Mixed Logit Models for Multiparty Elections

Garrett Glasgow

Harvard University

November 12, 1999

I thank R. Michael Alvarez, Fred Boehmke, Eric Lawrence, Jonathan Nagler, and Kenneth Train for their helpful comments. I can be reached at the Center for Basic Research in the Social Sciences (CBRSS), 34 Kirkland Street, Cambridge, MA 02138; email gglasgow@fas.harvard.edu.

1 Introduction

Recently those who study elections have focused a great deal of attention on the modeling of multicandidate and multiparty elections. The emergence of H. Ross Perot as a factor in American presidential elections no doubt was a major motivation for the increased interest in multicandidate elections, although the increased availability of survey data from many European countries was also an influence. Since many of the elections currently of interest to political scientists have more than two candidates competing, alternatives to the methods traditionally used to study vote choice (binomial logit and probit models) are necessary.

Whitten and Palmer (1996), in a study of voting behavior in Great Britain and the Netherlands, advocate the use of multinomial logit models for multicandidate elections. Previously, many political scientists analyzed multicandidate elections by estimating logit or probit models on various pairs of candidates in the election. Multinomial logit is superior to repeated applications of binomial logit in terms of efficiency (although the substantive interpretation is identical (Alvarez and Nagler 1998)).

Multinomial logit (MNL) begins by assuming that the utility yielded by party j to individual i can be represented by:

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

where V_{ij} represents the systemic (observed) portion of utility, and ε_{ij} represents the stochastic (unobserved) portion of utility.[1] Under the assumption that ε_{ij} is distributed independently and identically in accordance with the extreme value distribution, the probability that individual i will select party j is given by:

[1] 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.

$$P_i(j) = \frac{e^{V_{ij}}}{\sum_{k \in C_i} e^{V_{ik}}} \tag{2}$$

where V_{ik} is the utility yielded by party k to individual i and C_i is the choice set faced by individual i. MNL has many desirable properties. The probabilities that result from estimating MNL are necessarily between zero and one, and the probabilities sum to one over all alternatives.

However, multinomial logit models also have the independence of irrelevant alternatives property (IIA), which is undesirable when studying multiparty elections. The IIA property necessitates that for a specific individual the ratio of the choice probabilities of any two alternatives is independent of the utility of any other alternatives. For MNL the ratio of the choice probabilities for alternatives j and l is:

$$\frac{P_i(j)}{P_i(l)} = \frac{e^{V_{ij}} / \sum_{k \in C_i} e^{V_{ik}}}{e^{V_{il}} / \sum_{k \in C_i} e^{V_{ik}}} = \frac{e^{V_{ij}}}{e^{V_{il}}} = e^{V_{ij} - V_{il}} \tag{3}$$

which does not depend on the utility of any alternatives except j and l. The IIA property can lead to some troubling substantive implications. The most (in)famous example is the "red bus/blue bus paradox", but serious problems also arise in the context of multiparty elections. For instance, let us assume that an individual is torn between supporting one of two parties, one left-wing, and one right-wing. This individual has a 50% chance of supporting each party. Now imagine that a new right-wing party enters the race, identical to the existing right-wing party in every respect. We would expect this individual would now have a 50% chance of supporting the left-wing party, and a 25% chance of supporting each right-wing party. However, the IIA property requires the original probability ratio between the original right-wing party and the left-wing party to remain unchanged by the entry of the second right-wing party — any model with the IIA property would predict that our hypothetical individual would have a 1/3 chance of supporting each of the parties. This simple property of MNL has serious implications for the substantive interpretation of these models in situations where the researcher is examining the effect of adding or subtracting alternatives from the choice set (such as the impact of a third party on election outcomes).

Most advances in the empirical modeling of multiparty elections in recent years have focused on relaxing the IIA assumption. The most popular model choice in political science that does not assume IIA is the multinomial probit (MNP). MNL models assume that the unobserved components of utility are independent and identical extreme value distributions, which leads to the IIA property. MNP instead assumes that the unobserved components of utility are distributed jointly normal, with a general variance-covariance matrix. With the joint normal distribution the unobserved component of utility for each alternative can have a different variance, and can be correlated with the unobserved components of utility for other alternatives. Allowing the utility of different alternatives to be correlated enables MNP models to avoid the IIA assumption, as the ratio of the choice probabilities for two alternatives can now depend on the utility of other alternatives. The MNP model was first developed by Hausman and Wise (1978), and was first applied to multicandidate elections in a series of articles by Alvarez and Nagler (1995, 1998a, 1998b). The latter have found MNP models to have superior substantive results to those of MNL.

Although multinomial probit is superior in many ways to multinomial logit for studying multiparty elections, it does have a number of deficiencies. Estimation of MNP models is extremely cumbersome. Unlike MNL, MNP does not have a closed analytic form for the choice probabilities. This means that to estimate a MNP model with J alternatives, $J-1$ integrals must be evaluated. This greatly increases the time necessary to obtain parameter estimates relative to MNL. In practice it will be possible in many cases to model flexible substitution patterns (IIA violations) without having to estimate a fully specified covariance matrix; however, in MNP $J-1$ integrals must be evaluated regardless of the researcher's beliefs about how individuals view the alternatives in the choice set. Further, MNP is restricted in the form of the covariance matrix that can be estimated — the variances and covariances must be those from a multivariate normal distribution. There will be times when assuming a normal distribution for the unobserved portions of utility may be hard to justify theoretically. For instance, the researcher may wish to place bounds to the unobserved portions of utility, but this is not possible in MNP. This restriction is especially important if MNP is to be specified as a random-coefficients model (as in Hausman and Wise 1978), since there may be strong theoretical reasons to suspect that the random parameters are distributed non-normally.

Mixed logit (MXL) models represent a more flexible option for studying multicandidate elections than MNP. Since the choice probabilities are MNL combined with one or more general probability distributions, MXL does not face the restrictions described above. Any number of probability distributions can be combined with the MNL probabilities; thus, it may be possible in MXL to account for IIA violations at less computational cost than in MNP. Further, since the mixing distributions are general, MXL can handle non-normal IIA violations and random-coefficients. In the following section I will develop and explain the mixed logit model, and compare it to MNP in both an error-components and a random-coefficients framework. Section 3 discusses the estimation of both the MXL model. Section 4 presents a MXL model for the 1987 British general election. There I will demonstrate the utility of MXL in studying unobserved heterogeneity. Finally, I will conclude, explaining why mixed logit models offer a flexible and simple alternative to MNP in the study of multicandidate elections.

2 Mixed Logit and Multinomial Probit Formulations

In this section I will develop a common utility framework for both the multinomial probit and mixed logit models. Assume that an individual faces a choice set consisting of C 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 systemic component V_{ij} and a stochastic component e_{ij}. Rewrite the systemic component of utility as $V_{ij} = \beta' x_{ij} + \psi_j' a_i$, where x_{ij} is a vector of characteristics unique to alternative j relative to individual i, a_i is a vector of characteristics unique to individual i, and β and ψ_j are parameters to be estimated. Rewrite the stochastic component of utility as $e_{ij} = \delta_i' z_{ij} + \varepsilon_{ij}$, where z_{ij} is a vector of characteristics that are either unique to alternative j or to individual i (or both), δ_i is a random variable with mean vector μ that varies over individuals, and ε_{ij} is a random variable with mean zero that is independently and identically distributed (IID) over both individuals and alternatives. We then write the utility that individual i gets from alternative j as:

$$U_{ij} = \beta' x_{ij} + \psi_j' a_i + \delta_i' z_{ij} + \varepsilon_{ij} \tag{4}$$

Note that with the exception of the term $\delta_i' z_{ij}$ and the IID assumption about ε_{ij}, this specification of the utility function is identical to that presented in Alvarez and Nagler (1995) in their derivation of the MNP. There they assume $\delta_i = 0$ for all i, and assume that ε_{ij} is a multivariate normal distribution with a general covariance matrix. Below I will show how correct specification of δ_i and z_{ij} will lead to a model that is identical to that specified by Alvarez and Nagler, even if we maintain the assumption that ε_{ij} is IID.

In fact, a wide variety of qualitative choice models can be derived through different distributions of the stochastic components of utility δ_i and ε_{ij} and different specifications of the vector z_{ij}. For instance, in MNL, ε_{ij} is IID extreme value, and δ_i is a vector of zeros for all individuals i. With this specification the unobserved portion of utility is independent across alternatives, leading to the IIA property. Depending on the specification of z_{ij}, we can create qualitative choice models that specifically deal with the IIA assumption, or models designed to examine unobserved heterogeneity across individuals.

If the elements of $z_{ij} \neq x_{ij}$, then we are specifying an *error-components* model. The elements of z_{ij} are assumed to be error components that introduce heteroskedasticity and correlation across alternatives in the unobserved portion of utility. We assume that these error components have zero mean (μ is absorbed into the alternative specific constants). Note that even if the elements of δ_i are uncorrelated the unobserved portion of utility will be correlated across alternatives, since δ_i does not vary across alternatives.

If the elements of $z_{ij} = x_{ij}$ then we are specifying a *random-coefficients* model. In this instance the vector β gives the mean coefficient values for the elements in x_{ij}, while the vector δ_i gives the deviations of each individual from the mean vector β. In this instance $\mu = \beta$. Of course, combinations of the error-components and random-coefficients specifications are possible; elements of x_{ij} that do not enter z_{ij} are variables whose coefficients do not vary in the population, elements of z_{ij} that do not enter x_{ij} are variables whose coefficients vary in the population with mean zero, and elements that enter both x_{ij} and z_{ij} are variables whose coefficients vary in the population with mean β. Note that in either case we can assume that δ has mean zero, since μ is absorbed into the alternative-specific constants in the error-components models, or is an element of β in the random-coefficients models.

Below I present several distributions for δ and ε and specifications for z_{ij} that result in models that do not assume IIA and/or can examine heterogeneity across individuals. The following subsection presents the multinomial probit model, first discussing the specification usually employed by political scientists, and then demonstrating that MNP models can also be specified as random parameters models. The next subsection presents the mixed multinomial logit model, and demonstrates that it can be easily specified as an error-components model to allow for flexible substitution patterns across alternatives (IIA violations), or as a random-coefficients model to study unobserved heterogeneity across individuals.

2.1 Specification of the Multinomial Probit Model

Multinomial probit models have been described in the political science literature before (Alvarez and Nagler 1995, 1998a, 1998b; Lawrence 1997, Burden and Lacy 1999). However, here I will present a very general specification of MNP in order to facilitate comparisons between this model and MXL.

In order to get a multinomial probit model from the utility function in equation 4 we assume δ and ε have multivariate normal distributions:

$$\delta \sim MVN(0, \Sigma_\delta)$$
$$\varepsilon \sim MVN(0, I)$$

Since the sum of two normal distributions is also normally distributed, the observed and unobserved portions of utility become:

$$V_{ij} = \beta' x_{ij}$$
$$e_{ij} = z_{ij}' \Sigma_\delta z_{ij} + \Sigma_\varepsilon = \Sigma_\delta^*$$

The unobserved component of utility is distributed $N(0, \Sigma_\delta^*)$, which allows us to calculate the probability that individual i will select alternative j with a multinomial probit model. Specifically, the probability that individual i selects alternative j is given by:

$$P_i(j) = Pr[U_{ij} > U_{ik} \forall k] = \int_{U_{ij}=-\infty}^{\infty} \int_{U_{ik_1}=-\infty}^{U_{ij}-U_{ik_1}} \dots \int_{U_{ik_J}=-\infty}^{U_{ij}-U_{ik_J}} \phi(V_{ij}, \Sigma_\delta^*) \partial U_1 \dots \partial U_J \tag{5}$$

Depending on the specification of z_{ij}, the multinomial probit can be configured to deal with violations of IIA or with random coefficients. Note that most specifications of MNP in political science construct the multivariate normal distribution ϕ by allowing for a general covariance structure in the observation-specific error term ε, rather than through individual-specific errors as I have done here. Below I will demonstrate that either method will result in equivalent models.

Most applications of MNP in political science have been motivated by the desire to allow for more flexible substitution patterns between alternatives (that is, to relax the IIA assumption). Examples of this kind of model are found in Alvarez and Nagler (1995, 1998a, 1998b), Burden and Lacy (1999), and Lawrence (1997). These models place very little structure on the form of the unobserved portion of utility, instead seeking to measure the correlation of the unobserved portion of utility across alternatives. To specify this type of model in the framework presented here we define z_{ij} as an identity matrix of the same dimension as the number of alternatives. Then the unobserved portion of utility of individual i for alternative j becomes:

$$e_{ij} = z_{ij}' \Sigma_\delta z_{ij} + I = \Sigma_\delta + I = \Sigma_\delta^* \tag{6}$$

The unobserved portion of utility is distributed multivariate normal with mean zero and covariance matrix Σ_δ^*. This specification of the MNP is identical to those that have been employed in the political science literature in the past. This is best explained with an example. Let us assume that each individual faces three alternatives. The utility that individual i would receive from each of these alternatives is given by:

$$U_{i1} = \beta' x_{i1} + \psi_1' a_i + \delta_i' z_{i1} + \varepsilon_{i1}$$

$$U_{i2} = \beta' x_{i2} + \psi_2' a_i + \delta_i' z_{i2} + \varepsilon_{i2}$$

$$U_{i3} = \beta' x_{i3} + \psi_3' a_i + \delta_i' z_{i3} + \varepsilon_{i3}$$

If we define z_{ij} as above, then the unobserved portion of utility for each alternative becomes $\delta_{ij} + \varepsilon_{ij}$, the sum of two normal distributions. Denote this normally distributed variable δ^*_{ij}, which has mean zero and the covariance matrix described above.

Since only differences in utility matter in discrete choice models, we can normalize the utility of alternative 3 to zero by subtracting it from the utility for all alternatives. Then the utility of individual i for each alternative becomes:

$$U_{i1} = \beta'(x_{i1} - x_{i3}) + (\psi_1 - \psi_3)' a_i + (\delta^*_{i1} - \delta^*_{i3})$$

$$U_{i2} = \beta'(x_{i2} - x_{i3}) + (\psi_2 - \psi_3)' a_i + (\delta^*_{i2} - \delta^*_{i3})$$

$$U_{i3} = 0$$

Note that the unobserved portions of utility are still distributed normally (the difference of two normals is distributed normally). This is the normalization employed by Lawrence (1997), although he works with the observation-specific errors ε. This normalization can easily be shown to be equivalent to the slightly different normalization employed by Alvarez and Nagler (1995, 1998a, 1998b) and Burden and Lacy (1999), and this model specification will yield the same choice probabilities. This type of specification has broad appeal because the covariance matrix Σ^*_δ measures the variance and covariance of all unobserved portions of utility (although identification restrictions do not allow us to estimate the full covariance matrix in practice).

Alternatively, we might wish to specify a multinomial probit model that will account for unobserved heterogeneity in model coefficients. This was the motivation behind the MNP specification of Hausman and Wise (1978). The difference between a random-coefficients model and an error-components model such as the one described above lies purely in our definition of z_{ij}. Elements of z_{ij} that are also in x_{ij} will be estimated as random coefficients, with the mean of a particular coefficient given by the corresponding value of β and δ_i measuring each individual's deviations from the mean. Elements of z_{ij} that do not enter x_{ij} are error components; variables that have zero mean but vary normally in the population. If we set $z_{ij} = x_{ij}$ then we have a random coefficients specification. An estimate of β for a particular element in x_{ij} gives the mean coefficient estimate for that element of x_{ij}, while estimates of the properties of δ that correspond to an element in x_{ij} yield information about the distribution of the coefficient on that element of x_{ij}. The unobserved portion of utility of individual i for alternative j is then:

$$e_{ij} = x'_{ij}\Sigma_\delta x_{ij} + I \tag{7}$$

This specification allows us to estimate the variances and covariances of various coefficients on the variables in x_{ij}. Of course, this model is still subject to the same identification restrictions as any other MNP model. Thus, if there are J alternatives included in this model, we can estimate up to $J-1$ normally distributed random coefficients, or fewer if we are also interested in covariances.

2.2 Specification of the Mixed Logit Model

The mixed multinomial logit model is derived in much the same way as the multinomial probit model from the utility function specified in equation 4. The only differences are in how we define δ_i and ε_{ij}. For MXL each element of ε is assumed to be IID extreme value. However, δ is allowed to have any distribution that we desire. We can see that MNL is a special case of MXL, with $\delta_i = 0$ for all i. The choice probabilities for each individual are calculated as follows.

Denote the density of δ as $f(\delta|\theta)$, where θ are the fixed parameters of the distribution. Given δ_i, the probability that individual i selects alternative j is simply MNL.

$$P_i(j|\delta_i) = \frac{\mathrm{e}^{\beta' x_{ij} + \delta_i' z_{ij}}}{\sum_{k \in C_i} \mathrm{e}^{\beta' x_{ij} + \delta_i' z_{ij}}} \tag{8}$$

If we knew the value of δ_i for each voter the solution to equation 8 for each individual would be trivial. However, without this information estimation of equation 8 will generally not be possible. Since δ_i is a vector of Z random variables (the number of elements in z_{ij}, the choice probability is conditional on Z random variables. Without some information about these random variables it is not possible to determine the choice probability in equation 8.

However, if the distribution of the Z random variables is known then it is possible to determine the unconditional choice probability. To obtain the unconditional choice probability of individual i selecting alternative j we must integrate equation 8 with respect to the Z random variables.

$$P_i(j) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\frac{\mathrm{e}^{\beta' x_{ij} + \delta_i' z_{ij}}}{\sum_k \mathrm{e}^{\beta' x_{ik} + \delta_i' z_{ik}}} \right] f(\delta_1) \ldots f(\delta_Z) \partial\delta_1 \ldots \partial\delta_Z \tag{9}$$

From equation 9 it is easy to see how mixed multinomial logit receives its name. The choice probability is a mixture of MNL probabilities, with the weight of each particular MNL probability determined by the Z mixing distributions. The intuition behind this estimator is actually quite simple. Let us assume for a moment that δ_i consists of a single random variable, and that variable follows a known probability distribution $f(\delta_i|\theta)$. For each individual i we calculate the probability that individual i selects alternative j for every possible value of δ_i, and weight the contribution of each calculated probability to the likelihood function by the density $f(\delta_i|\theta)$. Changing the value of δ_i for individual i will result in different calculated probabilities of selecting alternative j; the impact each of these calculated probabilities has on the overall likelihood function is determined by the density of $f(\delta_i|\theta)$ at each particular value of δ_i. As the variance of $f(\delta_i|\theta)$ increases more weight is assigned to values of δ_i that are further from the mean (zero). This might improve the calculated likelihood for some individuals, since their particular value of δ_i may be far from the mean. If enough individuals are far from the mean then MXL will estimate a statistically significant coefficient for δ_i. The details of this estimation procedure are in the following section. As we shall see below, these estimates of δ are useful both in relaxing the IIA assumption and in examining unobserved heterogeneity across individuals. It is important to note that regardless of how MXL is specified (error-components or random-coefficients) it does not assume IIA. The unobserved component of utility for individual i for alternative j is $e_{ij} = \delta i' z_{ij} + \varepsilon_{ij}$. This term is necessarily correlated over alternatives, since δ_i is constant across alternatives for each individual. This means that $\text{Cov}(e_{ij}, e_{il})$ is not zero for $j \neq l$. Further, if the components of z vary across alternatives, $\text{Var}(e_{ij}) \neq \text{Var}(e_{il})$ for $j \neq l$. Thus by estimating the parameters of δ the unobserved portions of utility will no longer be distributed IID extreme value, and IIA will not hold.

Mixed logit is extremely flexible. For instance, MXL could be configured to represent a multinomial probit model by specifying δ to be a normal distribution. The MNP models correctly in use in political science can then be replicated as in the example in the previous subsection — the only difference is that ε is now specified to be IID extreme value rather than normal. Brownstone and Train (1999) find that MXL can approximate MNP probabilities very accurately (in fact, more accurately than the Geweke-Hajivassiliou-Keane (GHK) probability simulator when both are constrained to use the same amount of computer time). Nested logit models can also be replicated with MXL. To do this specify z_{ij} as a matrix of dummy variables of dimension $J \times N$, where N is the number of nests. Each dummy variable N has a value of 1 if alternative J is in that nest, and 0 otherwise. This will introduce correlations within each nest, but preserve the IIA property between nests. In fact, McFadden and Train (1998) demonstrate that MXL can be specified to approximate *any* discrete choice model derived from random utility maximization with the appropriate choices of δ and z_{ij}.

The choice probability for individual i selecting alternative j is given in equation 9. Solving this equation involves the evaluation of an integral of dimension Z, where Z is the number of random variables in δ. Because the distributions of the elements in δ are not tied to the number of alternatives (unless we specify them to be alternative-specific error terms) Z can be any number. This is unlike the MNP model, where integrals of dimension $J - 1$ must be estimated. This is one of the advantages of MXL over MNP.

In some cases it may be possible to account for IIA violations or sufficiently examine unobserved heterogeneity by estimating less than $J - 1$ elements of unobserved utility. Specifications of MXL in these situations that set $Z < (J - 1)$ will be easier to estimate than MNP models for the same problem, which must evaluate $J-1$ dimensional integrals regardless of the types of IIA violations or unobserved heterogeneity. Similarly, there may be situations where we wish to examine unobserved heterogeneity in more than $J-1$ variables, or believe that IIA violations are particularly complex. In this case MXL can be specified with $Z > (J-1)$.

Another important advantage of MXL over MNP is that the distribution of δ is general. Thus MXL can account for non-normal distributions of coefficients in a random-coefficients framework, or IIA violations that stem from non-normal disturbances. In contrast, random coefficients and error components in the MNP must be normally distributed.

3 Estimation of Mixed Logit Models

Both MNP and MXL models are difficult to estimate relative to MNL. The choice probabilities in both models depend on multidimensional integrals. For MNP, if there are J alternatives then we must evaluate an integral of dimension $J-1$. For MXL, if δ has Z random variables then we must evaluate an integral of dimension Z. As J increases for MNP, or as Z increases for MXL, the likelihood functions will no longer have closed form solutions, meaning that standard maximum likelihood techniques are no longer feasible[2]. Thus both MNP and MXL are generally estimated using simulated maximum likelihood (SMLE). The most common SMLE method for estimating MNP models in political science is the GHK probability simulator. This method is described elsewhere in the econometric and political science literature (Alvarez and Nagler 1998, Geweke 1991, Hajivassiliou and McFadden 1990, Keane 1990, Lawrence 1997), and will not be covered here. Below I describe the SMLE method for estimating MXL models.

To estimate the mixed logit model we set up the following log-likelihood function:

$$\mathcal{L}(\hat{\theta}) = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}\log\left[\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left\{\frac{e^{\beta' x_{ij}+\delta_i' z_{ij}}}{\sum_k e^{\beta' x_{ik}+\delta_i' z_{ik}}}\right\} f(\delta_1)\ldots f(\delta_Z)\partial\delta_1\ldots\partial\delta_Z\right] \tag{10}$$

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

[2]Note that this is why nobody has attempted a "mixed probit" model, with a general δ and ε IID standard normal. Estimating this model would involve the estimation of a $(J-1)+Z$. dimensional integral. Unless there is a strong theoretical reason to believe that the IID disturbances are normal the MXL is superior to a "mixed probit" model due to its greater feasibility of estimation.

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

The integral in the log-likelihood function can be evaluated numerically if $Z < 3$. However, when $Z \geq 2$ simulated maximum likelihood must be employed.

MXL choice probabilities are simulated by using a Monte Carlo simulation technique to approximate the integrals, and then by maximizing the resulting simulated log-likelihood function. This Monte Carlo simulation technique approximates the choice probabilities in equation 10 by computing the integrand at randomly chosen values for each δ_{i_z}. For each individual, for each random parameter, we first draw a random variable from the distribution we have assumed the particular element of δ follows. If the elements of δ are independent across individuals and alternatives we generate a matrix of $I \times Z$ independent random variables drawn from the appropriate distributions ($f(\delta_{i_z}|\theta)$) for the zth random parameter)[3]. The corresponding choice probabilities are then computed for a given value of the parameter vector θ. This process is repeated R times for the given value of the parameter vector.

Let $\hat{P}^r_{ij}(\theta)$ be the realization of the choice probability for the rth draw. The choice probabilities for a given parameter vector θ are approximated by averaging over the values of $\hat{P}^r_{ij}$:

$$\hat{P}_{ij}(\theta) = \frac{1}{R}\sum_{r=1}^{R}\hat{P}^r_{ij}(\theta) \tag{11}$$

$\hat{P}_{ij}(\theta)$ is the simulated choice probability of individual i choosing alternative j given θ. This simulated choice probability is an unbiased estimator of the actual probability $P_{ij}(\theta)$, with a variance that decreases as R increases. It is also twice differentiable and strictly positive for any realization of the finite R draws. These properties are especially appealing because they imply that log-likelihood functions constructed with $\hat{P}_{ij}(\theta)$ are always defined and can be maximized with

[3]Assuming that the elements of δ are independent is not necessary, however.

conventional gradient-based optimization methods.

Thus, we construct a simulated log-likelihood function:

$$\mathcal{SL} = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij} \log\left[\hat{P}_{ij}(\theta)\right] \tag{12}$$

The parameter vector θ is the vector that maximizes the simulated log-likelihood function. Under weak conditions this estimator is consistent, asymptotically efficient, and asymptotically normal (Lee 1992). However, this estimator does display some bias at low values of R, which decreases as R increases. The bias is exceedingly small when $R = 250$ (Brownstone and Train 1999); most empirical work uses R equal to 500 or 1000.

Although comparisons between MXL and MNP in terms of speed of estimation depend on many factors, such as the characteristics of the dataset and model specification, in general mixed logit models can be estimated at least as quickly as comparable multinomial probit models. The GHK simulator commonly used to estimate MNP models is recursive, meaning that the range for the random draw for one alternative depends on the value of previous draws for other alternatives. The probability simulator for MXL draws simultaneously for all probabilities from unrestricted ranges, speeding the process of creating the simulated log-likelihood. Brownstone and Train (1999) find that for a fixed amount of computer time, the mixed logit simulator has lower simulation variance that the GHK simulator, leading to more accurate probability estimates in the same amount of time. Unfortunately, the speed advantage of the MXL simulator over the GHK simulator is not very noticeable in many applications.

However, an alternative simulator proposed by Bhat (1999a) and Train (1999a) shows promise of dramatic decreases in 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 (1999), 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 N that defines the sequence and dividing a unit interval into N even parts. The dividing points become the first $N-1$ elements in the Halton sequence. Each of the N portions of the unit interval is divided as the entire unit interval was (creating $N \times (N-1)$ elements), and so on. Halton sequences give an even distribution of points across the unit interval.

Halton sequences can be used in place of random draws to estimate mixed logit models. One Halton sequence is defined for each random variable in the model, and the choice probabilities are calculated at each point in the Halton sequence, rather than at each randomly drawn value. Estimation is otherwise identical to that when using random draws to evaluate the integrals.[4] Preliminary tests of Halton sequences for estimation of mixed logit models found them to be vastly superior to random draws. Both Bhat (1999a) and Train (1999) find that the simulation error in estimated parameters is lower with 100 Halton draws than with 1000 random draws. In this example, using a Halton sequence in place of random draws allows us to obtain more accurate estimates at a fraction of the estimation cost of using random draws. Thus the use of Halton sequences in estimation gives MXL a clear advantage in estimation speed over MNP, while sacrificing none of the flexibility that makes mixed logit theoretically appealing.

4 An Empirical Application to the 1987 British Election

The advantages of mixed logit models in the study of multiparty elections is most easily demonstrated through an empirical application. For this application I have selected data from the 1987 British general election survey (Heath 1989). This is the same dataset used by Alvarez and Nagler in their demonstration of the properties of MNP relative to those of MNL (1998). In the empirical applications below I maintain the coding utilized in that paper and an earlier study (Alvarez, Bowler, and Nagler 1996), except where noted below.

[4] Bhat (1999a) refers to this estimation technique as "quasi-random simulated likelihood estimation".

The Labour Party did not fare well in the 1987 election, gathering only 30.8% of the vote, compared to 42.2% for the Conservatives and 22.6% for the Alliance. This was the second poor showing for Labour in the 1980s (they only won 27.6% of the vote in 1983), prompting many scholars to speculate on the causes of the Labour Party's decline.

One theory holds that the 1980s witnessed a fragmentation of the working class, the traditional supporters of the Labour Party. Crewe (1987) maintained that the old, Labour supporting working class was on the decline, and was being replaced by a new, more affluent and Conservative-leaning working class. In the past Labour had relied on support from the traditional working class communities in the North, with a heavy preponderance of trade union members and council estate residents. However, by the 1980s the picture of a typical member of the working class was changing: he or she was more likely to live in the South, own his or her own residence, and work in the private sector. This new, rising working class was less inclined to vote Labour than the old, declining working class. As the composition of the working class changed Labour support eroded. Crewe argues that by 1987 Labour "... was a party neither of one class nor one nation; it was a regional class party" (Crewe 1987).

In contrast, Heath and Jowell (1987) argue that there has been no clear fragmentation of the working class. Although it is undeniable that there has been a decline in *absolute* class voting, the working class has not splintered any more than any other class. The poor showing of the Labour Party in the 1983 and 1987 elections has more to do with the political positions supported by Labour rather than the dissolving of working class solidarity. In 1987, as in 1983, Labour adopted issue positions that were too far left to make them attractive to most voters. Instead, voters found the comparatively moderate positions of the Conservative and Alliance Parties more appealing. Working class voters who abandoned Labour did so because of the relative issue positions of the parties, not because of a shift in class interests.

Each of these theories offers a different prediction for the impact of class on voting. If the working class was indeed fragmenting along the lines described by Crewe in 1987 we should observe some segments of the working class continuing to vote Labour *because of their class*, while other segments abandon class voting. In particular, individuals from the North, trade union members, and those living on council estates (the old working class) should remain strong class voters and cast their votes for Labour, while other working class individuals (the new working class) should be less constrained by class and more willing to vote for the Conservative or Alliance Parties. Thus, we would expect the impact of class on the vote for members of the old working class to be very *homogeneous*, and favoring Labour, while the impact of class on the vote for the new working class will be more *heterogeneous* in nature.

However, if it is the issue positions of the parties that are driving the decline in Labour vote, and class voting is less prevalent than before, then we would expect to see *heterogeneity* in the impact of class on vote choice for both the old and new working classes. The driving force behind the decline in the Labour vote in this case in not the fragmentation of the working class, but the limited electoral appeal of the issue positions that Labour chose to endorse. More moderate working class individuals thus abandon Labour for one of the other parties, while others continue to vote their class interests.

To test between these two theories I estimated a mixed logit model of the 1987 British general election using the dataset described above. Issue distances are treated as alternative-specific variables, while all other variables are individual-specific. Unlike the MNL and MNP estimates presented by Alvarez, Bowler, and Nagler (1996) and Alvarez and Nagler (1998), I normalize the individual-specific coefficients with respect to the Labour party. This allows for direct examination of the impact of class on vote choices involving the Labour party in relation to both the Conservative Party and the Alliance.

In order to examine heterogeneity in the impact of class on the Labour vote, I estimate 4 coefficients as distributions rather than fixed. The first set of coefficients is on the effect of union membership on vote choice (for both Conservative relative to Labour and Alliance relative to Labour). This variable is expected to capture the impact of class on vote behavior for the old working class. The second set of variables that do not have fixed coefficients are for blue collar workers who are not union members (again, for the probability of voting Conservative relative to voting Labour, and the probability of voting Alliance relative to Labour). The distributions are specified as triangular distributions, with mean m and "spread" s. These distributions have zero density below $m - s$, linearly increasing density from $m - s$ to m, linearly decreasing density from m to $m + s$, and zero density above $m + s$. I selected triangular distributions rather than the more familiar normal because normal distributions have unbounded tails, which would imply that some individuals in the distribution could have an infinite coefficient for the impact of class on voting. Triangular distributions are bounded, and thus do not suffer from this problem.[5]

If the decline in Labour voting is due to the fragmentation of the working class, we would expect to see heterogeneity in the impact of class on non-union blue collar workers (the new working class) as individuals abandon their class ties, while union workers (the old working class) remain fairly homogeneous. However, if it is the relative extremity of Labour's issue positions that is leading to the decline in the Labour vote, we would expect to see heterogeneity in *both* groups, as individuals abandon their class ties to vote for the more moderate Conservative and Alliance Parties. The results of estimating the mixed logit model described above with 125 Halton draws are presented in table 1.

[Table 1 here]

The results for the fixed coefficients in this model are nearly identical to those reported by Alvarez and Nagler (1998) in a MNL model on this data, although it is hard to see except in the case of the alternative-specific variables (the issue distances). When one considers that MNL is a special case of MXL with all degenerate distributions this result is not surprising.

Examining the results for the random coefficients reveals a great deal of information about the impact of class on the Labour vote. The mean and spread of each of the 4 triangularly distributed random coefficients are reported near the bottom of the table. The spread coefficient indicates the distance the endpoints of the triangular distribution are from the mean. The mean coefficients for the impact of class for both the old and new working class are negative and significant, indicating that on average both groups are more likely to vote Labour than vote for another party. If we were

[5]Note that a random-coefficients MNP specification could not duplicate this model for 2 reasons. First, the number of random coefficients must equal $J - 1$; in this case a random-coefficients MNP would be limited to 2 random coefficients. Second, the random coefficients in a random-coefficients MNP must be normally distributed.

to restrict our attention to the means (as we are forced to do in fixed coefficient models) it would appear that working class support is firmly behind the Labour Party.

However, the mixed logit model also reveals a great deal of heterogeneity in the impact of class on vote choice for both the old and new working class. This indicates that even though a working class individual is expected to support Labour on average, many working class individuals are not voting for Labour. Thus the appearance of solid working class support that the means of the distributions hint at is revealed to be much less secure.

Further, notice that statistically significant heterogeneity is discovered in the impact of class on vote choice for both the old and the new working class. This indicates that the decline of Labour support is not due solely to a fragmentation of the working class. Although there is strong evidence that the new working class is beginning to drift away from Labour, many members of the old working class seem to be following. Thus the ability to estimate random-coefficients models with mixed logit can contribute to our understanding of politics.

However, the main motivation behind using the MNP in political science is to account for violations of IIA. If we wish to examine the impact the Alliance had on the electoral outcome in 1987 this is a critical point — if IIA is violated but not accounted for any inferences about a two-party race between the Labour and Conservative Parties are suspect. Alvarez and Nagler (1998) compare the predicted three- and two-party vote shares estimated by MNL and MNP on this dataset. They demonstrate that accounting for IIA violations changes the predicted vote shares, indicating that the IIA assumption of MNL is likely leading to inaccurate predictions.

Mixed logit models also allow for violations of IIA, and are thus also suitable for making inferences about a hypothetical two-party race. The predicted two-party vote stemming from the model in table 1 is likely to differ slightly from that of a MNP model because of the former's random-coefficients specification. MNP models as the are generally estimated in the political science literature allow for violations of IIA by estimating the covariance matrix of the unobserved portion of utility. This is a type of error-components specification, where the error components include all unobserved portions of utility. In the random-coefficients specification presented in table 1 the only portion of unobserved utility estimated is the heterogeneity in the 4 random coefficients. Some unobserved portions of utility will remain unaccounted for, and the effect this will have on our attempts to control for IIA will depend entirely on how much of the unobserved portions of utility were captured by the random coefficients. However, with the right specification MXL models can approximate any substitution pattern among alternatives. In order to demonstrate that MXL can approximate the choice probabilities of a MNP model I specified another MXL model, identical to the one in table 1 except that the alternative-specific constant terms are now estimated as normal distributions. The results of this estimation, using 125 Halton draws, are presented in table 2.

[Table 2 here]

This mixed logit model accounts for all unobserved portions of utility in estimation. Note that the normally distributed coefficients on the constants do not account for the unobserved portions of utility in exactly the same way as the covariance matrix in the MNP, since I have not allowed them to be correlated. However, models that do allow this are feasible.

The predicted vote shares in three- and two-party races are presented in table 3. The predicted vote shares for MNL and MNP come from Alvarez and Nagler (1998). I also report the predicted vote shares for the original MXL model (MXL1), and the MXL model that adds the normally distributed constant terms (MXL2).

[Table 3 here]

Clear differences emerge between the model which assumes IIA (the multinomial logit) and those which do not. The first MXL model has predicted two-party vote shares that fall between those calculated by MNL and MNP. This indicates that while this random-coefficients specification does account for IIA violations, it does not include all unobserved portions of utility that affect how voters view the parties relative to each other. The second MXL model, however, produces predicted two-party vote shares nearly identical to those estimated by MNP. This indicates that this specification of mixed logit is accounting for violations of IIA as effectively as MNP, and still yielding valuable information on the heterogeneity of several coefficients. Thus it is apparent that mixed logit models are a tractable and flexible alternative to multinomial probit in the study of multiparty elections.

5 Discussion

MXL offers two advantages over MNP. The first is computational — all multinomial probit models are estimated by evaluating a multivariate normal density of dimension $J - 1$. If substitution patterns can be specified correctly with fewer than $J-1$ error components, then a mixed multinomial logit model with fewer than $J-1$ elements in z will be easier to estimate than a multinomial probit. Further, the use of Halton sequences in the estimation of mixed logit models can make the estimation of a MXL model much faster that estimation of a comparable MNP model. Secondly, and more importantly, MXL is not tied to the normal distribution. In a random-coefficients context this is important, since there may be theoretical reasons to believe that the distribution of a parameter is non-normal. For instance, in a spatial voting model the estimated weights on the distances between candidates and voters should never be positive — however, with a normally distributed random parameter weights will be positive for some proportion of the population. The ability to incorporate distributions other than the normal into MXL also has advantages in modeling different substitution patterns (IIA violations). MXL is completely flexible in this respect. McFadden and Train (1998) have demonstrated that *any* random utility model, with any substitution patters, can be approximated arbitrarily closely by MXL with the appropriate choice of the z's and $f(\delta|\theta)$. The empirical application to the 1987 British general election demonstrates that mixed logit is a feasible and valuable tool for the study of multiparty elections.

6 Bibliography

Alvarez, R.M., S. Bowler, and J. Nagler (1996). "Issues, Economics, and the Dynamics of Multi-Party Elections: The British 1987 General Election". Pasadena, CA: California Institute of Technology Social Science Working Paper 949.

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.

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

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.

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

Bunch, D.S. (1991). "Estimability in the Multinomial Probit Model." *Transportation Research B*, 25:1, pp. 1-12.

Crewe, I. (1987). "A New Class of Politics" and "Tories Prosper from a Paradox", *The Guardian*, 15/16 June.

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. (1989). *British Election Study, 1987. A Computer File.* Colchester: ESRC Data Archive.

Heath, A. F. and R. Jowell (1991). *Understanding Political Change: The British Voter 1964-1987.* Pergamon Press: New York, NY.

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

Keane, M.P. (1992). "A Note on Identification in the Multinomial Probit Model." *Journal of Business and Economic Statistics*, 10:2, pp.193-200.

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.

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

McFadden, D. and K. Train (1998). "Mixed MNL Models for Discrete Response." Unpublished manuscript.

Revelt, D. and K. Train (1999). "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level." *The Review of Economics and Statistics*, forthcoming.

Train, K. (1986). *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand.* MIT Press: Cambridge, MA.

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

Whitten, G.D. and H.D. Palmer (1996). "Heightening Comparativists' Concern for Model Choice: Voting Behavior in Great Britain and the Netherlands." *American Journal of Political Science*, 40:1, pp. 231-60.

Table 1. Mixed Multinomial Logit Estimates, 1987 British Election

Independent Variables	Conservatives/Labour	Alliance/Labour
Defense	-0.20**	
	(0.02)	
Unemployment/Inflation	-0.13**	
	(0.03)	
Taxation	-0.18**	
	(0.03)	
Nationalization	-0.20**	
	(0.02)	
Redistribution	-0.09**	
	(0.02)	
Crime	-0.11**	
	(0.05)	
Welfare	-0.15**	
	(0.02)	
Constant	-2.31**	-3.03**
	(0.93)	(0.77)
South	0.30	0.43*
	(0.25)	(0.22)
Midlands	-0.13	0.18
	(0.25)	(0.22)
North	-0.77**	-0.72**
	(0.23)	(0.21)
Wales	-1.98**	-1.46**
	(0.41)	(0.33)
Scotland	-1.18**	-0.75**
	(0.32)	(0.27)
Public Sector Employee	0.13	0.02
	(0.19)	(0.17)
Female	0.39*	0.04
	(0.18)	(0.16)
Age	0.28**	0.24**
	(0.06)	(0.05)
Home Ownership	1.04**	0.58**
	(0.22)	(0.18)
Family Income	0.14**	0.06*
	(0.04)	(0.03)
Education	0.00	0.88**
	(0.44)	(0.36)
Inflation	0.32**	0.05
	(0.13)	(0.12)
Unemployment	0.12	0.11
	(0.08)	(0.07)
Taxes	0.29**	-0.03
	(0.08)	(0.08)
Union Member (Mean)	-1.38**	-0.86**
	(0.24)	(0.19)
Union Member (Spread)	2.64**	0.14
	(0.93)	(1.42)
Blue Collar (Mean)	-0.73**	-1.21**
	(0.23)	(0.37)
Blue Collar (Spread)	3.38**	3.32*
	(0.92)	(1.58)
Number of Observations	2131	
Log-Likelihood	-1473.97	

Standard errors in parentheses. ** indicates statistical significance at the 99% level; * indicates statistical significance at the 95% level. Random coefficients have triangular distributions.

Table 2. Mixed Multinomial Logit Estimates,
1987 British Election
(Alternative Substitution Patterns)

Independent Variables	Conservatives/Alliance	Labour/Alliance
Defense	-0.22**	
	(0.03)	
Unemployment/Inflation	-0.15**	
	(0.04)	
Taxation	-0.20**	
	(0.04)	
Nationalization	-0.23**	
	(0.03)	
Redistribution	-0.11**	
	(0.03)	
Crime	-0.12*	
	(0.06)	
Welfare	-0.18**	
	(0.03)	
South	0.33	0.48*
	(0.21)	(0.24)
Midlands	-0.17	0.20
	(0.28)	(0.24)
North	-0.85**	-0.77**
	(0.28)	(0.24)
Wales	-2.18**	-1.55**
	(0.50)	(0.38)
Scotland	-1.32**	0.80**
	(0.38)	(0.30)
Public Sector Employee	0.15	0.02
	(0.22)	(0.18)
Female	0.44*	0.05
	(0.21)	(0.17)
Age	0.30**	0.26**
	(0.08)	(0.06)
Home Ownership	1.18**	0.60**
	(0.29)	(0.20)
Family Income	0.15**	0.06*
	(0.05)	(0.04)
Education	-0.01	1.04**
	(0.50)	(0.42)
Inflation	0.36**	0.04
	(0.16)	(0.13)
Unemployment	0.13	0.11
	(0.09)	(0.08)
Taxes	0.34**	-0.06
	(0.10)	(0.09)
Union Member (Mean)	-1.55**	-0.88**
	(0.31)	(0.21)
Union Member (Spread)	2.95**	0.23
	(1.21)	(1.72)
Blue Collar (Mean)	-0.80**	-1.37**
	(0.21)	(0.43)
Blue Collar (Spread)	3.80**	4.11**
	(1.11)	(1.75)
Constant (Mean)	-2.65**	-3.38**
	(1.09)	(0.90)
Constant (Std. Deviation)	1.09*	0.57
	(0.54)	(0.55)
Number of Observations	2131	
Log-Likelihood	-1472.88	

Standard errors in parentheses. ** indicates statistical significance at the 99% level; * indicates statistical significance at the 95% level. Random coefficients have triangular distributions for Union and Blue Collar, and normal distributions for Constant.

Table 3. Estimated Aggregate Vote Shares, Three-Party and Two-Party Races

	MNL	MNP	MXL1	MXL2
Three-Party Race				
Conservative	45.2	44.9	45.0	45.0
Labour	29.5	29.8	29.9	29.9
Alliance	25.3	25.2	25.1	25.1
Two-Party Race				
Conservative	59.1	57.5	58.2	57.4
Labour	40.9	42.5	41.8	42.6