Cambridge University Press
978-1-107-09264-8 - Applied Choice Analysis: Second Edition
David A. Hensher, John M. Rose and William H. Greene
Excerpt
More information

Part I

Getting started

Cambridge University Press
978-1-107-09264-8 - Applied Choice Analysis: Second Edition
David A. Hensher, John M. Rose and William H. Greene
Excerpt
More information

1 In the beginning

Education is a progressive discovery of our own ignorance.

(Will Durant, 1885–1991)

1.1 Choosing as a common event

Why did we *choose* to write the first edition of this primer and then a second edition? Can it be explained by some inherent desire to seek personal gain or was it some other less self-centered interest? In determining the reason, we are revealing an underlying objective. It might be one of maximizing our personal **satisfaction** level or that of satisfying some community-based objective (or social obligation). Whatever the objective, it is likely that there are a number of reasons why we made such a *choice* (between writing and not writing this primer) accompanied by a set of constraints that had to be taken into account. An example of a reason might be to "promote the field of research and practice of choice analysis"; examples of constraints might be the time commitment and the financial outlay.

Readers should be able to think of choices that they have made in the last seven days. Some of these might be repetitive and even habitual (such as taking the bus to work instead of the train or car), buying the same daily newspaper (instead of other ones on sale); other choices might be a once-off decision (such as going to the movies to watch a latest release or purchasing this book). Many choice situations involve *more than one choice* (such as choosing a destination and means of transport to get there, choosing where to live and the type of dwelling, or choosing which class of grapes and winery in sourcing a nice bottle of red or white).

The storyline above is rich in information about what we need to include in a study of the choice behavior of individuals or groups of individuals (such as households, lobby groups, and organisations). To arrive at a choice, an individual must have considered a set of **alternatives**. These alternatives are usually called the **choice set**. Logically, one must evaluate at least two alternatives to be able to make a choice (one of these alternatives may be "not to make a choice" or "not participate at all"). At least one actual (or potential) **choice setting** must exist (e.g. choosing where to live, choosing who to vote for, or choosing among alternative future green sources of vehicle fuels) but there may be more than one choice (e.g. what type of dwelling to live in, whether to buy or rent, and how much to pay per week if rented). The idea that an individual may have to consider a number of choices leads to a set of interrelated choices. Some choice situations might also involve subjective responses on a psychological scale (such as the rating of a health scheme, the amenity of a suburb, or a bottle of wine); or on a best–worst scale in which they choose the most preferred (or best) alternative or attribute and the least preferred (or worst) alternative or attribute.

Determining the set of alternatives to be evaluated in a choice set is a crucial task in choice analysis. Getting this wrong will mean that subsequent tasks in the development of a choice model will be missing relevant information. We often advise analysts to devote considerable time to the identification of the choices that are applicable in the study of a specific problem. This is known as **choice set generation**. In identifying the relevant choices, one must also consider the range of alternatives, and start thinking about what influences the decision to choose one alternative over another. These influences are called **attributes** if they relate to the description of an alternative (e.g., travel time of the bus alternative, vintage of a bottle of wine), but an individual's prejudices (or tastes) will also be relevant and are often linked to **socio-economic characteristics** (SECs) such as personal income, age, gender, and occupation.

To take a concrete example, a common problem for transportation planners is to study the transport-related choices made by a sample of individuals living in an urban area. Individuals make many decisions related to their transportation needs. Some of these decisions are taken occasionally (e.g., where to live and work) while others are taken more often (e.g., departure time for a specific trip). These examples highlight a very important feature of choice analysis – the temporal perspective. Over what time period are we interested in studying choices? As the period becomes longer, the number of possible choices that can be made (i.e., are not fixed or pre-determined) are likely to increase. Thus if we are interested in studying travel behavior over a five-year period, then it is reasonable to assume that an individual can make choices related to the locations of both living and working, as well as the means of transport and departure time. That is, a specific choice of means of transport may indeed be changed as a consequence of the person changing where they reside or work. In a shorter period such as one year, choosing among modes of transport may be conditional on where one lives or works, but the latter is not able to be changed given the time that it takes to relocate one's employment.

The message in the previous paragraphs is that careful thought is required to define the choice setting so as to ensure that all possible behavioral responses (as expressed by a set of choice situations) can be accommodated when a change in the decision environment occurs. For example, if we increase fuel prices, then the cost of driving a car increases. If one has only studied the choice of mode of transport then the decision maker will be "forced" to modify the choice among a given set of modal alternatives (e.g., bus, car, train). However it may be that the individual would prefer to stay with the car but to change the time of day they travel so as to avoid traffic congestion and conserve fuel. If the departure time choice model is not included in the analysis, then experience shows that the modal choice model tends to force a substitution between modes, which in reality is a substitution between travel at different times of the day by car.

Armed with a specific problem or a series of associated questions, the analyst now recognizes that to study choices we need a set of choice situations (or outcomes), a set of alternatives and a set of attributes that belong to each alternative. But how do we take this information and convert it to a useful framework within which we can study the choice behavior of individuals? To do this, we need to set up a number of behavioral rules under which we believe it is reasonable to represent the process by which an individual considers a set of alternatives and makes a choice. This framework needs to be sufficiently realistic to explain past choices and to give confidence in likely behavioral responses in the future that result in staying with an existing choice or making a new choice. The framework should also be capable of assessing the likely support for alternatives that are not currently available, be they new alternatives in the market or existing ones that are physically unavailable to some market segments. These are some of the important issues that choice modelers will need to address and which are central to the journey throughout this book.

Before we overview the structure of the book, we thought it useful to go back in time and get an appreciation of the evolution of choice modeling, which began at least ninety years ago.

1.2 A brief history of choice modeling

It is eighty-seven years since Thurstone's classic 1927 paper on the *law of comparative judgment*, in which he assumes that the response of an individual to a pair of alternatives, *i*, *j*, is determined by the discriminal processes $v_i = f(\alpha_i) + \varepsilon_i$ and $v_j = f(\alpha_j) + \varepsilon_j$. The terms $f(\alpha_i)$ and $f(\alpha_j)$ represent a single-valued function of unknown parameters α_i and α_j, characteristics of the "objects participating in the *i*, *j* pair." These parameters are referred to by Thurstone as "affective values" of the corresponding objects (or alternatives); ε_i and ε_j are elements of the discriminal processes specific to the randomly selected individual and are assumed by Thurstone to obey a normal bivariate **distribution** function. The difference process $(v_i - v_j)$ is distributed normally with mean $f(\alpha_i) - f(\alpha_j)$ and **variance** $\sigma^2_{ij} = \sigma^2_i + \sigma^2_j + 2\rho_{ij}\sigma_i\sigma_j$ where ρ_{ij} is the correlation between the alternatives. The individual is assumed to judge $X_i > X_j$ when $(v_j - v_i) > 0$. Thus the probability that a randomly sampled individual will be observed to judge $X_i > X_j$ is $Prob_{ij} = \Phi\{f(\alpha_i) - f(\alpha_j) / \sigma_{ij}\}$. This response function is referred to as a statement of the law of comparative judgment. McFadden (2001) described the Thurstone contribution as a model of imperfect discrimination in which alternative *i* with true stimulus level V_i is perceived with a normal error as $V_i + \varepsilon_i$ and Thurstone showed that the probability $P_{\{i,j\}}(i)$ that alternative *i* is chosen over alternative *j* has a form that we now call binomial probit. The emphasis on probabilistic choice theory can be credited to both Thurstone (1945) and a lesser-known author, Hull (1943).

An alternative to the normal response function proposed by Bradley and Terry (1952) and Luce (1959) is of special interest because of its psychological interpretation (Restle 1961, Bock and Jones 1968). The authors proposed a model for the probability that X_i is ranked above X_j in the pair X_i, X_j, as $Prob_{ij} = \{\pi_i/(\pi_i + \pi_j)\}$, $i = 1,2,\ldots,n$; $j = 1,2,\ldots,n$. π_i and π_j are positive parameters characteristic of alternatives X_i and X_j. Bradley and Terry (1952) introduce $\pi_i + \pi_j$ in the denominator to **normalize** π_i so that $Prob_{ij} + Prob_{ji} = 1$. Luce (1959) developed the theoretical foundations in a precise form, in which π_i can be interpreted as the probability that X_i will be ranked first among all *n* alternatives. The probability that X_i will be ranked first in any subset of alternatives, and in particular in the subset $\{X_i, X_j\}$ in any subset of alternatives follows from Luce's *principle of independence from irrelevant alternatives*, which states that the ratio π_i/π_j is constant regardless of what other alternatives are in the subset. This became known as the IIA rule or constant shares assumption. Importantly, this model was transformed into the logistics

Cambridge University Press
978-1-107-09264-8 - Applied Choice Analysis: Second Edition
David A. Hensher, John M. Rose and William H. Greene
Excerpt
More information

response function by setting $\pi_i/\pi_j = \exp(\alpha_i - \alpha_j)$. Bradley and Terry (1952) were the first (in the psychological literature) to estimate the logit response function by using a maximum likelihood estimator, although the logistic form goes back many years in Bioassay (see Ashton 1972 for a review and summary of the contribution of Berkson). Estimates of the natural log of π_i/π_j were obtained by employing logistic deviates $y_{ij} = \ln\{prob_{ij}/(1 - prob_{ij})\}$. After exponential transformation of parameters (what later became the representative or observed component of *utility*), the Bradley–Terry–Luce (BTL) model becomes equivalent to Thurstone's Case V model, except that the logistic's density replaces the Gaussian density of Thurstone's response function. The principle of IIA has the exact same effect as constant correlation of discriminal processes for all pairs of alternatives (stimuli). This implies that the conditional probability of an individual's choice between any two alternatives, given their choice between any other two alternatives, is equal to the unconditional probability. The famous *red bus/blue bus* example introduced by Mayberry in Quandt (1970) and due to Debreu (1960), has been used extensively to highlight the risk of empirical **validity** of IIA, which became the springboard for many of the developments in discrete choice models to circumvent the rigidity of IIA.

Marschak (1959) generalized the BTL model to **stochastic utility maximization** over multiple alternatives, and introduced it to economics, referring for the first time to Random Utility Maximization (RUM) (also see Georgescu-Roegen 1954). Marschak explored the testable implications of maximization of random preferences, and proved for a finite set of alternatives that choice probabilities satisfying Luce's IIA axiom were consistent with RUM. An extension of this result established that a necessary and sufficient condition for RUM with independent errors to satisfy the IIA axiom was that the ε_i be identically distributed with a Type I Extreme Value distribution, $\text{Prob}(\varepsilon_i \le c) = \exp(-e^{-c/\sigma})$, where σ is a scale factor and c is a location parameter. The sufficiency was proved by Anthony Marley and reported by Luce and Suppes (1965).

In the 1960s a number of researchers realized that the study of choices among mutually exclusive (discrete) alternatives was not appropriate through the application of ordinary least squares (OLS) regression. Given that the dependent variable of interest was discrete, typically binary (1, 0), the use of OLS would result in predicted outcomes violating the boundary limits of probability. Although under a binary choice setting, probabilities in the range 0.3 to 0.7 tended to satisfy a common range of a linear OLS (or linear probability model form), any probabilities at the extremities were likely to be greater than 1.0 and less than 0. To avoid this, a transformation is required, the most popular being the logistic (log of the odds) transformation. Software to estimate a binary logit (or probit) model started to appear in the 1960s, replacing the popular discriminant analysis method. The early programs included PROLO (PRObit-LOgit) written by Cragg at the University of British Columbia and which was used in many PhD theses in the late 1960s and early 1970s (including Charles Lave 1970, Thomas Lisco 1967, and David Hensher 1974). Peter Stopher (at Northwestern University, and now at Sydney) in the late 1960s had written a program to allow for more than two alternatives, but as far as we are aware it was rarely used. During the period of the late 1960s and early 1970s there were a number of researchers developing logit software for multinomial logit, including McFadden's code that became the basis of QUAIL (programmed in particular by David Brownstone), Charles Manski's program (XLogit) used by MIT students such as Ben-Akiva, Andrew Daly's ALogit, Hensher and Johnson's BLogit, and Daganzo and Sheffi's TROMP. Bill Greene had a version of Limdep in the 1970s that began with Tobit and then Logit.

Despite the developments in software (mainly binary choice and some limited multiple choice capability), it was not until the link was made between McFadden's contribution at Berkeley (McFadden 1968) and a project undertaken by Charles River Associates to develop a joint mode and destination choice model (Domencich and McFadden 1975), that we saw a significant growth in research designed to deliver practical tools for modeling interdependent discrete choices involving more than two alternatives. By the late 1960s, McFadden had developed an empirical model from Luce's choice axiom (centered on IIA as described above). Letting $P_C(i)$ denote the probability that a subject confronted with a set of mutually exclusive and exhaustive alternatives C will choose alternative i, given the IIA property, Luce showed that if his axiom holds, then one can associate with each alternative a positive "strict utility" w_i such that $P_C(i) = w_i / \sum_{k \in C} w_k$. Taking the strict utility for alternative i to be a parametric exponential function of its attributes x_i, $w_i = \exp(x_i\beta)$, gave a practical statistical model for individual choice data. McFadden called this the *conditional logit model* because it reduced to a logistic in the two-alternative case, and had a ratio form analogous to the form for conditional probabilities (McFadden 1968, 1974). McFadden (1968, 1974) proved necessity (given sufficiency had already been shown), starting from the implication of the Luce axiom that multinomial choice between an object with strict utility w_1 and m objects with strict utilities w_2 matched binomial choice between an object with strict utility w_1 and an object with strict utility mw_2.

In McFadden (2001), the author explains that he "initially interpreted the conditional logit model as a model of a decision making bureaucracy, with random elements coming from **heterogeneity** of tastes of various bureaucrats. It was then transparent that in an empirical model with data across decision-makers, the randomness in utility could come from both inter-personal and intra-personal variation in preferences, and from variations in the attributes of alternatives known to the decision-maker but not to the observer." This led in his classic 1974 paper on the conditional logit model to introduce the idea of an *extensive* margin for discrete decisions in contrast to the *intensive* margin that operates for a representative consumer making continuous decisions. This was a defining distinction between the economist's and the psychologist's interpretation of randomness.

The 1970s saw much activity in finessing the multinomial logit model based on the form developed by Dan McFadden. In addition to the Charles River Associates project (published as the book by Domencich and McFadden 1975), which introduced inclusive value to connect levels calculated as probability weighted averages of systematic utility components at the next level down in the tree (with Ben Akiva separately developing the log sum formula for exact calculation of inclusive values – see Ben Akiva and Lerman 1979) McFadden directed the Travel Demand Forecasting Project (TDFP), which set out to develop a comprehensive framework for transportation policy analysis using disaggregate behavioral tools. TDFP used the introduction of BART to test the ability of disaggregate travel demand models to forecast a new transportation mode. On the methodological front, TDFP developed methods for choice-based sampling and for simulation, and statistical methods for estimating and testing **nested** logit models, that laid the foundation for later results. Some of the ideas that led to the eventual discovery of the nested logit had been laid down by Marvin Manheim (1973), and Alan Wilson barely missed it when proposing the combined distribution-mode split function for the famous SELNEC transport model (Wilson *et al.* 1969).

The concern with the limitation of the IIA condition led to the development of the nested logit model (referred to as tree logit by Andrew Daly) in which the idea of dissimilarity noted over forty-five years before in psychology finally was treated explicitly in the RUM framework through the recognition that the variance associated with the unobserved influences in the random component is likely to be different across the finite set of alternatives in a choice set, but possibly similar for subsets of alternatives. This had appeal to those interested in decision trees, although it must be pointed out that the nesting structure is a mechanism to accommodate differential variance in the unobserved effects that may not align with intuition in the construction of decision trees. With the knowledge that the distribution of the variance associated with the unobserved effects can be defined by a location and a **scale parameter**, the nested logit model had found a way of explicitly identifying and parameterizing this scale, which became known alternatively as composite cost, inclusive value, logsum and expected maximum utility. The contributions to this literature, in particular the theoretical justification under RUM, are attributable to Williams (1977) and Daly and Zachary (1978), with a later generalization by McFadden (2001). In particular, the Williams–Daly–Zachary analysis provides the foundation for derivation of RUM-consistent choice models from social surplus functions, and connects RUM-based models to **willingness to pay** (WTP) for projects.

The period from the mid 1970s to 2010 saw an explosion of contributions to theory, computation and empirical applications of **closed-form** discrete-choice models of the multinomial logit (MNL) and nested logit (NL) variety. The most notable development of closed-form models occurred when it was recognized that the nested logit model reveals crucial information to accommodate the pooling of multiple data sets, especially revealed and stated preference data. Although Louviere and Hensher (1982, 1983) and Louviere and Woodworth (1983) had recognized the role of stated choice data in the study of discrete choices in situations where new alternatives and/or existing alternatives with stretched attribute levels outside of these observed in real markets exist, it was the contribution of Morikawa (see Ben Akiva and Morikawa 1991) that developed a way to combine data sets while accounting for differences in scale (or variance) that was the essential feature of the choice model that had to be satisfied if the resulting model was able to satisfy the theoretical properties of RUM. Bradley and Daly (1997, but written in 1992) and Hensher and Bradley (1993) had shown how the nested logit method could be used as a "nested logit trick," to identify the scale parameter(s) associated with pooled data sets and to adjust the parameter estimates so that all absolute parameters can be compared across data sets.

Despite great progress in linking multiple choices and multiple data sets, some critical challenges remained. These centered initially on open-form models such as multinomial probit, which in the 1980s was difficult to estimate beyond a few alternatives, given the need to accommodate multiple integrals through analytical solutions. The need for numerical integration was required, but it was not until a number of breakthroughs associated with the notion of simulated **moments** (McFadden 1989) that the door opened to ways of accommodating more complex choice models, including models that could account for the fuller range of sources of unobserved heterogeneity in preferences.