Mixed Logit Models in Political Science

Garrett Glasgow
University of California, Santa Barbara

Presented at the Eighteenth Annual Political Methodology Conference
Emory University, July 19, 2001

Abstract

Mixed logit (MXL) is a general discrete choice model thus far unexamined in the study of multinomial political science problems. Mixed logit assumes the unobserved portions of utility are a mixture of an IID extreme value term and another multivariate distribution selected by the researcher. This general specification allows MXL to avoid imposing the independence of irrelevant alternatives (IIA) property on the choice probabilities. Further, MXL is a flexible tool for examining heterogeneity in behavior through random-coefficients specifications. Mixed logit is a more general discrete choice model than multinomial probit (MNP) in several respects, and can be applied to a wider variety of questions than MNP. Empirical examples examining voter substitution between parties in the 1987 British general election, heterogeneity in the influence of union membership on vote choice in the 1992 US presidential election, and heterogeneity in the career decisions of US Representatives over time demonstrate the utility of mixed logit in political science applications.

I thank Jim Adams, R. Michael Alvarez, Andy Baker, Fred Boehmke, Scott Desposato, Rod Kiewiet, Gary King, Eric Lawrence, Jonathan Nagler, Phil Paolino, Kevin Quinn, Kenneth Train, Chris Winship, Langche Zeng, and two anonymous reviewers for "Mixed Logit Models for Multi-party Elections" for their helpful comments.

1 Introduction

In this paper I describe the mixed logit (MXL), a flexible discrete choice model based on random utility maximization, and discuss its applicability to the study of multiparty elections.[1] Mixed logit models have seen application in marketing and transportation research (Algers, Bergström, Dahlberg, and Dillén 1999; Bhat 1998a, 1998b; Brownstone and Train 1999; Jain, Vilcassim, and Chintagunta 1994; Revelt and Train 1998; Train 1998), but have seen only limited application in political science (Glasgow 2001).

Mixed logit is a discrete choice model that estimates the structure of the error term. In this respect MXL is identical to multinomial probit (MNP), a discrete choice model that is more familiar to political scientists. In the study of multicandidate and multiparty elections MNP is an increasingly popular choice (Alvarez and Nagler 1995, 1998a, 1998b; Lacy and Burden 1999, 2000; Lawrence 1997; Quinn, Martin, and Whitford 1999; Schofield, Martin, Quinn, and Whitford 1998). The primary motivation for using MNP in the study of multiparty elections is a desire to avoid the independence of irrelevant alternatives (IIA) property. Models which assume IIA, such as multinomial logit (MNL), assume that the ratio of the probability of voting for party A over the probability of voting for party B remains unchanged when party C enters or leaves the election. MNP relaxes this assumption, and allows researchers to estimate how voters view parties as similar or different. This is an important point when estimating substitution patterns (how voters will react to a candidate or party entering or leaving an election).

Mixed logit also relaxes the IIA assumption, and like MNP can estimate substitution patterns that account for the unobserved similarities and differences of candidates and parties. Both models accomplish this by assuming a particular structure for the unobserved portions of utility and estimating components of this structure. In most political science applications the unobserved characteristics of candidates or parties are assumed to induce correlations across alternatives — however, this is only one of a wide variety of structures that can be placed on the unobserved portions of utility. Both MNP and MXL allow for many different specifications of the unobserved portions of utility, with MXL allowing a greater range of specifications than MNP. Thus, MXL is a more general discrete choice model than MNP in several respects.

Both MNP and MXL are based on a theory of random utility maximization. These models are based on the assumption that a voter's preferences among alternatives in the choice set can be described by a utility function. This utility function depends on the attributes of the alternatives and the characteristics of the individual. When chosing an alternative, individuals select the alternative that yields the highest utility.

The utility function is represented as the sum of two components — a systematic component that depends on the observed attributes of the alternatives and the characteristics of the individuals, and a stochastic component that represents the influence of unobserved factors on an individual's choice. The utility yielded by party j to individual i can be represented by:

[1] Mixed logit models are also known as "random parameters logit", "mixed multinomial logit", "random coefficients logit", and "error components logit".

$$U_{ij} = V_{ij} + e_{ij} \tag{1}$$

where V_{ij} represents the systematic (observed) portion of utility, and ε_{ij} represents the stochastic (unobserved) portion of utility. The probability that individual i selects alternative j is the probability that the utility for alternative j exceeds the utility of all other alternatives:

$$P_{ij} = \Pr(U_{ij} > U_{ik}) \, \forall K = \Pr(V_{ij} + e_{ij} > V_{ik} + e_{ik}) \, \forall K \tag{2}$$

Since utility depends in part on some unobserved factors, it is not possible to say with certainty which alternative an individual will choose. Instead, random utility models make assumptions about the distribution of the stochastic portion of utility and calculate the probability that an individual will select each alternative by estimating a discrete choice model. The type of discrete choice model estimated depends on the assumptions made about the distribution of ε_{ij}.

Multinomial logit assumes that the unobserved portions of utility ε_{ij} are identically and independently distributed (IID) in accordance with the extreme value distribution.[2] It is well known that the choice probabilities of MNL have the IIA property. For each individual, the ratio of the choice probabilities of any two alternatives is independent of the utility of any other alternatives. This property is not unique to MNL — IIA will hold in *any* discrete choice model that assumes the unobserved portions of utility are IID. IIA can lead to unrealistic estimates of individual behavior when alternatives are added to or deleted from the choice set (unrealistic substitution patterns). Most advances in the empirical modeling of multiparty and multicandidate elections in recent years have focused on relaxing the restrictiveness of the IIA property and estimating more natural substitution patterns. However, it is important to note that when IIA is an inappropriate assumption the model assumption that has actually been violated is that the stochastic portion of utility is IID.

There are reasons to account for IID violations in discrete choice models beyond estimating more realistic substitution patterns. The sources of the IID violations themselves are often of great substantive interest. In a sense, the ultimate goal in empirical modeling is to completely specify the relationship between the dependent and independent variables, and reduce the remaining error term to random noise. Thus, a preferable approach to models that account for violations of the IID assumption would be to explicitly model the sources of these violations, and reduce the remaining error to something that is irrelevant, and offers no information on substitution patterns or preferences for alternatives (Horowitz 1991). For example, observing and modeling what leads voters to view two candidates as similar is preferable to accounting for these similarities through the specification of the unobserved portion of utility. However, in many cases the relevant variables will be unknown or unobservable, and our only recourse will be to account for these variables as IID violations in the unobserved portion of utility.

[2] I use the term "multinomial logit" generically, intending it to refer to both multinomial logit (with variables specific to the individual) and conditional logit (with variables specific to the alternatives, and possibly the individual as well). The general form of both models is identical, with the differences between the two models arising through the choice of variables included (Maddala 1983). Referring to this class of models as "multinomial logit" brings this terminology into line with the accepted terminology for the equivalent probit model, which is referred to as "multinomial probit" regardless of what kind of variables are included.

Three possible violations of the IID assumption will be considered here. One is when alternatives share unobserved attributes that influence choice. These unobserved attributes cause correlation in the unobserved portion of utility across alternatives, leading individuals to violate the IID assumption. This violation is known as *common unobserved attributes.* Most MNPs in political science are specified to account for common unobserved attributes — for example, the common unobserved attributes that might lead a third candidate entering a US presidential election to be seen as a substitute for one of the two original candidates. The other violation of IID considered here occurs when unobserved characteristics of the individual influence how observed characteristics of the individual and attributes of the alternatives affect choice. For example, voters with the same observed characteristics may place different weights on the issue positions of candidates or parties. For some individuals, candidate or party positions on a particular issue may have a great influence on their vote choice, while for others this influence may be minimal or non-existent. Each individual places their own particular weight on these issue positions, which leads to correlation across the utility of alternatives for each individual and again leads individuals to violate the IID assumption. This violation is known as *random taste variation*, since "tastes" for the either the attributes of alternatives or for the relationship between individual characteristics and alternatives vary randomly across individuals. Finally, the IID assumption is also violated when two or more observed choices depend on common unobserved characteristics of individuals, causing correlation in the utility for alternatives over different choice sets. This violation is known as *unobserved heterogeneity.* This is primarily a problem in panel data. For example, an individual might have a predisposition to favor certain kinds of candidates for an unobserved reason. In a single-choice situation this would just be a part of the error term, but in a repeated choice situation this would lead to correlation between the choices made by this individual over time.

Both multinomial probit and mixed logit relax the assumption that the unobserved portions of utility are IID, and both can therefore address violations of IID that arise due to common unobserved attributes and random taste variation. Multinomial probit relaxes the IID assumption by specifying the distribution of the unobserved portions of utility as multivariate normal with a general covariance matrix. This allows MNP to estimate the heteroskedasdicity and correlation of the unobserved portions of utility through the covariance matrix of this multivariate normal distribution. MXL relaxes the IID assumption by specifying the unobserved portions of utility as a combination of the IID extreme value term of the multinomial logit and another distribution g that can take any form. Models of this form are called "mixed logit" because the choice probability for an individual is a mixture of multinomial logit models, with g as the mixing distribution. MXL is able to estimate the heteroskedasdicity and correlation of the unobserved portions of utility through the parameters that describe this general distribution. This specification is more general in its treatment of the unobserved portions of utility than MNP, which requires the unobserved portions of utility be distributed multivariate normal and estimates the covariance matrix of the unobserved portions of utility. Thus, MXL has advantages over MNP in many applications, whether the IID violations are due to unobserved attributes or random taste variation.

In section 2 I describe how multinomial probit and mixed logit can both be derived from a common utility framework. When presented in this way it is apparent that MNP and MXL (and most other discrete choice models familiar to political scientists) are in fact identical in basic structure, with the only differences arising through different assumptions about the distribution of the unobserved portions of utility. MXL is revealed to be a more general model than MNP in several respects, since it places fewer restrictions on the assumptions that can be made about the unobserved portions of utility. Section 3 presents three empirical examples. The first example is an estimate of substitution patterns between parties competing in the 1987 British general election. There it is shown that MXL can be specified in much the same way as MNP, and used to determine if certain alternatives are viewed as similar by individuals. The second example studies heterogeneity in the impact of union membership on presidential vote in the 1992 US presidential election. The third example studies Congressional career decisions, specifying correlations over time. Section 4 concludes, discussing the utility of mixed logit in political science research.

2 Specification of the Mixed Logit and Multinomial Probit Models

Assume an individual faces a choice set consisting of J alternatives. Let the utility that individual i receives from alternative j be denoted by U_{ij}, which is the sum of a linear-in-parameters systematic component V_{ij} and a stochastic component e_{ij}. Rewrite the systematic component of utility as $V_{ij} = x_{ij}\beta_j$, where x_{ij} is a vector of characteristics unique to alternative j relative to individual i, unique to individual i, or both. β_j is a vector of parameters to be estimated which are either fixed over individuals and alternatives, or vary over alternatives for those elements in x_{ij} unique to individual i. Rewrite the stochastic component of utility as $e_{ij} = z_{ij}\eta_i + \varepsilon_{ij}$, where z_{ij} is a vector of characteristics that can vary over individuals, alternatives, or both (z_{ij} and x_{ij} can have some or all elements in common). ε_{ij} is a random term with mean 0 that is IID over individuals and alternatives, and is normalized to set the scale of utility. η_i is a vector of random terms with mean 0 that varies over individuals according to the distribution $g(\eta|\Omega)$, where Ω are the fixed parameters of the distribution g. We then write the utility that individual i gets from alternative j as $U_{ij} = x_{ij}\beta_j + (z_{ij}\eta_i + \varepsilon_{ij})$. Stacking the utilities yields:[3]

$$U = X\beta + (Z\eta + \varepsilon) \tag{3}$$

If IIA holds, $\eta = 0$ for all i, so U depends only on the systematic portion of utility and an IID stochastic portion of utility. Discrete choice models that assume IIA do not estimate $Z\eta$, implicitly assuming $\eta = 0$. However, by considering the impact of the term $Z\eta$ on utility, discrete choice models can be specified that are able to consider the effects of unobserved attributes and random taste variation, and thus avoid the IIA assumption. These models will estimate Ω (the parameters of the distribution of η) as well as β.

If the elements of Z are also contained in X this is a *random-coefficients* model. In this instance the appropriate elements in the vector β give the mean values for the random coefficients (the coefficients on those variables contained in both Z and X), while Ω gives the other parameters of the distribution of the random coefficients (such as the variance). Random-coefficients models are usually specified to examine random taste variation, allowing for the study of heterogeneity in the impact of the independent variables on the dependent variable. This is an underexplored area of research in voting behavior — many studies have established the *mean* relationships of numerous variables to vote choice, but little is known about the *heterogeneity* of those relationships. One

[3] I will generally drop subscripts from the notation from here on unless they are necessary to avoid confusion.

exception is a study by Rivers (1988), who examined heterogeneity in voter behavior by estimating separate coefficients for each individual in the dataset. In a sense the random-coefficients models described here, which estimate the parameters of the distribution of each random coefficient, are a compromise between the approach of Rivers and the approach of most other models, which assume all coefficients in the model are identical for all individuals.

If the elements of Z are not contained in X this is an *error-components* model. The elements of Z are assumed to be error components that introduce heteroskedasdicity and correlation across alternatives in the unobserved portion of utility. Error-components models are usually specified to estimate more realistic substitution patterns, and can do this through considering either random taste variation or unobserved attributes. In many cases these error components are simply random coefficients which are assumed to have a mean 0, thus examining the effect of random taste variation on substitution patterns. If the elements of Z are set as constants across individuals, but vary across alternatives, then an error-components model can estimate the effect of unobserved attributes. Of course, combinations of the error-components and random-coefficients specifications are possible; elements of X that do not enter Z are variables whose coefficients do not vary in the population, elements of Z that do not enter X are variables whose coefficients vary in the population with mean 0, and elements that enter both X and Z are variables whose coefficients vary in the population with means represented by the appropriate elements in β. The elements contained in Z determine if the model specifies IID violations as unobserved alternatives or random taste variation.

A wide variety of discrete choice models can be derived by specifying different distributions for the stochastic components of utility η and ε and by including different information in the vector Z. Below I demonstrate how both multinomial probit and mixed logit models can be derived from this utility framework, and discuss the relative merits of each in the study of multicandidate elections.

2.1 Multinomial Probit

To derive a multinomial probit model from Eq. 3 assume η and ε have multivariate normal distributions. The random term η is normally distributed with mean 0 and a general covariance matrix, while the random term ε is distributed IID standard normal. Note that if $\eta = 0$ this is an independent probit model, which has the IIA property.

The unobserved portion of utility in this model is $\xi = Z\eta + \varepsilon$. Since the sum of two normal distributions is also normally distributed, ξ is distributed as a multivariate normal with mean 0 and a general covariance matrix Σ. Estimation of the multinomial probit generally involves estimating β and Σ.

Most applications of MNP in political science have been motivated by the desire to relax the IIA assumption and allow for more flexible substitution patterns between alternatives. Contrary to the popular belief in political science, MNP *does* impose a priori constraints on how individuals view alternatives (as do all other discrete choice models). For example, most MNPs in political science assume that the IID violations are due to common unobserved attributes, so Z is defined as an identity matrix of the same dimension as the number of alternatives. This is an error-components specification with normally distributed alternative-specific dummy variables as the error components. Although this specification is very general, it is inaccurate to say no a priori constraints are imposed. This particular specification of MNP is dominant in political science (Alvarez and Nagler 1995, 1998a, 1998b; Alvarez, Nagler, and Bowler 2000; Lacy and Burden 1999, 2000; Lawrence 1997; Quinn, Martin, and Whitford 1999; Schofield, Martin, Quinn, and Whitford 1998). Alternative specifications are of course possible through different specifications of Z. For instance, the original specification of MNP by Hausman and Wise (1978) specified $Z = X$, and estimated a random-coefficients MNP with the elements of the covariance matrix of the unobserved portions of utility depending upon the values of the variables included in Z. Such specifications have yet to be examined in political science.

The probability that individual i selects alternative j is estimated by integrating over the multivariate normal distribution of the unobserved portions of utility. Since only differences in utility are relevant for the choice probabilities, the dimension of integration is reduced from J to $J-1$ by subtracting the utility for one alternative from all other utilities and integrating over the resulting utility differences. Thus, estimating a MNP model generally requires the evaluation of a $(J-1)$-dimensional integral. If the dimension of integration is greater than two numerical techniques cannot compute the integrals with sufficient speed and precision for maximum likelihood estimation. In this case simulation techniques must be applied to estimate the MNP — for example, the GHK probability simulator or MCMC simulation. Advances in computational power and simulation techniques have reduced the costs of estimating MNPs to the point where they are becoming a popular choice for examining multicandidate and multiparty elections.

Multinomial probit allows for a more realistic formulation of models of political behavior than discrete choice models that assume the unobserved portions of utility are distributed IID, such as MNL. However, MNP has two properties that limit the range of models available for study with this method. The first limitation is in the number of random terms that may be estimated with a MNP. As only the differences in utility matter, and because one element in Σ must be fixed to set the scale of utility, only $\frac{J\times(J-1)}{2} - 1$ elements in Σ are identified. This means that MNP can estimate at most $\frac{J\times(J-1)}{2} - 1$ random coefficients or error components. This is true regardless of the number of elements in Z. There may be circumstances, particularly in a random-coefficients setting, where it is desirable to examine random taste variation in a greater number of coefficients. However, adding additional elements to Z in a MNP will not allow for the study of random taste variation in more coefficients — this will simply make the identified terms in Σ linear combinations of the variance and covariance of the various elements in Z. Alternatively, in some situations the substitution patterns between alternatives can be captured with fewer error components than there are alternatives. However, reducing the number of elements in Z will not reduce the dimension of integration, as Σ is still $J \times J$, requiring the evaluation of a $(J-1)$-dimensional integral to solve for the choice probabilities.

The second limitation of MNP, and perhaps the more restrictive in the study of multiparty elections, is the requirement that all of the terms in η be distributed normally. There are many instances in which non-normal distributions on error components or random coefficients are appropriate. For instance, the spatial model of voting maintains that individuals dislike candidates who are "far" from their ideal issue positions, so a negative sign on a coefficient that measures the impact of "issue distance" on vote choice is expected. If the goal is to examine random taste variation in issue distance with a MNP the coefficient on "issue distance" could be specified as a random coefficient with a normal distribution. Unfortunately, some individuals in the dataset would have coefficients of the "wrong" sign, since the normal distribution has infinite tails. A better specification for this random coefficient would be a distribution constrained to take the "correct" sign, such as the negative of a log-normal or beta distribution. Multinomial probit does not allow for non-normal distributions on error components or random coefficients, and would thus be restricted to estimating a random-coefficients model with unsatisfactory empirical implications.

2.2 Mixed Logit

To derive a mixed logit model from Eq. 3 assume ε is IID extreme value, while η follows a general distribution $g(\eta|\Omega)$. If $\eta = 0$ this is MNL, which has the IIA property. Estimation of the mixed logit generally involves estimating the vectors β and Ω.

For each individual, the choice probabilities will depend on β and η. Conditional on η, the probability that individual i selects alternative j is simply MNL:

$$P(j|\eta) = \frac{\mathrm{e}^{X\beta+Z\eta}}{\sum_{k\in J}\mathrm{e}^{X\beta+Z\eta}} \tag{4}$$

If the value of η was known for each individual, the solution to Eq. 4 would be straightforward. However, η is unobserved, although it is drawn from a known joint density function g. Thus, in order to obtain the unconditional choice probability for each individual the logit probability must be integrated over all values of η weighted by the density of η.

$$P(j) = \int_{\eta}\left[\frac{\mathrm{e}^{X\beta+Z\eta}}{\sum_{k}\mathrm{e}^{X\beta+Z\eta}}\right] g(\eta|\Omega)\partial\eta \tag{5}$$

Examination of Eq. 5 reveals the choice probability is a mixture of MNL probabilities, with the weight of each particular MNL probability determined by the mixing distribution g (thus the term "mixed logit"). The IIA property does not hold for MXL, even if the covariance matrix of g is diagonal. This is because η is constant across alternatives, introducing correlation in the utility across alternatives at the individual level (note this is also true of the MNP specification in the previous subsection).

The likelihhod function in a repeated choice setting is similar to the single choice likelihood function in Eqs. 4 and 5. The probability of each individual's sequence of observed choices over T time periods is:

$$P(j|\eta) = \prod_{t}\frac{\mathrm{e}^{X_t\beta+Z_t\eta}}{\sum_{k\in J}\mathrm{e}^{X_t\beta+Z_t\eta}} \tag{6}$$

with an unconditional choice probability given by:

$$P(j) = \int_{\eta} \left[\prod_{t} \frac{e^{X_t\beta + Z_t\eta}}{\sum_k e^{X_t\beta + Z_t\eta}} \right] g(\eta|\Omega) \partial\eta \tag{7}$$

Like the single choice specification, IIA does not hold for a MXL over a series of choices. Further, since η is constant across time, this introduces correlation for individuals over time.

The unconditional probability that individual i selects alternative j in either case is estimated by integrating over η. This integral cannot be evaluated analytically since it does not have a closed-form solution. If the dimension of integration is greater than two quadrature techniques cannot compute the integrals with sufficient speed and precision for maximum likelihood estimation. Thus simulation techniques are usually applied to estimate mixed logit models.

The integrals in the choice probabilities are approximated using a Monte Carlo technique, and then the resulting simulated log-likelihood function is maximized. For a given Ω a vector of values for η is drawn from $g(\eta|\Omega)$ for each individual. The draws of η can be taken randomly or by using Halton sequences (a quasi-random method of drawing that ensures more even coverage of the interval over which the integration is to be performed). The values of this draw can then be used to calculate $\hat{P}(j|\eta)$, the conditional choice probability given in Eqs. 4 or 6. This process is repeated R times, and the integration over $g(\eta|\Omega)$ is approximated by averaging over the R draws. The resulting simulated choice probability $\hat{P}(j|\beta, \Omega)$ is then inserted into the simulated log-likelihood function, which is maximized with conventional gradient-based optimization methods. See the Appendix for details on the estimation of mixed logit models.

Mixed logit is a more general discrete choice model than MNP in two respects. First, any number of elements may be included in the random term η. Unlike MNP, which reflects the effects of unobserved attributes and random taste variation in the covariance matrix of the unobserved portions of utility, MXL includes the elements of η as additional coefficients in the utility function. This means that the number of elements in η are not subject to the identification restrictions of the covariance matrix of the unobserved portions of utility, and thus MXL can estimate any number of random coefficients or error components. Further, the integration required to solve for the choice probabilities in MXL is over the elements in η, while in MNP it is over the differences in the unobserved portions of utility. If there are Q elements in η, solving for the choice probabilities in MXL will require the evaluation of a Q-dimensional integral. If substitution patterns or random taste variation can be captured with fewer error components such that Q is less than $J - 1$, estimating MXL will be easier than estimating an equivalent MNP, which requires solving an integral of dimension $J - 1$. This also means that MXL can estimate random coefficients in a model that has only two alternatives — MNP cannot do this, since there are no free elements in the covariance matrix when $J = 2$. Second, in mixed logit the elements of η can follow any distribution. MNP requires that $g(\eta|\Omega)$ be multivariate normal, which can be undesirable in many situations. Elements in η which have restricted signs or a finite support are easily handled in MXL.

These advantages make mixed logit a more general discrete choice model than multinomial probit. A mixed logit model can be specified that will estimate the same substitution patterns or random coefficients as any MNP — McFadden and Train (2000) demonstrate that MXL can be specified to approximate any discrete choice model derived from random utility maximization (to an arbitrary degree of closeness) with the appropriate choices of g and Z.[4] Conversely, MNP is unable to estimate models that approximate MXL under many specifications. Thus, mixed logit can be specified to answer many questions about multicandidate and multiparty elections that multinomial probit cannot, and therefore have not previously been addressed.

3 Empirical Applications

The best way to demonstrate the various uses of mixed logit models is through empirical examples. Below are three empirical examples, each intended to demonstrate a different use for mixed logit. The first example is an estimate of substitution patterns between parties competing in the 1987 British general election. There it is shown that MXL can be specified in much the same way as MNP, and used to determine if certain alternatives are viewed as similar by individuals. The second example studies heterogeneity in the impact of union membership on presidential vote in the 1992 US presidential election. The third example studies Congressional career decisions, specifying correlations over time.

3.1 Substitution Patterns in the 1987 British General Election

In this first example I use data from the 1987 British general election survey (Heath, Jowell, and Curtice 1989). This same dataset has also been examined by Alvarez and Nagler (1998) and Alvarez, Nagler, and Bowler (2000), with a multinomial probit model. In those papers they use an error-components MNP to account for common unobserved attributes and relax the IIA property. Below I will estimate an error-components MXL designed to account for unobserved attributes in the same way as the MNPs commonly estimated in political science. This reveals that like MNP, MXL relaxes the IIA property, and can even uncover the same substitution patterns as MNP.

The error-components MXL presented below is designed to account for the same violations of the IID assumption that lead to the MNP specification most common in political science. In order to facilitate this comparison I use the same variable coding as described in Alvarez, Nagler, and Bowler (2000) (see the web appendicies for this paper at http://www.polsci.ucsb.edu/faculty/glasgow). The impact of issue distance is assumed to be constant across all three parties included in the model (Conservative, Labour, Alliance), while the individual-specific variables are normalized such that the coefficients for voting Alliance are zero.

In order to specify a mixed logit that will treat IID violations in the same way as the MNP models that are popular in political science, first note that the unobserved portion of utility in

[4] Note that this result is stronger than the "universal logit" theorem (McFadden 1984), which states that any discrete choice model can be approximated by a model that takes the form of a standard logit. However, this requires that the attributes of each alternative be allowed to enter into the utility functions of other alternatives, meaning that logit models of this form are no longer consistent with random utility maximization. In contrast, MXL can approximate any discrete choice model while remaining consistent with theories of individual behavior.

a mixed logit can be decomposed into two parts, with u a general distribution that carries the correlation and heteroskedasdicity across alternatives, and ε an IID extreme value term:

$$\begin{aligned}
U_1 &= X_1\beta_1 + e_1 = X_1\beta_1 + u_1 + \varepsilon_1 \\
U_2 &= X_2\beta_2 + e_2 = X_2\beta_2 + u_2 + \varepsilon_2 \\
U_3 &= X_3\beta_3 + e_3 = X_3\beta_3 + u_3 + \varepsilon_3
\end{aligned}$$

Since I am interested in replicating the substitution patterns estimated by a MNP, I specify the us as multivariate normal. In order to identify the model u_3 is subtracted from all utilities to yield:

$$\begin{aligned}
U_1 &= X_1\beta_1 + (u_1 - u_3) + \varepsilon_1 \\
U_2 &= X_2\beta_2 + (u_2 - u_3) + \varepsilon_2 \\
U_3 &= X_3\beta_3 + 0 + \varepsilon_3
\end{aligned}$$

This model is estimated by specifying three dummy variables — one that enters the first utility function, one that enters the second utility function, and one that enters both the first and second utility function. These dummy variables are estimated as normally distributed error components, with the first two measuring the standard deviations of $U_1 - U_3$ and $U_2 - U_3$, and the third measuring the covariance between these differences in utility. In order to set the scale of utility the standard deviation of one of the random coefficients must be set to a constant. Note that this setup will estimate the $(J-1) \times (J-1)$ covariance matrix of the *differences* in utility, while most applications of MNP in political science have estimated the elements of the $J \times J$ covariance matrix of the utility functions themselves (but see Lawrence 1997). This does not make any substantive difference in the interpretation of the model, although it is important to note that the unobserved portion of utility will be scaled differently in the MXL when compared to the MNP (due to the addition of the IID extreme value term to the utility functions). In the model presented below the standard deviation of the error component for voting Conservative relative to the Alliance was constrained to $\sqrt{\pi^2/3 + 2}$ (the sum of the assumed variance of the difference between the IID extreme value terms and the variance of the difference between the normal error components usually assumed by political scientists).

The results of estimating this error-components MXL are presented in Table 1. The coefficient estimates on issue distance (which do not vary over parties) are presented in the first seven rows. The coefficients for the individual-specific variables are presented below the issue distance coefficients, with the coefficients for voting Conservative relative to the Alliance in the first column, and the coefficients for voting Labour relative to the Alliance in the second column. The error components are presented below the individual-specific coefficients.[5]

Table 1 here.

[5] This model and all other models in this paper were estimated using 125 Halton draws in the simulation of the integrals in the choice probabilities. See the appendix for details.

Examination of Table 1 reveals that the mean effects of the variables in the model are similar to those estimated in Alvarez, Nagler, and Bowler (2000). Short-term political effects, the relative issue positions of the parties and perceptions of the state of the national economy, and several demographic variables emerge as important factors in the 1987 British general election.

The major motivation for using MNP in political science has been to avoid the unrealistic substitution patterns that are estimated when IIA is assumed but does not hold. The substitution patterns estimated by the mixed logit model in Table 1 are similar to those estimated by the MNP specification common in political science. The vote shares reported here are the mean of the probabilities calculated using 1000 draws from the multivariate normal distribution of the estimated parameters. This was done to obtain standard errors on the predictions, which are reported in parentheses. In a three-party race this MXL predicted 44.56% (0.88%) would vote Conservative, 29.75% (0.79%) Labour, and 25.68% (1.08%) Alliance. The equivalent MNP predictions were 44.93% (0.81%) Conservative, 29.74% (0.71%) Labour, and 25.32% (0.89%) Alliance, while a MNL model estimated on the same data predicted 44.92% (0.78%) Conservative, 29.89% (0.77%) Labour, and 25.19% (0.87%) Alliance (this model is the equivalent of the MXL model without the error components). The predicted vote shares with the Alliance removed from the choice set are 57.60% (0.90%) Conservative and 42.40% (0.90%) Labour for MXL, 57.73% (1.09%) Conservative and 42.27% (1.09%) Labour for MNP, and 58.68% (0.79%) Conservative and 41.32% (0.79%) Labour for MNL.

The estimated vote shares of all three models in the three-party race are close to one another. However, when the Alliance is removed from the choice set the different substitution patterns estimated by the three models result in three different predicted vote shares in the two-party race. MNL, which assumes the stochastic portion of utility is IID, generates a higher prediction for the Conservative share of the vote in a two-party race than either MNP or MXL, which do not assume the stochastic portion of utility is IID.

The MNP and MXL results suggest that for many voters Labour was seen as a better substitute for the Alliance than the Conservatives, and this is reflected in the estimated substitution pattern when the Alliance is removed from the choice set. The predicted substitution pattern for the MXL model is quite close to that predicted by MNP, and falls between that predicted by MNL and that predicted by MNP. Note this does not suggest that MNP is generating the "correct" substitution pattern, as the underlying assumptions (our particular choice of g and Z) may be wrong. Different numbers and types of error components might result in a model that more accurately reflects the true substitution patterns than the models considered here. Since MXL imposes fewer restrictions on the number and distribution of the error components it is better able to explore alternative (and possibly more accurate) error-components specifications than MNP.

3.2 Heterogeneity and Union Voting Behavior in the 1992 US Presidential Election

Both multinomial probit and mixed logit can be specified as random-coefficients models as well. However, just as in an error-components specification, MXL imposes fewer restrictions than MNP on the number and distribution of the random coefficients in a random-coefficients specification. Thus MXL allows more flexibility than MNP in the specification of random-coefficients models. The choice between an error-components or a random-coefficients specification, and the choice of what information to include in Z (constants to pick up unobserved attributes, or other variables to examine random taste variation) will generally be decided by determining if the primary focus of the model is on exploring substitution patterns or random taste variation. Below I present a random-coefficients MXL designed to study heterogeneity in voting behavior among union members in the 1992 US presidential election. The dataset I used was the 1992 American National Election Study (Miller, Kinder, and Rosenstone 1993). For variable coding see the web appendix for this paper.

I examined random taste variation among union members in this dataset. Union member is coded as a one if someone in the respondent's household is a member of a labour union and not an employee of the government, and zero otherwise. Government employees were screened out in the belief that this group would be less sensitive to foreign competition than non-government union members. A significant amount of random taste variation in the impact of union membership on the vote between Clinton and Perot would tend to confirm taste variation among union members. To test for random taste variation in the impact of union membership on the vote I estimated a random-coefficients MXL. I specified two independent random coefficients on the dummy variables that indicate if a voter was a member of a labour union, both for voting Bush relative to Perot, and for voting Clinton relative to Perot. As the hypothesis I am testing maintains that some union members weighed union endorsements more heavily, while others were attracted to Perot's anti-NAFTA, pro-labor stance, random coefficients are required that allow individuals to be either above or below a mean. A normal distribution is one possibility. However, normal distributions have infinite tails, which would require that some individuals have implausible (near-infinite) coefficient values. Thus I utilized triangular distributions for the random coefficients. Triangular distributions have a density function that is zero before some endpoint $m-a$, rises linearly to a mean m, descends linearly to the other endpoint $m+a$, and is zero beyond $m+a$. The parameters that describe this distribution are the mean (m) and the distance between the mean and the endpoints (a). Triangular distributions are similar to normals in that the density function is symmetric and has more mass in the middle than in the tails, but avoid the substantive implications of a random coefficient with an infinite support.

The results of estimating this random-coefficients MXL are presented in Table 2. The first column lists the independent variables in the model, while the next 3 columns list the coefficient estimate, the standard error, and the t-statistic. The fifth and sixth columns list diagnostics for assessing how the missing data that was imputed contribute to uncertainty about the coefficient estimates. R is a measure of the relative increase in variance in each coefficient due to missing data, while λ is an estimate of the fraction of missing information about each coefficient (Rubin 1987, Schafer 1997).

Table 2 here.

The substantive findings among the fixed coefficients in Table 3 are similar to those reported in other studies. Ideological distance had a strongly negative impact on vote choice. Assesments of both personal finances and the state of the national economy had an impact on voting for Bush relative to Perot, with individuals who had more positive assesments more likely to vote for Bush. These variables did not have a statistically significant impact on the vote choice between Clinton and Perot. Assesments of the ability of the U.S. to compete in the global economy did not have a statistically significant impact on vote choice. Individuals who felt Japan was competing unfairly were more likely to favor Perot over Clinton. This indicates that individuals who harbored more protectionist feelings were more likely to vote for Perot. Younger people were more likely to vote for Perot. This is likely due to weaker partisanship among younger voters who have not yet established a voting history with one of the major parties. Minorities heavily favored Clinton and shunned Perot, as did women. Partisanship had the expected impact on vote behavior, with individuals who stated they identified with one of the major parties more likely to vote for that party's candidate.

Examination of the coefficient for the mean impact of union membership reveals that, on average, union membership did not influence vote choice. However, the estimated distance between the mean and the endpoint of the triangularly distributed coefficient for the impact of union membership on the vote choice between Clinton and Perot reveals there was a statistically significant amount of random taste variation in this coefficient. Although union membership did not affect the vote choice between Clinton and Perot on average, there was a significant degree of heterogeneity in this impact. For some individuals, union membership was a factor in the vote choice between Clinton and Perot. Examining the mean impact of union membership on this vote choice without considering variation in this impact would lead to the mistaken conclusion that union membership had no impact on voting in the 1992 presidential election.

These estimates of random taste variation reveal a great deal about union voting behavior in the 1992 presidential election. To demonstrate this I created a hypothetical voter, with characteristics set to the mean or modal values of the individuals in the dataset. I then computed the probability of voting for each candidate for this hypothetical voter both if he was a union member and if he was not a union member, using the MXL presented in Table 2. With this random-coefficients MXL the impact of union membership on vote choice depends on where the hypothetical voter falls in the distributions on the coefficients on union membership. Below I present 3 different predictions from the MXL model in Table 2 for the hypothetical voter; with the impact of union membership on the vote at the mean value for the Clinton versus Perot vote, and with the Clinton versus Perot union membership coefficient at the 1^{st} and 3^{rd} quartiles. The mean and standard deviation of the predicted probabilities were calculated using 1000 draws from the multivariate normal distribution of the estimated parameters, and are presented in Table 3.

Table 3 here.

The predictions for the hypothetical voter vary widely with this mixed logit model, depending on the impact of union membership on the vote for this particular individual. At the mean of the distribution on the Clinton versus Perot union membership coefficient the predicted vote probabilities shift slightly towards Perot when the hypothetical voter is moved into a union, with both Bush and Clinton losing slightly. However, if the coefficient value for this voter is at the 1^{st} quartile in the triangular distribution instead of the mean, this voter is far more likely to support Perot and less likely to support Clinton if he is a union member. Switching this voter out of a union leads to a large drop in support for Perot and a smaller drop in support for Bush, with Clinton gaining that support. If this coefficient is at the 3rd quartile in the distribution, if the hypothetical voter is in a union he is more likely to support Clinton, and switching this voter out of a union leads to a large gain for Perot at the expense of Clinton, with Bush also gaining some support.

These different predictions generated by this mixed logit for the effect of union membership on the vote are all for the same hypothetical voter. All individuals who match this profile would appear to be identical in the dataset, and most models currently used in political science would predict that these individuals would behave in exactly the same way. However, this random-coefficients specification of MXL reveals that there was a great deal of random taste variation in the impact of union membership on the vote choice between Clinton and Perot in 1992. Some individuals in labor unions followed the endorsements of their leadership, while others voted against these endorsements and supported Perot. This led to a great deal of individual-level variation in the impact of union membership on the vote choice between these two candidates. This heterogeneity in the behavior of union members is not apparent in models that assume there is no random taste variation.

Another way to demonstrate the effect of random taste variation in the impact of union membership on the vote is to examine the impact of this variable on vote probabilities over the entire range of the random coefficient. In Figure 1 I graph the means and standard deviations of the vote probabilities for the hypothetical voter if he is a union member. The means and standard deviations of the predicted probabilities were calculated as in Table 3 across the values of the random coefficient. The three heavy curves represent the mean probabilities of voting for each of the three candidates as the coefficient value on the impact of union membership changes, while the lighter curves represent a two standard deviation confidence interval around the mean values. Note the graphs are not smooth because each point was calculated individually using 1000 draws from the multivariate normal distribution of the estimated parameters.

Figure 1 here.

It is obvious from Figure 1 that the vote probabilities for the hypothetical voter vary widely, depending on where in the coefficient distribution he is. The impact of union membership on the vote is very different across individuals — even those individuals who are indistinguishable in the dataset in terms of the demographic, social, and political variables included in most models of vote choice.

3.3 Congressional Career Decisions and Individual Correlation over Time

The last empirical example studies Congressional career decisions over time. The dataset I used was the Congressional career decision data compiled by Kiewiet and Zeng in their study of factors that influence Members of Congress in the decision to retire, run for reelection, or seek higher office (Kiewiet and Zeng 1993). For variable coding see the web appendix for this paper.

Kiewiet and Zeng hypothesize that age, holding a leadership position, minority party status, ideological location in the party caucus, and the institution of House reforms would all affect the utility that a member of Congress receives from serving in the House. The previous vote share in the last general election, scandals, and harmful redistrictings (that cause the member to face another incumbent for reelection) would all affect the likelihood of winning reelection. Thus, these variables should have an impact on the decision between running for relection and the other options (retire or seek higher office).

Four variables are also expected to affect the expected utility of running for higher office (here regarded as a Senate seat or the Governorship). These variables are the overlap between a member's district and the entire state, whether or not there is a Senate election in that year in the member's state that is either an open seat or has an incumbent of another party, , whether the Senate seat is open, and whether the Governorship is open. These variables are expected to affect the choice between running for reelection or seeking higher office, but not the choice between running for reelection or retiring.

The model employed by Kiewiet and Zeng was a "mother logit" model, designed to relax the IIA assumption while still maintaing a logit form for the probabilities. The mother logit allows any choice probability to be expressed in logit form (McFadden 1975, 1984). Kiewiet and Zeng estimated their mother logit by entering factors expected to influence the utility of seeking higher office into the utility functions for retirement and running for reelection — thus, the four variables expected to influence the choice between seeking higher office or running for reelection were also entered into the estimation of the choice between retiring or running for reelection. This model was estimated by pooling the data for all Members of Congress (those Members of Congress who did not have an opportunity to seek higher office were omitted from this model, and examined in a separate binary logit).

I reestimated the model described by Kiewiet and Zeng using a mixed logit. Mixed logit represents an improvement over a mother logit in this application in at least two respects. First, as Kiewiet and Zeng point out (p. 939), estimating a mother logit on the pooled data makes the assumption that individual choices are not correlated over time. Mixed logit relaxes this assumption, and allows for the choices made by an individual to be correlated over time. This happens because η is constant across time as well as across alternatives, so not only does this allow for correlations across alternatives (relaxing the IIA property), but it also allows for correlation across time. Second, although mother logit models can approximate the choice probabilities of any discrete choice model with a logit form, they are not consistent with random utility maximization (Brownstone and Train 1999; McFadden and Train 2000). This is because that attributes of each alternative are allowed to enter the utility functions of other alternatives. In this example, components of the utility function for seeking higher office are entering into the choice between retirement and running for reelection, even though that choice should only depend on the utility of retirement and the utility of running for reelection. Thus, while a mother logit does not have the IIA property, it is also not consistent with random utility maximization. Mixed logit does not have the IIA property, and it is also consistent with our theories of individual behavior.

I specified a mixed logit model similar to the one estimated in Table 1, by specifying three dummy variables — one that enters the utility function for retirement, one that enters the utility function for seeking higher office, and one that enters both utility functions. The model is normalized with respect to running for reelection. Unlike the model in Table 1, this model was not set to the same scale as an equivalent MNP. Factors thought by Kiewiet and Zeng to affect the utility of House membership enter into both comparisons (retire vs. reelection and higher office versus reelection), while the scandal variable only enters into the retirement decision, and the four variables that affect the expected utility of seeking higher office enter into the seek higher office decision. Note that this data was censored, since not all members of Congress had an opportunity to seek higher office. The results of estimating this MXL are presented in Table 4.

Table 4 here.

Examination of Table 4 reveals that the substantive effects of the variables in the model are similar to those estimated in Kiewiet and Zeng (1993). Further, none of the error components reaches standard levels of statistical significance, so I cannot reject the hypothesis that there is no individual level variation in how members of Congress view their career options. Although MXL allows decisions to be correlated over time, there is little evidence of any such correlation here. Kiewiet and Zeng hypothesized that such correlation might be mitigated by the much larger cross-sectional size of the dataset as compared to its length (there are 1796 individuals in this dataset, who served an average of five terms each). Alternatively, the error structure I specified here might be incorrect. Regardless, this MXL represents an improvement over the mother logit model since it is consistent with random utility maximization.

4 Discussion

A wide variety of structures can be placed on the unobserved portions of utility in order to answer different questions about voting behavior in multicandidate and multiparty elections. While the multinomial probit models currently in favor in political science focus on estimating accurate substitution patterns through a single type of error-components specification, there are many other ways to specify the unobserved portions of utility in a discrete choice model that can improve our substantive knowledge of voter behavior. Both multinomial probit and mixed logit can be specified to explore many of these questions, although mixed logit is the more general model of the two.

Mixed logit holds two advantages over multinomial probit that are of consequence. First, since the random components (error components or random coefficients) in MNP are captured in the covariance matrix of the unobserved portions of utility, their number is limited by the number of alternatives in the model and the identification restrictions on the covariance matrix. In contrast, the number of random components that can be specified in a mixed logit is unlimited. Second, all of the random elements in MNP must be normally distributed. Conversely, the random elements in MXL can follow any distribution. These advantages of mixed logit mean that a wider variety of specifications are possible, and MXL can thus be applied to a wider set of questions than MNP.

It is primarily concern with accurate substitution patterns that has led to the increasing popularity of MNP in political science. Many interesting substantive questions can only be addressed if researchers know if voters view certain parties or candidates as substitutes. Both MNP and MXL relax the IIA property by considering the unobserved portion of utility in calculating substitution patterns, and are thus suitable for answering these types of questions. However, mixed logit is more general in its specification of the unobserved portion of utility — any number of error components may be specified, and they can follow any distribution. Thus mixed logit is more flexible in the study of substitution patterns across alternatives than multinomial probit.

The advantages of mixed logit in the study of random taste variation are even more clear. Although models of voting behavior have existed for decades, little is understood about heterogeneity in the impact of issue positions, demographic variables, and other factors on vote choice. More generally, individual-level variation in many settings is substantively interesting and merits further exploration. Both MNP and MXL can be specified to explore random taste variation and thus problems of this type. However, multinomial probit is limited in the number of random coefficients that can be estimated, and these random coefficients must be normally distributed. Mixed logit can include any number of random coefficients, and these random coefficients can follow any distribution. Random coefficients that theory tells us are non-normal (such as the impact of issue positions on the vote in a spatial model) are easily handled in MXL, while the equivalent random coefficients in a MNP would necessarily contradict our theory. Thus mixed logit is an unambiguously superior tool for exploring heterogeneity.

Thus far political scientists have primarily been interested in the impact of non-IID unobserved utility on substitution patterns, and have treated the unobserved utility itself as a nuisance. However, the structure of unobserved utility is also of substantive interest. Understanding if a variable has a homogeneous or heterogeneous impact on a decision is useful information when trying to explain individual behavior. Of course, a model that is able to incorporate this information into the systematic portion of utility would be preferred — however, our data and theories are often lacking. Until better theories are developed, or better data is obtained, models that are able to uncover some of the structure of the unobserved portions of utility offer the best chance to improve our knowledge in this area. Both multinomial probit and mixed logit are useful in this line of research. However, mixed logit is more general in its specification of the unobserved portions of utility, and thus allows us to explore individual behavior more broadly.

Appendix: Estimation of Mixed Logit Models

The parameters to be estimated in a mixed logit model are β, the vector of fixed coefficients, and Ω, the parameters that describe the distribution of η. The mixed logit log-likelihood function for given values of the parameters β and Ω is:

$$\mathcal{L}(\beta, \Omega) = \sum_{I} \sum_{J} y_{ij} \log \left[\int_{\eta} \left\{ \frac{e^{X\beta + Z\eta}}{\sum_k e^{X\beta + Z\eta}} \right\} g(\eta|\Omega) \partial\eta \right] \tag{8}$$

where I is the set of all individuals, J is the set of all alternatives, and

$$y_{ij} = \begin{cases} 1 & \text{if } i \text{ chooses } j \\ 0 & \text{otherwise} \end{cases}$$

The dimension of η is $1 \times Q$, where Q is the number of variables in Z. Thus, the log-likelihood function in Eq. 8 involves the estimation of a Q-dimensional integral.[6] This integral cannot be evaluated analytically since it does not have a closed-form solution. If $Q = 1$ or $Q = 2$ the log-likelihood can be evaluated with numerical methods such as quadrature. However, if Q is greater than two quadrature techniques cannot compute the integrals with sufficient speed and precision for maximum likelihood estimation (Hajivassiliou and Ruud 1994, Revelt and Train 1998). In this case simulation techniques must be applied to estimate the log-likelihood function.

The integrals in the choice probabilities are approximated using a Monte Carlo technique, and then the resulting simulated log-likelihood function is maximized. For a given Ω a vector of values for η is drawn from $g(\eta|\Omega)$ for each individual. The values of this draw can then be used to calculate $\hat{P}(j|\eta)$, the conditional choice probability given in Eq. 4. This process is repeated R times, and the integration over $g(\eta|\Omega)$ is approximated by averaging over the R draws. Let $\hat{P}_r(j|\eta_r)$ be the realization of the choice probability for individual i for alternative j for the r^{th} draw of η. The choice probabilities given the parameter vectors β and Ω are approximated by averaging over the values of $\hat{P}_r(j|\eta)$:

$$\hat{P}(j|\beta, \Omega) = \frac{1}{R} \sum_{r=1}^{R} \hat{P}_r(j|\eta_r) \tag{9}$$

$\hat{P}(j|\beta, \Omega)$ is the simulated choice probability of individual i choosing alternative j given β and Ω. This simulated choice probability is an unbiased estimator of the actual probability $P(j)$, with

[6] Note that this is why a "mixed probit" model (with $g(\eta|\Omega)$ following a general distribution and ε IID standard normal) is generally regarded as impractical. Estimating this model would involve the evaluation of a $(Q + 1)$-dimensional integral, since the choice probabilities conditional on η are not closed-form as they are in MXL, but instead require the evaluation of a univariate normal density. Unless there is a strong theoretical reason to believe that the IID disturbances are normal, MXL is superior to a "mixed probit" model due to the lower dimension of integration required for estimation.

a variance that decreases as R increases. It is also twice differentiable and strictly positive for any realization of the finite R draws, which means that log-likelihood functions constructed with $\hat{P}(j|\beta,\Omega)$ are always defined and can be maximized with conventional gradient-based optimization methods. Under weak conditions this estimator is consistent, asymptotically efficient, and asymptotically normal (Hajivassiliou and Ruud 1994; Lee 1992). When R increases faster than the square root of the number of observations, this estimator is asymptotically equivalent to the maximum likelihood estimator. However, this estimator does display some bias at low values of R, which decreases as R increases. The bias is very low when $R = 250$ (Brownstone and Train 1999); most empirical work uses R equal to 500 or 1000.

The choice probabilities above depend on β and Ω, which need to be estimated. In order to estimate Ω the distributions in η are re-expressed in terms of standardized, independent distributions. That is, $g(\eta|\Omega)$ is re-expressed as $\mu + Ws$, where $\mu = 0$ (the mean vector of η), W is the Choleski factor of Ω_η, and s consists of IID deviates drawn from standardized, independent distributions. Under this specification, $Z\eta$ becomes $(sZ)\Omega$, where $\Omega = W$ and is to be estimated. A simulated log-likelihood function can then be constructed:

$$\mathcal{SL}(\beta,\Omega) = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}\log\left[\hat{P}(j|\beta,\Omega)\right] \tag{10}$$

The estimated parameter vectors $\hat{\beta}$ and $\hat{\Omega}$ are the vectors that maximize the simulated log-likelihood function.

An alternative estimation procedure proposed by Bhat (1999a) and Train (1999a) dramatically reduces estimation time for mixed logit models. This alternative simulation technique uses non-random draws from the distributions to be integrated over, rather than random draws. By drawing from a sequence designed to give fairly even coverage over the mixing distribution, many fewer draws are needed to reduce simulation variance to an acceptable level. In both Bhat (1999a) and Train (1999a), Halton sequences are used to create a series of draws that are distributed evenly across the domain of the distribution to be integrated.

Halton sequences are created by selecting a number h that defines the sequence (where h is a prime number) and dividing a unit interval into h equal parts.[7] The dividing points on this unit interval become the first $h - 1$ elements in the Halton sequence. Each of the h sub-portions of the unit interval is divided as the entire unit interval was and these elements are added on to the end of the sequence. This process is continued until the desired number of elements in the sequence is reached. Halton sequences result in a far more even distribution of points across the unit interval than random draws. Figure 2 presents 1000 random draws from a 2-dimensional uniform space, and compares them to 1000 Halton draws from the same space. It is obvious that Halton draws result in more even coverage of the space, with less "clumping".

Figure 2 here.

[7] Prime numbers are used to define Halton sequences, since the Halton sequence for a non-prime number will divide the unit interval in the same way as the Halton sequences based on the prime numbers that constitute the non-prime number.

Halton sequences can be used in place of random draws to estimate mixed logit models. For each element of η a different prime number is selected, and a Halton sequence of length $(R \times I) + 10$ is created (where R is the number of Halton draws desired for each observation and I is the number of individuals in the dataset). The first ten elements of the sequence are discarded because the first elements tend to be correlated over Halton sequences defined by different prime numbers. The first individual in the dataset is assigned the first R elements of each Halton sequence, the second individual is assigned the next R elements, and so on. For each element of each Halton sequence the inverse of the cumulative distribution for that element of η is calculated. The resulting values become the IID deviates s in the simulated log-likelihood function. Estimation is otherwise identical to that when using random draws to evaluate the integrals.

Both Bhat (1999a) and Train (1999a) found that in estimating mixed logits, the simulation error in estimated parameters is lower with 100 Halton draws than with 1000 random draws. Thus, using Halton sequences in place of random draws allows us to obtain more accurate estimates of model parameters at a fraction of the estimation cost.

I am aware of two software packages currently available that allow for estimation of mixed logit models. All mixed logit models in this paper were estimated using GAUSS. The GAUSS code used to estimate the models in this paper is available at http://www.polsci.ucsb.edu/faculty/glasgow. This code is a modified version of the GAUSS code made available by Kenneth Train on his website at http://elsa.berkeley.edu/users/train. Mixed logits can also be estimated in Limdep, although the options for estimation are more limited than those in the GAUSS code — only normally or lognormally distributed random coefficients are permitted. In comparing both software packages I found that in many applications Limdep produced results that agreed with the equivalent specification in the GAUSS code. However, in some applications Limdep failed to converge while the GAUSS code did. Overall, the GAUSS code seems more reliable than Limdep, and offers more options for estimation (more distributions are available for the random components, error-components specifications are possible, and Halton draws are available as an estimation option).

For further details on the estimation of mixed logit models see Appendix A to the paper "Mixed Logit Models for Multicandidate Elections" on my homepage.

References

Algers, S., P. Bergström, M. Dalhberg, and J.L. Dillén (1999). "Mixed Logit Estimation of the Value of Travel Time." Unpublished manuscript.

Alvarez, R.M., S. Bowler, and J. Nagler (2000). "Issues, Economics, and the Dynamics of Multi-Party Elections: The British 1987 General Election". *American Political Science Review*, 94:1, pp. 131-49.

Alvarez, R.M. and J. Nagler (1995). "Economics, Issues, and the Perot Candidacy: Voter Choice in the 1992 Presidential Election." *American Journal of Political Science*, 39:3, pp. 714-44.

Alvarez, R.M. and J. Nagler (1998a). "When Politics and Models Collide: Estimating Models of Multiparty Elections." *American Journal of Political Science*, 42:1, pp.,55-96.

Alvarez R.M. and J. Nagler (1998b). "Economics, Entitlements, and Social Issues: Voter Choice in the 1996 Presidential Election." *American Journal of Political Science*, 42:4, pp. 1349-63.

Bhat, C.R. (1998a). "Accommodating Variations in Responsiveness to Level-of-Service Measures in Travel Mode Choice Modeling." *Transportation Research A*, 32:7, pp.,495-507.

Bhat, C.R. (1998b). "Accommodating Flexible Substitution Patterns in Multi-Dimensional Choice Modeling: Formulation and Application to Travel Mode and Departure Time Choice." *Transportation Research B*, 32:7, pp. 455-66.

Bhat, C.R. (1999a). "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model." Unpublished manuscript.

Brownstone, D. and K. Train (1999). "Forecasting New Product Penetration with Flexible Substitution Patterns." *Journal of Econometrics*, 89:1 pp. 109-29.

Glasgow, G. (2001). "Mixed Logit Models for Multicandidate and Multiparty Elections." *Political Analysis*, 9:2, pp. 116-136.

Hajivassiliou, V.A., and P.A. Ruud (1994). "Classical Estimation Methods for LDV Models Using Simulation." In *Handbook of Econometrics*, Vol. 4, eds. R.F. Engle and D.L. McFadden, North Holland: New York, pp. 2383-441.

Hausman, J.A. and D.A. Wise (1978). "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." *Econometrica*, 46:2, pp. 403-26.

Heath, A.F., R.M. Jowell, and J.K. Curtice (1989). *British Election Study, 1987. A Computer File.* Colchester: ESRC Data Archive.

Honaker, J., A. Joseph, G. King, and K. Scheve (1999). *Amelia: A Program for Missing Data*

(Gauss version). Cambridge, MA: Harvard University, http://GKing.Harvard.edu/.

Horowitz, J.L. (1991). "Reconsidering the Multinomial Probit Model." *Transportation Research B*, 25:6, pp. 433-8.

Jain, D.C., N.J. Vilcassim, and P.K. Chintagunta (1994). "A Random-Coefficients Logit Brand-Choice Model Applied to Panel Data." *Journal of Business and Economic Statistics*, 12:3, pp. 317-28.

Kiewiet, D.R. and L. Zeng (1993). "An Analysis of Congressional Career Decisions, 1947-1986." *American Political Science Review*, 87:4, pp. 928-941.

King, G., J. Honaker, A. Joseph, and K. Scheve (2000). "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review*, in press.

Lacy, D. and B. Burden (1999). "The Vote-Stealing and Turnout Effects of Ross Perot in the 1992 U.S. Presidential Election." *American Journal of Political Science*, 43:1, pp. 233-55.

Lacy, D. and B. Burden (2000). "The Vote-Stealing and Turnout Effects of Third-Party Candidates in U.S. Presidential Elections, 1968-1996." Unpublished manuscript.

Lawrence, E.D. (1997). "Simulated Maximum Likelihood via the GHK Simulator: An Application to the 1988 Democratic Super Tuesday Primary." Unpublished manuscript.

Lee, L. (1992). "On Efficiency of Methods of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Choice Models." *Econometric Theory*, 8, pp. 518-52.

Maddala, G.S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University Press: New York.

McFadden, D. (1984). "Econometric Analysis of Qualitative Response Models." In *Handbook of Econometrics*, II, Z. Griliches and M. Intriligator, eds., North Holland: Amsterdam, pp. 1395-1457..

McFadden, D. and K. Train (2000). "Mixed MNL Models for Discrete Response." *Journal of Applied Econometrics*, 15:5, pp. 447-470.

Miller, W.E., D.R. Kinder, S.J. Rosenstone, and the National Election Studies (1993). *American National Election Study, 1992: Pre- and Post-Election Survey.* Ann Arbor: Inter-university Consortium for Political and Social Research.

Quinn, K.M., A.D. Martin, and A.B. Whitford (1999). "Voter Choice in a Multi-Party Democracy: A Test of Competing Theories and Models." *American Journal of Political Science* 43:4, pp. 1231-47.

Revelt, D. and K. Train (1998). "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level." *The Review of Economics and Statistics*, 80:4, pp. 647-57.

Rivers, D. (1988). “Heterogeneity in Models of Electoral Choice.” *American Journal of Political Science*, 32:3, pp. 737-57.

Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman and Hall: New York.

Schofield, N., A.D. Martin, K.M. Quinn, and A.B. Whitford (1998). “Multiparty Electoral Competition in the Netherlands and Germany: A Model Based on Multinomial Probit.” *Public Choice*, 97:3, pp. 257-93.

Sherman, R.P. (2000). “Tests of Certain Types of Ignorable Nonresponse in Surveys Subject to Item Nonresponse or Attrition.” *American Journal of Political Science*, 44:2, pp. 362-374.

Train, K. (1998). “Recreation Demand Models with Taste Differences over People.” *Land Economics*, 74:2, pp. 230-9.

Train, K. (1999a). “Halton Sequences for Mixed Logit.” Unpublished manuscript.

Table 1: Mixed Logit Estimates (MNP Replication), 1987 British General Election (Alliance Coefficients Normalized to Zero)

Independent Variables	Conservatives/Alliance				Labour/Alliance	
Defense			-0.30**	(0.04)		
Phillips Curve			-0.19**	(0.04)		
Taxation			-0.27**	(0.04)		
Nationalization			-0.30**	(0.03)		
Redistribution			-0.14**	(0.03)		
Crime			-0.17*	(0.08)		
Welfare			-0.23**	(0.03)		
South	-0.21	(0.28)			-0.64*	(0.30)
Midlands	-0.48	(0.28)			-0.26	(0.30)
North	-0.22	(0.29)			0.93**	(0.29)
Wales	-0.95	(0.55)			2.01**	(0.48)
Scotland	-0.86*	(0.41)			1.03**	(0.38)
Public Sector Employee	0.17	(0.24)			0.00	(0.23)
Female	0.45*	(0.23)			-0.06	(0.22)
Age	0.07	(0.08)			-0.35**	(0.08)
Home Ownership	0.75**	(0.28)			-0.78**	(0.25)
Family Income	0.12*	(0.05)			-0.10*	(0.05)
Education	-1.28*	(0.53)			-0.98	(0.53)
Inflation	0.49**	(0.16)			-0.02	(0.16)
Taxes	0.04	(0.11)			-0.15	(0.10)
Unemployment	0.48**	(0.10)			0.01	(0.11)
Working Class	0.07	(0.25)			1.03**	(0.25)
Union Member	-0.91**	(0.27)			0.56*	(0.24)
Constant	0.74	(1.16)			3.92**	(1.14)
Error Components (Mean)	0	—			0	—
Error Components (Std. Dev.)	2.30	—			1.72**	(0.30)
Error Components (Sqrt Corr.)			0.43	(0.69)		
Number of Observations			2131			
Log-Likelihood			-1474.70			

Standard errors in parentheses. ** indicates statistical significance at the 99% level; * indicates statistical significance at the 95% level.

Table 2: Mixed Logit Estimates,
1992 US Presidential Election
(Perot Coefficients Normalized to Zero)

Independent Variables	Coeff. Value	Std. Error	T Stat.	R	λ
Ideological Distance	-0.13	0.03	-4.99	0.38	0.30
Bush/Perot Coeffs.					
Constant	0.55	0.36	1.55	0.00	0.00
Personal Finances	0.23	0.11	2.01	0.00	0.00
National Economy	0.48	0.16	3.01	0.00	0.00
US in Global Economy	0.05	0.12	0.43	0.00	0.00
Democrat	-0.14	0.32	-0.45	0.01	0.01
Republican	1.45	0.29	4.92	0.00	0.00
Income	-0.01	0.02	-0.50	0.04	0.04
Age 18-29	-1.09	0.29	-3.77	0.01	0.01
Age 30-44	-0.64	0.26	-2.51	0.00	0.00
Age 45-59	-0.48	0.28	-1.73	0.00	0.00
Minority	0.98	0.47	2.07	0.00	0.00
Education	0.12	0.06	1.83	0.05	0.04
Female	0.60	0.18	3.32	0.00	0.00
Japan Unfair	0.07	0.10	0.68	0.17	0.15
Import Limits	-0.10	0.11	-0.97	0.01	0.01
Abortion	-0.35	0.09	-4.02	0.00	0.00
Union Member					
Mean	-0.11	0.26	-0.42	0.00	0.00
|Mean - Endpoint|	0.39	1.86	0.21	0.01	0.01
Clinton/Perot Coeffs.					
Constant	0.38	0.35	1.07	0.02	0.02
Personal Finances	0.03	0.11	0.23	0.00	0.00
National Economy	-0.23	0.18	-1.25	0.01	0.01
US in Global Economy	-0.20	0.13	-1.55	0.01	0.01
Democrat	1.41	0.26	5.43	0.01	0.01
Republican	-0.66	0.29	-2.26	0.00	0.00
Income	-0.05	0.02	-2.75	0.07	0.07
Age 18-29	-1.10	0.28	-3.89	0.00	0.00
Age 30-44	-0.79	0.26	-3.01	0.01	0.01
Age 45-59	-0.23	0.28	-0.83	0.01	0.01
Minority	2.20	0.42	5.20	0.01	0.01
Education	0.05	0.06	0.78	0.01	0.01
Female	0.30	0.18	1.71	0.01	0.01
Japan Unfair	0.22	0.10	2.27	0.03	0.03
Import Limits	-0.05	0.11	-0.44	0.01	0.01
Abortion	0.02	0.10	0.16	0.04	0.04
Union Member					
Mean	-0.03	0.27	-0.10	0.00	0.00
|Mean - Endpoint|	3.13	1.19	2.63	0.02	0.02
Number of Observations	1441				

Random coefficients have triangular distributions.

Table 3: Vote Probabilities for a Hypothetical Voter

	Not a Union Member	Union Member	Difference
(C/P At Mean)			
Bush	9.01 (2.81)	8.33 (3.10)	0.69 (4.15)
Clinton	43.64 (7.37)	43.53 (7.49)	0.11 (10.42)
Perot	47.35 (7.36)	48.14 (7.33)	-0.79 (10.32)
(C/P at 1st Quartile)			
Bush	9.01 (2.81)	11.35 (4.16)	-2.33 (4.98)
Clinton	43.64 (7.37)	23.95 (5.60)	19.69 (9.36)
Perot	47.35 (7.36)	64.70 (6.65)	-17.35 (9.93)
(C/P at 3rd Quartile)			
Bush	9.01 (2.81)	5.12 (2.04)	3.89 (3.46)
Clinton	43.64 (7.37)	65.24 (7.09)	-21.60 (10.07)
Perot	47.35 (7.36)	29.64 (6.60)	17.71 (9.73)

Standard errors in parentheses.

Figure 1: Distribution of Vote Probabilities over Union Coefficient (Clinton/Perot)

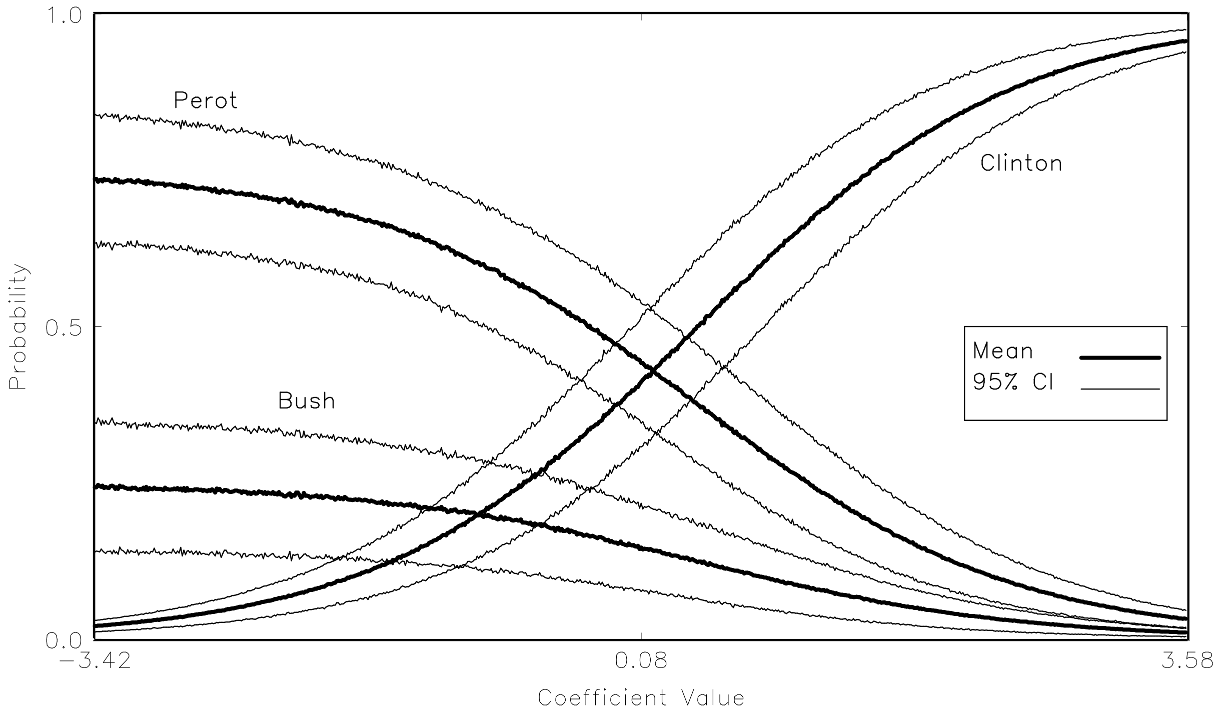

Table 4: Mixed Logit Estimates, Congressional Career Decisions (Reelection Coefficients Normalized to Zero)

Independent Variables	Retire/Reelection		Higher Office/Reelection	
Age	0.07**	(0.01)	-0.04**	(0.01)
Republican	0.93**	(0.24)	-0.29	(0.27)
Chair/Leader	-0.20	(0.21)	-0.99	(0.73)
Dem. Conservatism	1.42**	(0.28)	-1.06**	(0.37)
Rep. Conservatism	0.22	(0.27)	0.06	(0.34)
Institutional Reform	0.30*	(0.13)	0.44	(0.19)
Previous Margin	-0.00	(0.00)	0.00	(0.00)
Scandal	1.94**	(0.36)	—	—
Redistricting	0.87**	(0.33)	1.79**	(0.36)
District/State Ratio	—	—	0.94**	(0.08)
Senate Election	—	—	1.59**	(0.18)
Open Seat Senate	—	—	0.76**	(0.16)
Open Governorship	—	—	0.52**	(0.14)
Constant	-7.73**	(0.42)	-0.23	(0.50)
Error Components (Mean)	0	—	0	—
Error Components (Std. Dev.)	0.01	(0.25)	0.19	(0.54)
Error Components (Sqrt Corr.)	-0.00		(0.20)	
Number of Observations	8370			
Number of Individuals	1796			
Log-Likelihood	-2383.91			

Standard errors in parentheses. ** indicates statistical significance at the 99% level; * indicates statistical significance at the 95% level.

Figure 2: Distribution of 1000 Random Draws Compared to 1000 Halton Draws