Mixed Logit Models for Multiparty Elections
Author(s): Garrett Glasgow
Source: *Political Analysis*, Vol. 9, No. 2 (Spring 2001), pp. 116-136
Published by: Oxford University Press on behalf of the Society for Political Methodology
Stable URL: http://www.jstor.org/stable/25791636

Mixed Logit Models for Multiparty Elections

Garrett Glasgow
Department of Political Science,
University of California, Santa Barbara,
Santa Barbara, CA 93106
e-mail: glasgow@sscf.ucsb.edu
http://www.polsci.ucsb.edu/faculty/glasgow

Mixed logit (MXL) is a general discrete choice model thus far unexamined in the study of multicandidate and multiparty elections. Mixed logit assumes that 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 voter behavior through random-coefficients specifications. MXL is a more general discrete choice model than multinomial probit (MNP) in several respects, and can be applied to a wider variety of questions about voting behavior than MNP. An empirical example using data from the 1987 British General Election demonstrates the utility of MXL in the study of multicandidate and multiparty elections.

1 Introduction

ONE OF THE most active fields of research in political methodology in recent years has been the modeling of multicandidate and multiparty elections. As most elections in the world have more than two candidates or parties competing for office, accurate modeling of such elections is crucial to our understanding of politics. 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] MXL models have seen application in marketing and transportation research (Algers et al. 1998; Bhat 1998a, b; Brownstone and Train 1999; Jain et al. 1994; Revelt and Train 1998; Train 1998) but have not been applied to political science problems to date.

In the study of multiparty elections the multinomial probit (MNP) is an increasingly popular choice (Alvarez and Nagler 1995, 1998a, b; Lacy and Burden 1999, 2000; Lawrence 1997; Quinn et al. 1999; Schofield et al. 1998). The primary motivation for using MNP in the

Author's note: I thank Jim Adams, R. Michael Alvarez, Andy Baker, Fred Boehmke, Scott Desposato, Gary King, Eric Lawrence, Jonathan Nagler, Phil Paolino, Kevin Quinn, Kenneth Train, Chris Winship, and two anonymous reviewers for their helpful comments.

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

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 to 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).

MXL 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. Random utility models are predominant in the empirical study of multiparty and multicandidate elections. These models are based on the assumption that a voter's preferences among candidates or parties can be described by a utility function. This utility function depends on the attributes of the alternatives (the candidates or parties) and the characteristics of the individual (the voter). When voting, individuals select the candidate or party that yields the highest utility.

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

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

where V_{ij} represents the systematic (observed) portion of utility, and e_{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 e_{ij}.

MNL assumes that the unobserved portions of utility e_{ij} are identically and independently distributed (IID) in accordance with the extreme value distribution.[2] It is well known that

[2]I use the term "multinomial logit" generically, intending it to refer to both MNL (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

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 specify completely the relationship between the dependent and the 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 model explicitly 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.

Two possible violations of the IID assumption are considered here.[3] 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 U.S. 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 nonexistent. Each individual places his or her 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" either for the attributes of alternatives or for the relationship between individual characteristics and alternatives vary randomly across individuals.

this terminology into line with the accepted terminology for the equivalent probit model, which is referred to as "multinomial probit" regardless of which kind of variables are included.

[3]There is a third possibility. 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, and since most studies of multiparty and multicandidate elections are cross sectional, I do not address this possible violation here.

Both MNP and MXL 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. MNP 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 MNL 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 MNL 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 that 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 to the study of multicandidate and multiparty elections, whether the IID violations are due to unobserved attributes or random taste variation.

In Section 2 I describe how MNP and MXL can both be derived from a common utility framework. When presented in this way it is apparent that MNP and MXL (and all 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 two MXL models for the 1987 British General Election. The first model is designed to estimate realistic substitution patterns similar to those of a MNP, while the second examines heterogeneity across individuals in the impact of social class on the vote. This section reveals that MXL is able to address a wide range of questions about multiparty elections. Section 4 concludes, explaining why MXL is a useful alternative to MNP in the study of multicandidate and multiparty elections.

2 Specification of the Mixed Logit and Multinomial Probit Models

Assume that an individual faces a choice set consisting of J alternatives. Let the utility that individual i receives from alternative j be denoted 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 either are 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 $e_{ij} = z_{ij}\eta_i + \varepsilon_{ij}$, 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 $g(\eta \mid \Omega)$, 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[4]

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

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

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 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 that 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 MNP and MXL 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 that 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 MNP 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 that no a priori constraints are imposed. This particular specification of MNP is dominant in political science (Alvarez and Nagler 1995, 1998a, b; Alvarez et al. 2000; Lacy and Burden 1999, 2000; Lawrence 1997; Quinn et al. 1999; Schofield et al. 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 that $Z = X$ 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.

MNP allows for a more realistic formulation of models of political behavior than discrete choice models that assume that 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 $[(J \times (J - 1))/2] - 1$ elements in Σ are identified. This means that MNP can estimate at most $[(J \times (J - 1))/2] - 1$ 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 nonnormal 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 data set 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 a beta distribution. MNP does not allow for nonnormal 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 that ε is IID extreme value, while η follows a general distribution, $g(\eta \mid \Omega)$. If $\eta = 0$ this is MNL, which has the IIA property. Estimation of the MXL generally involves estimating 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:

$$P(j \mid \eta) = \frac{e^{X_j\beta_j + Z_j\eta}}{\sum_{k\in J} e^{X_k\beta_k + Z_k\eta}} \tag{4}$$

If the value of η were 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, 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.

$$P(j) = \int_\eta \left[\frac{e^{X_j\beta_j + Z_j\eta}}{\sum_{k\in J} e^{X_k\beta_k + Z_k\eta}}\right] g(\eta \mid \Omega)\partial\eta \tag{5}$$

Examination of Eq. (5) reveals that 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, introducting correlation in the utility across alternatives at the individual level (note that this is also true of the MNP specification in the previous subsection).

The unconditional probability that individual i selects alternative j 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 MXL 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 $g(\eta \mid \Omega)$ for each individual. The draws of η can be taken randomly or by using Halton sequences (a nonrandom 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 $\hat{P}(j \mid \eta)$, the conditional choice probability given in Eq. (4). This process is repeated R times, and the integration over $g(\eta \mid \Omega)$ is approximated by averaging over the R draws. The resulting simulated choice probability $\hat{P}(j \mid \beta, \Omega)$ 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 MXL models.

MXL 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 is 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 a 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 MXL the elements of η can follow any distribution. MNP requires that $g(\eta \mid \Omega)$ 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 MXL a more general discrete choice model than MNP. A MXL 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.[5] Conversely, MNP is unable to estimate models that approximate MXL under many specifications. Thus, MXL can be specified to answer many questions about multicandidate and multiparty elections that MNP cannot, and therefore have not been addressed previously.

3 An Empirical Application to the 1987 British General Election

The advantages of mixed logit models in the study of multiparty elections are most easily demonstrated through an empirical application. In this example I use data from the 1987 British General Election survey (Heath et al. 1989). This same data set has also been examined by Alvarez and Nagler (1998) and Alvarez et al. (2000) with a MNP model. Those papers use an error-components MNP to account for common unobserved attributes and relax the IIA property. Below I 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

[5]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.

same substitution patterns as MNP. MXL is also capable of studying random taste variation. Thus, I also estimate a random-coefficients MXL designed to study heterogeneity in the impact of social class on voting.

I begin by presenting an error-components MXL designed to account for the same violations of the IID assumption that lead to the MNP specification most common in political science. To facilitate this comparison I use the same variable coding as described by Alvarez et al. (2000) throughout this section (see Appendix B at the Political Analysis website). The impact of issue distance is assumed to be constant across all three parties included in the model (Conservative, Labour, and Alliance), while the individual-specific variables are normalized such that the coefficients for voting Alliance are zero.

To specify a MXL 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 a MXL 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. 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 the 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. 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 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 second column and the coefficients for voting Labour relative to the Alliance

Table 1 Mixed logit estimates (MNP Replication), 1987 British General Election (Alliance coefficients normalized to zero)

Independent variable	*Conservative/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 (—)
SD	2.30 (—)		1.72** (0.30)
$\sqrt{\text{Cov.}}$		0.43 (0.69)	
Number of observations		2131	
Log-likelihood		−1474.70	

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

in the fourth column. The error components are presented below the individual-specific coefficients.[6]

Examination of Table 1 reveals that the mean effects of the variables in the model are similar to those estimated by Alvarez et al. (2000). Short-term political effects, the relative issue positions of the parties, 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 MXL model in Table 1 are similar to those estimated by the MNP specification common in political science. The vote shares reported here are

[6]This model and the model in Table 2 were estimated using 125 Halton draws in the simulation of the integrals in the choice probabilities. See the Appendix for details.

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 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 that 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 that 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.

Both MNP and MXL 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 and 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 estimated using the same data set as above, which reveals random taste variation in the impact of social class on vote choice.

From World War I until the late 1960s social class was recognized as a powerful influence on party choice in Britain. Members of the "working class" (defined as those holding a manual labor occupation) voted predominantly for the Labour Party, while members of the "middle class" (those who held nonmanual occupations) voted predominantly for the Conservative Party. The strength of the relationship between class and political preference during this time prompted one scholar to write that "class is the basis of British party politics; all else is embellishment and detail" (Pulzer 1967).

However, by the 1980s the common wisdom held that Britain was experiencing a "class dealignment," especially within the working class (Crewe 1983a; Sarlvik and Crewe 1983; but see Heath et al. 1985). Most recent studies of elections in Britain have emphasized the growing impact of electorate-wide, short-term factors on vote choice, such as party issue positions, perceptions of the economy, and social reforms (Alvarez et al. 2000; Crewe 1992; Garrett 1992).

The decline of class as a determinant of the vote could also be deeply rooted in individual-level changes. By 1987 it is possible that many individuals no longer identified with "their" social class, or even with the concept of a class structure. For these individuals the impact of social class on the vote would be low or nonexistent. It is equally likely that many individuals clung to their class ties, and for these individuals the impact of social class on the vote would be high. Proxy variables for these differences in class identification (usually geographic region, home ownership, and working in the public sector) are crude measures of individual-level differences. Much preferred would be measures of the individual-level factors that lead voters to stop identifying with their social class, such as the relative affluence of the voter's family during childhood, the level of social mobility within a family or local geographic region leading to mixed-class households, and misgivings about trade union power in the Labour party. Unfortunately, measures of variables such as these are often unreliable or unavailable. This makes the impact of social class on the vote observationally random, with a mean impact measured by the variables that are observed and individual-level variation around this mean. This can be expressed as random taste variation for the relationship of social class to the vote (some voters prefer a strong relationship, while others prefer a weak relationship).

I examined random taste variation in two measures of social class in this dataset. The first is a measure of working-class occupation, coded 1 if the survey respondent held a working class occupation (any kind of manual, personal service, or agricultural work) and 0 otherwise. Many scholars believe that class ties were becoming weak within the working class by 1987, and a significant amount of random taste variation in the impact of this measure on the vote would tend to confirm this. The second measure is one of trade union membership, coded 1 if the respondent was a member of a trade union or staff association and 0 otherwise. Trade union members at this time were generally regarded as the core of the working class and staunch Labour supporters—a significant amount of random taste variation in the impact of this measure on the vote would indicate that class ties were weak even among those regarded as the strongest Labour supporters.

To test for random taste variation in the impact of social class on the vote, I estimated a random-coefficients MXL. The variables included and the coding of these variables were identical to those used in the MXL presented in Table 1. However, in place of the error components I specified four independent random coefficients. Two are the coefficients on the dummy variables that indicate if a voter was a member of the working class (both for voting Conservative relative to the Alliance and for voting Labour relative to the Alliance), and two are on the dummy variables that indicate if a voter was a member of a trade union (both for voting Conservative relative to the Alliance and for voting Labour relative to the Alliance). As the hypothesis I am testing maintains that some individuals have clung to their class ties while others have abandoned them, 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 $-a$, rises linearly to a mean m, descends linearly to the other endpoint a, and is zero beyond a. The parameters that describe this distribution are the mean and the distance between the mean and the endpoints $(m - |a|)$. 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.[7]

Note that this specification relaxes two of the constraints necessarily imposed on a random-coefficients MNP. First, the maximum number of error components or random coefficients for a MNP with three alternatives is two, while there are four random coefficients in this model. Second, the random coefficients in this MXL are nonnormal, while MNP can estimate only normally distributed random coefficients. The results of estimating this MXL are presented in Table 2.

Table 2 Mixed logit estimates (examining heterogeneity), 1987 British General Election (Alliance coefficients normalized to zero)

Independent variable	*Conservative/Alliance*		*Labour/Alliance*
Defense		−0.21** (0.02)	
Phillips curve		−0.13** (0.03)	
Taxation		−0.19** (0.03)	
Nationalization		−0.21** (0.02)	
Redistribution		−0.09** (0.02)	
Crime		−0.11* (0.05)	
Welfare		−0.16** (0.02)	
Constant	0.57 (0.76)		2.85** (0.80)
South	−0.13 (0.19)		−0.49* (0.23)
Midlands	−0.31 (0.19)		−0.23 (0.23)
North	−0.07 (0.20)		0.71** (0.21)
Wales	−0.59 (0.40)		1.46** (0.34)
Scotland	−0.48* (0.28)		0.78** (0.28)
Public sector employee	0.10 (0.17)		−0.02 (0.17)
Female	0.38** (0.16)		−0.03 (0.16)
Age	0.04 (0.05)		−0.24** (0.06)
Home ownership	0.53** (0.21)		−0.59** (0.18)
Family income	0.08** (0.03)		−0.07* (0.03)
Education	−0.83** (0.34)		−0.69* (0.37)
Inflation	0.28** (0.11)		−0.05 (0.13)
Taxes	0.01 (0.08)		−0.10 (0.07)
Unemployment	0.31** (0.07)		0.01 (0.08)
Working class			
Mean	0.09 (0.19)		0.80** (0.18)
Mean − endpoint	4.10** (0.71)		2.45** (0.80)
Union member			
Mean	−0.58** (0.19)		0.42** (0.17)
Mean − endpoint	0.10 (1.70)		0.14 (1.20)
Number of observations		2131	
Log-likelihood		−1467.61	

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

[7]The model in Table 2 was also estimated with normally distributed random coefficients, with similar substantive results.

Overall, the mean results produced by this MXL model are substantively similar to those produced by the MXL in Table 1 and the MNP discussed earlier.[8] In particular, the estimated mean impact of the social class variables on vote choice was substantively identical to that in those models. Working-class voters were more likely on average to vote for the Labour party relative to the Alliance, while appearing indifferent on average between the Conservatives and the Alliance. Trade union members were more likely to vote for the Labour party relative to the Alliance on average and more likely to vote for the Alliance relative to the Conservatives on average.

The estimated distance between the mean and the endpoints of the triangularly distributed coefficients reveals that there was a statistically significant amount of random taste variation in the impact of a working-class occupation on the vote. Although holding a manual labor occupation does not have an impact on the vote choice between the Conservatives and the Alliance on average, there is a statistically significant degree of variance in this impact. This indicates that for some individuals in this sample, considerations of social class played a role in the voting decision between the Conservatives and the Alliance. Examining the mean impact of class on the vote choice between the Conservatives and the Alliance without considering random taste variation leads to the conclusion that social class had no impact on that vote choice, when in fact it had a substantial impact for some individuals. Heterogeneity in the impact of class on voting is similarly found in the vote choice between the Labour party and the Alliance. The estimated distance between the mean and the endpoints of the triangularly distributed coefficients for the impact of trade union membership is not statistically distinguishable from zero, indicating that the hypothesis that the impact of trade unionism on vote choice was homogeneously pro-Labour cannot be rejected. Thus, class ties appear weak within the working class as a whole, but not among trade union members.

Estimates of random taste variation can reveal a great deal about voter behavior in an election. To demonstrate this I created a hypothetical voter, with characteristics set to the mean or modal values of the individuals in the data set. I then computed the probability of voting for each party for this hypothetical voter if he was a member of the working class, and if he was not a member of the working class, using both a MNP as it is usually specified in political science applications and the MXL presented in Table 2. For the MNP specification considered here there is only one set of predictions, since the effect of working-class membership on vote choice is assumed to be identical across all voters. However, with the random-coefficients MXL the impact on the vote of being in the working class will depend on where the hypothetical voter falls in the distributions of the coefficients on working class. Below I present five predictions from the MXL model presented in Table 2 for the hypothetical voter; with the impact of working class on the vote at the mean value for both the Conservative versus the Alliance vote and the Labour versus the Alliance vote, with the Conservative versus the Alliance working-class coefficient at the first and third quartiles (holding the Labour versus Alliance working-class coefficient at its mean), and with the Labour versus the Alliance working-class coefficient at the first and third quartiles (holding the Conservative versus Alliance working-class coefficient at its mean). 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.

[8]The substitution patterns estimated by this MXL model were similar to those estimated by both the MNP and the MXL in Table 1. In a three-party race this model predicted 44.97% (0.82%) Conservative, 29.82% (0.74%) Labour, and 25.21% (0.86%) Alliance. In a two-party race this model predicted 58.25% (0.81%) Conservative and 41.75% (0.81%) Labour. This result demonstrates that even if the random effects in η are independent, IIA does not hold for MXL due to the common influence of η across alternatives for each individual.

Table 3 Vote probabilities for a hypothetical voter

	Working class	*Non-working class*	*Difference*
Multinomial probit			
Conservative	29.89 (4.36)	33.64 (4.60)	3.75
Labour	38.03 (4.88)	24.84 (4.31)	−13.19
Alliance	32.08 (4.46)	41.52 (4.89)	9.44
Mixed logit			
At means			
Conservative	24.97 (5.21)	30.00 (5.29)	5.03
Labour	42.15 (6.23)	25.87 (4.81)	−16.28
Alliance	32.88 (4.80)	44.13 (5.09)	11.25
C/A at 1st quartile			
Conservative	9.33 (2.46)	30.23 (5.10)	20.90
Labour	50.93 (6.26)	25.61 (4.85)	−25.32
Alliance	39.74 (5.68)	44.16 (5.01)	4.42
C/A at 3rd quartile			
Conservative	52.07 (6.74)	30.11 (5.14)	−21.96
Labour	27.00 (5.43)	26.02 (4.94)	−0.98
Alliance	20.93 (3.81)	43.87 (5.10)	22.94
L/A at 1st quartile			
Conservative	31.76 (5.83)	30.27 (5.28)	−1.49
Labour	26.38 (4.92)	25.69 (4.76)	−0.69
Alliance	41.87 (5.31)	44.04 (5.18)	2.17
L/A at 3rd quartile			
Conservative	17.50 (4.25)	30.32 (5.28)	12.82
Labour	59.61 (6.20)	25.70 (4.86)	−33.91
Alliance	22.89 (4.04)	43.97 (5.21)	21.08

Note: Standard errors in parentheses.

MNP predicts that switching the hypothetical voter from a working-class occupation to a non-working-class occupation makes him less likely to support the Labour party, with the Alliance gaining the most probability of support. However, the predictions for the hypothetical voter vary widely in the MXL presented in Table 2, depending on the impact of social class on the vote for this particular individual. At the means of both distributions of the working-class coefficients the predicted vote probabilities display a pattern similar to that predicted by MNP when the hypothetical voter is moved out of the working class. However, if the coefficient value for the Conservative versus the Alliance vote is at the first quartile in the triangular distribution, this voter is far more likely to support Labour and less likely to support the Conservatives. Switching this voter out of the working class leads to a large drop in Labour support, with the Conservatives gaining most of that support. If this coefficient is at the third quartile in the distribution, if the hypothetical voter is in the working class he is *more* likely to support the Conservatives than any other party, and switching this voter out of the working class leads to a large gain for the Alliance, primarily at the expense of the Conservatives. If the coefficient value for the Labour versus the Alliance vote is at the first quartile of the triangular distribution, switching the hypothetical voter out of the working class has little effect on the predicted vote probabilities. However, if this coefficient is at the third quartile of the distribution, switching from the working class to the non-working class causes a large drop in the probability of supporting Labour, with the Alliance gaining about two-thirds of this.

These predictions generated by this MXL for the effect of social class on the vote are all for the same hypothetical voter. All individuals who match this profile would appear to be identical in the data set, and most models (such as the specification of MNP examined here) would predict that these individuals would behave in exactly the same way. However, this random-coefficients specification of MXL reveals that there is a great deal of random taste variation in the impact of social class on the vote. Some individuals in the working class view the relationship of the parties to social class as a key component of their vote choice, while others feel that this is irrelevant. Still other members of the working class

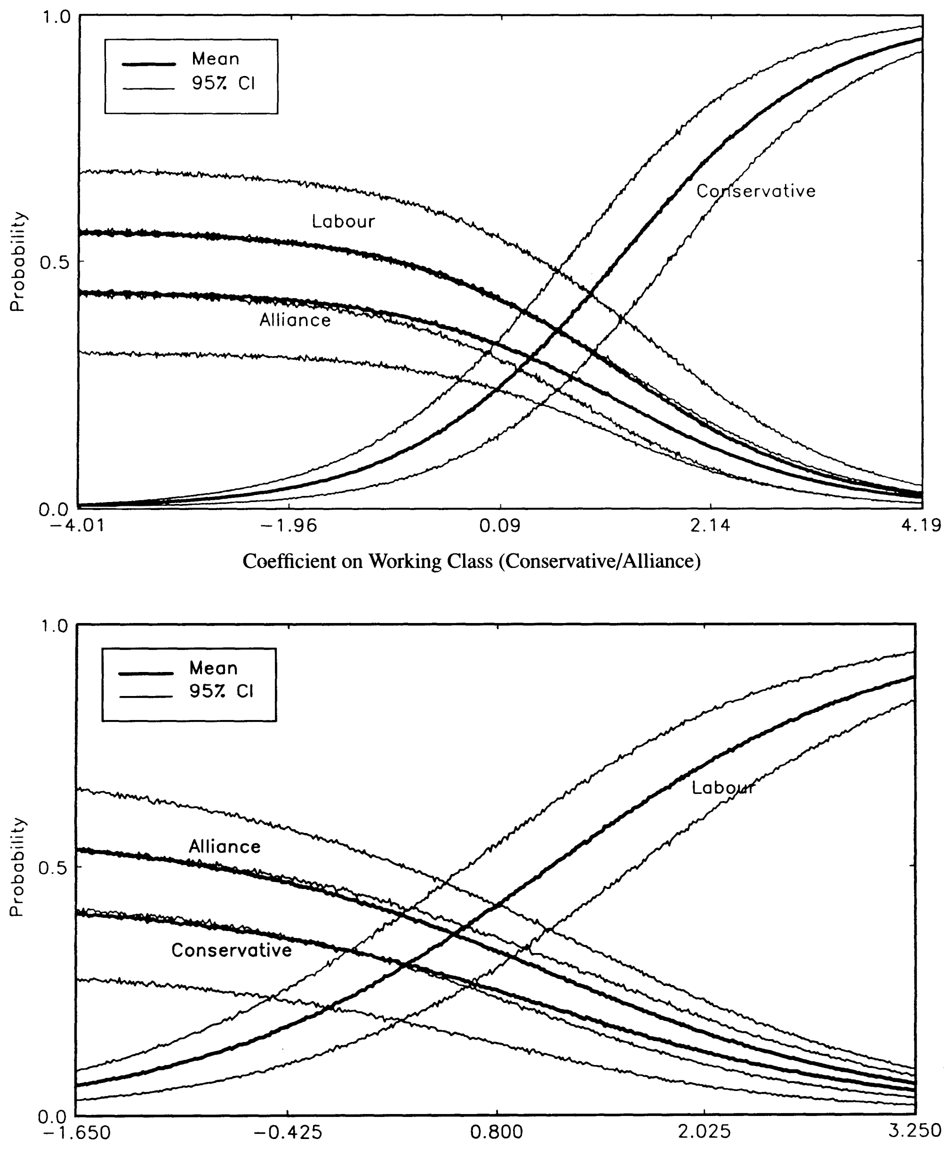

Fig. 1 Distribution of vote probabilities for a hypothetical voter.

vote against their traditional class ties. This heterogeneity in the behavior of the working class is not apparent in models that assume that there is no random taste variation.

Another way to demonstrate the effect of random taste variation in the impact of working-class membership on the vote is to examine the impact of this variable on vote probabilities over the entire range of each random coefficient (holding the other random coefficient at its mean). In Fig. 1 I graph the means and 95% confidence intervals of the vote probabilities for the hypothetical voter if he is a member of the working class. The means and standard deviations of the predicted probabilities were calculated as in Table 3 across the values of each random coefficient. The three heavy curves represent the mean probabilities of voting for each of the three parties as the coefficient value on the impact of working class changes, while the lighter curves represent a two-standard deviation confidence interval around the mean values. Note that the graphs are not smooth because each point was calculated individually using 1000 draws from the multivariate normal distribution of the estimated parameters.

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

The results of the MXL models presented here reveal that social class was still a powerful influence on vote choice for many individuals in the 1987 general election. Most scholars maintain that Britain was experiencing a "class dealignment" in the 1980s, especially within the working class. However, it is apparent that the political cleavages between the working class and the middle class still carried weight at the time of the 1987 General Election for many individuals.

4 Discussion

A wide variety of structures can be placed on the unobserved portions of utility to answer different questions about voting behavior in multicandidate and multiparty elections. While the MNP 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 MNP and MXL can be specified to explore many of these questions, although MXL is the more general model of the two.

MXL holds two advantages over MNP that are of consequence in the study of multicandidate and multiparty elections. 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 MXL 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 MXL mean that a wider variety of specifications is 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 be addressed only 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, MXL 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 MXL is more flexible in the study of substitution patterns across alternatives than MNP.

The advantages of MXL 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. Both MNP and MXL can be specified to explore random taste variation. However, MNP is limited in the number of random coefficients that can be estimated, and these random coefficients must be normally distributed. MXL can include any number of random coefficients, and these random coefficients can follow any distribution. Random coefficients that theory tells us are nonnormal (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 MXL is an unambiguously superior tool for exploring heterogeneity in the impact of variables on vote choice.

Thus far political scientists have been interested primarily 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 a heterogeneous impact on vote choice is useful information when trying to explain voting 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 are obtained, models that are able to uncover some of the structure of the unobserved portions of utility offer the best chance to increase our knowledge of voter behavior in this area. Both MNP and MXL are useful in this line of research. However, MXL is more general in its specification of the unobserved portions of utility and thus allows us to explore voter behavior in multicandidate and multiparty elections more broadly.

Appendix: Estimation of Mixed Logit Models

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

$$\mathcal{L}(\beta, \Omega) = \sum_{I} \sum_{J} y_{ij} \log \left[\int_{\eta_i} \left\{ \frac{e^{X_{ij}\beta_j + Z_{ij}\eta_i}}{\sum_{k \in J} e^{X_{ik}\beta_k + Z_{ik}\eta_i}} \right\} g(\eta_i \mid \Omega) \partial \eta_i \right] \tag{6}$$

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. (6) involves the estimation of a Q-dimensional integral.[9] This

[9]Note that this is why a "mixed probit" model (with $g(\eta \mid \Omega)$ 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.

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 $g(\eta \mid \Omega)$ for each individual. The values of this draw can then be used to calculate $\hat{P}(j \mid \eta)$, the conditional choice probability given in Eq. (4). This process is repeated R times, and the integration over $g(\eta \mid \Omega)$ is approximated by averaging over the R draws. Let $\hat{P}_r(j \mid \eta_r)$ be the realization of the choice probability for individual i for alternative j for the rth draw of η. The choice probabilities given the parameter vectors β and Ω are approximated by averaging over the values of $\hat{P}_r(j \mid \eta)$:

$$\hat{P}(j \mid \beta, \Omega) = \frac{1}{R}\sum_{r=1}^{R} \hat{P}_r(j \mid \eta_r) \tag{7}$$

$\hat{P}(j \mid \beta, \Omega)$ 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 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 $\hat{P}(j \mid \beta, \Omega)$ 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. To estimate Ω the distributions in η are reexpressed in terms of standardized, independent distributions. That is, $g(\eta \mid \Omega)$ is reexpressed 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:

$$\mathcal{SL}(\beta, \Omega) = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}\log[\hat{P}_i(j \mid \beta, \Omega)] \tag{8}$$

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 (1999) and Train (1999) dramatically reduces estimation time for MXL models. This alternative simulation technique uses nonrandom 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. Both Bhat (1999) and Train (1999) use Halton sequences 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.[10] The dividing points on this unit interval become the first $h - 1$ elements in the Halton sequence. Each of the h subportions of the unit interval is divided as the entire unit interval was, and these elements are added 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.

Halton sequences can be used in place of random draws to estimate MXL 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 data set). The first 10 elements in 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 data set is assigned the first R elements in each Halton sequence, the second individual is assigned the next R elements, and so on. For each element in 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 (1999) and Train (1999) found that in estimating MXLs, 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 MXL models. All MXL models in this paper were estimated using GAUSS. The GAUSS code used to estimate the models in this paper is available at the Political Analysis website or at http://www.polsci.ucsb.edu/faculty/glasgow. This code is a modified version of the GAUSS code made available by Kenneth Train at his website, http://elsa.berkeley.edu/users/train. MXLs 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 MXL models see Appendix A at the Political Analysis website.

References

Algers, S., P. Bergström, M. Dalhberg, and J. L. Dillén (1998). "Mixed Logit Estimation of the Value of Travel Time," Unpublished manuscript.

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):714–744.

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

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

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):1349–1363.

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):131–149.

Bhat, C. R. (1998a). "Accommodating Variations in Responsiveness to Level-of-Service Measures in Travel Mode Choice Modeling." *Transportation Research A* 32(7):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):455–466.

Bhat, C. R. (1999). "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model." *Transportation Research* (in press).

Brownstone, D., and K. Train (1999). "Forecasting New Product Penetration with Flexible Substitution Patterns." *Journal of Econometrics* 89(1):109–129.

Crewe, I. (1992). "The 1987 General Election." In *Issues and Controversies in British Electoral Behavior*, eds. D. Denver and G. Hands. London: Harvester Wheatsheaf, pp. 343–354.

Crewe, I. (1983). "The Electorate: Partisan Dealignment Ten Years On." *West European Politics* 6(4):183–215.

Garrett, G. (1992). "The Political Consequences of Thatcherism." *Political Behavior* 14(4):361–382.

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. New York: North Holland, pp. 2383–2441.

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):403–426.

Heath, A. F., R. M. Jowell, and J. K. Curtice (1985). *How Britain Votes*. New York: Pergamon Press.

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

Horowitz, J. L. (1991). "Reconsidering the Multinomial Probit Model." *Transportation Research B* 25(6):433–438.

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):317–328.

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):233–255.

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* 8:518–552.

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

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

McFadden, D., and K. Train (2000). "Mixed MNL Models for Discrete Response." *Journal of Applied Econometrics* (in press).

Pulzer, P. G. (1967). *Political Representation and Elections in Britain*. London: Allen and Unwin.

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):1231–1247.

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):647–657.

Rivers, D. (1988). "Heterogeneity in Models of Electoral Choice." *American Journal of Political Science* 32(3):737–757.

Sarlvik, B., and I. Crewe (1983). *Decade of Dealignment: The Conservative Victory of 1979 and Electoral Trends in the 1970s*. Cambridge: Cambridge University Press.

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):257–293.

Train, K. (1998). "Recreation Demand Models with Taste Differences over People." *Land Economics* 74(2):230–239.

Train, K. (1999). "Halton Sequences for Mixed Logit," Unpublished manuscript.