Simulated Maximum Likelihood via the GHK Simulator:
An Application to the 1988 Democratic Super Tuesday Primary

Eric Lawrence

The Brookings Institution

I would like to thank Michael Keane for his general advice on estimation and the MNP model, David Runkle for providing the Gauss code to estimate the model, Jim Stimson for the use of his computer, and Mike Alvarez, Larry Bartels, Katherine Harris, Brad Jones, and Barbara Norrander for suggestions and comments. An earlier version of this paper was presented as a poster at the

1996 Political Methodology summer meeting in Ann Arbor, MI. Please direct all comments to lawrence@polisci.umn.edu.

Few political science applications of discrete choice models with more than two alternatives have estimated multinomial probit (MNP) models. This has been the case despite the fact that the MNP model allows for correlated error terms across alternatives and unequal error term variances, properties that may be implied by theories of political choice processes.[1] The tendency to avoid MNPs has been reinforced by the difficulty of estimating these models, with few user friendly software packages available for estimating the models. In a recent survey of vote choice models for comparative politics scholars, for example, Whitten and Palmer (1996, 256) argue that estimating MNP models with three or more alternatives imposes such great computational burdens that they are "impractical."[2] Theoretical advances on the computational side combined with the incorporation of MNP models into standard software packages[3] have begun to undercut the long standing "nice, but too hard" criticism of MNP estimation.

In this paper, I explain a method for estimating MNPs that weakens this argument against using MNPs when they are theoretically appropriate. The computational demands of estimating MNPs, once insurmountable, remain high but are reasonable for researchers with fast desktop machines or access to a workstation. The method explained here, simulated maximum likelihood combined with an accurate method for simulating probabilities (the GHK simulator), is not the only possible way to estimate MNPs. It may be the most easily implemented classical method, however, given that existing maximum likelihood routines can be modified to accommodate the GHK simulator.

Multinomial probit models have long been considered infeasible for empirical researchers because calculating the choice probabilities requires the evaluation of multiple integrals with no simple closed form solutions. About fifteen years ago, simulated maximum likelihood was first proposed as a way around the direct evaluation of the multidimensional integrals required by the MNP model (Lerman and Manski 1981). Lerman and Manski showed that instead of directly calculating the multivariate normal choice probabilities, the probabilities could be simulated. While this was a conceptual advance, the method of simulating probabilities used by Lerman and

---

1 Discrete choice models are appropriate when the dependent variable is observed in the form of a dichotomy or polychotomy. In political science, the most frequent application of discrete choice models has been vote choice. For an early but very thorough survey of alternative discrete choice models, see McFadden (1981). For a discussion of situtations where MNP models have distinct advanatages over logit models, see Alvarez and Nagler (1996).

2 A small number of political science papers have overcome the computational barriers and successfully applied the MNP model. See Alvarez and Nagler (1995), who modeled the 1992 Bush-Clinton-Perot Presidential race by directly evaluating bivariate normal integrals and Quinn, Martin, and Whitford (1996), who modeled vote choice in the Netherlands by using Gibbs sampling.

3 As of this writing, both LIMDEP 7.0 with the NLOGIT 1.1 module and Gauss-X have added MNP options to their packages, both using the GHK probability simulator described in this paper. I have no experience with either, however, so can offer no evaluation of the programs.

Manski, a frequency simulator, was not accurate enough to satisfy empirical researchers.[4] Considerable effort was expended by econometricians to develop more accurate simulation methods, and one such method, the GHK simulator, will be demonstrated in this paper. By using SML with the GHK simulator, it becomes possible to estimate MNPs with large choice sets. Simulated maximum likelihood provides a framework for statistical inference, and the GHK simulator makes it possible to approximate the integrals required by MNP models.

To establish a firmer motivation for the utility of simulated maximum likelihood based on the GHK simulator (SML-GHK), a brief sketch of recent developments in discrete choice estimation precedes the technical details of SML. Following the background section, the MNP model and the GHK simulator are discussed. First, the necessary assumptions for a three alternative MNP are outlined. It is then shown how the GHK simulator can be applied to the three alternative MNP through the use of simulated maximum likelihood. Next, the logic of the GHK simulator for a slightly more complicated case, an MNP with four alternatives, is demonstrated. After the mathematical examples, the results of a substantive application are presented--an MNP for the 1988 Democratic Super Tuesday primary, with a choice set containing three candidates. These results are discussed with an emphasis on the estimates of the covariance matrix, information that can not be obtained in logit models. Finally, the paper concludes with a general discussion of simulated maximum likelihood and other potential applications of the methods presented in the paper.

Recent Developments in Discrete Choice Estimation

To understand why simulated maximum likelihood and the GHK simulator are important developments for estimating discrete choice models, a little background is required. When estimating discrete choice models with more than two alternatives, researchers have historically chosen to work with multinomial logit or conditional logit models[5] because they are much easier to program and demand fewer computing resources. These models are simple to estimate with

---

4 A discussion of the problems with the frequency simulator method used by Lerman and Manski can be found in Börsch-Supan and Hajivassiliou (1993, 352). Essentially, the problem with the frequency simulator is that the probabilities are not a smooth function of the model parameters and that gradient optimization methods cannot be used. It should be noted that Lerman and Manski acknowledged the shortcomings of the frequency simulator method.

5 To simplify notation, I will denote both multinomial logit models and conditional logit models as simply logit models. The original distinction between the two models is as follows. In MNL models, only characteristics of individual choosers are included as independent variables. In "pure" CL models, only characteristics of the alternatives are included as independent variables, with the name coming from the conditional utility function implying this set of independent variables. Of course, both individual specific variables and alternative specific variables can be included in the same model. I will denote these "general" CL models simply as "logits" throughout the paper, with the assumption that I am referring to a choice specification with more than two alternatives.

standard likelihood maximization routines, but they impose rigid covariance structures that may be undesirable on theoretical grounds. In particular, these models suffer from the assumption of the "independence of irrelevant alternatives" (IIA).[6] Under the IIA assumption, the error terms associated with the utility from the alternatives are uncorrelated across alternatives. In the vote choice context, this amounts to the substantive assumption that the error terms for each candidate are pairwise independent. This is problematic if several candidates are perceived as ideologically "close", for example. Put another way, if Bill Clinton had a twin brother (Will) with an identical political background who decided to enter the presidential race, the entry of Will Clinton into the race could not affect the relative likelihood of voting for Bill Clinton over Bob Dole.[7] The logit models also assume homoscedasticity of the errors, an assumption that rarely has a substantive rationale.

Multinomial probit (MNP), in contrast to logit, allows for a flexible covariance structure, permitting correlations across alternatives and unequal variances across alternatives.[8] The drawback of estimating an MNP model, however, is that it requires the evaluation of multiple integrals for which no closed form solutions exist. Specifically in the MNP case, for J alternatives, multiple integrals of dimension J-1 must be calculated. This is hard and computationally intensive. When using traditional computational methods, such as quadrature[9], adding an order of integration increases the computational costs by roughly an order of magnitude. This phenomenon has been called the “curse of dimensionality.” Models that demand the evaluation of such integrals are therefore not typically implemented empirically. Instead, researchers tend to either estimate logits

---

6 Hausman and Wise (1978, 404) prefer "independence of relevant alternatives" or "independence among alternatives" as more appropriate characterizations of the assumption because these phrases connote the problem more precisely. This seems to be a reasonable claim.

7 This, of course, is a political variant of the “red bus/blue bus” analogy of McFadden (1974). Some researchers object to the concern with IIA and the goal of estimating covariance parameters, making a two-pronged argument. First, "good" models will reduce error correlations by including independent variables that explain choice behavior, thereby reducing the unobserved heterogeneity captured in the covariance matrix. Second, covariance parameters should be considered nuisance parameters, and to the extent that they drive choice behavior, we have less understanding of the substance of the choice process. These critiques deserve to be taken seriously, but they do not eliminate the need for models that get around the IIA assumption. Consistent with the first objection, one approach that has been taken is to keep adding independent variables to logit specifications until IIA tests (Hausman-McFadden, Small-Hsiao) can be passed. This solution is quite ad hoc, however, and it is not clear that the resulting estimates provide a solid understanding of choice behavior. Regarding the second objection, it is plausible that some variables that influence choice behavior will always be unobservable. For example, election choices may be influenced by the "campaigning quality" of candidates and that variable may not be measurable. If so, MNP may be an appropriate model.

8 MNP is not the only specification that avoids IIA. Heterogeneous logit (sometimes called "random-parameters logit") provides another way around the IIA assumption (Erdem and Keane 1996). Like MNP, this method is computationally intensive. I will only discuss MNP here.

9 See Greene (1997, 190-192) for a brief discussion of approximating integrals by quadrature.

as second best solutions or simply confine their research questions to examining the effects of independent variables on choice behavior, for which logits can provide useful estimates if the model is well specified and the assumption of independent errors is not severely violated. Because of the IIA assumption, the logit models do not allow the researcher to answer substantive questions about adding or removing an alternative from the choice set.

About five years ago, the feasibility of estimating high dimensional ($J > 3$) MNPs came within the reach of applied researchers due to theoretical breakthroughs and increases in computing power. The theoretical breakthroughs yielded both Bayesian and Classical (frequentist) statistical methods for the simulation of high dimensional choice probabilities. Bayesian estimation methods, such as Markov Chain Monte Carlo methods and Gibbs sampling, provide one general approach. Applications of Gibbs sampling in applied statistics have proliferated rapidly in the last five years, leading some observers to dub the trend "Gibbsmania." This technique has proven feasible for econometric applications (Geweke, Keane, and Runkle 1994a; McCulloch and Rossi 1994) and is gaining a toehold in political science (Jackman 1995; Helland and Whitford 1995; Quinn, Martin, and Whitford 1996).

The second general approach, which will be discussed in this paper, is to use probability simulation (also called Monte Carlo integration) methods. By simulating the choice probabilities, one can significantly reduce computational costs because in simulation methods, computing costs increase only linearly with the number of alternatives. Two classical statistical methods have been developed that accommodate simulated choice probabilities, simulated maximum likelihood (SML[10], Manski and Lerman 1981) and the method of simulated moments (MSM, due to McFadden 1989 and Pakes and Pollard 1989). The simulated probabilities required for both MSM and SML can be estimated by the GHK simulator so that one can estimate either SML-GHK or MSM-GHK models. Monte Carlo analysis (Geweke, Keane, and Runkle 1994a) suggests that MSM is harder to implement than SML because of the difficulty of obtaining good starting values for MSM, so SML-GHK is the method that is explained in detail in this paper.[11] The GHK simulator method was chosen here because a recent exhaustive survey of probability simulators (Hajivassiliou, McFadden, and Ruud 1996) reached the conclusion that the GHK simulator is the most reliable and accurate way to simulate multivariate normal probabilities for classical estimation.

---

10 A note for the jargon watchers: SML has also been called MSL on ocassion, but the two are the same.

11 Also, the Monte Carlo results presented by Lee (1995) suggest that the inherent bias of SML is typically quite small.

The Multinomial Probit Model

This paper uses notation that is often used for MNPs, although there is no accepted universal standard for subscripts and parameters. It also uses the language of random utility models (RUMs) in describing the model, but this is not the only possible justification for the MNP functional form. For a first example, consider the case where there are J=3 alternatives and i=1...N individuals. Each individual i derives utility from each alternative j, U_{ij}, that is a sum of systematic components $X_i\beta_j$ and $Z_{ij}\gamma_j$, and a stochastic term ε_{ij}. For the systematic components, the X_i terms vary across individuals, while the Z_{ij} terms vary across alternatives or across alternatives *and* individuals. The assumption that the ε_{ij}s are distributed multivariate normal makes this a MNP model.[12] Notice that in this model, the variances need not be equal and that the covariances need not be zero.

$$U_{i1} = X_i\beta_1 + Z_{i1}\gamma_1 + \varepsilon_{i1}$$

$$U_{i2} = X_i\beta_2 + Z_{i2}\gamma_2 + \varepsilon_{i2}$$

$$U_{i3} = X_i\beta_3 + Z_{i3}\gamma_3 + \varepsilon_{i3}$$

$$\Sigma = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{21} & \sigma_{32} & \sigma_{33} \end{pmatrix}\right)$$

In order to formally identify the model, however, several restrictions must be imposed (for a full discussion of identification issues in MNP models, see Bunch 1991). There are several possible ways to achieve identification, so the following restrictions are not the only possible restrictions. First, the utility derived from one of the alternatives is set to zero. Here, the third alternative is set to zero. This is necessary because all the theory says that matters for choosing between two alternatives is the difference in utility between the alternatives. That is, adding an arbitrary constant to each alternative does not affect the relative ranking of the alternatives. Once this restriction is imposed, alternative 1 will be chosen if the utility from alternative one is positive (greater than alternative 3) and is greater than that from alternative 2. Second, the scales of the β and σ parameters are not separately identified from observed choice behavior. This scale issue does not arise in logit models because the researcher implicitly restricts the covariance structure through the assumption of Type I extreme value distribution on the errors.

---

12 In a bivariate (J=2) probit model, the error term is distributed as a univariate normal. The corresponding distributional assumption for logit models was derived by McFadden (1974), who showed that logit models assume the errors are independent across alternatives and follow a type I extreme value distribution.

In this MNP (J=3) example, in order to set the scale on the σ parameters, σ_{11} is set to 1. The choice of 1 is arbitrary,[13] but it pins down the scale for all the estimated parameters. Note that an arbitrary β parameter could also be held constant to achieve identification, but some researchers may feel less comfortable with their prior knowledge regarding the sign of a structural parameter. The only constraint is that a positive number should be chosen if a variance parameter is fixed, something that should be "known" with a high level of confidence by the researcher. The restrictions made here produce the following formally identified specification, with the third alternative as the baseline category.

$$U_{i1}^{*} = X_i\beta_1^{*} + Z_{i1}\gamma_1 - Z_{i3}\gamma_3 + \varepsilon_{i1}^{*}$$

$$U_{i2}^{*} = X_i\beta_2^{*} + Z_{i2}\gamma_2 - Z_{i3}\gamma_3 + \varepsilon_{i2}^{*}$$

$$U_{i3}^{*} = 0$$

$$\Sigma^{*} = \begin{pmatrix} \varepsilon_{i1}^{*} \\ \varepsilon_{i2}^{*} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{12}^{*} \\ \sigma_{21}^{*} & \sigma_{22}^{*} \end{pmatrix}\right)$$

This "transformed" or "differenced" model was formed by subtracting the utility for the third alternative from that of the first and second alternatives. The terms in the differenced model are starred to denote that the utilities, parameters, and error terms are not the same as in the original model. Note that in order to combine the alternative specific variables, it would be necessary to restrict the alternative specific parameters to be equal. Now let

$$d_{ij} = \begin{cases} 1 \text{ if } u_{ij} \geq u_{ik} \ \forall \ k \\ 0 \text{ otherwise} \end{cases}$$

which indicates that individual i will choose alternative j if the utility derived from alternative j is greater than that for all other alternatives. The choice probabilities for the J alternatives can be illustrated mathematically in the following way. For notational simplicity, we combine the alternative specific and individual specific variables and parameters into a single $X_{ij}^{*}\beta^{*}$ term. Then the probability that individual i chooses alternative 1, for example, can be written as:

---

13 Except for the fact that it is a positive number, given that variances must be positive.

$$P(d_{i1}=1) = \text{Prob}(u_{i1}^* > 0, u_{i1}^* > u_{i2}^*)$$

$$= \text{Prob}(x_{i1}\beta^* + \varepsilon_{i1}^* > 0, x_{i1}\beta^* + \varepsilon_{i1}^* > x_{i2}\beta^* + \varepsilon_{i2}^*)$$

$$= \text{Prob}(\varepsilon_{i1}^* > -x_{i1}\beta^*, \varepsilon_{i2}^* < x_{i1}\beta^* - x_{i2}\beta^* + \varepsilon_{i1}^*)$$

$$= \int_{-x_{i1}\beta^*}^{\infty} \int_{-\infty}^{x_{i1}\beta^* - x_{i2}\beta^* + \varepsilon_{i1}^*} f(\varepsilon_{i1}^*, \varepsilon_{i2}^*) \partial\varepsilon_{i2}^* \partial\varepsilon_{i1}^* \qquad (1)$$

The choice region for alternative 1 corresponding to equation (1) is shown in figure 1. The shaded region is the area of the space where alternative 1 will be chosen over alternatives 2 and 3, and the size of the region will determine the probability that individual i will choose alternative j. The log likelihood function is then derived by summing the logged probabilities across the alternatives of the differenced model for each individual.

$$\log L(\hat{\beta}, \hat{\Sigma}) = \sum_{i=1}^{N} \sum_{j=1}^{J-1} d_{ij} \ln P(d_{ij} = 1 | x_i \hat{\beta}) \qquad (2)$$

Estimating the MNP model, entails maximizing the log likelihood function specified in (2). The advantage of MNP over logit is that it allows for flexible error correlation structures and does not suffer from the IIA assumption. A difficulty created by MNP, however, is that the resulting choice probabilities, as in (1), are multiple integrals. What can we do, then, to estimate the more substantively rich MNP models? To solve this problem, several researchers developed accurate methods of simulating choice probabilities, the subject of the next section.

Simulation of Choice Probabilities

The method for simulating multivariate normal integrals in discrete choice models described here was developed independently by Geweke (1991), Keane (1990, 1994), and Hajivassiliou and McFadden (1990, 1996). For expository purposes, the three alternative example discussed above provides the initial example. By matrix algebra, it is possible to decompose the multivariate normal errors in the following way.

$$\begin{pmatrix} \varepsilon_{i1}^* \\ \varepsilon_{i2}^* \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \sigma_{12}^* & \sqrt{\sigma_{22}^* - \sigma_{12}^{*\,2}} \end{pmatrix} \begin{pmatrix} \eta_{i1} \\ \eta_{i2} \end{pmatrix}$$

The matrix premultiplying the η vector, which is distributed iid standard normal, is the unique lower triangular choleski factorization for Σ^*. Call this matrix A, defined for notational purposes as $A = \begin{pmatrix} 1 & 0 \\ a_{21} & a_{22} \end{pmatrix}$ Now assume that the η_{ij} terms are

$$\begin{pmatrix} \eta_{i1} \\ \eta_{i2} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]$$

Then,

$$\varepsilon^*_{i1} = \eta_{i1}$$

$$\varepsilon^*_{i2} = \eta_{i1}\sigma^*_{21} + \eta_{i2}\sqrt{\sigma^*_{22} - \sigma^{*\,2}_{21}}$$

$$\mathrm{var}(\varepsilon^*_{i1}) = 1$$

$$\mathrm{cov}(\varepsilon^*_{i1}, \varepsilon^*_{i2}) = E(\sigma^*_{21}\eta_{i1}{}^2) = \sigma^*_{21}$$

$$\mathrm{var}(\varepsilon^*_{i2}) = \sigma^{*\,2}_{21} + (\sigma^*_{22} - \sigma^{*\,2}_{21}) = \sigma^*_{22}$$

Now, using these transformed errors and the lower triangular choleski matrix A, we can rewrite the choice model as:

$$U^*_{i1} = X_{i1}\beta^* + \eta_{i1}$$

$$U^*_{i2} = X_{i2}\beta^* + a_{21}\eta_{i1} + a_{22}\eta_{i2}$$

$$U^*_{i3} = 0$$

Then,

$$\mathrm{Prob}(d_{i1} = 1) = p\left(\eta_{i1} > -X_{i1}\beta^*, \eta_{i2} < \frac{X_{i1}\beta^* - X_{i2}\beta^* + \eta_{i1}(1 - a_{21})}{a_{22}}\right) \qquad (3)$$

Given these transformations, it is possible to condition on the η 's instead of the ε's in order to generate a choice. The logic that Geweke (1991), Keane (1990, 1994), and Hajivassiliou and McFadden (1990, 1996) used to solve the problem of estimating the MNP was to think of the η's as *sequential* events that could be simulated recursively,[14] in effect breaking a complicated problem into simple components.

---

[14] Because these authors made the similar insight independently and at roughly the same time, this method of estimation has been designated the "GHK recursive simulator."

Using this logic in the three alternative case, one first asks if η_{i1} satisfies the first condition in (3) above, then one asks if η_{i2} satisfies the second condition in (3). If so, then the alternative (alternative 1 in this case) will be chosen. The probability that alternative 1 will be chosen is the joint probability of both events occurring. By breaking up the problem in this way, the researcher no longer needs to evaluate multivariate integrals. Instead, the GHK simulator only requires draws from truncated normal distributions and the evaluation of univariate integrals. This is illustrated below.

Suppose η_{i1} was known and equal to η^*_{i1} such that $\eta^*_{i1} > -X_{i1}\beta^*$. Then we could rewrite the probability of alternative 1 being chosen as:

$$\text{Prob}(d_{i1}=1|\eta^*_{i1}) = p\left(\eta_{i2} < \frac{X_{i1}\beta^* - X_{i2}\beta^* + \eta^*_{i1}(1-a_{21})}{a_{22}}\right) \tag{4}$$

If we knew η_{i1}, this would be an easy problem. This is clearer when the probability of choosing alternative 1 is written as the product of the unconditional probability of η_{i1} satisfying the first condition in (3) above times the conditional probability that η_{i2} falls in the right choice region, given that η_{i1} satisfies the first condition in (3). That is:

$$\text{Prob}(d_{i1}=1) = p(\eta_{i1} > -X_{i1}\beta^*)p(d_{i1}=1|\eta_{i1} > -X_{i1}\beta^*)$$

The latter probability can be expressed as:

$$p(d_{i1}=1|\eta_{i1} > -X_{i1}\beta^*) = \int_{-X_{i1}\beta}^{\infty} p(d_{i1=1}|\eta^*_{i1})\frac{f(\eta^*_{i1})}{p(\eta_{i1} > -X_{i1}\beta^*)}\partial\eta^*_{i1}$$

Here, we are still stuck with a bivariate integral because the second term in the product requires the evaluation of an integral. This integral, however, is in the form of the expectation of a simple object (E(x) = ∫ xf(x)dx) that can be written as:

$$\underset{\eta^*_{i1}\text{ s.t. }\eta^*_{i1} > -X_{i1}\beta^*}{E}[\text{prob}(d_{i1}=1|\eta^*_{i1})]$$

Now, if we go out and draw $\tilde{\eta}_{i1}$ from the density $\frac{f(\eta^*_{i1})}{p(\eta_{i1} > -X_{i1}\beta)}$ then $E[\text{prob}(d_{i1}=1|\tilde{\eta}_{i1})]$ is an unbiased estimator of the double integral (1) above.[15] This leaves us with a two step process for simulating the choice probability $p(d_{i1}=1)$.

---

15 Note that this density is a truncated normal density when $f(\cdot) \equiv \phi(\cdot)$, as it is in the MNP case.

Step 1: make D draws of $\tilde{\eta}_{i1}^{d}$ from the truncated normal density.

Step 2: calculate $\text{prob}(d_{i1} = 1 | \tilde{\eta}_{i1})$

$$\hat{p}_{GHK}(d_{i1} = 1 | \eta_{i1} > -X_{i1}\beta^{*}) = \frac{1}{D}\sum_{d=1}^{D} p(d_{i1} = 1 | \tilde{\eta}_{i1}^{d})$$

$$\hat{p}_{GHK}(d_{i1} = 1) = p(\eta_{i1} > -X_{i1}\beta^{*})\frac{1}{D}\sum_{d=1}^{D} p(d_{i1} = 1 | \tilde{\eta}_{i1}^{d})$$

In the GHK estimation procedure, step 1 is repeated D times in order to obtain $(\eta_{i1}^{d}),\ d = 1,\ldots,D$. In step 2, the results of D draws are averaged in order to calculate the value of the log likelihood function. These two steps are conducted for each alternative, and once simulated probabilities are calculated for all alternatives, parameter estimation proceeds as in standard ML. Because estimation depends on simulating the choice probabilities, however, the general procedure is called *simulated maximum likelihood* . At each step 2, the simulated LLF is calculated as:

$$\log L(\hat{\beta}, \hat{\Sigma}) = \sum_{i=1}^{N}\sum_{j=1}^{J-1} d_{ij} \ln \hat{p}_{GHK}(d_{ij} = 1 | x_i\hat{\beta})$$

The innovation developed by Geweke, Hajivassiliou, and Keane was to figure out how to simulate the probabilities as described above. Once the simulated probabilities are obtained, maximum likelihood techniques are employed. Steps in the iterative search to maximize the likelihood function are made using standard ML procedures; using the BHHH algorithm is standard practice. Standard errors can also be calculated through the BHHH algorithm. The BHHH algorithm uses the result that the hessian can be approximated with the outer product of the gradients:

$$V\hat{A}R(\hat{\beta}) = \left(\frac{\partial L(\hat{\beta})}{\partial \hat{\beta}} . \frac{\partial L(\hat{\beta})}{\partial \hat{\beta}'}\right)^{-1}$$

The GHK simulator produces unbiased estimates $\hat{p}$ of p (Börsch-Supan and Hajivassiliou 1993, 359) but ln($\hat{p}$) is biased down for an estimate of ln(p). This results from the nonlinearity of logarithms. Suppose one has an unbiased estimator $\hat{p}$ of p, such that $\hat{p}$ = p + ε with probability .5 and at $\hat{p}$ = p - ε with probability .5. It is easily seen that the E[ln($\hat{p}$)] is less than ln[E($\hat{p}$)] because of the concavity of the logarithmic function.. Therefore, simulated maximum likelihood, which uses the expectation of the logged simulated probabilities, is biased down. In addition to the bias result, the following asymptotic properties of SML have been derived by Lung-Fei Lee (1995):

$\hat{\beta}_{SML}$ is biased for finite D as $N \to \infty$.

$\hat{\beta}_{SML}$ is consistent and asymptotically efficient if $D/\sqrt{N} \to \infty$ as $N \to \infty$.

D must grow faster than $\sqrt{N}$ as $N \to \infty$.

The ordinary ML case is that maximum likelihood estimators are consistent as N approaches infinity. The SML requirement for consistency is more stringent in the sense that increasing N alone will not achieve consistency. Instead, the ratio of the number of draws to the square root of the sample size must grow arbitrarily large. These analytical results should not scare off attempts to use SML, however, as Monte Carlo evidence suggests that SML bias is not too severe and decreases with the number of draws. The obvious practical question is how many draws to make in order to make the bias in SML trivial. Fortunately, the number of draws needed to minimize bias is well within the range of modern computing power. In their comparison of alternative computational strategies for MNP, Geweke, Keane, and Runkle use D=30 (1994a, 630) and suggest that it is a reasonable number of draws to make. In independent research, both Keane (1994) and Börsch-Supan and Hajivassiliou (1993) present Monte Carlo evidence suggesting that setting D at 20 yields negligible bias.[16] The SML-GHK method, therefore, provides a feasible and accurate technique for estimating MNP models.[17]

Before presenting the results of an application with real data, it will be illustrative to show how the four alternative case would be estimated. For notational simplicity, only the transformed model is presented here.

---

[16] Geweke, Keane, and Runkle (1994a) discuss another strategy for dealing with SML bias. The parameter estimates from SML can be used as starting values for method of simulated moments (MSM) estimation (McFadden 1989), which is unbiased. In Monte Carlo analyses, Geweke et al. (1994a) show that a disadvantage of MSM-GHK is that it is very slow if the starting values are not close to the true values. SML does not have this same problem regarding starting values, so if the small amount of bias inherent in SML is a serious concern, the researcher can first estimate an MNP via SML-GHK, then use those estimates as starting values for MSM-GHK. This two-step procedure requires a significant amount of work, however, so would probably not be undertaken if computing resources are a constraint.

[17] It should be noted that SML-GHK performs better in the cross-sectional MNP case than in the panel MNP case. In a paper comparing alternative estimation strategies for panel MNPs, Geweke et al. (1994b) find that SML-GHK is outperformed (on RMSE grounds) by MSM-GHK, which is in turn outperformed by Gibbs sampling.

$$U_{i1} = X_{i1}\beta + \varepsilon_{i1}$$

$$U_{i2} = X_{i2}\beta + \varepsilon_{i2}$$

$$U_{i3} = X_{i3}\beta + \varepsilon_{i3}$$

$$U_{i4} = 0$$

$$\varepsilon_i \sim N(0, \Sigma)$$

$\Sigma = AA'$ where A is the choleski decomposition matrix

$$\Sigma = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \end{pmatrix} = \begin{bmatrix} 1 & & \\ a_{12} & a_{22} & \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \begin{pmatrix} \eta_{i1} \\ \eta_{i2} \\ \eta_{i3} \end{pmatrix} \text{ where } \eta_{ij} \overset{iid}{\sim} N(0,1)$$

Identifying restrictions parallel to the 3 alternative case were made, with J=4 as the baseline alternative, and the a_{11}. element normalized to 1. The probabilities were calculated as previously, accommodating the additional alternative.

$$\begin{aligned} P(d_{i1} = 1) &= P(u_{i1} > 0, u_{i1} > u_{i2}, u_{i1} > u_{i3}) \\ &= P(x_{i1}\beta + \eta_{i1} > 0, x_{i1}\beta + \eta_{i1} > x_{i2}\beta + a_{21}\eta_{i1} + a_{22}\eta_{i2}, \\ &\qquad x_{i1}\beta + \eta_{i1} > x_{i3}\beta + a_{31}\eta_{i1} + a_{32}\eta_{i2} + a_{33}\eta_{i3}) \\ &= P(\eta_{i1} > -x_{i1}\beta, \eta_{i2} < 1/a_{22}[x_{i1}\beta - x_{i2}\beta + \eta_{i1}(1 - a_{21})], \\ &\qquad \eta_{i3} < 1/a_{33}[x_{i1}\beta - x_{i3}\beta + \eta_{i1}(1 - a_{31}) - a_{32}\eta_{i2}]) \end{aligned}$$

Notice that the a_{ii} terms are in the denominators of the probabilities. This implies that the diagonal elements of the choleski matrix can not take on zero values. Furthermore, if the optimization algorithm sends values of the diagonal element close to zero, this can create problems for the calculation of the simulated probabilities, an issue discussed in more detail below. With an additional step for the additional alternative, probabilities are simulated as in the J=3 case.

step (1) draw η_{i1}^{d} such that $\eta_{i1}^{d} > -x_{i1}\beta$

step (2) draw η_{i2}^{d} such that $\eta_{i2}^{d} < 1/a_{22}\left[x_{i1}\beta - x_{i2}\beta + \eta_{i1}^{d}(1-a_{21})\right]$

repeat (1) and (2) D times to get $(\eta_{i1}^{d}, \eta_{i2}^{d}), d = 1,\ldots,D$

step (3) form $\hat{p}_{GHK}(d_{i1} = 1) = p(\eta_{i1} > -x_{i1}\beta)\cdot\frac{1}{D}\sum_{d=1}^{D}\left\{\begin{matrix} p(\eta_{i2} < 1/a_{22}\left[x_{i1}\beta - x_{i2}\beta + \eta_{i1}^{d}(1-a_{21})\right]\cdot \\ p(\eta_{i3} < 1/a_{33}\left[x_{i1}\beta + x_{i3}\beta + \eta_{i1}^{d}(1-a_{31}) - a_{32}\eta_{i2}^{d}\right]\end{matrix}\right\}$

Now we can do SML as above by constructing the log likelihood function, calculating the simulated probabilities for each alternative, then summing across alternatives within individuals, then summing across individuals.

Vote Choice in the 1988 Democratic Super Tuesday Primary

In general elections in the United States, model choice generally poses few problems for political researchers. The structure of the two-party system and the electoral laws make it very difficult for third party or minor party candidates to gain significant vote shares, with occasional exceptions like Ross Perot in 1992. When modeling vote choice between two candidates, either binomial probits or binomial logits are typically used, with little to distinguish between the two functional forms.[18] For two candidate elections, then, the choice of functional form is typically an issue of secondary importance for the researcher.

In elections with more than two candidates, however, model choice becomes more salient, because different functional forms yield different information. When more than two candidates are in the set of alternatives, researchers must decide how to model the more complicated choice faced by voters. In the case of U.S. presidential primaries, scholars have used a range of model specifications of vote choice. In his important book on the dynamics of presidential primaries, Bartels (1988) concentrates on changes in evaluations and expectations within the campaign. When vote choice is explained at the individual level, Bartels examines choice between two candidates at a time, such as Hart vs. Mondale in 1984 (Bartels 348). In her comprehensive study of the Super Tuesday primary, Norrander (1992) analyses aggregate vote data and exit poll data, but does not specify an individual level vote choice model. In other work, Norrander (1993) explains vote shares of individual candidates by state and by year, with variables such as money

---

18 The general election models can become complicated by political circumstances such as 1992 or if the analyst desires to model turnout and vote choice simultaneously in a single model.

spent by the candidate, previous vote totals by the candidate, and state ideology predicting vote percentages.

Other work in the area explicitly develops individual level vote choice models, but tends to remain within the logit framework. In one such effort, vote choice in the 1988 primaries is estimated as an MNL using NES data, with vote a function of candidate evaluation and perceived viability (Abramson et al. 1992, tables 8 and 9). Aldrich and Alvarez (1994) used exit poll data to account for the effects of issue priorities in MNL models of the 1988 Super Tuesday primaries. Both of these works restrict their analyses of the Democratic primaries to the four leading candidates-- Dukakis, Jackson, Gore, and Gephardt. Finally, another group of researchers (Stone et al. 1995) explains primary vote choice as a mixture of an elimination by aspects (EBA) model[19] and an expected utility model. In this application, the model is simulated and the transition probabilities of the EBA model are not directly estimated. These models produce useful information regarding the relationship between the independent variables and vote outcomes. The assumptions embedded in these models, however, do not allow the researchers to address correlations across alternatives nor unequal error variances across alternatives.

By allowing for a flexible error correlation structure, the MNP model allows for substantive insights that are not possible in the ordinary logit framework. The MNP model also allows the researcher to determine whether the assumptions inherent in the logit model are reasonable in the particular choice problem that is being modeled. In classical statistical inference, all inferences depend on having a well specified model. Unfortunately, not all variables of interest are available to researchers in every application. When relevant exogenous variables are omitted from a discrete choice specification, the error terms will be correlated if the omitted variables are determinants of choice behavior. In research on vote choice, several types of variables such as "overall candidate quality" or "candidate campaign organizational effectiveness" are likely to be unavailable to the analyst. If these variables can not be measured but they influence vote choice across candidates, the impact of the variables will be manifested in the covariance structure in the MNP model. When an analyst knows that theoretically relevant variables have been omitted from the model specification, the covariance matrix in the MNP can provide evidence regarding the effects of these omitted variables.

Other information can be obtained from the MNP models as well. To assess whether the IIA assumption is plausible, the effects of removing or adding a candidate from or to the choice set can be simulated. For example, it becomes possible to estimate whether Perot took more votes away from Clinton or Bush in 1992 without relying on self-report ranking data (Alvarez and

---

[19] McFadden (1981, 219) classifies probabilistic choice models into three "families" of functional forms: Thurstonian forms, containing probit models; Lucean forms, containing logit models; and Tverskian forms, containing several models developed by Amos Tversky. The EBA specification falls into the latter family.

Nagler 1995). These questions regarding IIA can not be addressed in the logit model. Furthermore, one rarely has a substantive reason to impose the assumption of homoscedasticity of the errors in discrete choice models. Estimating an MNP permits the researcher to free up the variance terms and investigate whether the assumption of equal variances is a plausible one. For all these reasons, the MNP model gives the researcher more information and permits the testing of assumptions that must be assumed by logit models.

For the specific case of the American presidential primaries, MNPs have the potential to assess the implications of the dynamics of the winnowing process. Bartels (1988, 304) raises the intriguing possibility that the presidential primary process solves the well known problem of preference aggregation by allowing voter preferences to change over time. Bartels examines the concept of momentum by using, among other data sources, a "continuous monitoring" survey conducted by the NES, that interviewed random samples of voters throughout the primary process. One difficulty with measuring preference change, however, is that each primary voter only gets to vote once, so it is not possible to observe how individual voters might change their preferences when faced with different choice sets or when obtaining additional information about the candidates. Early primaries have larger number candidates, some of whom drop out (are winnowed) before the later primaries. Therefore, it is difficult to determine whether a voter in a primary late in the primary season has genuinely changed preferences or whether the voter has adapted preferences to account for the smaller choice set. This is purely speculative, but the MNP model may make it possible to explore counterfactual primary scenarios by utilizing the correlations between primary candidates. As Alvarez and Nagler (1996) suggest, for political scientists, the MNP model may have its greatest advantages when examining issues of removing or adding alternatives to the choice set.

In this paper, the Super Tuesday exit polls (CBS/New York Times 1988) used previously by Aldrich and Alvarez (1994) and Norrander (1992) provide the data[20] for estimating a three alternative MNP.[21] In MNPs, as in logits, the effects of both individual specific and alternative specific variables can be estimated. While logits do not demand the inclusion of both types of

---

[20] These data have samples of individual states and a regional sample of Super Tuesday states. The latter sample is used in my analysis. The regional sample contains respondents from Alabama, Arkansas, Florida, Georgia, Louisiana, Maryland, Mississippi, Missouri, North Carolina, Oklahoma, Tennessee, Texas, and Virginia. Only those respondents who voted for the six main candidates and who had non-missing data on all the independent variables were included, yielding a sample size of 8044.

[21] In the intial cut at estimation, I specified a six alternative MNP, with 47 estimable parameters. Difficulties in converging to stable paramter estimates for that model forced me to scale back the model and build it back up to the higher dimensional problem. At the time of this writing, the estimates for the four alternative model are unstable, so they are not reported here. With these types of models, at times it can be hard to diagnose whether problems stem from the lack of information in the data about particular parameters or from computational issues.

variables, however, the difficulty of sorting out the separate effects of individual specific variables and correlations across alternatives makes it imperative to include alternative specific variables in MNP models (Keane 1992).[22] The size of the regional Super Tuesday sample, over 9000 Democratic voters, makes it feasible to estimate an MNP with several independent variables. Although only voters for Gephardt, Gore, and Dukakis are included in the results presented here, the sample was chosen to allow for comparability once additional alternatives are added. Therefore, because of the strong support for Jesse Jackson among black voters, only non black Democratic voters are included.[23] The independent variables used in the analysis are addressed in turn.

The independent variables in this analysis generally follow the variable choice of other research on presidential primaries. Each individual specific independent variable that is added to the model, however, adds J-1 parameters to the model, and given the computational intensity of estimating an MNP of this size, not all theoretically relevant variables are included. An *ideology* variable is included; in the exit poll data, ideology is measured on a three point scale.[24] The ideology variable is of particular interest, given the large literature on the spatial theory of vote choice. Lacking a objective location of the candidates in ideological, the expectation is that more liberal voters will favor Dukakis over Gephardt and Gore.

The next three variables are all dummy variables. One of the intentions of the political elites who developed the Democratic Super Tuesday structure was to make it more likely that a moderate candidate won the Democratic nomination. By stacking early primaries in the South, it was believed that the more conservative southern voters would vote for centrist candidates. In particular, Senator Albert Gore pursued a southern strategy, choosing to play down Iowa and New Hampshire and concentrate efforts in the South. A *grew up in the South* dichotomous variable

---

22 Keane calls this condition "fragile" or "tenuous" identification, which can occur even when a model is formally identified. He also provides an intuitive explanation of the underlying logic of the problem. The remedy for this problem is to have "exclusion restrictions", which are restrictions that "certain exogenous variables in the model do not affect the utility level of certain alternatives" (193). Alternative specific variables meet this requirement, as do individual specific variables that do not affect the utility of all alternatives. It can be difficult to come up with examples of the latter, however, that make sense theoretically.

23 Initially, for the six alternative model, a dummy variable indicating whether or not the respondent was black was included in the model specification to account for the overwhelming support Jesse Jackson received from black voters. It was later determined, however, that there probably is not enough variation in that variable across candidates to estimate the parameters. In the sample, for example, only one of the 172 voters who voted for Paul Simon was black, creating estimation problems. Dropping the *black* variable was not appealing because it would entail omitting a variable that clearly influences whether or not a respondent voted for Jesse Jackson. Therefore, the sample was restricted to non black voters.

24 The question wording is as follows "On most political matters, do you consider yourself...?" (1 = liberal, 2 = moderate, 3 = conservative).

allows for the exploration of whether there was variation in the types of voters in the south that may not have been fully appreciated by the creators of the Super Tuesday system. Among the states in the regional Super Tuesday sample are the *home states* of Richard Gephardt and Al Gore, Missouri and Tennessee, respectively. Those states were generally conceded to the home state candidates by their competitors, so that competitive efforts were not made in either state. To account for this, two dummy variables indicating whether a voter was from Missouri or from Tennessee were included in the Gephardt and Gore utility indices. Therefore, only two home state parameters were estimated.

Finally, the model accounts for the level of spending by each candidate in each state. In the 1988 Democratic primaries, the candidate spending the most in each state tended to win that state (see Wilcox 1991 for a thorough analysis). To control for the level of spending by the candidate in each state, a *money spent per capita in the voter's state* was constructed for each candidate.[25] This results in a candidate specific variable that varies across candidates and across individuals in different states. The parameter estimates on the money variable will allow for inferences on the effects of money, but the usual problems of endogeneity apply.[26] Given these exogenous variables, the model is transformed in the following way.

---

25 These data came from the *FEC Reports on Financial Activity 1987-88: Final Report, Presidential Pre-Nomination Campaigns*, FEC August 1989. The data are from Table A9, "State-by-State Expenditures," pp.12-17. I thank Frank Sorauf for providing the data source, which can also be found in the appendix of Norrander (1992). It should be noted that these data are somewhat noisy, as a candidate campaigning in Missouri, for example, could stay across the state line with those expenses not counting against the Missouri cap. These tactics make the spending data imperfectly reflect the level of effort by the candidate in the state.

26 Candidates are likely to spend more money in those states where they think they have a good chance of doing well, so correlation between the independent variable and the error term could result. One way around this problem would be to include state intercepts (fixed effects) for each candidate and each state to absorb the effects of candidate expectations. In the model here, however, that would result in (J-1)*13 parameters (26) for the three candidates in the 14 states. Even with modern computing power, that would make an already complicated model very hard to estimate.

$U_{ij} = X_i\beta_j + Z_{ij}\gamma_j + \varepsilon_{ij},\ (j = 1,\ldots,3)$

where the β_j are (4 x 1) vectors of parameters for the 4 individual specific variables,

and the γ_j are scalar parameters for the candidate expenditure variables.

After transforming the data by making Jackson the baseline candidate, the transformed model is:

$U^*_{ij} = X_i\beta^*_j + Z_{ij}\gamma^*_j - Z_{i3}\gamma_3 + \varepsilon^*_{ij},\ (j = 1,2)$

$U^*_{i3} = 0,$

$\varepsilon^*_{ij} = \varepsilon_{ij} - \varepsilon_3$

For each j, $(j = 1,2),\ \beta^*_j = \beta_j - \beta_3$

The identified covariance matrix in the transformed model is:

$$\Sigma^* = \begin{bmatrix} 1 & \\ \sigma^*_{21} & \sigma^*_{22} \end{bmatrix}$$

The transformed model has 11 possible parameters to estimate, with the 4 (2 variables by 2 candidates) parameters for the individual specific variables in each candidate's utility index, the 2 home state parameters for Gore and Gephardt, the 3 money parameters, and the 2 covariance elements. In general, [J(J-1)/2] - 1 covariance parameters can be estimated in MNP models (Bunch 1991), although it is possible to estimate one additional covariance parameter if the utility scale is set by fixing one of the non-covariance parameters at a non-zero value. In the model estimated here, the σ^*_{11} term is set to one. The results from the model estimation are presented in table 1 below. The number of draws for the probability simulator is 15, and the a_{11} parameter is fixed at 1 for identification purposes.

Table 1: Multinomial Probit Estimates, Three Candidate Model

	parameter label	θ	$\hat{\theta}$	Standard Error
Covariance Parameters		a11	1	
		a21	.867	.030
		a22	.230	.032
Ideology (1,2,3)	Gephardt	b11	.181	.025
	Gore	b21	.204	.021
Born in South (0,1)	Gephardt	b12	.294	.042
	Gore	b22	.369	.036
Home State (0,1)	MO-Gephardt	b13	.337	.064
	TN-Gore	b24	.428	.070
$ Per Cap. in State	Gephardt	g1	.00266	.00066
	Gore	g2	.00350	.0053
	Dukakis	g3	-.0134	.0062

Final value of LLF = -4644.991
N = 5282
D = 15

Michael Dukakis is the baseline candidate, so parameters on all individual specific variables are relative to Dukakis.

These parameter estimates will not be discussed in detail, but they conform to expectations. The positive coefficients on the ideology, and south variables indicate that more conservative voters and voters born in the south were more likely to vote for Gephardt and Gore than for Dukakis. The home state variables were also of the expected sign, with Gephardt and Gore receiving larger than average support in their respective home states. The candidate expenditure coefficients all indicate significant marginal impacts of spending by the candidates. Notice that because the Dukakis money enters the utility functions for Gephardt and Gore, the negative coefficient is the expected sign. From the estimated elements of the choleski matrix, we can calculate the covariance terms of the differenced model:

$$\sigma^*_{21} = 1{*}a_{21} = .867 \text{ and } \sigma^*_{22} = a^2_{21} + a^2_{22} = .867^2 + .230^2 = .805$$

Both of these estimates suggest deviation from an MNP with uncorrelated errors.[27] When the same independent variables are included in a conditional logit specification, the effects of the Gephardt ideology and raised in south variables are estimated very imprecisely. The MNP estimates do not have this feature, an indication that freeing up the covariance parameters in the MNP specification clarifies the effects of these variables.[28] Although the estimates suggest the potential benefits of using SML for the estimation of vote choice models, at this point the effects of the independent variables should be considered provisional until the model is estimated with the additional members of the choice set (Jackson, Simon, and Hart) faced by the Super Tuesday voters.

Discussion

As MNP routines are added to standard software packages, they will likely be estimated with increasing frequency. Because they can be very tricky to estimate, however, it will be worthwhile to provide some practical advice that may speed up the estimation process for high

---

27 This is discussed at greater length in the appendix. If the errors in the original model were uncorrleated and of equal variance, the estimates of the two covariance elements would be close to .5 and 1, respectively.

28 Of course, a fairer comparison of the two discrete choice approaches would be to compare the predicted probailities for more completely specified models, a task that will be taken up in later versions of this work.

dimensional MNPs. This advice is by no means comprehensive, but it may prove useful for those frustrated by difficulties with achieving convergence. Of course, there are no guarantees that these strategies will solve every problem, but they may help in some situations.

First, start small. Although it is tempting to start with a fully specified model with no constraints on the covariance matrix, this strategy can end up costing more time in the long run if the optimization routine gets bogged down when trying to sort out the competing effects of independent variables and error covariances. One can start small by restricting the size of the choice set, restricting the number of parameters to be estimated, or by starting with a small number of draws for the GHK simulator. On this last point, the results presented above were obtained by first converging with the number of draws set at five, then using those estimates as starting values for a new set of fifteen draws.

Second, when first starting up, iterate across sets of parameters a few iterations at a time, then keep those parameters fixed when iterating over the next set.[29] In particular, iterating across alternative-specific variables first can be helpful, especially if those variables are very informative in predicting choice behavior. It can be helpful to free up the covariance elements once only small improvements are found when iterating on individual specific and alternative specific variables.

Third, make reasonable restrictions whenever feasible.[30] The model above estimates separate parameters for each candidate for the alternative specific candidate spending variable. Monte Carlo evidence (Geweke, Keane, and Runkle 1994a) suggests that restricting these coefficients to be equal facilitates estimation and convergence, but this restriction was not imposed on the model here because of expectations that some candidates, especially Jesse Jackson, would have differing marginal productivities of campaign spending. For the covariance matrix, it may facilitate estimation to first allow unequal variances, then to allow the covariances to be estimated.

Fourth, when estimating the covariance elements, it may be helpful to estimate one factor at a time. This can be achieved by first estimating the a_{jj} elements, then the a_{j1} elements, the a_{j2} elements, etc. (see appendix for more details). By proceeding in this manner, the structure of the errors can be made incrementally more complex.

This paper largely sets aside the question of precisely when the MNP model is appropriate for examining vote choice. For that question, no simple answer exists. Given current technology, MNPs are becoming much more feasible to estimate, though far from as simple or quick to estimate as models with uncorrelated errors. The recent developments on both the Bayesian and

---

29 This is a strategy employed by Poole and Rosenthal in their estimation of D-Nominate scores. They did it because they had so many parameters, but it can also be useful in highly nonlinear models like the MNP model.

30 Keane (1992) is highly recommended reading for understanding how the effects of individual specific varaibles can "mimic" the effect of error covariances.

classical fronts have made it possible for researchers to use the MNP model to address political science questions in vote choice and other discrete choice situations.

Appendix: The MNP Covariance Matrix

Because the standard MNP specification uses utility differences, the interpretation of the covariance matrix of the estimable model is not always simple or straightforward. Consider the following expressions for the error terms of a five alternative MNP, with the choleski transformation of the error terms included as well.

$$\varepsilon_{i1} - \varepsilon_{i5} = \varepsilon^*_{i1} = a_{11}\eta_1$$

$$\varepsilon_{i2} - \varepsilon_{i5} = \varepsilon^*_{i2} = a_{21}\eta_1 + a_{22}\eta_2$$

$$\varepsilon_{i3} - \varepsilon_{i5} = \varepsilon^*_{i3} = a_{31}\eta_1 + a_{32}\eta_2 + a_{33}\eta_3$$

$$\varepsilon_{i4} - \varepsilon_{i5} = \varepsilon^*_{i4} = a_{41}\eta_1 + a_{42}\eta_2 + a_{43}\eta_3 + a_{44}\eta_4$$

Notice that each of the transformed errors contains the error term for the baseline alternative. Therefore, even if the original model has a covariance matrix with uncorrelated errors, the transformed errors will be correlated. For example:

$$\sigma^*_{21} = \text{cov}(\varepsilon^*_{i1}, \varepsilon^*_{i2}) = E[(\varepsilon_{i1} - \varepsilon_{i5})(\varepsilon_{i2} - \varepsilon_{i5})]$$

$$= \varepsilon_{i1}\varepsilon_{i2} - \varepsilon_{i1}\varepsilon_{i5} - \varepsilon_{i5}\varepsilon_{i2} + \varepsilon_{i5}\varepsilon_{i5}$$

$$= \sigma_{12} - \sigma_{15} - \sigma_{52} + \sigma_{55}$$

In this case, if the errors in the original model are uncorrelated, the covariance element will equal the variance of the baseline alternative. It is simple to show that if the original model has i.i.d. errors, the transformed model will have equal variances, but correlations of .5 for all the off diagonal elements.[31] The fact that the transformed covariance terms contain multiple covariance elements from the untransformed model makes is difficult to interpret the covariance elements directly. Given that $\eta \overset{iid}{\sim} N(0,1)$, however, the transformed covariance elements can be calculated by multiplying the appropriate errors as described above.

$$\sigma^*_{11} = a_{11}$$

$$\sigma^*_{21} = a_{11}a_{21}$$

$$\sigma^*_{22} = a^2_{21} + a^2_{22}$$

....

$$\sigma^*_{43} = a_{31}a_{41} + a_{32}a_{42} + a_{33}a_{43}$$

$$\sigma^*_{44} = a^2_{41} + a^2_{42} + a^2_{43} + a^2_{44}$$

---

31 This holds for multinomial logit, where the error correlations in the transformed model are fixed at .5.

As discussed above, to facilitate estimation, it may sometimes be desirable to fix elements of the choleski matrix, which in turn reduces the implied number of factors in the error covariance matrix.

References

Abramson, Paul, John Aldrich, Phil Paolino, and David Rohde. 1992. "'Sophisticated'" Voting in the 1988 Presidential Primaries." *American Political Science Review* 86: 55-69.

Aldrich, John and R. Michael Alvarez. 1994. "Issues and the Presidential Primary Voter." *Political Behavior* 16: 289-317.

Alvarez, R. Michael and Jonathan Nagler. 1995. "Economics, Issues, and the Perot Candidacy: Voter Choice in the 1992 Presidential Election." *American Journal of Political Science* 39:714-744.

Alvarez, R. Michael and Jonathan Nagler. 1996. "Correlated Disturbances in Discrete Choice Models: A Comparison of Multinomial Probit Models and Logit Models." Typescript.

Bartels, Larry. 1988. *Presidential Primaries and the Dynamics of Public Choice*. Princeton: Princeton University Press.

Börsch-Supan, Axel and Vassilis Hajivassiliou. 1993. "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models." *Journal of Econometrics* 58: 347-368.

Bunch, David. 1991. "Estimability in the Multinomial Probit Model." *Transportation Research B*, 25B: No. 1, 1-12.

CBS News, *New York Times*. 1988. CBS News *New York Times Super Tuesday Primary Election Exit Polls,* 1988 [computer file]. New York: CBS News, *New York Times* [producer]. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor, ICPSR #9139].

Erdem, Tulin and Michael Keane. 1996. "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets." *Marketing Science* 15(1): 1-20.

Geweke, John. 1991. "Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints." In *Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface* , 571-578. Alexandria, VA: American Statistical Association.

Geweke, John, Michael Keane, and David Runkle. 1994a. "Alternative Computational Approaches to Inference in the Multinomial Probit Model." *Review of Economics and Statistics,* 76: 609-632.

-----. 1994b. "Statistical Inference in the Multinomial Multiperiod Probit Model." Federal Reserve Bank of Minneapolis Research Department, Staff Report 177. [Also forthcoming in *Journal of Econometrics*.]

Greene, William. 1997. *Econometric Analysis*, 3rd ed. Saddle River, NJ: Prentice Hall.

Helland, Eric and Andy Whitford. 1995. "Bayesian Specification Tests and Fenno's Core Constituency." Paper presented at the annual meeting of the Midwest Political Science Association.

Hajivassiliou, Vassilis and Daniel McFadden. 1990. "The Method of Simulated Scores for the Estimation of LDV Models with an Application to External Debt Crises." Cowles Foundation Discussion Paper 967, Yale University. [forthcoming in *Econometrica* 1997]

Hajivassiliou, Vassilis, Daniel McFadden, and Paul Ruud. 1996. "Simulation of Multivariate Normal Orthant Probabilities: Methods and Programs."*Journal of Econometrics* 72: 85-134.

Hausman, Jerry and David Wise. 1978. "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." *Econometrica* 46: 403-426.

Jackman, Simon. 1995. "Bayesian Tools for Social Scientists." Paper presented at the annual meeting of the American Political Science Association.

Keane, Michael. 1990. *Four Essays in Empirical Macro and Labor Economics*. Ph.D. Dissertation, Brown University.

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

Keane, Michael. 1994. "A Computationally Practical Simulation Estimator for Panel Data." Econometrica 62(1): 95-116.

Lee, Lung-Fei. 1995. "Asymptotic Bias in Simulated Maximum Likelihood Estimation of Discrete Choice Models." *Econometric Theory* 11: 437-483.

Lerman, Steven and Charles Manski. 1981. "On the Use of Simulated Frequencies to Approximate Choice Probabilities." In *Structural Analysis of Discrete Data with Econometric Applications*, Charles Manski and Daniel McFadden, eds. Cambridge, MA: MIT Press.

McCulloch, Robert and Peter Rossi. 1994. "An Exact Likelihood Analysis of the Multinomial Probit Model." *Journal of Econometrics*, 64(1): 207-240.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, Paul Zarembka, ed. Academic Press.

McFadden, Daniel. 1981. "Econometric Models of Probabilistic Choice." In *Structural Analysis of Discrete Data with Econometric Applications*, Charles Manski and Daniel McFadden, eds. Cambridge, MA: MIT Press.

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

Norrander, Barbara. 1992. *Super Tuesday: Regional Politics and Presidential Primaries*. Lexington: University of Kentucky Press.

Norrander, Barbara. 1993. "Nomination Choices: Caucus and Primary Outcomes, 1976-1988." *American Journal of Political Science* 37: 343-364.

Pakes, Ariel and David Pollard, 1989. "Simulation and the Asymptotics of Optimization Estimators." *Econometrica*, 57: 1027-1058.

Quinn, Kevin, Andrew Martin, and Andrew Whitford. 1996. "Explaining Voter Choice in a Multiparty Democracy: A Look at Data from the Netherlands." Presented at the Annual Meeting of the American Political Science Association, San Francisco.

Stone, Walter, Ronald Rappaport, and Lynn Atkenson. 1995. "A Simulation Model of Presidential Nomination Choice." *American Journal of Political Science*. 39: 135-161.

Whitten, Guy and Harvey Palmer. 1996. "Heightening Comparativists' Concern for Model Choice: Voting Behavior in Great Britain and the Netherlands." *American Journal of Political Science*. 40: 231-260.

Wilcox, Clyde. 1991. "Financing the 1988 Prenomination Campaigns." In *Nominating the President*., Emmett Buell Jr. and Lee Sigelman, eds. Knoxville: University of Tennessee Press.