives *deterministically* available.

[11] For simplicity we assume these parameters are fixed. In principle, one can allow random taste parameters for each class.

[12] For notational simplicity, we assume that the choice set is class-specific, though some of the subsets of the universal choice set may *not* be available to individuals due to availability restrictions.

R_s = decision protocol specific to latent class s.

D = dimension of the class membership vector.

T_n = D-dimensional random vector which denotes the class membership, i.e., $T_n = [l_1, \ldots, l_D]'$ where l_d is the level in dimension d.[13]

H_{dkn} = k^{th} criterion function for latent class dimension $d = 1, \ldots, D$, $\forall k = 1, \ldots, K_d$, where K_d is the number of criterion functions for dimension d.

$$l^*_{sn} = \begin{cases} 1 & \text{if individual } n \text{ is in latent class } s \\ 0 & \text{otherwise} \end{cases}$$

Z_n = $Q_Z \times 1$ vector of observable characteristics of individuals.

Z^*_n = $M_Z \times 1$ vector of latent characteristics of individuals.

$A_{Z;n}$ = $P_Z \times 1$ vector of indicators of the latent vector Z^*_n.

$A_{S;n}$ = $P_S \times 1$ vector of indicators of latent class for individual n.

U_{isn} = utility of alternative i to individual n in latent class s.

$$y_{in} = \begin{cases} 1 & \text{if alternative } i \text{ is chosen by individual } n \\ 0 & \text{otherwise} \end{cases}$$

X_{in} = $Q_X \times 1$ vector of observable attributes of alternatives and characteristics of individual for alternative $i \in C_n$.

X^*_{in} = $M_X \times 1$ vector of latent attributes of alternatives and characteristics of individual for alternative $i \in C_n$.

$A_{X;n}$ = $P_X \times 1$ vector of indicators of the latent vector X^*_n.

W_n = $Q_W \times 1$ vector of observed explanatory variables[14] affecting the latent vectors Z^*_n, X^*_n.

---

[13] Note that $T_n = [T_{1n} = l_1, \ldots, T_{dn} = l_d, \ldots, T_{Dn} = l_D]'$ where T_{dn} denotes the d^{th} component of T_n. l_d may be a binary or an ordered categorical variable.

[14] These variables may include *both* individual characteristics and attributes of alternatives.

The modeling framework is schematized in Figure 6-5. In the figure, rectangles represent observable set of variables (explanatory variables as well as indicators of unobserved constructs), while ellipses represent unobserved constructs. The key features are:

1. Incorporation of *latent variables* such as attitudes and perceptions; and
2. Incorporation of discrete or categorical latent constructs such as choice sets, decision protocols, and taste variations through *latent classes*.

To this end, we encompass latent variable models and latent class models under the umbrella of latent structure models. Further, since we develop a choice model with explicit links to latent structure models purportedly to enrich the choice model, we refer to the choice model as the *latent structure choice model*.

The primary focus of our interest is in modeling the underlying process governing the individual's choices and preferences indicated[15] by y_{in}. The choice process is hypothesized to vary systematically across a *finite* set of "unobserved" groups, and to be homogeneous within each such group[16]. Since each homogeneous group is unobserved, the groups are characterized by latent classes which are mutually exclusive and collectively exhaustive (i.e., an individual is a member of at most one class and at least one class). It must be emphasized that the *homogeneity* of the classes is with respect to the *unobserved* constructs such as tastes, decision protocol, and choice set, and *not* with regard to explanatory variables.

The observable individual characteristics, Z_n, and the individual's attitudes, Z_n^*, are postulated to explain or affect the individual's class membership indicated by l_{sn}^*. The latent classes manifest themselves through observable indicators $A_{S;n}$[17]. The

[15] It must be noted that y_{in}, the indicator of the underlying preference, may include the choice indicator in a revealed preference context, and stated preferences such as alternative ratings and rankings and stated choice in hypothetical choice experiments.

[16] In principle we can also allow for unobserved *heterogeneity* within each group, to capture individual-specific preference biases and idiosyncratic taste variations which may not be captured through unobserved heterogeneity in tastes, decision protocol, and choice set considered, especially, in the presence of multiple responses per individual such as in discrete panel data and stated preference tasks.

[17] Note that the indicators, which are typically responses to attitudinal questions, may also depend directly on characteristics of individual to capture response biases.

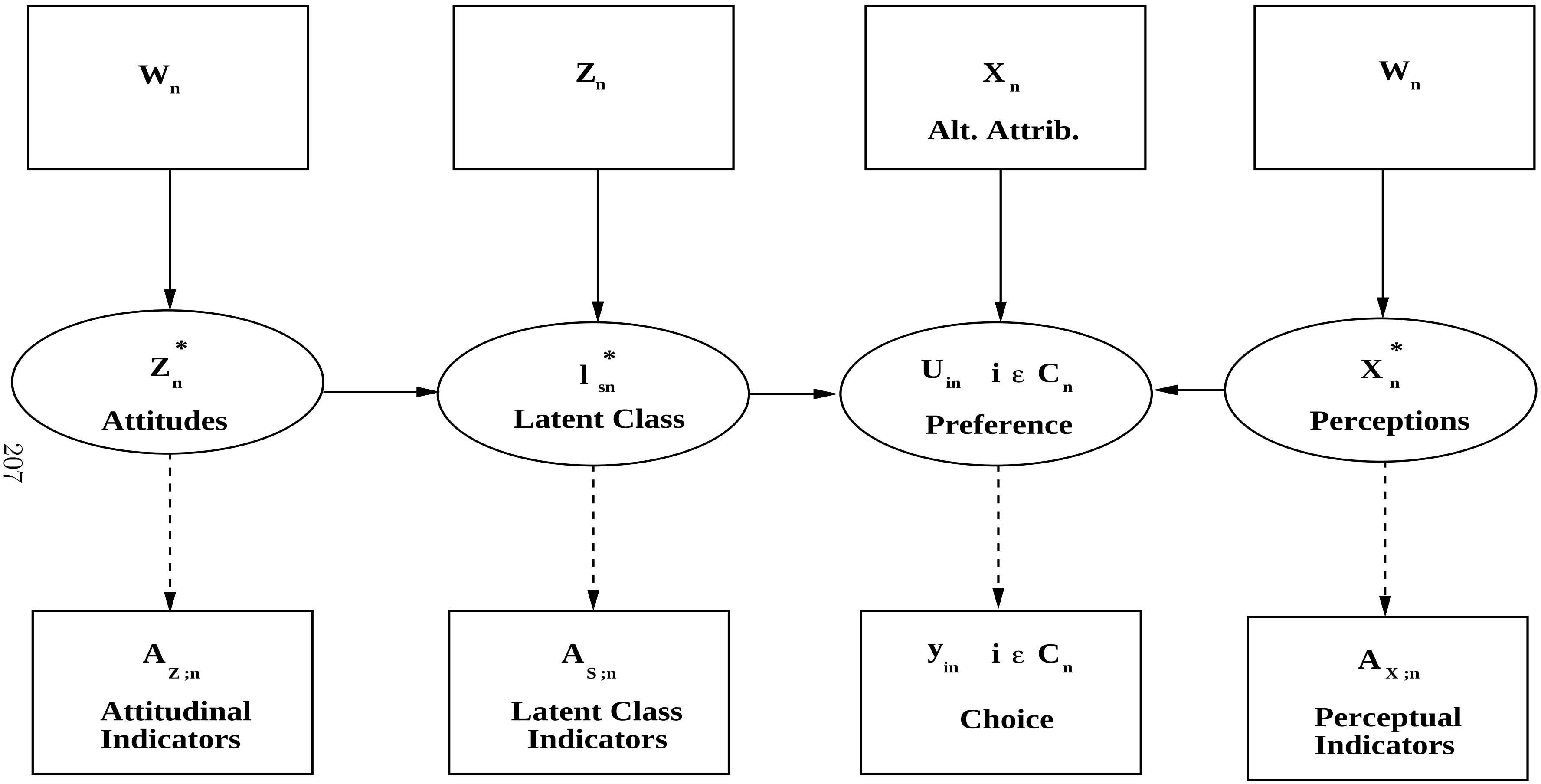


Figure 6-5: Latent Structure Choice Model: A Schematic Representation

individual's attitudes Z_n^* are postulated to be formed through an *attitude formation process* represented by a mapping from observable explanatory variables, W_n (which may include both individual characteristics and attributes of alternatives). Further, these attitudes are manifested through the observable attitudinal indicators $A_{Z;n}$.

The utility of alternative i for individual n, U_{isn}, depends on the observable attributes of alternatives and individual characteristics, X_{in}, the intangible or perceptual attributes, X_{in}^*, and the latent class s to which the individual belongs. Specifically, X_{in}^* includes attributes of alternative i as "perceived" by the individual, and hence, affect the choice process, but are not directly observable to the analyst. X_{in}^* are formed through a *perception formation process* which is represented by a mapping from W_n to the perceptual indicators $A_{X;n}$. The perception formation process captures the notion that the *same* objective reality, as represented by the decision-making environment, may be *perceived differently* by individuals depending on how the individual processes the information, and consequently are purportedly incorporated through individual's socio-economic and demographic characteristics.

It must be noted that Z_n^* may represent attitudes and/or perceptions. For example, if the latent class characterizes the choice actually considered by an individual in a particular choice situation, then Z_n^* may include perceptions of the attributes of alternatives in terms of "cognitive" categories or thresholds which may be utilized in a non-compensatory choice set formation process. On the other hand, if the latent class characterizes taste heterogeneity, then Z_n^* denotes attitudes such as individual's sensitivity to attributes of alternatives. For simplicity of exposition, we assume that Z_n^* represents attitudes.

Before we proceed with the formulation and specification of the LSCM, it is useful to crystallize, at the very outset, the aforementioned concepts through a shipper's freight mode choice example. Herein X_{in} includes observable service attributes such as transit time, rate, etc. X_{in}^* includes the shipper's perceptions about intangible service attributes such as "service-quality", "reliability", which may affect the choice process, but are not directly observable. Shipper's form perceptions about carriers and their attributes depending on past experiences, communications of carriers, along with the objective service-related attributes. Further, perceptions are formed in conjunction with the shipper's service requirements such as delivery schedule, shipment size, special needs and customizations, etc. Hence, it must be emphasized that shippers with different characteristics and requirements may perceive identical unobservable service attributes of carriers differently. Consequently, W_n may include shipper's characteristics as well as "engineering" or objective attributes of the different carriers.

Perceptions are manifested through *satisfaction ratings* of service attributes[18] such as equipment (e.g., equipment availability, condition of equipment), consistency (e.g., consistency of pick up and delivery, transit time reliability), convenience (e.g., shipment tracing, responsiveness to inquiries), integrity (e.g., accuracy of billing, responsiveness to claims), and flexibility (e.g., rerouting, rescheduling, handling of emergency shipments, special pickup and delivery schedules), and which form $A_{X;n}$. These satisfaction ratings may be collected on a Likert-type scale in surveys of shippers.

Z_n^* may characterize shipper's attitudes or sensitivity to attributes such as time-sensitivity, cost-sensitivity, service-quality sensitivity, etc. Attitudes are formed by past experiences, shipper's needs and requirements, and shipper's characteristics which may include earliest acceptable delivery time, annual sales, maximum acceptable delay, electronic data interchange (EDI) usage, annual tonnage shipped, average length of haul, value of shipments, etc. The indicators of attitudes, $A_{Z;n}$, may include shipper's *importance ratings* of service attributes, including intrinsic preferences towards different carriers. These importance ratings may be also be collected on a Likert-type scale.

The latent class, l_{sn}^*, may characterize the carriers actually considered by the shipper for a particular shipment. There may also exist a carrier which is the shipper's core carrier. The indicators, $A_{S;n}$, of the latent class may include responses to questions such as: Would you consider carrier j as being available to you?. The latent class may also characterize groups of shippers with similar response patterns to changes in service attributes (i.e., market segments) wherein the importance ratings of service attributes are used as latent class indicators.

[18] See the study by La Londe and Cooper [1989] for more details.

The preference indicator y_{in} may include shipper's shipment shares across competing carriers in particular corridors or origin-destination markets, and/or ratings, rankings and stated choices of transportation alternatives in hypothetical situations settings.

Going back to the formulation of the LSCM, we postulate the following relationships among the observable and latent variables/classes of interest. The structural equations relate the explanatory variables to the latent variables/classes, while the measurement equations relate the latent variables/classes to the observed indicators. Specifically, the model system used in the implementation of the schema as illustrated in Figure 6-5 includes:

1. *Structural Model* consisting of the relationships between the relevant problem characteristics such as attributes of alternatives, individual's socio-economic and demographic characteristics, and the individual"s attitudes, perceptions and class membership, and the underlying preferences towards alternatives.

2. *Measurement Model* consisting of the relationships between the underlying preferences and the revealed preferences & the stated preferences in surveys with different response formats which capture the response biases and protocol effects; the mapping from the latent attitudes, perceptions and classes to the corresponding attitudinal, perceptual and class indicators.

Given the overview of the LSCM, we outline the specification of the structural and measurement models of LSCM. The specification of LSCM boils down to the specification of the relationships (the "arrows") in Figure 6-5.

Structural Model

Let $f_1(\cdot)$ denote the mapping from W_n to Z_n^* with associated parameter vector Γ_Z[19]. Similarly, $f_2(\cdot)$ denotes the mapping from W_n to X_n^* with associated parameter vector

[19]In the notation for the structural and measurement equations and the following discussion we denote only the "structural" or "systematic" parameters and deliberately suppress parameters associated with "random" components.

Γ_X. The utilities U_{isn} for individual n in latent class s are specified as functions $U(\cdot)$ of observable attributes X_{in} and unobservable attributes X^*_{in} with taste parameters β_s[20]. Since we assume that attitudes, perceptions, and utilities are *continuous* variables, the mappings $f_1(\cdot)$, $f_2(\cdot)$ and $U(\cdot)$ are real functions.

The specification of the probabilistic mapping from Z_n and Z^*_n to l^*_{sn} is through the specification of criterion functions[21] H_{dkn}, $\forall k = 1, \ldots, K_d$, $\forall d = 1, \ldots, D$, and consequently, through a class membership model. Depending on the latent class characterization, a class membership model denoted by $Q_s(\cdot)$ with parameters θ is postulated. The generic specification of the structural model can be written as:

$$Z^*_n = f_1(W_n; \Gamma_Z) \tag{6.13}$$

$$H_{dkn} = H(Z_n, Z^*_n; \theta_{dk}), \quad \forall k = 1, \ldots, K_d, \ \forall d = 1, \ldots, D \tag{6.14}$$

$$\begin{aligned} \mathrm{P}(l^*_{sn} = 1) &= \tilde{Q}_s(\tilde{H}_{dkn}, \ \forall k = 1, \ldots, K_d, \ \forall d = 1, \ldots, D) \\ &= Q_s(Z_n, Z^*_n; \theta), \quad \forall s = 1, \ldots, S \end{aligned} \tag{6.15}$$

$$X^*_n = f_2(W_n; \Gamma_X) \tag{6.16}$$

$$U_{isn} = U(X_{in}, X^*_{in}; \beta_s), \quad i \in C_s; \ \forall s = 1, \ldots, S \tag{6.17}$$

Measurement Model

The measurement equations are specified in a similar vein. The function $g_1(\cdot)$ with parameters Λ_Z maps from the latent vector Z^*_n to the associated indicators $A_{Z;n}$. Similarly, the function $g_3(\cdot)$ with parameters Λ_X maps from the latent vector X^*_n to the associated indicators $A_{X;n}$. The indicators of the latent classes are specified as functions $g_2(\cdot)$ of l^*_{sn} with parameters ϕ_s. The class-specific choice model is the probabilistic mapping from the class-specific utility functions to choice, and is given by a function $g^i_4(\cdot)$ which maps the class-specific utility functions to the choice i given the associated class-specific decision protocol and choice set. The generic specification

[20]The effects of latent concepts such as decision protocol and choice sets considered are subsumed in the choice model conditional on these latent concepts, and hence not explicitly denoted in the utility functions.

[21]$\tilde{H}_{dkn}$, $\forall k = 1, \ldots, K_d$, $\forall d = 1, \ldots, D$ are the corresponding systematic components.

of the measurement model can be written as:

$$
\begin{aligned}
A_{Z;n} &= g_1(Z_n^*; \Lambda_Z) && (6.18)\\
\mathrm{P}(A_{S;n}|l_{sn}^* = 1) &= g_2(\phi_s), \quad \forall s = 1, \ldots, S && (6.19)\\
A_{X;n} &= g_3(X_n^*; \Lambda_X) && (6.20)\\
\mathrm{P}(y_{in} = 1|l_{sn}^* = 1) &= g_4^i(U_{jsn},\ \forall j \in C_s; R_s), \quad i \in C_n;\ \forall s = 1, \ldots, S && (6.21)
\end{aligned}
$$

It must be noted that the functions $g_1(\cdot)$ and $g_3(\cdot)$ depend on how the indicators $A_{Z;n}$ and $A_{X;n}$ are measured. Consequently, we may adopt functions mapping from continuous variables into continuous or ordered categorical variables. For simplicity of exposition we assume $A_{Z;n}$ and $A_{Z;n}$ are continuous variables. In appendix C we present a measurement model wherein the indicators are ordered categorical. The function $g_2(\cdot)$ depends on the characterization of the latent class, and correspondingly we may adopt the measurement model specification as discussed in section 6.2. The mapping $g_4(\cdot)$ is either deterministic or probabilistic depending on the class-specific choice set and decision protocol.

For simplicity, we did not allow the functions $g_1(\cdot)$, $g_2(\cdot)$ and $g_3(\cdot)$ to include 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 perceptual, attitudinal or class indicators. For example, if the individual is responding to the question purported to elicit the perceived availability of an alternative, then the response may be affected by the individual's desirability of the alternative (see Ben-Akiva and Boccara [1993]). To illustrate another example of this issue, consider a situation wherein the individual is requested to provide an overall measure of satisfaction with an alternative. This indicator may be tainted depending on whether the alternative was chosen in a recent choice situation.

6.5.1 Model Formulation & Specification

In this section, we formulate and specify the structural and measurement equations in more detail. The sequence followed for the presentation is:

1. Latent Variable Sub-model;
2. Latent Class Sub-model; and
3. Discrete Choice Sub-model.

Latent Variable Sub-model

Herein we specify the the relationships among the latent variables Z_n^* and X_n^*, the corresponding indicators $A_{Z;n}$ and $A_{X;n}$, and the explanatory variables W_n. The mappings from the exogenous variables to the latent variables form the structural equations, while the mappings from the latent variables to the indicators form the measurement equations.

Structural equations for latent vector Z^* and X^*:

Assuming *linear in parameters* functional forms for the structural equations of Z^* and X^*, we have

$$Z_n^* = B_Z Z_n^* + \Gamma_Z W_n + \zeta_Z \tag{6.22}$$

$$X_n^* = B_X X_n^* + \Gamma_X W_n + \zeta_X \tag{6.23}$$

The matrices B_Z and B_X allow for the latent variables to affect each other to capture structural dependencies among the components of the latent vector Z^* and X^*. The matrices $(I - B_Z)$ and $(I - B_X)$ are assumed to be non-singular. The matrices Γ_Z and Γ_X capture the effects of W_n on the latent vectors. ζ_Z and ζ_X are disturbance vectors with $\mathrm{E}(\zeta_Z) = 0$, $\mathrm{E}(\zeta_X) = 0$, and which are uncorrelated with W_n. We assume, without any loss of generality, W_n to be in deviation form.

Measurement equations for latent vector Z^* and X^*:

Assuming *linear in parameters* functional forms for the measurement equations for Z^* and X^*, we have

$$A_{Z;n} = \Lambda_Z Z_n^* + \epsilon_Z \tag{6.24}$$

$$A_{X;n} = \Lambda_X X_n^* + \epsilon_X \tag{6.25}$$

where Λ_Z and Λ_X are coefficient matrices that capture the relationship from Z_n^* to $A_{Z;n}$, and from X_n^* to $A_{X;n}$ respectively, and ϵ_Z and ϵ_X are the errors of measurement for $A_{Z;n}$ and $A_{X;n}$, respectively. The errors of measurement are assumed to be uncorrelated with ζ_Z, ζ_X, Z_n^*, X_n^*, and with each other. The expected values of ϵ_Z and ϵ_X are zero. To simplify matters, $A_{Z;n}, A_{X;n}$ are written as deviations from their respective means.

Latent Class Sub-model

In section 6.3, a latent class model analogous to the latent variable model was presented. In the LSCM, the latent class model forms a sub-model. The process of assignment of an individual to a latent class is governed by a class membership model. As before, we postulate the existence of criterion functions which map from Z_n and Z_n^* to a vector of latent variables H_{dkn} $\forall k = 1, \ldots, K_d; \forall d = 1, \ldots, D$.

Structural equation for criterion functions:

Assuming a linear functional for the criterion functions, H_{dkn} can be specified as:

$$H_{dkn} = \theta'_{zdk} Z_n + \theta'_{z^*dk} Z_n^* + \delta_{dkn}, \quad \forall k = 1, \ldots, K_d; \forall d = 1, \ldots, D \tag{6.26}$$

where θ_{zdk} and θ_{z^*dk} are unknown parameters to be estimated and δ_{dkn} is a random error component.

Structural equation for latent class: class membership model:

The class membership model, $Q_s(Z_n, Z_n^*; \theta)$, maps from Z_n and Z_n^* to the latent class probabilities through the criterion functions, H_{dkn}, i.e.,

$$\begin{aligned} \mathrm{P}(l_{sn}^* = 1 | Z_n, Z_n^*; \theta) &= \tilde{Q}_s(\tilde{H}_{dkn},\ \forall k = 1, \ldots, K_d, \forall d = 1, \ldots, D) \\ &= Q_s(Z_n, Z_n^*; \theta), \quad \forall s = 1, \ldots, S \end{aligned} \tag{6.27}$$

The specification of the function $Q_s(\cdot)$ depends on the specific problem context, and the latent class characterization. Following along the lines of the class membership models developed in section 3.3, three types of membership models are briefly presented.

Categorical Criterion Model

Herein each latent class s is associated with a criterion function H_{sn}, and a criterion maximizing rule associates a latent class with the criterion functions. The indicator function for the latent class s is written as:

$$l_{sn}^* = \begin{cases} 1 & \text{if} \ \ H_{sn}(Z_n, Z_n^*; \theta_s) = \max\limits_{\forall s'=1,\ldots,S} \{H_{s'n}(Z_n, Z_n^*; \theta_{s'})\} \\ 0 & \text{otherwise} \end{cases} \tag{6.28}$$

Then by assuming different parametric distributions for $(\delta_{1n}, \ldots, \delta_{Sn})$, different class membership models such as the MNL-type and MNP-type models can be constructed.

Ordinal Criteria Model

Suppose the latent class can be characterized by ordered levels along each of the D dimensions of the class. The modeling approach is to assume that each dimension d is captured by a criterion function H_d. Let L_d represent the levels along dimension d.

The membership model is formulated under the assumption that Z_n and Z_n^* affect the level of latent dimension d through a "threshold crossing" model wherein a particular level is triggered if the corresponding criterion function falls between two thresholds. Let the criterion function $H_{dn}, \forall d = 1, \ldots, D$ for individual n be written as:

$$H_{dn} = \theta'_{zd} Z_n + \theta'_{z^*d} Z_n^* + \delta_{dn}, \quad \forall d = 1, \ldots, D \tag{6.29}$$

Further, correlations among the random components of H_d's (i.e., δ_d's) are allowed to capture the unobserved interrelationships among the dimensions. Thus, the probability of individual n being in latent class $T_n = [l_1, \ldots, l_D]'$ (with index s such that $(l^*_{sn} = 1) \Leftrightarrow T_n = [l^s_1, \ldots, l^s_D]' \Leftrightarrow l^s_d = l_d \forall d)$, denoted by $Q_s(Z_n, Z^*_n, \theta)$, equals:

$$\begin{aligned}
& \mathrm{P}\left(\bigcap_{d=1}^{D} (\tau^d_{l_d-1} \leq H_{dn} \leq \tau^d_{l_d})\right) \\
= \ & \mathrm{P}\left(\bigcap_{d=1}^{D} (\tau^d_{l_d-1} \leq \theta'_{zd} Z_n + \theta'_{z^*d} Z^*_n + \delta_{dn} \leq \tau^d_{l_d})\right) \\
= \ & \mathrm{P}\left(\bigcap_{d=1}^{D} (\tau^d_{l_d-1} - \theta'_{zd} Z_n - \theta'_{z^*d} Z^*_n \leq \delta_{dn} \leq \tau^d_{l_d} - \theta'_{zd} Z_n - \theta'_{z^*d} Z^*_n)\right)
\end{aligned} \tag{6.30}$$

where τ^d are the threshold parameters for dimension d. By specifying different parametric distributions for $(\delta_{1n}, \ldots, \delta_{Dn})$, different class membership models can be constructed.

Binary Criteria Model

In this case the latent class is identified by an D-dimensional vector, with each dimension represented by a binary variable. The d^{th} dimension takes the value 1 if, and only if, $H_{dkn} \geq 0$, $\forall k = 1, \ldots, K_d$. Assume that the random components of criterion functions across dimensions of the latent class are independent (i.e., if $\delta_d = [\delta_{1d}, \ldots, \delta_{K_d d}]$ and $\delta_{d'} = [\delta_{1d'}, \ldots, \delta_{K_{d'} d'}]$, then δ_d and $\delta_{d'}$ are independent $\forall d \neq d'$). The probability of the individual being in latent class $T_n = [l_1, \ldots, l_D]'$ equals

$$\begin{aligned}
= \ & \mathrm{P}\left(T_{dn} = l_d, \ \forall d = 1, \ldots, D\right) \\
= \ & \mathrm{P}\left(\left\{\bigcap_{k=1}^{K_d} \left(\delta_{dkn} \geq v^+_{dkn}\right), \forall d | l_d = 1\right\} \bigwedge \right. \\
& \left. \left\{\left(\exists k \in \{1, \ldots, K_d\} : \left(\delta_{dkn} < v^+_{dkn}\right)\right), \forall d | l_d = 0\right\}\right)
\end{aligned} \tag{6.31}$$

where $v^{+}_{dkn} = -\theta'_{zdk}Z_n - \theta'_{z^*dk}Z^*_n, \quad \forall k = 1, \ldots, K_d$.

By the assumption of independence of random components of criterion functions across latent dimensions the above equation reduces to:

$$Q_s(Z_n, Z^*_n; \theta) = \prod_{d=1}^{D} \left[\mathrm{P}(T_{dn} = 1|Z_n, Z^*_n; \theta)\right]^{l_d} \left[1 - \mathrm{P}(T_{dn} = 1|Z_n, Z^*_n; \theta)\right]^{1-l_d} \qquad (6.32)$$

Measurement equation for latent class:

The measurement equation for the class indicators, $A_{s;n}$ can be expressed as discussed in section 6.2.

Discrete Choice Sub-model

Structural equation for latent preferences:

The structural equation for the utilities of alternatives is expressed as a *linear in parameters* function of the observable attributes of alternatives and individual's characteristics, X_{in}, and the perceptual attributes X^*_{in} given the latent class s.

$$U_{isn} = \beta'_{xs}X_{in} + \beta'_{x^*s}X^*_{in} + \nu_{isn}, \qquad \forall i \in C_s; \forall s = 1, \ldots, S \qquad (6.33)$$

The utility specification is meaningful only for those latent classes wherein the class-specific decision protocol is utility maximization.

Measurement equation for latent preferences:

We assume for simplicity that the indicator of the preferences of alternatives is the choice indicator[22]. If in a particular latent class, say s, an alternative is hypothesized to be chosen according to a random utility maximizing protocol[23] then the

[22] Indicators such as stated preference rankings and ratings can be suitably incorporated.

[23] It must be noted that if the latent class characterizes the decision protocol then the protocol would vary across the latent classes. It may include "deterministic" protocols such as "always pick alternative i" (if the individual in that class is captive to i), "pick the alternative with minimum cost or price", etc.

measurement equation for the underlying utilities is written as:

$$y_{in}|(l^*_{sn} = 1) = \begin{cases} 1 & \text{if } \; U_{isn}(X_{in}, X^*_{in}, \beta_s) = \max\limits_{j \in C_s}\{U_{jsn}(X_{jn}, X^*_{jn}, \beta_s)\} \\ 0 & \text{otherwise} \end{cases} \tag{6.34}$$

If it is hypothesized that each latent class s has its own parameter vector β_s in the choice situation under consideration, and ν_{isn}'s are independently and identically distributed Gumbel (0,1) across alternatives and individuals, we obtain a class-specific MNL model, where the probability of individual n in latent class s choosing alternative i is expressed as:

$$\mathrm{P}(y_{in} = 1 | l^*_{sn} = 1, X_n, X^*_n; \beta_s) = \frac{\exp(\beta'_{xs} X_{in} + \beta'_{x^*s} X^*_{in})}{\sum\limits_{j \in C_s} \exp(\beta'_{xs} X_{jn} + \beta'_{x^*s} X^*_{jn})} \tag{6.35}$$

As it is well known, by suitable assumptions for the distributions of ν_{isn}, we obtain the MNP and Nested Logit models.

Distributional Assumptions

As noted in the latent class sub-model, a class of parameterized distributions may be adopted for the random components of the criterion functions, to construct different types of class membership models. In the latent variable sub-models we may allow for the random components of the structural equations of Z^*_n and X^*_n to be correlated (i.e, ζ_Z and ζ_X are correlated). All other error vectors appearing across the structural and measurement equations in the latent variable sub-models are assumed to be independent. But correlations may be allowed within the components of the same error vector. Further, in the latent variable sub-models, we assume all the error vectors to have continuous supports.

Distributional assumptions in each of the sub-models must ensure that each of the sub-models are identified. It must be noted that for the class-specific discrete choice sub-model necessary and sufficient conditions are available, while for the latent variable models *sufficient* conditions for identification must be checked[24]. For the latent class sub-model general necessary and sufficient conditions do not exist at this time. For this sub-model, necessary scaling restrictions on the criterion functions to reflect the fact that the criterion functions are latent should be imposed. We allow for arbitrary specifications of the measurement model of the latent class sub-model depending on how the data on class indicators is collected.

The Choice Model: Primary model of interest

The primary model of interest is the choice model conditional on X_n, Z_n and W_n. If $f_{X^*,Z^*}(X^*, Z^*|W_n; B_X, B_Z, \Gamma_X, \Gamma_Z)$ represents the joint density of (X^*, Z^*) conditional on W_n, the choice model is written as, $\mathrm{P}(y_{in} = 1|Z_n, X_n, W_n; \theta, \beta)$

$$= \int_{X^*}\int_{Z^*} \sum_{s=1}^{S} \mathrm{P}(y_{in} = 1|l^*_{sn}, X_n, X^*_n; \beta_s, C_s, R_s)\ Q_s(Z_n, Z^*_n; \theta)\ \cdot \\ f_{X^*,Z^*}(X^*, Z^*|W_n; B_X, B_Z, \Gamma_X, \Gamma_Z)\, dX^*\, dZ^* \qquad (6.36)$$

6.5.2 Model Estimation

Now we turn our attention to the derivation of the sample likelihood function. The probability of observing the response vector, $[Y_n, A_{X;n}, A_{Z;n}, A_{S;n}]$, where $Y_n = [y_{1n}, \ldots, y_{J_n n}]$, conditional on the explanatory variables, $[X_n, Z_n, W_n]$, i.e., $\mathrm{P}(Y_n, A_{X;n}, A_{Z;n}, A_{S;n}|X_n, Z_n, W_n; \theta, \beta, \phi, B_X, B_Z, \Gamma_X, \Gamma_Z, \Lambda_X, \Lambda_Z)$, equals

$$\int_{X^*}\int_{Z^*} \mathrm{P}(Y_n, A_{X;n}, A_{Z;n}, A_{S;n}|X^*, Z^*, X_n, Z_n, W_n; \theta, \beta, \phi, \Lambda_X, \Lambda_Z)\ \cdot \\ f_{X^*,Z^*}(X^*, Z^*|W_n; B_X, B_Z, \Gamma_X, \Gamma_Z)\, dX^*\, dZ^*$$

$$= \int_{X^*}\int_{Z^*} \mathrm{P}(Y_n, A_{S;n}|X^*, Z^*, X_n, Z_n, W_n; \theta, \beta, \phi)\ g_X(A_X|X^*; \Lambda_X)\ \cdot \\ g_Z(A_Z|Z^*; \Lambda_Z)\ f_{X^*,Z^*}(X^*, Z^*|W_n; B_X, B_Z, \Gamma_X, \Gamma_Z)\, dX^*\, dZ^*$$

[24]See Bollen [1989] for an extensive discussion of identification issues in latent variable models, though it must be noted that no general necessary and sufficient conditions for identification exist.

where $g_X(\cdot)$ and $g_Z(\cdot)$ denote the conditional distributions of the indicators of X^* and Z^*, respectively. The second equation follows from the distributional assumptions since conditional on Z^* and X^*, $[Y_n, A_{S;n}]$, $A_{Z;n}$ and $A_{X;n}$ are independent.

For notational simplicity define:

$$\mathrm{P}(Y_n|l^*_{sn} = 1, X^*, X_n; \beta_s, C_s, R_s) \equiv \mathrm{P}(Y_n|l^*_{sn}, X^*) \tag{6.37}$$

$$\mathrm{P}(A_{S;n}|l^*_{sn} = 1; \phi_s) \equiv \mathrm{P}(A_{S;n}|l^*_{sn}) \tag{6.38}$$

$$\mathrm{P}(l^*_{sn} = 1|Z^*, Z_n; \theta) \equiv \mathrm{P}(l^*_{sn}|Z^*) \tag{6.39}$$

$$f_{X^*,Z^*}(X^*, Z^*|W_n; B_X, B_Z, \Gamma_X, \Gamma_Z) \equiv f(X^*, Z^*) \tag{6.40}$$

$$g_X(A_X|X^*; \Lambda_X) \equiv g(A_X|X^*) \tag{6.41}$$

$$g_Z(A_Z|Z^*; \Lambda_Z) \equiv g(A_Z|Z^*) \tag{6.42}$$

Then the likelihood function reduces to:

$$\begin{aligned}
&\int_{X^*}\int_{Z^*} \left\{\sum_{s=1}^{S} \mathrm{P}(Y_n|l^*_{sn}, X^*)\mathrm{P}(A_{S;n}|l^*_{sn})\mathrm{P}(l^*_{sn}|Z^*)\right\} \; g(A_X|X^*) \;\; g(A_Z|Z^*) \;\cdot \\
&\qquad f(X^*, Z^*)\, dX^*\, dZ^* \\
=\; & \sum_{s=1}^{S} \left\{\int_{X^*}\int_{Z^*} \mathrm{P}(Y_n|l^*_{sn}, X^*)\mathrm{P}(A_{S;n}|l^*_{sn})\mathrm{P}(l^*_{sn}|Z^*) \; g(A_X|X^*) \; g(A_Z|Z^*) \;\cdot \right. \\
&\qquad \left. f(X^*, Z^*)\, dX^*\, dZ^* \right\} \\
=\; & \sum_{s=1}^{S} \left\{\mathrm{P}(A_{S;n}|l^*_{sn}) \int_{X^*}\int_{Z^*} \mathrm{P}(Y_n|l^*_{sn}, X^*) \; \mathrm{P}(l^*_{sn}|Z^*) \; g(A_X|X^*) \cdot \right. \\
&\qquad \left. g(A_Z|Z^*) \; f(X^*, Z^*)\, dX^*\, dZ^* \right\}
\end{aligned} \tag{6.43}$$

As seen in equation (6.43) the likelihood function is a complex multi-dimensional integral. The dimensionality of the integral equals the number of hypothesized attitudes and perceptions (i.e., $M_Z + M_X$). If the random vectors ζ_X and ζ_Z are assumed to be independent, then the latent vectors X^* and Z^* are independent conditional on the explanatory variables, and equation (6.43) reduces to:

$$\sum_{s=1}^{S} \left\{ \mathrm{P}(A_{S;n}|l^*_{sn}) \left[\int\limits_{X^*} \mathrm{P}(Y_n|l^*_{sn}, X^*) \; g(A_X|X^*) \; f(X^*) \, dX^* \right] \cdot \right.$$
$$\left. \left[\int\limits_{Z^*} \mathrm{P}(l^*_{sn}|Z^*) \; g(A_Z|Z^*) \; f(Z^*) \, dZ^* \right] \right\} \qquad (6.44)$$

Thus, the dimensionality of the integral decreases alleviating the difficulties in calculating the likelihood function (to a certain extent).
The problem then reduces to the specification of the density functions of X^* and Z^*. If

1. B_Z and B_X matrices are zeros (the structural equations of the latent variable sub-models reduce to *seemingly unrelated system of equations*); or

2. B_Z and B_X matrices are lower triangular *and* the components of ζ_X and ζ_Z are independent, i.e., the variance-covariance matrices of ζ_X and ζ_Z are diagonal (*recursive structural system*)

then, the probability density functions are sufficiently identified and can be used without any difficulty in a maximum likelihood estimation procedure under the assumption of multivariate normal distribution for ζ_Z and ζ_X. Else, the usual identification problems associated with the estimation of simultaneous equations apply (see Greene [1990]).

An alternate estimation procedure is to estimate the two latent variable sub-models corresponding to equations (6.22) and (6.24), and equations (6.23) and (6.25), respectively. The likelihoods for these latent variable sub-models are:

$$f(A_{Z;n}|W_n; B_Z, \Gamma_Z, \Lambda_Z) = \int\limits_{Z^*} g_Z(A_{Z;n}|Z^*; \Lambda_Z) \; f_{Z^*}(Z^*|W_n; B_Z, \Gamma_Z) \, dZ^* \qquad (6.45)$$

and

$$f(A_{X;n}|W_n; B_X, \Gamma_X, \Lambda_X) = \int\limits_{X^*} g_X(A_{X;n}|X^*; \Lambda_X) \; f_{X^*}(X^*|W_n; B_X, \Gamma_X) \, dX^* \qquad (6.46)$$

Using these estimated models, the latent variables are fitted[25] to obtain $\hat{X}^*$ and $\hat{Z}^*$ and used in a conditional maximum likelihood estimation to estimate the *latent class choice model with class indicators*, i.e., maximizing sample log-likelihood corresponding to the observation likelihood function:

$$\sum_{s=1}^{S} \mathrm{P}(Y_n|l^*_{sn}, \hat{X}^*)\ \mathrm{P}(A_{S;n}|l^*_{sn})\ \mathrm{P}(l^*_{sn}|\hat{Z}^*) \tag{6.47}$$

It must be noted that introducing the fitted values, $\hat{X}^*$ and $\hat{Z}^*$, in the conditional likelihood function implies some degree of inconsistency (let alone efficiency) in estimating the parameters of the conditional likelihood function. More formally, the sampling distribution of $\hat{X}^*$ and $\hat{Z}^*$, $F_{\hat{X}^*,\hat{Z}^*}$ must be used in the conditional likelihood function. Then, the conditional likelihood to be maximized can be written as:

$$\mathcal{L}(\Theta) = \int_{\hat{X}^*}\int_{\hat{Z}^*} \sum_{s=1}^{S}\ \mathrm{P}(Y_n|l^*_{sn}, \upsilon)\ \mathrm{P}(A_{S;n}|l^*_{sn})\ \mathrm{P}(l^*_{sn}|\omega)\, dF(\upsilon, \omega) \tag{6.48}$$

To alleviate the computational difficulty in obtaining the conditional likelihood function, a simulation approach can be adopted to obtain an estimate of the likelihood function. Since after the estimation of the latent variable sub-models the sampling distribution F is known, R draws from F (i.e., $[\upsilon^{(r)}, \omega^{(r)}]$, for $r = 1, \ldots, R$) can be used to approximate the conditional likelihood function, i.e.,

$$\mathcal{L}_R(\Theta) = \frac{1}{R}\sum_{r=1}^{R}\left\{\sum_{s=1}^{S}\mathrm{P}(Y_n|l^*_{sn}, \upsilon^{(r)})\ \mathrm{P}(A_{S;n}|l^*_{sn})\ \mathrm{P}(l^*_{sn}|\omega^{(r)})\right\} \tag{6.49}$$

Under the usual regularity conditions of the integrand in equation (6.48),

$$\lim_{R\to\infty} \mathcal{L}_R(\Theta) = \mathcal{L}(\Theta). \tag{6.50}$$

If

$$\hat{\Theta}_R = \mathrm{argmax}\ \ \mathcal{L}_R(\Theta) \tag{6.51}$$

[25]See Appendix I for methods for the extraction of latent variables.

then the consistency and asymptotic normality of $\hat{\Theta}_R$ is ensured for sufficiently large R under some technical conditions (see Pakes and Pollard [1989]). The estimated standard errors of $\hat{\Theta}_R$ are still inconsistent due to simulation error.

Likelihood function for a Specific Case:

Assuming that the choice process generating Y_n, given that the individual is in latent class s and the explanatory variables X_n and X_n^*, and the class membership are governed by multinomial logit models, then

$$\mathrm{P}(y_{in} = 1 | l_{sn}^* = 1, X^*, X_n; \beta_s) = \frac{\exp(\beta_{xs}' X_{in} + \beta_{x^*s}' X_{in}^*)}{\sum_{j \in C_s} \exp(\beta_{xs}' X_{jn} + \beta_{x^*s}' X_{jn}^*)} \tag{6.52}$$

$$Q_s(Z^*, Z_n; \theta) = \frac{\exp(\theta_{zs}' Z_n + \theta_{z^*s}' Z_n^*)}{\sum_{s'=1}^{S} \exp(\theta_{zs'}' Z_n + \theta_{z^*s'}' Z_n^*)}. \tag{6.53}$$

Further, assuming the P_S indicators of the latent classes are independent conditional on the latent class, the likelihood function is written as:

$$\sum_{s=1}^{S} \left\{ \left(\prod_{p=1}^{P_S} g_{ps}(A_{S;pn} | l_{sn}^* = 1; \phi_{ps}) \right) \left[\int_{X^*} \int_{Z^*} \left(\prod_{i \in C_s} \left[\frac{\exp(\beta_{xs}' X_{in} + \beta_{x^*s}' X_{in}^*)}{\sum_{j \in C_s} \exp(\beta_{xs}' X_{jn} + \beta_{x^*s}' X_{jn}^*)} \right]^{y_{in}} \right) \right. \right.$$
$$\left(\frac{\exp(\theta_{zs}' Z_n + \theta_{z^*s}' Z_n^*)}{\sum_{s'=1}^{S} \exp(\theta_{zs'}' Z_n + \theta_{z^*s'}' Z_n^*)} \right) g_X(A_X | X^*; \Lambda_X) \; g_Z(A_Z | Z^*; \Lambda_Z) \; \cdot$$
$$\left. \left. f_{X^*,Z^*}(X^*, Z^* | W_n; B_X, B_Z, \Gamma_X, \Gamma_Z) \, dX^* \, dZ^* \right] \right\} \tag{6.54}$$

Once the model parameters are estimated, the analyst may be interested in obtaining the class sizes in the population. We need to extract or estimate the latent class probabilities, and to this end, we present different approaches for their extraction in Appendix J.

6.6 Summary

In this chapter, we extended the latent class choice model to incorporate class indicators wherein we viewed the class indicators as “attributes” of the latent class. We also advanced a general class of choice models called the latent structure choice model which incorporates attitudinal, perceptual and class indicators.

In chapter 7 we focus our attention on incorporating attitudes characterized through attitudinal indicators, and present various ways to capture taste variations stemming from variations in attitudes. We also present a case study in a shipper’s freight mode choice situation.

Chapter 7

Incorporating Attitudinal Data in Choice Models: Case Study – Application to Shipper's Freight Mode Choice Context

7.1 Introduction

The new competitive environment in the marketplace, increasing complexity of the logistics process, and innovations in the production and inventory control technologies have catalyzed the interest in shippers to view transportation as an important link in the supply chain. Consequently, shippers demand from the carriers (or suppliers of transportation services) specialized transportation service to meet these challenges. From the perspective of the carrier, its ability to design tailored services and to meet the changing needs of the shippers is all the more important as it faces increasing competition from other carriers. Further, transportation deregulation has changed the way business is conducted between shippers and carriers. The terms and conditions in a freight transportation contract include in addition to transportation rate (or price), detailed standards of the service to be provided and associated penalties in case of default of the contractual commitments (see for example, La Londe and Cooper [1989]).

In principle, the carrier with the ability to predict accurately the demand effects of changes in service levels can develop sound service design and marketing strategies to enhance its revenues at the expense of other carriers who lack this ability. In this regard, carriers can utilize a freight demand model, which is sensitive to service attributes, to analyze the effects of changes in service levels on the service demanded by shippers. Specifically, the demand model should represent the behavior of shippers at the disaggregate level to enable the identification of service design and marketing strategies.

In the remainder of this section, we highlight the developments in freight demand models in an effort to motivate our work[1]. The reader is directed to Vieira [1992] (see also Winston [1983] and Zlatopper and Austrian [1989]) for a more comprehensive literature review. Early freight demand studies used aggregate data on mode shares and characteristics of transportation modes.

The developments in discrete choice analysis in the 70's changed the focus to disaggregate freight demand models which utilize information from individual shipments. As in any discrete choice modeling exercise, it has been argued that disaggregate freight demand models provide more precise elasticity measures. Vieira [1992] categorizes disaggregate models into:

1. Models with *ad hoc behavioral specifications*: Herein the emphasis is on the freight mode choice decision, with the behavior of shippers derived from a random utility model (see Allen [1977], Daughety and Inaba [1979] and Winston [1981]).

2. Models with *logistic cost specification*: Herein the mode choice decision is linked to other decisions the firm has to make while coordinating production and distribution decisions. Baumol and Vinod [1970] analyzed freight demand as derived from the total logistic costs function of the firm by explicitly including the in-

[1] This review section relies heavily on the work of Vieira [1992].

ventory carrying costs and safety-stock needs. Chiang *et al.* [1981] extended the total logistic costs specification to include shipment size being simultaneously determined in the mode choice process. McFadden *et al.* [1985] proposed a discrete-continuous model to address the same problem.

More recent work by Vieira [1992] extends the traditional logistic cost minimization model to include shipper's perceptions of service and other intangible attributes. Vieira [1992] recognizes that shippers have different needs, and presents evidence that their responses towards changes in service quality differ considerably. Further, therein the emphasis is on the estimation of freight demand models from both revealed preference and stated preference data using the ideas developed in Morikawa [1989].

Given the focus of this case study on incorporating attitudinal data in freight demand models, it is instructive to note how such data has been utilized in the freight demand context. Several researchers interviewed shippers from different industries and tried to identify the importance ranking of factors affecting mode choice. Seidenfus [1985], who reviewed some such studies in Europe, observed that the importance of transportation rate reduced from first/second place in studies in 1959 and 1972, to fourth/sixth place in studies in 1983. While factors such as speed and reliability maintained their importance over time, a significant increase in importance of customer service and security was observed. The suggested trend was a shift in focus from price to quality of service. The emphasis was on the *qualitative* assessment of the shippers' attitudes and not on the construction of *quantitative* demand models which utilize such attitudinal data.

Shippers are diverse and demand specialized service to meet their transportation needs. Though trade-offs between service levels are different for each shipper, it is hypothesized that groups of shippers may have similar response behavior to changes in service levels. Consequently, segmenting the market of shippers is a valuable input in the carrier's effort to provide differentiated service. Specifically, the carriers can use *market segmentation* to tailor their service design strategies to meet the needs of certain segments, and transform untapped opportunities in others to enhance revenues/profits. The managerial rationale underlying market segmentation is the identification of groups of shippers who have similar needs and preference structures.

Traditionally, the bases of market segmentation are attributes of the goods or commodity being transported, such as density and value, and shipper's characteristics such as annual sales and volume of shipments. Specifically, the focus of attention is on the observable dimensions of shippers to aid in the identification of the market segments, measure the "size" of each market segment, and to be of strategic and operational value for service design and marketing. But the segments based on commodity and shipper's characteristics ignored the *underlying process* which governed the mode choice, and how shippers value different service attributes differently.

To this end, market segmentation studies have tried to use attitudinal indicators to identify the important factors/causes for differences in shippers' transportation mode choice process. Specifically two approaches have been used in market segmentation which utilize attitudinal indicators which include:

1. An approach which recognizes that differences in the shipper's attitudes towards service quality are the underlying motivation for defining market segments. The concept is operationalized through a clustering algorithm which builds "clusters" or segments of shippers having *similar attitudinal indicators*. Once segments are identified, the average attitudinal indicators in each segment aid in the interpretation and development of a service design and marketing strategy for that segment. An ad hoc comparison with the average characteristics of shippers in each segment is usually proposed as a means of identifying the segment in which a new shipper belongs, since attitudinal indicators are not available for the new shipper. A classification model, such as a discrete choice model potentially may also be used to represent the relationship between shipper's characteristics and segment assignment[2].

   McGinnis [1978] used 32 attitudinal questions regarding mode choice in a survey of transportation managers of major U.S. firms. Subsequent to market

[2]The classification model built in this manner has a contradiction of sorts, and to illustrate we relate it to the latent class model. Consider that the clusters are characterized by latent classes and one adopts a latent class model. Specifically, let $g(A|s)$ denote the conditional distribution of attitudinal indicators given the latent class s, and $\mathrm{P}(s|Z)$ denote the probability of a shipper

segmentation, service design strategies for each market segment were also discussed. Collision [1984] conducted a similar study specific to marine services, while Cooper and Rose [1986] studied the benefits of market segmentation to a particular carrier.

2. Realizing the importance of the link between observable characteristics that are strongly related to attitudes, in the industrial marketing literature, several approaches have been to suggested to improve the definition of market segments (see, for example, Wind and Cardozo [1974], Bonoma and Shapiro [1978]). The theme adopted is a macro-segmentation stage based on product's and firm's characteristics, followed by micro-segmentation stage based on the attitudinal indicators. Even here the segmentation is ad hoc without a behavioral link between observable characteristics and attitudinal indicators.

---

belonging to latent class s given shipper's characteristics Z. Then $f(A|Z)$ is written as:

$$f(A|Z) = \sum_{s=1}^{S} g(A|s)\ \mathrm{P}(s|Z) \tag{7.1}$$

If one ignores the effects of Z, we have the equivalent assumption of constant prior probability π_s of a shipper being in class s, and we obtain:

$$f(A) = \sum_{s=1}^{S} g(A|s)\ \pi_s \tag{7.2}$$

If in reality there is a causal assignment process as in equation (7.1) and the analyst instead uses equation (7.2), then the model parameters in the conditional distribution will be biased. The bias is expected to disappear if the sampling protocol is such that Z is randomly sampled from its population distribution $h(Z)$. This is because:

$$\begin{aligned} f(A) &= \int_Z f(A|Z)\ h(Z)\, dZ \\ &= \int_Z \sum_{s=1}^{S} g(A|s)\ \mathrm{P}(s|Z)\ h(Z)\, dZ \\ &= \sum_{s=1}^{S} g(A|s) \int_Z \mathrm{P}(s|Z)\ h(Z)\, dZ \\ &= \sum_{s=1}^{S} g(A|s)\ \pi_s \end{aligned} \tag{7.3}$$

since $\pi_s = \int_Z \mathrm{P}(s|Z)\ h(Z)\, dZ$.

More recently, Vieira [1992] has taken the view that information such as shipper's importance ratings of service attributes are manifestations of the shipper's sensitivity to attributes or latent attitudes. Shipper's characteristics are used to determine latent attitudes. The identification of market segments based on similarity in attitudes is suggested to provide the basis for differentiating the service offered to shippers. Specifically, market segmentation is operationalized by applying a clustering algorithm which builds clusters having similar *attitudes*. In this case, the classification model is derived implicitly from the clustering process.

Before we turn to the approach taken in this study, the above two approaches deserve critical assessment. We note that market segmentation is expected to be identified with reference to similarity in the preference structure and associated response to changes in service levels, and consequently, a freight mode choice model should play an integral part in the segmentation approach. But both approaches do not explicitly recognize this simple goal. Further, it is important to appreciate the notion that observed variations in attitudinal indicators for *all* the indicators *may not necessarily* be reflected in variations in preference structure since variations in attitudinal indicators may be due to response or measurement errors. Consequently, the focus of attention must be on the subset of the attitudinal indicators which are relevant. The second approach partially recognizes this notion by prescribing a causal representation for the generation of the attitudinal indicators through the existence of attitudes, and allowing for measurement errors in attitudinal indicators. In this approach, the operationalization of market segmentation entails the estimation of separate choice models in each cluster[3]. One might be tempted to proceed with the viewpoint that, if a causal representation is indeed important, then clustering using the causal variables such as shipper's characteristics would suffice. The pitfalls in this viewpoint are:

1. A priori there may not exist a behavioral/psychological theory to suggest which

[3]Vieira [1992] does not follow up the clustering procedure with separate estimation using observations in each cluster due to small sample sizes in each cluster, but instead estimates a single freight mode choice model, and proceeds on to interpret prediction tests in each of the clusters.

of the causal variables may be utilized in the clustering algorithm. Consequently, all the potential variables are used in the clustering algorithm, while in reality, only a subset of causal variables may be relevant.

2. Most clustering algorithms weigh equally variations among the causal variables. There may exist some causal variables, which are relevant from the perspective of market segmentation, but their variability may be limited in the sample, while shipper's preference structure may be sensitive to these variables, i.e., small changes in these variables may lead to substantial differences in the preference structure.

3. At an operational level, it must be noted the basic idea behind the clustering algorithm is the (dis)similarity in a multivariate space using a "distance-based" metric. This is intuitively acceptable if the causal variables are metrically scalable, while for categorical causal variables the use of a distance-based metric is debatable.

In this study, in principle we extend the work of Vieira [1992] by linking the choice model with an explicit causal model for attitude formation, and by specifying responses to attitudinal questions in surveys as indicators of attitudes. Methodologically, the work is a significant departure from previous work as we develop models wherein the freight transportation choice model forms an important sub-model. Further, we allow for the market segments to be characterized either on "continuous" scale of attitudes or "discretized" versions with finite groups of segments.

The remainder of this chapter is organized as follows: In section 7.2 we outline the modeling framework and develop the class of models for incorporating attitudinal data with particular reference to the shipper's freight mode choice context. In section 7.3 we discuss the data set utilized in our case study. In section 7.4 we present estimation results and elaborate on the substantive findings of the study.

7.2 A Class of Models for Capturing Attitudinal Indicators

In the latent structure choice model, we endeavored to incorporate a gamut of psychological factors such as attitudes, perceptions and other latent categorical concepts. In this section we restrict our attention to modeling approaches wherein only attitudinal data is utilized. Consequently, the models can be construed as special cases of the latent structure choice model. The key feature of these approaches is the use of shipper's attitudes or sensitivity to the service attributes, and which are manifested through attitudinal indicators such as shipper's importance ratings of the different attributes (including alternative-specific intrinsic preferences). Operationally, shipper's attitudes affect the taste parameters of the choice model rather judiciously.

The modeling framework in presented Figure 7-1. The framework assumes that Z_n^*, a $M \times 1$ vector of shipper's attitudes towards freight service attributes, is not observable and hence *latent*. Shipper's attitudes are determined by Z_n, a $Q \times 1$ vector of shipper's characteristics, which may include variables such as: number of employees, density of shipments, earliest acceptable delivery time, annual sales, maximum acceptable delay, electronic data interchange (EDI) usage, annual tonnage shipped, average length of haul and average price. The shipper's importance ratings on the following freight service attributes: transit time, reliability, rate, payment terms and billing, loss and damage, usability of equipment, and responsiveness – form the $P \times 1$ vector of attitudinal indicators A_n. The mapping from the shipper's characteristics to the indicators is postulated to capture the *attitude formation process*, and is referred to as the *attitude formation sub-model*.

The figure also illustrates a *choice sub-model*, which maps from the attributes of alternatives and shipper's characteristics, X_n, to the utility of alternative i, denoted U_{in}, and from the utilities to the choice indicator, y_{in}, as in any random utility model. The choice indicator may include shipper's shipment shares across competing carriers in particular corridors or origin-destination markets, and/or ratings, rankings and stated modal choices in hypothetical scenarios.

The important link between the attitude formation sub-model and the choice sub-model is provided through the mapping from the latent attitudes to the utilities through the taste parameters. Specifically, the shipper's attitudes are expected to affect the choice process through the sensitivities of shippers to freight service attributes, and consequently through the taste parameters of the freight mode choice model.

Assuming that the importance ratings are *metrically scalable* the sub-model relating the shipper's characteristics to the latent attitudes, and the latent attitudes to attitudinal indicators may be represented by a linear latent variable sub-model[4]. The attitude formation sub-model consists of: (1) structural sub-model which determines the shipper's attitudes, and (2) measurement sub-model which determines the indicators given attitudes. The structural sub-model is written as[5]:

$$Z_n^* = \Gamma Z_n + \zeta_n \tag{7.5}$$

where Γ is a $M \times Q$ parameter matrix and ζ_n is a $M \times 1$ random vector[6]. Further, we specify the measurement sub-model as:

$$A_n = \Lambda Z_n^* + \xi_n \tag{7.6}$$

where Λ is a $P \times M$ parameter matrix, and ξ_n is a $P \times 1$ random vector. Assume without loss of generality that the vectors A_n and Z_n are written as deviations from their respective means. In appendix C, the measurement model for Z_n^* is presented which recognizes the ordered categorical nature of indicators.

It must be noted that different relationships between the shipper's characteristics

[4]Specifically, we formulate a MIMIC model since we assume that the shipper's characteristics are perfectly measured.

[5]The structural sub-model may also be specified to capture structural relationships among the latent attitudes, i.e,

$$Z_n^* = B Z_n^* \Gamma Z_n + \zeta_n \tag{7.4}$$

where B is $M \times M$ parameter matrix.

[6]The number of attitudes postulated is usually much smaller than the number of shipper's characteristics, i.e., $M \ll Q$.

and the indicators can be represented by appropriate specification of the non-zero parameters in Γ and Λ. In particular, the analyst after characterizing the latent attitudes which are relevant in differentiating shipper's behavior, identifies the indicators that potentially "reflect" each latent attitude, and the relevant shipper's characteristics which may determine each latent attitude. For example, if the latent attitude characterizes shipper's cost sensitivity then the attitudinal indicators may include importance ratings of transportation rate, loss & damage, payment terms and billing, etc. Further, the shipper's characteristics which determine the latent attitude may include annual sales, annual tonnage, value of the commodity shipped, etc.

Given the specification of the attitude formation sub-model, we need to "link" it to the choice sub-model to reflect the effects of attitudes on the underlying preferences. Herein we postulate that the variations in the shipper's sensitivity to K freight service attributes can be "generated" through the M latent attitudes with $M \ll K$.

To highlight the main arguments for this postulate, consider a situation wherein the importance rating for each attribute forms the attitudinal indicator of an associated attribute sensitivity. Consequently, we have K attitudinal indicators and K attribute sensitivities. If the importance ratings are measured on a Likert-type scale with L levels, then we may assume that the indicators are ordered categorical variables. If the k^{th} indicator, A_{kn}, takes the level $l_k \in \{1, \ldots, L\}$, the measurement model which links the associated sensitivity Z^*_{kn} with the indicator, is based on a "threshold crossing" idea as in the ordinal probability model of McKelvey and Zavoina [1975], i.e.,

$$A_{kn} = \begin{cases} 1 & \text{if} \quad \vartheta^k_0 = -\infty < Z^*_{kn} \leq \vartheta^k_1 \\ 2 & \text{if} \quad \vartheta^k_1 < Z^*_{kn} \leq \vartheta^k_2 \\ 3 & \text{if} \quad \vartheta^k_2 < Z^*_{kn} \leq \vartheta^k_3 \\ & \vdots \\ L & \text{if} \quad \vartheta^k_{L-1} < Z^*_{kn} < \infty = \vartheta^k_L \end{cases} \tag{7.7}$$

where $[\vartheta^k_1, \ldots, \vartheta^k_{L-1}]$ form a set of estimable threshold parameters. Let Y_n denote the

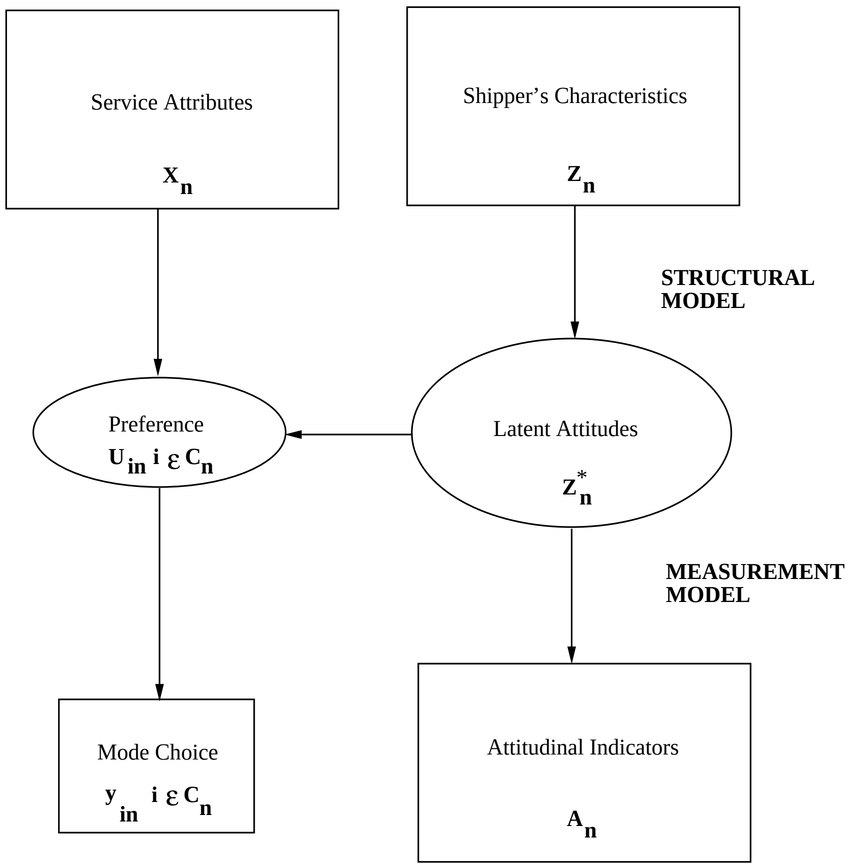


Figure 7-1: Framework for Incorporating Attitudinal Data in Freight Demand Model

choice indicator vector. If $A_n = [l_1, \ldots, l_K]'$, note that

$$\mathrm{P}(Y_n, A_n|Z_n, X_n) = \int \mathrm{P}(Y_n|X_n, Z^*)\ \mathrm{P}(A_n|Z^*)\ f(Z^*|Z_n)\, dZ^* \tag{7.8}$$

where $\mathrm{P}(Y_n|X_n, Z^*)$ is the choice model conditional on Z^*, $f(Z^*|Z_n)$ is the density function of Z^*, and $\mathrm{P}(A_n|Z^*)$ is an indicator function taking the value 1 if Z^* falls between two K-dimensional threshold vectors $\boldsymbol{\vartheta}^+ = [\vartheta^1_{l_1-1}, \ldots, \vartheta^K_{l_K-1}]'$ and $\boldsymbol{\vartheta}^- = [\vartheta^1_{l_1}, \ldots, \vartheta^K_{l_K}]'$, and zero otherwise. Hence, $\mathrm{P}(Y_n, A_n|Z_n, X_n)$ can be written as:

$$\int_{\boldsymbol{\vartheta}^+ \leq Z^* \leq \boldsymbol{\vartheta}^-} \mathrm{P}(Y_n|X_n, Z^*)\ f(Z^*|Z_n)\, dZ^* \tag{7.9}$$

First, it is difficult to estimate the model since the likelihood function entails a K-dimensional integral. Second, the number of parameters estimated may be quite large (i.e., each attribute sensitivity Z^*_{kn} is associated with a parameter vector Γ_k,[7] the parameters of the distribution[8] of ζ, and the threshold parameters ϑ).[9]

[7] Γ_k is the k^{th} row of Γ.

[8] The usual scaling restrictions apply to the distribution of ζ.

[9] If we allow structural relationships among the sensitivities to attributes, the structural sub-model is specified as:

$$Z^*_n = BZ^*_n + \Gamma Z_n + \zeta_n \tag{7.10}$$

where B is a $K \times K$ parameter matrix. The estimation of such a model may be conducted in four stages. In the first stage, we estimate the reduced form model of equation (7.10), i.e.,

$$Z^*_n = \Pi Z_n + \tilde{\zeta}_n \tag{7.11}$$

where $\Pi = (I - B)^{-1}\Gamma$, and $\tilde{\zeta}_n = (I - B)^{-1}\zeta_n$. The reduced form model is estimated consistently, equation by equation as in the ordinal probability model, by maximizing the marginal likelihood of the k^{th} indicator with respect to Π_k where Π_k is the k^{th} row of Π. In the second stage, the estimated reduced form equation is used to fit the sensitivities, $\hat{Z}^*_n$, which are utilized in equation (7.10) to estimate the structural parameters B and Γ consistently by maximizing the marginal likelihood of each indicator with respect to B_k and Γ_k, where B_k and Γ_k are the k^{th} rows of B and Γ, respectively. In the third stage, we estimate the reduced form covariance (i.e., $\tilde{\Sigma} = (I - B)^{-1}\Sigma[(I - B)^{-1}]'$ where Σ is the correlation matrix of ζ). Taking two indicators at a time, say A_{kn} and $A_{k'n}$, the covariance term, $\tilde{\sigma}_{kk'}$, is consistently estimated by maximizing the bivariate marginal likelihood, $\mathrm{P}(A_{kn} = l_k, A_{k'n} = l_{k'}|Z_n; \hat{\Pi}_k, \hat{\Pi}_{k'})$, with respect to $\tilde{\sigma}_{kk'}$. In the fourth and final stage, the choice model parameters are estimated by maximizing the conditional likelihood

$$\int \mathrm{P}(Y_n|X_n, Z^*) \quad dF(\hat{Z}^*|Z_n) \tag{7.12}$$

where $F(\hat{Z}^*|Z_n)$ is the sampling distribution of $\hat{Z}^*$. The conditional likelihood may be approximated

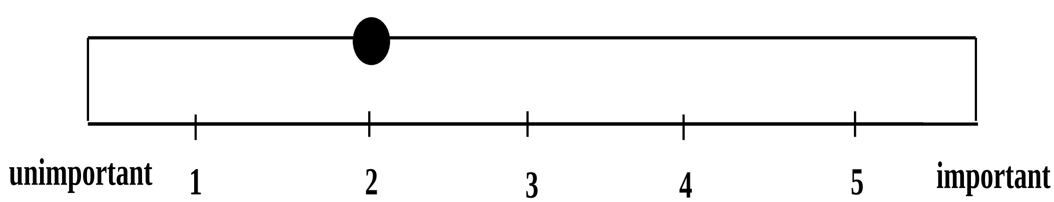
The shipper's sensitivity to two or more attributes may be interrelated, i.e., a shipper may have high (or low) sensitivity to two or more attributes. For example, a shipper with high sensitivity to transit time may also be expected to be highly sensitive to travel time reliability, or vice versa. Thus, postulating M latent attitudes, where $M \ll K$, allows the analyst to capture *prior* information or substantive knowledge of the choice context through the identification of a specific structure for the formation of shipper's sensitivity to attributes. Further, if the analyst assumes that the importance ratings are continuous, from the standpoint of model identification and estimation we need two or more indicators for each latent attitude. Consequently, the indicators which are specified to reflect a particular latent attitude include a subset of importance ratings of attributes.

In the remainder of this section, we present three approaches to link the attitude formation sub-model and the choice sub-model including:

- Random Coefficients Model with Latent Attitudes;
- Scaled Coefficient Choice Model; and
- Latent Class Choice Model for Taste Heterogeneity with Attitudinal Indicators.

We focus on attitudinal data collected as responses to the following types of questions:

`How important is attribute` k `to you?`



by Monte Carlo integration. This estimation procedure is derived from the approach proposed by Mallar [1977] for the estimation of simultaneous probability models. Similar estimation procedures may be developed following along the lines of other procedures for estimating simultaneous probability models (see Amemiya [1978], Lee [1981], and Sobel and Arminger [1992]).

In the traditional random coefficients model reviewed in section 2.5.1 the key idea adopted was that each shipper's parameter vector differs from the population mean by some "unobserved" amount. In our first approach, we extend the random coefficients framework to incorporate latent attitudes.

In the second approach, we postulate that the scale of the taste parameters depends on the latent attitudes. The basic idea of this approach is that the latent attitudes reflect sensitivity to different alternatives, and can be adequately incorporated as scale effects on shipper's taste parameters.

In the third approach, we hypothesize, as in the latent class choice model, that the underlying choice process varies systematically across a finite set of groups of shippers in the population, and to be homogeneous within each such group. Since each homogeneous group of shippers is unobserved, the groups are characterized by *latent classes*. The only exception being that we utilize the latent attitudes in the criterion functions of the class membership model.

Given this overview of the three approaches, we develop the three models in more detail.

7.2.1 Random Coefficients Model with Latent Attitudes

The utility of alternative i for shipper n depends on the observed vector of attributes of alternative i and shipper's characteristics, X_{in}. Using a linear functional form for the utility functions,

$$U_{in} = \beta_n' X_{in} + \epsilon_{in}, \quad \forall i \in C_n \tag{7.13}$$

where β_n is a $K \times 1$ taste parameter vector specific to shipper n. The shipper-specific parameter vector is written as:

$$\beta_n = \beta_0 + \Theta Z_n^* + \nu_n \tag{7.14}$$

where β_0 is a $K \times 1$ *base parameter* vector, Θ is a $K \times M$ parameter matrix which captures the "additive" effects of attitudes on taste parameters, and ν_n represents the $K \times 1$ vector shipper-specific idiosyncratic taste variations which even the latent attitudes fail to explain.

As the first step in the construction of the sample likelihood we are interested in obtaining the probability $\mathrm{P}(Y_n, A_n|Z_n, X_n; \beta_0, \Theta, \Gamma, \Lambda)$[10]. Assuming that the random vectors ϵ, ν, ζ, and ξ are independent, this probability equals

$$\int\limits_{Z^*}\int\limits_{\nu} \mathrm{P}(Y_n|Z^*, \nu, X_n; \beta_0, \Theta)\ f(A_n|Z^*; \Lambda)\ f(Z^*|Z_n; \Gamma)\, g(\nu)\, \mathrm{d}Z^*\, \mathrm{d}\nu \tag{7.15}$$

where $f(A_n|Z^*; \Lambda)$ is the distribution of the indicators given attitudes, $f(Z^*|Z_n; \Gamma)$ is the distribution of attitudes given Z_n, and $g(\nu)$ is the probability density of ν. The problem reduces to the specification of the choice sub-model denoted by $\mathrm{P}(Y_n|Z^*, \nu, X_n; \beta_0, \Theta)$. Assuming that the ϵ_{in}'s are independently and identically distributed Gumbel (0,1) across alternatives and shippers, this probability is expressed as:

$$\mathrm{P}(y_{in} = 1|Z^*, \nu, X_n; \beta_0, \Theta) = \frac{\exp((\beta_0 + \Theta Z^* + \nu)' X_{in}}{\sum\limits_{j \in C_n} \exp((\beta_0 + \Theta Z^* + \nu)' X_{jn})}. \tag{7.16}$$

We can test whether the effects of attitudes on the taste parameters are significant by testing $H_0 : \Theta = 0$. Note that when $\Theta = 0$ we have the traditional random coefficients model. The observation likelihood function involves a $(K + M)$-dimensional integral making the estimation non-trivial. By judiciously specifying the formation and manifestation of attitudes, and the effects of attitudes on the taste parameters (i.e., the fixed and free parameters of Θ), it may be reasonable to assume that the idiosyncratic taste variations are negligible[11]. Thus, when $\nu \equiv 0$, the probability of an observation equals

$$\int\limits_{Z^*} \mathrm{P}(Y_n|Z^*, X_n; \beta_0, \Theta)\ f(A_n|Z^*; \Lambda)\ f(Z^*|Z_n; \Gamma)\, \mathrm{d}Z^* \tag{7.17}$$

[10]For notational convenience only the structural parameters are explicitly denoted, while the parameters associated with the random components are implicitly assumed.

[11]Note that some degree of randomness in taste parameters is introduced due to the random component of Z^*.

After estimation of the model parameters, the choice model of interest is obtained as:

$$\mathrm{P}(Y_n|Z_n, X_n; \beta_0, \Theta, \Gamma) = \int\limits_{Z^*} \mathrm{P}(Y_n|Z^*, X_n; \beta_0, \Theta)\ f(Z^*|Z_n; \Gamma)\, \mathrm{d}Z^* \tag{7.18}$$

7.2.2 Scaled Coefficient Choice Model

Herein the utility of alternative i for shipper n depends on X_{in}, and is specified as:

$$U_{in} = \beta_n' X_{in} + \epsilon_{in}, \quad \forall i \in C_n \tag{7.19}$$

where β_n is a parameter vector specific to shipper n. This parameter vector is written as:

$$\beta_n = M_n \beta \tag{7.20}$$

where M_n is a positive and diagonal scaling matrix and β is the base parameter vector, i.e.,

$$M_n = \begin{pmatrix} \mu_{1n} & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \mu_{1n} & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \mu_{1n} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \mu_{Dn} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & \mu_{Dn} \end{pmatrix}$$

Assume that the parameter vector β is constructed by concatenating D sub-vectors, where $D \leq K$. Then β is written as:

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_D \end{pmatrix} \tag{7.21}$$

where β_d is the $K_d \times 1$ parameter vector corresponding to the d^{th} sub-vector with $\sum_{d=1}^{D} K_d = K$. It is instructive to understand the rationale for this construction. We argued earlier that the taste variations to two or more attributes may be interrelated, i.e., a shipper may have high (or low) sensitivity to two or more attributes. If each latent attitude is associated with a unique subset of the attributes, and consequently we assume that the latent attitude affects only the corresponding taste parameters, then the D sub-vectors correspond to the taste parameters of each subset. In this case, D equals M, the number of latent attitudes. In addition to the situation considered in the previous case, if a taste parameter is affected by two or more latent attitudes, then D is greater than M since a scale factor specific to the particular parameter must be specified.

The matrix M_n, essentially scales the d^{th} sub-vector by the scale factor μ_{dn}. Since we desire the scalars μ_{dn}'s to be positive as well as reflect the effects of attitudes, they are parameterized as:

$$\mu_{dn} = \exp(\alpha_d' Z_n^*) \quad \forall d = 1, \ldots, D. \tag{7.22}$$

The basic idea in the scaling approach is the incorporation of monotonic relationships between the parameters and the attitudes paying heed to the desirable feature that the signs of base parameters and the scaled parameters are identical. The significance of the effects of attitudes on the taste parameters can be checked by testing $H_0 : \alpha_d = 0 \, \forall d$.

The attitude formation sub-model representing the interrelationships between Z_n^*, Z_n and A_n is specified as in equations (7.5) and (7.6). Assuming that the random vectors ζ, ξ, and ϵ are independent, the probability of an observation, $\mathrm{P}(Y_n, A_n | Z_n, X_n; \beta, \alpha, \Gamma, \Lambda)$, equals

$$\int_{Z^*} \mathrm{P}(Y_n | Z^*, X_n; \beta, \alpha) \; f(A_n | Z^*; \Lambda) \; f(Z^* | Z_n; \Gamma) \, \mathrm{d}Z^* \tag{7.23}$$

The problem reduces to the specification of the choice sub-model, $\mathrm{P}(Y_n | Z^*, X_n; \beta, \alpha)$. Assuming that the ϵ_{in}'s are independently and identically distributed Gumbel (0,1) across alternatives and shippers, this probability is expressed as:

$$\mathrm{P}(y_{in} = 1|Z^*, X_n; \beta, \alpha) = \frac{\exp((M_n\beta)' X_{in})}{\sum\limits_{j \in C_n} \exp((M_n\beta)' X_{jn})} \tag{7.24}$$

We may also allow idiosyncratic taste variations as in the random coefficients model, and this may be operationalized along two approaches:

1. *Additive specification of randomness* wherein the shipper-specific parameter vector is written as:

   $$\beta_n = M_n\beta + \nu_n \tag{7.25}$$

   where ν_n is $K \times 1$ random vector which represents the idiosyncratic taste variations. As in the random coefficients model, the shipper-specific parameters may take on behaviorally implausible values if ν_n has unbounded support.

2. *Multiplicative specification of randomness* wherein the scale factors are parameterized as:

   $$\mu_{dn} = \exp(\alpha_d' Z_n^* + \nu_d) \tag{7.26}$$

   where ν_d is a random component associated with the d^{th} component of the taste parameter vector. Note that the multiplicative specification is more parsimonious, and unlike the additive specification, maintains the signs of the parameters even if ν_d has unbounded support.

7.2.3 Latent Class Choice Model for Taste Heterogeneity with Attitudinal Indicators

Herein the utility of alternative i for shipper n depends on X_{in} and the latent class s to which the shipper belongs. Using a linear functional form for the utility functions,

$$U_{isn} = \beta_s' X_{in} + \epsilon_{isn}, \quad \forall i \in C_n \tag{7.27}$$

where β_s is parameter vector specific to class s, with $s = 1, \ldots, S$. As in the latent class choice model, we postulate the class membership model and class-specific choice sub-model to operationalize the model. The attitude formation sub-model is specified as in equations (7.5) and (7.6).

The latent class is characterized by a D-dimensional binary vector. Let l^*_{sn} indicate the class membership of shipper n taking the value 1 if shipper belongs to class s and zero otherwise. We assume the existence of criterion functions, H_{dn}'s, which are specified as:

$$H_{dn} = \theta'_d Z^*_n + \delta_{dn}, \quad \forall d = 1, \ldots, D \tag{7.28}$$

By assuming a parametric distribution for δ and adopting a "threshold crossing" approach (i.e., d^{th} dimension of the latent class takes the value 1 if, and only if, $H_{dn} \geq 0$) for mapping from the criterion functions to the D-dimensional binary vector, $T_n = [l_1, \ldots, l_D]'$, the class membership model denoted by $\mathrm{P}(l^*_{sn} = 1 | Z^*; \theta)$ equals

$$\mathrm{P}\left((\delta_{dn} \geq -\theta'_d Z^*_n, \forall d | l_d = 1) \bigwedge (\delta_{dn} < -\theta'_d Z^*_n, \forall d | l_d = 0) \right) \tag{7.29}$$

Assuming that ζ, ξ, δ, and ϵ are independent, the probability of an observation, $\mathrm{P}(Y_n, A_n | Z_n, X_n; \beta, \theta, \Gamma, \Lambda)$, equals

$$\sum_{s=1}^{S} \mathrm{P}(Y_n | l^*_{sn} = 1, X_n; \beta_s) \left\{ \int_{Z^*} \mathrm{P}(l^*_{sn} = 1 | Z^*; \theta) \; f(A_n | Z^*; \Lambda) \; f(Z^* | Z_n; \Gamma) \, \mathrm{d}Z^* \right\} \tag{7.30}$$

7.3 Survey Data

In 1988[12], a major US railroad as part of an effort to determine the effects of changes in service quality on market share and revenues retained a marketing research firm (henceforth called ABC) to survey shippers and to determine their sensitivities to service and price. ABC's approach was based on service quality elasticities derived from conjoint experiments on several dimensions of service. The elasticities estimated then, reflected what shippers say they would do, and not on choices actually made.

[12] The discussion of the data is from a Vieira [1992] and must be referred to for more details.

The primary focus of the ABC study was to determine the demand elasticities of service attributes in the transport of five commodities: paper, aluminum, pet food, plastics and tires.

Decision makers from companies manufacturing the selected commodities were screened for participation. Besides having the responsibility to select the mode or carrier in different shipments, the decision maker should ship at least $1 Million of that commodity annually. Given 348 shippers meeting this criteria and agreeing to participate, the actual response rate was around 50% (166). Of these 166 respondents, complete data was available for 146 respondents.

The questionnaires were administered via mail due to the geographical dispersion of the selected shippers. As described in the following paragraphs, revealed and stated preferences were collected together with perceptions of mode service.

Revealed Preference Data

Revealed preference data describes the shipper's current practices regarding mode choice situations. For each shipper, information was collected for two major corridors used in outbound shipments of the selected commodities. These corridors were defined by shippers in terms of the origin and destination region, as well as the percentage of their annual tonnage shipped in that corridor.

Five modes were included in the questionnaire: truck only, single carrier rail, multiple carrier rail, intermodal (with at least one transshipment between truck and rail) and piggyback. In terms of service attributes, the shipper reported the typical rate per ton and "dock-to-dock" transit time by truck and rail in each corridor. Finally, questions about characteristics of shippers, or firms they represent, were also included in the survey. On one hand, information detailing the shipper's service requirements included acceptable early and late delivery times. On the other hand, annual sales, number of employees and the decision maker's experience in the position provided the demographics of shippers surveyed.

Stated Preference Data

Stated preference data elicited shippers' behavior in hypothetical transportation scenarios. Shippers were presented with two offerings in terms of transportation service features, and asked to rate their preference relative to the left or the right profile on a 9-point Likert type scale. In this scale, one or nine meant strong preference of one or the other profile, and five meant indifference between the offerings. The scenarios did not refer to any particular corridor and, in fact shippers were asked to think in general about their outbound shipments.

The transportation offerings or profiles in each scenario were described by the same attributes, but at different levels. In the ABC's study, each scenario used only two or three service attributes at a time to describe the profile, with the underlying assumption that the missing attributes were at the same level in both offerings.

Further, shipper's attitudes towards the different freight service attributes are manifested through corresponding "importance ratings" on a 5-point Likert-type scale.

Figure 7-2 presents an example of such a trade-off question. Each shipper answered forty trade-off questions, thinking about their outbound shipments of the selected commodity. The attributes and levels in each offering were chosen for each shipper based on his/her current perceptions of mode services to keep the realism of the question, while trying to better elicit the trade-offs among service attributes.

In this case study we focus our attention on the stated preference data and the use of importance ratings, and consequently the variables relevant for our analysis are described in Table 7.1 in detail.

Table 7.2 presents a descriptive summary of the importance ratings of shippers in the survey. On average transit time reliability is considered as the most important service dimension. As seen in the frequency distribution of ratings for each service dimension, shippers may tend to overstate their importance ratings, appearing that all service attributes are very important to all shippers. This seemingly makes the process of market segmentation based on differences in attitudinal indicators inappropriate. We take the view that the exact location of the scale of the importance ratings is inconsequential. We pose the conceptual question whether it is possible to explain a component of the observed variations in importance ratings through the shipper's

Which service offering would you prefer?

10% lower than average rate I pay now 80% of shipments arrive when I want them to	10% higher than average rate I pay now 90% of shipments arrive when I want them to

1	2	3	4	5	6	7	8	9
Strongly Prefer Left								Strongly Prefer Right

Source: ABC Study

Figure 7-2: Example of Trade-off Question latent attitudes, and consequently through variations in shipper's characteristics. If so, the latent attitudes are postulated to play an important part in how shippers weigh the different freight service attributes in the freight mode choice process.

7.4 Estimation Results

In this section we present different shipper's preference models estimated on the stated preference data. Given our intention to assess the extent of taste variations in the shipper's freight choice process through shipper's attitudes, we assume for simplicity a *linear specification* vis-a-vis a logistic cost specification of the systematic utility function as seen in Vieira [1992]. Further, the estimated models for incorporating importance ratings are categorized into:

1. Models with a single latent attitude: *Overall Attribute Sensitivity*.

2. Models with two latent attitudes: *Time Sensitivity* and *Cost Sensitivity*.

Further, in the estimated models we *do not allow* for any idiosyncratic taste variations, and thus restrict our attention to taste variations stemming from differences in shippers' attitudes.

7.4.1 Model 0: Ordinal Probit Model

This is the simplest of the estimated models. Since the preference ratings are measured on a 9-point Likert type scale, we estimate an ordinal probit model with symmetric thresholds. The symmetry restriction was imposed to reflect the nature of the response scale since there is no reason to assume that strongly preferring one transportation offering is different from strongly preferring the other one since the order of presentation of the alternatives is arbitrary. Thus we estimate only four thresholds

Variable Type	NAME	DESCRIPTION
Profile related Variables	**PREFER**	Preference indicator (1-9)
	LRTT	Transit time (days)[a]
	LRCTT	Consistency of transit time[b]
	LRRATE	Freight rate ($/ton)[c]
	LRPTB	Payment terms & billing[d]
	LRLOD	Loss and damage[e]
	LRUEQ	Usability of equipment[f]
	LRRSP	Responsiveness[g]
Shipper's Characteristics	**NEMP**	Number of employees (10^3)
	DENSITY	Average density of shipments (10^2 ton/m^3)
	EARLYD	Earliest acceptable delivery time (day)
	SALES	Annual sales (10^3 Million $)
	LATED	Maximum acceptable delay (day)
	EDI	Shipper uses EDI
	TONNES	Annual tonnage shipped (10^5 ton)
	AVHAUL	Average length of haul (10^3 miles)
	ADAVPR	Average price of shipment ($/ton)
Importance Ratings (1-5)	**RTIME**	Transit time
	RCTIME	Consistency of transit time
	RRATE	Freight rate
	RPTB	Payment terms and billing
	RLOD	Loss and damage
	RUEQ	Usability of equipment
	RRSP	Responsiveness
	RLOE	Level of effort

[a]Transit time varies from 50% faster to 50% slower than average dock to dock time at present.

[b]Consistency is defined as the fraction of shipments which arrive when the shipper wants to, and this fraction varies between 0.4 and 0.95.

[c]Rate varies from 50% higher to 50% lower than average present rates.

[d]Loss and damage is defined as % of shipment value lost or damaged, and it varies from 0.1% to 3%.

[e]Usability of equipment is defined as the fraction of time sufficient quantity of acceptable equipment is provided, and it varies from 0.5 to 0.99.

[f]Payment terms and billings is defined as the fraction of time payment terms and billings are satisfactory, and it varies from 0.25 to 1.0.

[g]Responsiveness is defined as the fraction of time the carrier satisfactorily responds to inquiries, and it varies from 0.25 to 1.0.

Table 7.1: Names and Definition of Variables – Freight Demand Study

Service Attribute	**Average**	**Frequency Distribution** Essential **1**	**2**	**3**	**4**	Not Import. **5**
Transit Time	1.48	88	47	10	1	-
Reliability	1.28	110	32	3	1	-
Rate	1.53	84	49	11	1	1
Payment Terms & Billing	1.64	70	61	13	1	1
Loss and Damage	1.33	102	40	4	-	-
Usability of Equipment	1.42	100	36	6	3	1
Responsiveness	1.42	88	54	4	-	-
Level of Effort	1.58	76	58	10	1	1

1=essential; 2=very important; 3=somewhat important; 4=not very important; 5=not important at all.

Table 7.2: Importance Ratings for Different Service Dimensions

parameters. The utility function is specified as[13]:

$$\begin{aligned} U &= \beta_1 LRTT + \beta_2 LRCTT + \beta_3 LRRATE + \beta_4 LRPTB + \beta_5 LRLOD + \\ &\qquad \beta_6 LRUEQ + \beta_7 LRRSP + \epsilon \end{aligned}$$

where[14] $\epsilon \sim \mathcal{N}(0,1)$. The association between the utility function, the threshold parameters and the ordinal response is illustrated in Figure 7-3. Specifically, the response probabilities are given by:

$$\begin{aligned} \mathrm{P}(y=1) &= \Phi(-\kappa_4 - \beta' X) \\ \mathrm{P}(y=2) &= \Phi(-\kappa_3 - \beta' X) - \Phi(-\kappa_4 - \beta' X) \\ \vdots &= \vdots \\ \mathrm{P}(y=5) &= \Phi(\kappa_1 - \beta' X) - \Phi(-\kappa_1 - \beta' X) \\ \vdots &= \vdots \\ \mathrm{P}(y=9) &= 1 - \Phi(\kappa_4 - \beta' X) \end{aligned} \tag{7.31}$$

where X is the attribute vector (or rather the attribute difference vector) and $\Phi(\cdot)$ is the cumulative distribution function of standard normal variate. Whenever an attribute was not included in a given scenario, it was assumed that it was at the same level in both offerings.

The estimated model is presented in Table 7.3. All the coefficients are significant. The standard errors are calculated from the estimated information matrix. Further, to estimate the standard errors correctly as multiple responses from the same shipper are likely to be correlated, we utilize the variance-covariance matrix for extremum estimators (Amemiya [1985]), and consequently we refer to the corrected t-statistics in conducting simple hypothesis tests. As expected the transit time and transportation rate coefficients are positive. The consistency of transit time, which is measured as

[13] All the variables are in difference form, i.e., attribute level of left profile - attribute level of right profile.

[14] Fixing the variance of ϵ is necessary to set the scale of the utility function.

Choice Model

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.
Service Attributes	LRTT	0.072	0.006	11.79	11.60
	LRCTT	-3.642	0.210	-17.35	-16.98
	LRRATE	0.021	0.001	18.22	19.08
	LRPTB	-0.388	0.035	-11.13	-11.55
	LRLOD	0.228	0.011	20.83	21.23
	LRUEQ	-1.280	0.074	-17.38	-18.11
	LRRSP	-0.337	0.038	-8.98	-10.06
Threshold Parameters	κ_1	0.249	0.007	35.11	35.18
	κ_2	0.530	0.010	54.41	54.07
	κ_3	0.850	0.012	70.97	69.69
	κ_4	1.286	0.015	84.71	82.55

Log-likelihood at zero= -12831.79
Log-likelihood at convergence= -12234.16
$\bar{\rho}^2 = 0.046$
Number of observations = 5840

Table 7.3: Model 0: Ordinal Probit Model

the fraction of shipments arriving in time, has a large negative coefficient reflecting a strong preference for reliable freight service[15]. The significance of coefficients of loss and damage, usability of equipment and responsiveness reflect the importance of non-traditional of service-quality in the choice process. Further, it must be noted that the service-quality attributes in order of decreasing importance in terms of shipper's sensitivity are: consistency of transit time, usability of equipment, payment terms and billing and responsiveness.[16]

7.4.2 Single Attitude Models

Herein we assume the existence of a single latent attitude which represents an "overall attribute sensitivity" to service attributes. The path diagram for the attitude

[15]This is in comparison to other attributes such as payment terms and billing, usability of equipment, and responsiveness which have the same "units" as fractions.
[16]See footnote 15.

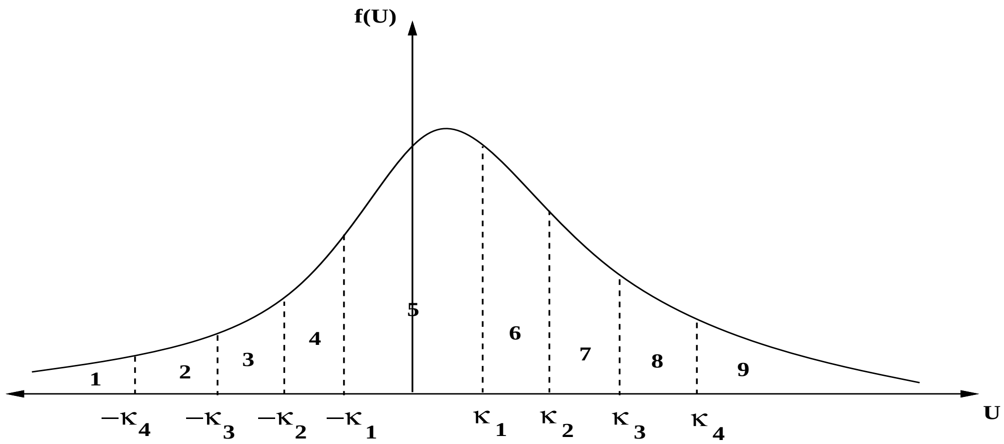


Figure 7-3: Illustration of Utility Function and Symmetric Threshold Parameters

formation sub-model is represented in Figure 7-4. Shipper's characteristics such as number of employees, density of shipments, earliest acceptable delivery time, annual sales, maximum acceptable delay, EDI usage, annual tonnage shipped, average length of haul and average price are postulated to determine the shipper's attitude. All the importance ratings are utilized as indicators of the latent attitude, and we assume that they are continuous. Assuming that the shipper's characteristics and importance ratings are written as deviations from their respective means, the structural and measurement sub-models of the attitude formation sub-model are specified as:

Structural Sub-model

$$\begin{aligned} Z^* \;=\;& \gamma_1 NEMP + \gamma_2 DENSITY + \gamma_3 EARLYD + \gamma_4 SALES + \gamma_5 LATED + \\ & \gamma_6 EDI + \gamma_7 TONNES + \gamma_8 AVHAUL + \gamma_9 ADAVPR + \zeta \end{aligned}$$

Measurement Sub-model

$$\begin{aligned} RTIME &= Z^* + \xi_1 \\ RCTIME &= \lambda_2 Z^* + \xi_2 \end{aligned}$$

$$
\begin{aligned}
RRATE &= \lambda_3 Z^* + \xi_3 \\
RPTB &= \lambda_4 Z^* + \xi_4 \\
RLOD &= \lambda_5 Z^* + \xi_5 \\
RUEQ &= \lambda_6 Z^* + \xi_6 \\
RRSP &= \lambda_7 Z^* + \xi_7 \\
RLOE &= \lambda_8 Z^* + \xi_8
\end{aligned}
$$

The scale[17] of the latent attitude is set to that of the importance rating for transit time. Further, we assume that the random components of the measurement model are independent.

We also assume that ζ, ξ and ϵ are independent. The distributions of ζ and ξ are specified as:

$$\zeta \sim \mathcal{N}(0, \sigma_\zeta^2) \tag{7.32}$$

and

$$\xi_p \sim \mathcal{N}(0, \sigma_{\xi_p}^2) \quad \forall p = 1, \ldots, 8. \tag{7.33}$$

In the three models presented in this section, the above latent variable model forms a sub-model. Consequently, as a first step in the maximum likelihood estimation of these models, we estimated the one latent attitude model (ignoring the choice sub-model) and used the estimated parameters as starting values for the corresponding parameters of the latent variable sub-model of the complete model system which consists of the attitude formation sub-model and the choice sub-model given the latent attitude[18]. In each of the subsequent models, we outline the specification of the choice model with special attention to the judicious manner in which the overall attribute sensitivity induces heterogeneity in taste parameters.

[17] Since the latent variable is unobserved it does not have a definite scale. Hence it is *necessary* to fix one parameter in the column of the Λ matrix in the measurement model to unity. This defines the unit of measurement of the latent variable to be the same as the corresponding indicator.

[18] Estimation results for the one latent attitude sub-model are presented in Appendix B. It must be noted that this MIMIC model was estimated using MLE under the assumption of multivariate normal vector $[A, Z]$ using the LINCS program, instead of the conditional likelihood $f(A|Z)$.

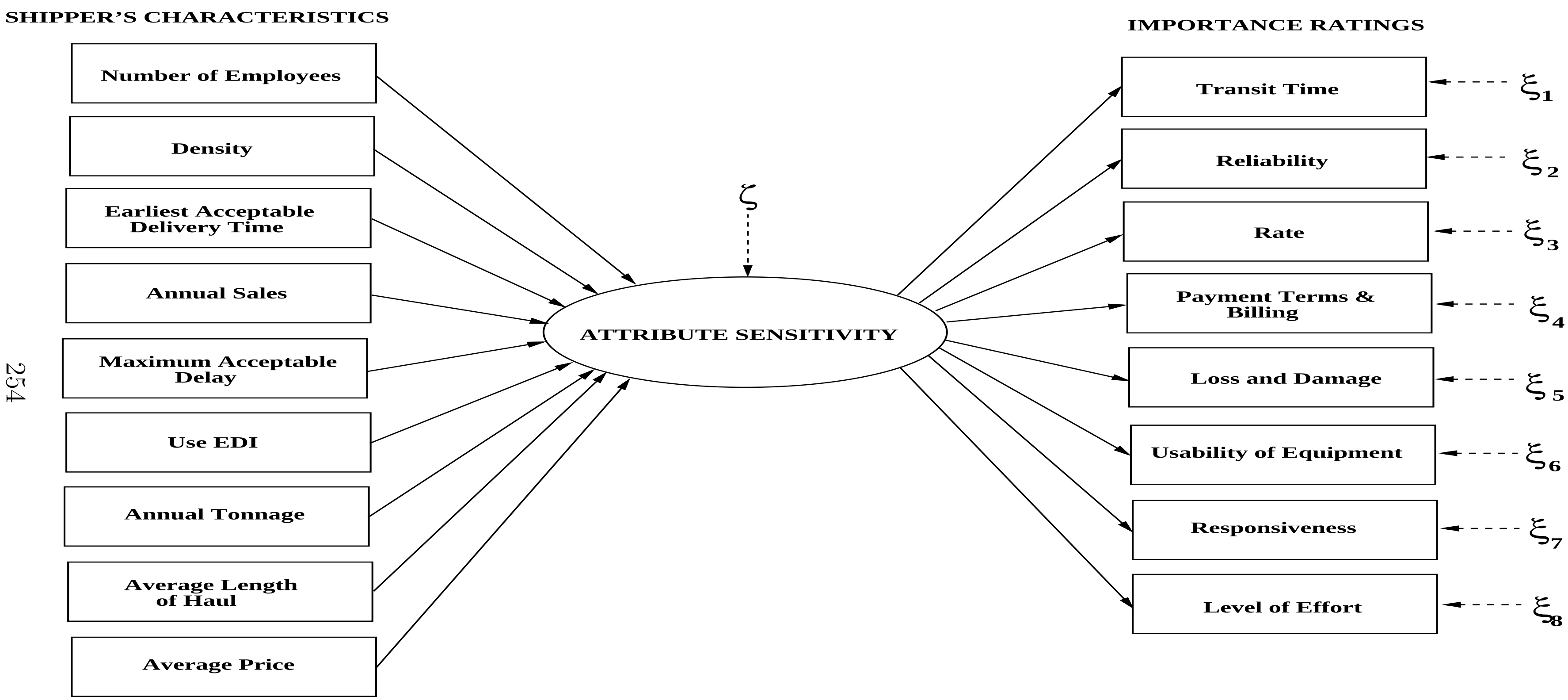


Figure 7-4: Path Diagram for One Attitude Model

Model 1-1: Choice Model with Interaction Variables between Attributes and Attitude

In this model the utility function of the choice sub-model is specified as in Model 0, i.e.,

$$\begin{aligned} U &= \beta_1 LRTT + \beta_2 LRCTT + \beta_3 LRRATE + \beta_4 LRPTB + \beta_5 LRLOD + \\ & \qquad \beta_6 LRUEQ + \beta_7 LRRSP + \epsilon \end{aligned}$$

But instead of fixed taste parameters, we capture the effects of the latent attitude on the taste parameters through additive effects on a set of "base taste parameters". Specifically, we have $\beta_k = \beta_{0,k} + \alpha_k Z^*, \quad \forall k = 1, \ldots, 7$ where $\beta_{0,k}$ is the base parameter for attribute k and α_k captures the effect of the latent attitude, and hence referred to as "taste modifiers".

The estimated choice sub-model and the structural sub-model of the attitude formation sub-model are presented in Table 7.4. The choice model parameters in Model 0 and the estimated base parameters in Model 1-1 are quite similar. Except for the taste modifiers associated with consistency of transit time, payment terms & billing and usability of equipment, all others are insignificant. Noting the monotonic relationship between the latent attitude and the importance ratings, an increase in the latent attitude is reflected through a decrease in the magnitude of the taste parameters[19]. Consequently, the signs of the estimated taste modifiers are expected to be opposite to the signs of the base parameters[20]. This is indeed the case for all the taste modifiers (including the insignificant ones).

Now we turn our attention to the attitude formation process as represented by the structural sub-model. In general, since all the variables in the structural model

[19]Note that the importance ratings range from essential to unimportant, and consequently higher values of latent attitude are associated with lower sensitivity to attributes.

[20]By utilizing the "fitted" latent attitude from the structural sub-model, the signs of the taste parameters were behaviorally acceptable in the sample. Note that the latent attitude has unbounded support as ζ is assumed be a normal random variable, and this naturally implies that the taste parameters will take on behaviorally implausible values. Restricting the randomness of ζ to within 3 standard deviations, the occurrence of such parameter values in the sample appear negligible.

are positive, a positive coefficient for a specific characteristic implies that the latent attitude *increases* with an increase in the corresponding variable, and hence the sensitivity to attributes *decreases*. As the number of employees increase the shipper's sensitivity to freight service attributes decreases. This may be due in part to shippers with larger workforce have distributed production and warehousing facilities leading to lesser sensitivity to freight service attributes. On the other hand, the negative coefficient for annual sales indicates that firms with higher annual sales are more sensitive. Shippers with higher acceptable delays are less sensitive and this notion is reflected in the associated positive coefficient. Surprisingly, users of EDI are less sensitive as one would expect them to be more sensitive to freight service attributes, especially service-quality attributes such as payment terms and billing, responsiveness, etc. The negative coefficients for average length of haul and average price of shipment indicate that shippers transporting high value goods over longer distances are more sensitive. Annual tonnage shipped and early acceptable delivery time do not seem to have an affect on the shipper's sensitivity. The fit of the structural model as measured by the squared multiple correlation[21] is 0.08. Although this measure may not suggest good fit, the model does have the unique capability of providing useful insights as to how shipper's characteristics relate to shipper's sensitivity to freight service attributes.

The estimated measurement sub-model is presented in Table 7.5. All the parameters in Λ are significant implying that the latent attitude does explain part of the observed variations in the attitudinal indicators.

The log-likelihood of the complete model, which includes the observed preference response and the attitudinal indicators, is -54227.15. Using the structural sub-model and the choice sub-model, the log-likelihood of the choice model component[22]

[21]A useful measure of fit of the individual equations of the model system is a measure similar to the R^2 in linear regression. This measure is called the *squared multiple correlation* (SMC), and is defined for each equation in the latent variable sub- model, whether it is a structural equation or a measurement equation, as follows:

$$SMC = 1 - \frac{\texttt{error variance}}{\texttt{variance of variable on LHS of equation}} \tag{7.34}$$

.

[22]It must be noted that the main objective of the whole exercise of incorporating attitudinal data

Choice Sub-model

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.
Base Parameters	LRTT	0.072	0.006	11.78	10.38
	LRCTT	-3.705	0.213	-17.41	-16.14
	LRRATE	0.021	0.001	18.28	18.34
	LRPTB	-0.388	0.035	-11.10	-11.89
	LRLOD	0.229	0.011	20.91	18.76
	LRUEQ	-1.289	0.074	-17.42	-16.49
	LRRSP	-0.340	0.038	-9.05	-10.56
Taste Modifiers	LRTT	-0.018	0.017	-1.06	-0.71
	LRCTT	1.164	0.541	2.15	2.74
	LRRATE	0.003	0.003	0.89	0.34
	LRPTB	0.195	0.095	2.05	2.35
	LRLOD	-0.039	0.029	-1.32	-1.11
	LRUEQ	0.650	0.197	3.31	2.94
	LRRSP	0.077	0.101	0.76	0.56
Threshold Parameters	κ_1	0.249	0.007	35.11	34.97
	κ_2	0.531	0.010	54.40	52.89
	κ_3	0.852	0.012	70.93	68.21
	κ_4	1.289	0.015	84.64	83.24

Structural Sub-model

	Parameter	Estimates	Std. err.	t-stat	t-stat corr
Structural Parameters	NEMP	0.012	0.002	6.23	5.92
	DENSITY	-0.055	0.031	-1.76	-1.56
	EARLYD	0.005	0.008	0.64	0.79
	SALES	-0.046	0.004	-12.44	-12.93
	LATED	0.064	0.009	7.34	6.12
	EDI1	0.113	0.014	8.36	7.91
	TONNES	-9×10^{-5}	6×10^{-4}	-0.14	-0.19
	AVHAUL	-0.087	0.015	-5.91	-5.72
	ADAVPR	-0.023	0.006	-3.89	-3.41
Noise Parameter	σ_ζ	0.407	0.009	44.16	42.90

Squared multiple correlation = 0.08

Table 7.4: Model 1-1: Model with Interaction Variables between Attribute and Attitude

is -12230.11. Compared to Model 0, this model improves on the choice model component only marginally. On the other hand, the model provides valuable insight into the varying sensitivities of shippers to freight service attributes which may be of strategic importance. Further, a more parsimonious model can be estimated by eliminating insignificant taste modifiers.

It must be noted that one of the drawbacks of the modeling approach is that since Z^* is a random variable with unbounded support, the taste parameters with non-zero taste modifiers may take on behaviorally implausible values with positive probabilities.

Model 1-2: Latent Class Choice Model with Attitudinal Indicators

We postulate the existence of two groups of shippers with differing sensitivities to service attributes. The assignment of a shipper to the latent classes is governed by a threshold crossing model with a criterion function specified as[23]:

$$H = \theta Z^* + \delta \tag{7.35}$$

where[24] $\delta \sim \mathcal{N}(0, 1)$. Then the probability of the shipper being in class 1 is given by $\Phi(-\theta Z^*)$.

The estimated choice sub-model and the structural sub-model of the attitude formation sub-model are presented in Table 7.6. The choice model parameters in Model 0 are sandwiched between corresponding parameters in the two classes, and all the parameters are significant and have expected signs. The magnitudes of the taste parameters in class 1 are greater than the corresponding parameters in class 2, and hence class 1 is interpreted as the high sensitive segment and class 2 as the low sensitive segment. Further, the class-specific parameters of consistency of transit time and usability of equipment show qualitative differences across the two classes[25]. A

is to have a "better" choice model, and from which stems our need to check the improvement, if any, of the preference model.

[23]A constant in the criterion function is not specified since Z^* is in deviation form.

[24]Fixing the variance of δ is necessary to set the scale of the criterion function.

[25]In fact, t-tests comparing the equality of two corresponding parameters at a time across classes

Measurement Sub-model

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.	SMC
Structural Parameters	λ_1	1.000	–[a]	–	–	0.41
	λ_2	0.639	0.020	32.39	31.24	0.25
	λ_3	0.538	0.029	18.74	18.91	0.10
	λ_4	0.885	0.031	28.97	27.38	0.26
	λ_5	0.708	0.023	30.20	30.56	0.32
	λ_6	0.748	0.030	25.07	24.22	0.19
	λ_7	0.582	0.023	25.14	25.61	0.20
	λ_8	0.933	0.032	29.59	28.12	0.30
Noise Parameters	σ_{ξ_1}	0.501	0.006	77.35	76.22	–
	σ_{ξ_2}	0.460	0.005	91.79	89.91	
	σ_{ξ_3}	0.687	0.007	103.16	102.43	
	σ_{ξ_4}	0.626	0.007	92.99	92.87	
	σ_{ξ_5}	0.432	0.005	87.85	86.76	
	σ_{ξ_6}	0.657	0.007	98.66	97.89	
	σ_{ξ_7}	0.489	0.005	96.84	96.23	
	σ_{ξ_8}	0.591	0.007	89.56	89.41	

[a]Fixed parameter.

Complete Log-likelihood at convergence = -54227.15
Log-likelihood of choice model at zero= -12831.79
Log-likelihood at choice model at convergence= -12230.11
$\bar{\rho}^2 = 0.045$
Number of observations = 5840

Table 7.5: Model 1-1: Model with Interaction Variables between Attribute and Attitude (cont'd)

similar pattern of the significance of the effects of latent attitude on the same set of parameters was observed in Model 1-1.

In the structural sub-model, the estimated coefficients are similar to those in Model 1-1 with one exception. In Model 1-2 we have a significant and positive coefficient for annual tonnage indicating that high volume shippers are less sensitive to freight service attributes. At this time, it is instructive to note an important caveat of our modeling approach which explicitly links the choice model with the attitude formation model. The (mis)specification of the choice sub-model will affect[26] the estimated attitude formation sub-model since the preference response can be construed as another "indicator" of latent attitude. Further, compared to Model 1-1, the fit of the structural sub-model is marginally better.

In Table 7.7 the estimated criterion function and the measurement sub-model are presented. The coefficient θ is significant and positive indicating the presence of effect of latent attitude in the class membership model[27]. Since larger values of latent attitude (and lower attribute sensitivities) are associated with higher importance ratings, and class 1 is the high sensitive class compared to class 2, θ is expected to be positive. The estimated measurement sub-model is similar to that of Model 1-1.

The log-likelihood of the complete model, which includes the observed preference response and the attitudinal indicators, is -54187.14, which betters Model 1-1 by 40 units with the addition of only 1 parameter. Using the structural sub-model, the estimated criterion function, and the choice sub-model the log-likelihood of the choice model component is -12219.06 which betters that of Model 1-1 by 11 units with an additional parameter. It must be noted this model addresses the drawback of the previous approach as the taste parameters are fixed within each class, and thus behaviorally implausible values are unlikely.

at 5% significance revealed that class-specific parameters of consistency of transit time, payment terms & billing, and usability of equipment are different across classes, while all other parameters are the same.

[26]Of course, the reverse argument also holds as the (mis)specification of the attitude formation sub-model affects choice sub-model.

[27]It must be noted that when $\theta \equiv 0$, we have two classes with *every* shipper having equal probability of being in each class.

Choice Sub-model

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.
High Sensitive Segment	LRTT	0.095	0.018	5.23	5.14
	LRCTT	-5.236	0.715	-7.33	-6.78
	LRRATE	0.020	0.002	8.73	7.89
	LRPTB	-0.538	0.088	-6.12	-6.34
	LRLOD	0.272	0.029	9.32	8.91
	LRUEQ	-1.756	0.181	-9.71	-9.67
	LRRSP	-0.432	0.080	-5.39	-5.23
Low Sensitive Segment	LRTT	0.050	0.017	2.95	2.34
	LRCTT	-2.448	0.472	-5.19	-5.93
	LRRATE	0.021	0.002	9.12	8.88
	LRPTB	-0.246	0.083	-2.97	-2.36
	LRLOD	0.191	0.026	7.34	7.11
	LRUEQ	-0.847	0.155	-5.47	-5.32
	LRRSP	-0.257	0.077	-3.34	-3.83
Threshold Parameters	κ_1	0.251	0.007	34.61	32.31
	κ_2	0.534	0.010	52.61	50.89
	κ_3	0.857	0.013	67.06	66.27
	κ_4	1.298	0.017	78.03	77.56

Structural Sub-model

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.
Structural Parameters	NEMP	0.010	0.001	7.36	7.12
	DENSITY	-0.041	0.028	-1.47	-1.11
	EARLYD	-0.004	0.006	-0.69	-0.52
	SALES	-0.030	0.004	-7.35	-7.98
	LATED	0.056	0.007	7.95	7.35
	EDI	0.099	0.012	8.58	7.91
	TONNES	0.001	4×10^{-4}	2.84	3.02
	AVHAUL	-0.085	0.013	-6.41	-6.07
	ADAVPR	-0.039	0.005	-8.29	-7.71
Noise Parameter	σ_ζ	0.340	0.006	53.38	52.79

Squared multiple correlation of structural equation= 0.09

Table 7.6: Model 1-2: Latent Class Choice Model with Attitudinal Indicators

Criterion Function

Parameter	Estimates	Std. err.	t-stat	t-stat corr.
θ	2.280	1.103	2.07	2.23

Measurement Sub-model

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.	SMC
Structural Parameters	λ_1	1.000	–[a]	–	–	0.33
	λ_2	0.721	0.023	32.03	31.29	0.23
	λ_3	0.642	0.032	20.01	19.78	0.10
	λ_4	1.063	0.034	30.83	29.42	0.27
	λ_5	0.875	0.027	32.80	32.93	0.35
	λ_6	0.918	0.035	26.12	26.31	0.20
	λ_7	0.713	0.025	28.30	26.97	0.21
	λ_8	1.147	0.034	33.54	33.13	0.33
Noise Parameters	σ_{ξ_1}	0.535	0.005	104.66	102.21	–
	σ_{ξ_2}	0.468	0.005	96.79	96.71	
	σ_{ξ_3}	0.686	0.007	103.66	102.48	
	σ_{ξ_4}	0.623	0.007	92.49	91.64	
	σ_{ξ_5}	0.423	0.005	82.13	81.02	
	σ_{ξ_6}	0.652	0.007	94.70	95.11	
	σ_{ξ_7}	0.485	0.005	95.97	96.77	
	σ_{ξ_8}	0.581	0.007	87.51	86.89	

[a]Fixed parameter.

Complete Log-likelihood at convergence = -54187.14
Log-likelihood of choice model at zero= -12831.79
Log-likelihood at choice model at convergence= -12219.06
$\bar{\rho}^2 = 0.046$
Number of observations = 5840

Table 7.7: Model 1-2: Latent Class Choice Model with Attitudinal Indicators (cont'd)

Model 1-3: Scaled Coefficient Choice Model

As before the utility function of the choice sub-model is specified as:

$$\begin{aligned} U &= \beta_1 LRTT + \beta_2 LRCTT + \beta_3 LRRATE + \beta_4 LRPTB + \beta_5 LRLOD + \\ &\quad \beta_6 LRUEQ + \beta_7 LRRSP + \epsilon \end{aligned}$$

where $\beta_k = \beta_{0,k}\mu, \;\; \forall k = 1, \ldots, 7$, with $\beta_{0,k}$ the base parameter for attribute k and the scale factor is given by $\mu = \exp(\alpha Z^*)$. The main advantage of this approach is the relative parsimony of specification[28], and since the latent attitude is a generic sensitivity variable, the scaling approach is behaviorally sensible.

The estimated sub-choice model and the structural sub-model of the attitude formation sub-model are presented in Table 7.8. The base taste parameters are very similar to those in Model 0 and Model 1-1. The coefficient of the scale factor is insignificant indicating measurable effect of the latent attitude on the taste parameters. In tune with our expectations this coefficient is negative with higher values of the latent attitude reflecting lower sensitivity and consequently, the scale factor should scale down the base taste parameters. The estimated structural sub-model is very similar to that in Model 1-1 with the same fit. Further, the estimated measurement sub-model presented in Table 7.9 is similar to those of Model 1-1 and Model 1-2.

The log-likelihood of the complete model, which includes the observed preference response and the attitudinal indicators, is -54269.08, which is worse compared to Model 1-1 and Model 1-2. Using the structural sub-model and the choice sub-model, the log-likelihood of the choice model component is -12231.06 which betters that of Model 0 by 3 units with 1 additional parameter.

Given the relative parsimony of the scaling approach we proceed to estimate the two attitude model using the scaling approach.

[28]Note that in principle, the coefficient α may be specific for each attribute, but the analyst can judiciously group attributes such that the latent attitude affects the shipper's sensitivity to each group in a similar manner.

Choice Sub-model

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.
Base Parameters	LRTT	0.072	0.006	11.83	11.21
	LRCTT	-3.678	0.211	-17.44	-16.89
	LRRATE	0.021	0.001	17.93	17.12
	LRPTB	-0.390	0.035	-11.27	-10.79
	LRLOD	0.227	0.011	20.73	19.83
	LRUEQ	-1.292	0.073	-17.60	-17.59
	LRRSP	-0.338	0.037	-9.08	-9.45
Scale Effect	α	-0.236	0.092	-2.57	-2.32
Threshold Parameters	κ_1	0.249	0.007	35.11	34.98
	κ_2	0.531	0.010	54.40	52.01
	κ_3	0.852	0.012	70.93	70.89
	κ_4	1.289	0.015	84.64	82.49

Structural Sub-model

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.
Structural Parameters	NEMP	0.012	0.002	6.18	5.91
	DENSITY	-0.033	0.031	-1.08	-1.23
	EARLYD	0.013	0.008	1.64	1.15
	SALES	-0.045	0.004	-12.23	-11.57
	LATED	0.058	0.009	6.80	6.23
	EDI1	0.106	0.013	7.94	8.47
	TONNES	7×10^{-5}	7×10^{-4}	0.11	0.21
	AVHAUL	-0.087	0.014	-6.06	-6.91
	ADAVPR	-0.023	0.006	-4.06	-3.82
Noise Parameter	σ_ζ	0.409	0.009	43.60	42.44

Squared multiple correlation of structural equation= 0.08

Table 7.8: Model 1-3: Scaled Coefficient Choice Model (one attitude)

Measurement Sub-model

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.	SMC
Structural Parameters	λ_1	1.000	–[a]	–	–	0.40
	λ_2	0.632	0.020	31.53	31.78	0.24
	λ_3	0.554	0.029	18.94	19.33	0.10
	λ_4	0.905	0.031	29.02	27.56	0.27
	λ_5	0.725	0.024	30.34	29.77	0.33
	λ_6	0.758	0.030	25.10	24.66	0.19
	λ_7	0.595	0.024	25.24	24.13	0.20
	λ_8	0.947	0.032	29.52	30.11	0.31
Noise Parameters	σ_{ξ_1}	0.506	0.006	78.90	77.32	–
	σ_{ξ_2}	0.464	0.005	93.27	92.81	
	σ_{ξ_3}	0.686	0.007	102.88	101.74	
	σ_{ξ_4}	0.624	0.007	92.25	91.57	
	σ_{ξ_5}	0.430	0.005	87.31	86.23	
	σ_{ξ_6}	0.657	0.007	98.73	97.67	
	σ_{ξ_7}	0.488	0.005	96.49	95.93	
	σ_{ξ_8}	0.591	0.007	89.46	89.33	

[a]Fixed parameter.

Complete Log-likelihood at convergence = -54269.08
Log-likelihood of choice model at zero = -12831.79
Log-likelihood at choice model at convergence = -12231.90
$\bar{\rho}^2 = 0.046$
Number of observations = 5840

Table 7.9: Model 1-3: Scaled Coefficient Choice Model (one attitude contd)

7.4.3 Model with Two Attitudes: Time sensitivity and Cost sensitivity

In this model we postulate the existence of two latent attitudes – *time sensitivity* and *cost sensitivity* – which capture heterogeneity of the shippers to service attributes. In the specification of the attitude formation sub-model, it is necessary to hypothesize which importance ratings indicate each latent attitude, and which observed shipper's characteristics determine them. The path diagram for the attitude formation sub-model is represented in Figure 7-5.

In the measurement sub-model, the importance ratings for transit time, reliability and rate are specified as indicators of time sensitivity. The indicators of cost sensitivity are importance ratings for rate, payment terms and billing, and loss and damage. In the structural model, the specification of the fixed and free parameters is based on a combination of prior hypothesis and a sequence of statistical tests of all the potential factors in each dimension. Mathematically, the final specification of structural sub-model is specified as:

$$\begin{aligned} Z_T^* &= \gamma_{T1}NEMP + \gamma_{T2}DENSITY + \gamma_{T3}EARLYD + \gamma_{T4}SALES + \\ &\qquad \gamma_{T5}LATED + \gamma_{T6}EDI + \zeta_T \\ Z_C^* &= \gamma_{C1}LATED + \gamma_{C2}EDI + \gamma_{C3}TONNES + \gamma_{C4}AVHAUL + \\ &\qquad \gamma_{C5}ADAVPR + \zeta_C \end{aligned}$$

where

$$\begin{pmatrix} \zeta_T \\ \zeta_C \end{pmatrix} \sim \mathcal{BVN}\left(0, \begin{bmatrix} \sigma^2_{\zeta_T} & \rho_{\zeta_T,\zeta_C}\sigma_{\zeta_T}\sigma_{\zeta_C} \\ \rho_{\zeta_T,\zeta_C}\sigma_{\zeta_T}\sigma_{\zeta_C} & \sigma^2_{\zeta_C} \end{bmatrix}\right) \tag{7.36}$$

Further, the measurement sub-model is specified as:

$$\begin{aligned} RTIME &= Z_T^* + \xi_1 \\ RCTIME &= \lambda_{21}Z_T^* + \xi_2 \\ RRATE &= \lambda_{31}Z_T^* + \lambda_{32}Z_C^* + \xi_3 \end{aligned}$$

$$
\begin{aligned}
RPTB &= Z_C^* + \xi_4 \\
RLOD &= \lambda_{52} Z_C^* + \xi_5
\end{aligned}
$$

where we assume that ξ_p $\forall p$ are independently distributed, and $\xi_p \sim \mathcal{N}(0, \sigma^2_{\xi_p})$. The scale of time sensitivity is set to that of the importance rating of transit time, while the scale of cost sensitivity is set to that of the importance rating of payment terms and billing.

As before the utility generating the preference response is given by:

$$
\begin{aligned}
U \;=\;& \beta_1 LRTT + \beta_2 LRCTT + \beta_3 LRRATE + \beta_4 LRPTB + \beta_5 LRLOD + \\
& \beta_6 LRUEQ + \beta_7 LRRSP + \epsilon
\end{aligned}
$$

We postulate that the taste parameters of time-related service attributes such as transit time and consistency of transit time will be affected by the shipper's time sensitivity. Similarly, the taste parameters of cost-related service attributes such as rate, payment terms and billing, and loss & damage are postulated to be affected by shipper's cost sensitivity. Consequently, the taste parameters are scaled by positive scalars which are parameterized functions of the latent attitude, and are written as:

$$
\begin{aligned}
\beta_1 &= \beta_{0,1}\mu_T \\
\beta_2 &= \beta_{0,2}\mu_T \\
\beta_3 &= \beta_{0,3}\mu_C \\
\beta_4 &= \beta_{0,4}\mu_C \\
\beta_5 &= \beta_{0,5}\mu_C \\
\beta_6 &= \beta_{0,6} \\
\beta_7 &= \beta_{0,7}
\end{aligned}
$$

with $\mu_T = \exp(\alpha_T Z_T^*)$ and $\mu_C = \exp(\alpha_C Z_C^*)$. It must be noted that in our specification we assume that the taste parameters corresponding to usability of equipment and responsiveness are constant[29].

Assuming that ζ, ξ and ϵ are independent, we adopt the maximum likelihood criterion in the estimation of the model parameters. As a first step only the attitude formation sub-model was estimated (see Appendix B) and the estimated parameters were used as starting values in the full information maximum likelihood estimation procedure. The estimated choice sub-model and the structural sub-model of the attitude formation sub-model is presented in Model 7.10. The base parameters of the choice model are very similar to those of Model 0, Model 1-1 and Model 1-3. The coefficients associated with the scaling parameters are significant and negative. This is in tune with our expectations since higher values of the latent attitudes are associated with lower sensitivities, and consequently the corresponding taste parameters are scaled down.

The structural sub-model represents the causal formulation for the determination of time sensitivity and cost sensitivity. In the time sensitivity dimension, the positive coefficient for number of employees indicate that shippers with larger workforce are less time sensitive. Shippers with earlier acceptable delivery times are less time sensitive. In contrast, it must be noted when only one latent attitude – overall attribute sensitivity – was postulated the effect of earliest acceptable delivery time variable on sensitivity was insignificant. As expected shippers with higher acceptable delays are less time sensitive. Thus, we can conclude that shippers with wider acceptable time windows are expected to be less time sensitive. Shippers using EDI are less time sensitive. The fit of the time sensitivity equation is 0.05.

Turning to the cost sensitivity dimension, the negative coefficient for maximum acceptable delivery time implies that shippers with higher acceptable delays are more cost sensitive. Shippers using EDI are less cost sensitive. Further, shippers with high annual tonnage are less cost sensitive. It must be noted that in the single

[29] In principle, a third latent attitude such as *service-quality sensitivity* can be postulated with the importance ratings of loss and damage, usability of equipment, responsiveness and level of effort utilized as relevant indicators of service-quality. Consequently, the service quality sensitivity can be postulated to affect the taste parameters of loss and damage, usability of equipment and responsiveness. Such an exercise was not conducted due to the computational complexity since calculation of the likelihood function entails three-dimensional integration.

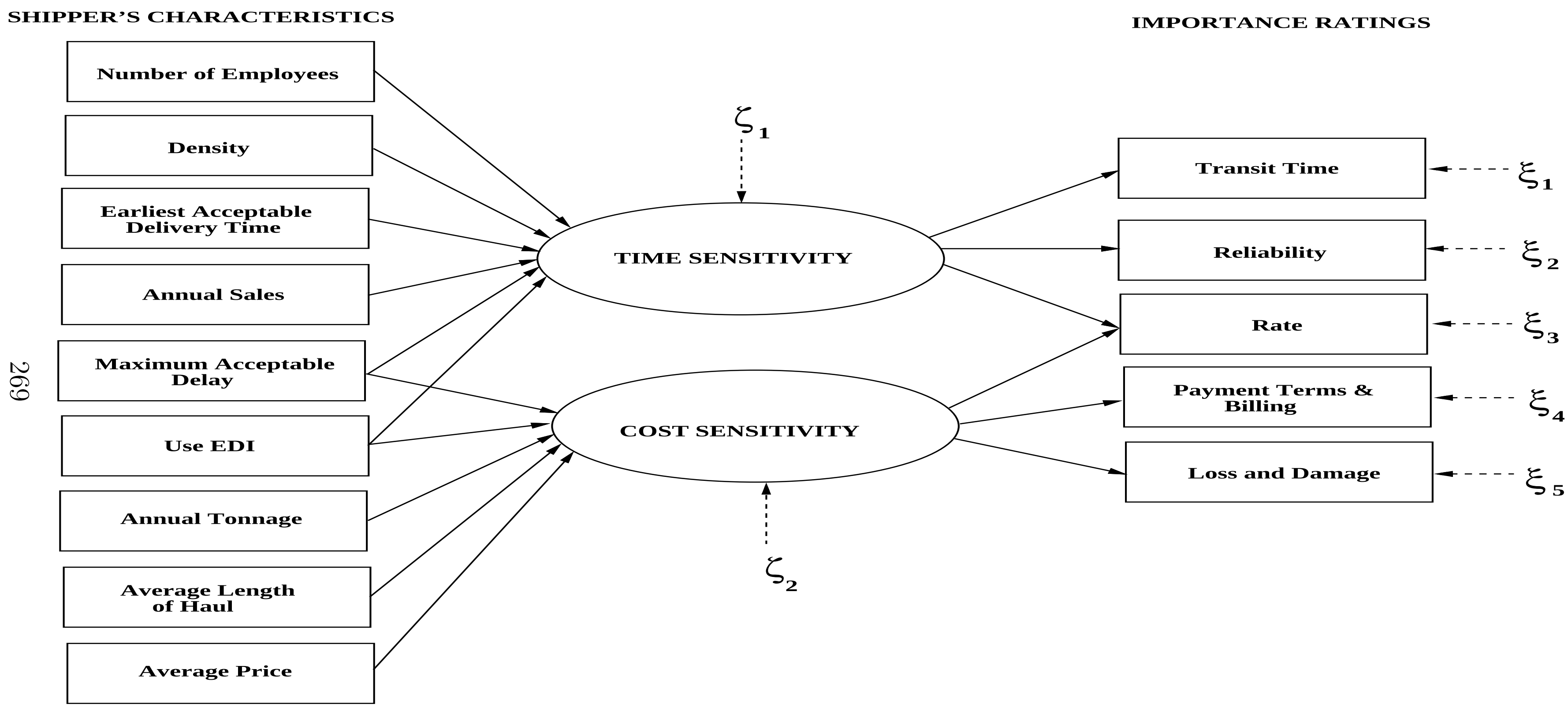


Figure 7-5: Path Diagram for Two Attitude Model

attitude models (except for Model 1-2), the effect of annual tonnage on sensitivity was negligible. Further, the negative coefficients for average length of haul and average price of shipment indicate that shippers transporting high value goods over long distances are more cost sensitive. The fit of the cost sensitivity equation is 0.10.

The estimated measurement sub-model is presented in Table 7.11. In the measurement equation of importance rating for transportation rate, the coefficient of time sensitivity is negative implying that the importance rating decreases as the latent attitude increases, i.e., the shipper with low time sensitivity tends to consider transportation rate more important. Noting the "inverse" relationship between shipper's preference towards transportation rate and transit time, this relationship is in tune with our expectations. Further, the fits of the measurement equations improved quite considerably indicating that the observed variations in the importance ratings is better explained through a two latent attitude model compared to the one latent attitude model.

The log-likelihood of the complete model[30], which includes the preference response and the attitudinal indicators is -37245.89. Using the structural sub-model of the attitude formation sub-model, and the choice sub-model, the log-likelihood of the choice model component is -12222.13 which betters that of Model 0 by 12 units with two additional parameters.

7.4.4 Summary of Estimated Models

Table 7.12 presents the summary of the estimated models. Looking at the complete log-likelihood, among the one latent attitude models, the latent class choice model fits the data the best, followed by the model with the interactions between attributes and latent attitude, and the scaled coefficient choice model. Comparing the choice model component of the one latent attitude models, the latent class choice model fits the best, although the improvement in model fit compared to the ordinal probit model is not considerable.

[30]It must be noted this complete log-likelihood is not comparable to the log-likelihoods of Model 1-1, Model 1-2, and Model 1-3, since in Model 2 only a subset of the attitudinal indicators are utilized.

Choice Sub-model

	Parameter	Estimates	Std. err.	t-stat	t-stat corr.
Base Parameters	LRTT	0.071	0.006	11.75	11.61
	LRCTT	-3.634	0.215	-16.89	-15.77
	LRRATE	0.021	0.001	18.13	19.11
	LRPTB	-0.391	0.035	-11.31	-11.71
	LRLOD	0.228	0.011	20.69	20.98
	LRUEQ	-1.288	0.074	-17.48	-17.11
	LRRSP	-0.340	0.038	-9.07	-9.75
Scale Effect	α_T	-0.371	0.127	-2.91	-2.38
	α_C	-0.224	0.086	-2.61	-2.45
Threshold Parameters	κ_1	0.249	0.007	35.11	34.91
	κ_2	0.531	0.010	54.39	53.67
	κ_3	0.852	0.012	70.92	70.14
	κ_4	1.289	0.015	84.60	82.87

Structural Sub-model

	Parameter	Estimate	t-stat	t-stat corr.	SMC
Time Sensitivity Dimension	NEMP	0.011	4.88	4.44	0.05
	DENSITY	0.045	1.22	1.14	
	EARLYD	0.042	4.67	4.73	
	SALES	-0.050	-11.67	-10.87	
	LATED	0.069	6.28	6.11	
	EDI	0.053	3.10	2.98	
Cost Sensitivity Dimension	LATED	-0.083	-8.11	-8.34	0.10
	EDI	0.136	7.87	7.23	
	TONNES	0.006	7.00	6.83	
	AVHAUL	-0.217	-12.34	-11.89	
	ADAVPR	-0.049	-6.97	-7.03	
Noise Parameters	σ_{ζ_T}	0.531	40.83	39.11	–
	σ_{ζ_C}	0.500	45.09	44.68	
	ρ_{ζ_T,ζ_C}	0.670	42.58	41.33	

Table 7.10: Model 2: Scaled Coefficient Choice Model with Time Sensitivity and Cost Sensitivity

Measurement Sub-model

Indicator	Time sens.			Cost sens.			σ_ξ	SMC
	estimate	t-stat	t-stat corr.	estimate	t-stat	t-stat corr.		
RTIME	1.000	–[a]	–	–	–	–	0.38	0.66
RCTIME	0.646	26.67	25.37	–	–	–	0.41	0.41
RRATE	-0.219	-7.00	-6.84	0.811	21.68	20.37	0.63	0.24
RPTB	–	–	–	1.000	–[a]	–	0.51	0.51
RLOD	–	–	–	0.556	27.80	26.33	0.44	0.30

[a]Fixed parameters or not estimated.

Complete Log-likelihood at convergence = -37245.89
Log-likelihood of choice model at zero= -12831.79
Log-likelihood at choice model at convergence= -12222.13
$\bar{\rho}^2 = 0.046$
Number of observations = 5840

Table 7.11: Model 2: Scaled Coefficient Choice Model with Time Sensitivity and Cost Sensitivity (cont'd)

Model	Complete		Choice Component			
	Log-lik.	# of par.	Log-like.	Akaike	# of par.	$\bar{\rho}^2$
Ordinal Probit	-12234.15	11	-12234.15	-12245.15	11	0.046
Interactions between attrib. and 1 latent att.	-54227.15	43	-12230.11	-12248.11	18	0.045
Latent class with 1 latent att.	-54187.14	44	-12219.06	-12238.06	19	0.046
Scaled coeff. with 1 latent att.	-54269.08	37	-12231.90	-12243.90	12	0.046
Scaled coeff. with 2 latent att.	-37245.89	35	-12222.90	-12235.90	13	0.046

Table 7.12: Summary of Estimated Models

The parsimonious scaled coefficient choice model with two latent attitude model has the best fit in terms of the Akaike criterion among all the models. Although the improvements in fit with the use of the sophisticated and computationally demanding modeling approaches appear negligible, the latent attitude formation model may provide considerable strategic information to the marketing manager of a railroad.

7.5 Summary

In this chapter we presented a class of choice models for incorporating attitudinal data, and capturing heterogeneity in choice processes stemming from attitudinal variations. The key feature of the methodology is the representation of the attitude formation process, and its explicit link to the choice process. This class of models must be viewed as a special case of the latent structure choice model. We also applied the class of choice models in a shipper's freight mode choice context.

Chapter 8

Conclusions and Future Research

In this chapter we summarize the key conclusions from the thesis and suggest avenues for future research. In future research directions, we highlight the theoretical open problems, and emphasize applications and possible extensions of the tools developed in this thesis.

8.1 Summary and Conclusions

We developed the latent class choice model (LCCM), wherein the latent classes characterize sources of heterogeneity such as segments of the population with varying tastes, choice set considered, and decision protocol adopted. We formulated three types of class membership models (categorical criterion model, binary criteria model and ordinal criteria model) with special references to the aforementioned types of heterogeneity. Further, for the latent class choice model for taste heterogeneity, we emphasized a parsimonious specification for the "generation" of unobserved variations in tastes to attributes of alternatives through a smaller number of sensitivity dimensions, and provided illustrative examples.

We applied the latent class choice model for taste heterogeneity in the estimation of travel choice models with distributed value of time. We demonstrated the efficacy of the model compared to extant approaches of introducing interaction variables in the utility functions, and random coefficient models. The case study evidenced the significance of the unobserved variations in value of time in the sample which persisted even after the systematic variations due to socio-economic and demographic variables were accounted for. Specifically, the application of the estimated models revealed the existence of considerably higher values of times for certain segments of the population, and presence of heterogeneity along travel time and travel cost sensitivity dimensions.

We applied the latent class choice model for decision protocol heterogeneity in a transportation mode choice context with data from simulated choice experiments. It must be noted that in principle, the idea of allowing for decision protocol variations through a model-based approach is a significant departure from traditional random utility models. Since decision protocols in revealed preference (RP) and stated preference (SP) settings may differ for the *same individual*, we also discussed the need to combine RP and SP data, and outlined an approach to validate decision protocols exhibited in SP analysis with those of RP data, if both RP and SP data are available. This approach is built on previous work wherein choice models utilize both RP and SP data (see Ben-Akiva and Morikawa [1990a]).

It must be noted that we did not specify explicit indicators for the classes in LCCM. In chapter 6, as a first step in the incorporation of indicators of latent classes which may be viewed as *attributes* which characterize the class, we elaborated on the different types of specification of the measurement model depending on the characterization of the latent class. These measurement models are linked to the different class membership models to obtain a more refined latent class model. Subsequently, the latent class model is integrated with the choice model to form the LCCM with indicators. We also developed the latent structure choice model (LSCM) which incorporates the gamut of attitudinal and perceptual indicators through latent attitudes, perceptions and classes, and discussed estimation of the model system parameters.

As a special case of LSCM, we elaborated on a class of choice models which incorporates attitudinal indicators such as individual's importance ratings of attributes of alternatives in the choice process. The emphasis was on "generating" unobserved taste variations from variations in individual attitudes. We applied such models in a shipper's freight mode choice study. In principle, we extended the work of Vieira [1992] by linking the choice model with an explicit causal model for attitude formation, and specifying responses to attitudinal questions in surveys as indicators of attitudes. Specifically, the shipper's importance ratings of service attributes were utilized as the attitudinal indicators. Although, the effects of the shipper's attitudes on the sensitivity to service attributes were statistically significant, the overall improvement in the fit of the choice model is limited. Given the considerable computational burden in the estimation of such models vis-a-vis standard choice models such as multinomial logit model, the results are apparently disappointing. This is partly attributable to negligible variations in the importance ratings of attributes of alternatives.

8.2 Future Research Directions

This thesis has endeavored to advance discrete choice modeling techniques with an emphasis on the incorporation of the psychological factors affecting the underlying choice process. We also note that the empirical case studies have only surfaced the potential for the modeling approaches, and further work is needed to assess their ramifications and to transcribe the methodological developments from an academic setting to practical applications. In this regard, we foresee future research to be conducted along two directions:

1. *Theoretical Developments*; and
2. *Empirical Applications*.

8.2.1 Theoretical Developments: Existence and Identification Issues

As noted in chapter 3, we have not yet derived the conditions for the existence of the estimates of model parameters in LCCM. Further, even under the assumption of existence, we have only partially addressed the issue of model identification. Specifically, no general necessary and sufficient conditions currently exist for the existence and unicity of model parameters.

The above issues are not adequately addressed for LSCM too. It must be noted that sufficient conditions for the identification of the LSCM sans latent classes (i.e., LSCM incorporating only attitudes and perceptions "measured" on a continuous scale) can be obtained. In this case, we note that the sufficient conditions for the attitude formation and perception formation sub-models exist since they correspond to those of the linear latent variable model. Further, necessary and sufficient conditions of the identification of the choice sub-model are well known. It follows that if the component sub-models are *sufficiently* identified, the complete model is also *sufficiently* identified.

If the LSCM incorporates psychological factors such as *continuous* attitudes and perceptions as well as categorical or discrete concepts characterized as latent classes, then no general necessary and sufficient identification conditions exist. Further, even sufficient conditions do not exist since we have not developed, unlike the latent variable model, sufficient conditions for the latent class sub-model.

More theoretical research is needed to address these important open problems related to the existence and identification of model parameters.

8.2.2 Modeling Enhancements and Empirical Applications

Lifestyle Concept in Travel Demand Models

The concept of lifestyle is hypothesized to capture long term decisions of individuals and households which guide their preferred pattern of mobility, activity and travel choices, and is expected to be a key "higher" level factor substituting for traditional social class and economic status variables. The main arguments for using the lifestyle concept in the disaggregate travel demand modeling approach, wherein the travel behavior of an individual is the basic building block, are:

- to capture the constraints imposed on the travel decisions of individuals by the mobility decisions;
- to incorporate implicitly the intra-household interactions and constraints in the choice behaviors;

- hypothesized that the primary changes in travel behavior due to new information technologies (IT) such as Advanced Traveler Information Systems (ATIS) and Advanced Traffic Management Systems (ATMS), and tele-options such as tele-commuting, tele-conferencing, and tele-banking, would stem from the consequent changes in lifestyles and activity patterns;
- to account for heterogeneity in choice behaviors in a systematic manner; and
- to incorporate "soft" or subjective information such as individual's attitudes, values and opinions, and perceptions.

It must be noted that the lifestyle concept is not directly measurable or observable, and hence may be characterized through latent classes.

The above considerations are captured in the conceptual framework depicted in Figure 8-1. In the figure, the ellipses represent unobservable concepts. The key elements of the framework are:

1. The explicit incorporation of a *latent* lifestyle concept as the composite outcome of the life decisions an individual makes. These life decisions are the pragmatic decisions an individual makes in order to fulfill his/her feasible aspirations within three important aspects of life: patterns of interpersonal relations (family formation), economic activity (participation in the labor force) and patterns of leisure (orientation towards different non-economic activity types, duration and frequency). Jointly, they can be interpreted as a choice reflecting one's aspirations for a life pattern.

2. The individual's *latent information state* concept which represents the experiential knowledge of the individual along three dimensions:

   (a) *Activity and transportation network*;

   (b) *Information technologies*: perceived benefits of accessing or purchasing information technologies (e.g., tele-options).

(c) *Information sources*: reliability and usefulness, as perceived by the individual, of different information sources providing information on activity and transportation network characteristics (e.g., ATIS).

The feedback from the information state occurs at two levels:

1. In the short term, activity/travel scheduling choices are affected. These relate towards access and processing of real-time information source such as ATIS.

2. In the long term, the individual changes the lifestyle choice through activity program shifts. These relate to shifts in usage of tele-options, purchase/subscribe to new IT, etc.

The remaining components in Figure 8-1 and their implications are self-explanatory.

Latent Class Choice Model for Choice Set Heterogeneity

In the LCCM for choice set heterogeneity reviewed in section 2.7 the class indicators utilized correspond to the recorded responses to alternative availability questions. For example, if one considers the choice set formation process as one of elimination of those alternatives not satisfying certain criteria or rules, the availability responses reflect the final outcome of the process which is the choice set considered. Specifically, information as to which of the attributes did not satisfy the individual's criteria are not utilized.

Noting that the rules adopted by individuals may be operationalized through a set of criterion functions satisfying a set of inequalities, in principle, responses to questions related to "satisfaction" or "acceptability" of the attributes may meaningfully indicate these criterion functions. For example, consider an individual who eliminates an alternative depending on a threshold such as a reservation price. The response to a question such as: "Is the price of alternative j acceptable to you?", would reflect whether or not the alternative satisfies the individual's price threshold, and consequently, may be utilized as the indicator of a price related criterion function. It is

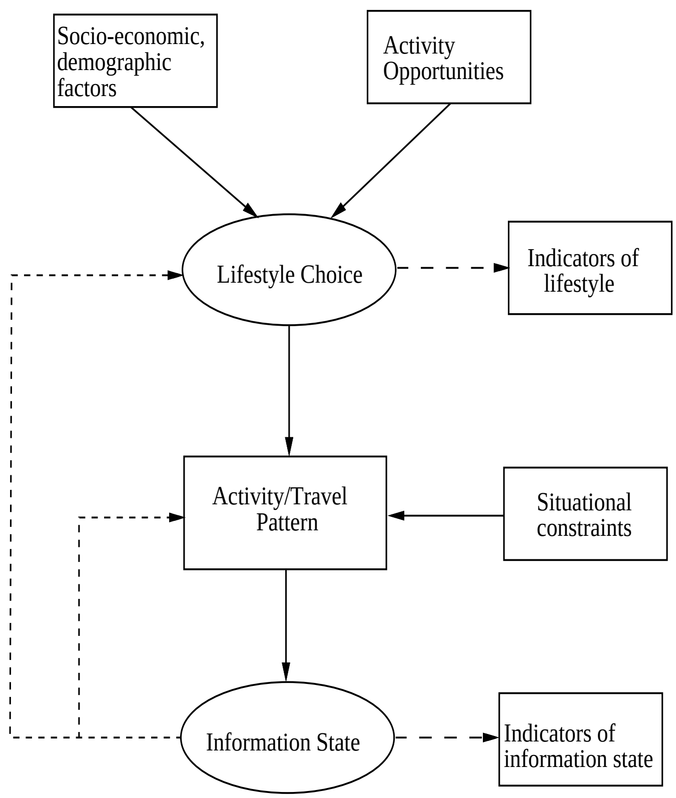


Figure 8-1: Latent Lifestyle and Travel demand: A conceptual framework

relatively transparent to extend previous work in choice set formation to include such indicators.

It must also be noted that an exploratory data analysis of the availability and acceptability indicators would reveal whether a *conjunctive* or a *disjunctive* rule is adopted in the screening of alternatives.

Enhancements in Data Collection Efforts

The fruition of the statistically advanced and conceptually sophisticated set of tools for discrete choice analysis put forth in this thesis depends heavily on the availability of "richer" data such as attitudinal and perceptual indicators. Such data, routinely collected in the marketing research context, are rarely collected in travel demand analysis. Consequently, more substantive research is needed in survey and questionnaire design to reflect the changing needs of travel demand analysis, and more importantly, to assess the practical significance and benefits of such modeling approaches.

Combining RP and SP Data

It must be noted that the case study for illustrating decision protocol heterogeneity is in the context of simulated choice experiments, while the modeling approach is fairly general and may be adopted for RP data, and also to combine RP and SP data.

If the preference model estimated on SP data suggests the existence of more than one decision protocol, the question remains as to whether such heterogeneity may exist in the *actual market environment*. If RP data is not available, then it is left to the analyst's judgment on how the SP model is utilized in forecasting and other model applications.

Since decision protocols in RP and SP settings may differ, application of the approach outlined in section 5.4 to validate decision protocols exhibited in SP analysis with those of RP data would be a useful exercise. Specifically, we need to assess the stability of decision protocols in both data sets in different choice situations.

Appendix A

Simulation Estimator for Ordinal Criteria Model

The latent class probability in an ordinal criteria model cannot be obtained in closed form as it entails non-trivial multi-dimensional integration, and tractable numerical procedures exist only if the number of dimensions of the latent class is limited to 3. To handle large number of dimensions, we propose the use of the simulator proposed by Geweke [1989] (see Bolduc [1993] for an application of the GHK simulator for the estimation of the Multinomial Probit model).

Let the criterion functions $H_{dn}, \forall d = 1, \ldots, D$ for individual n be written as:

$$H_{dn} = \theta_d' Z_n + \delta_{dn}, \quad \forall d = 1, \ldots, D \tag{A.1}$$

where Z_n is a $Q \times 1$ vector of individual characteristics, θ_d is a $Q \times 1$ parameter vector, and δ_n is a $Q \times 1$ random vector. We allow for the random components of H_d's, i.e., δ_d's, to be correlated to capture the unobserved interrelationships among the latent class dimensions[1]. The criterion functions can be written in a compact form as:

$$H_n = \Theta Z_n + \delta_n \tag{A.2}$$

[1]If the random components are independent the class membership model reduces to the product of ordinal probit models corresponding to each dimension and hence is tractable even for large dimensions.

where H_n is the stacked $D \times 1$ vector of criterion functions, δ_n is the $D \times 1$ random vector, and Θ is the $D \times Q$ parameter matrix with the d^{th} row corresponding to θ'_d. Assume that $\delta_n \sim \mathcal{N}(0, \Sigma)$ where Σ is a D-dimensional correlation matrix[2]. If Γ is the lower triangular Cholesky decomposition of Σ such that $\Sigma = \Gamma\Gamma'$, the equivalent model of equation (A.2) is written as:

$$H_n = \Theta Z_n + \Gamma \xi_n \tag{A.3}$$

where $\xi_n \sim \mathcal{N}(0, I_D)$ with I_D a D-dimensional identity matrix. It must be noted the parameters of Γ are interrelated due to the correlation matrix it generates[3]. The constraint set associated with the event $T_n = [l_1, \ldots, l_D]'$ is written as:

$$\begin{pmatrix} \tau^1_{l_1-1} \\ \tau^2_{l_2-1} \\ \vdots \\ \tau^D_{l_D-1} \end{pmatrix} \leq \begin{pmatrix} H_{1n} \\ H_{2n} \\ \vdots \\ H_{Dn} \end{pmatrix} = \begin{pmatrix} \theta'_1 Z_n \\ \theta'_2 Z_n \\ \vdots \\ \theta'_D Z_n \end{pmatrix} + \begin{pmatrix} \gamma_{11} & 0 & \ldots & 0 \\ \gamma_{21} & \gamma_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{D1} & \gamma_{D2} & \ldots & \gamma_{DD} \end{pmatrix} \begin{pmatrix} \xi_{1n} \\ \xi_{2n} \\ \vdots \\ \xi_{Dn} \end{pmatrix} \leq \begin{pmatrix} \tau^1_{l_1} \\ \tau^2_{l_2} \\ \vdots \\ \tau^D_{l_D} \end{pmatrix} \tag{A.5}$$

and can be explicitly specified for each criterion function as:

$$\begin{aligned}
\tau^1_{l_1-1} \leq H_{1n} \leq \tau^1_{l_1} \quad &\Rightarrow \quad \tau^1_{l_1-1} - \theta'_1 Z_n \leq \xi_{1n} \leq \tau^1_{l_1} - \theta'_1 Z_n \\
\tau^2_{l_2-1} \leq H_{2n} \leq \tau^2_{l_2} \quad &\Rightarrow \quad \frac{\tau^2_{l_2-1} - \theta'_2 Z_n - \gamma_{11}\xi_{1n}}{\gamma_{22}} \leq \xi_{2n} \leq \frac{\tau^2_{l_2} - \theta'_2 Z_n - \gamma_{11}\xi_{1n}}{\gamma_{22}} \\
\vdots \qquad\qquad &\qquad\qquad\qquad \vdots \\
\tau^D_{l_D-1} \leq H_{Dn} \leq \tau^D_{l_D} \quad &\Rightarrow \quad \frac{\tau^D_{l_D-1} - \theta'_D Z_n - \sum_{d=1}^{D-1} \gamma_{Dd}\xi_{dn}}{\gamma_{DD}} \leq \xi_{Dn} \leq \frac{\tau^D_{l_D} - \theta'_D Z_n - \sum_{d=1}^{D-1} \gamma_{Dd}\xi_{dn}}{\gamma_{DD}}
\end{aligned}$$

[2] Since the criterion functions are latent, the scale of each criterion function is set by fixing the variance of its random component to 1. Consequently, Σ is a correlation matrix.

[3] For example, consider a two-dimensional class membership model with

$$\begin{pmatrix} \delta_{1n} \\ \delta_{2n} \end{pmatrix} \sim \mathcal{BVN}\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) \tag{A.4}$$

and $\Gamma = \begin{pmatrix} \gamma_{11} & 0 \\ \gamma_{12} & \gamma_{22} \end{pmatrix}$ then $\Sigma = \Gamma\Gamma'$ implies $\gamma_{11} = 1$, $\gamma_{22} = \sqrt{1 - \gamma_{12}^2}$, and $\gamma_{12} = \rho < 1$.

The above constraints can be written in more compact form as:

$$
\begin{aligned}
&\nu_{1n}^{+} \leq \xi_{1n} \leq \nu_{1n}^{-} \\
&\nu_{2n}^{+}(\xi_{1n}) \leq \xi_{2n} \leq \nu_{2n}^{-}(\xi_{1n}) \\
&\quad\vdots \qquad\qquad\qquad \vdots \\
&\nu_{Dn}^{+}(\xi_{1n}, \ldots, \xi_{D-1;n}) \leq \xi_{Dn} \leq \nu_{Dn}^{-}(\xi_{1n}, \ldots, \xi_{D-1;n})
\end{aligned}
\tag{A.6}
$$

where

$$
\nu_{d'n}^{+} = \begin{cases} \tau_{l_1-1}^{1} - \theta_1' Z_n, & d' = 1 \\ \frac{\tau_{l_{d'}-1}^{d'} - \theta_{d'}' Z_n - \sum_{d=1}^{d'-1} \gamma_{d'd} \xi_{dn}}{\gamma_{d'd'}}, & d' = 2, \ldots, D \end{cases}
\tag{A.7}
$$

and

$$
\nu_{d'n}^{-} = \begin{cases} \tau_{l_1}^{1} - \theta_1' Z_n, & d' = 1 \\ \frac{\tau_{l_{d'}}^{d'} - \theta_{d'}' Z_n - \sum_{d=1}^{d'-1} \gamma_{d'd} \xi_{dn}}{\gamma_{d'd'}}, & d' = 2, \ldots, D \end{cases}
\tag{A.8}
$$

Given the above, the observed probability, denoted by $\mathrm{P}(T_n|Z_n;\Theta,\tau,\Gamma)$ is written as:

$$
\begin{aligned}
\Pr\Big(&\nu_{1n}^{+} \leq \xi_{1n} \leq \nu_{1n}^{-}, \nu_{2n}^{+}(\xi_{1n}) \leq \xi_{2n} \leq \nu_{2n}^{-}(\xi_{1n}), \ldots, \\
&\nu_{Dn}^{+}(\xi_{1n}, \ldots, \xi_{D-1;n}) \leq \xi_{Dn} \leq \nu_{Dn}^{-}(\xi_{1}, \ldots, \xi_{D-1;n})\Big)
\end{aligned}
\tag{A.9}
$$

Noting that the ξ's are assumed independent, by recursive conditioning, the above probability reduces to

$$
\Pr\left(\nu_{1n}^{+} \leq \xi_{1n} \leq \nu_{1n}^{-}\right) \cdot \Pr\left(\nu_{2n}^{+}(\xi_{1n}) \leq \xi_{2n} \leq \nu_{2n}^{-}(\xi_{1n}) | \nu_{1n}^{+} \leq \xi_{1n} \leq \nu_{1n}^{-}\right) \cdots
\tag{A.10}
$$

Now we turn our attention to approximating the above probability through importance sampling. Let r denote a particular random draw and ξ_{nr} a given realization of the random vector ξ_n such that equation (A.6) is satisfied. Based on R such draws, the required latent class probability can be approximated as, $\tilde{\mathrm{P}}_R(T_n|Z_n;\Theta,\tau,\Gamma)$, which equals

$$
\begin{aligned}
& \frac{1}{R}\sum_{r=1}^{R} \mathrm{P}_r(T_n|Z_n;\Theta,\tau,\Gamma) \\
= & \sum_{r=1}^{R} \Pr\left(\nu_{1n}^{+} \leq \xi_{1nr} \leq \nu_{1n}^{-}\right) \cdot \Pr\left(\nu_{2n}^{+}(\xi_{1nr}) \leq \xi_{2nr} \leq \nu_{2n}^{-}(\xi_{1nr})|\xi_{1nr}\right) \cdots \\
& \quad \Pr\left(\nu_{Dn}^{+}(\xi_{1nr},\ldots,\xi_{D-1;nr}) \leq \xi_{Dnr} \leq \nu_{Dn}^{-}(\xi_{1nr},\ldots,\xi_{D-1;nr})|\xi_{1nr},\ldots,\xi_{D-1;nr}\right)
\end{aligned}
\tag{A.11}
$$

If $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal variable, $\mathrm{P}_r(T_n|Z_n;\Theta,\tau,\Gamma)$ in the above equation can be compactly written as:

$$
\begin{aligned}
\mathrm{P}_r(T_n|Z_n;\Theta,\tau,\Gamma) = & \left(\Phi(\nu_{1nr}^{-}) - \Phi(\nu_{1nr}^{+})\right) \cdot \left(\Phi(\nu_{2nr}^{-}) - \Phi(\nu_{2nr}^{+})\right) \cdots \\
& \left(\Phi(\nu_{Dnr}^{-}) - \Phi(\nu_{Dnr}^{+})\right)
\end{aligned}
\tag{A.12}
$$

where the ν's for each draw are computed as:

$$
\nu_{d'nr}^{+} = \begin{cases} \tau_{l_1-1}^{1} - \theta_1' Z_n, & d' = 1 \\ \frac{\tau_{l_{d'}-1}^{d'} - \theta_{d'}' Z_n - \sum_{d=1}^{d'-1} \gamma_{d'd}\xi_{dnr}}{\gamma_{d'd'}}, & d' = 2,\ldots,D \end{cases}
\tag{A.13}
$$

and

$$
\nu_{d'nr}^{-} = \begin{cases} \tau_{l_1}^{1} - \theta_1' Z_n, & d' = 1 \\ \frac{\tau_{l_{d'}}^{d'} - \theta_{d'}' Z_n - \sum_{d=1}^{d'-1} \gamma_{d'd}\xi_{dnr}}{\gamma_{d'd'}}, & d' = 2,\ldots,D \end{cases}
\tag{A.14}
$$

and ξ_{dnr} is drawn according to $\Phi^{-1}\left(u\left(\Phi(\nu_{dnr}^{-}) - \Phi(\nu_{dnr}^{+})\right) + \Phi(\nu_{dnr}^{+})\right)$ where $u \sim \mathcal{U}[0,1]$.

Appendix B

LISREL Model Estimation Results for Shipper's Freight Mode Choice Case Study

In this appendix we present estimation results for the LISREL model systems utilized in chapter 7. The estimation was conducted using standard software (LINCS from RJ Software Inc. [1993]) for the estimation of the LISREL model. It must be noted that the assumptions in the maximum likelihood estimation of these models are slightly different from the assumptions in the sub-models presented in chapter 7. Herein we assume that *both* the indicators and the explanatory variables determining latent attitudes are multivariate normal vectors, while in the sub-models in chapter 7 we allow only for the distribution of the indicators conditional on the latent attitudes to be multivariate normal.

B.1 One latent attitude model

Herein we assume the existence of a single latent attitude which denotes the an "overall attribute sensitivity" of the shipper to service attributes. The path diagram for the latent variable model is represented in Figure 7-4 of chapter 7. Shipper's characteristics such as number of employees, density of shipments, earliest acceptable delivery time, annual sales, maximum acceptable delay, EDI usage, annual tonnage shipped, average length of haul and average price are postulated to determine the shipper's attitude. Also all the importance ratings are utilized as indicators of the latent attitude. The estimated model system is presented in Table B.1.

B.2 Two latent attitude model

In this model we assume the existence of two latent attitudes – *time sensitivity* and *cost sensitivity* – which capture the heterogeneity of the shipper to service related attributes. In the specification of the formation of the latent attitudes through the latent variable model, it is necessary to hypothesize which importance ratings indicate each dimension, and which shipper's characteristics determine them. After some experimentation, the best representation of the latent variable model is illustrated in path diagram in Figure 7-5 of chapter 7. In the measurement model, the importance ratings for transit time, reliability and rate are utilized as indicators of time sensitivity, while the importance ratings of rate, payment terms and billing, and loss and damage are utilized as the indicators of cost sensitivity. The scale of time sensitivity is set to that of the importance rating of transit time, while the scale of cost sensitivity is set to that of the importance rating of payment terms and billing. The estimated model system is presented in Table B.2.

STRUCTURAL MODEL - LATENT ATTITUDE

Independent variable	Estimate	t-stat
Number of employees	0.012	6.49
Density	-0.072	-2.30
Earliest acceptable delivery time	-0.004	-0.52
Sales	-0.048	-13.14
Maximum acceptable delay	0.071	8.10
EDI usage	0.117	8.54
Annual tonnage	-3×10^{-4}	-0.41
Average haul	-0.080	-5.40
Average price	-0.023	-3.94

Squared multiple correlation of structural equation= 0.08

MEASUREMENT MODEL

Indicator	Estimate	t-stat	squared multiple correlation
Transit Time	1.000	–[a]	0.43
Consistency of transit time	0.647	30.97	0.27
Rate	0.522	19.69	0.10
Payment terms & billing	0.860	30.29	0.26
Loss & damage	0.684	32.69	0.31
Usability of equipment	0.727	26.34	0.18
Responsiveness	0.563	27.05	0.19
Level of effort	0.912	32.35	0.30

[a]Fixed parameter.

Total number of observations= 485

Table B.1: One latent attitude model

STRUCTURAL MODEL - TIME SENSITIVITY

Independent variable	Estimate	t-stat
Number of employees	0.011	5.37
Density	0.011	0.32
Earliest acceptable delivery time	0.022	2.44
Sales	-0.050	-12.82
Maximum acceptable delay	0.084	7.83
EDI usage	0.067	4.07

Squared multiple correlation of TIME SENSITIVITY model= 0.06

STRUCTURAL MODEL - COST SENSITIVITY

Independent variable	Estimate	t-stat
Maximum acceptable delay	-0.080	-7.84
EDI usage	0.146	8.46
Annual tonnage	0.006	7.20
Average haul	-0.220	-12.70
Average price	-0.052	-7.60

Squared multiple correlation of COST SENSITIVITY model= 0.10

MEASUREMENT MODEL

Indicator	Time sens.		Cost sens.		squared multiple correlation
	estimate	t-stat	estimate	t-stat	
Transit Time	1.000	–[a]	–	–	0.66
Consistency of transit time	0.672	30.20	–	–	0.44
Rate	-0.209	-6.98	0.797	21.92	0.25
Payment terms & billing	–	–	1.000	–[a]	0.51
Loss & damage	–	–	0.544	29.66	0.29

[a]Fixed parameter.

Total number of observations= 485

Table B.2: Two latent attitude model

Appendix C

Latent Variable Model with Ordered Categorical Indicators

In chapter 7 we assumed that the importance ratings of attributes of alternatives are metrically scalable, although these ratings are measured on a Likert-type scale and hence are ordered categorical variables. In this appendix, seeking a more realistic representation of the latent variable model, we outline an approach which parallels the work of Muthén [1983, 1984] to incorporate ordered categorical variables as indicators of latent variables. The presentation of the model is unique with an emphasis on the specification of the fixed and free parameters of the model necessary to aid in the identification of the model. The necessary identification conditions presented have not been explicitly enunciated in Muthén [1983, 1984] and deserve special attention.

A traditional latent variable model with continuous indicators consists of two parts: a measurement model and a structural model. The first of these specifies how the latent variables are related to the observed or measured variables (i.e., indicators which are manifestations of the underlying latent variables) and the second specifies the relationship from the explanatory variables to the latent variables. In similar vein, we focus on the latent variable model with ordered categorical indicators which consists of: a structural model representing the latent variables as a function of a set of causal variables, and a measurement model which is the mapping from the latent variables to the ordered categorical indicators.

It must be noted that latent variable models with categorical indicators are important in many applications, particularly in the social and the behavioral sciences, as the indicators frequently have a small number of categories. The assumption of continuous indicators in such cases may not be appropriate, especially when the distributions of the indicators are skewed.

C.1 Structural model

As in the latent variable model with continuous indicators, the $M \times 1$ vector of latent variables denoted by Z_n^* is assumed to be *continuous* and determined by the $Q \times 1$ vector of explanatory variables Z_n through a linear structural model, i.e.,

$$Z_n^* = \gamma_0 + \Gamma Z_n + \zeta_n \tag{C.1}$$

where γ_0 is a $M \times 1$ parameter vector of intercepts, Γ is a $M \times Q$ parameter matrix, and ζ_n is a $M \times 1$ random vector. In the traditional latent variable model, the scale of each latent variable is usually set to that of one of the continuous indicators by fixing the coefficient of the latent variable in the measurement equation for that indicator to 1. Consequently, the unit of measurement for the latent variable is the corresponding indicator. Herein, since the indicators are ordered categorical, we set the scale of the latent variable $Z_{mn}^* \ \ \forall m = 1, \ldots, M$ by fixing the variance of ζ_m to 1. Specifically, we assume $\zeta_n \sim \mathcal{N}(0, \Psi)$ where Ψ is a *correlation* matrix.

C.2 Measurement model

In the latent variable with continuous indicators, one postulates a direct mapping from the latent variables to the indicators through a measurement model such as:

$$A_n = \lambda_0 + \Lambda Z_n^* + \xi_n \tag{C.2}$$

where λ_0 is a $P \times 1$ parameter vector of intercepts, Λ is a $P \times M$ parameter matrix, and ξ_n is a $P \times 1$ random vector representing measurement errors. In a similar vein, noting that the indicators are ordered categorical, we postulate an "inner" measurement model with an *intermediate* $P \times 1$ latent vector A_n^*, referred to as *latent indicator vector*, specified as a linear function of Z_n^*, i.e.,

$$A_n^* = \lambda_0 + \Lambda Z_n^* + \xi_n \tag{C.3}$$

As noted earlier in the latent variable model with continuous indicators, one element of each column of Λ is fixed to 1. But in the above specification, such restrictions are not imposed allowing the conduct of statistical tests for the significance of the effects of different latent variables on the latent indicator vector.

An "outer" measurement model maps from the latent indicator vector A_n^* to the indicators A_n. This mapping is based on a "threshold crossing" idea as in the ordinal probability model of McKelvey and Zavoina [1975], i.e.,

$$A_{pn} = \begin{cases} 1 & \text{if} \quad \tau_0^p = -\infty < A_{pn}^* \leq 0 = \tau_1^p \\ 2 & \text{if} \quad \tau_1^p < A_{pn}^* \leq \tau_2^p \\ 3 & \text{if} \quad \tau_2^p < A_{pn}^* \leq \tau_3^p \\ & \vdots \\ L & \text{if} \quad \tau_{L-1}^p < A_{pn}^* < \infty = \tau_L^p \end{cases} \tag{C.4}$$

where A_{pn} denotes the p^{th} categorical indicator, L denotes the number of ordered levels such that $A_{pn} \in \{1, \ldots, L\}$[1], and $(\tau_2^p, \ldots, \tau_{L-1}^p)$ are estimable threshold parameters[2]. Since the indicators are ordered categorical, it is necessary to fix the diagonal elements of the covariance matrix of ξ_n, denoted by Θ, to 1. Further, if we assume $\xi_n \sim \mathcal{N}(0, I_P)$, where I_P is a P-dimensional identity matrix, then conditional on Z^*, the measurement model can be written as a product of ordinal probit proba-

[1]For simplicity we assume the number of ordered levels for different indicators to be the same.

[2]It must be noted that the set of threshold parameters may be assumed to be the same across the different indicators.

bilities. It must be noted that if the set of threshold parameters are the same across indicators, then it is necessary to fix only one of the diagonal elements in Θ to 1, and all other diagonal elements may be free parameters.

C.3 Likelihood function and Estimation

Assuming that the set of threshold parameters are indicator-specific, and ζ and ξ are independent with $\xi_n \sim \mathcal{N}(0, I_P)$, the probability of observing $A_n = [l_{1n}, \ldots, l_{pn}, \ldots, l_{Pn}]$ where $l_{pn} \in \{1, \ldots, L\}$ given the explanatory variables Z_n, $\mathrm{P}(A_n|Z_n; \gamma_0, \Gamma, \lambda_0, \Lambda, \tau, \Psi)$, is written as:

$$\int_{Z^*} \prod_{p=1}^{P} \left\{ \Phi(\tau_{l_{pn}} - \lambda_{0p} - \lambda_p Z^*) - \Phi(\tau_{l_{pn}-1} - \lambda_{0p} - \lambda_p Z^*) \right\} \quad f(Z^*|Z_n; \gamma_0, \Gamma, \Psi)\, dZ^* \tag{C.5}$$

where λ_{0p} is the p^{th} component of λ_0, and λ_p is the p^{th} row of Λ, $f(Z^*|.)$ is the distribution of the latent variables given Z_n, and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal variate. Consequently, the log-likelihood of a random sample of N observations for the latent variable model with ordered categorical indicators, $\mathcal{L}(\gamma_0, \Gamma, \lambda_0, \Lambda, \tau, \Psi)$, equals

$$\sum_{n=1}^{N} \log \left[\int_{Z^*} \prod_{p=1}^{P} \left\{ \Phi(\tau_{l^p_{pn}} - \lambda_{0p} - \lambda_p Z^*) - \Phi(\tau_{l^p_{pn}-1} - \lambda_{0p} - \lambda_p Z^*) \right\} f(Z^*|Z_n; \gamma_0, \Gamma, \Psi)\, dZ^* \right] \tag{C.6}$$

When the threshold parameters are shared across the P indicators, then one needs to fix only one of the diagonal elements of Θ to 1. For example, assuming that $\xi_n \sim \mathcal{N}(0, \Theta)$ where Θ is a diagonal matrix, with say $\theta_{11} = 1$, the log-likelihood function, $\mathcal{L}(\gamma_0, \Gamma, \lambda_0, \Lambda, \tau, \Psi, \Theta)$ equals

$$\sum_{n=1}^{N} \log \left[\int_{Z^*} \prod_{p=1}^{P} \left\{ \Phi\left(\frac{\tau_{l_{pn}} - \lambda_{0p} - \lambda_p Z^*}{\theta_{pp}} \right) - \Phi\left(\frac{\tau_{l_{pn}-1} - \lambda_{0p} - \lambda_p Z^*}{\theta_{pp}} \right) \right\} \right.$$

$$f(Z^*|Z_n;\gamma_0,\Gamma,\Psi)\,dZ^*\Bigg] \tag{C.7}$$

Estimation of the model parameters by maximizing the likelihood function is numerically feasible for latent variable models with up to three latent variables. Approaches to estimate models with large dimensions of the latent vector may include estimation methods by simulation wherein the likelihood function is approximated by Monte Carlo simulation.

Appendix D

Latent Choice Choice Model for Taste Variations: Estimation Using the EM Algorithm

D.1 Overview of the EM Algorithm

Dempster *et al.* [1977] presented a broadly applicable algorithm for computing maximum likelihood estimates from incomplete data. The term "incomplete data" in its general form implies the existence of two sample spaces $\mathcal{Y}$ and $\mathcal{X}$ and a many-one mapping from $\mathcal{X}$ to $\mathcal{Y}$. The observed data $\boldsymbol{y}$ are a realization from $\mathcal{Y}$. The corresponding $\boldsymbol{x}$ in $\mathcal{X}$ is not observed directly, but only indirectly through $\boldsymbol{y}$. More specifically, assume that there is a mapping $\boldsymbol{x} \rightarrow \boldsymbol{y}$ from $\mathcal{X}$ to $\mathcal{Y}$, and that $\boldsymbol{x}$ is known only to lie in $\boldsymbol{\mathcal{X}}(\boldsymbol{y})$, the subset of $\mathcal{X}$ determined by the equation $\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{x})$, where $\boldsymbol{y}$ is the observed data. $[\boldsymbol{x}, \boldsymbol{y}]$ is referred to as the *complete data*. The general idea behind the algorithm is the approximation of the maximum likelihood estimates through a particular iterative procedure. Each iteration of the algorithm consists of an expectation (E) step followed by a maximization (M) step, and hence is dubbed the EM algorithm. When the underlying complete data comes from an exponential family whose maximum likelihood estimates are easily computed, then each maximization step of an EM algorithm is likewise easily computed. We begin by reviewing the formulation of the general EM algorithm given in Dempster *et al.* [1977].

A family of sampling densities $f([\boldsymbol{x}, \boldsymbol{y}]|\boldsymbol{\Phi}) = h(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\Phi})\tilde{f}(\boldsymbol{x}|\boldsymbol{\Phi})$ depending on parameters $\boldsymbol{\Phi} \in \Omega$ is postulated and its corresponding family of sampling densities $g(\boldsymbol{y}|\boldsymbol{\Phi})$ is derived. The complete-data specification $f(\cdot|\cdot)$ is related to the incomplete-data specification by

$$g(\boldsymbol{y}|\boldsymbol{\Phi}) = \int\limits_{\mathcal{X}(\boldsymbol{y})} h(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\Phi})\tilde{f}(\boldsymbol{x}|\boldsymbol{\Phi})\, d\boldsymbol{x} \tag{D.1}$$

The EM algorithm attempts to find a value of $\boldsymbol{\Phi}$ which maximizes $g(\boldsymbol{y}|\boldsymbol{\Phi})$ given an observed $\boldsymbol{y}$, by making essential use of the associated family $f(\boldsymbol{x}|\boldsymbol{\Phi})$. The conditional density $k(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\Phi})$ defined on $\mathcal{X}(\boldsymbol{y})$ is given by $f([\boldsymbol{x}, \boldsymbol{y}]|\boldsymbol{\Phi})/g(\boldsymbol{y}|\boldsymbol{\Phi})$, so that the incomplete log-likelihood, $L(\boldsymbol{\Phi}) = \log g(\boldsymbol{y}|\boldsymbol{\Phi})$, can be written as:

$$L(\boldsymbol{\Phi}) = \log f([\boldsymbol{x}, \boldsymbol{y}]|\boldsymbol{\Phi}) - \log k(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\Phi}) \tag{D.2}$$

For $\boldsymbol{\Phi}$ and $\boldsymbol{\Phi'}$ in Ω then one has,

$$L(\boldsymbol{\Phi}) = Q(\boldsymbol{\Phi}|\boldsymbol{\Phi'}) - H(\boldsymbol{\Phi}|\boldsymbol{\Phi'}) \tag{D.3}$$

where $Q(\boldsymbol{\Phi}|\boldsymbol{\Phi'}) = E(\log f([\boldsymbol{x}, \boldsymbol{y}]|\boldsymbol{\Phi})|\boldsymbol{y}, \boldsymbol{\Phi'})$ and $H(\boldsymbol{\Phi}|\boldsymbol{\Phi'}) = E(\log k(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\Phi})|\boldsymbol{y}, \boldsymbol{\Phi'})$. Given a current best estimate $\boldsymbol{\Phi}^{(p)}$, the EM iteration $\boldsymbol{\Phi}^{(p)} \rightarrow \boldsymbol{\Phi}^{(p+1)}$ is:

1. **E-step**: Compute $Q(\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(p)})$.

2. **M-step**: Choose $\boldsymbol{\Phi}^{(p+1)}$ to be a value in $\boldsymbol{\Phi} \in \Omega$ which maximizes $Q(\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(\boldsymbol{p})})$.

As noted earlier the algorithm is most useful in situations wherein the maximization of $\log f([\boldsymbol{x}, \boldsymbol{y}]|\boldsymbol{\Phi})$ over $\boldsymbol{\Phi} \in \boldsymbol{\Omega}$ is easy. In such situations, the M-step maximization of $Q(\boldsymbol{\Phi}|\boldsymbol{\Phi'})$ over $\boldsymbol{\Phi} \in \boldsymbol{\Omega}$ can be carried out with relative ease. A fundamental property of the EM algorithm is the monotonicity of the likelihood function along successive iterations (see Lemma 1 and Theorem 1 of Dempster *et al.* [1977]). Wu [1983] studies the aspects of convergence of sequence of estimates generated by the EM algorithm and whether the algorithm finds a global maximum, local maximum or stationary value of the incomplete likelihood function $L(\mathbf{\Phi})$. If the likelihood function is bounded from above then the likelihood sequence $L(\mathbf{\Phi^{(p)}})$ converges to some L^*. There is no guarantee that L^* is the global maximum of $L(\mathbf{\Phi})$. This is because even though a global maximization of $Q(\cdot)$ is involved in the M-step, the other term $H(\cdot)$ in $L(\mathbf{\Phi}) = Q(\cdot) - H(\cdot)$ may not cooperate. Further, even the convergence to a local maximum cannot be satisfactorily addressed without further assumptions[1]. Specifically, Wu [1983] shows that if $Q(\mathbf{\Phi}|\mathbf{\Phi'})$ is continuous in $\mathbf{\Phi}$ and $\mathbf{\Phi'}$, then all the limit points of the sequence $\mathbf{\Phi^{(p)}}$ are stationary points of $L(\mathbf{\Phi})$ and $L(\mathbf{\Phi^{(p)}})$ converges to $L^* = L(\mathbf{\Phi^*})$ for some stationary point $\mathbf{\Phi^*}$. Further, Wu [1984] shows that if $Q(\mathbf{\Phi}|\mathbf{\Phi'})$ has *continuous derivatives* in $\mathbf{\Phi}$ and $\mathbf{\Phi'}$ and $L(\mathbf{\Phi})$ is *unimodal* in Ω with a unique stationary point, then $\mathbf{\Phi^{(p)}}$ converges to the unique maximizer $\mathbf{\Phi^*}$ of $L(\mathbf{\Phi})$.

Further, slightly different sufficient conditions for convergence and a general description of the rate of the convergence of the algorithm close to a stationary point are provided in Dempster *et al.* [1977] (see Theorem 2 of Dempster *et al.* [1977] and Boyles [1983] for the correct versions of the sufficient conditions).

In the above setting, the support for $\boldsymbol{x}$ may either be discrete or continuous, and accordingly $g(\boldsymbol{y}|\mathbf{\Phi})$ is dubbed a finite mixture model or an infinite mixture model. The EM algorithm in the context of finite mixture models, has been derived and studied by a number of authors. Hasselblad [1966, 1969] (arbitrary finite mixtures of univariate normal densities and mixtures of univariate densities from exponential families), Day [1969] and Wolfe [1970] (mixture of two multivariate normal densities with common covariance matrix and arbitrary finite mixtures of multivariate normal densities, respectively).

D.2 Adoption for the Latent Class Choice Model

For the latent class choice model, one can regard the available data as an incom-

[1]If the sequence $\mathbf{\Phi^{(p)}}$ converges to some $\mathbf{\Phi^*}$, then the hessian matrices of $Q(\cdot)$ and $H(\cdot)$ are negative definite at $\mathbf{\Phi^*}$ (the negative definiteness of the Hessian of $H(\cdot)$ follows from Lemma 2 of Dempster *et al.* [1977]). But nothing can be said about the Hessian matrix of $L(\cdot)$ at $\mathbf{\Phi^*}$.

plete data by considering individual's choice indicator $\tilde{y}_n$ to be the "known" part of an observation $y_n = (\tilde{y}_n, s_n)$, where s_n is the index of the individual's latent class. For notational simplicity we drop the socio-economic and demographic characteristics and attributes of alternatives in the $f(\cdot)$ and $g(\cdot)$ density functions. For $\boldsymbol{\Phi} = (\boldsymbol{\theta}, \boldsymbol{\rho}, \boldsymbol{\tau}) \in \Omega$ the sample probability mass functions $g(\tilde{\boldsymbol{y}}|\boldsymbol{\Phi}) = \prod_{n=1}^{N} \Pr(\tilde{y}_n|\boldsymbol{\Phi})$ and $f(\boldsymbol{y}|\boldsymbol{\Phi}) = \prod_{n=1}^{N} \Pr(s_n|\boldsymbol{\theta}, \boldsymbol{\rho}, \boldsymbol{\tau}) \Pr(\tilde{y}_n|s_n, \boldsymbol{\beta_{s_n}})$ respectively[2]. Then for $\boldsymbol{\Phi'} = (\boldsymbol{\theta'}, \boldsymbol{\rho'}, \boldsymbol{\tau'}) \in \Omega$, the conditional probability mass function $k(\boldsymbol{y}|\tilde{\boldsymbol{y}}, \boldsymbol{\Phi'})$ equals

$$\prod_{n=1}^{N} \frac{\Pr(s_n|\boldsymbol{\theta'}, \boldsymbol{\rho'}, \boldsymbol{\tau'}) \Pr(\tilde{y}_n|s_n, \boldsymbol{\beta'_{s_n}})}{\Pr(\tilde{y}_n|\boldsymbol{\Phi'})} \tag{D.4}$$

and the function $Q(\boldsymbol{\Phi}|\boldsymbol{\Phi'})$ is determined to be:

$$\begin{aligned} Q(\boldsymbol{\Phi}|\boldsymbol{\Phi'}) &= \sum_{n=1}^{N} \sum_{s_n=1}^{S} \log \Pr(s_n|\boldsymbol{\theta}, \boldsymbol{\rho}, \boldsymbol{\tau}) \Pr(\tilde{y}_n|s_n, \boldsymbol{\beta_{s_n}}) \frac{\Pr(s_n|\boldsymbol{\theta'}, \boldsymbol{\rho'}, \boldsymbol{\tau'}) \Pr(\tilde{y}_n|s_n, \boldsymbol{\beta'_{s_n}})}{\Pr(\tilde{y}_n|\boldsymbol{\Phi'})} \\ &= \sum_{n=1}^{N} \sum_{s_n=1}^{S} \left[\frac{\Pr(s_n|\boldsymbol{\theta'}, \boldsymbol{\rho'}, \boldsymbol{\tau'}) \Pr(\tilde{y}_n|s_n, \boldsymbol{\beta'_{s_n}})}{\Pr(\tilde{y}_n|\boldsymbol{\Phi'})} \right] \log \Pr(s_n|\boldsymbol{\theta}, \boldsymbol{\rho}, \boldsymbol{\tau}) \\ &\quad + \sum_{n=1}^{N} \sum_{s_n=1}^{S} \left[\frac{\Pr(s_n|\boldsymbol{\theta'}, \boldsymbol{\rho'}, \boldsymbol{\tau'}) \Pr(\tilde{y}_n|s_n, \boldsymbol{\beta'_{s_n}})}{\Pr(\tilde{y}_n|\boldsymbol{\Phi'})} \right] \log \Pr(\tilde{y}_n|s_n, \boldsymbol{\beta_{s_n}}) \end{aligned} \tag{D.5}$$

As seen in equation (D.5) the maximization problem has some attractive features. Now we have a separable maximization problem with the first term in equation (D.5) involves the parameters $\boldsymbol{\theta}, \boldsymbol{\rho}, \boldsymbol{\tau}$ alone and the second term involves the class-specific choice model parameters $\boldsymbol{\beta_s}, \ \forall s = 1, \ldots, S$. Under the assumption that the random components in the criterion functions are multivariate normally distributed, by fixing the $\boldsymbol{\rho}$ parameter vector, the log-concavity of the class membership model in $[\boldsymbol{\theta}, \boldsymbol{\tau}]$ is ensured due to the log-concavity of the generating multivariate density (see appendix E for a proof). Consequently, the unicity of the maximum point is ensured if the fixed correlation matrix does not lie on the boundary of the space of $D \times D$ correlation matrices, where D equals the number of criterion functions.

Further, if the class-specific choice model parameters are not shared across the

[2] $\boldsymbol{\theta}$ are the structural parameters in the criterion functions, $\boldsymbol{\rho}$ are the parameters in the density function of the random components of the criterion functions, and $\boldsymbol{\tau}$ are threshold parameters.

classes, then the second term separates into S component problems, each of which involves only one of the parameter vectors $\boldsymbol{\beta_s}$. Even if the class-specific parameters are shared, if one assumes that the choice process in each class is governed by a multinomial logit model (MNL) then the second maximization problem is a concave function since it is a convex combination of concave functions since the likelihood function for each class is concave with respect to the class-specific parameters.

One can view the component problems and the maximization problem as a weighted maximum likelihood estimation problem involving sums of logarithms weighted by posterior probabilities that sample observations belong to appropriate classes, given the current best estimate of $\mathbf{\Phi}$.

Appendix E

Issues of Identification and Estimation in Latent Class Choice Models and Agent-effects Models for Panel Data

E.1 Latent Class Choice Model with MNL-type Class Membership Model

Here we assume that each latent class s is associated with an underlying criterion function H_{sn}, where

$$H_{sn} = \theta'_s Z_n + \delta_{sn}, \qquad \forall s = 1, \ldots, S. \tag{E.1}$$

where Z_n are the characteristics of individual n which affect class membership. Further, assuming a criterion function maximizing assignment process the indicator function for the latent class s is written as:

$$l^*_{sn} = \begin{cases} 1 & \text{if} \quad H_{sn}(Z_n, \delta_{sn}; \theta_s) = \max\limits_{\forall s'=1,\ldots,S} \{H_{s'n}(Z_n, \delta_{s'n}; \theta_{s'})\} \\ 0 & \text{otherwise} \end{cases} \tag{E.2}$$

Then by specifying a probability density function for $(\delta_{1n}, \ldots, \delta_{Sn})$, a probabilistic class membership model can be constructed. For example, if the random variables, δ_{sn}, $\forall s = 1, \ldots, S$, are independently and identically distributed Gumbel (0,1) random variables we obtain the Multinomial Logit Model-type class membership model, i.e.,

$$\begin{aligned} Q_s(Z_n;\theta) &= \mathrm{P}(l^*_{sn} = 1|Z_n, \theta) \\ &= \frac{\exp(\theta'_s Z_n)}{\sum\limits_{s'=1}^{S} \exp(\theta'_{s'} Z_n)}. \end{aligned} \tag{E.3}$$

where $Q_s(Z_n;\theta)$ denotes the class membership model. It must be noted that in order to identify the class membership model we need to set $\gamma_s = 0$ for some s, say $s = 1$. For notational simplicity the data matrix can be modified so that $\theta = [\theta'_2, \ldots, \theta'_J]'$, and

$$Q_s(Z_n;\theta) = \frac{\exp(\theta' Z_{sn})}{\sum\limits_{s'=1}^{S} \exp(\theta' Z_{s'n})} \tag{E.4}$$

If every class s has its own taste parameter vector, then the latent class choice model which expresses the probability of choosing alternative i is written as:

$$P(i|X_n, Z_n; \theta, \beta) = \sum_{s=1}^{S} P(i|X_n; \beta_s) Q_s(Z_n; \theta) \tag{E.5}$$

Let $P(i|\beta,\theta) \equiv P(i|X_n, Z_n; \theta, \beta)$, $P(i|s) \equiv P(i|X_n; \beta_s)$, and $Q(s) \equiv Q_s(Z_n;\theta)$. Differentiating equation (E.5) with respect to β_s and θ we obtain:

$$\nabla_{\beta_s} P(i|\theta,\beta) = Q(s) \nabla_{\beta_s} P(i|s), \ \forall s = 1, \ldots, S \tag{E.6}$$

$$\nabla_\theta P(i|\theta,\beta) = \sum_{s=1}^{S} P(i|s) \nabla_\theta Q(s) \tag{E.7}$$

The second derivatives are given by:

$$\nabla^2_{\beta_s\beta_s} P(i|\theta,\beta) = Q(s) \nabla^2_{\beta_s\beta_s} P(i|s), \ \forall s = 1, \ldots, S \tag{E.8}$$

$$\nabla^2_{\beta_s\beta_{s'}} P(i|\theta,\beta) = 0 \; \forall s \neq s' \tag{E.9}$$

$$\nabla^2_{\beta_s\theta} P(i|\theta,\beta) = \nabla_{\beta_s} P(i|s) \nabla_{\theta'} Q(s) \tag{E.10}$$

$$\nabla^2_{\theta\theta} P(i|\theta,\beta) = \sum_{s=1}^{S} P(i|s) \nabla^2_{\theta\theta} Q(s) \tag{E.11}$$

Hence the Hessian of the likelihood of single observation is written as:

$$\begin{pmatrix} \nabla^2_{\beta_1\beta_1} P(i|\theta,\beta) & 0 & \cdots & 0 & \nabla^2_{\beta_1\theta} P(i|\theta,\beta) \\ & \nabla^2_{\beta_2\beta_2} P(i|\theta,\beta) & \cdots & 0 & \nabla^2_{\beta_2\theta} P(i|\theta,\beta) \\ & & \ddots & \vdots & \vdots \\ \text{symmetric} & & & \nabla^2_{\beta_S\beta_S} P(i|\theta,\beta) & \nabla^2_{\beta_S\theta} P(i|\theta,\beta) \\ & & & & \sum_{s=1}^{S} P(i|s) \nabla^2_{\theta\theta} Q(s) \end{pmatrix}$$

Given θ, and assuming that the class-specific choice model is an MNL then the likelihood function is concave in β. Further, noting that the MNL-type class membership model is log-concave in θ, and the class-specific MNL model is log-concave in β, the likelihood function of the latent class choice model is a summation of log-concave functions since the product of log-concave functions is log-concave. To characterize the sum of log-concave functions, we prove the following lemma.

Lemma 1 *If $f(X)$ and $g(X)$ are log-concave functions, their sum $f(X) + g(X)$ is also log-concave.*

Proof: By log-concavity of $f(X)$ and $g(X)$ we have for two points X_1 and X_2

$$f\left(\frac{X_1 + X_2}{2}\right)^2 \geq f(X_1) f(X_2) \tag{E.12}$$

and

$$g\left(\frac{X_1 + X_2}{2}\right)^2 \geq g(X_1) g(X_2). \tag{E.13}$$

Our objective is to show that

$$\left(f(\frac{X_1 + X_2}{2}) + g(\frac{X_1 + X_2}{2})\right)^2 \geq (f(X_1) + g(X_1))(f(X_2) + g(X_2)) \tag{E.14}$$

Let $f(X_1) = a_1$, $f(X_2) = a_2$, $f(\frac{X_1+X_2}{2}) = a_3$, $g(X_1) = b_1$, $g(X_2) = b_2$, and $g(\frac{X_1+X_2}{2}) = b_3$. Then, we need to show that

$$(a_3 + b_3)^2 \geq (a_1 + b_1)(a_2 + b_2)$$
$$(a_3 + b_3)^2 - (a_1 + b_1)(a_2 + b_2) \geq 0$$

Since we have

$$\begin{aligned} & a_1 b_1((a_3 + b_3)^2 - (a_1 + b_1)(a_2 + b_2)) \\ = \; & a_1(a_1 + b_1)(b_3^2 - b_1 b_2) + b_1(a_1 + b_1)(a_3^2 - a_1 a_2) + (a_1 b_3 - b_1 a_3)^2 \end{aligned} \quad \text{(E.15)}$$

and noting that $a_1 > 0$, $b_1 > 0$, $(a_3^2 - a_1 a_2) \geq 0$ and $(b_3^2 - b_1 b_2) \geq 0$, the lemma follows. ∎

Employing Lemma 1 to the log-likelihood function, the following proposition follows:

Proposition 1 *The log-likelihood function for the latent class choice model with an MNL-type categorical criterion class membership model and class-specific MNL model is strictly concave in* θ*, where* θ *includes the parameters in the class-specific choice model, and the parameters in the class membership model.*

Proof: Trivial. ∎

E.2 Latent Class Choice Model with Ordinal Criteria Class Membership Model

As the first step in the characterization of the likelihood function of the latent class choice model with ordinal criteria class membership model, we analyze the multivariate ordinal probit model, also referred to as the Grouped Continuous Model for Multivariate Ordered Categorical Variables in the biometrics literature (see Anderson and Pemberton [1985]). Herein we *observe* for individual n , the ordinal categorical vector $\boldsymbol{y_n} = [y_{1n}, \ldots, y_{Dn}]'$, with $y_{dn} \in \mathcal{L} = \{1, \ldots, L_d\}$, $\forall d$, where L_d is the number of levels in the d^{th} dimension. We postulate the existence of an unobservable (latent) variable for each dimension d, denoted H_{dn}'s, such that

$$H_{dn} = \theta_d' Z_n + \delta_{dn}, \quad \forall d = 1, \ldots, D \tag{E.16}$$

where θ_d is an unknown parameter vector and δ_{dn} is a random component. Further, correlations between the random components of H_d's (i.e., δ_d's) are allowed to capture the *unobserved* interrelationships between the different dimensions. The observed levels in each dimension are associated with the latent variables as follows:

$$\boldsymbol{y_n} = [y_{1n}, \ldots, y_{Dn}]' \;\Leftrightarrow\; \left\{ \left(\tau^d_{y_{dn}-1} \leq H_{dn} \leq \tau^d_{y_{dn}} \right),\ \forall d = 1, \ldots, D \right\},$$
$$\forall [y_{1n}, \ldots, y_{Dn}]' \in \{\mathcal{L}_1 \times \mathcal{L}_2 \times \cdots \times \mathcal{L}_D\} \tag{E.17}$$

where

$\tau^d_{y_{dn}-1}$	-	lower bound value (threshold) for the latent variable of dimension d when the corresponding level is y_{dn}
$\tau^d_{y_{dn}}$	-	upper bound value (threshold) for the latent variable of dimension d when the corresponding level is y_{dn}

Assuming

$$\begin{pmatrix} \delta_{1n} \\ \vdots \\ \delta_{Dn} \end{pmatrix} \sim \mathcal{N}(0, \boldsymbol{R}) \tag{E.18}$$

where $\boldsymbol{R}$ is a correlation matrix[1], the probability of observing $[y_{1n}, \ldots, y_{Dn}]'$ is given by:

$$\int_{v^+_{1n}}^{v^-_{1n}} \cdots \int_{v^+_{Dn}}^{v^-_{Dn}} f_{\delta_1, \ldots, \delta_D}(u_1, \cdots, u_D)\, du_1 \cdots du_D \tag{E.19}$$

[1]It is necessary to set the scale of each latent variable, and this is usually done by fixing the variance of the random components of the latent variables to unity.

where $f(\cdot)$ is the density function of the multivariate normal density, $v_{dn}^{+} = \tau_{y_{dn}-1}^{d} - \theta_d' Z_n$ and $v_{dn}^{-} = \tau_{y_{dn}}^{d} - \theta_d' Z_n$ for $d = 1, \ldots, D$. In addition, τ_0^d $\forall d$ are set to $-\infty$, $\tau_{L_d}^1$ $\forall d$ to $+\infty$, and τ_1^d $\forall d$ are arbitrarily set to zero to fix the origin of the latent variables.

It must be noted that multivariate ordinal probit model is a multivariate generalization of the ordinal probit model of McKelvey and Zavoina [1975], an extension of the multivariate probit model of Ashford and Sowden [1970] which allowed for two levels in each dimension precluding the specification of estimable thresholds.

Haberman [1980] and Pratt [1981] have shown that the log-likelihood function of the ordinal probit model of McKelvey and Zavoina [1975] is concave. We generalize the result for the multivariate ordinal probit model in Proposition 2. It must be noted that the uniqueness of the correlation parameters are not yet established.

Proposition 2 *The log-likelihood function for the multivariate ordinal probit model is strictly concave in θ.*

Proof: First we show the concavity of the function:

$$h(v_{1n}^{+}, \ldots, v_{Dn}^{+}, v_{1n}^{-}, \ldots, v_{Dn}^{-}) = \log \int_{v_{1n}^{+}}^{v_{1n}^{-}} \cdots \int_{v_{Dn}^{+}}^{v_{Dn}^{-}} f_{\delta_1, \ldots, \delta_D}(u_1, \cdots, u_D)\, du_1 \cdots du_D \quad \text{(E.20)}$$

as a function of $(v_{1n}^{+}, \ldots, v_{Dn}^{+}, v_{1n}^{-}, \ldots, v_{Dn}^{-})$ for $v_{dn}^{-} > v_{dn}^{+}$, $\forall d$. Defining indicator functions $\mathcal{I}_d(u_d, v_{dn}^{-}, v_{dn}^{+})$ such that:

$$\mathcal{I}_d(u_d, v_{dn}^{-}, v_{dn}^{+}) = \begin{cases} 1 & v_{dn}^{+} < u_d < v_{dn}^{-} \\ 0 & \text{otherwise} \end{cases} \quad \text{(E.21)}$$

the probability can be written as:

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathcal{I}_1(u_1, v_{1n}^{-}, v_{1n}^{+}) \cdots \mathcal{I}_D(u_D, v_{Dn}^{-}, v_{Dn}^{+}) f_{\delta_1, \ldots, \delta_D}(u_1, \cdots, u_D)\, du_1 \cdots du_D \quad \text{(E.22)}$$

Noting that $\log \mathcal{I}_d(u_d, v_{dn}^{-}, v_{dn}^{+})$ is concave in $(u_d, v_{dn}^{-}, v_{dn}^{+})$, and $\log f$ is concave in $(u_1, \cdots, u_D)$ for the multivariate normal density (Prékopa [1971]) we have employing Theorem 6 of Prékopa[2] [1972, pp. 342] , that h is concave in $(v^+_{1n}, \ldots, v^+_{Dn}, v^-_{1n}, \ldots, v^-_{Dn})$. Since the v's are linear functions of θ, as long as the inequality restrictions are satisfied, it follows that $\log P$ in concave in θ. ■

From Proposition 2, we note that if the class-specific choice model is MNL, then the likelihood function of latent class choice model with ordinal criteria membership model is a summation of log-concave functions for fixed $\boldsymbol{R}$.

Employing Lemma 1 to the log-likelihood function, the following proposition follows:

Proposition 3 *The log-likelihood function for the latent class choice model with ordinal criteria membership model and class-specific MNL model is strictly concave in θ, where θ includes the parameters in the class-specific choice model, and the structural parameters in the class membership model.*

Proof: Trivial. ■

E.3 Discrete Panel Data Models

In the presence of panel data, it is possible to capture heterogeneity in preferences among individuals through individual-specific effects since we have multiple responses from the same individual. There are two approaches to estimating the model parameters in such a case. The first approach, referred to as the "fixed-effects" model in the literature (Chamberlain [1980]), assumes that the alternative specific constants (ASCs) in the choice model are individual-specific. If the model has the full set of ASCs and N individuals, the model necessitates availability of sufficiently long spells per individual and the estimation of large number of parameters $((J-1)N+K)$ where J is the number of alternatives, and K is the number of "other" parameters in the utility function. The second, more tractable approach, is a model wherein the individual-specific effects are assumed to be distributed in the sample. Such a specification is referred to as the "random-effects" model (also called "agent effects" model)

[2]The integral of a log concave function with respect to some of the arguments is a log-concave function of its remaining arguments.

(Heckman [1981], Hsiao [1986]). Herein we focus on the random-effects model and discuss some necessary identification restrictions not enunciated in the literature. To motivate the necessary restrictions for identification consider a binary choice problem. The utility functions for the two alternatives are written as:

$$U_{itn} = \beta' X_{itn} + \nu_{in} + \epsilon_{itn}, \quad i = 1, 2 \tag{E.23}$$

where ν_{in} is the individual-specific error component for alternative i representing individual's unobserved intrinsic preferences towards the alternative, and subscript t denotes the time period of observation. Assume that the ASC's are included in the β parameter vector and that

$$\begin{pmatrix} \nu_{1n} \\ \nu_{2n} \end{pmatrix} \sim \mathcal{BVN}\left(0, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right) \tag{E.24}$$

Assuming that $\epsilon_{itn} \sim$ Gumbel$(0,1)$, the conditional choice model given (ν_{1n}, ν_{2n}) is represented by a binary logit model:

$$\begin{aligned} \mathrm{P}(y_{tn}(2) = 1 | \nu_{1n}, \nu_{2n}) &= \frac{\exp(\beta' X_{2tn} + \nu_{2n})}{\exp(\beta' X_{1tn} + \nu_{1n}) + \exp(\beta' X_{2tn} + \nu_{2n})} \\ &= \frac{\exp(\beta' X_{2tn} + (\nu_{2n} - \nu_{1n})}{\exp(\beta' X_{1tn}) + \exp(\beta' X_{2tn} + (\nu_{2n} - \nu_{1n}))} \end{aligned} \tag{E.25}$$

where $y_{tn}(i)$, for $i = 1, 2$ denotes the choice indicator taking the value 1 if alternative i is chosen at time t and 0 otherwise. Let $\tilde{\nu}_n = (\nu_{2n} - \nu_{1n})$, and by the distributional assumptions for ν's, $\tilde{\nu}_n \sim \mathcal{N}(0, \tilde{\sigma}^2)$ where $\tilde{\sigma}^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$. Therefore, we can estimate from the data only $\tilde{\sigma}$, while σ_1, σ_2 and ρ are not estimable.

The preceding argument stems from the fact that the choice probabilities depend only on the utility differences, and it naturally extends over to the multinomial choice case. Consider a situation wherein an individual n picks an alternative from the index set $\{1, \ldots, J\}$ over t_n occasions. We assume for simplicity that at each occasion the entire set of alternatives is available to each individual. Let $Y_n = (Y_{1n}, \ldots, Y_{t_n n})'$ denote the choice history, where $Y_{tn} = (y_{tn}(1), y_{tn}(2), \ldots, y_{tn}(J))'$ is a vector of binary variables indicating the alternative chosen by the individual at time t.

The utilities can be written in a compact form as:

$$\boldsymbol{U_{tn}} = \boldsymbol{X_{tn}\beta} + \boldsymbol{\nu_n} + \boldsymbol{\epsilon_{tn}} \tag{E.26}$$

where $\boldsymbol{\nu_n} \sim MVN(0, \Sigma)$. Given ϵ_{itn}'s are independently and identically distributed Gumbel (0,1), the probability of individual n choosing alternative i on occasion t conditioned on the individual-specific error component $\boldsymbol{\nu_n}$ is given by the multinomial logit model[3], i.e.,

$$\begin{aligned}
\mathrm{P}(y_{tn}(i) = 1 | X_{tn}, \boldsymbol{\nu_n}) &= \frac{\exp(\beta' X_{itn} + \nu_{in})}{\sum\limits_{j=1}^{J} \exp(\beta' X_{jtn} + \nu_{jn})} \\
&= \frac{\exp(\beta' X_{itn} + (\nu_{in} - \nu_{Jn}))}{\sum\limits_{j=1}^{J} \exp(\beta' X_{jtn} + (\nu_{jn} - \nu_{Jn}))} \\
&= \frac{\exp(\beta' X_{itn} + \tilde{\nu}_{in})}{\sum\limits_{j=1}^{J} \exp(\beta' X_{jtn} + \tilde{\nu}_{jn})}
\end{aligned} \tag{E.27}$$

There exists a linear mapping matrix $\Delta_J : (J-1) \times J$ that operates on $\boldsymbol{\nu_n}$ to obtain the differenced $\tilde{\boldsymbol{\nu}}_n$. Further $\tilde{\boldsymbol{\nu}}_n \sim MVN(0, \tilde{\Sigma})$ where

$$\tilde{\Sigma} = \Delta_J \Sigma \Delta_J' \tag{E.28}$$

Let $G(\cdot)$ and $\tilde{G}(\cdot)$ denote the cumulative distribution functions[4] of $\boldsymbol{\nu_n}$ and $\tilde{\boldsymbol{\nu}}_n$. The unconditional probability of observing the sequence $(Y_{1n}, \ldots, Y_{t_n n})'$ can be written as:

$$\begin{aligned}
&= \int\limits_{\Re^J} \left\{ \prod_{t=1}^{t_n} P_n(Y_{tn} | X_{tn}, \boldsymbol{\nu}) \right\} dG(\boldsymbol{\nu}) \\
&= \int\limits_{\Re^{J-1}} \left\{ \prod_{t=1}^{t_n} P_n(Y_{tn} | X_{tn}, \tilde{\boldsymbol{\nu}}) \right\} d\tilde{G}(\tilde{\boldsymbol{\nu}})
\end{aligned}$$

---

[3] For notational simplicity, we assume that all the J alternatives are available to each individual.

[4] The difference induces a new distribution whose density $\tilde{g}(\cdot)$ can be easily derived from the density $g(\cdot)$ by the linear transformation with Jacobian equal to unity.

Since from the data we can estimate $\tilde{\Sigma}$ the question reduces to whether we can identify Σ from $\tilde{\Sigma}$. Also, noting that Σ has $J(J+1)/2$ nonredundant parameters, and $\tilde{\Sigma}$ has $J(J-1)/2$ parameters, we need to restrict J parameters in Σ. The necessary and sufficient conditions for this problem closely resemble conditions for the identifiability of the variance-covariance matrix of random component in the Multinomial Probit Model (see Ben-Akiva and Bolduc [1991]). Obviously restricting the ν's to only $J-1$ utility functions ensures the identification of the variance-covariance matrix of the individual-specific error components. But the question remains as to which alternative may be set as the base alternative. This issue is important if one notices that even if the pure random components of the utility functions are assumed to be homoscedastic, the error-component structure induces heteoscedasticity in utilities. Consequently, by setting one alternative as the base sets the corresponding random component's variance lower than the variance of the random components of all other utility functions. It must be noted this problem is analogous to the problem of setting the base alternative in choice models estimated on non-rectangular data[5] wherein the precision with which the alternative specific constants are estimated usually depends on the base alternative. The ad hoc approaches to address the issue may include:

1. Estimating choice models with each of the alternatives set as the base alternative. Goodness of fit statistics from each of the models may be used to pick the "best" model.

2. It is computationally convenient to pick the alternative which is available to *all* the individuals in the sample as the base alternative.

It must be noted such considerations are often ignored with agent-effects models (see for example, Chintagunta *et al.* [1991]).

[5]In rectangular choice data, the universal set of alternatives is available to *every* individual. If each choice set varies across individuals then the choice data is usually referred to as non-rectangular. In non-rectangular data it is often suggested to use as the base the alternative which is most often *available*.

Appendix F

Random Coefficients Multinomial Logit Model: A Factor Analytic Representation

The basic idea in the random coefficients model is the assumption that each individual n has his/her own taste parameter β_n vector which differs from the "average" parameter vector $\bar{\beta}$ for the "representative" individual by an unknown (hence random) amount. Assuming a parametric distribution $f(\beta; \Theta)$ for the taste parameter vector, the choice model is given as:

$$\mathrm{P}(i|X;\Theta) = \int \mathrm{P}(i|X;\beta)\, f(\beta;\Theta)\, d\beta \tag{F.1}$$

It must be noted that the estimation of the MNL model with random coefficients is difficult as the choice probability calculation entails the evaluation of a multi-dimensional integral since a model with K random coefficients is expressed as a K-dimensional integral. Consequently, to reduce the dimensionality as well as to capture the interrelationships between the different random coefficients in a systematic manner we postulate that the randomness in K coefficients may be generated through

$M \ll K$ factors which may be assumed to be white noise. Specifically, let

$$\beta_n = \tilde{\beta} + \Lambda\xi \tag{F.2}$$

where Λ is a $K \times M$ parameter matrix, and ξ is a $M \times 1$ random vector. The specification of the free and fixed parameters in Λ depends on the the correlation structure postulated between the random coefficients, and we illustrate it with an example.

Consider a utility specification with five random taste parameters. Further, assume that parameters 1, 3 and 5 may be correlated with each other, parameters 2 and 3 may be correlated, while there is no correlation between any two parameters from the two different subsets, $\{1, 3, 5\}$ and $\{2, 4\}$. Consequently, the random coefficient vector may be written as:

$$\begin{pmatrix} \beta_{1n} \\ \beta_{2n} \\ \beta_{3n} \\ \beta_{4n} \\ \beta_{5n} \end{pmatrix} = \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \tilde{\beta}_3 \\ \tilde{\beta}_4 \\ \tilde{\beta}_5 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ \lambda_{51} & 0 \end{pmatrix} \begin{pmatrix} \xi_{1n} \\ \xi_{2n} \end{pmatrix} \tag{F.3}$$

If we assume that $\xi_{mn} \sim \mathcal{N}(0, \sigma_m^2)$, we should fix one element in each column of Λ to 1. To see why this is necessary, the variance-covariance matrix of β_n equals

$$\begin{pmatrix} \sigma_1^2 & 0 & \lambda_{31}\sigma_1^2 & 0 & \lambda_{51}\sigma_1^2 \\ 0 & \sigma_2^2 & 0 & \lambda_{42}\sigma_2^2 & 0 \\ \lambda_{31}\sigma_1^2 & 0 & \lambda_{31}^2\sigma_1^2 & 0 & \lambda_{31}\lambda_{51}\sigma_1^2 \\ 0 & \lambda_{42}\sigma_2^2 & 0 & \lambda_{42}^2\sigma_2^2 & 0 \\ \lambda_{51}\sigma_1^2 & 0 & \lambda_{31}\lambda_{51}\sigma_1^2 & 0 & \lambda_{51}^2\sigma_1^2 \end{pmatrix} \tag{F.4}$$

Thus, if λ_{11} is a free parameter we can atmost identify $\text{var}(\beta_{1n}) = \lambda_{11}^2\sigma_1^2$, but not λ_{11} and σ_1^2 uniquely. Hence, it is necessary to either fix λ_{11} or σ_1^2 to some constant. Consequently, if σ_1^2 is a free parameter, then λ_{11} may be conveniently fixed to 1. In similar vein, λ_{22} is set to 1.

The main advantage of the factor analytic representation for the generation of the random coefficients is the reduction of the dimensionality of the integration involved in the calculation of the likelihood function. Further, if the analyst can meaningfully impose some correlation structure among the random coefficients, then the factor analytic representation significantly reduces the number of parameters estimated. For example, if an alternative attribute is specified as alternative-specific then we might expect correlations among the corresponding J alternative-specific random coefficients.

Appendix G

Some Characterizations of a Class of Dynamic Choice Models

Before proceeding with the details of the class of dynamic choice models, we present a general overview of the underlying processes in choice dynamics. The key necessary features of dynamic choice models include:

- They must capture systematically the effects of *past* choice behavior on *future* choice behavior; and
- At each time period or choice occasion the individual *updates* his/her *information state* and has varying propensities to either continue with the previous choice or switch to a new alternative or the analyst uses each individual's choice history to update the prior preference structure assumed for that individual.

Different types of assumptions on the *prior information* lead to specific types of dynamic choice models. Some of the examples of prior information might include:

- Prior distribution of taste parameters;
- Prior distribution of latent (unobserved) classes such as choice sets, alternative loyalty, market segments, etc.

It must be noted that the prior information is statistical in nature and hence can be characteristized through some probability measure with a set of unknown parameters.

As an illustrative example, we restrict our attention to dynamic choice model with update on taste parameter distributions, while adoption to other types is transparent.

Dynamic choice model with update on taste parameters

Consider a situation wherein an individual n picks an alternative from the index set $\{1, \ldots, J\}$ over t_n time periods. We assume for simplicity that at time period the entire set of alternatives is available for each individual. Let $Y_n = (Y_{1n}, \ldots, Y_{t_n n})'$ denote the choice history, where $Y_{t'n} = (y_{t'n}(1), y_{t'n}(2), \ldots, y_{t'n}(J))'$ is a vector of binary variables indicating the alternative chosen by the individual at time $t' \in \{1, \ldots, t_n\}$.

At time $t = 1$, when the analyst starts making observations of the choices made by individuals, unobserved heterogeneity in the underlying choice process may exist across the members of the population. Assuming a random utility model for the choice process, the utility of alternative i to individual n at time t' can be written as:

$$U_{it'n} = \beta'_{t'n} X_{it'n} + \epsilon_{it'n} \tag{G.1}$$

We assume that $\epsilon_{it'n}$ are independent and identically distributed across time for the same individual, while these error components may be correlated across alternatives. The utilities can be written in a compact form as:

$$U_{t'n} = X_{t'n} \beta_{t'n} + \epsilon_{t'n} \tag{G.2}$$

It may be reasonable to assume, given the lack of prior information, that at the first time period, the taste parameters are randomly distributed across the sample. After the first time period, one expects to "learn" or gain experience, and thereby leading to an "update" of the distribution of the taste parameters. This updating mechanism is expected to be function of the first choice. The alternating sequence of choice and updating mechanisms continues till we reach the end of the choice history for an individual. Hence, even if the distribution of the taste parameters are identical across the individuals at the start of the process, they vary at subsequent time periods depending on each individual's choice history up to that time. Let $f_{t'n}(\beta)$ denote the density of the taste parameter vector at time t'. We need a relationship such as:

$$f_{t'n}(\beta) = g(f_{1n}(\beta), Y_{1n}, Y_{2n}, \ldots, Y_{t'-1,n}, X_{1n}, \ldots, X_{t'-1,n}), \quad t' \in \{1, \ldots, t_n\} \tag{G.3}$$

More specifically, we are looking for a *sufficient statistic* which summarizes all the essential bits of information up to time t' such as choices and the explanatory variables. Such a relationship can be constructed by using a Bayesian updating scheme. For example, if the initial density function is $f_{1n}(\beta) = f(\beta)\ \forall n$, then the density function at the second time period, $f_{2n}(\beta)$ is obtained as:

$$f_{2n}(\beta) = \frac{\Pr(Y_{1n}|X_{1n};\beta)f(\beta)}{\int \Pr(Y_{1n}|X_{1n};\beta)f(\beta)d\beta} \tag{G.4}$$

and the choice probability at the second time period equals

$$\begin{aligned} \Pr(Y_{2n}|X_{2n};f_{2n}(\beta)) &= \int \Pr(Y_{2n}|X_{2n};\beta)f_{2n}(\beta)d\beta \\ &= \frac{\int \Pr(Y_{2n}|X_{2n};\beta)\Pr(Y_{1n}|X_{1n};\beta)f(\beta)d\beta}{\int \Pr(Y_{1n}|X_{1n};\beta)f(\beta)d\beta} \end{aligned} \tag{G.5}$$

In this fashion, the density function $f_{t'n}(\beta)$ at any time t' is written as:

$$f_{t'n}(\beta) = \frac{\prod_{s=1}^{t'-1}\Pr(Y_{sn}|X_{sn};\beta)f(\beta)}{\int \prod_{s=1}^{t'-1}\Pr(Y_{sn}|X_{sn};\beta)f(\beta)d\beta} \tag{G.6}$$

and the choice probability at t' equals

$$\Pr(Y_{t'n}|X_{t'n};f_{t'n}(\beta)) = \frac{\int \prod_{s=1}^{t'}\Pr(Y_{sn}|X_{sn};\beta)f(\beta)d\beta}{\int \prod_{s=1}^{t'-1}\Pr(Y_{sn}|X_{sn};\beta)f(\beta)d\beta} \tag{G.7}$$

Assuming that the attributes are time-invariant (i.e., $X_{itn} = X_{in}, \forall t$) then one would expect the probability of choosing an alternative i to increase to 1 as the individual choice history contains a sufficiently large number of alternative i choices. Consider a situation wherein an individual chooses alternative i every time until time $t' - 1$.

The probability of choosing alternative i at time t' is written as[1]:

$$\Pr(y_{t'n}(i) = 1|X_{tn}; f_{t'n}(\beta)) = \frac{\int \{\Pr(y_n(i) = 1|X_n;\beta)\}^{t'} f(\beta)d\beta}{\int \{\Pr(y_n(i) = 1|X_n;\beta)\}^{t'-1} f(\beta)d\beta} \tag{G.8}$$

The change in the conditional choice probability from t' to $t'+1$ equals

$$\begin{aligned} &\frac{\int \{\Pr(y_n(i) = 1|X_n;\beta)\}^{t'+1} f(\beta)d\beta}{\int \{\Pr(y_n(i) = 1|X_n;\beta)\}^{t'} f(\beta)d\beta} - \frac{\int \{\Pr(y_n(i) = 1|X_n;\beta)\}^{t'} f(\beta)d\beta}{\int \{\Pr(y_n(i) = 1|X_n;\beta)\}^{t'-1} f(\beta)d\beta} \\ = &\frac{\left[\begin{array}{c} \{\int \Pr(y_n(i) = 1|X_n;\beta)\}^{t'+1} f(\beta)d\beta \quad \int \{\Pr(y_n(i) = 1|X_n;\beta)\}^{t'-1} f(\beta)d\beta \;- \\ \left(\int \{\Pr(y_n(i) = 1|X_n;\beta)\}^{t'} f(\beta)d\beta\right)^2 \end{array}\right]}{\int \{\Pr(y_n(i) = 1|X_n;\beta)\}^{t'} f(\beta)d\beta \quad \int \{\Pr(y_n(i) = 1|X_n;\beta)\}^{t'-1} f(\beta)d\beta} \end{aligned} \tag{G.9}$$

Noting that $\ln \int \{\Pr(y_n(i) = 1|X_n;\beta)\}^{t'} f(\beta)d\beta$ is a convex[2] function in t' (see Hardy *et al.* [1952]), we obtain that the numerator in equation (G.9) is positive. Hence the conditional probability monotonically increases in t'. Since the conditional probability is bounded between 0 and 1, the limit of the conditional probability sequence exists, i.e.,

$$\lim_{t'\to\infty} \frac{\int \{\Pr(y_n(i) = 1|X_n;\beta)\}^{t'} f(\beta)d\beta}{\int \{\Pr(y_n(i) = 1|X_n;\beta)\}^{t'-1} f(\beta)d\beta} = 1 \tag{G.10}$$

Therefore the model behaves consistently with our behavioral expectations. Note further that the foregoing analysis applies to any choice history wherein all alternatives other than alternative i are chosen only a countably finite number of time periods.

The likelihood for the choice history Y_n, $\Pr(Y_n|X_{1n},\ldots,X_{t_n n})$, equals

$$\begin{aligned} &\Pr(Y_{1n}|X_{1n})\Pr(Y_{2n}|Y_{1n},X_{1n},X_{2n})\cdots\Pr(Y_{t_n n}|Y_{1n},\ldots,Y_{t_n-1,n},X_{1n},\ldots,X_{t_n-1,n}) \\ = &\left\{\int \Pr(Y_{1n}|X_{1n};\beta)f(\beta)d\beta\right\}\left\{\frac{\int \Pr(Y_{2n}|X_{2n};\beta)\Pr(Y_{1n}|X_{1n};\beta)f(\beta)d\beta}{\int \Pr(Y_{1n}|X_{1n};\beta)f(\beta)d\beta}\right\}\cdots \\ &\left\{\frac{\int \prod_{s=1}^{t}\Pr(Y_{sn}|X_{sn};\beta)f(\beta)d\beta}{\int \prod_{s=1}^{t-1}\Pr(Y_{sn}|X_{sn};\beta)f(\beta)d\beta}\right\} \end{aligned}$$

[1] Note that by the assumption of time-invariant attributes, for any time t $\Pr(y_{tn}(i) = 1|X_{tn};\beta) = \Pr(y_n(i) = 1|X_n;\beta)$.

[2] The convexity result follows from a generalization of Hölder's inequality.

$$= \int \prod_{s=1}^{t} \Pr(Y_{sn}|X_{sn};\beta) f(\beta) d\beta \tag{G.11}$$

Thus, through a bayesian updating scheme on the distribution of the taste parameters β the resulting likelihood function of the choice history is the same if one notes that conditional on β, the choices are independent since we assumed that ϵ_{itn}'s are independent across t. It must be noted that the exact choice sequence does not affect the likelihood function, and all that matters is the product of the choice probabilities conditional on β. So by directly allowing for interdependencies among the choices across time for the same individual through the random taste parameters one is actually capturing a dynamic choice model with a bayesian updating scheme on the distribution of taste parameters.

Appendix H

Approximations for Central Moments of Function of Random Variables

In this appendix, we outline a standard method to approximate the first two central moments of a function of random variables. Consider a multivariate random vector, X, with finite first and second moments, and an analytic function $f(\cdot)$ which generates the random variable Y such that $Y = f(X)$. Let $\mathrm{E}(X) = \mu$ and $\mathrm{E}((X-\mu)(X-\mu)') = \Sigma$. By a Taylor's series expansion of the $f(\cdot)$ about μ, we can approximate Y as:

$$Y = f(X) \approx f(\mu) + \nabla f(\mu)'(X-\mu) + \frac{1}{2}(X-\mu)'\nabla^2 f(\mu)(X-\mu) \tag{H.1}$$

where $\nabla f(\cdot)$ and $\nabla^2 f(\cdot)$ denote the first and second derivatives of $f(\cdot)$, respectively. Since $\mathrm{E}(\nabla f(\mu)'(X-\mu)) = 0$, we have

$$\mathrm{E}(Y) \approx f(\mu) + \frac{1}{2}\mathrm{E}((X-\mu)'\nabla^2 f(\mu)(X-\mu)) \tag{H.2}$$

Thus, E(Y) is a function of the mean vector *and* the variance-covariance matrix of the generating random variables. Further, by a first order Taylor's series expansion

$$\begin{aligned} \mathrm{Var}(Y) &\approx \mathrm{Var}(\nabla f(\mu)'(X-\mu)) \\ &= \nabla f(\mu)'\Sigma\nabla f(\mu) \end{aligned} \tag{H.3}$$

Appendix I

Methods for the Extraction of Latent Variables

In this appendix, we briefly review the different methods for extracting the latent variables after the latent variable model system has been estimated. We also elaborate on constructing or estimating a model, referred to as the *factor score model*, which maps from the indicators to the latent variables in a MIMIC model. Applying similar tools for the more general LISREL model will be transparent.

In the MIMIC model, the *linear* structural model is written as:

$$\eta = B\eta + \Gamma x + \zeta \tag{I.1}$$

In equation (I.1), η is the $m \times 1$ vector of latent endogenous variables; x is the $q \times 1$ vector of latent exogenous variables; B is the $m \times m$ coefficient matrix capturing the influence of the latent variables on each other; and Γ is the $m \times q$ parameter matrix reflecting the effects of x on η. The matrix $(I - B)$ is assumed to non-singular. ζ is the disturbance vector that is assumed to have an expected value of zero $[E(\zeta) = 0]$ and which is uncorrelated with x.

The *linear* measurement model is written as:

$$y = \Lambda\eta + \epsilon \tag{I.2}$$

The $p \times 1$ vector of observed variables, y, form the indicators of η; Λ is a $p \times m$ parameter matrix capturing the mapping from η to y; and ϵ is a $p \times 1$ random vector representing the errors of measurement for y. ϵ is assumed to be uncorrelated with η and ζ. The expected value of ϵ is zero. To simplify matters y and x are written as deviations from their respective means (without any loss of generality).

Let the unknown parameters be stacked in a vector θ. S_{yy} denotes the observed covariance matrix of y, S_{yx} denotes the observed covariance between y and x, and S_{xx} denotes the covariance matrix of x. Then the covariance matrix of the observed $[y', x']'$ is given by

$$S = \begin{bmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{bmatrix} \tag{I.3}$$

Let $\Sigma(\theta)$ denote the covariance matrix of the vector $[y', x']'$ implied by the model system, i.e., as a function of the unknown parameter vector θ, where

$$\Sigma(\theta) = \begin{bmatrix} \Sigma_{yy}(\theta) & \Sigma_{yx}(\theta) \\ \Sigma_{xy}(\theta) & \Sigma_{xx}(\theta) \end{bmatrix} \tag{I.4}$$

It must be noted that $\Sigma_{xx}(\theta)$ is not a function of the unknown parameters, since $\Sigma_{xx}(\theta) \equiv E(xx')$. Note that $\Sigma_{yy}(\theta)$ is written as

$$\Sigma_{yy}(\theta) = E(yy') \tag{I.5}$$

Substituting equation (I.2) in equation (I.5) we obtain

$$\begin{aligned} \Sigma_{yy}(\theta) &= E[(\Lambda\eta + \epsilon)(\eta'\Lambda' + \epsilon')] \\ &= \Lambda E(\eta\eta')\Lambda' + \Theta \end{aligned} \tag{I.6}$$

where $\Theta = E(\epsilon\epsilon')$. $E(\eta\eta')$ can be further broken down by substituting the reduced form of equation (I.1) – $\eta = (I - B)^{-1}(\Gamma x + \zeta)$ – for η in equation (I.6) and by simplifying to obtain

$$\Sigma_{yy}(\theta) = \Lambda(I-B)^{-1}(\Gamma E(xx')\Gamma' + \Psi)[(I-B)^{-1}]'\Lambda' + \Theta \tag{I.7}$$

where $\Psi = E(\zeta\zeta')$. Thus, the covariance matrix of y is a complex function of 5 model parameter matrices.

Similarly, $\Sigma_{yx}(\theta)$ is given by

$$\Sigma_{yx}(\theta) = E(yx') \tag{I.8}$$

Substituting equations (I.2) in equation (I.8) gives

$$\Sigma_{yx}(\theta) = E[(\Lambda\eta + \epsilon)x'] \tag{I.9}$$

Again making use of the reduced form of η leads to

$$\Sigma_{yx}(\theta) = \Lambda(I-B)^{-1}\Gamma E(xx') \tag{I.10}$$

Assembling equations (I.7) and (I.10) in equation (I.4) gives

$$\Sigma(\theta) = \begin{bmatrix} \Lambda(I-B)^{-1}(\Gamma E(xx')\Gamma' + \Psi)[(I-B)^{-1}]'\Lambda' + \Theta & \Lambda(I-B)^{-1}\Gamma E(xx') \\ E(xx')\Gamma'[(I-B)^{-1}]'\Lambda' & E(xx') \end{bmatrix} \tag{I.11}$$

The covariance matrix of η is obtained from equation (I.1) as:

$$\Sigma_{\eta\eta} = (I-B)^{-1}(\Gamma E(xx')\Gamma' + \Psi)[(I-B)^{-1}]' \tag{I.12}$$

Primarily two methods may be adopted to extract the latent variables, in addition to the obvious method of extracting the latent variables using the structural model. If the extraction of the latent variables is based on the information from the indicators y, then the extraction procedure is defined as the Partial Information Extraction Method. If the extraction is based on information from both the explanatory variables x and the indicators, then the extraction method is defined as the Full Information

Extraction Method.

In the popular extraction methods, only the information from the measurement model is used in the extraction procedure, while the information from the structural model is essentially ignored. To this end, we present an extraction method which utilizes information contained in the structural and measurement models. Furthermore, we compare the partial information and the full information extraction methods for the single latent variable case.

I.1 Partial Information Extraction Methods

These methods are the usual extraction methods adopted in a Factor Analytic model. Based on the assumptions made in the extraction procedure, the methods are further classified into:

1. Weighted Least Squares Method; and

2. Regression Method.

I.1.1 Weighted Least Squares Method

Assume that the parameters in the MIMIC model have been obtained through some estimation procedure (see, for example, Bollen [1989]). The measurement model is given by equation (I.2) i.e.,

$$y = \Lambda\eta + \epsilon \tag{I.13}$$

Regard the specific factors $\epsilon = (\epsilon_1, \ldots, \epsilon_p)'$ as errors. Since $\text{var}(\epsilon_i) = \theta_{ii}$, $\forall i = 1, 2, \ldots, p$, need not be equal, Bartlett [1937] has suggested that weighted least squares be used to estimate the latent variables. The sum of the squared errors, weighted by the reciprocal of their variances, is given by

$$\mathcal{A} = \sum_{i=1}^{p} \frac{\epsilon_i^2}{\theta_{ii}} = (y - \Lambda\eta)'\Theta^{-1}(y - \Lambda\eta) \tag{I.14}$$

Then the coefficients of the factor score model are obtained by minimizing the function given in equation (I.14) with respect to η. The first order conditions lead to:

$$\begin{aligned}\eta &= (\Lambda'\Theta^{-1}\Lambda)^{-1}\Lambda'\Theta^{-1}y \\ &= \kappa y \end{aligned} \tag{I.15}$$

where κ is the $m \times p$ parameter matrix in the factor score model. The second order condition is given by:

$$\frac{\partial^2 \mathcal{A}}{\partial\eta\partial\eta'} = 2\Lambda'\Theta^{-1}\Lambda \tag{I.16}$$

which is positive definite since Θ is a positive definite matrix, implying that the solution for η in equation (I.15) corresponds to the minimum of $\mathcal{A}$. The main disadvantages of this method are: (1) information regarding the correlations between the latent variables is neglected, and (2) extraction is based only on the measurement model. The factor score model is unaffected by the assumptions on the explanatory variables x.

I.1.2 Regression Method

In this method the latent variables are extracted using the measurement model, and the information regarding the correlations between the latent variables. Assume that the joint distribution of the indicators and the latent variables is *multivariate normal*, i.e., $(y', \eta')' \sim N(0, \Sigma)$ where Σ equals

$$\begin{bmatrix} \Sigma_{yy} & \Sigma_{y\eta} \\ \Sigma_{\eta y} & \Sigma_{\eta\eta} \end{bmatrix} \tag{I.17}$$

The covariance matrix Σ_{yy} is given by equation (I.7), while the covariance matrix $\Sigma_{\eta\eta}$ is given by equation (I.12). $\Sigma_{y\eta}$ is obtained from equation (I.2) as

$$\Sigma_{y\eta} = \Lambda\Sigma_{\eta\eta} \tag{I.18}$$

Therefore, considering the conditional distribution of η given the indicators y, the $m \times p$ parameter matrix in the factor score model, denoted by κ, is given by:

$$\kappa = \Sigma_{\eta\eta}\Lambda'(\Lambda\Sigma_{\eta\eta}\Lambda' + \Theta)^{-1} \tag{I.19}$$

If one uses the sample covariance matrix S_{yy} for the indicators instead of the estimated covariance matrix, κ equals

$$\Sigma_{\eta\eta}\Lambda'(S_{yy})^{-1} \tag{I.20}$$

One of the main drawbacks of this method, is the possibility of non-zero coefficients for some indicators in the extraction of a particular latent variable, even though the corresponding indicators are not specified as indicators of that particular latent variable (in the case of multiple latent variables). Such inconsistencies do not arise in the Weighted Least Squares Method if Θ is a diagonal matrix. Even here it must be noted that we do not make any specific assumptions on the distribution of the explanatory variables x.

I.2 Full Information Extraction Method Under Normally Distributed Explanatory Variables

The drawbacks in the methods described in section I.1 motivate the extraction of latent variables using both the measurement model and the structural model, leading to the full information extraction method. The extraction procedure is described separately for the *single* latent variable case and the latent *vector* case, to illustrate the close relationship between the Regression Method described in section I.1.2 and the full information extraction method for the single latent variable MIMIC model.

I.2.1 Single latent variable case

The structural model for the single latent variable MIMIC model is specified as:

$$\eta = \gamma' x + \zeta \tag{I.21}$$

The basic idea in this method is to assume that $(\eta, y', x')'$ is distributed multivariate normal and use the conditional distribution of η given $(y', x')'$ to fit the latent variable as function of y and x. These fitted values are then "regressed" on the indicators y to obtain the $1 \times p$ parameter vector κ of the factor score model. The conditional expectation of η is given by (using sample covariance matrices):

$$E[\eta|y,x] = \begin{bmatrix} \Sigma_{\eta y} & \Sigma_{\eta x} \end{bmatrix} \begin{bmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{bmatrix}^{-1} \begin{bmatrix} y \\ x \end{bmatrix} \tag{I.22}$$

$$= \begin{bmatrix} \hat{\alpha} & \hat{\beta} \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} \tag{I.23}$$

where $\Sigma_{\eta y} = \sigma_\eta^2 \Lambda'$ and $\Sigma_{\eta x} = \gamma' E(xx')$. Therefore, $\hat{\eta}$ can be written as

$$\hat{\eta} = \hat{\alpha} y + \hat{\beta} x. \tag{I.24}$$

On regressing $\hat{\eta}$ on y leads to the factor score model's parameter vector κ given by

$$\kappa = \hat{\alpha} + \hat{\beta} S_{xy} S_{yy}^{-1} \tag{I.25}$$

The standard error of the factor score model is calculated as follows. We know that in a multiple regression model of Y on X, the unbiased estimate of the variance of the error term is given by (see Greene 1990)

$$s^2 = \frac{e'e}{N-K}. \tag{I.26}$$

where e is the residual vector, N is the number of observations and K is the number of parameters estimated, and the estimated variance-covariance matrix equals $s^2(X'X)^{-1}$. Similarly, the estimated variance-covariance matrix of $\hat{\kappa}$ is written as

$$\mathrm{Var}(\hat{\kappa}) = s^2(Y'Y)^{-1} \tag{I.27}$$

where Y is the stacked matrix of y'. In this case $e'e = \hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} - \hat{\gamma}'Y'Y\hat{\gamma}$ where $\hat{\boldsymbol{\eta}}$ is the stacked $\hat{\eta}$. Therefore,

$$s^2 = \frac{(N-1)(\sigma^2_\eta - \hat{\gamma}'S_{yy}\hat{\gamma})}{N-P} \tag{I.28}$$

and

$$\mathrm{Var}(\hat{\gamma}) = \frac{s^2}{(N-1)}S^{-1}_{yy} \tag{I.29}$$

Comparison of Full Information Extraction Method and the Regression Method for a single latent variable MIMIC model

The regression method yields the factor score vector κ_R given by equation (I.20). The full information extraction method yields the factor score vector κ_F given by

$$\kappa_F = \hat{\alpha} + \hat{\beta}S_{xy}S^{-1}_{yy} \tag{I.30}$$

Let

$$\begin{bmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{bmatrix}^{-1} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{I.31}$$

where

$$A = S^{-1}_{yy}(I + S_{yx}(S_{xx} - S_{xy}S^{-1}_{yy}S_{yx})^{-1}S_{xy}S^{-1}_{yy}) \tag{I.32}$$

$$B = -S^{-1}_{yy}S_{yx}(S_{xx} - S_{xy}S^{-1}_{yy}S_{yx})^{-1} \tag{I.33}$$

$$C = B' = -(S_{xx} - S_{xy}S^{-1}_{yy}S_{yx})^{-1}S_{xy}S^{-1}_{yy} \tag{I.34}$$

$$D = (S_{xx} - S_{xy}S^{-1}_{yy}S_{yx})^{-1} \tag{I.35}$$

Then,

$$\hat{\alpha} = \Sigma_{\eta y} A + \Sigma_{\eta x} C \tag{I.36}$$

$$\hat{\beta} = \Sigma_{\eta y} B + \Sigma_{\eta x} D \tag{I.37}$$

Substituting equations (I.32), (I.33), (I.34), and (I.35) in equations (I.36) and (I.37), and equations (I.36) and (I.37) in equation (I.30), and simplifying yields $\kappa_F = \kappa_R$. Therefore both the full information and the regression methods yield the same factor score model.

I.2.2 Latent vector case

Here the extraction is similar to the single latent variable case. Assume that the vector $(\eta', y', x')'$ is distributed multivariate normal with variance-covariance matrix Σ, i.e.,

$$\Sigma = \begin{bmatrix} \Sigma_{\eta\eta} & \Sigma_{\eta y} & \Sigma_{\eta x} \\ \Sigma_{y\eta} & \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{x\eta} & \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \tag{I.38}$$

Therefore, using the conditional distribution of η given y and x, the fitted values $\hat{\eta}$ is obtained, using sample variance-covariance matrix for $(y', x')'$, as

$$\hat{\eta} = \begin{bmatrix} \Sigma_{\eta y} & \Sigma_{\eta x} \end{bmatrix} \begin{bmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{bmatrix}^{-1} \begin{bmatrix} y \\ x \end{bmatrix} \tag{I.39}$$

$$= A \begin{bmatrix} y \\ x \end{bmatrix} \tag{I.40}$$

Considering the set of equations describing η as a function of y, as a set of *seemingly unrelated regression equations*, the only information needed for the maximum likelihood estimation of the parameters of the model system are the variance-covariance matrix of the variables concerned, i.e,.

$$\Sigma^* = \begin{bmatrix} \Sigma_{\hat{\eta}\hat{\eta}} & \Sigma_{\hat{\eta}y} \\ \Sigma_{y\hat{\eta}} & S_{yy} \end{bmatrix} \tag{I.41}$$

Using equation (I.39),

$$\Sigma_{\hat{\eta}y} = A \begin{bmatrix} S_{yy} \\ \Sigma_{xy} \end{bmatrix} \tag{I.42}$$

and

$$\Sigma_{\hat{\eta}\hat{\eta}} = A \begin{bmatrix} \Sigma_{y\eta} \\ \Sigma_{x\eta} \end{bmatrix} \tag{I.43}$$

Substitute equations (I.42) and (I.43) in equation (I.41), and run the seemingly unrelated regression model *taking care to specify the free and fixed parameters depending on the relevant indicators for each of the latent variables as specified in the measurement model.* Therefore, in this method one controls the free and fixed parameters in the model, thereby preventing undesirable non-zero parameter values.

I.3 Full Information Extraction Method Conditional on Explanatory Variables

In the full information extraction method discussed earlier we assumed that x is normally distributed. But this may be an unrealistic assumption as some of the explanatory variables may be dummy or categorical variables. Further, even if all the explanatory variables are continuous and normally distributed in the population, it may be not be normally distributed in the sample. Herein, we present the extraction method under such circumstances.

Given the structural and measurement models are specified as in equations (I.1) and (I.2), we may assume the distribution of $(\eta', y')'$ given x to be normally distributed, i.e., $(\eta', y')'|x \sim \mathcal{N}(M, \Sigma)$, where

$$M = \begin{pmatrix} (I-B)^{-1}\Gamma x \\ \Lambda(I-B)^{-1}\Gamma x \end{pmatrix} \tag{I.44}$$

and

$$\Sigma = \begin{pmatrix} (I-B)^{-1}\Psi[(I-B)^{-1}]' & (I-B)^{-1}\Psi[(I-B)^{-1}]'\Lambda' \\ \Lambda(I-B)^{-1}\Psi[(I-B)^{-1}]' & \Lambda(I-B)^{-1}\Psi[(I-B)^{-1}]'\Lambda' + \Theta \end{pmatrix} \tag{I.45}$$

The conditional distribution of η given y and x is also normally distributed with $\eta|(y, x) \sim \mathcal{N}(M_1, \Sigma_1)$, where

$$\begin{aligned} M_1 \;=\; & (I-B)^{-1}\Gamma x + (I-B)^{-1}\Psi[(I-B)^{-1}]'\Lambda'(\Lambda(I-B)^{-1}\Psi[(I-B)^{-1}]'\Lambda' + \Theta)^{-1} \\ & (y - \Lambda(I-B)^{-1}\Gamma x) \end{aligned} \tag{I.46}$$

and

$$\begin{aligned} \Sigma_1 \;=\; & (I-B)^{-1}\Psi[(I-B)^{-1}]' - (I-B)^{-1}\Psi[(I-B)^{-1}]'\Lambda' \\ & (\Lambda(I-B)^{-1}\Psi[(I-B)^{-1}]'\Lambda' + \Theta)^{-1}\Lambda(I-B)^{-1}\Psi[(I-B)^{-1}]' \\ =\; & (I-B)^{-1}\Psi[(I-B)^{-1}]'\Big(I - \Lambda'(\Lambda(I-B)^{-1}\Psi[(I-B)^{-1}]'\Lambda' + \Theta)^{-1} \\ & \Lambda(I-B)^{-1}\Psi[(I-B)^{-1}]'\Big) \end{aligned} \tag{I.47}$$

Appendix J

Methods for the Extraction of Latent Class Probabilities

J.1 Extraction of Latent Class Probabilities for Latent Class Model

In latent variable models, methods exist for extracting the latent variables from the structural model, measurement model, and from both the structural and measurement models. The primary objective of the extraction methods is to *predict* the latent variables when the inputs to the model system change. For example, assume that the analyst has estimated the latent variable model system. The analyst may conduct future measurements of the explanatory variables, indicators or both, either from the same population for which the model system was estimated or from a different population wherein the analyst expects the estimated latent variable model system (and hence the psychological "laws") to hold. The analyst can predict the latent variables from:

- Structural Model: If data is available only on the explanatory variables the analyst can predict the latent variables using the estimated structural model. This is the easiest and the most intuitive approach.

- Measurement Model: If data is available only on the indicators, the analyst can predict the latent variables using a model, referred to as the factor score model, which maps from the indicators to the latent variables, and which is constructed primarily from the measurement model using the "weighted least squares method" or "regression method" (see Johnson and Wichern [1982]).

- Structural and Measurement Models: Herein information from both the estimated structural and measurement models are utilized in the construction of the factor score model (see Gopinath [1992]).

In similar vein, in the latent class model, one can either use the class membership model, the measurement model, or both to estimate the latent class probabilities depending on the availability of Z_n or A_n. For notational simplicity, let $\mathrm{P}(s|Z_n) \equiv Q_s(Z_n;\theta)$, and $g(A_n|s) \equiv g(A_n|l^*_{sn} = 1; \phi_s)$. If one observes only Z_n, then $\mathrm{P}(s|Z_n)$ is trivially obtained from the class membership model.

J.1.1 Extraction from the Measurement Model

Suppose one observes or measures only A_n and we need to obtain $\mathrm{P}(s|A_n)$.

Deterministic Assignment Approach:

Assuming that we have estimated the parameters of the latent class model and these estimated parameters as "fixed" parameters[1], the latent class assignment can be conducted by maximizing the likelihood of observation of A_n, i.e.,

$$\begin{aligned} \max_{\eta_s \forall s} \quad & \sum_{s=1}^{S} \eta_s g(A_n|s) \\ \text{subject to} \quad & \sum_{s=1}^{S} \eta_s = 1 \\ & \eta_s \geq 0 \qquad \forall s \end{aligned} \tag{J.1}$$

where η_s's are the probabilities of interest. This formulation leads to the "determin-

[1]In reality, the estimated parameters are statistics and hence are random variables.

istic" assignment of the individual to the latent class s^* such that

$$s^* = \underset{s}{\operatorname{argmax}}\{g(A_n|s)\} \tag{J.2}$$

The above assignment is obvious, since the optimal solution to a linear program is at an extreme point of the polyhedron $\{\sum_s \eta_s = 1; \eta_s \geq 0\ \forall s\}$.

Probabilistic Assignment Approach

Using a Bayesian approach,

$$\mathrm{P}(s|A_n) = \frac{g(A_n|s)\ \mathrm{P}(s)}{\sum\limits_{s'=1}^{S} g(A_n|s')\ \mathrm{P}(s')} \tag{J.3}$$

The unconditional probability mass function, $\mathrm{P}(s)$, may be calculated as:

$$\mathrm{P}(s) = \int\limits_Z \mathrm{P}(s|z) f_Z(z)\, dz \tag{J.4}$$

where $f_Z(z)$ denotes the probability density of the explanatory variables in the population. Thus, $\mathrm{P}(s|A_n)$ can be (numerically) obtained as:

$$\mathrm{P}(s|A_n) = \frac{g(A_n|s) \int\limits_Z \mathrm{P}(s|z) f_Z(z)\, dz}{\sum\limits_{s'=1}^{S} g(A_n|s') \int\limits_Z \mathrm{P}(s'|z) f_Z(z)\, dz} \tag{J.5}$$

J.1.2 Extraction from Class Membership and Measurement Models

Here we address the question of how to estimate the latent class probabilities given information on Z_n and A_n. One might be tempted to trivially adopt the class membership model and obtain $\mathrm{P}(s|Z_n)$. The main drawback of this approach is that one does not use the information from A_n. Using a Bayesian approach,

$$\mathrm{P}(s|Z_n, A_n) \quad = \quad \frac{\mathrm{P}(s, A_n|Z_n)}{\mathrm{P}(A_n|Z_n)}$$

$$= \frac{g(A_n|s, Z_n)\ \mathrm{P}(s|Z_n)}{\sum_{s'=1}^{S} g(A_n|s', Z_n)\ \mathrm{P}(s'|Z_n)}$$

$$= \frac{g(A_n|s)\ \mathrm{P}(s|Z_n)}{\sum_{s'=1}^{S} g(A_n|s')\ \mathrm{P}(s'|Z_n)} \quad \text{(J.6)}$$

Thus, in general $\mathrm{P}(s|Z_n, A_n) \neq \mathrm{P}(s|Z_n)$.

J.2 Extraction of Latent Class Probabilities for Latent Structure Choice Model

As in the latent class model, one can use either the structural model, or the structural and measurement model, to estimate the latent class probabilities depending on the availability of W_n, Z_n, $A_{Z;n}$.

J.2.1 Extraction using Structural Model

If only W_n and Z_n are measured/observed, then $\mathrm{P}(l^*_{sn} = 1|W_n, Z_n)$ is trivially obtained from the estimated structural equations for Z^* and l^*_{sn}, i.e.,

$$\hat{Z}^*_n = (I - \hat{B}_Z)^{-1}\hat{\Gamma}_Z W_n \quad \text{(J.7)}$$

and

$$\mathrm{P}(l^*_{sn} = 1|W_n, Z_n) = Q_s(Z_n, \hat{Z}^*_n; \hat{\theta}) \quad \text{(J.8)}$$

Note that some degree of inconsistency is introduced due to ignoring the sampling error in $\hat{Z}^*_n$. More precisely,

$$\mathrm{P}(l^*_{sn} = 1|W_n, Z_n) = \int_{\hat{Z}^*_n} Q_s(Z_n, \omega; \hat{\theta})\, \mathrm{d}F(\omega) \quad \text{(J.9)}$$

where $F(\cdot)$ is the sampling distribution of $\hat{Z}_n^*$. By drawing R draws from the sampling distribution F, an estimate of $\mathrm{P}(l_{sn}^* = 1|W_n, Z_n)$ is obtained as:

$$\frac{1}{R}\sum_{r=1}^{R}\left\{Q_s(Z_n, \omega^{(r)}; \hat{\theta})\right\} \tag{J.10}$$

J.2.2 Extraction using Structural and Measurement Models

Suppose one observes the explanatory variables W_n and Z_n, and the indicators $A_{Z;n}$ and $A_{S;n}$, then one might be tempted to trivially adopt the structural equations. The main drawback in this approach is that one does not use the information from $A_{Z;n}$ and $A_{S;n}$. Using a Bayesian approach,

$$\mathrm{P}(l_{sn}^*|W_n, Z_n, A_{Z;n}, A_{S;n}) = \frac{f(A_{Z;n}, A_{S;n}|l_{sn}^*, W_n, Z_n)\ P(l_{sn}^*|W_n, Z_n)}{\sum\limits_{s'=1}^{S} f(A_{Z;n}, A_{S;n}|l_{s'n}^*, W_n, Z_n)\ P(l_{s'n}^*|W_n, Z_n)} \tag{J.11}$$

The reader can easily verify that $f(A_{Z;n}, A_{S;n}|l_{sn}^*, W_n, Z_n)$ is written as:

$$\frac{\int\limits_{Z^*} g(A_{Z;n}|Z^*)\ \mathrm{P}(A_{S;n}|l_{sn}^*)\ \mathrm{P}(l_{sn}^*|Z_n, Z^*)\ f(Z^*|W_n)\, dZ^*}{P(l_{sn}^*|W_n, Z_n)} \tag{J.12}$$

Thus, the required probability, $\mathrm{P}(l_{sn}^*|W_n, Z_n, A_{Z;n}, A_{S;n})$ reduces to:

$$\frac{\mathrm{P}(A_{S;n}|l_{sn}^*) \int\limits_{Z^*} g(A_{Z;n}|Z^*)\ \mathrm{P}(l_{sn}^*|Z_n, Z^*)\ f(Z^*|W_n)\, dZ^*}{\sum\limits_{s'=1}^{S} \mathrm{P}(A_{S;n}|l_{s'n}^*) \int\limits_{Z^*} g(A_{Z;n}|Z^*)\ \mathrm{P}(l_{s'n}^*|Z_n, Z^*)\ f(Z^*|W_n)\, dZ^*} \tag{J.13}$$

In a similar vein, if one observes the explanatory variables (X_n, Z_n, W_n), the choice indicator Y_n, and the indicators of attitudes, perceptions and latent class, i.e., $A_{Z;n}$, $A_{X;n}$, and $A_{S;n}$, the latent class probability, $\mathrm{P}(l_{sn}^*|X_n, Z_n, W_n, Y_n, A_{X;n}, A_{Z;n}, A_{S;n})$, equals

$$\frac{\mathrm{P}(A_{S;n}|l^*_{sn}) \int\limits_{X^*}\int\limits_{Z^*} \mathrm{P}(Y_n|l^*_{sn}, X^*)\mathrm{P}(l^*_{sn}|Z^*)g(A_X|X^*)g(A_Z|Z^*)f(X^*, Z^*)\, dX^*\, dZ^*}{\sum\limits_{s'=1}^{S} \left\{ \mathrm{P}(A_{S;n}|l^*_{s'n}) \int\limits_{X^*}\int\limits_{Z^*} \mathrm{P}(Y_n|l^*_{s'n}, X^*)\mathrm{P}(l^*_{s'n}|Z^*)g(A_X|X^*)g(A_Z|Z^*)f(X^*, Z^*)\, dX^*\, dZ^* \right\}} \tag{J.14}$$

Bibliography

[1] Abelson, R. P., and Levy, A. Decision making and decision theory. In *Handbook of Social Psychology, Vol. 1*, G. Lindzey and E. Aronson, Eds. Random House, New York, 1985.

[2] Aitkin, M., Anderson, D., and Hinde, J. Statistical Modeling of Data on Teaching Style. *Journal of the Royal Statistical Society B 47* (1981), 67–75.

[3] Akaike, H. A New Look at Statistical Model Identification. *IEEE Transactions on Automatic Control 6* (1974), 716–723.

[4] Allen, W. B. The Demand for Freight Transportation: A Micro approach. *Transportation Research 11B* (1977), 9–14.

[5] Amemiya, T. The Estimation of a Simultaneous Equation Generalized Probit Model. *Econometrica 46* (1978), 1193–1205.

[6] Amemiya, T. *Advanced Econometrics.* Harvard University Press, Cambridge, Mass, 1985.

[7] Anderson, J. A., and Pemberton, J. D. The Grouped Continuous Model for Multivariate Ordered Categorical Variables and Covariate Adjustment. *Biometrics 41* (1985), 875–885.

[8] Anderson, J. J. Normal Mixtures and the Number of Clusters Problem. *Computational Statistics Quarterly 2* (1985), 3–14.

[9] Anderson, N. H. Algebraic Models in Perception. In *Handbook of Perception*, E. C. Carterette and M. P. Friedman, Eds. Academic Press, New York, 1974a.

[10] Anderson, N. H. Cognitive algebra. In *Advances in Experimental Social Psychology*, L. Berkowitz, Ed. Academic Press, New York, 1974b.

[11] Anderson, S. P., de Palma, A., and Thisse, J. F. *Discrete Choice Theory of Product Differentiation.* The MIT Press, Cambridge, Mass, 1992.

[12] Aptech Systems. GAUSS: User's Guide, 1993.

[13] Ashford, J. R., and Sowden, R. R. Multivariate Probit Analysis. *Biometrics 41* (1970), 875–885.

[14] Bartholomew, D. J. *Latent Variable Models and Factor Analysis.* Oxford University Press, 1987.

[15] Bartlett, M. S. The statistical conception of mental factors. *British Journal of Psychology 28* (1937), 97–104.

[16] Bates, J. J. Econometric Issues in Stated Preference Analysis. *Journal of Transport Economics and Policy 22*, 1 (1988), 59–69.

[17] Bates, J. J., and Roberts, M. Value of Time Research: Summary of Results and Recommendations. In *Proc. of the PTRC Annual Summer Meeting, University of Sussex* (1987).

[18] Bates, J. J., and Terzis, G. Surveys Involving Adaptive Stated Preference Techniques. In *Survey and Statistical Computing*, A. Westlake, Ed. Elsevier Science Pubslishers, 1992.

[19] Beach, L. R., and Mitchell, T. R. A Contingency Model for the Selection of Decision Strategies. *Acad. Manage. Review 3* (1978), 439–449.

[20] Ben-Akiva, M., and Boccara, B. *Integrated Framework for Travel Behavior Analysis.* World Conference in Transporation Research, Aix-en-Provence, France, 1987.

[21] Ben-Akiva, M., and Boccara, B. *Discrete Choice Models with Latent Choice Sets*. Banff Invitational Symposium on Consumer Decision Making and Choice Behavior, May 1990.

[22] Ben-Akiva, M., and Bolduc, D. Multinomial Probit with Taste Variations and Generalized Auto-regressive Errors, Working Paper, 1991.

[23] Ben-Akiva, M., and Gopinath, D. Travel Demand Model System for the Information Era. Final Report submitted to Federal Highways Administration, March 1993.

[24] Ben-Akiva, M., Morikawa, T., and Shiroishi, F. Analysis of the Reliability of Preference Ranking Data. *Journal of Business Research 23* (1991), 253–268.

[25] Ben-Akiva, M. E. Choice Models with Simple Choice Set Generation Processes. Tech. rep., Department of Civil Engineering, 1977.

[26] Ben-Akiva, M. E., Bolduc, D., and Bradley, M. Estimation of Travel Choice Models with Randomly Distributed Values of Time. *Transportation Research Record 1413* (1993).

[27] Ben-Akiva, M. E., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H., and Rao, V. Combining Revealed and Stated Preference Data. *Marketing Letters 5*, 4 (1994), 335–350.

[28] Ben-Akiva, M. E., and Moriwaka, T. Estimation of Travel Demand Models from Multiple Data Sources. In *Proceedings of the Eleventh International Symposium on Transportation and Traffic Theory, Yokohama, Japan* (July 1990a).

[29] Bentler, P. M. Multivariate Analysis with Latent Variables. *Annual Review of Psychology 31* (1980), 419–456.

[30] Bernardino, A., Ben-Akiva, M. E., and Salomon, I. A Stated Preference Approach to Modeling the Adoption of Telecommuting. *Transportation Research Record 1413* (1993), 22–30.

[31] Bettman, J. R., and Kakkar, P. Effects of Information Presentation Format on Consumer Information Acquisition Strategies. *Journal of Consumer Research 3* (1977), 233–240.

[32] Blalock, H. M. Making Causal Inferences for Unmeasured Variables from Correlations among Indicators. *American Journal of Sociology 74* (1968), 119–137.

[33] Bolduc, D. Estimation of MNP models using Maximum Simulated Likelihood with Analytical Derivatives and GHK Choice Probability Simulators, Working Paper, 1993.

[34] Bollen, K. A. *Structural Equations with Latent Variables.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1989.

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

[36] Box, G. E., and Cox, D. R. An Analysis of Transformations. *Journal of Royal Statistical Society B26* (211-243), 1964.

[37] Boyles, R. A. On the Convergence of the EM Algorithm. *Journal of the Royal Statistical Society, Ser. B 45*, 1 (1983), 47–50.

[38] Bozdogan, H. Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions. *Psychometrica 52* (1987), 345–370.

[39] Bozdogan, H., and Sclove, S. L. Multi-sample Cluster Analysis using Akaike's Information Criterion. *Annals of the Institute of Statistics and Mathematics 36* (1984), 163–180.

[40] Bradley, M. Realism and Adaptation in Designing Hypothetical Travel Choice Concepts. *Journal of Transport Economics and Policy 22*, 1 (1988), 121–137.

[41] Bradley, M. A., and Daly, A. J. *New Analysis Issues in Stated Preference Research.* PTRC Summer Annual Conference, 1993.

[42] Bradley, M. A., and Gunn, H. A Stated Preference Analysis of Values of Time in the Netherlands. *Transportation Research Record*, 1285 (1989), 78–88.

[43] Bradley, M. A., and Gunn, H. *Further Applications and Validation of The Netherlands Value of Time Study.* Sixth International Conference on Travel Behavior, Quebec, Canada, 1991.

[44] Bruner, J. S. Social Psychology and Perception. In *Readings in Social Psychology*, E. F. Maccoby, T. M. Newcomb, and E. L. Hartley, Eds. Holt, Rinehart and Winston, New York, 1958.

[45] Bruner, J. S., Goodnow, J. J., and Austin, G. A. *A Study of Thinking.* Science Editions, New York, 1962.

[46] Bruzelius, L. H. *The Value of Travel Time.* Croom Helm, London, 1979.

[47] Cambridge Systematics. Customer Preference and Behavior. Prepared for the Electric Power Research Institute. Tech. rep., Cambridge Systematics, 1986.

[48] Cattin, P., and Wittink, D. R. Commercial Use of Conjoint Analysis: A Survey. *Journal of Marketing 46* (1982), 44–53.

[49] Chaiken, S., and Stangor, C. Attitudes and Attitude Change. *Annual Review of Psychology 38* (1987), 575–630.

[50] Chamberlain, G. Analysis of Covariance with Qualitative Data. *Review of Economic Studies 47* (1980), 225–238.

[51] Cherlow, J. R. Measuring Values of Travel Time Savings. *Journal of Consumer Research 7* (1981), 360–371.

[52] Chiang, Y., Roberts, P. O., and Ben-Akiva, M. Development of a Policy Sensitive Model for Forecasting Demand. Tech. rep., Center for Transportation Studies, Massachusetts Institute of Technology, 1981.

[53] Chintagunta, P. K., Jain, D. C., and Vilcassim, N. J. Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data. *Journal of Marketing Research 28* (1991), 417–428.

[54] Claffey, P. J. Characteristics of Passenger Car Travel on Toll Roads and Comparable Free Roads. In *Studies in Highway Engineering Economy, Bulletin 306*. Highway Research Board, National Academy of Sciences, Washington, D.C, 1961, pp. 1–22.

[55] Clogg, C. C., and Goodman, L. A. Latent Structure Analysis of a Set of Multidimensional Contingency Tables. *Journal of the American Statistical Association 79* (1984), 762–771.

[56] Clogg, C. C., and Goodman, L. A. On Scaling Models Applied to Data from Several Groups. *Psychometrika 51* (1986), 123–135.

[57] Clogg, C. C., and Goodman, L. A. Simultaneous Latent Structure Analysis in Several Groups. In *Sociological Methodology*, N. B. Tuma, Ed. Jossey-Bass, San Francisco, 1986.

[58] Cooper, J., and Croyle, R. T. Attitudes and Attitude Change. *Annual Review of Psychology 35* (1984), 395–426.

[59] Daughety, A. F., and Inaba, K. Freight Transport Demand Revisited: A Microeconomic View of Multimodal, Multicharacteristics Service Uncertainty and the Demand for Freight Transport. *Transportation Research 13* (1979), 281–288.

[60] Day, N. E. Estimating the Components of a Mixture of Normal Distributions. *Biometrica 56* (1969), 463–474.

[61] Dayton, M., and Macready, G. B. A Scaling Model with Response Errors and Intrinsically Unscalable Respondents. *Psychometrika 45* (1980), 343–356.

[62] Dayton, M., and Macready, G. B. Concomitant Variable Latent-Class Models. *Journal of the American Statistical Association 83* (1988), 173–178.

[63] Dempster, A. P., Laird, N. M., and Rubin, D. B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Ser. B 30* (1977), 1–38.

[64] Dillon, W., Kumar, A., and de Borrero, M. S. Capturing Individual Differences in Paired Comparisons: An Extended BTL Model Incorporating Descriptor Variables. *Journal of Marketing Research 30* (1993), 42–51.

[65] Doyle, P., and Sanders, J. Market Segmentation and Positioning in Specialized Industrial Markets. *Journal of Marketing 49* (1985), 24–32.

[66] Duncan, O. D., Haler, A. O., and Portes, A. Peer Influences on Aspirations: A Reinterpretation. *American Journal of Sociology 74* (1968), 119–137.

[67] Edwards, W. The Theory of Decision Making. *Psychological Bulletin 51* (1954), 380–417.

[68] Ehrenberg, A. S. C. *Repeat Buying: Facts, Theory and Applications.* Charles Griffin and Co., London, 1988.

[69] Einhorn, H. J., and Hogarth, R. M. Behavioral Decision Theory: Processes of Judgment and Choice. *Annual Review of Psychology 32* (1981), 53–88.

[70] EPRI. Methodology for Predicting the Demand for New Electricity-Using Goods. Tech. rep., EA-53, Project 488-1, Electric Power Research Institute, Palo Alto, California, 1977.

[71] Everitt, B. S. *An Introduction to Latent Variable Models.* Monographs on Statistics and Applied Probability. Chapman and Hall, 1984.

[72] Fader, P. S. Integrating the Dirichlet-Multinomial and Multinomial Logit Models of Brand Choice. *Marketing Letters (forthcoming)* (1993).

[73] Fader, P. S., and Lattin, J. M. Accounting for Heterogeneity and Nonstationarity in a Cross-section Model of Consumer Purchase Behavior. *Marketing Science 12*, 3 (1993), 304–317.

[74] Fader, P. S., and McAlister, L. An Elimination by Aspects Model of Consumer Response to Promotion Calibrated on UPC Scanner Data. *Journal of Marketing Research 27* (1990), 322–332.

[75] Fischer, G. W., and Nagin, D. Random Versus Fixed Coefficient Quantal Choice Models. In *Structural Analysis of Discrete Data with Econometric Applications*, C. Manski and D. McFadden, Eds. The MIT Press, Cambridge, Mass, 1981.

[76] Fischer, G. W., and Nagin, D. Simultaneous Equations Models with Discrete and Censored Dependent Variables. In *Structural Analysis of Discrete Data with Econometric Applications*, C. Manski and D. McFadden, Eds. The MIT Press, Cambridge, Mass, 1981, pp. 347–363.

[77] Fishbien, M., and Ajzen, I. Attitudes and Opinions. *Annual Review of Psychology 23* (1972), 487–544.

[78] Formann, A. K. Linear Logistic Latent Class Analysis for Polytomous Data. *Journal of the American Statistical Association 87*, 448 (1992), 476–486.

[79] Fowkes, A. S. *Recent Developments in Stated Preference Techniques in Transport Research*. PTRC Summer Annual Conference, 1991.

[80] Fowkes, T., and Wardman, M. The Design of Stated Preference Travel Choice Experiments with Special Reference to Interpersonal Taste Variations. *Journal of Transport Economics and Policy 22*, 1 (1988), 27–44.

[81] FRANK, R., MASSY, W., AND WIND, Y. *Market Segmentation.* Englewood Cliffs, NJ: Prentice-Hall, 1972.

[82] FRANK, R. E., AND GREEN, P. E. Numerical Taxonomy in Marketing Analysis: A Review Article. *Journal of Marketing Research 5* (1968), 83–94.

[83] GAUDRY, M., AND WILLS, M. Testing the Dogit Model with Aggregate Time-Series and Cross-Sectional Travel Data. *Transportation Research 13B* (1979), 155–166.

[84] GENSH, D. H. A Two-stage Disaggregate Attribute Choice Model. *Marketing Science 6*, 3 (1987), 223–239.

[85] GEWEKE, J. Efficient Simulation from the Multivariate Normal Distribution Subject to Linear Inequality Constraints and the Evaluation of Constraint Probabilities, Mimeo. Duke University, Durham, NC, 1989.

[86] GEWEKE, J., KEANE, M., AND RUNKLE, D. Alternative Computational Approaches to Inference in the Multinomial Probit Model. working paper, October 1992.

[87] GHOSH, J. K., AND SEN, P. K. On the Asymptotic Performance of the Log-likelihood Ratio Statistic for the Mixture Model and Related Results. In *Proc. Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Volume II* (Montery, 1985), L. M. le Cam and R. A. Olshen, Eds., Wadsworth, pp. 789–806.

[88] GÖNÜL, F., AND SRINIVASAN, K. Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues. *Marketing Science 12*, 3 (1993), 213–229.

[89] GOPINATH, D. Modeling Infrastructure Performance and User costs. Master's thesis, Department of Civil Engineering, Massachusetts Institute of Technology, January 1992.

[90] Gopinath, D., and Ben-Akiva, M. Estimation of Discrete Choice models from Ratings-based Conjoint Data. working paper, February 1993.

[91] Green, P. E., and Rao, V. R. Conjoint Measurement for Quantifying Judgmental Data. *Journal of Marketing Research 8* (1971), 355–363.

[92] Green, P. E., and Srinivasan, V. Conjoint Analysis in Consumer Research: Issues and Outlook. *Journal of Consumer Research 5* (1978), 103–123.

[93] Green, P. E., and Wind, Y. New Way to Measure Consumers' Judgments. *Harvard Business Review* (July 1975), 107–117.

[94] Greene, W. H. *Econometric Analysis*. Macmillan Publishing Company, New York, 1990.

[95] Grover, R., and Srinivasan, V. A Simultaneous Approach to Market Segmentation and Market Structuring. *Journal of Marketing Research 24* (1987), 139–153.

[96] Guadagni, P. M., and Little, J. D. C. A Logit Model of Brand Choice Calibrated on Scanner Data. *Marketing Science 2* (1983), 1–20.

[97] Haberman, S. J. Discussion of McCullagh's Lecture. *Journal of the Royal Statistical Society, Ser. B 42* (1980), 132–137.

[98] Hague Consulting Group. *The Netherlands 'Value of Time' Study*: Final Report to Dienst Verkeerskunde, Rijkswarterstaat, Netherlands.

[99] Hague Consulting Group. ALOGIT: User's Guide, 1992.

[100] Hajivassiliou, V. A., and McFadden, D. The Method of Simulated Scores for the Estimation of LDV models, working paper, December 1992.

[101] Haney, D. G. The Value of Time for Passenger Cars: A Theoretical Analysis and Description of Preliminary Experiments. Vol. I of a final report to the U.S. Bureau of Public Roads, Stanford Research Institute, Menlo Park, California, May 1967.

[102] Hardy, G. H., Littlewood, J. E., and Polya, G. *Inequalities*. Cambridge University Press, Cambridge, UK, 1952.

[103] Hasselblad, V. Estimation of Parameters for a Mixture of Normal Distributions. *Technometrics 8* (1966), 431–444.

[104] Hasselblad, V. Estimation of Finite Mixtures of Distributions from the Exponential Family. *Journal of the American Statistical Association 64* (1969), 1459–1471.

[105] Hauser, H. R., and Wernerfelt, B. An Evaluation Cost Model of Consideration Sets. *Journal of Consumer Research 16* (1990), 393–408.

[106] Hausman, J. Specification Tests in Econometrics. *Econometrica 46* (1978), 1251–1271.

[107] Hausman, J., and Wise, D. A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences. *Econometrica 46* (1978), 1219–1240.

[108] Heckman, J. J. The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process. In *Structural Analysis of Discrete Data with Econometric Applications*, C. Manski and D. McFadden, Eds. The MIT Press, Cambridge, Mass, 1981.

[109] Heckman, J. J., and Singer, B. A Method of Minimizing the Impact of Distributional Assumptions in Econometric Models of Duration Data. *Econometrica 52*, 2 (1984), 271–320.

[110] Heckman, J. J., and Willis, R. J. A Beta-logistic Model for the Analysis of Sequential Labor Force Participation by Married Woman. *Journal of Political Economy 85*, 1 (1977), 27–58.

[111] Hensher, D. A. Behavioral and Resource Values of Travel Time Savings: a Bicentenniel Update. *Australian Road Research 19*, 3 (1989), 223–229.

[112] Hildreth, C., and Houck, J. Some Estimators for a Linear Model with Random Coefficients. *Journal of the American Stastitical Association 63* (1969), 584–595.

[113] Hogarth, R. M. Process Tracing in Clinical Judgment. *Behavioral Science 19* (1974), 298–313.

[114] Hsiao, C. Some Estimation Methods for a Random Coefficient Model. *Econometrica 43* (1975), 305–325.

[115] Hsiao, C. *Analysis of Panel Data.* Cambridge University Press, Cambridge, UK, 1986.

[116] Johnson, R. A., and Wichern, D. W. *Applied Multivariate Statistical Analysis.* Prentice-Hall, 1982.

[117] Jöreskog, K. G. A General Method for Estimating a Linear Structural Equation System. In *Structural Models in the Social Sciences*, A. S. Goldberger and O. D. Duncan, Eds. Academic Press, New York, 1973.

[118] Just, M. A., and Carpenter, P. A. Eye Fixations and Cognitive Processes. *Cognitive Psychology 8* (1976), 441–480.

[119] Kahneman, D., and Snell, J. Predicting Utility. In *Insights in Decision Making*, R. M. Hogarth, Ed. University of Chicago Press, Chicago, IL, 1990.

[120] Kahneman, D., and Tversky, A. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica 47* (1979), 263–291.

[121] Kamakura, W. A., and Russell, G. J. A Probabilistic Choice Model for Market Segmentation and Elasticity Structure. *Journal of Marketing Research 25* (1989), 379–390.

[122] Katz, D. A Functional Approach to the Study of Attitudes. *Public Opinion Quarterly 24* (1960), 163–204.

[123] KATZ, D., AND SCOTLAND, E. A Preliminary Statement to Theory of Attitude Structure and Change. In *Psychology: A Study of a Science*, S. Koch, Ed. McGraw-Hill, New York, 1959.

[124] KEESLING, J. W. *Maximum Likelihood Approaches to Causal Analysis.* PhD thesis, Department of Education, University of Chicago, 1972.

[125] KITAMURA, R., AND LAM, T. A Model of Constrained Binary Choice. In *Ninth International Conference on Transportation and Traffic Theory* (VNU Science Press, 1984), pp. 493–512.

[126] KOPPELMAN, F., AND HAUSER, J. Destination Choice Behavior for Non-Grocery-Shopping Trips. *Transportation Research Record 673* (1979), 157–165.

[127] KROES, E. P., AND SHELDON, R. J. The Use of Attitude Models and Stated Preference Models in Practical Transport Analysis. In *Behavioral Research for Transport Policy.* VNU Science Press, Utrecht, The Netherlands, 1986.

[128] KROES, E. P., AND SHELDON, R. J. Stated Preference Methods: An Introduction. *Journal of Transport Economics and Policy 22*, 1 (1988), 11–25.

[129] LA LONDE, B. J., AND COOPER, M. C. *Partnerships in Providing Customer Service: A Third-Party Perspective.* Council of Logistics Management, Oak Brook, IL, 1989.

[130] LAWLEY, D. N., AND MAXWELL, A. E. *Factor Analysis as a Statistical Method.* Butterworth, London, 1971.

[131] LAZARSFELD, P. F., AND HENRY, N. W. *Latent Structure Analysis.* Houghton Mifflin, Boston, 1968.

[132] LEHMANN, D. R. *Market Research and Analysis.* Irwin, Homewood, IL, 1989.

[133] LISCO, T. E. *The Value of Commuter's Travel Time- A Study in Urban Transportation.* PhD thesis, Department of Economics, University of Chicago, June 1967.

[134] Louviere, J., and Sheldon, R. J. Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event. *Journal of Consumer Research 10* (1986), 348–361.

[135] Louviere, J. J. *Analyzing Decision Making: Metric Conjoint Analysis*. Sage University Paper Series on Quantitative Applications in the Social Sciences. Sage Publications, Newbury Park, 1988a.

[136] Louviere, J. J. Conjoint Analysis Modeling of Stated Preference. *Journal of Transport Economics and Policy 22*, 1 (1988b), 93–118.

[137] Luce, R. D., and Tukey, J. W. Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement. *Journal of Mathematical Psychology 1* (1964), 1–27.

[138] Luenberger, D. *Linear and Nonlinear Programming*. Addison-Wesley, Reading, Mass, 1984.

[139] Mahmassani, H., Yen, J., Herman, R., and Sullivan, M. A. Employee Attitudes and Stated Preferences Toward Telecommuting: An Exploratory Analysis. *Transportation Research Record 1413* (1993), 42–48.

[140] Mallar, C. D. The Estimation of Simultaneous Probability Models. *Econometrica 45*, 7 (1977), 1717–1722.

[141] Manski, C. *The Analysis of Qualitative Choice*. PhD thesis, Department of Economics, MIT, Cambridge, Mass, 1973.

[142] Manski, C. The Structure of Random Utility Models. *Theory and Decision 8* (1977), 229–254.

[143] March, J. G. Bounded Rationality, Ambiguity, and the Engineering of Choice. *Bell Journal of Economics 9* (1978), 587–608.

[144] March, J. G., and Simon, H. A. *Organizations*. Wiley, New York, 1958.

[145] McCutheon, A. L. *Latent Class Analysis*. Sage Publications, Newbury Park, 1987.

[146] McFadden, D. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*, P. Zarembka, Ed. Academic Press, New York, 1973.

[147] McFadden, D. The Choice Theory Approach to Marketing Research. *Marketing Science 5*, 4 (1986), 275–297.

[148] McFadden, D., Winston, C., and Börsch-Supan, A. Joint Estimation of Freight Transportation Decisions Under Nonrandom Sampling. In *Analytical Studies in Transport Economics*, A. F. D. Iowa, Ed. Cambridge University Press, New York, 1985, pp. 137–157.

[149] McFadden, D. L., and Leonard, G. K. Issues in the Contingent Valuation of Environmental Goods: Methodologies for Data Collection and Analysis. In *Contingent Valuation: A Critical Assessment*, J. A. Hausman, Ed. North-Holland, Amsterdam, 1993.

[150] McGinnis, M. A. Segmenting Freight Markets. *Transportation Journal 18*, 1 (1978), 59–71.

[151] McKelvey, R., and Zavoina, W. A Statistical Model for the Analysis of Ordinal Level Dependent Variables. *Journal of Mathematical Sociology 4* (1975), 103–120.

[152] McLachlan, G. J. On Bootstrapping the Likelihood Ratio Test Statistic for the Number of Components in a Normal Mixture. *Applied Statistics 36* (1987), 318–324.

[153] McLachlan, G. J., and Basford, K. E. *Mixture Models: Inference and Application to Clustering*. Marcel Dekker, 1988, New York.

[154] Miller, G. A. The Magic Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review 63* (1956), 81–97.

[155] Montgomery, H., and Svenson, O. On Decision Rules and Information Processing Strategies for Choices Among Multiattribute Alternatives. *Scandinavian Journal of Psychology 17* (1976), 283–291.

[156] Moore, W. L., Pessemier, E. A., and Lehmann, D. R. Hierarchical Representations of Market Structures and Choice Processes through Preference Trees. *Journal of Business Research 14* (1986), 107–111.

[157] Morikawa, T. *Incorporating Stated Preference Data in Travel Demand Analysis*. PhD thesis, Department of Civil Engineering, MIT, 1989.

[158] Morikawa, T., Ben-Akiva, M., and McFadden, D. Incorporating Psychometric Data in Econometric Travel Demand Models, Working Paper, 1990.

[159] Morikawa, T., Ben-Akiva, M., and Yamada, K. Forecasting Intercity Rail Ridership Using Revealed Preferences and Stated Preferences. *Transportation Research Record*, 1328 (1991), 30–35.

[160] Muthén, B. Latent Variable Structural Equation Modeling with Categorical Data. *Journal of Econometrics 22* (1983), 43–65.

[161] Muthén, B. A General Structural Equation Model with Dichotomous, Ordered Categorical and Continuous Latent Variable Indicators. *Psychometrika 49* (1984), 115–132.

[162] Newell, A., and Simon, H. A. *Human Problem Solving*. Prentice-Hall, Englewood Cliffs, NJ, 1972.

[163] Nisbett, R. E., and Wilson, T. D. Telling more than we know: Verbal report on mental processes. *Psychological Review 84* (1977), 231–259.

[164] Olson, J. M., and Zanna, M. P. Attitudes and Attitude Change. *Annual Review of Psychology 44* (1993), 117–154.

[165] Pakes, A., and Pollard, D. Simulation and the Asymptotics of Optimization Estimators. *Econometrica 57*, 5 (1989).

[166] Park, C. W. A Seven-Point Scale and a Decision-Maker's Simplifying Choice Strategy: An Operationalized Satisficing-Plus Model. *Organizational Behavior and Human Performance 21* (1978), 252–271.

[167] Payne, J. W. Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis. *Organizational Behavior and Human Performance 16* (1976), 366–387.

[168] Payne, J. W., Bettman, J. R., and Johnson, E. J. Behavioral Decision Research: A Constructive Processing Perspective. *Annual Review of Psychology 43* (1992), 87–131.

[169] Pratt, J. W. Concavity of the Log Likelihood. *Journal of the American Statistical Association 76*, 373 (1981), 103–106.

[170] Prékopa, A. Logarithmic Concave Measures with Application to Stochastic Programming. *Acta Scientarium Mathematicarum 33* (1971), 301–316.

[171] Prékopa, A. On Logarithmic Concave Measures and Functions. *Acta Scientarium Mathematicarum 34* (1972), 335–343.

[172] Punj, G., and Stewart, D. W. Cluster Analysis in Marketing Research: Review and Suggestions for Application. *Journal of Marketing Research 20* (1983), 134–148.

[173] Quandt, E. Estimates of Modal Split. *Transportation Research 2*, 1968 (1968), 41–50.

[174] Rapoport, A., and Wallsten, T. S. Individual Decision Behavior. *Annual Review of Psychology 24* (1972), 131–176.

[175] RJ Software Inc. LINCS: Linear Covariance Structure: User's Guide, 1993.

[176] Roberts, J. H., and Lattin, J. Development and Testing of a Model of Consideration Set Composition. *Journal of Marketing Research 28* (1991), 429–440.

[177] Rosenberg, M. J., and Hovland, C. I. Cognitive, Affective and Behavioral Components of Attitudes. In *Attitude Organization and Change*, M. J. Rosenberg *al.*, Ed. Yale University Press, 1960, 1960.

[178] Russo, J. E., and Rosen, L. D. An Eye Fixation Analysis of Multialternative Choice. *Memory and Cognition 3* (1975), 267–276.

[179] Salomon, I. *Lifestyle as a Factor in Explaining Travel Behavior*. PhD thesis, Department of Civil Engineering, MIT, 1980.

[180] Salomon, I., and Ben-Akiva, M. Lifestyle as a Factor in Explaining Travel Behavior. *Environment and Planning A 15* (1983), 623–638.

[181] Samuelson, P. A. *Foundations of Economic Analysis*. Harvard University Press, Cambridge, Mass, 1983.

[182] Schwartz, G. Estimating the Dimensions of a Model. *Annals of Statistics 6* (1987), 461–464.

[183] Sclove, S. L. Population Mixture Models and Clustering Algorithms. *Communications in Statistics A6* (1977), 417–434.

[184] Sclove, S. L. Application of Model-selection Criteria to some Problems in Multivariate Analysis. *Psychometrica 52* (1987), 333–343.

[185] Seidenfus, H. S. Changes in Transport User's Modal Split - Paris: ECMT Round Table 69, 1985.

[186] Shanteau, J. An Information-Integration Analysis of Risky Decision Making. In *Human Judgment and Decision Processes*, M. Kaplan and S. Schwartz, Eds. Academic Press, New York, 1977.

[187] Sharp, C. Developments in Transport Policy in United Kingdom: The Value of Time Savings and of Accident Prevention. *Journal of Transport Economics and Policy 22* (1988), 235–238.

[188] Shepard, R. N. On Subjectively Optimum Selections Among Multi-attribute Alternatives. In *Human Judgments and Optimality*, M. W. Shelly and G. L. Bryan, Eds. Wiley, New York, 1964, pp. 257–281.

[189] Shocker, A. D., Ben-Akiva, M., Boccara, B., and Nedungadi, P. Consideration Set Influences on Consumer Decision-Making and Choice: Issues, Models, and Suggestions. *Marketing Letters 2:3* (1991), 181–197.

[190] Shugan, S. M. The Cost of Thinking. *Journal of Consumer Research 7* (1980), 121–134.

[191] Shyr, F.-Y. *Combining Laboratory and Field Data in Rail Fatigue Analysis*. PhD thesis, Department of Civil and Environmental Engineering, MIT, 1993.

[192] Simon, H. A. A Behavioral Model of Rational Choice. *Quarterly Journal of Psychology 69* (1955), 99–118.

[193] Simon, H. A. Rational Choice and the Structure of the Environment. *Psychological Review 63* (1956), 129–138.

[194] Simon, H. A. *Administrative Behavior*. Macmillan Company, New York, 1957.

[195] Simon, H. A. *The New Science of Management Decision*. New York University, New York, 1960.

[196] Simon, H. A. Invariance of Human Behavior. *Annual Review of Psychology 41* (1990), 1–19.

[197] Slovic, P. Differential Effects of Real Versus Hypothetical Payoffs on Choices Among Gambles. *Journal of Experimental Psychology 80* (1969), 434–437.

[198] Slovic, P., Fischoff, B., and Lichtenstein, L. Behavioral Decision Theory. *Ann. Rev. Psychology* (1977).

[199] Slovic, P., Lichtenstein, S., and Edwards, W. Boredom-induced Changes in Preferences Among Bets. *American Journal of Psychology 78* (1965), 324–329.

[200] Smith, M. B. The Personal Setting of Public Opinions: A Study of Attitudes towards Russia. *Public Opinion Quarterly 11* (1947), 507–523.

[201] Sobel, M. E., and Arminger, G. Modeling Household Fertility Decisions: A Nonlinear Simultaneous Probit Model. *Journal of the American Statistical Association 87*, 417 (1992), 38–47.

[202] Srivastava, R. K., Leone, R. P., and Shocker, A. D. Market Structure Analysis: Hierarchical Clustering of Products Based on Substitution-in-Use. *Journal of Marketing 45* (1981), 38–48.

[203] Stopher, P. Captivity and Choice in Travel-behavior Models. *ASCE Journal of Transportation Engineering 106* (1980), 4270–435.

[204] Sullivan, M. A., Mahmassani, H. S., and Yen, J. Choice Model of Employee Participation in Telecommuting Under Cost-Neutral Scenario. *Transportation Research Record 1413* (1993), 31–41.

[205] Svenson, O. A Note on Think Aloud Protocols Obtained During the Choice of a Home. Reports from the Psychological Laboratories, University of Stockholm, No. 421, 1974.

[206] Svenson, O. Process Descriptions of Decision-making. *Org. Behavior Human Performance 23* (1979), 86–112.

[207] Swait, J. *Probabilistic Choice Set Generation in Transportation Demand Models*. PhD thesis, Department of Civil Engineering, MIT, 1984.

[208] Swait, J. A Structural Equation Model of Latent Segmentation and Product Choice for Cross-sectional Revealed Preference Choice Data. working paper, June 1993.

[209] Swait, J., and Ben-Akiva, M. Empirical Test of a Constrained Discrete Choice Model. In *Behavioral Research for Transport Policy*. VNU Science Press, Utrecht, The Netherlands, 1986.

[210] Swait, J., and Ben-Akiva, M. Incorporating Random Constraints in Discrete Models of Choice Set Generation. *Transportation Research 21B*, 2 (1987a), 91–102.

[211] Swait, J., and Ben-Akiva, M. Empirical Test of a Constrained Discrete Choice Model: Mode Choice in Sao-Paulo, Brazil. *Transportation Research 21B*, 2 (1987b), 103–115.

[212] Swamy, P. *Statistical Inference in Random Coefficient Regression Models.* Springer-Verlag, New York, 1971.

[213] Swamy, P. Linear Models with Random Coefficients. In *Frontiers in Econometrics*, P. Zarembka, Ed. Academic Press, New York, 1974.

[214] Tesser, A., and Shaffer, D. R. Attitudes and Attitude Change. *Annual Review of Psychology 41* (1990), 479–523.

[215] Thaler, R. H. Mental Accounting and Consumer Choice. *Marketing Science 4* (1985), 199–214.

[216] Thaler, R. H. *The Winner's Curse: Paradoxes and Anamolies of Economic Life.* Free Press, New York, 1992.

[217] Thomas, T. C. The Value of Time for Passenger Cars: An Experimental Study of Commuters' Values. Vol. II of a final report to the U.S. Bureau of Public Roads, Stanford Research Institute, Menlo Park, California, May 1967.

[218] Thorngate, W., and Maki, J. Decision Heuristics and the Choice of Political Candidates. Tech. rep., University of Alberta, Canada, 1977.

[219] Titterington, D. M. Some Recent Results in the Analysis of Mixture Distributions. *Statistics 4* (1990), 619–641.

[220] Titterington, D. M., Smith, A. F. M., and Makov, U. E. *Statistical Analysis of Finite Mixture Distributions.* Wiley, New York, 1985.

[221] TRAIN, K. *Qualitative Choice Analysis: Theory, Econometrics and an Application to Automobile Demand*. MIT Series in Transportation Studies. The MIT Press, Cambridge, Mass, 1986.

[222] TRAIN, K., AND MCFADDEN, D. The Goods/Leisure Tradeoff and Disaggregate Work Trip Mode Choice Models. *Transportation Research 12* (1978), 349–353.

[223] TRIANDIS, H. C. *Attitude and Attitude Change*. John Wiley & Sons, New York, 1971.

[224] TVERSKY, A. Intransitivity of Preferences. *Psychological Review 76* (1969), 31–48.

[225] TVERSKY, A. Elimination by Aspects: A Theory of Choice. *Psychological Review 79* (1972a), 281–299.

[226] TVERSKY, A., AND KAHNEMAN, D. Judgment Under Uncertainty: Heuristics and Biases. *Science 185* (1974), 1124–1131.

[227] TVERSKY, A., AND KAHNEMAN, D. Rational Choice and the Framing of Decisions. *Journal of Business 59* (1986), 251–278.

[228] TVERSKY, A., SATTATH, S., AND SLOVIC, P. Contingent Weighting in Judgment and Choice. *Psychological Review 95* (1988), 371–384.

[229] VIEIRA, L. *The Value of Service in Freight Industry*. PhD thesis, Department of Civil Engineering, MIT, 1992.

[230] VON WINTERFELDT, D., AND FISCHER, G. Multiattribute utility theory: Models and assessment procedures. In *Utility, probability, and human decision making*, D. Wendt and C. Vlek, Eds. Reidel, Dordrecht, 1975.

[231] WALLSTEN, T. S. Proceses and Models to Describe Choice and Inference. In *Cognitive Processes in Choice and Decision Behavior*, T. S. Wallsten, Ed. Erlbaum, Hillsdale, 1980.

[232] Wardman, M. A Comparison of Revealed Preference and Stated Preference Models of Travel Behavior. *Journal of Transport Economics and Policy 22*, 1 (1988), 71–91.

[233] Waters II, W. G. Values of Travel Time Savings Used in Road Project Evaluation: A Cross Country/Jurisdiction Comparison, *australasian road research forum*, 1992.

[234] Waters II, W. G. Variations in the Value of Travel Time Savings: Empirical Studies and the Values for Road Project Evaluation. Faculty of Commerce, The University of British Columbia and Institute of Transport Studies, The University of Sydney Working paper: ITS-WP-93-18, 1993.

[235] Wermuth, M. *Structure and Calibration of a Behavioral and Attitudinal Binary Mode Choice Model between Public Transportation and Private Car.* PTRC Summer Annual Conference, 1978.

[236] Widlert, S. *Stated Preference Studies: The Design Affects the Results.* Seventh International Conference on Travel Behavior, Valle Nevado, Santiago, Chile, June, 1994.

[237] Wiley, D. E. The Identification Problem for Structural Equation Models with Unmeasured Variables. In *Structural Models in the Social Sciences*, A. S. Goldberger and O. D. Duncan, Eds. Academic Press, New York, 1973.

[238] Wind, Y. Issues and Advances in Segmentation Research. *Journal of Marketing Research 5*, 3 (1978), 329–350.

[239] Winfrey, R. *Economic Analysis of Highways.* International Textbook Company, Scranton, Pennsylvania, 1969.

[240] Winston, C. A Disaggregate Model of the Demand for Intercity Freight Transportation. *Econometrica 49*, 4 (1981), 981–1006.

[241] Winston, C. The Demand for Freight Transportation: Models and Applications. *Transportation Research 17A*, 6 (1983), 419–427.

[242] Wolfe, J. H. Pattern Clustering by Multivariate Mixture Analysis. *Multivariate Behavioral Research 5* (1970), 329–350.

[243] Wright, P., and Barbour, F. Phased Decision Strategies: Sequels to an Initial Screening. In *Studies in Management Sciences: Multiple Criteria Decision Making*, M. K. Starr and M. Zeleny, Eds. North-Holland, New York, 1977, pp. 91–110.

[244] Wu, C. F. J. On the Convergence Properties of the EM Algorithm. *The Annals of Statistics 11*, 1 (1983), 95–103.

[245] Yarmal-Vuarl, F., and Ataman, E. Noise Histogram and Cluster Validity for Gaussian-mixtured data. *Pattern Recognition 20* (1987), 385–401.

[246] Zenor, M., and Srivastava, R. K. Inferrring Market Structure with Aggregate Data: A Latent Segment Logit Approach. *Journal of Marketing Research 30* (August 1993).

[247] Zlatopper, T. J., and Austrian, Z. Freight Transportation Demand: A Survey of Recent Econometric Studies. *Transportation Journal 16* (1989), 27–46.