Walker, J.L. and M. Ben-Akiva, "Advances In Discrete Choice: Mixtures Models," in (A. de Palma, R. Lindsey, E. Quinet and R. Vickerman, eds.) Handbook in Transport Economics, Edward Elgar, 2011, pp. 160-187.

HANDBOOK IN TRANSPORT ECONOMICS

EDITED BY ANDRE DE PALMA, ROBIN LINDSEY, EMILE QUINET, ROGER VICKERMAN

CHAPTER 8

ADVANCES IN DISCRETE CHOICE: MIXTURES MODELS

BY JOAN L. WALKER AND MOSHE BEN-AKIVA

INTRODUCTION

Recent advances in discrete choice models have been driven by the growth in computer power and use of simulation, which have allowed for unprecedented flexibility in distributional form. In this chapter, we review both the basic discrete choice models and the latest formulations. In particular, we focus on the concept of mixture models. Mixture models are currently being used in a wide array of statistical modeling procedures as a way to relax restrictive assumptions and generalize model forms. As mixing allows for any distributional form to be approximated, this represents a powerful and important advancement in discrete choice analysis.

We will first briefly review the foundations of discrete choice analysis and the classic model forms of probit and the GEV family (logit, nested logit, cross-nested logit). Then we will move onto mixture models, beginning with basic formulations and then covering more advanced forms, including what we are calling *behavioral (or structural) mixture models*.

FOUNDATIONS OF DISCRETE CHOICE ANALYSIS

We start by providing the foundations of choice analysis, including the choice modeling framework and the random utility model. This section is based on Ben-Akiva and Lerman (1985), where further details can be obtained.

CHOICE MODELING FRAMEWORK

This section presents the basic elements that are used to model a decision maker's choice among a set of mutually exclusive and collectively exhaustive alternatives. A specific theory of choice requires definitions of the following elements of the choice problem:

Who (or what) is the decision maker? The decision making entity can be an individual person or group of persons (such as a household) or an organization such as a firm or government agency.

What are the characteristics of the decision maker? Individuals have different tastes, and therefore we must explicitly treat the differences in the decision-making processes among individuals. Therefore the characteristics of the decision maker (for example, gender, age, and firm size) become an important part of the problem.

What are the alternatives? This is a finite and countable set of alternatives from which a decision maker chooses. The environment of the decision maker determines the *universal* set of alternatives. Any single decision maker considers a subset of this universal set, termed a *choice set*, which includes those alternatives that are both feasible to the decision maker and known during the decision process.

What are the attributes of the alternatives? In discrete choice analysis, we characterize each alternative by its attributes, for example price, time, brand name, color, or size. The attractiveness of an alternative is then evaluated in terms of a vector of attribute values. Attributes may be measured on a continuous scale (time, price) or be categorical (brand). Furthermore, categorical attributes can be ordinal (1, 2, 3 bathrooms) or nominal (red, green, blue). Some may be more straightforward to measure (price, number of rooms), whereas others may be more complex (comfort, reliability).

What is the decision rule that the decision maker uses to make a choice? The decision rule describes the internal mechanisms used by the decision maker to process the information available and arrive at a unique choice. There exists a wide array of decision rules, including dominance, satisfaction, lexicographic, elimination by aspect, habitual, imitation, and utility. It is this latter class that is most often associated with discrete choice analysis due to its extensive use in the development of the predictive models of human behavior emphasized here. A utility-based decision rule means that the attractiveness of an alternative expressed by a vector of attribute values is reducible to a scalar. This index of attractiveness is denoted utility. The supposition of a single index is based on the notion of trade-offs, or compensatory offsets, that a decision maker is using explicitly or implicitly in comparing different attributes. Utility theory derives from microeconomic consumer theory.

Extensions of Consumer Theory for Discrete Choice Analysis

While utility theory is not necessary for the methods described in this chapter, microeconomic consumer theory does provide a helpful means of interpreting the framework and deriving choice models. Here we briefly discuss how economic consumer theory extends to discrete choice analysis. The seminal work linking discrete choice analysis with consumer theory is McFadden (1974).

Classic economic consumer theory assumes that consumers are rational decision makers. When faced with a set of possible consumption bundles, consumers assign preferences to each of the various bundles and then choose the most preferred bundle from the set of feasible bundles. Under the assumptions of complete, transitive, and continuous preferences, there exists an ordinal utility function that expresses mathematically the consumer's preferences and associates a real number with each possible bundle such that it summarizes the preference orderings of the consumer. Consumer behavior can then be expressed as an optimization problem in which the consumer selects the consumption bundle such that her utility is maximized subject to her budget constraint. The optimization is solved to obtain the demand functions in the form of quantities of goods demanded as a function of income and prices. The demand function can then be substituted back into the utility equation to derive the indirect utility function, defined as the maximum utility that is achievable under given prices and income. The indirect utility function is what is used in discrete choice analysis, and will be referred to simply as *utility* throughout this chapter.

There are several extensions to classic consumer theory that are important to discrete choice models.

First, classic consumer theory is concerned with continuous (that is, infinitely divisible products), which is necessary for the calculus-based derivations of many of the key results. Discrete choice theory is concerned with a choice among a set of mutually exclusive alternatives, and therefore a different mathematical approach is necessary.

Second, classic consumer theory assumes homogeneous goods (for example, a car is a car), and therefore the utility is a function of quantities only and not attributes. Lancaster (1966) proposed that it is the attributes of the goods that determine the utility they provide, and therefore utility can be expressed as a function of the attributes of the commodities (for example, a car has a make, model, and level of fuel efficiency).

Finally, classic consumer theory assumes deterministic behavior. The earliest probabilistic choice models were originated in psychology by Thurstone (1927), and further developed by Luce (1959) and Marschak (1960). These introduced the concept that individual choice behavior is intrinsically probabilistic as a way to explain observed intransitivity of preferences. Marschak (1960) provided a link to economics with the concept of random utility theory. McFadden (1974) then built the connection between econometric models of probabilistic choice and utility theory. Random utility theory attributes the source of the stochasticity to incomplete information of the analyst, and Manski (1977) identified four sources of uncertainty: unobserved alternative attributes, unobserved individual characteristics (or taste variations), measurement errors, and proxy (or instrumental) variables. Due to these uncertainties, utility is modeled as a random variable, consisting of an observable (that is, measurable) component and an unobservable (that is, random) component.

These modifications lead to the operational framework for discrete choice analysis that is described in the next section. That is, a framework that works for discrete choices among related commodities, and makes use of a stochastic utility broken into a systematic and random component and which is a function of attributes of the alternatives and characteristics of the decision maker.

Random Utility Model

Random utility theory, the origins of which were described above, is the most common theoretical basis of discrete choice models. The random utility model (or RUM) operationalizes this theory. In RUM, the utility is treated as a random variable. Specifically, the utility that individual n associates with alternative i in the choice set C_n is given by

$$U_{in} = V_{in} + \varepsilon_{in} .$$

V_{in} is the deterministic (or systematic) part of the utility, and is a function of the attributes of the alternative itself and the characteristics of the decision maker (together denoted as x_{in}). ε_{in} is the random term, capturing the uncertainty. As we assume the alternative with the highest utility is chosen, the probability that alternative i is chosen by decision maker n from choice set C_n is

$$P(i|C_n) = P\left[U_{in} \geq U_{jn}, \forall\, j \in C_n\right] = P[U_{in} = \max_{j \in C_n} U_{jn}] \,.$$

Or, substituting in the deterministic and random components of the utility,

$$\begin{aligned} P(i|C_n) &= P\left[V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}, \forall j \in C_n\right] \\ &= P\left[(V_{in} + \varepsilon_{in}) = \max_{j \in C_n}(V_{jn} + \varepsilon_{jn})\right] \\ &= P\left[\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn}, \forall\, j \in C_n\right] \quad . \end{aligned}$$

This is a multivariate cumulative distribution function of a vector of epsilon differences (with respect to alternative i) evaluated for the vector of systematic utility differences (also with respect to alternative i).

The output of a random utility model is the probability of an individual selecting each alternative. These individual probabilities can be aggregated to produce forecasts for the population.

Simplifying assumptions are often made in order to maintain parsimonious and tractable discrete choice models. An assumption already highlighted above is that of utility maximization. Other simplifying assumptions include deterministic choice sets, the use of straightforward explanatory variables x_{in} (for example, relatively easy to measure characteristics of the decision maker and attributes of the alternative), and linearity in the unknown parameters β (that is, $V_{in} = \beta x_{in}$). Another key assumption is made on the distribution of the disturbances, ε, and this is addressed in the next section.

THE CLASSIC DISCRETE CHOICE MODELS

Here we describe the most common, operational discrete choice models. As mentioned above, each makes a different assumption on the distribution of the disturbance term, epsilon. While such assumptions are made on the utilities in level form, the derivation of the model is based on the utility differences. The distribution of epsilon differences is derived from the assumed distribution of the epsilons in level form. The probability equation for the choice model is then simply the cumulative distribution function of the epsilon differences evaluated at the differences of systematic utilities.

PROBIT

The name probit comes from *probability unit*. The probit model is based on the most obvious assumption to make, which is that the vector of disturbances $\varepsilon_n = \left(\varepsilon_{1n}, \varepsilon_{2n}, \dots, \varepsilon_{J_n n}\right)'$ is multivariate normal distributed with a vector of means 0 and a $J_n \times J_n$ variance-covariance matrix Σ_ε, where J_n is the number of alternatives in C_n. For example, in the binomial case,

$$\begin{bmatrix}\varepsilon_{1n}\\ \varepsilon_{2n}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{bmatrix}\right),$$

where σ_i^2 and σ_{ij} are the variance and covariance elements of Σ_ε. The difference in epsilons is

$(\varepsilon_{2n} - \varepsilon_{1n}) \sim N(0, \sigma_1^2 + \sigma_2^2 - 2\sigma_{12})$,

which we denote as $\Delta\varepsilon_n \sim N(0, \sigma^2)$.

The probability density function (pdf) is then

$$f(\Delta\varepsilon) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\Delta\varepsilon}{\sigma}\right)^2} .$$

And the probit choice probability is the cumulative normal distribution, which is the integral over the pdf:

$$P(1|C_n) = F_q(V_n) = \int_{-\infty}^{V_n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{q}{\sigma}\right)^2} \, dq = \Phi\left(\frac{V_n}{\sigma}\right),$$

where $\Phi(\cdot)$ is the standard cumulative normal distribution.

For trinomial probit, where we have taken the utility differences with respect to the first alternative, the distribution of differenced epsilons is

$$\begin{bmatrix} \varepsilon_2 - \varepsilon_1 \\ \varepsilon_3 - \varepsilon_1 \end{bmatrix} \sim N\left(0, \Sigma_1 = \begin{bmatrix} \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} & \sigma_1^2 + \sigma_{23} - \sigma_{12} - \sigma_{13} \\ \sigma_1^2 + \sigma_{23} - \sigma_{12} - \sigma_{13} & \sigma_1^2 + \sigma_3^2 - 2\sigma_{13} \end{bmatrix}\right).$$

Let $n(q; 0, \Sigma_1)$ denote the above bivariate density evaluated at the column vector $q = (q_1, q_2)'$, where Σ_1 is the variance-covariance matrix of epsilon differences taken with respect to alternative 1. The probability is then

$$P(1|C_n) = P(\varepsilon_{2n} - \varepsilon_{1n} \leq V_{1n} - V_{2n} \; and \; \varepsilon_{3n} - \varepsilon_{1n} \leq V_{1n} - V_{3n})$$

$$= \int_{-\infty}^{V_{1n} - V_{2n}} \int_{-\infty}^{V_{1n} - V_{3n}} n(q; 0, \Sigma_1) dq_1 dq_2 .$$

More generally, a probit equation for a choice amongst J_n alternatives is based on the assumption that the disturbance terms of the utility functions are multivariate normal distributed with mean vector zero. The disturbance differences, $\Delta\varepsilon_1$, taken with respect to the first alternative is also multivariate normal distributed $\Delta\varepsilon_1 \sim MVN(0, \Sigma_1)$, with a distribution as follows:

$$f(\Delta\varepsilon_1) = (2\pi)^{-\frac{J_n - 1}{2}} |\Sigma_1|^{-\frac{1}{2}} e^{-\frac{1}{2}(\Delta\varepsilon_1{}' \Sigma_1^{-1} \Delta\varepsilon_1)}.$$

Defining $n(q; 0, \Sigma_1)$ as the above multivariate density evaluated at the $(J_n - 1) \times 1$ column vector q, the probability of choosing alternative 1 is then

$$P(1|C_n) = \int_{-\infty}^{V_{1n} - V_{2n}} \int_{-\infty}^{V_{1n} - V_{3n}} \cdots \int_{-\infty}^{V_{1n} - V_{Jn}} n(q; 0, \Sigma_1) \, dq.$$

This model requires evaluation of a $J_n - 1$ integral to evaluate the choice probabilities.

The main advantage of probit is its ability to capture all correlations among alternatives via the fully specified variance-covariance matrix Σ_1 of epsilon differences. However, it has no closed form expression, and therefore is impractical for large number of alternatives.

GEV Family: Logit, Nested Logit, Cross-Nested Logit

The other set of classic discrete choice models are all members of the GEV, or Generalized Extreme Value family. (More recently GEV is being referred to as MEV or Multivariate Extreme Value to be more consistent with the nomenclature for the statistical distribution employed). This is a large family of models that include the logit and the nested logit models. The logit model, the simplest of the GEV family, has been more popular than probit because of its tractability. However, it imposes restrictions on the covariance structure, which may be unrealistic in some contexts. The derivations of other models in the GEV family are aimed at relaxing restrictions, while maintaining tractability.

The term logit comes from *Logistic Probability Unit*. The model was first introduced in the context of binomial choice where the difference in two Extreme Value distributed random variables has a logistic distribution. The logit model is derived from the assumption that the disturbance terms of the utility functions are independent and identically Extreme Value distributed. That is, ε_{in} for all i and n is distributed as

$F(\varepsilon) = exp\left[-e^{-\mu(\varepsilon-\eta)}\right]$ (the cdf) and

$f(\varepsilon) = \mu e^{-\mu(\varepsilon-\eta)} exp\left[-e^{-\mu(\varepsilon-\eta)}\right]$ (the pdf),

where η is a location parameter (irrelevant as it cancels out) and μ is a strictly positive scale parameter. The probability that a given individual n chooses alternative i within a choice set C_n is given by

$$P(i|C_n) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}} .$$

An important property of the logit model is Independence from Irrelevant Alternatives (IIA). That is, the ratio of the probabilities of any two alternatives (i and j below) is independent of the choice set (C_1 and C_2 below) as follows:

$$\frac{P(i|C_1)}{P(j|C_1)} = \frac{P(i|C_2)}{P(j|C_2)} \ \forall\ i, j, C_1, C_2 ,$$

where $i, j \in C_1, i, j \in C_2, C_1 \subseteq C_n$, and $C_2 \subseteq C_n$.

The nested logit model relaxes some of the restrictive assumptions of logit, capturing some correlations among alternatives. It overcomes the IIA problem of logit when utilities of alternatives are correlated (such as the red bus and blue bus problem) or when multidimensional choices are considered (such as departure time and route choice). It is based on the partitioning of the choice set C_n into M mutually exclusive and collectively exhaustive nests C_{mn}. Here we present the model in the case of two levels, although the model can be extended to any number of levels (see Ben-Akiva and Lerman, 1985). The probability of decision maker n to choose alternative i within nest C_{mn} is given by

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

where

$$P(C_{mn}|C_n) = \frac{e^{\mu V_{C_m n}}}{\sum_{l=1}^{M} e^{\mu V_{C_l n}}},$$

$$P(i|C_{mn}) = \frac{e^{\mu_m V_{in}}}{\sum_{j \in C_{mn}} e^{\mu_m V_{jn}}}, \text{ and}$$

$$V_{C_m n} = \frac{1}{\mu_m} ln \sum_{j \in C_{mn}} e^{\mu_m V_{jn}}.$$

This last term $V_{C_m n}$ is the expected maximum utility (also known as the logsum, inclusive value, or accessibility). Parameters μ and μ_m reflect the correlation among alternatives within the nest C_{mn}. The correlation between the utility of two alternatives is

$$Corr(U_{in}, U_{jn}) = \begin{cases} 1 - \frac{\mu^2}{\mu_m^2} & \text{if } i \text{ and } j \in C_{mn} \\ 0 & \text{otherwise} \end{cases},$$

where $0 < \mu/\mu_m \leq 1$ (see Ben-Akiva and Lerman, 1985). If $\mu = \mu_m$ for all m then $Corr(U_{in}, U_{jn}) = 0$ for all i and j and the model reverts to logit. It is not possible to identify μ and all μ_m in the model, and one must be constrained (typically one is set to 1).

More generally, logit and nested logit are members of the GEV family of models, which were proposed by McFadden (1978). The GEV family is attractive because it allows for a multivariate distribution of the disturbance terms (that is, it relaxes the independence assumption) and yet retains a closed form expression for the choice probability.

A GEV model depends on a function G, which is a non-negative, differentiable function defined for positive arguments with the following properties:

1. G is homogeneous of degree $\mu > 0$ (McFadden's original formulation with $\mu = 1$ was generalized to $\mu > 0$ by Ben-Akiva and François, 1983.)
2. $\lim_{z_i \to \infty} G(z_1, \dots, z_i, \dots, z_{J_n}) = \infty, \forall\, i = 1, \dots, J_n$
3. The kth partial derivative with respect to k distinct z_i is non-negative if k is odd, and non-positive if k is even.

A function G that satisfies these conditions leads to the cumulative distribution function

$$F(\varepsilon_1, \dots, \varepsilon_{J_n}) = e^{-G(e^{-\varepsilon_1}, \dots, e^{-\varepsilon_{J_n}})},$$

and the resulting choice model

$$P(i|C_n) = \frac{e^{V_{in}} \frac{\partial G\left(e^{V_{1n}}, \dots, e^{V_{J_n n}}\right)}{\partial e^{V_{in}}}}{\mu G\left(e^{V_{1n}}, \dots, e^{V_{J_n n}}\right)}.$$

McFadden (1978) shows that the choice model defined by this equation is consistent with random utility maximization.

Further, the expected maximum utility (V_C) is as follows:

$$V_C = \frac{\ln G\left(e^{V_1},\dots,e^{V_{J_n}}\right)+\gamma}{\mu},$$

where γ is Euler's constant (~0.577).

From the GEV theorem, the known family of logit models can be derived. The logit model is a GEV model with G-function

$$G(z) = \sum_{i=1}^{J_n} z_i^{\mu}.$$

Nested logit has a G-function of

$$G(z) = \sum_{m=1}^{M}\left(\sum_{i\in C_{mn}} z_i^{\mu_m}\right)^{\frac{\mu}{\mu_m}}.$$

The only other commonly used GEV form is the cross-nested logit model (and its variants, including paired combinatorial logit and generalized nested logit, see Koppelman and Sethi, 2000), which has a G-function of

$$G(z) = \sum_{m=1}^{M}\left(\sum_{j\in C_n} \alpha_{jm}\, z_i^{\mu_m}\right)^{\frac{\mu}{\mu_m}}.$$

While the GEV model does allow for multivariate disturbance distribution, it still has limitations. The mixture models described in the next section further relax the restrictions.

MIXTURE MODELS

There has been an evolving sophistication of choice models in which restrictions of earlier model forms are relaxed. The latest wave of advances has been in the area of mixture models. Mixtures have been used for a long time as a way to generate flexible distributional forms. For example, Figure 1 demonstrates how a mixture of two normal distributions (40% of a $N(2,1)$ and 60% of a $N(6,2)$) yields a bimodal distribution that could not otherwise be captured by a standard parametric distribution. The resulting probability density function is a convex combination of the two normal probability densities as follows:

$$f(y) = 0.4\,\phi\left(\frac{y-2}{1}\right) + 0.6\,\phi\left(\frac{y-6}{2}\right),$$

where $\phi(\cdot)$ is the standard normal density distribution.

Similarly, in discrete choice modeling, the motivation of mixture models has been to obtain more flexible forms. The mixture model of interest for choice modeling is a *probability mixture model*, which is a probability distribution that is a convex combination of other probability distributions.

An example of a discrete probability mixture is shown in Figure 2. This is a mixture of two binomial logit probabilities, which are explained by a constant and a continuous explanatory variable x (plotted along the x-axis). The vertical axis is the probability of choosing alternative 1. One of the binomial logits ($s = 1$) has an intercept of -1 and a coefficient of x of 1. The other binomial logit ($s = 2$) has an intercept of -6 and the the coefficient of x is 2. The mixing distribution is discrete with 40% from one logit and 60% from the other. The resulting mixed probability distribution is then

$$P(i = 1|x) = 0.4\, P(i = 1|s = 1, x) + 0.6\, P(i = 1|s = 2, x) = 0.4\, \frac{1}{1+e^{-(1+x)}} + 0.6\, \frac{1}{1+e^{-(-6+2x)}}.$$

As with the mixture of normals in Figure 1, the mixed probability (the solid line in Figure 2) is of a form that could otherwise not be captured by a standard discrete choice model.

While these are both examples of mixtures with a discrete mixing distribution, the mixing distribution can also be continuous. For example, a choice model with a normally distributed random parameter has a continuous mixing distribution, namely, the normal density distribution.

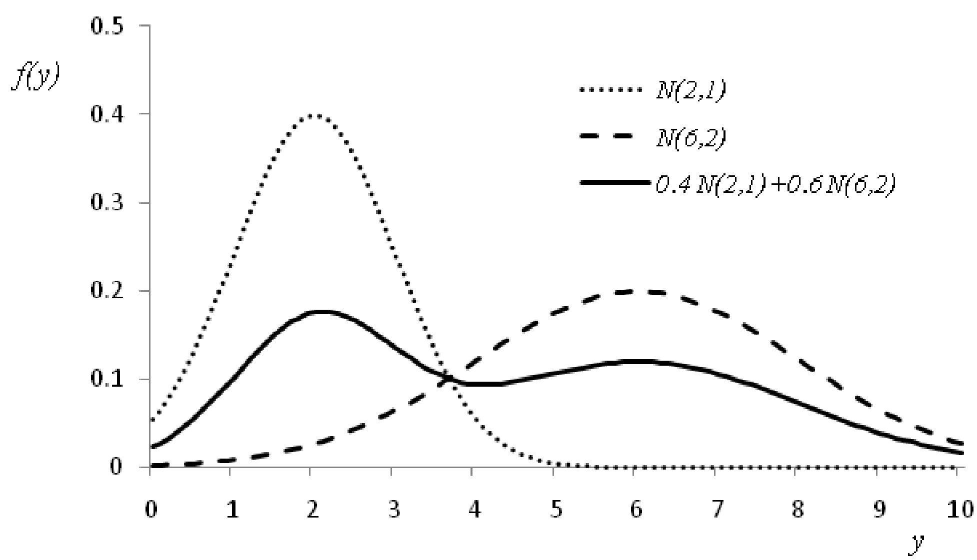

Figure 1: Discrete Mixture of Normal Distributions

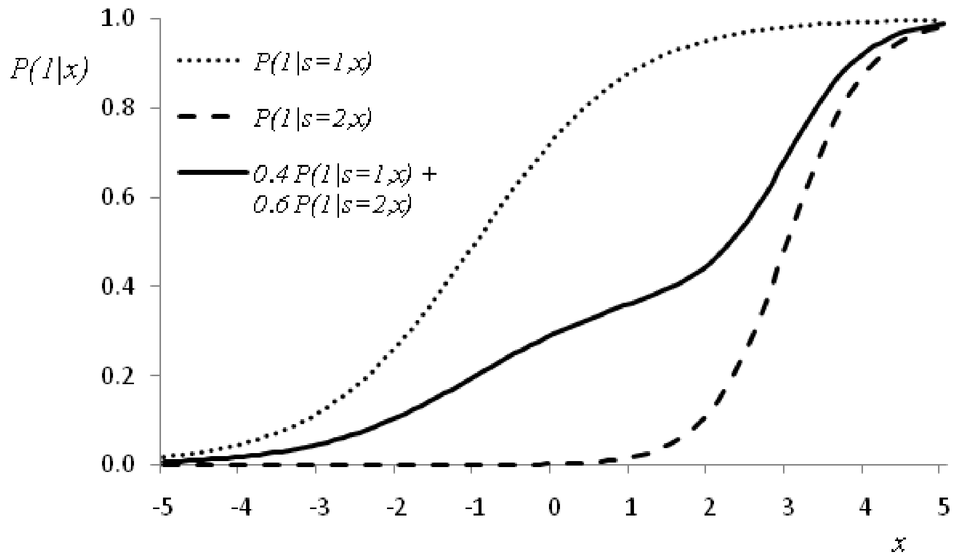

Figure 2: Discrete Mixture of Binomial Logit Probability Distribution

Early exploration of mixture models in discrete choice arose through the quest for a smooth probability simulator to estimate a probit model (McFadden, 1989, Stern, 1992, and Bolduc and Ben-Akiva, 1991, and described in Walker et al., 2007). From a behavioral perspective, an early motivation of mixture models in discrete choice was to capture heterogeneity of choice set; the logit captivity model (Ben-Akiva, 1977; Gaudry and Dagenais, 1979) is a discrete mixture of *(i)* a "captive" model in which the choice set is reduced to a single alternative and *(ii)* a compensatory logit model. Another early motivation was to account for unobserved taste heterogeneity. Boyd and Mellman (1980) and Cardell and Dunbar (1980) employed mixture models of logit on market share data and Ben-Akiva et al. (1993) on individual-level data. Other motivations include allowing for flexible substitution patterns, to capture panel effects, and to account for unobserved effects; example references for each of these are provided below.

TYPOLOGY OF MIXTURE MODELS

Probability mixture models are a broad class of model that can vary on a number of dimensions. The distribution being mixed can be any type of choice model (logit, GEV, probit, etc. and dynamic or static). The mixing distribution can be continuous or discrete, and can be specified with covariates or without. The mixing distribution can be motivated by either statistical or behavioral principles. Further, behavioral specifications can be in regard to different aspects of the choice model such as decision rules, tastes, and omitted factors. The models may employ only the choice response as an indicator of the preference (utility), or they may employ additional indicators specific to the mixing aspect. And so on.

Variation along these dimensions lead to an endless number of mixture models that can, indeed, approximate any distributional form. McFadden and Train (2000) demonstrate, for example, that a mixed logit model can approximate any random utility model to any degree of accuracy. Mixture models are typically more complex to estimate than choice models without mixing. However increases in computational power, improvement in optimization techniques, and use of simulation and alternative estimators (e.g., Bayesian) have made very complex models possible and have lead to a proliferation of their use.

In the remainder of this chapter, we will first review the most typical form of discrete and continuous mixtures found in the literature. Following this review, the next section will motivate and discuss 'behavioral' mixture models.

CONTINUOUS PROBABILITY MIXTURES

The continuous mixture choice model takes the form

$$P(i|C_n, \gamma) = \int_\theta P(i|C_n, \theta) f(\theta|\gamma) d\theta \ ,$$

where $P(i|C_n, \gamma)$ is the probability of interest (and γ the unknown parameters), $P(i|C_n, \theta)$ is the choice probability distribution being mixed (with parameters θ) and $f(\theta|\gamma)$ is the mixing distribution. This form has been widely used for a variety of purposes. The most common use is to capture random taste variation as in random parameter logit (for example, Hess et al., 2005). Other uses include modeling alternative-specific variances (for example, Gönül and Srinivasan, 1993), flexible substitution patterns (for example, Bhat, 1998), and correlations over multiple responses (for example, Revelt and Train, 1998), correlations over time (for example, Toledo, 2003) and correlations over space (for example, Vichiensan et al., 2005). Various choice model kernels have been used, most typically logit but also other GEV forms such as nested logit (for example, Dugundji and Walker, 2007). The primary specification issue is the specification of the mixing distribution. Various distributions have been used such as normal, lognormal, triangular and Johnson's SB (see, for example, the discussion in Train, 2003) and these distributions can either be independent univariate distributions over the choice model parameters or be multivariate distributions. Typically, the mixing distribution is not a function of covariates, although it can include covariates (for example, Greene et al., 2006). The primary estimation issue is the costly computation of the integral, for which simulation techniques are primarily employed (see Train, 2003).

DISCRETE PROBABILITY MIXTURES

The discrete mixture choice model takes the form

$$P(i|C_n, \gamma) = \sum_{s=1}^{S} \gamma_s P(i|C_n, \theta_s) \ ,$$

where $P(i|C_n, \gamma)$ is the probability of interest, $s = 1, \ldots, S$ denote distinct classes, the distributions being mixed are the class-specific choice probabilities, $P(i|C_n, \theta_s)$ (with unknown class-specific coefficients θ_s), and the mixing distribution γ_s are the class membership probabilities. Discrete choice mixtures are employed when there is believed to be distinct, yet unknown, classes of behavior within the population. The most basic form is shown here in which the mixing distribution is not a function of covariates. This model has been popular in market research to do market segmentation. The covariate mixture model will be presented in the next section on behavioral mixture models, and further references will be provided. The primary advantages of discrete choice mixture models are that the probability distribution does not require an integral (if it's a mixture of choice models that do not require an integral such as GEV) and it does not require a-priori specification of the distribution of the parameters.

BEHAVIORAL MIXTURE MODELS

The mixture models described above are powerful; the model forms can fit any distribution and significantly increase model fit. However, there are also issues that arise with such mixture approaches. First is that the sophisticated models of the covariance are a black box in that, while they may provide an excellent fit to the data, the cause or source of the distribution is not readily apparent. Difficulties arise from the explosion of parameters that are difficult to interpret, a tendency to overfit the data, and the lack of insight provided for policy and marketing analysis. Another issue is that using such mixture models for forecasting requires the assumption that the distributions have temporal stability, which is highly suspect.

These limitations lead to the direction of *behavioral (or structural) mixture models*. The concept of the behavioral mixture model is to provide a behavioral rational to the mixture rather than an ad-hoc distribution. This is done by modeling the covariance structure via explicit latent variable constructs as the method to capture the source of behavioral heterogeneity. That is, make the mixing distribution behavioral.

A framework for behavioral mixture models is shown in Figure 3. In addition to the latent utility, there are explicit latent constructs to capture attitudes and perceptions. An attitude is “a disposition to respond favorably or unfavorably to an object, person, institution, or event” (Ajzen, 2005). In the context of choice modeling, it is a predisposition towards performing a particular behavior. Attitudes are formed over time and are affected by social norms and past experience. Kahneman (1997) and other behavioral scientists have emphasized the importance of attitudes as the “emotional core” driving many decision-making processes. More generally in this framework, the attitude latent variable is a latent characteristic of the respondent. The perception latent variable captures attributes of the alternatives as perceived by the decision maker. The mixing distribution is thus defined by these explicit and interpretable latent constructs.

The framework also shows the use of indicators of the underlying latent variables. Just as stated and revealed preferences are indicators of latent utility, we may also have attitudinal and perceptual indicators to provide information on latent attitudes and perceptions. For example, the psychometric indicators shown in Figure 4 are indicators of underlying latent constructs. The latent variables can be either discrete or continuous, and each form is taken in turn in the next two sections of the chapter.

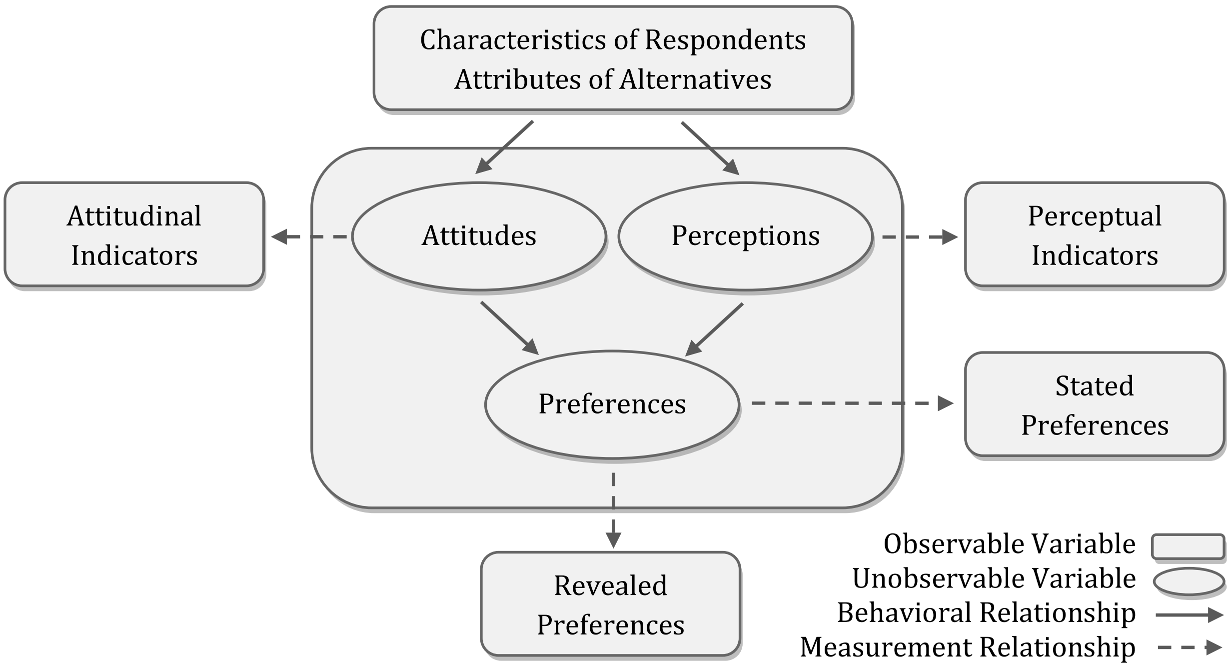

Figure 3: Conceptual Framework for Behavioral Mixture Models

Environmental Awareness

- I am willing to pay more for travel if it would help the environment.
- Use of transit can help improve the environment.

Sensitivity to Schedules

- I need to make trips to a wide variety of locations each week.
- I need the flexibility to make many trips during the day.
- I generally make the same types of trips at the same times.

Urban Form Preferences

- I am willing to travel longer to have a big house and a garden.
- I like to live within walking distance to shops and restaurants.
- I enjoy the hustle and bustle of the city.

(Responses collected on a Likert scale.)

Figure 4: Psychometric Indicators for Attitudes Toward Transportation and Urban Form

Continuous Behavioral Mixtures

Continuous behavioral mixture models refer to cases in which the latent variables are continuous constructs. With continuous constructs, the model integrates choice models with the structural and measurement models of the latent variables as shown in Figure 5.

The preferences are a function of both observed variables x and latent variables x^*. Conditioning on x and x^*, the choice probability of i is

$$P(i|x, x^*; \theta),$$

where θ is a set of unknown parameters to be estimated. The probability can take on the form of any choice model.

The latent variables are not known and must be integrated out. Therefore, one needs a distribution of the latent variables as a function of observable variables and structural parameters γ, denoted as $f_s(x^*|x; \gamma)$. This is the structural model of a latent variables model. The probability is then

$$P(i|x; \theta, \gamma) = \int_{x^*} P(i|x, x^*; \theta) f_s(x^*|x; \gamma) dx^*.$$

Thus, the mixing distribution is the structural equation model, which explicitly captures the behavioral latent constructs.

While it is possible to stop at the equation above, it is difficult to obtain identification of the unknown parameters related to the latent variables. Therefore additional measurements in the form of psychometric indicators are incorporated. This is done via measurement models. From assumptions of the form of the measurement model, we obtain the distribution of indicators as a function of the latent variables and a set of parameters α, or $f_m(I|x^*; \alpha)$. This leads to a joint likelihood of the choice and the latent variable indicators

$$P(i, I|x; \theta, \gamma, \alpha) = \int_{x^*} P(i|x, x^*; \theta) f_m(I|x^*; \alpha) f_s(x^*|x; \gamma) dx^*.$$

Additional information on this framework, including applications, can be obtained from Ben-Akiva et al. (2002a, 2002b) and Walker and Ben-Akiva (2002). McFadden (1986) presented the methodological seed for introducing latent variables into choice models. Additional applications of continuous behavioral mixtures can be found in Cambridge Systematics (1986), Train et al. (1987), Boersch-Supan et al. (1996), Ashok et al. (2002), and Morikawa et al (2002).

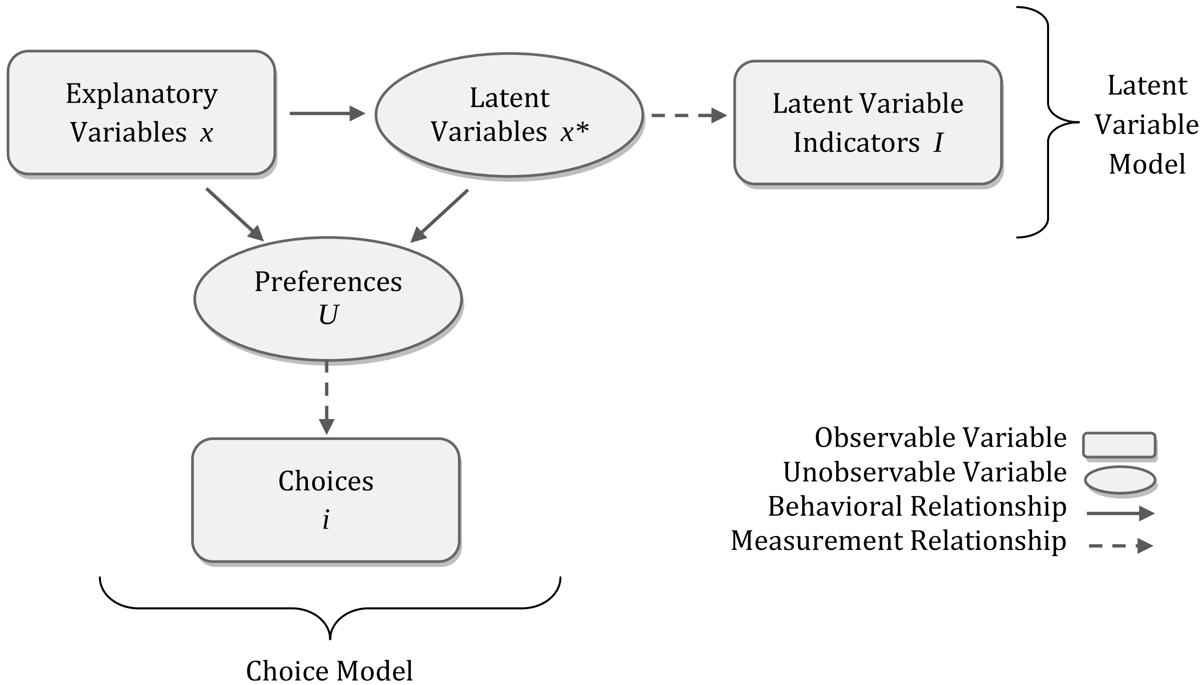

Figure 5: Behavioral Mixture Model Framework

Discrete Behavioral Mixtures

The discrete behavioral mixture is closely related to the discrete mixture model described above. The *behavioral* discrete mixture choice model is one in which the class-membership model is parameterized with socio-economic and situational data to provide causal structure to the heterogeneity. The model then takes the form

$$P(i|C_n,\gamma) = \sum_{s=1}^{S} Q_n(s|\gamma)P(i|C_n,\theta_s)\ ,$$

where $P(i|C_n,\gamma)$ is the probability of interest, $s = 1, \ldots$,S denote distinct behavioral classes, the distributions being mixed $P(i|C_n,\theta_s)$ are the class-specific choice probabilities, and the mixing distribution $Q(s|\gamma)$ are class membership probabilities parameterized by coefficients γ. The discrete behavioral mixtures are used when there is believed to be distinct, yet unknown, classes of behavior within the population. The behavior can vary across classes in terms of the choice set considered (for example, Manski, 1977; Swait and Ben-Akiva, 1987; Ben-Akiva and Boccara 1995), decision protocol (for example, Gopinath, 1995), or taste parameters (for example, Kamakura and Russell, 1987; Greene and Hensher, 2003; and Walker and Li, 2007). Among other advantages, the insights on market segmentation derived from the discrete behavioral mixture model provide a powerful analysis tool. An example of such will be provided in the empirical results section. Here we do not show the form that includes indicators of the discrete classes, and the reader is referred to Gopinath (1995).

EMPIRICAL APPLICATION

In this empirical application we demonstrate the various choice model formulations using a single dataset. For this purpose, we make use of a stated preference survey of household residential location choice decisions conducted in Portland, Oregon in 1994. The full data collection effort is described in Cambridge Systematics (1996). Each survey question asked for a preference among five hypothetical housing options, with alternatives varying across price, size, community amenities, accessibility, and several other factors influencing residential choices. An example choice experiment is provided in Figure 6. Each choice experiment consisted of the 5 alternatives shown (buy single-family, buy multi-family, rent single-family, rent multi-family, and move out of the metro area) and the list of attributes shown (type of dwelling, residence size, etc.). The fifth alternative of moving out of the metro area is a so-called "opt-out" alternative. Each individual was presented with 8 different choice experiments of the format shown in Figure 1, and the values of the attributes that describe each alternative varied across each choice experiment. There were 16 different versions (blocks) of the questionnaire, that is 16 different sets of 8 choice experiments, each appearing as in Figure 1 but with differing attributes selected based on the experimental design. The survey was administered to 611 individuals (we use a cleaned set of 507 individuals) and each survey respondent was assigned randomly to one of the 16 blocks.

We estimate 5 different models:

1. Logit
2. Nested Logit
3. Logit mixture – error component
4. Logit mixture – random parameter
5. Behavioral logit mixture – latent class choice model

The estimation results are presented in Table 1. (All results are shown except for the alternative specific constants due to lack of space.) Each model estimates the probability of choosing each of the 5 alternatives: buy single family, buy multiple family, rent single family, rent multiple family, and move out of the metropolitan area. The alternatives are described by the variables listed in the left column of the table, which have been grouped into attributes related to the house, neighborhood, and accessibility. The household characteristic of income is interacted with the price variables (either purchase price for the two buy alternatives or monthly rent for the two rent alternatives).

The "classic" models of logit and nested logit are presented first. The nested logit model is specified with the two buy alternatives in one nest, the two rent alternatives in a second nest, and the move out of the metro area in a third nest. In this case, the nested logit model is not significantly different from the logit model. (Other nesting structures also were not significant.) The logit model performs well in that a large proportion of this laundry list of attributes are significant, the signs are as expected, and relative magnitude of parameters make sense.

Next the logit mixture models are presented. First is the error component formulation, which is specified to be similar in concept to the nested logit. However, the key difference is that the error components capture the correlation across multiple responses from a single individual, which is ignored in nested logit. Further, nested logit has identical alternative specific variances, whereas the error component formulation allows for the variances of alternative utilities to differ. As the results show, these differences lead to a significant increase in fit in the error component model over the nested logit model.

Another major advantage of logit mixture model is the ability to capture heterogeneity through random parameters. The fourth model presents results of such a specification in which standard deviations are estimated for a subset of model parameters. Most parameters demonstrate unobserved heterogeneity and the fit of the model increases substantially. While the mixed logit estimation results thus far indicate substantial heterogeneity, the models are not particularly informative as to the behavioral drivers of the heterogeneity. It's the fifth model, our behavioral mixture model, that tries to address this.

Model five is a latent class choice model with three distinct behavioral classes. This model is explained in detail in Walker and Li (2007), and here described briefly to demonstrate the concept of behavioral mixtures. The behavioral driver of the model is that there are "deep-rooted and embedded, prevalent attitudes towards different types of residential areas" (Aeroe, 2001). In this context of residential location choices, such lifestyle differences lead to differences in considerations, criterion, and preferences for location. This model employs the framework of latent class choice models to infer both lifestyle groups (the latent variable) and how lifestyle preferences impact residential location decisions (the choice model). The behavioral mixture is the mixture over these lifestyle types.

The estimation results are shown in Table 1 and Table 1b. There are three lifestyle classes predicted by the model (that is, $S = 3$), where the number of classes was determined endogenously by estimating models with different number of classes. There are two sets of estimated coefficients. First is the residential choice model conditional on lifestyle, which is shown in Table 1. The coefficients of the class-specific residential choice probability vary across the three classes, which denote different trade-offs being made by the different lifestyle groups. The second set of coefficients is the class membership model, shown in Table 1b. It is a logit equation where the alternatives are the different lifestyle choices and the explanatory variables include information on household structure, employment, the age of the head of household, and resources available to the household.

Examination of the estimation results suggests the following three lifestyle segments (where the description of the lifestyle is interpreted from the class-specific parameters in Table 1 and the character of members in the class is interpreted from Table 1b):

Lifestyle 1: Households that are suburban, auto, and school oriented. These tend to be affluent, more established families.

Lifestyle 2: Households that are transit oriented and also prefer a suburban setting. These tend to be less affluent, younger families and non-families.

Lifestyle 3: Households that are urban and auto oriented. These tend to be older, non-family professionals.

While the socio-economic variables related to lifecycle are shown to have significant explanatory power, there remained significant aspects of lifestyle preferences that could not be explained by such observable explanatory variables. This is an indication that the latent variable construct is critical; the variation in preferences cannot be captured systematically.

In this empirical demonstration, all three mixture models resulted in significant improvements of fit over the conventional logit and nested logit models. Further, the behavioral mixture model provides the best fit to the data. However, this is not always the case and, indeed, is not the most important advantage. (Note that these data could not support additional correlation parameters in the random parameter model beyond what is reported here.) A larger issue is that the covariance structures of the error component and random parameter models do not provide much insight into the behavior. On the other hand, the behavioral mixture provides a behavioral rationale for the mixture through the concept of latent lifestyles. This behavioral structure provides a better foundation for model development (for example, rational to search for more parsimonious models), policy and marketing analysis (for example, how to influence the behavior of the different classes), and forecasting (for example, in dealing with issue of temporal stability).

FUTURE DIRECTIONS

Enabled by computer processing power and use of simulation and mixture methods, discrete choice models have evolved considerably to where very complex formulations of the covariance structure can be captured. In this chapter, we describe this evolution from underlying behavioral principles, to basic IIA choice models, to probit and GEV, to mixture models. We conclude with the need to further pursue the concept of behavioral mixture models, which provide a behavioral rational to the mixing structure through explicit latent variable constructs. Rather than just providing a better fit to the data, these models have potential to provide greater temporal stability and behavioral insight for policy and marketing analysis.

REFERENCES

Aeroe T (2001) *Residential Preferences, Choice of Housing, and Lifestyle*. PhD dissertation (English Summary), Aalborg University.

Ajzen I (2005) *Attitudes, Personality, and Behavior (2nd Edition)*. Milton-Keynes, England: Open University Press, McGraw-Hill.

Ashok K, Dillon WR, Yuan S (2002) Extending Discrete Choice Models to Incorporate Attitudinal and Other Latent Variables. *Journal of Marketing Research* 39, 31-46.

Ben-Akiva M, McFadden D, Train K, Walker J, Bhat C, Bierlaire M, Bolduc D, Boersch-Supan A, Brownstone D, Bunch D, Daly A, de Palma A, Gopinath D, Karlstrom A, Munizaga M (2002a) Hybrid Choice Models: Progress and Challenges. *Marketing Letters* 13(3), 163-175.

Ben-Akiva M, Walker J, Bernardino A, Gopinath D, Morikawa T, Polydoropoulou A (2002b) Integration of Choice and Latent Variable Models. In *In Perpetual Motion: Travel Behaviour Research Opportunities and Application Challenges* (Mahmassani H, ed.), Elsevier Science, 431-470.

Ben-Akiva ME, Boccara B (1995) Discrete Choice Models with Latent Choice Sets. *International Journal of Research in Marketing* 12, 9–24.

Ben-Akiva M, Bolduc D, Bradley M (1993) Estimation of Travel Model Choice Models with Randomly Distributed Values of Time. *Transportation Research Record* 1413, 88–97.

Ben-Akiva ME, Lerman SR (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, MA.

Ben-Akiva M, François B (1983) μ Homogeneous Generalized Extreme Value Model. Working paper, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Ben-Akiva M (1977) Choice Models with Simple Choice Set Generating Processes. Working Paper, Deptartment of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Bhat CR (1998) Accommodating Flexible Substitution Patterns in Multi-dimensional Choice Modeling: Formulation and Application to Travel Mode and Departure Time Choice. *Transportation Research B* 32(7), 455-466.

Boersch-Supan A, McFadden DL, Schnabel R (1996) Living Arrangements: Health and Wealth Effects. In *Advances in the Economics of Aging* (Wise DA, ed.), The University of Chicago Press, 193-216.

Bolduc D, Ben-Akiva M (1991) A Multinomial Probit Formulation for Large Choice Sets. Proceedings of the Sixth International Conference on Travel Behavior 2, 243-258.

Boyd J, Mellman J (1980) The Effect of Fuel Economy Standards on the U.S. Automotive Market: A Hedonic Demand Analysis. *Transportation Research A* 14,367–378.

Cambridge Systematics Inc (1996) Data Collection in the Portland, Oregon Metropolitan Area, Travel Model Improvement Program, Track D Data Research Program. Prepared for the U.S. Department of Transportation.

Cambridge Systematics Inc (1986) *Customer Preference and Behavior Project Report*. Prepared for the Electric Power Research Institute.

Cardell S, Dunbar F (1980) Measuring the Societal Impacts of Automobile Downsizing. *Transportation Research A* 14, 423–434.

Dugundji ER, Walker JL (2005) Discrete Choice with Social and Spatial Network Interdependencies. *Transportation Research Record* 1921, 70-78.

Gaudry MJI, Dagenais MG (1979) The Dogit Model. *Transportation Research* 13B (2), 105–112.

Gopinath, DA (1995) *Modeling Heterogeneity in Discrete Choice Processes: Application to Travel Demand*. Ph.D. Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Gönül F, Srinivasan K (1993) Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues. *Marketing Science* 12(3), 213-229.

Greene WH, Hensher DA, Rose J (2006) Accounting for Heterogeneity in the Variance of Unobserved Effects in Mixed Logit Models. *Transportation Research Part B* 40, 75-92.

Greene WH, Hensher DA (2003) A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit. *Transportation Research Part B* 37, 681-698.

Hess S, Bierlaire M, Polak JW (2005) Estimation of Value of Travel-time Savings Using Mixed Logit Models. *Transportation Research Part A* 39(2-3), 221-236.

Kahneman D (1997) Economists Have Preferences, Psychologists Have Attitudes. Presented at NSF Symposium on Preference Elicitation, UC Berkeley, Berkeley, CA.

Kamakura WA, Russell GJ (1989) A Probabilistic Choice Model for Market Segmentation and Elasticity Structure. *Journal of Marketing Research* 25, 379-390.

Koppelman FS, Sethi V (2000) Closed-form Discrete-choice Models. In *Handbook of Transport Modeling* (Hensher DA, Button KJ, eds.), Emerald Group Publishing. 211-227.

Lancaster K (1966) A New Approach to Consumer Theory. *Journal of Political Economy* 74, 132-157.

Luce RD (1959) *Individual Choice Behavior*. Wiley, New York.

Manski C (1977) The Structure of Random Utility Models. *Theory and Decision* 8, 229-254.

Marschak J (1960) Binary Choice Constraints on Random Utility Indicators. In *Stanford Symposium on Mathematical Methods in the Social Sciences* (Arrow K, ed.), Stanford University Press, Stanford.

McFadden D, Train K (2000) Mixed MNL Models of Discrete Response. *Journal of Applied Econometrics* 15, 447–470.

McFadden D (1989) A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica* 57(5), 995-1026.

McFadden D (1986) The Choice Theory Approach to Marketing Research. *Marketing Science* 5(4), 275-297.

McFadden D (1978) Modeling the Choice of Residential Location. In *Spatial Interaction Theory and Residential Location* (Karlqvist A, Lundqvist L, Snickars F, Weibull J, eds.), North-Holland, Amsterdam, 75–96.

McFadden D (1974) Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers of Econometrics* (Zarembka P, ed.), Academic Press, 105-142.

Morikawa T, Ben-Akiva M, McFadden D (2002) Discrete Choice Models Incorporating Revealed Preferences and Psychometric data. *Econometric Models in Marketing, Advances in Econometrics* 16, Elsevier, Amsterdam, 29-55.

Revelt D, Train K (1998) Mixed Logit with Repeated Choices: Households' Choice of Appliance Efficiency Level. *Review of Economics and Statistics* 80(4), 647-657.

Stern S (1992) A Method for Smoothing Simulated Moments of Discrete Probabilities in Multinomial Probit Models. *Econometrica* 60(4), 943-952.

Swait JD, Ben-Akiva M (1987) Incorporating Random Constraints in Discrete Models of Choice Set Generation. *Transportation Research B* 21(2), 91-102.

Thurstone L (1927) A Law of Comparative Judgement. *Psychological Review* 34, 273-286.

Toledo T (2003) *Integrating Driver Behavior Modeling*. Ph.D. Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Train K (2003) *Discrete Choice Methods with Simulation*. Cambridge University Press.

Train K, McFadden D, Goett A (1987) Consumer Attitudes and Voluntary Rate Schedules for Public Utilities. *Review of Economics and Statistics* 69, 383–391.

Vichiensan V, Miyamoto K, Tokunaga Y (2005) Mixed Logit Modeling Framework with Structuralized Spatial Effects: A test of Applicability with Area Unit Systems in Location Choice Analysis. *Journal of the Eastern Asian Society for Transportation Studies* 6, 3789-3802.

Walker JL, Ben-Akiva M, Bolduc D (2007) Identification of Parameters in Normal Error Component Logit-Mixture (NECLM) Models. *Journal of Applied Econometrics* 22, 1095-1125.

Walker JL, Li J (2007) Latent Lifestyle Preferences and Household Location Decisions. *Journal of Geographical Systems* 9(1), 77-101.

Walker J, Ben-Akiva M (2002) Generalized Random Utility Model. *Mathematical Social Sciences* 43(3), 303-343.

Figure 6: Choice Experiment Example

	(Alternative 1)	(Alternative 2)	(Alternative 3)	(Alternative 4)	(Alt. 5)
	Buy Single Family	Buy Multi-Family	Rent Single Family	Rent Multi-Family	Move out of the Metro Area
Type of Dwelling :	*single house*	*apartment*	*duplex / row house*	*condominium*	
Residence Size :	*< 1,000 sq. ft.*	*500-1,000 sq. ft.*	*1,500 - 2,000 sq. ft.*	*< 500 sq. ft.*	
Lot Size :	*< 5,000 sq. ft.*	*n/a*	*5,000 - 7,500 sq. ft.*	*n/a*	
Parking :	*street parking only*	*street parking only*	*driveway, no garage*	*reserved, uncovered*	
Price or Monthly Rents :	*< $75K*	*$50K - $100K*	*> $1,200*	*$300 - $600*	
Community Type :	*mixed use*	*mixed use*	*rural*	*urban*	
Housing Mix :	*mostly single family*	*mostly multi-family*	*mostly multi-family*	*mostly multi-family*	
Age of Development :	*10-15 years*	*0-5 years*	*10-15 years*	*0 - 5 years*	
Mix of Residential Ownership :	*mostly own*	*mostly own*	*mostly rent*	*mostly own*	
Shops/Services/Entertainment :	*community square*	*basic shops*	*community square*	*basic, specialty shops*	
Local Parks :	*none*	*yes*	*none*	*none*	
Bicycle Paths :	*none*	*yes*	*yes*	*yes*	
School Quality :	*very good*	*very good*	*fair*	*fair*	
Neighborhood Safety :	*average*	*average*	*average*	*average*	
Shopping Prices Relative to Avg :	*20% more*	*20% more*	*same*	*10% more*	
Walking Time to Shops :	*20-30 minutes*	*20-30 minutes*	*< 10 minutes*	*10 - 20 minutes*	
Bus Fare, Travel Time to Shops :	*$1.00, 15-20 minutes*	*$1.00, > 20 minutes*	*$0.50, 5 - 10 minutes*	*$0.50, < 5 minutes*	
Travel Time to Work by Auto :	*> 20 minutes*	*15-20 minutes*	*15 - 20 minutes*	*< 10 minutes*	
Travel Time to Work by Transit :	*> 45 minutes*	*30-45 minutes*	*30 - 45 minutes*	*15 - 30 minutes*	

Table 1a: Estimation Results

		Model 1		Model 2		Model 3		Model 4				Model 5							
		Logit		Nested Logit		Logit Mixture 1 Error Component		Logit Mixture 2 Random Parameter				Behavioral Logit Mixture							
								Mean		Standard Dev.		Class Independent		Class1 "Suburban"		Class2 "Transit"		Class3 "Urban"	
	Variable	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat
Housing Attributes	Monthly rent ($00) - low/middle income	-0.135	**-13.0**	-0.126	**-10.0**	-0.162	**-13.8**	-0.212	**-11.4**	0.107	**6.0**	-0.152	**-11.5**						
	Monthly rent ($00) - high income	-0.077	**-5.3**	-0.069	**-4.1**	-0.051	**-2.9**	-0.122	**-3.2**	0.149	**3.8**	-0.079	**-3.6**						
	Monthly rent ($00) - income not available	-0.136	**-7.7**	-0.128	**-6.8**	-0.160	**-6.6**	-0.171	**-5.4**	0.021	0.5	-0.160	**-5.0**						
	Purchase price ($000) - low income	-0.905	**-13.4**	-0.853	**-10.9**	-1.192	**-13.7**	-1.520	**-9.3**	0.846	**5.4**	-1.007	**-9.8**						
	Purchase price ($000) - middle income	-0.617	**-11.2**	-0.573	**-8.9**	-0.719	**-10.7**	-0.925	**-11.2**	0.380	**5.1**	-0.928	**-11.2**						
	Purchase price ($000) - high income	-0.307	**-4.7**	-0.272	**-3.9**	-0.249	**-3.0**	-0.310	**-2.5**	0.488	**1.7**	-0.534	**-4.8**						
	Purchase price ($000) - income not available	-0.582	**-7.2**	-0.537	**-6.1**	-0.644	**-5.7**	-0.780	**-4.9**	0.582	**4.8**	-0.783	**-4.4**						
	Single house (v. Duplex)	0.323	**5.9**	0.293	**4.9**	0.382	**6.3**	0.428	**6.1**					*0.503*	***4.3***	*0.840*	***5.8***	*-0.318*	***-2.1***
	Condo (v. Apartment)	0.162	**2.3**	0.143	**2.1**	0.170	**2.3**	0.223	**2.6**					0.302	1.7	0.468	**2.0**	0.036	0.3
	Residential size (square feet/1000)	0.380	**7.7**	0.352	**6.7**	0.440	**8.3**	0.485	**7.2**	0.630	**9.5**			*1.377*	***12.9***	*-0.335*	***-2.4***	*0.049*	*0.4*
	Lot size (square feet/1000)	0.008	1.3	0.008	1.4	0.008	1.2	-0.001	-0.1	0.092	**10.2**			*0.009*	*0.8*	*0.059*	***3.7***	*-0.052*	***-3.4***
Neighborhood Attributes	Mostly owners (v. Mostly renters)	0.151	**3.4**	0.141	**3.4**	0.181	**3.8**	0.183	**3.4**					*0.226*	***2.3***	*-0.070*	*-0.6*	*0.278*	***2.7***
	Mostly multi-family housing (v. Mostly single family)	-0.026	-0.6	-0.027	-0.7	-0.041	-0.9	-0.056	-1.1					*-0.179*	*-1.9*	*0.204*	*1.6*	*-0.126*	*-1.3*
	Schools - 75 percentile (v. below 60)	0.268	**5.0**	0.246	**4.6**	0.190	**2.8**	0.391	**5.4**					*0.618*	***4.3***	*0.381*	***2.3***	*0.174*	*1.4*
	Schools - 60-75 percentile (v. below 60)	0.166	**2.7**	0.155	**2.6**	0.301	**5.1**	0.243	**3.1**					0.336	**2.3**	0.294	1.6	0.029	0.2
	Above average safety (v. average)	0.103	**2.5**	0.102	**2.6**	0.127	**2.8**	0.141	**2.8**					*0.226*	***2.4***	*-0.235*	*-1.9*	*0.295*	***2.9***
	Mixed use (v. Rural)	0.057	0.9	0.054	1.0	0.073	1.1	0.064	0.8					0.133	1.0	-0.160	-1.0	0.261	1.8
	Urban (v. Rural)	0.013	0.2	0.020	0.3	0.040	0.6	0.050	0.6					*0.000*	*0.0*	*-0.271*	*-1.6*	*0.407*	***2.8***
	Suburban (v. Rural)	-0.075	-1.2	-0.066	-1.1	-0.062	-0.9	-0.091	-1.2					-0.199	-1.4	-0.128	-0.7	0.106	0.7
	Local bike path (v. no local bike path)	0.079	1.8	0.069	1.7	0.083	1.8	0.118	**2.2**					*-0.100*	*-1.0*	*0.415*	***3.2***	*0.135*	*1.3*
	Local park (v. no local park)	0.020	0.5	0.016	0.4	0.021	0.5	0.051	1.0					0.073	0.8	0.154	1.2	-0.110	-1.2
	Local community square (v. no shops)	0.190	**3.2**	0.175	**3.1**	0.213	**3.3**	0.216	**2.9**					0.301	**2.1**	0.115	0.7	0.240	1.6
	Basic plus specialty shops (v. no shops)	0.146	**2.4**	0.144	**2.5**	0.173	**2.6**	0.131	**1.6**					*0.453*	***3.1***	*-0.540*	***-2.8***	*0.374*	***2.7***
	Basic shops (v. no shops)	0.129	**2.1**	0.116	**2.0**	0.153	**2.3**	0.148	**1.9**					*0.198*	*1.4*	*-0.170*	*-1.0*	*0.404*	***2.8***
Transport Attributes	Walk time to local shops (minutes)	-0.007	**-3.7**	-0.007	**-3.7**	-0.009	**-4.0**	-0.011	**-4.5**	0.019	**4.7**			*-0.010*	***-2.3***	*0.006*	*1.0*	*-0.019*	***-4.1***
	Travel time to work by auto (minutes)	-0.004	-0.9	-0.004	-0.9	-0.001	-0.2	-0.005	-0.8	0.015	1.3			*-0.015*	*-1.5*	*0.029*	***2.1***	*-0.014*	*-1.1*
	Travel time to work by transit (minutes)	-0.006	**-2.6**	-0.005	**-2.5**	-0.006	**-2.6**	-0.006	**-2.4**	0.014	**3.8**			*-0.003*	*-0.6*	*-0.021*	***-3.5***	*0.006*	*1.1*
	Off street parking available (v. no off street parking)	0.364	**6.3**	0.331	**5.3**	0.427	**6.9**	0.484	**6.4**					0.633	**5.2**	0.333	**2.1**	0.408	**3.0**
Correlation Terms	Nesting parameter - Buy alternatives			1.150	1.0 vs 1														
	Nesting parameter - Rent alternatives			1.130	0.8 vs 1														
	Nesting parameter - Move out alternative			1.000	fixed														
	Standard deviation - Buy alternatives *(σB)*					0.840	**7.0**			0.519	**3.7**	0.932	**6.4**						
	Standard deviation - Rent alternatives *(σR)*					1.182	**10.8**			1.200	**9.4**	1.251	**9.3**						
	Standard deviation - Move out alternative *(σM)*					2.178	**20.5**			2.740	**16.3**	2.166	**16.5**						
	Number of observations	4056		4056		4056		4056				4056							
	Number of parameters	34		36		37		49				115							
	Null Log-likelihood	-6527.88		-6527.88		-6527.88		-6527.88				-6527.88							
	Final Log-likelihood	-5729.35		-5728.48		-5044.40		-4917.74				-4796.64							
	Rho-square	0.122		0.122		0.227		0.247				0.265							
	Adjusted rho-square	0.117		0.117		0.222		0.239				0.248							

Table 1b: Estimation results (continued)

		Model 5 continued Behavioral Logit Mixture - Class Membership					
		Class 1 (43%)		Class 2 (30%)		Class 3 (27%)	
		Affluent, more established families		Less affluent, younger families and non-families		Older, non-family, professionals	
	Variable	Coef	t-stat	Coef	t-stat	Coef	t-stat
	Intercept	-1.673	-1.9	1.513	1.6	0.161	0.2
Household Structure	Number of children under 5 years old	1.273	1.0	1.431	1.1	-2.704	-1.0
	Number of children from 5 to 11	*0.426*	***2.4***	*-0.117*	*-0.5*	*-0.309*	*-1.2*
	Number of children from 12 to 17	0.300	1.1	0.097	0.4	-0.398	-1.0
	Number of persons 18 and over	0.444	1.5	-0.163	-0.5	-0.282	-0.9
	Non-family dummy	-0.706	-1.7	0.124	0.3	0.582	1.7
Employment	Number of employed persons	-0.096	-0.5	-0.048	-0.2	0.145	0.5
	Number of retired persons	-0.480	-1.4	0.076	0.2	0.404	1.3
	Dummy if at least one "manager/professional"	*0.131*	*0.6*	*-0.571*	***-2.1***	*0.440*	*1.6*
	Maximum number of work hours	-0.002	-0.3	-0.004	-0.4	0.006	0.9
Age of the Head of Household	Piecewise linear age of HOH: age 20-35	-0.030	-0.6	0.040	0.9	-0.010	-0.2
	Piecewise linear age of HOH: age 36-60	*0.020*	*1.1*	*-0.058*	***-2.7***	*0.037*	*1.8*
	Piecewise linear age of HOH: age 60 plus	0.035	0.9	0.003	0.1	-0.038	-1.5
Resources	*Dummy for medium income*	*1.760*	***3.4***	*-0.904*	*-1.9*	*-0.855*	***-2.2***
	Dummy for high income	*1.607*	***4.4***	*-1.026*	***-3.7***	*-0.581*	*-1.9*
	Dummy for income not reported	*2.183*	***4.6***	*-2.151*	***-3.7***	*-0.032*	*-0.1*

Italics indicates parameters that vary significantly across classes (Wald statistic, 90% confidence).