NOTAT

Fundamentals of Choice Models: A Synthesis of Theoretical, Methodological and Practical Issues

Bharat P. Bhatta

N-NR 1/2011
Avdeling for samfunnsfag

HØGSKULEN i SOGN OG FJORDANE

HØGSKULEN i SOGN og FJORDANE

NOTAT

Postboks 133, 6851 SOGNDAL, telefon 5767600 telefaks 57676100

TITTEL	NOTATNR.	DATO
Fundamentals of Choice Models: A Synthesis of Theoretical, Methodological and Practical Issues	1/2011	19.09.2011
PROSJEKTTITTEL Discrete Choice Analysis with Applications to Travel Demand	TILGJENGE	TAL SIDER 24
FORFATTAR Bharat P. Bhatta	PROSJEKTLEIAR/-ANSVERLEG	
OPPDRAGSGJEVAR	EMNEORD Discrete choice models; Random utility maximization; Decision makers; Utility function; Model formulation	

SAMMANDRAG

This paper synthesizes the fundamentals of discrete choice models. This paper also discusses the basic concepts and theory underlying the econometrics of discrete choice, specific choice models, estimation method, model building and tests, and applications of discrete choice models. The work highlights the relationship between economic theory and discrete choice models: how economic theory contributes to choice modeling and vice versa.

PRIS	ISSN	ANSVARLEG SIGNATUR
kr 90,-	0806-1696	Bente Karin Hvsaby

Fundamentals of Choice Models: A Synthesis of Theoretical, Methodological and Practical Issues

Bharat P. Bhatta

Sogn og Fjordane University College, P.O. Box 133, N-6851 Sogndal, Norway

Bharat.bhatta@hisf.no

Abstract
This paper synthesizes the fundamentals of discrete choice models. This paper also discusses the basic concepts and theory underlying the econometrics of discrete choice, specific choice models, estimation method, model building and tests, and applications of discrete choice models. The work highlights the relationship between economic theory and discrete choice models: how economic theory contributes to choice modeling and vice versa.

Keywords: Discrete choice models; Random utility maximization; Decision makers; Utility function; Model formulation

1 Introduction

Discrete choice models, which belong to random utility maximization (RUM) models, are widely used to analyze individual choice behavior (see Ben-Akiva and Lerman, 1985; Train, 1986 and 2003; McFadden, 2000 and 2001; Hensher et al., 2005). In the discrete choice framework, a decision maker facing a mutually exclusive and collective exhaustive set of finite number of alternatives obtains utility from each alternative and chooses the one with the highest utility. Discrete choice models thus deal with discrete or qualitative[1] outcomes involving a behavioral choice such as choice of occupation or mode of travel[2]. The discrete outcomes can be ordered, e.g., number of cars a household owns or a respondent's level of agreement to a statement in a Likert scale[3], or unordered, that is, the ordering of the outcomes has no effect on the choice process, e.g., choice of travel mode or type of housing.

[1] They are nominal scale variables which simply denote categories, so they are also called categorical variables. Conceptually, we can classify the discrete outcomes as those involving a behavioral choice (e.g. choice of occupation or mode of travel) or those simply describing discrete outcomes of a physical event (e.g. type of an accident) (Washington et al., 2003).

[2] There are numerous examples on discrete outcomes involving a behavioral choice in diverse fields, including transportation for choice of travel mode, route, destination, car brand, type of a vehicle to own, and so on; economics for choice of technology, market participation, production plant/plan, etc; sociology for choice of marital status (single-married-living together), marriage partner, etc; housing for choice of residential location, type of housing (rent-single own-company own), etc; marketing for choice of brand, menu, ad media (radio-TV-newspaper), etc; business for choice of costumer, portfolio, securities, etc; and education for choice of college, degree, subject/discipline choice, etc., to name only a few.

[3] A Likert scale is a psychometric scale widely used in a survey research in order to know a respondent's attitude and level of agreement (or disagreement) to a statement. The scale is named after its discoverer Rensis Likert

The discrete choice models based on the RUM framework are important tools for analysis of individual choice behavior and have successfully been applied in diverse fields, including, transportation (c.f. McFadden, 2000; Hess, 2005; Bhatta, 2010)[4], consumer behavior (Train et al., 1987; Ashok et al., 2002), education (DesJardins et al., 1999), political science (Glasgow, 2001), economics (Herriges and Phaneuf, 2002) and peace and conflict (Barros and Proenca, 2005) to name only a few. Discrete choice analysis is an increasingly popular tool to model individual choice behavior.

Transportation is the most important field of research and application of discrete choice models. Initially, the discrete choice models in transportation were used in analyses of binary choice of travel modes in the 1960s (e.g., Warner, 1962; Lave, 1969; Lisco, 1967; Quarmby, 1967; cited in Ben-Akiva and Lerman, 1985). The models were used in estimating a value of travel time savings and predicting the market shares of alternative travel modes. The choice models have been extensively applied in transportation for nearly five decades[5]. Travel demand modeling is one of the well-researched topics. There is an extensive and lively body of literature on the topic (see Domencich and McFadden, 1975; Ben-Akiva and Lerman, 1985; Train, 1986 and 2003; McFadden, 2000; Hess, 2005; Bhatta, 2010 and relevant references therein). Highly advanced models such as complex generalized extreme value (GEV) models (e.g. GEV models allowing for cross-nesting, multi-level GEV models, recursive GEV models, and so on) and models with mixed distributions (e.g. mixed logit) are developed in the RUM framework of travel demand (see Train, 2003; Hess et al., 2007; Hess, 2005; McFadden, 2000; and relevant references therein).

The remainder of the paper is organized as follows. Section 2 discusses the theoretical foundations of choice models followed by commonly used models in section 3. Section 4 presents the maximum likelihood estimation procedures. Section 5 gives a detailed exposition of model building and tests in discrete choice analysis. Section 6 briefly explains validation of a choice model followed by a brief overview of applications of choice models in section 7. Finally, Section 8 concludes the paper.

2 The Theoretical Foundation of Discrete Choice Models

This section discusses discrete choice theory and how it relates to economic consumer theory and presents the basics of the random utility maximization (RUM) models.

2.1 Discrete choice theory

Discrete choice theory, which typically involves the following elements in the choice process, concerns the behavioral choice of discrete alternatives (c.f. Ben-Akiva and Lerman, 1985).

(1932). The Likert scale is normally used to describe items with five or seven ordered response options. For example, the response options in a typical five-point Likert scale are: (1) strongly disagree, (2) disagree, (3) neutral, (4) agree, and (5) strongly agree.

[4] There are enumerable applications in transportation.

[5] McFadden (2000) excellently reviews a historical account of the development of the state-of-the-art in the field of travel behavior research and its connection to RUM models.

- The decision makers are the individual persons or households or firms or governments or any other decision making units that possess preferences or tastes over alternatives. For example, travelers are the decision makers for a trip to work.
- The characteristics of the decision makers are income, age, sex, and so on the decision makers.
- The alternatives are competing products, course of action, or any other option or items over which a decision must be made. The alternatives form the choice set The choice set in the RUM framework of discrete choice analysis exhibits three characteristics, viz., mutually exclusive, collectively exhaustive and finite (c.f., e.g., Train, 2003).The choice set that includes all the alternatives from the perspective of population (or an analyst) is called the universal choice set The set of alternatives that is viable for a decision maker is the feasible choice set The feasible choice set is thus the subset of the universal choice set
- The attributes are something that make the alternatives useful (*or just opposite*) for a decision maker. For example, travel time, travel cost, comfort and so on are the attributes of a travel mode.
- The decision rule is the criteria followed by the decision maker to come up with the actual choice. Discrete choice theory uses random utility maximization as the decision rule.

As utility is the fundamental concept in economic theory (c.f., e.g., Varian, 1992; Silberberg and Suen, 2001) and utility maximization is the decision rule in discrete choice analysis, the discrete choice models are based on the economic theory of utility maximization (Train, 2003; Ben-Akiva and Lerman, 1985). The utilities are, however, latent variables. The actual choice, which we can observe as analysts, is a manifestation of the underlying utilities of the alternatives.

Since choice behavior of a decision maker is probabilistic from the perspective of an analyst, the discrete choice theory incorporates this probabilistic behavior through the concept of random utility. According to random utility theory, a concept first proposed by Thurstone (1927) and subsequently developed by Luce (1959) and Marschak (1960), the utility of an alternative is a random variable which consists of observable and unobservable parts from the perspective of analyst. The observable part of utility is assumed to be a function of attributes of alternatives (following Lancaster, 1966[6]) and characteristics of the decision makers. The characteristics of decision makers are included in utility function in order to capture heterogeneity across decision makers since all the decision makers are not alike. The final component of the utility is a random term introduced in order to account for uncertainty due to analyst's incomplete information about the choice process. The random utility maximization is thus the basic principle in discrete choice theory which directly follows from microeconomic consumer theory.

2.2 Operationalization of the RUM model

In the RUM framework of discrete choice analysis, a decision maker facing a mutually exclusive and collective exhaustive set of finite number of alternatives obtains utility from each alternative and chooses the one with the highest utility. But the analyst, who is an observer of the choice process, is not able to observe the utility of the alternatives, therefore, decomposes the utility into

[6] Lancaster proposed a new approach to consumer theory. According to the approach, it is not a good itself but its attributes that determine its utility. Utility is therefore a function of attributes of the goods.

two parts for analytical purposes: (i) an observable part and (ii) an unobservable part. Consider there are J alternatives available for a decision maker n. The utility of the alternative $i \in J$ for the decision maker n, U_{in}, can therefore be written as:

$$U_{in} = V_{in} + \varepsilon_{in} \quad (1)$$

where V_{in} and ε_{in} represent the observed and unobserved parts of the utility of the alternative i for the decision maker n respectively from the point of view of an analyst. V_{in} is the systematic or representative utility. The systematic utility is deterministic in the sense that it is broadly a function of a vector of attributes of the alternative, Z_{in}, and a vector of characteristics of the decision maker, S_n, so:

$$V_{in} = V(Z_{in}, S_n) \quad (2)$$

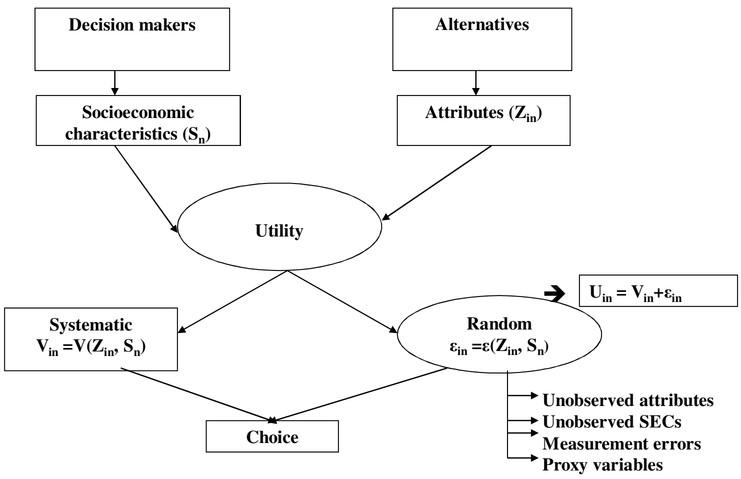

Figure 1. Diagrammatic representation of theory underlying a discrete choice model

Representing both attributes of the alternative and characteristics of the decision maker by x_{in} and assuming a linear formulation[7] of the systematic utility function, equation (2) can be rewritten as:

$$V_{in} = V(x_{in}) = \beta' x_{in} \quad (3)$$

where β is a vector of parameters representing the tastes of the decision makers to be estimated from the data. Figure 1 clearly illustrates the diagrammatic representation of theory underlying a discrete choice model.

[7] Alternatively, we can use non-linear formulations such as Box-Cox or Box-Tukey transformations, see, e.g., Ben-Akiva and Lerman (1985), page 179.

On the other hand, the analyst is not able to observe ε and treats this term as a random variable or random utility (a disturbance or error term). The total utility is thus a random variable from the perspective of the analyst. Incorporating the disturbances in the utility function imply that discrete choice models explicitly take into account the uncertainty in modeling in the sense that we cannot explain the behavior of individuals only by observable variables. The disturbances are assumed to capture unobserved attributes of the alternatives, unobserved taste variations of the decision makers, errors in measuring the variables and use of proxy (or instrumental) variables (Manski, 1977). The inclusion of a random component in the model recognizes the fact that the choice process dealing with human behavior is inherently probabilistic from the perspective of the analyst, leading to the RUM model where the alternative with the maximum systematic utility has the highest probability of being chosen.

Consider that C is the choice set for the population in a study area, the universal choice set, and that C_n is the choice set for an individual decision maker n, the feasible choice set[8]. The probability that an individual decision maker n chooses the alternative i is equal to the probability that the utility of alternative i, U_{in}, is greater than or equal to the utilities of all other available alternatives in the choice set (Ben-Akiva and Lerman, 1985 (p. 101); Train, 2003 (p. 19)) as given by:

$$\begin{aligned} P_{in} &= \Pr(U_{in} \geq U_{jn}, \forall j \in C_n, j \neq i) \\ &= \Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}, \forall j \in C_n, j \neq i) \\ &= \Pr(\varepsilon_{jn} - \varepsilon_{in} \geq V_{in} - V_{jn}, \forall j \in C_n, j \neq i) \\ &= \int_{\varepsilon} I(\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn}, \forall j \in C_n, j \neq i) f(\varepsilon_n) d\varepsilon_n \qquad (4) \end{aligned}$$

where I(.) is the indicator function which equals 1 if the expression inside parentheses is true and 0 otherwise. Equation (2.4) shows that only differences in utility matter leading to the conclusion that the choice process is relative. Consequently, the addition of the same constant to utilities of all the alternatives, or the multiplication to all the utilities by the same positive constant, does not change the probability of the alternative being chosen.

The mean of the error terms can be included in the deterministic part of utility as an additional parameter called alternative specific constant (ASC). When the deterministic part of the utility includes an ASC for all the alternatives except one, the error terms have zero mean by construction. The number of ASCs included in the model is thus equal to the number of alternatives minus one since one of the ASCs is normalized to zero. Apart from providing the mean of the error terms, the ASCs ensure the average probability of each alternative in the sample exactly equals the proportion of the decision makers actually choosing the alternative.

Different assumptions on the joint distribution of the random component of the utility function results in different structures of discrete choice models (see section 3). Discrete choice models generally postulate that the probability of an individual decision maker choosing an alternative from a mutually exclusive and collectively exhaustive set of finite number of alternatives is a function of the socioeconomic characteristics and the relative attractiveness of the alternative.

[8] The feasible choice set includes all the alternatives that are possible for a decision maker to choose, that is, the choice set from the perspective of the decision maker.

2.3 Microeconomic consumer theory and discrete choice models

As discussed above, the discrete choice theory is fundamentally based on microeconomic theory of consumer behavior. In microeconomic consumer theory, a consumer is assumed to maximize utility subject to budget constraints where utility is a function of quantities of goods (c.f., e.g., Varian, 1992; Silberberg and Suen, 2001). The utility maximization subject to budget constraints is a typical constrained optimization problem which we can solve in order to derive the demand functions that express the choice of the consumer for given prices and income. If we substitute the demand functions back into the utility function, we obtain a function called indirect utility function. The indirect utility, which is the maximum utility achievable at given prices and income, is a function of prices of the goods and the income of the consumer. The discrete choice theory uses this indirect utility function in operationalization of the RUM model.

Traditional consumer theory assumes deterministic behavior of consumers. Discrete choice theory on the other hand recognizes that human behavior is probabilistic from the perspective of analyst. The discrete choice theory deals with a choice of discrete alternative among a set of finite and mutually exclusive alternatives in contrast to continuous goods in traditional consumer theory. As a result, we cannot apply the usual first order conditions to derive the demand functions in discrete choice theory although we retain the principle of utility maximization. The outcomes of the discrete choice models are the probabilities of a decision maker choosing an alternative as opposed to the quantity of goods chosen in traditional consumer theory.

3 Commonly Used Choice Models

There are many state-of-the-art discrete choice models. Review of the choice models is beyond the scope of this work. I refer to Hess (2005) for the excellent review of choice models. This section only discusses the commonly used models such as logit, the generalized extreme value family and the nested logit, probit, and the mixed logit models.

3.1 The logit model

The logit model is derived under the assumption that all the disturbances ε_{in} are independently and identically Gumbel or type I extreme value (EV) distributed. The logit choice probability[9] of alternative i for decision maker n can be given by (c.f. Ben-Akiva and Lerman, 1985; Train, 2003):

$$P_{in} = \frac{e^{\beta' x_{in}}}{\sum_{j \in C_n} e^{\beta' x_{jn}}} \tag{5}$$

This equation shows that the logit choice probability has the analytically tractable expression[10]. The logit model is by far the easiest and most widely used model of discrete choice models mainly due to its closed form, ease of estimation and interpretation, and the simplicity to add or remove the alternatives.

---

[9] Luce (1959) derived the formula of the logit choice probability first under the assumption of independence from irrelevance alternatives and Marschak (1960) showed that the model was consistent with the RUM hypothesis. I consistently use the term ‘logit’ model instead of the ‘multinomial logit’ model in this paper. So I use the logit models, the nested logit models, and so on in this paper.

[10] The analytically tractable expression is commonly called ‘closed form’. I will also use this term to refer to an analytically tractable expression throughout the paper.

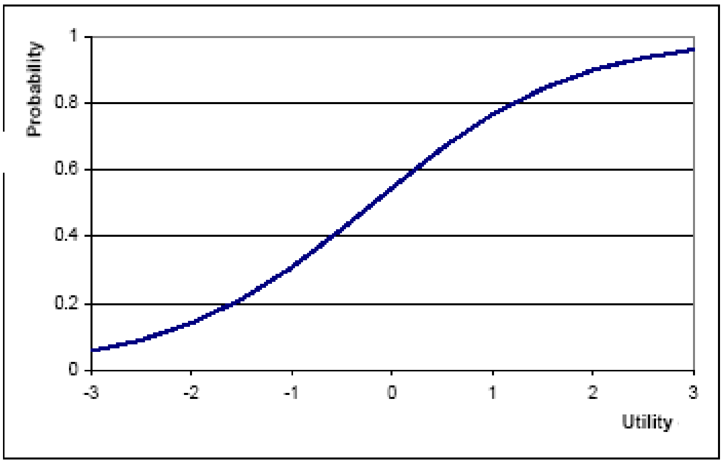

Figure 2. Shape of the logit choice probability

The relation between the logit choice probability and the systematic utility is sigmoid or S-shaped[11] as shown in figure 2 (see Train, 1986). The shape implies that a small change in the utility of an alternative due to a change in its attributes has little effect on its choice probability if the systematic utility of the alternative is very high (or very low) compared to that of other alternatives in the choice set The impact of a change in the attributes is the highest when the choice probability is close to 0.5 and the impact gets smaller as the choice probability approaches one or zero. The shape of the choice probability has thus important policy implications.

The independently and identically distributed (i.i.d.) assumption of error terms requires zero correlation of unobserved factors across the alternatives and homoscedastic variance for all the alternatives. However, the logit model has been widely criticized for its independence from irrelevant attributes (IIA) property resulting from the i.i.d. assumption of disturbance terms. The IIA property restricts the ratio of the choice probabilities of any two alternatives in the choice set to be independent of the existence of other alternatives and/or their attributes[12]. This restriction implies that introduction of a new alternative or a change in an attribute of any alternative will change the probability of existing alternatives in proportion to their probabilities before the change. The IIA property is a major limitation of the logit model since it implies equal competition across the alternatives. This does not hold true in choice situations if some alternatives are more similar in unobserved factors than other alternatives in the sense that the random terms of the alternatives share a common factor (and hence are correlated). The 'red bus/blue bus paradox' is the extreme example of the problem due to the IIA property[13].

[11] It is not only the logit model but most discrete choice models share this relation between the choice probability and the systematic utility (c.f. Train, 2003, page 42).

[12] Mathematically, the IIA property of any two alternatives i and k for decision maker n can be expressed as:

$$\frac{P_{in}}{P_{kn}} = \frac{e^{V_{in}} \Big/ \sum_{j \in C_n} e^{V_{jn}}}{e^{V_{kn}} \Big/ \sum_{j \in C_n} e^{V_{jn}}} = \frac{e^{V_{in}}}{e^{V_{kn}}} = e^{V_{in} - V_{kn}}$$

[13] See, for example, Train (1986), page 19, more about the 'red bus/blue bus paradox'.

The IIA property is also the power of the logit model because the IIA provides basically two practical advantages to the logit model. Firstly, it allows sampling of alternatives for estimation because the estimated parameters of a logit model for the sample of alternatives are the same as the parameters of the logit model for the full choice set for many sampling procedures. When the choice set is too large, sampling a subset of alternatives leads to reduced cost of data collection and computation. Secondly, it allows forecasting demand for a new alternative. The demand for a new alternative can be forecast in the logit model simply by adding an e^V term of the new alternative to the denominator of the formula of the choice probability[14].

The logit model is restrictive because it cannot allow for random taste heterogeneity across decision makers, unrestricted substitution patterns across alternatives and correlation in unobserved factors over time and space (c.f. Train, 2003). Nevertheless, the logit model is widely used in the large scale real world applications because of its ease of estimation, interpretation and application.

3.2 The generalized extreme value (GEV) family and the nested logit model

The GEV family of models, developed by McFadden (1978), assumes that the unobserved factors for all the alternatives are jointly distributed as a generalized extreme value. The GEV models retain most of the computational advantages of the logit model, while at the same time they allow a general pattern of inter-dependence among the unobserved factors across the alternatives, which the logit models cannot. The GEV family thus consists of analytically tractable closed form discrete choice models based on the RUM framework. The choice probability of alternative i for decision maker n with GEV models is given by (adapted from McFadden (1978), p. 81):

$$P_{in} = \frac{e^{V_{in}} \dfrac{\partial G}{\partial e^{V_{in}}}(e^{V_{1n}}, e^{V_{2n}}, \ldots, e^{V_{J_n n}})}{\mu G(e^{V_{1n}}, e^{V_{2n}}, \ldots, e^{V_{J_n n}})} \tag{6}$$

where G is a homogeneous of degree μ and non-negative differentiable function with certain properties (see Ben-Akiva and Lerman, 1985 and Hess, 2005 for its properties). Readers are referred to McFadden (1978), Ben-Akiva and Lerman (1985), Train (2003), and Daly and Bierlaire (2005) (among others) for detail discussions about the GEV family of models.

The nested logit[15] (NL) model is the basic member of the GEV family that overcomes some of the limitations of the logit model. The NL model is formed by dividing the choice set into subsets of alternatives, called nests, which are more similar to each other with respect to unobservable factors than they are to other alternatives in the choice set Figure 2 illustrates a simple NL model where metro and bus are grouped into a nest since they are both public transport modes, presumably sharing the common unobserved factors. The alternatives in a common nest exhibit a higher degree of similarity and competitiveness than the alternatives in different nests. This level of competitiveness, represented by cross-elasticities between pairs of alternatives, is identical for all pairs of alternatives in the nest. The NL model is useful if some alternatives are more similar to other alternatives in unobserved factors. The NL model is thus a natural modeling process for interrelated alternatives. The NL model partially relaxes the IIA

[14] It must be the case that the estimated parameters are sufficient to calculate the V for the new alternative.
[15] The NL models are sometimes called the hierarchical (or sequential) logit (see Ortuzar and Willumsen, 2001).

property but retains the closed form. The IIA property holds within nests but not across nests. Readers are referred to Ortuzar (2001) for a historical account of the development of the NL models.

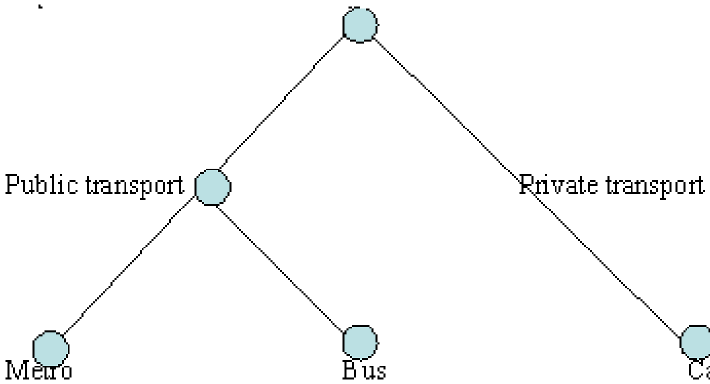

Figure 3. Illustration of a simple NL model

The two-level NL model is the simplest of the NL models where the alternatives in the choice set C_n are divided into mutually exclusive subsets of alternatives (see figure 3). In the NL model, the alternatives within a nest share a random term. The utility of alternative i in nest C_{mn} is given (suppressing n) by:

$$U_i = V_i + \varepsilon_m + \varepsilon_{im} \tag{7}$$

ε_m are independent across nests by an assumption and ε_{im} are i.i.d. EV. The choice probability with the NL model can be expressed as a product of the logit choice probabilities of the alternative in the nest and the nest containing the alternative. The choice probability of alternative i belonging to nest C_{mn} for decision maker n with the NL model is thus given by:

$$P(i|C_n) = P(C_{mn}|C_n)\, P(i|C_{mn})$$

$$= \frac{e^{\lambda_m I_{mn}}}{\sum_{k=1,\dots,M} e^{\lambda_k I_{kn}}} \frac{e^{V_{in}/\lambda_m}}{\sum_{j \in C_{mn}} e^{V_{jn}/\lambda_m}} \tag{8}$$

where $P(C_{mn}|C_n)$ is the probability that the nest is chosen (marginal probability), $P(i|C_{mn})$ is the probability of the alternative being chosen conditional on the nest is chosen (conditional probability) and I_{mn} is the logsum variable[16] that serves as an attribute of the nest C_{mn} defined as:

$$I_{mn} = \ln \sum_{j \in C_{mn}} e^{V_{jn}/\lambda_m} \tag{9}$$

λ_m is the logsum parameter associated with the nest C_{mn} that measures the degree of correlation of unobserved factors among the alternatives in this nest. The logsum parameter is associated with each nest of the NL structure. λ_m is restricted in between 0 and 1 to ensure consistency with

[16] Logsum, as the term implies, is the log of a sum of systematic utilities of lower level alternatives in an NL model. It is also called inclusive value. It is the maximum expected utility from the available alternatives in a logit framework. It is also a measure of consumer's surplus in the context of the logit models of discrete outcomes.

the RUM model[17]. If λ_m=1, the alternatives are not correlated and the NL model reduces to the logit model.

3.3 The probit model

The equation (4) results in the probit model if the error terms are jointly normally distributed. The probit model was initially developed in psychology (Thurstone, 1927). Probit is a very flexible model because it allows for random taste heterogeneity of decision makers, unrestricted substitution patterns across alternatives and correlation in unobserved factors over time and space. Despite the flexibility of the probit model, the logit models are chosen instead of the probit model because of ease of estimation, interpretation and application of the logit models. I refer to Horowitz (1991), Weeks (1997) and Train (2003) (among others) regarding the probit model.

3.4 The mixed logit model

The mixed logit (ML) is a highly flexible model for examining discrete choices (c.f. McFadden and Train, 2000; Train, 2003; Hess, 2005; Hess et al., 2007; Bhatta, 2010). The ML model allows for random taste heterogeneity of decision makers, unrestricted substitution patterns across alternatives and correlation in unobserved factors over time (Train, 2003). Unlike the logit model, the choice probability of the ML model does not have a closed form. Consequently, the model has to be estimated with the help of simulation. The ML model allows for any distributions for the random coefficients of the model. McFadden and Train have shown that the ML model can represent any RUM model. The reader is referred to Bhatta (2010, chapter 7) for formulation, more detailed discussions and applications of the ML models.

4 Estimation of Discrete Choice Models

Discrete choice models with a closed form may be estimated by any search algorithm or iterative solution method. They are generally estimated by the method of maximum likelihood. The method of maximum likelihood, as the term implies, gives the values of the parameter estimates that maximize the likelihood of observing the actual choices with the observed data. I present the estimation procedures with the method of maximum likelihood for exogenous sample[18] in this section.

Consider N is the sample of decision makers choosing an alternative i from the choice set C_n. The likelihood function (L) to be maximized for this multinomial choice problem (c.f. Ben-Akiva and Lerman, 1985; Train, 2003) is given by:

[17] Interpretation of the logsum parameter:

- $\lambda_m>1$ => Not consistent with the RUM model. Reject the NL model in favor of the logit model.
- $\lambda_m=1$ => No correlation across the alternatives in the nest. The NL model reduces to the logit model.
- $0<\lambda_m<1$ => Correlation across the alternatives in the nest which requires the NL model.
- $\lambda_m=0$ => Perfect correlation across the alternatives in the nest. The alternatives in the nest can be treated as one alternative.
- $\lambda_m<0$ => Not consistent with the RUM model. Reject the NL model in favor of the logit model.

[18] See Manski and McFadden (1981) and Cosslett (1981) who discuss estimation methods for different sampling methods.

$$L = \prod_{n=1}^{N} \prod_{i \in C_n} P_{in}^{y_{in}} \tag{10}$$

where $P_{in} = \dfrac{e^{\beta' x_{in}}}{\sum_{j \in C_n} e^{\beta' x_{jn}}}$ for the logit model

where y_{in} = 1 if the decision maker n chooses the alternative i, 0 otherwise. Taking the logarithm of equation (2.10) gives the log-likelihood function as follows:

$$LL = \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \ln P_{in} = \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \left(\beta' x_{in} - \ln \sum_{j \in C_n} e^{\beta' x_{jn}} \right) \tag{11}$$

Differentiating equation (11) with respect to β_k and setting the expression equal to zero gives the first order condition for maximization of the likelihood function (10) as:

$$\frac{\partial (LL)}{\partial \hat{\beta}_k} = \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \left(x_{ink} - \frac{\sum_{j \in C_n} e^{\beta' x_{jn}} x_{jnk}}{\sum_{j \in C_n} e^{\beta' x_{jn}}} \right) = 0$$

$$or \tag{12}$$

$$\sum_{n=1}^{N} \sum_{i \in C_n} [\ y_{in} - P_{in}\ ] x_{ink} = 0$$

Solving equation (12)[19] in β_k gives the values of β_k that maximizes the likelihood function provided that the second order condition[20] is met

The equation (12) is solved iteratively using the Newton-Raphson or BHHH or BFGC or PRCG algorithms[21]. McFadden (1974b) shows that the log-likelihood function in equation (11) is globally concave giving the unique solution (i.e., the parameter estimates). He also confirms that the parameter estimators are consistent, asymptotically normal and asymptotically efficient under fairly general conditions.

The models which do not have a closed form are estimated with the help of simulation. For example, the mixed logit model is estimated by the method of maximum simulated likelihood (see Train, 2003).

5 Model Building and Tests

A model building process in discrete choice analysis typically involves the following steps:

- Formulation of the systematic utility function

---

[19] There is one equation like (12) for each parameter and we have to solve a simultaneous nonlinear equation system to find the optimum values of parameters. This has to be done iteratively that is what the algorithms do.

[20] The second order condition is the positive value of the second derivative of equation (11) with respect to β_k estimated at β_k.

[21] BHHH, BFGC, DFP, PRCG stands for the Berndt, Hall, Hall & Hausman, Broyden-Fletcher-Goldfarb-Shanno, Davidon-Fletcher-Powell, Polak-Ribiere-type Conjugate Gradient respectively (see, e.g., GAUSS manual for maximum likelihood estimation (www.aptech.com).

- Testing the formulation of systematic utility function given the model structure (for example, the logit model)
- Testing the model structure given the formulation of systematic utility function

The alternative formulations and model structures are tested to select the 'best' formulation and model structure.

5.1 Formulation of systematic utility function

Formulation of systematic utility function concerns the choice of functional form, explanatory variables and the form in which they enter the model. Since utility maximization is the fundamental principle underlying the discrete choice models, it is important to discuss how different variables entering the model affect the utility based on economic theory, behavioral rationale and intuition.

The linear in the parameters[22] is a normally applied functional form in formulating the utility function. Alternatively, we can use non-linear formulations such as Box-Cox or Box-Tukey transformations[23].

Theory and previous studies, purpose of model building, judgment of model builder, behavioral rationale and statistical techniques generally guide the modeler about the choice of variables and the form in which they enter the model. In addition to the variables influencing the choice, an alternative specific constant for each alternative except one must be included in the utility function. The formulation of systematic utility function is discussed in the following paragraphs with an example of travel mode choice model (see appendix 1 for a brief discussion about the typical variables that enter a travel mode choice model).

An attribute of the alternative entering a model as a generic or an alternative specific variable has important theoretical and behavioral implications. If an attribute has the same marginal utility for all the alternatives, it enters the model as a generic variable. Otherwise, the attribute enters as an alternative specific variable. For example, if a minute of travel time (or a dollar of cost) has the same marginal disutility regardless of mode, then travel time (or travel cost) enters the model as a generic variable. Otherwise, it enters as an alternative specific variable. In a travel demand model, travel cost generally enters the utility function as a generic variable while the travel time enters as an alternative specific variable.

Disaggregating total travel time, for example with public transport, into different components such as access/egress time, invehicle time, waiting time and so on implies that different components of time may have the different marginal utility and hence different effects on choice probability. This is in fact consistent with experiences and behavioral rationale because travelers consider wait time to be more troublesome than invehicle time. The use of total travel time, on the other hand, assumes that any of the components has the same marginal utility.

---

[22] A function is linear in the parameter if the parameter has a power of 1 only and is not multiplied or divided by any other parameter in the model.

[23] The Box-Cox transformation $\beta\frac{x^{\lambda}-1}{\lambda}$ for x>0. The Box-Tuky transformation $\beta\frac{(x+\alpha)^{\lambda}-1}{\lambda}$ for x+α>0 (see, Ben-Akiva and Lerman, 1985, page 179).

Socioeconomic characteristics of a decision maker such as income, age, sex, car ownership and so on capture heterogeneity and taste variation of the decision makers. They cannot appear in the same way in all the utility functions as the attributes of alternative since they do not vary over the alternatives. They can enter the model if they are specified in ways that create differences in utility over the alternatives. Either the coefficients of a characteristic are introduced in all the utility functions but one (by normalizing the coefficient of the characteristic to zero in one of the utility functions) (or some of the relevant utility functions only) or the characteristic is interacted with the attributes of the alternatives, e.g., cost/income. In normalization, coefficients are interpreted as the differential effect of characteristics on the utility of the alternative compared to the normalized one. In interaction, the characteristics affect the differences in utility through their interaction with attributes of the alternative.

The form in which a variable enters a model also has important theoretical and behavioral implications. Linear or logarithmic are the commonly used forms in practice. The linear form assumes that a variable has the same marginal utility regardless of its current value. For example, changes in disutility for a minute change in travel time in a 10 minutes trip or a one hour trip are the same according to the linear form. The logarithmic form, on the other hand, takes into account the value of the variable before the change. Consequently, changes in disutility for a minute change in travel time in the 10 minutes or one hour trips are entirely different because travelers perceive the marginal effects of those trips differently. Other transformations such as power series expansion and piecewise linearization can also be applied without violating the linear in the parameters formulation.

5.2 Testing the formulations of systematic utility function

The formulations of systematic utility function of competing models are tested using 'informal tests' and likelihood ratio test to select the 'best' model (c.f. Ben-Akiva and Lerman, 1985; Train, 1986 and 2003; Hensher et al., 2005). The informal tests help to evaluate the reasonableness of the model with respect to signs and relative magnitudes of parameter estimates according to economic theory (or some other theory), behavioral rationale and intuition. The likelihood ratio test helps to identify the better formulation of the systematic utility function based on the goodness of fit of the model to the data.

Informal tests

Model builder's judgments with respect to signs and values of parameter estimates are the important aspects of model building since we usually have a priori expectations about the signs and relative magnitudes of the parameter estimates based on theory, previous studies, behavioral rationale and intuition.

Signs of coefficients. The most essential test is to examine the signs of coefficients based on theory, behavioral rationale and intuition about the expected effect of the corresponding variables. For example, we expect that the coefficients of travel time and cost variables should have negative signs because travel time and cost have negative effects on utility of an alternative according to economic theory. Similarly, the coefficient of income variable associated to car driving should have positive sign.

The ratio of pairs of coefficients within a utility function. We normally have a priori information about the relative magnitude, at least a range, of coefficients within a utility function. For example, we expect the implicit value of time (VOT) to be within a range that can be supported by economic theory in a mode choice model. The model is questionable if the ratio is unreasonable (i.e., outside the expected range) according to economic theory. The VOT, which is one of the relative magnitudes of parameter estimates within a model, is one of the informal tests for evaluating the reasonableness of a travel demand model. Similarly, we have prior knowledge with respect to a reasonable range for the ratio of the parameters for different travel time components in public transport in the mode choice model. The VOTs are therefore normally used to check the reasonableness of a travel demand model.

Comparisons of coefficients of alternative specific variables across utility functions. Comparing the coefficients of alternative specific variables across utility functions is another informal test since we normally have expectations about the effect of characteristics of decision makers on different alternatives. Economic theory or intuition or behavioral rationale usually provides an indication about the differences in coefficients of alternative specific variables. For example, we expect a positive coefficient for the income of car driving, negative coefficient for public transport, positive coefficient but less than that of car driving for car passenger and so on. The differences are judged whether they are reasonable.

Statistical significance of coefficients

The statistical significance of a coefficient is the important test in model building. A model is questionable if most of the parameter estimates of the model are statistically insignificant although they carry expected signs and have reasonable relative values according to economic theory.

Likelihood ratio test

The likelihood ratio test (LRT) is a general test to test the models estimated using the maximum likelihood methods. The LRT serves the same function for the models estimated by the method of maximum likelihood as the F-test for the linear regression models estimated by the method of ordinary least squares. The LRT[24] is used to simultaneously test the values of several parameters of a model. It is widely used in discrete choice modeling because it tests different specifications of systematic utility functions given the model structure. It also guides which variables have to be included and how they enter the model (for example, generic vs. alternative specific, linear vs.

[24] There are two applications of the LRT. One is the specific application to test whether all the parameters are simultaneously zero, i.e., $\beta_1=\beta_2=\ldots=\beta_k=0$. The test statistic is $-2(L(0)-L(\beta))$ where $L(0)$ and $L(\beta)$ are the values the log likelihood function at 0 and maximum (i.e. at convergence) respectively. The test statistic, i.e., $-2(L(0)-L(\beta))$, is asymptotically distributed as χ^2 with degrees of freedom equal to the number of parameters in the model. If the test statistic exceeds the critical value, the null hypothesis is rejected.

Second is the general application and the LRT is defined as: LRT = -2(LLR-LLU) where LLR and LLU are the maximum values of the log-likelihood functions of the restricted and unrestricted models respectively. If the LRT exceeds the critical value of χ^2 with degrees of freedom equal to the number of restrictions, then the null hypothesis is rejected. The LRT is a very useful test in discrete choice modeling. It is used to test the different specifications, e.g., linear vs. non-linear specification of systematic utility function as well as testing nested vs. non-nested hypotheses, market segmentation and so on.

log-linear vs. nonlinear and so on) based on the fit of the model to the observed data because the inclusion of a variable may not always result in a significantly better fit of the model.

The likelihood ratio index, which compares the value of log likelihood function with its estimated parameters to the value of log likelihood function when all the parameters are simultaneously zero, is typically used goodness of fit measure in discrete choice models. The rho-squared (ρ^2)[25] and adjusted rho-squared ($\bar{\rho}^2$)[26] are two types of likelihood ratio indexes which measure how well the model fits the data. Though the ρ^2 and $\bar{\rho}^2$ are used in a similar way to the R^2 and $\bar{R}^2$ in regression analysis, the likelihood ratio indexes have limited use because they can only be used to compare the nested models. Their interpretations are also different[27]. Everything else being equal, the model with a higher value of the $\bar{\rho}^2$ is better.

Market segmentation tests

It is important to test whether the models are identical for different market segments of socioeconomic characteristics such as male and female or ownership of car or level of income or place of residence and so on. For the market segmentation test, the total sample is divided into different groups based on socioeconomic characteristics and same models are formulated and estimated for different market segments and the LRT is used. If we assume that market segmentation exits based on income, we divide the total sample into different groups according to income level, for example, high, medium and low income groups. We assume identical specification across the groups and estimate the models on the subset of data for each market segment. We also estimate the model with same specification on all the observations. Lastly, the LRT test[28] is used to check if the market segment exits across the different decision makers.

5.3 Test of the model structure

Testing for the IIA property is testing for the model structure of the logit model. Hausman and McFadden (1984) test, Small and Hsiao (1985) test and McFadden's (1987) omitted variable test are widely used to test the model structure. The first two tests check if the parameter estimates are significantly different after removing some alternatives from the choice set while

---

[25] $\rho^2 = 1 - \frac{L(\hat{\beta})}{L(0)}$ where $L(\hat{\beta})$ and $L(0)$ are the values of the log-likelihood functions at maximum and when all parameters are simultaneously zero respectively, and $0 \leq \rho^2 \leq 1$.

[26] $\bar{\rho}^2 = 1 - \frac{L(\hat{\beta}) - K}{L(0)}$ where K is the number of parameter estimates in the model, and $0 \leq \bar{\rho}^2 \leq 1$.

[27] R^2 measures the proportion of the total variation in dependent variable explained by the model (c.f. Gujarati, 2003). But ρ^2 has no equivalent interpretation (c.f. Train, 2003).

[28] If LR, LR_H, LR_M and LR_L denote the final log-likelihood values of models estimated on full data set, subsets of data sets of high income, medium income and low income decision makers respectively, the LRT = $-2[LR - (LR_H + LR_M + LR_L)] \sim \chi^2$ with degrees of freedom equal to number of restrictions. If the calculated LRT exceeds the critical value of χ^2 for the degree of freedom, the null hypothesis is rejected.

McFadden's omitted variable test[29] checks if the cross-alternative variables enter the model. If IIA is not valid, a better formulation of the systematic utility function is investigated and/or the alternatives with missing or miss-specified variables are searched or an acceptable nesting structure is investigated or more advance models are considered. However, the commonly used tests of IIA such as Hausman and McFadden, Small and Hsiao and McFadden's omitted variable tests are less demanding. If IIA fails, the less demanding tests provide a little guidance about further improvement of the model. A model structure can therefore be tested within a GEV or a mixed logit framework where the logit model is a special case given that the general framework is estimable.

5.4 Prediction tests

The outlier analysis and predicting market shares for different segments of a market are basically two types of tests related to prediction (c.f. Ben-Akiva and Lerman, 1985). In outlier analysis, the estimated models are applied to the sample to predict the choice probability of each alternative for each decision maker. If the predicted probability is the lowest for the chosen alternative, it has a large effect on log-likelihood function. The possible reasons might be the miss-specification of the model, coding or measurement errors in the data or unexplained variation in choice behavior. Market segmentation prediction test examines whether the predicted and observed market shares in a market segment are equal with full set of alternative specific constants (c.f., e.g., ibid). Large deviations call for investigation to improve the model formulation.

5.5 More on model building

Model building is not a trivial task. It requires the knowledge of theory, statistical methods and the judgment of the modeler. It is a systematic iterative process. We generate a set of 'reasonable models' based on the knowledge regarding theory and previous studies. Then we use 'goodness of fit', statistical tests and own judgment to select among the models. Since good fit does not necessarily imply a good model, we do not rely only on goodness of fit measures to select among competing models. A theory not only indicates the variables and how they enter the model, but

---

[29] First, the logit model is estimated using all the observations. Then systematic utilities (V_{in}) and choice probabilities $P_n(i|C_n)$ are estimated for all the alternatives and the decision makers using all the observations as follows:

$$\hat{V}_{in} = \hat{\beta}' x_{in} \qquad \hat{P}_n(i \mid C_n) = \frac{e^{\hat{\beta}' x_{i_n}}}{\sum_{j \in C_n} \hat{\beta}' x_{jn}}$$

Second, a new variables is created for a set of alternatives A_n belonging to C_n as follows:

$$\hat{V}_{A_n n} = \frac{\sum_{j \in A_n} \hat{V}_{jn} . \hat{P}_n(j \mid C_n)}{\sum_{j \in A_n} \hat{P}_n(j \mid C_n)}, \quad n = 1,\ldots,N$$

$$Z_{in}^{A_n} = \hat{V}_{in} - \hat{V}_{A_n n} \, if \; i \in A_n, 0 \; otherwise, n = 1,\ldots,N$$

Third, a model containing the original variables plus just created new variable, Z_{in}, is estimated. The coefficient of Z_{in} is looked at whether it is statistically significant. If it is significant, the IIA cannot be rejected.

also, perhaps more importantly, suggests the expected signs and relative magnitudes of the coefficients of the variables in the model. We subject a model to different tests including informal and statistical tests (c.f., section 5). Modeler's subjective judgment with respect to signs and relative values of the parameters are even more important.

Sometimes, the inclusion of a variable in a model may badly affect the results due to wrong signs, unreasonable relative magnitudes and statistical significance of the coefficients. Some of the implications of a variable in a model result are as follows:

- The inclusion of the variable does not improve the model fit significantly.
- The coefficient of the variable carries a wrong sign.
- The t-statistics is very low indicating that the variable does not have a significant impact.
- The inclusion of the variable results in wrong signs of other coefficients and unreasonable relative magnitudes of the coefficients in the model and so on.

If one of the above conditions holds, it is important to consider about including the variable in the model (if the variable is not a variable of particular interest in the analysis). There can be multiple reasons for such type of unexpected results. One of the reasons could be coding or measurement errors in the data. The possible solutions can be correction for errors in the variable (if possible) or removal of the observation or exclusion of the variable in an analysis if it is not the variable of particular interest in the study. If the variable, which cannot be corrected for errors, and which is not a variable of particular interest in an analysis, it might be better to exclude the variable in such situations. The same may apply to income. If the purpose of the model is, for example, to examine the impact of income on market shares of different travel modes, we, however, cannot exclude income from the model.

6 Model Validation[30]

A model must be validated before it is accepted to support decision making. Model validation is probably the most important part of the model building process because validation ensures that the model meets its intended requirements in terms of the results obtained and model's predictive capabilities. It is argued that that less attention is paid to validate a model.

We subject a model to statistical and practical/theoretical significance of model results including violation of assumptions of the model and predicting capability of the model discussed above. Statistical significance of the coefficients, the likelihood ratio test and goodness-of-fit measures such likelihood ratio indexes are the basic measures of statistical significance of a model. Other measures such as expected signs of the coefficients and their relative magnitudes examine the practical or theoretical significance of a model. We cannot validate a model unless the model is robust against structural change, model structure, prediction and policy tests[31] and so on.

The estimated models are applied to the sample that was not used in the estimation of the models and examined whether the models perform as intended. This is often called the external

[30] Since model validation itself is a very broad and important topic, the purpose in this section is not to review but rather to briefly introduce the topic relevant to the study in the paper.

[31] Tests against theoretical/statistical significance, structural change, model structure, and so on are briefly discussed in section 5.2 above.

validation of a model. The external validation of a model requires a validation sample which is a subsample of the data other than that was used in estimating the model or entirely independent data. The model is then estimated on this data. The external validation thus involves comparison of the model results including the forecasts on validation data against results and forecasts obtained from the original data. If the model results estimated on validation sample match the results obtained from original sample, then the model can be good enough for application.

The external validation also helps to generalize the results of a model in other places or at other times. External validation is the important step of model building to validate the results to the target population of the study. However, external validation is the least-practiced step of model building because it is time-consuming and expensive.

7. Applications of Choice Models

The discrete choice models are typically applied to investigate individual choice behavior, estimate demand for a product or service, willingness to pay, and impact of a policy change to people's welfare.

7.1 Investigating individual choice behavior

Discrete choice models are very important tools to investigate the choice behavior of individual decision makers such as persons, firms, households and government units. The choice models explore why an individual decision maker choose a particular product or service and what are the factors affecting the choice.

7.2 Demand estimation

The discrete choice models are demand models which estimate demand of discrete alternatives in terms of probabilities.

Disaggregate demand forecasting

The choice models are the demand models. P_{in} is the demand for alternative i of decision maker n. This is so called disaggregate or individual demand. The models are disaggregate in the sense that they are concerned with the decisions of individual units.

Aggregate demand forecasting

The disaggregate models can be used for aggregate forecasting purposes since the planners and the policy makers are not interested to an individual decision maker rather to some aggregate population. Estimated models are therefore frequently used to predict the impact of a policy change on market shares of different alternatives. We expect a priori that the increased fare of public transport, for example, reduces its users because demand of a good decreases if its price rises, ceteris paribus (c.f., e.g., Varian, 1992).

The disaggregate models are used in forecasting for some aggregate population of interest. The use of disaggregate models in aggregate forecasting can be prone to aggregation bias (c.f., e.g., Train, 1986 or 2003) if care is not taken. Aggregation bias in forecasting is due to nonlinearity of the model. Because the logit model is nonlinear, forecasting cannot be performed using only the averages of explanatory variables as in the linear regression models. Since the choice probabilities, derivatives and elasticities are nonlinear functions of the explanatory variables in the model, average value of the nonlinear function is not equal to the value of the function evaluated at the average of the explanatory variables. If we do, we will commit an error due to aggregation bias. Market segmentation and sample enumeration are the widely used approaches for aggregate forecasting using the disaggregate models[32].

In market segmentation, the population is divided into different groups representing the distinct values of explanatory variables. The proportion of decision makers and choice probabilities are estimated for each segment. Finally, the choice probabilities are weighted together with the segments share of the exposed population as weights.

In sample enumeration, population is represented by a sample. The proportion of decision makers choosing each alternative is estimated before and after a policy change using the estimated disaggregate model. The impact of the policy is therefore the difference between the proportions after and before the change. Sample enumeration is an approximation to the complete market segmentation approach. The approximation gets better as the sample size increases.

Elasticity

Elasticity is the important concept in economic theory (see, e.g., Varian, 1992; Silberberg and Suen, 2001). In general, elasticity is the ratio of the percentage change in one variable (typically dependent variable) to the percentage change in another variable (typically independent variable). It is the important outcome of a demand model. Elasticity in the context of demand models is defined as the responsiveness of demand to the change in factor/s affecting the demand. Modelers in travel demand are typically interested in two types of elasticities, namely, direct and cross elasticity. The former type measures the responsiveness of the choice probability of an alternative with respect to a change in one of its own attributes. The latter type on the other hand measures the responsiveness of the choice probability of an alternative with respect to a change in attributes of other competing alternatives.

In the context of discrete choice modeling, one can again distinguish between disaggregate (for an individual decision maker) and aggregate (for a group of decision makers) elasticities. Since a policy maker is mainly interested to know the impact of a policy to some group of decision makers, the aggregate elasticity is important from the policy perspective. The aggregate arc elasticity[33] is defined as (c.f. Train, 1986):

$$E = \frac{\Delta MS}{\Delta X}\frac{X}{MS} \tag{13}$$

where MS and X are the market shares and the variable of interest respectively. According to economic theory, the signs of direct and cross demand elasticities depend on the attributes of an alternative. For example, a priori the direct demand elasticity has negative sign and the cross demand elasticity has positive sign for the attributes such as travel time and travel cost of a travel

[32] See, for example, Ben-Akiva and Lerman (1985), chapter 6, for a brief overview of other aggregate forecasting techniques.
[33] The elasticity may be divided into arc and point elasticities. But I do not go into detail in the paper.

mode. The signs of the direct and cross elasticities will be exactly the opposite for the attributes such as frequency of public transport. Both direct and cross demand elasticities will be used in this study.

7.3 Willingness to pay

Estimating the willingness to pay (WTP) is one of the important applications of discrete choice models. WTP is closely related to the marginal rate of substitution which is the rate at which a decision maker is willing to exchange one good (or one attribute of an alternative) for another while maintaining the same level of utility (see, e.g., Varian, 1992; Silberberg and Suen, 2001).

A value of time (VOT) is one of the measures of WTP of a travel demand model where a traveler is willing to pay for the travel time reduction. It is the important outcome implied by the travel demand model. Given the utility function for a travel mode, the value of travel time savings is defined as the marginal rate of substitution between travel time and travel cost. VOT is thus the ratio of coefficients of travel time and travel cost variables[34]. VOT may depend on characteristics of the trip and the traveler. As stated before, the VOT can also serve as one of the informal tests for evaluating the reasonableness of a model.

7.4 Impact of a policy change to people's welfare

A policy maker is often interested to know the impact of a change in policy on welfare changes. For example, a reduction in waiting time for public transport can have impact on welfare changes. Consumer's surplus (CS) is a measure of welfare changes (c.f., e.g., Varian, 1992). Consumer welfare is directly related to expected utility. The logsum is the maximum expected utility from the available alternatives in the logit framework (c.f., e.g., Train, 2003). Logsum is, thus, a measure of consumer's surplus in the logit framework (ibid):

$$CS = \frac{1}{\mu}\left[\ln \sum_{j \in C_n} e^{\beta' x_{jn}}\right] \tag{14}$$

where μ converts utility into some monetary units and other symbols are already defined. The total CS in the population is then estimated as the weighed sum of the CS over a sample of decision makers. The change in CS due to a change in a policy is:

$$\begin{aligned}\Delta CS &= CS^1 - CS^0 \\ &= \frac{1}{\mu}\left[\ln \sum_{j \in C_n} e^{\beta' x_{in}^1} - \ln \sum_{j \in C_n} e^{\beta' x_{in}^0}\right]\end{aligned} \tag{15}$$

[34] A value of time (VOT) is the trade-off between travel time and travel cost. Given the utility function of travel mode, VOT is the ratio of coefficients of travel time and travel cost variables. In general,

$$VOT = \frac{\partial V_i}{\partial Time_i} \bigg/ \frac{\partial V_i}{\partial Cost_i}$$

where the subscripts 0 and 1 of CS and x refer the consumer surpluses and attributes before and after the change in the policy. The policy results in an increase in welfare if the change in the consumer surplus is positive.

8. Concluding Comments

Discrete choice models are the important analytical tools to investigate individual choice behavior. The choice models involve the behavioral process that leads to a decision maker's choice among discrete outcomes such as choice of travel mode or occupation. The decision makers and their characteristics, the alternatives and their attributes, and the decision rule are the typical elements in the choice process of discrete choice modeling. Random utility maximization is the theory underlying the discrete choice modeling where a decision maker is assumed to obtain utility from each alternative and chooses the alternative having the maximum utility among the available alternatives.

As we saw at the outset, the choice models are widely applied in diverse fields including transportation, economics, marketing, environment, and so on. The choice models are increasingly popular tools and consequently application area is expanding. For example, the choice models can be applied in marketing for choice of brand, menu, advertizing media (radio/TV/newspaper) and so on. Choice of marital status (whether a person wants to be single or married or living together), marriage partner, and so on are a few examples of choice in sociology. In business, choice of costumer, portfolio, securities, contract, and so on are the possible choice problems. Additionally, the choice models are successfully applied in political science, education, peace and conflict, environment, and so on.

The work presented in this paper synthesizes the fundamentals of discrete choice models. The paper also highlights the relationship between economic theory and discrete choice models: how economic theory contributes to choice modeling and vice versa. Economic theory lends theory (e.g. utility maximization), and methodology (e.g., formulation of systematic utility function, which variables and how enter the utility function, expected signs, relative magnitudes of coefficients) to discrete choice modeling. In return, the choice models provide the important analytical tool to investigate individual choice behavior and demand. The work also contrasts outcome of discrete choice theory to microeconomic consumer theory.

Acknowledgements

I am thankful to Odd I. Larsen, Otto A. Nielsen, Terje K. Tretvik and Harald M. Hjelle for their help and compliments on earlier draft of this work. Any errors and omissions are the sole responsibility of the author.

References

Aptech Systems, Inc., 2005. Maximum Likelihood Estimation for GAUSS User Guide, Maple Valley, WA.

Ashok, K., Dillon, W.R., Yuan, S., 2002. Extending discrete choice models to incorporate attitudinal and other latent variables. Journal of Marketing Research 39, 31-46

Barros, C.P., Proenca, I., 2005. Mixed logit estimation of radical Islamic terrorism in Europe and North America: a comparative study. Journal of Conflict Resolution 49, 298-314.

Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge, Mass.

Bhatta, B.P., 2009. Discrete Choice Models with Emphasis on Problems of Network Level of Service Attributes in Travel Demand Analysis. Ph.D. thesis, Molde University College, Molde, Norway

Cosslett, S., 1981. Efficient estimation of discrete choice models, in C. Manski and D. McFadden, eds., Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, MA, p. 51-111.

Daly, A., Bierlaire, M., 2005. A general and operational representation of GEV models. Transportation Research Part B: Methodological 40, 285-305.

DesJardins, S.L., Dundar, H., Hendel, D.D., 1999. Modeling the college application decision process in a Land-Grant University. Economics of Education Review. 18(1). p. 117-132.

Domencich, T., McFadden, D., 1975. Urban Travel Demand: A Behavioral Analysis. North-Holland Publishing Company, Amsterdam.

Glasgow, G., 2001. Mixed logit models for multiparty elections. Political Analysis 9, 116-136.

Gujarati, D.N., 2003. Basic Econometrics 4th edition. McGraw-Hill Higher Education, New York.

Hausman, J., McFadden, D., 1984. Specification tests for the multinomial logit model. Econometrica 52, 1219-1240.

Hensher, D. and Greene, W.H., 2003. The mixed logit model: the state of practice. Transportation 30, 133-176.

Hensher, D.A., Rose, J.M., Greene, W.H., 2005. Applied Choice Analysis: A Primer. Cambridge University Press, Cambridge, UK.

Hensher, D., Steward, J., Greene, W.H., 2007. An error component analysis of corporate bankruptcy and insolvency risk in Australia. Economic Record 83, 86-103.

Herriges, J.A., Phaneuf D.J., 2002. Inducing patterns of correlation and substitution in repeated logit models of recreation demand. American Journal of Agricultural Economics 84, 1076-1090.

Hess, S., 2005. Advanced Discrete Choice Models with Applications to Transport Demand. Ph.D. thesis. Centre for Transport Studies, Imperial College London, London.

Hess, S., Polak J.W., 2005. Accounting for random taste heterogeneity in airport-choice modelling. Transportation Research Record 1915, 36-43.

Hess, S., Polak, J.W, Daly, A., Hayman, G., 2007. Flexible substitution patterns in models of mode and time of day choice: new evidence from the UK and the Netherlands. Transportation 34, 213-238.

Horowitz, J.L., 1991. Reconsidering the multinomial probit model. Transportation Research Part B 25, 433-438.

Kohn, M., Manski, C., Mundel, D., 1976. An empirical investigation of factors influencing college going behavior. Annals of Economic and Social Measurement. 5, 391-419.

Lancaster, K., 1966. A new approach to consumer theory. Journal of Political Economy, 74, 132-157.

Lave, C.A. 1969. A behavioral approach to modal split forecasting. Transportation Research 3, 463-480.

Likert, R., 1932. A technique for the measurement of attitudes. Archives of Psychology 22, 1-55.

Lisco, T., 1967. The Value of Commuter's Travel Time: A Study in Urban Transportation. Ph.D. thesis. Department of Economics, University of Chicago, Chicago, III.

Luce, R., 1959. Individual Choice Behavior: A Theoretical Analysis. Wiley, New York.

Manski, C., McFadden, D., 1981. Alternative estimators and sample designs for discrete choice analysis, in Manski, C., and McFadden, D., (eds.), Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, MA, p. 2-50.

Manski, C., 1977. The structure of random utility models. Theory and Decision 8, 229-254.

Marschak, J., 1960. Binary choice constrains on random utility indications, in Arrow, K. (ed.), Stanford Symposium on Mathematical Methods in the Social Sciences, Stanford University Press, Stanford, CA, 312-329.

McFadden, D., 1974a. The measurement of urban travel demand. Journal of Public Economics 3, 303-328.

McFadden, D., 1974b. Conditional logit analysis of qualitative choice behavior. in Zarembka, P., (ed.). Frontiers in Econometrics. Academic Press, New York, chapter 4, 105-142.

McFadden, D., 1978. Modeling the choice of residential location. In A. Karlquist (ed.). Spatial Interaction Theory and Planning Models, North Holland, Amsterdam, Chapter 5, 198-272.

McFadden, D., 1987. Regression-based specification tests for the multinomial logit model. Journal of Econometrics 34, 63-82.

McFadden, D., 2000. Disaggregate behavioral travel demand's RUM side: a 30-year perspective. Paper presented at Conference on the International Association of Travel Behavior Research, Brisbane, July 2, 2000, Australia.

McFadden, D., 2001. Economic choices. The American Economic Review 91, 351-371.

McFadden, D., Train, K., 2000. The mixed MNL models for discrete response. Journal of Applied Econometrics 15, 447-470.

Ortúzar, J. de D., 2001. On the development of the nested logit model. Transportation Research Part B: Methodological 35, 213-216.

Ortúzar, J. de D., Willumsen, L.G., 2001. Modeling Transport 3rd edition. John Wiley & Sons Ltd, England.

Quarmby, D.A., 1967. Choice of travel mode for the journey to work: some findings. Journal of Transport Economics and Policy 1, 273-314.

Small, K., Hsiao, C., 1985. Multinomial logit specification tests. International Economic Review 26, 619-627.

Silberberg, E., Suen, W., 2001. The Structure of Economics: A Mathematical Analysis 3rd edition. McGraw-Hill Higher Education, Singapore.

Thurstone, L.L., 1927. The law of comparative judgment. Psychology Review 34, 273-286.

Train, K., 1986. Qualitative Choice Analysis: Theory, Econometrics and Application to Automobile Demand. MIT Press, Cambridge, Mass.

Train, K., 2003. Discrete Choice Methods with Simulation Cambridge. University Press, Cambridge, UK.

Train, K., McFadden, D., Ben-Akiva, M., 1987. The demand for local telephone service: a fully discrete model of residential calling patterns and service choices. Rand Journal of Economics 18, 109-123.

Varian, H.R., 1992. Microeconomic Analysis, 3rd edition. W.W. Norton & Company, New York.

Warner, S.L. 1962. Stochastic Choice of Mode in Urban Travel: A Binary Choice. Northwestern University Press, Evanston, II.

Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2003. Statistical and Econometric Methods for Transportation Data Analysis, Chapman & Hall/CRC, New York.

Weeks, M., 1997. The multinomial probit model revisited: A discussion of parameters estimability, identification and specification test. Journal of Economics Surveys 11, 297-320.