I THEORY AND ECONOMETRICS OF QUALITATIVE CHOICE MODELS

1 Qualitative Choice Models in General

1.1 Motivation

In recent years the emphasis in econometrics has shifted from aggregate models that describe markets as a whole to disaggregate models of the individual decisionmaking units that underlie market demand and supply. There are several reasons for this shift. First, economically relevant behavior is necessarily at the individual level: market supply and demand are simply the aggregate of many individuals' actions. Consequently, attempts to capture the structure, or causal relations, inherent in behavior are more naturally pursued at the individual level. Microeconomic theory provides a way of looking at the actions of individual decisionmaking units, as well as a rich set of hypotheses concerning these actions. This theory can be drawn upon in specifying and interpreting disaggregate econometric models to a degree that is not possible with aggregate models. Second, survey data on households and individual firms are becoming more and more available, making it possible to estimate disaggregate models in situations that would previously have been impossible to examine at the individual level. Furthermore, with these data on individual decisionmaking units, more precise estimation of underlying parameters is possible. Data on individual units necessarily contain greater variation in each factor, and usually less covariation among factors, than aggregate data, simply because the latter are sums or averages of the former. This fact is important in estimating econometric models since the precision with which each parameter in a model can be estimated generally increases with the variance of the variable entering the model and decreases with the covariance among variables. As a result, disaggregate models are often able to capture effects that cannot be incorporated accurately in aggregate models.

With the movement toward disaggregate modeling came a need for new methodologies. Standard econometric methods like regression were designed for analyzing variables that can assume any value within a range, that is, for continuous variables. These methods are usually appropriate for examining aggregate data. When the underlying behavior of the individual decisionmaking units is examined, however, it is often found that the outcome of the behavior is not continuous and standard regression procedures are inappropriate. Automobile demand is a case in point. The aggregate demand for a particular make and model of automobile, say, Honda Accord, can be considered a continuous variable that varies over time and geographic regions. But the demand for Honda Accord by any one household is not continuous; at any given set of prices and other factors, each household either buys a Honda Accord or not. Examining this demand with methods developed for continuous variables ignores the structure of the behavioral situation.

A variety of methods have been developed for examining the behavior of individuals when continuous methods are inappropriate. Qualitative choice analysis is among these. It is designed for describing decisionmakers' choices in certain types of situations. These situations arise in a variety of contexts in such areas as transportation, energy, telecommunications, housing, criminology, and labor, to name a few. However, just as regression is inappropriate in some contexts, qualitative choice analysis is applicable in only particular types of situations. Defining these situations is the topic of the next section.

1.2 Situations Described by Qualitative Choice Models

In any choice situation, the person making the choice has two or more different items, courses of action, or, more generally, "alternatives" among which to choose. A qualitative choice situation, which qualitative choice models are used to describe, is defined as one in which a decisionmaker faces a choice among a set of alternatives meeting the following criteria: (1) the number of alternatives in the set is **finite**; (2) the alternatives are **mutually exclusive**: that is, the person's choosing one alternative in the set necessarily implies that the person does not choose another alternative; and (3) the set of alternatives is **exhaustive**: that is, all possible alternatives are included, and so the person necessarily chooses one alternative from the set. Examples of choice situations fitting these criteria are a worker's choice of mode for travel to work (with the alternatives being auto, transit, walk, etc.); a household's choice of make and model of automobile (VW Rabbit, Olds Omega, etc.); a household's choice of type of convection oven (electric versus gas); and a businessperson's choice of long distance telecommunication service (Allnet, Sprint, AT&T, etc.). Examples of choice situations that do not fit these criteria, and hence are not qualitative choice situations, are a consumer's choice of how many pounds of beef to buy at the store (since the set of alternative weights is, at least theoretically, infinite: the person can buy 1 pound, 1.1 pounds, 1.11 pounds, 1.111 pounds, etc.); a businessperson's choice between two different life insurance policies being offered by a salesperson (since the set is not exhaustive: the person could choose a policy offered by another salesperson); and a worker's choice of mode of travel when there is the possibility of driving to the bus stop and then taking the bus (since the alternatives of auto, bus, walk, etc., are not mutually exclusive: the person could take both auto and bus).

These examples of situations that are not qualitative choice situations suggest an important point. When a choice situation cannot be described as "qualitative" because either its alternatives are not mutually exclusive or its set of alternatives is not exhaustive, it is usually possible to redefine the set of alternatives in such a way that the redefined set meets all three criteria. Consequently, situations that might at first appear not to meet the criteria, can qualify with a redefinition of alternatives. For example, the businessperson's choice between two life insurance policies offered by a particular salesperson does not seem applicable because the alternatives are not exhaustive; however, if a third alternative is added, namely, the possibility of choosing neither of the two policies, then the set of three alternatives (choosing either of the two policies or neither) is exhaustive, and the choice situation is a qualitative one. Similarly, the worker's alternative modes of auto, bus, walk, etc., are not mutually exclusive if it is possible that the worker drives to the bus stop and then takes the bus; however, if the set of alternative modes is redefined to be auto only, bus only, bus with auto access to bus, etc., then the alternatives are mutually exclusive and the worker is in a qualitative choice situation.

Since careful delineation of alternatives can usually assure that the second and third criteria are met, the only truly restrictive criterion is the first one, namely, that the number of alternatives be finite. A distinction established by this criterion is that between continuous and discrete variables. The set of alternatives available to the decisionmaker can be denoted by a variable; for example, the amount of beef a consumer chooses can be denoted by x, where x is any nonnegative number; the type of convection oven available to a household can be denoted as y, with y taking either of two values: either one for electric or two for gas. The variable x is obviously continuous in that, within any range, it can take an infinite number of values, whereas the variable y discrete. Any choice situation in which the set of alternatives can be denoted by a continuous variable is not a qualitative choice situation.

The term "discrete choice situation," which is often used to denote the same thing as "qualitative choice situation," arose from this distinction between continuous and discrete variables for denoting the set of alternatives. This distinction also gives meaning to the use of the term "qualitative": most choices that concern how many or how much of something (which are choice of quantity) have alternative sets that are denoted by continuous variables, whereas choices of which item (such as which type of oven) are nonquantitative, or qualitative, and have alternative sets that can be denoted by discrete variables.

Unfortunately, both of these terms can be misleading. "Qualitative choice situation" would seem to exclude choices regarding quantity. However, some choice situations are quantitative in that the choice is in regard to "how many" yet nevertheless qualify as qualitative choice situations. An example is the choice of how many autos to own; if it can be assumed that a household cannot own more than some N autos, then the set of alternatives is $0, 1, 2, \ldots, N$, which is clearly a finite, exhaustive set of mutually exclusive alternatives.

The difficulty with the term "discrete choice situation" is that, in practice, the distinction between discrete and continuous variables is not always meaningful and hence does not serve to guide the researcher in determining whether to use qualitative choice methods. Many continuous variables can be represented, without loss of accuracy and sometimes with an increase in accuracy, by discrete variables. For example, the amount of beef a consumer buys is a continuous variable, since weight is a continuous variable. However, in practice the variable can be treated as discrete, since a consumer never asks for, nor gets charged for, fractional quantities such as 1.111 pounds; the number is always rounded off to some convenient figure. Similarly, since it is not possible to pay fractions of cents, one's expenditures on any good, while often treated as continuous, are actually discrete, with the set of alternatives being every one-cent interval above zero. If there is some conceivable maximum for these discrete variables, then the number of alternatives is finite and the choice situations are qualitative.

Whether or not to utilize qualitative choice methods in situations such as these is a matter of taste, or, more precisely, a strategy decision by the researcher. Usually, choices of "how many" or "how much" are more fruitfully analyzed with qualitative choice methods if the number of alternatives is fairly small. When there are a large number of quantitative alternatives such that the discrete dependent variable is essentially indistinguishable from a continuous one, then standard econometric methods for continuous variables, such as regression, can be used adequately to present the choice. Thus, qualitative choice methods are appropriate for analyzing households' choices of how many cars to own, since only a small number of alternatives are involved, whereas regression techniques are probably more appropriate for examining households' expenditures on particular goods, since, even though the alternatives are truly discrete, there are enough quantitative alternatives to be adequately represented by a continuous variable.

In short, qualitative choice models are used to analyze situations in which a decisionmaker can be described as facing a choice among a finite and exhaustive set of mutually exclusive alternatives. Furthermore, if the choice is one of how many or how much and nevertheless meets these criteria, then qualitative choice models are used only if the number of alternatives is fairly small.

1.3 Specification

The general term "qualitative choice models" designates a class of models; specific qualitative choice models, such as logit and probit, are members of this class. This section presents the specification of the general class and the manner in which specific models within the class are distinguished.

All qualitative choice models calculate the probability that a decisionmaker will choose a particular alternative from a set of alternatives, given data observed by the researcher. The models differ in the functional form that relates the observed data to the probability. These concepts are now elaborated.

Denote the decisionmaker in a qualitative choice situation by n and the set of alternatives he faces by J_n. This set, sometimes called the choice set, is subscripted by n to represent the fact that different decisionmakers might face different sets of alternatives in similar choice situations. For example, in workers' choices of mode of travel, a person who cannot drive does not have the alternative of taking an auto, which is available to most workers.

The alternatives that the decisionmaker faces differ in their characteristics (otherwise there would essentially be only one alternative and no choice). For example, the time and cost of travel by auto is different than by bus; purchase price, fuel efficiency, interior space, and so on vary across makes and models of cars; monthly and per minute charges vary over long distance telecommunication services. Some of these characteristics are observed by the researcher and some are not.

Label the **observed** characteristics of alternative i as faced by decisionmaker n as the vector z_{in}, for all i in J_n. Note that the characteristics of each alternative are subscripted by n to reflect the fact that different decisionmakers can face alternatives with different characteristics. For example, the cost of travel to work by auto will vary across workers depending on the distance from their homes to work.

The decisionmaker's choice of alternative obviously depends on the characteristics of each of the available alternatives. Different decisionmakers, however, can make different choices when facing the same alternatives because the relative value that they place on each characteristic is different. The differences in the valuation of each characteristic of the alternatives depend on the characteristics of the decisionmaker, both those observed by the researcher and those not observed. Label the **observed** characteristics of decisionmaker n as s_n. Usually elements of s_n are income, age, education level, and so on.

The probability that decisionmaker n chooses alternative i from set J_n (labeled P_{in}) depends on the observed characteristics of alternative i compared with all other alternatives (i.e., on z_{in} relative to all z_{jn} for j in $J_n, j \neq i$) and on the observed characteristics of the decisionmaker (s_n). Qualitative choice models specify this probability as a parametric function of the general form

$$P_{in} = f(z_{in}, z_{jn} \text{ for all } j \text{ in } J_n \text{ and } j \neq i, s_n, \beta), \qquad (1.1)$$

where f is the function that relates the observed data to the choice probabilities. This function is specified up to some vector of parameters, β. In specific contexts, these parameters will usually have a particular meaning. For example, in the choice of mode for the commute to work, the parameters might represent the relative importance, or value, of travel cost and time in the workers' choices. These parameters are usually estimated by the researcher in the manner described in the chapters to follow, but can conceivably be determined on the basis of a priori information or expert judgment.

In some sense, the general description of qualitative choice models is completely contained in equation (1.1). All qualitative choice models have this general form. Specific qualitative choice models, such as logit or probit, are obtained by specifying f.

It is useful, however, to elaborate upon this general description with concepts from the standard microeconomic theory of utility maximization. Relating the general specification of qualitative choice models to utility theory offers three benefits. First, in the discussion thus far, the meaning of the choice probability P_{in} is not clear. Usually a probability is interpreted as the proportion of times a particular event would occur if a situation were repeated numerous times (or, more accurately, the fraction to which this proportion converges as the number of repetitions increases without bound). In the previous discussion, the source of repetitions is not delineated. A clear meaning of the choice probabilities emerges from the derivation of probabilities from utility theory. Second, since different qualitative choice models are obtained through different specifications of the function f, the motivation for different specifications and the manner in which they are obtained is critical information for a researcher. Utility theory provides a context for motivating and deriving various specifications of f. Third, and largely because of the first two reasons, the literature on qualitative choice models uses terms that only have meaning in the context of utility theory. Understanding this literature requires knowledge of the relation of qualitative choice models to utility maximization.

The derivation of qualitative choice models from utility theory is based on a precise distinction between the behavior of the decisionmaker and the analysis of the researcher. Consider first the decisionmaker. As stated, decisionmaker n has a choice among the alternatives in set J_n. The decisionmaker would obtain some relative happiness or "utility" from each alternative if he were to choose it. Designate the utility from alternative i in J_n as U_{in}, and similarly for each other alternative in J_n. This utility depends on various factors, including the characteristics of the alternative and the characteristics of the decisionmaker. For example, when a person chooses which automobile to buy, the utility of each alternative make and model depends on the price, fuel efficiency, seating capacity, and other characteristics of the vehicle as well as the income, number of dependents, and other characteristics of the decisionmaker. Label the vector of all relevant characteristics of alternative i as faced by person n as x_{in} and the vector of all relevant characteristics of person n as r_n. Since x_{in} and r_n include all relevant factors, we can write utility as a function of these factors,

$$U_{in} = U(x_{in}, r_n), \qquad \text{for all} \quad i \text{ in } J_n, \tag{1.2}$$

where U is a function.

The decisionmaker chooses, of course, the alternative from which he derives the greatest utility. That is, the decisionmaker chooses alternative i in J_n if and only if[1]

$$U_{in} > U_{jn}, \quad \text{for all} \quad j \text{ in } J_n, \quad j \neq i.$$

Substituting (1.2), we have

$$n \text{ chooses } i \text{ in } J_n \quad \text{iff} \quad U(x_{in}, r_n) > U(x_{jn}, r_n), \quad \text{for all} \quad j \text{ in } J_n, \quad j \neq i. \tag{1.3}$$

This completes the specification of how the decisionmaker behaves. Note that the decisionmaker's choice is deterministic: he chooses the alternative that provides the highest utility. If one were to define at this point the probability that person n would choose alternative i, then the probability would necessarily be either one or zero depending on whether or not alternative i provided the greatest utility.

To specify the choice probabilities, we focus on the researcher. Suppose that a researcher is interested in predicting this decisionmaker's choice. If the researcher observed all the relevant factors, i.e., x_{in} for all i in J_n and r_n, and knew the decisionmaker's utility function U, then the researcher could use relation (1.3) perfectly to predict the decisionmaker's choice. However, the researcher does not observe all the relevant factors and does not know the utility function exactly.

Partition the elements of x_{in} into two subvectors: those characteristics of the alternative that are observed by the researcher, denoted by vector z_{in}, and those that are not (not labeled). Similarly, partition r_n into observed characteristics of the person, labeled s_n, and characteristics that are not observed by the researcher. Finally, decompose $U(x_{in}, r_n)$ for each i in J_n into two subfunctions, one that depends only on factors that the researcher observes and whose form is known by the researcher up to a vector of parameters, β, to be estimated, with this component labeled $V(z_{in}, s_n, \beta)$, and another that represents all factors and aspects of utility that are unknown by the researcher, which is labeled e_{in}. That is,

$$U_{in} = U(x_{in}, r_n) = V(z_{in}, s_n, \beta) + e_{in}. \tag{1.4}$$

Note that this equation holds exactly, since e_{in} is simply the difference between true utility $U(x_{in}, r_n)$ and the part of utility the researcher knows, $V(z_{in}, s_n, \beta)$.

Since the researcher does not know $U(x_{in}, r_n)$ entirely, he cannot perfectly predict the decisionmaker's choice. However, the researcher knows part of the decisionmaker's utility, namely, the part denoted by $V(z_{in}, s_n, \beta)$, and with this information is able to make educated guesses as to the decisionmaker's choice. In particular, the researcher can (with one additional piece of information to be described) state the probability that the decisionmaker will choose each alternative.

Let us now define the choice probabilities. Suppose a researcher observed a group of decisionmakers all of whom faced the same alternatives with the same values for the **observed** portion of utility for each alternative. The unobserved part of each decisionmaker's utility is, by definition, not known by the researcher and, in general, will vary across decisionmakers in the group. Therefore, even though the observed part of utility is the same for all decisionmakers in the group, different decisionmakers would choose different alternatives depending on the values of the unobserved components of their utility. For example, if e_{in} is much larger than e_{im} for two decisionmakers, n and m, then person n might choose alternative i while person m does not. Within any group, a certain proportion of the decisionmakers will choose alternative i. The probability of choosing i is the fraction to which this probability converges as the size of the group increases without bound.

The choice probabilities can now be defined precisely. The probability that person n chooses alternative i, denoted P_{in}, is the limit of the proportion of times, as the number of times increases without bound, that the researcher would observe a decisionmaker who faces the same alternatives as person n, and with the same values of observed utility for each alternative, to choose alternative i. Note that this probability is defined on the researcher, reflecting the researcher's lack of information regarding all factors affecting the decisionmaker's choice. It is not defined on the decisionmaker, whose choice is deterministic.[2]

Given the definition of the choice probabilities, the function f that relates observed data to the probabilities is derived as follows. P_{in} is the probability that relation (1.3) holds, that is, that the utility of alternative i is higher than that of any other alternative, given the observed components of utility for each alternative.

$$P_{in} = \text{Prob}(U_{in} > U_{jn}, \text{ for all } j \text{ in } J_n, j \neq i). \tag{1.5}$$

By substitution of (1.4) and letting V_{in} denote $V(z_{in}, s_n, b)$ for notational simplicity,

$$P_{in} = \text{Prob}(V_{in} + e_{in} > V_{jn} + e_{jn}, \text{ for all } j \text{ in } J_n, j \neq i).$$

Rearranging,

$$P_{in} = \text{Prob}(e_{jn} - e_{in} < V_{in} - V_{jn}, \text{ for all } j \text{ in } J_n, j \neq i). \tag{1.6}$$

Let us examine the right-hand side of this equation. The researcher observes V_{in} and V_{jn}, and so he can calculate their difference, $V_{in} - V_{jn}$. The researcher does not observe e_{jn} or e_{in}; as explained, these terms are random, varying across decisionmakers with the same observed components of utility. Since e_{jn} and e_{in} are random variables, their difference $e_{jn} - e_{in}$ is also a random variable. Consequently, the right-hand side of (1.6) is simply a cumulative distribution: the probability that the random variable $e_{jn} - e_{in}$ is below the known value $V_{in} - V_{jn}$. More precisely, it is a joint cumulative distribution, namely, the probability that each random variable $e_{jn} - e_{in}$ is below $V_{in} - V_{jn}$, respectively, for all j in $J_n, j \neq i$.

By knowing the distribution of the random e's (though not knowing their particular values), the researcher can derive the distribution of each difference $e_{jn} - e_{in}$ for all j in $J_n, j \neq i$, and by using equation (1.6) calculate the probability that the decisionmaker will choose alternative i as a function of $V_{in} - V_{jn}$ for all j in $J_n, j \neq i$. This function is f in equation (1.1).

All qualitative choice models are obtained by specifying some distribution for the unknown component of utility and deriving functions for the choice probabilities. Different qualitative choice models are obtained by specifying different distributions for the e's, giving rise to different functional forms for the choice probabilities.

The meaning of equation (1.6) can be visualized if we restrict ourselves to examples with only two alternatives. Suppose the decisionmaker has a choice between alternatives i and j and that the observed component of utility for alternative i is one unit greater than that for alternative j. (Say, $V_{in} = 4$ and $V_{jn} = 3$, so that $V_{in} - V_{jn} = 1$.) Alternative i will be chosen if total utility, both observed and unobserved, is higher for alternative i than for alternative j, that is, if $V_{in} + e_{in} > V_{jn} + e_{jn}$. The values that e_{in} and e_{jn} take determine whether this occurs. If e_{in} exceeds e_{jn}, then alternative i will obviously be chosen, since both the observed and unobserved parts of its utility are greater than those for alternative j. However, even if instead e_{jn} exceeds e_{in}, alternative i will still be chosen as long as e_{jn} is no more than one unit greater than e_{in}. Alternative i will **not** be chosen (and alternative j will be) only if e_{jn} is more than one unit larger than e_{in}.

Equation (1.6) states these facts succinctly:

$$P_{in} = \text{Prob}(e_{jn} - e_{in} < V_{in} - V_{jn}),$$

which in this case, since $V_{in} - V_{jn} = 1$, becomes

$$P_{in} = \text{Prob}(e_{jn} - e_{in} < 1).$$

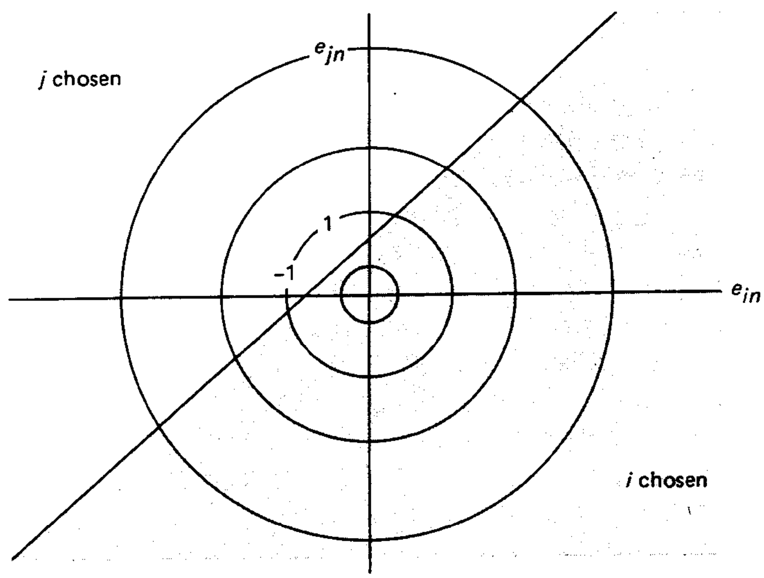

Figure 1.1
Probability that alternative i is chosen given that $V_{in} - V_{jn} = 1$.

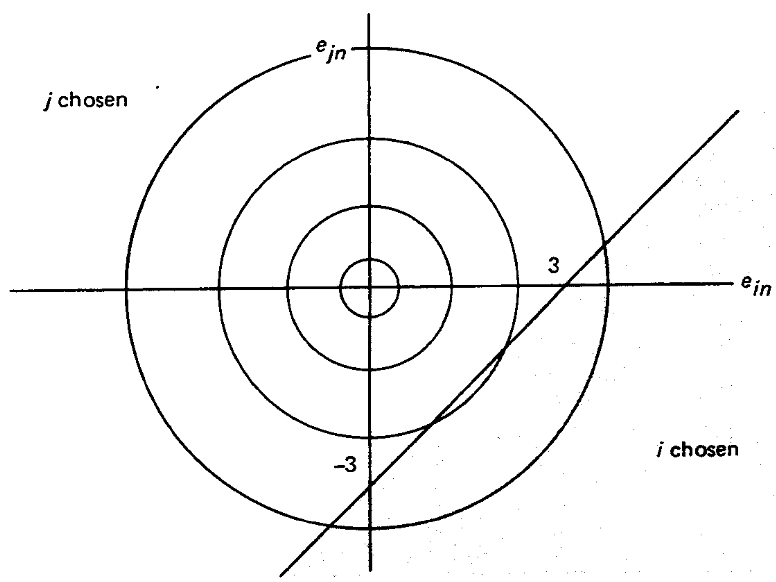

Figure 1.2
Probability that alternative i is chosen given that $V_{in} - V_{jn} = -3$.

This is, alternative i will be chosen as long as e_{jn} is no more than one unit larger than e_{in}.

Figure 1.1 depicts this probability. Each point in the two-dimensional graph represents a particular value of e_{in} and e_{jn}. The concentric circles represent the joint density of e_{in} and e_{jn}. The diagonal line connects those points for which e_{jn} is exactly one unit larger than e_{in}. For all points below this line, e_{jn} is less than one unit greater than e_{in} and so alternative i is chosen. For all points above the line, e_{jn} exceeds e_{in} by more than one unit and so alternative j is chosen. The probability that alternative i is chosen is the probability that e_{in} and e_{jn} fall below the line. That is, P_{in} is the volume under the shaded part of the joint density function. Different density functions for e_{in} and e_{jn} will obviously give rise to different probabilities.

The same type of analysis applies if the observed part of utility for alternative i is less than that for alternative j. For example, suppose $V_{in} = 2$ and $V_{jn} = 5$. Then alternative i will be chosen only if e_{in} is at least three units larger than e_{jn}. Stated conversely and in closer accordance with equation (1.6), alternative i will be chosen only if e_{jn} is no more than three units below e_{in}. The probability of this occurring is depicted in figure 1.2.