Chapter 7

Introduction to Discrete Choice Models

7.1 Introduction

It has been mentioned that the conventional selection bias model requires estimation of two structural models, namely the selection model where the probability of participation for each person needs to be calculated and the primary model where the relationship between individuals' outcomes and their attributes is sought to be evaluated. In terms of the selection model, when decisions are represented as a discrete (binary) variable –1 for participation and 0 for non-participation– the corresponding probabilities can be estimated by Logit or Probit models. However, in most practical situations, an individual may be faced with more than two decisions, expressed as alternative choices. In this case, the above models cannot provide consistent answers of the probability of selecting a specific alternative k over a set of M alternatives. A useful generalization that accounts for this situation is the *Multinomial Discrete Choice Models*.

The origins of discrete choice models are rooted in the early studies of psychophysics (the physical study of the relations between physical stimuli and sensory response) at 1860. These models were later applied in biology with the expression *models with discrete responses*. Today they frequently occur in product market demand theory, in labor econometrics, in economics as well as in other social sciences.

Suppose that an individual has to choose among K mutually exclusive alternatives of a product or of a service. The *neoclassical economists' approach* to this problem is that every individual has a utility function, which allows him to rank the alternatives in a consistent and unambiguous manner. The individual then chooses the alternative that is ranked first. Therefore, from the economists' point of view, the choice problem is a problem of maximization of a utility function while the choice process is deterministic since each unit just selects the alternative that maximizes its utility. These models have introduced, initially, by Thurstone (1927) and latter by McFadden (1973, 1974a, 1975, 1976, 1981) and McFadden and Reid (1975) and termed *Random Utility Models (RUM)*.

An important feature in modern discrete choice modeling that is first implied by Luce (1959) is the *Independence of Irrelevant Alternatives (IIA) property*. According to this for all sets of alternatives $S \subseteq A$ and $T \subseteq A$ such that $S \subseteq T$ and for all alternatives a and b such as $a \in S$ and $b \in S$

$$\frac{P_S(a)}{P_S(b)} = \frac{P_T(a)}{P_T(b)}. \tag{7.1}$$

This expression implies that the ratio of probabilities of choosing any two alternatives is independent of the attributes of all other alternatives. Moreover, the IIA property implies that the ratio of the probabilities of choosing any two alternatives is independent of the availability of a third alternative.

However, the IIA assumption is restrictive in many applications. For example, it is unlikely that the odds ratio of any two choices would be invariant to the introduction of a new alternative that is a close substitute for an existing alternative. A typical example of this restriction is the ''blue bus/red bus paradox'', fist pointed out by Debreu (1960) and discussed again by Anderson, de Palma and Thisse (1992). Assuming that the probability for two alternatives to have equal utility values is zero (that is there are no ties), we write:

$$u_{ik} = V_{ik} + \varepsilon_{ik} = \beta\,\mathrm{X}_{ik} + \varepsilon_{ik} \tag{7.2}$$

where X_{ik} is the $(M\text{-}1) \times J$ vector of the attributes of the k^{th} alternative $(k= 1,\ldots,M)$.

β is a $J \times 1$ vector of parameters.

V_{ik} is a function of the observed utilities.

ε_{ik} is a $J \times 1$ vector of residuals that capture unobserved variations in tastes of the individuals, in attributes of the alternatives as well as errors in the perception and optimisation by the consumers.

Note that u_{ik} can be thought of as the set of utilities for individual i, given that the utility of the last alternative is zero, or in other words, the utilities of the first $(M-1)$ alternatives (the value that is formed for these utilities) with respect to the last utility if it is fixed to the constant value 0. This specification is needed in order to avoid the identification problem, indicated by McCullogh and Rossi (1994).

The aim is the estimation of vector β, of the unobserved individual utilities u_{ik} and of the probability that the individual i will choose the alternative s. In this way, the relative frequency of choosing alternative s over a population of individuals can be estimated.

The probability that the i^{th} individual will choose the s^{th} alternative is:

$$\begin{aligned} P_{ij} &= P\left(V_{is} + \varepsilon_{is} \succ V_{ik} + \varepsilon_{ik}\right) \\ &= P\left(\varepsilon_{ik} - \varepsilon_{is} \prec V_{is} - V_{ik}\right) \end{aligned} \tag{7.3}$$

and generally,

$$P_{is} = P\left\{V_{is} + \varepsilon_{is} \succ \max_{k=1,2,\ldots M-1} (V_{ik} + \varepsilon_{ik})\right\} \quad \mathrm{i} = 1,2,\ldots\mathrm{N}$$

Defining $F(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M)$ as the joint cumulative distribution function over the values ε_{ik}, $(k = 1, 2, \ldots, M)$, and F_{ik} as the partial derivative of F with respect to its k^{th} argument, the probability P_{ik} can be written:

$$P_{ik} = \int_{\varepsilon=-\infty}^{+\infty} F_k\left(\varepsilon_{ik} + V_{ik} - V_{i1}, \varepsilon_{ik} + V_{ik} - V_{i2}, \ldots, \varepsilon_{ik} + V_{ik} - V_{ip}\right) d\varepsilon \tag{7.4}$$

Different specifications of F yield different families of discrete choice models. The most known ones are the *Multinomial Logit (MNL)*, the *Nested Multinomial Logit (NMNL)* and the *Multinomial Probit (MNP)*. A presentation of these models is provided in Theil (1969), McFadden (1978) and Manski and McFadden (1981).

7.2 Multinomial Logit Model (MNL)

An econometric choice model is specified by choosing a parametric form for V_{ik}, *$(k = 1, 2, ...M)$*, and a parametric distribution F. An important specialization of (7.4) is the Multinomial Logit model (MNL) with choice probabilities:

$$P_{ik} = \frac{exp(u_{ik})}{\sum_{j=1}^{M} exp(u_{ik})} \tag{7.5}$$

where $u_{ik} = V_{ik} + \varepsilon_{ik}$ is the unknown utility or value of interest for choice alternative k and for individual i $(i = 1, 2, ..., N)$.

V_{ik} is the systematic observable component or mean utility value of choice alternative k and for individual i.

ε_{ik} is the random error component associated with choice alternative k and individual i.

The MNL or *conditional Logit model* is obtained by assuming that ε_{ij} are independently and identically distributed (iid) with the *Extreme Value Type-I* distribution[1] (see McFadden, 1973):

$$P\left(\varepsilon_{ik} \leq \varepsilon\right) = exp\left(-e_{ik}\right)$$

However, the assumption of independent ε_{ij} yields counterintuitive forecasts for alternative sets containing choices that are close substitutes. Such an example is the "red bus / blue bus" example. In this case the model is inappropriate because the IIA axiom is implausible.

McFadden (1973) also discusses the Maximum Likelihood Estimator (MLE) of the regression parameter of interest β for the MNL. Hausman (1978), Hausman and

[1] The name Extreme Value comes from the interest to find the maximum of a series of random variables (their extreme value).

McFadden (1984) and McFadden (1987) provide some specification tests for this model.

McFadden (1978, 1981) describes a useful generalization of the MNL model and a way to relax the restrictive IIA assumption, namely the Nested Multinomial Logit model (NMNL) that uses a nested structure to estimate the probability of choosing a specific alternative. For a detailed presentation of the NMNL model the reader is referred to Maddala (1983) and Anderson et. al. (1992).

Another, more general type of MNL models that also relaxes IIA assumption is the *Mixed Multinomial Logit model (MMNL)* introduced by Boyd and Mellman (1980). This model allows for another source of variation to be present, apart from random disturbance. McFadden and Train (2000) establish by a theorem that MNL mixtures can closely approximate a very broad class of Random Utility models that have zero probability of ties. Estimation methods of MMNL are referred in Hausman and Wise (1978), Manski and McFadden (1981), McFadden (1989) and Hajivassiliou and McFadden (1997).

Another common approach is that of *δ-MNL model*. Georgescu-Roegen (1958) describe the problem of thresholds in stochastic consumer choices. Suppose that an individual is faced with two alternatives with approximately the same attributes. As a result their utility values are *approximately* the same. Given that the individual is able to recognize and respond to differences in utilities however small they may be, they relax utility maximization assumption, with the assumption that an individual i selects s alternative over k if and only if their difference in utilities is greater than a threshold δ; else the individual is indifferent between the choices. Lioukas (1984) also describes the problem of thresholds in stochastic consumer choice and approaches it by applying an MNL model to real data a three alternatives choice situation. He calls this model δ – MNL model.

7.3 Multinomial Probit Model

An alternative family of Discrete Choice Models is achieved by assuming different distribution for the joint cumulative distribution function $F(\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iM})$. More specifically, by assuming that the random components ε_{ik} are jointly Normally distributed with corresponding density:

$$f(\varepsilon_{ik}) = (2\pi)^{-M/2} \times |\Sigma|^{-1/2} \times exp\left[-\frac{1}{2}\varepsilon'_{ik}\ \Sigma\ \varepsilon_{\text{ik}}\right]$$

the Multinomial Probit Model is derived. Here M is the number of alternatives and the variance – covariance matrix Σ is normalized:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{21} & \cdots & \sigma_{M1} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{M2} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{M1} & \sigma_{M2} & \cdots & 1 \end{bmatrix}$$

to avoid the identification problem.

Despite the comparative advantage of MNP against MNL model in relaxing the IIA assumption the former cannot be represented in a closed form as MNL model (equation 7.5). Thus, the calculation of its choice probabilities requires evaluation of high dimensional numerical integration, specifically when the number of alternatives is higher than four. The most common methods to avoid such evaluation are the simulation methods. Specifically, the *Clark method* of Clark (1961), the *Method of Simulated Moments (MSM)* discussed in Lerman and Manski (1981), McFadden (1989), McFadden and Ruud (1994) and Geweke et. al. (1994), the *Simulated Maximum Likelihood estimation procedure (SML)* proposed by Manski and Lerman (1981) and the *method of Simulated Scores (MSS)* proposed by McCulloch and Rossi (1994), Hajivassiliou (1996) and Hajivassiliou and McFadden (1997) are some methods for the estimation of the parameters and the probabilities of the selection model.

A detailed description of this model is found in Hausman and Wise (1978), Geweke et. al. (1994) and Maddala (1983). Currim (1982) is referred in two different forms of MNP models, namely the *Independent Probit* and the *Generalized Covariance Probit* model. These models are obtained by properly reformulating the variance-covariance matrix Σ of the random errors ε_{ik}.