



# Multinomial probit and multinomial logit: a comparison of choice models for voting research

Jay K. Dow \*, James W. Endersby

*Department of Political Science, University of Missouri, Columbia, MO 65201-0630, USA*

---

## Abstract

Several recent studies of voter choice in multiparty elections point to the advantages of multinomial probit (MNP) relative to multinomial/conditional logit (MNL). We compare the MNP and MNL models and argue that the simpler logit is often preferable to the more complex probit for the study of voter choice in multi-party elections. Our argument rests on three areas of comparison between MNP and MNL. First, within the limits of typical data—a small sample of revealed voter choices among a few candidates or parties—neither model will clearly appear to have generated the observed data. Second, MNP is susceptible to a number of estimation problems, the most serious of which is that the MNP is often weakly identified in application. Weak identification is difficult to diagnose and may lead to plausible, yet arbitrary or misleading inferences. Finally, the logit model is criticized because it imposes the independence of irrelevant alternatives (IIA) property on voter choice. For most applications the IIA property is neither relevant nor particularly restrictive. We illustrate our arguments using data from recent US and French presidential elections.

© 2003 Elsevier Ltd. All rights reserved.

---

## 1. Introduction

Models of voter choice among more than two parties or candidates rest on underlying assumptions about the nature of individual decision-making. Statistical methods commonly used to estimate models of multiparty vote choice often impose restrictive assumptions about these choices that render inferences suspect. For example, binary logit or probit analyses are sometimes used to represent voter choice as a decision

---

\* Corresponding author. Tel.: +1-573-882-0047; fax: +1-573-884-5131.

E-mail addresses: [dowj@missouri.edu](mailto:dowj@missouri.edu) (J.K. Dow); [endersby@missouri.edu](mailto:endersby@missouri.edu) (J.W. Endersby).

between government and opposition. This ignores potentially important differences within the opposition. Less restrictive are ordered models including ordered logit, ordered probit and least squares. These represent voter choice among multiple candidates or parties, but impose the assumption that political competition takes place along a single, ordered axis. The unidimensionality assumption is typically invalid for multiparty systems (Lijphart, 1984; Taagepera and Shugart, 1989: pp. 92–98). Stokes (1963) provided an early criticism of the ordered axis assumption.

Alvarez and Nagler (1995, 1998, 2001), Alvarez et al. (2000), Schofield et al. (1998), Lacey and Burden (1999) and Quinn et al. (1999) identify advantages of multinomial probit (MNP) analysis in these applications. MNP fits within a class of multinomial choice models that includes the multinomial/conditional logit (MNL), and more flexible logit-type specifications including the generalized extreme value (GEV) and mixed logit (MXL) models. These represent voter choice as a decision among unordered alternatives (parties, candidates) as a function of chooser (voter) and choice (party/candidate) attributes. Since the MNL specification is tractable and simple to estimate, it is included in many commercial software packages and is used regularly in electoral research.<sup>1</sup> MNL, however, imposes the restrictive assumption that choices are independent across alternatives. MNP does not impose the independence assumption and advances in computer technology make its estimation increasingly practical. Thus, one might reasonably argue that MNP should be the benchmark methodology in the study of voter choice in multiparty elections.

Although political methodologists understand the relative merits of MNP and MNL in application, our experience suggests that many scholars investigating voter choice in multiparty elections are not fully aware of the assumptions and implications of the two models. We seek to inform applied researchers about the benefits and liabilities of MNP and MNL and to compare the two models in typical application. We assess MNP and MNL in the study of voter choice in multicandidate/multiparty elections and argue that for most purposes the simpler logit is preferable to probit. This is particularly true for applications estimating the probability that a voter casts a ballot for a candidate or a party selected from a fixed, stable pool of alternatives and the marginal changes in the probability associated with changes in the values of explanatory variables. We are motivated by our perception that electoral scholars are sometimes uncertain about MNP and associated issues such as the independence of irrelevant alternatives property (IIA). In particular, researchers are unsure as to whether their statistical analyses provide useful insights or are subject to challenge if based on MNL. We demonstrate that for the most common application in which these concerns arise—the study of voter choice in multiparty/candidate elections—these concerns are exaggerated and, under most circumstances, logit estimation performs as well or better than MNP.

Our argument rests on three areas of comparison between the MNP and MNL models. First, since the statistical properties motivating MNP and MNL are asymp-

---

<sup>1</sup> We do not distinguish between strictly multinomial (MNL) and conditional (CL) logit models as most electoral applications use a logit specification that combines the two.

totic, there are few a priori reasons to believe that one method or the other is superior in typical application. The relative merits of any statistical model primarily depend on the method that best represents the underlying choice process that generates the observed data. Within the limits of typical data—a relatively small sample vector of observed choices among a few candidates or parties—neither model will clearly appear to have produced the sample of observed data. The estimated MNP and MNL coefficients and standard errors (adjusted for scale), probabilities and marginal effects will be indistinguishable.

Second, MNP imparts a number of potentially serious problems. These are sufficiently difficult to detect that, in the absence of investing exceptional effort in model diagnostics, researchers are justified in using the MNL specification. The most important problem is that even formally identified MNP specifications are often weakly identified in application. This is serious because weak identification is difficult to diagnose and may lead to plausible, yet arbitrary or misleading inferences. The MNP presents a difficult maximum likelihood optimization problem that sometimes fails to converge at a global optimum or produces parameter estimates that are sufficiently imprecise as to make statistical inferences suspect. Except for cases of profound mis-specification, the logit likelihood will optimize at its global maximum and is not prone to optimization errors.

Finally, the logit model is criticized because it imposes the IIA property on voter choice. The IIA property imposes the restriction that the relative odds of selecting between any two candidates or parties is independent of the number of alternatives. Behavioral theories of voter choice sometimes assume IIA is violated, so logit estimation is suspect. For much of empirical voting research, the IIA property is not particularly restrictive. We illustrate our arguments using data from recent US and French presidential elections.

## 2. The multinomial logit and multinomial probit models

### 2.1. Overview

Scholars of voting behavior must adopt theoretically appropriate models of voter choice among multiple alternatives (Whitten and Palmer, 1996; Alvarez and Nagler, 1998; Adams and Merrill, 2000). The MNL and MNP models fit this requirement as each may be derived from economic theories of utility maximization. In multicandidate elections, assume voter  $i$ 's utility for candidate  $j$ ,  $U_{ij}$ , ( $i = 1, \dots, n$ ;  $j = 1, \dots, p$ ) is a function of both voter and candidate attributes and a stochastic error. A typical representation (Maddala, 1983: p. 61; Alvarez and Nagler, 1998; Powers and Xie, 2000: p. 248) is:

$$U_{ij} = \beta'X_{ij} + \alpha'Z_i + \varepsilon_{ij}$$

where  $X_{ij}$  is a vector of candidate attributes as perceived by voter  $i$ , and  $Z_i$  is a vector of voter characteristics. We seek to estimate the coefficients  $\beta$  and  $\alpha_j$  and their standard errors.

If each voter is a utility maximizer, the probability that voter  $i$  casts a ballot for candidate 1 is:

$$P_{i1} = P[U_{i1} > U_{i2}, U_{i1} > U_{i3}, \dots, U_{i1} > U_{ip}]$$

For any “ $m$ ” in the set of  $1, \dots, p$  candidates:

$$P(m) = P[\varepsilon_{im} - \varepsilon_{ij} < (\beta' X_{ij} + \alpha' Z_i) - (\beta' X_{im} + \alpha' Z_i), j \neq m]$$

MNL and MNP differ according to the assumed error structure for the  $\varepsilon_{ij}$ . The logit assumes the errors are independent, identically distributed (iid) with type I extreme value distribution (log Weibull). The probability that a voter  $i$  votes for candidate  $j$  is given by:

$$P(\text{vote} = j | \beta, \alpha_j, X_{ij}, Z_i) = \frac{\exp(\beta' X_{ij} + \alpha' Z_i)}{\sum_{k=1}^p \exp(\beta' X_{ik} + \alpha' Z_i)}$$

The MNP assumes the errors are distributed multivariate normal, with mean 0 and

covariance matrix  $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \sigma_{1n} & \dots & \dots & \sigma_n^2 \end{bmatrix}$ . The probabilities are written:

$$P(\text{vote} = j | \beta, \alpha_j, X_{ij}, Z_i, \Sigma^*) \\ = \int_{-\infty}^{\beta^* X_1^* + \alpha_1^* Z^*} \dots \int_{-\infty}^{\beta^* X_{j-1}^* + \alpha_{j-1}^* Z^*} f(\varepsilon_1^*, \dots, \varepsilon_{j-1}^*) \partial \varepsilon_1^*, \dots, \partial \varepsilon_{j-1}^*$$

where  $f(\cdot)$  is the probability density function of the multivariate normal distribution.<sup>2</sup>

## 2.2. General comparisons

Given the effort invested to argue the merits of one method over the other, the MNP and MNL are remarkably similar in important respects. Each returns sigmoid probabilities that sum to one over all alternatives. The desirable statistical properties of the MNP and MNL estimators include consistency, normality and efficiency. These are asymptotic properties and relatively little is known about the small sample behavior of either estimator (King, 1989; Long, 1997).

Probit offers a potential advantage over MNL in that the MNP error specification

<sup>2</sup> The MNL/CL and MNP models have to be identified by placing restrictions on the model before estimation. We subtracting the reference candidate/party utility from the remaining utility to identify the MNL/CL and MNP models estimated in this study. We also place necessary identification restrictions on the differenced MNP covariance structure. See, for example, Long (1997: pp. 153–154).

allows correlations between the errors. Examination of the MNP covariance structure, however, underscores the first of several potential problems in applying this statistical method in electoral research. The model proliferates parameters (Horowitz, 1991). A ‘ $p$ ’ alternative model adds  $K = [(p)(p + 1)/2]$  covariance parameters to the model. Of these, at most  $K^* = [(p)(p-1)/2-1]$  may be estimated.<sup>3</sup> For example, the three party model discussed above presents  $K = [(3)(4)/2] = 6$  distinct covariance terms, only  $K^* = [(3)(2)/2-1] = 2$  of which may be estimated. These terms are a function of all of the elements of the error covariance matrix and may not be readily interpreted in terms of variances and covariances associated with specific candidates or choices.

The MNP presents a difficult computational problem relative to the logit. The probit likelihood function is often flat near its optimum. In this case, the MNP model may produce arbitrary parameter estimates within the tolerance of the estimation procedure (Keane, 1992; Alvarez and Nagler, 1998). A typical sample identification strategy is to include one alternative specific variable in each utility. While this often helps, this restriction does not guarantee convergence at a global optimum within the tolerance of the software. The MNP likelihood simply presents a difficult optimization problem because (1) even with restrictions it is still relatively flat and (2) because it generally requires numerical approximation for the multivariate integrals. The logit does not require numerical integration and almost always converges to a global optimum.

The primary advantage of MNP relative to MNL centers on the IIA property. To assess the importance of the IIA property, recall that it comes to us from Arrow (1951) as a *desirable* condition imposed on axiomatic choice behavior. The idea (Arrow 1951: p. 26; Luce and Raiffa, 1957: p. 338; Luce, 1959: p. 9) is that if a chooser is comparing two alternatives according to a preference relationship, the ordinal ranking of these alternatives *should* not be affected by the addition or subtraction of other alternatives from the choice set. As Arrow describes the IIA property, it is a minimal condition for logical consistency. The probabilistic analog imposed by MNL, but not MNP, strengthens this by requiring that the odds ratio of choosing any two alternatives be independent of the addition or subtraction of other alternatives from the choice set. Specifically, the ratio of choice probabilities for any two alternatives does not depend on the characteristics of any of the other alternatives. It is quite easy to see why this is the case. Consider an election contested by three parties: D, R and G. The probability ratio of voting for parties D and R is:

$$\begin{aligned} \frac{P(\text{vote} = D | \beta, \alpha_D, X_{iD}, Z_i)}{P(\text{vote} = R | \beta, \alpha_R, X_{iR}, Z_i)} &= \frac{\exp(\beta' X_{iD} + \alpha'_D Z_i)}{\sum_{k=1}^3 \exp(\beta' X_{ik} + \alpha'_k Z_i)} \bigg/ \frac{\exp(\beta' X_{iR} + \alpha'_R Z_i)}{\sum_{k=1}^3 \exp(\beta' X_{ik} + \alpha'_k Z_i)} \\ &= \exp(\beta'(X_{iD} - X_{iR})) \end{aligned}$$

<sup>3</sup> The number of covariance parameters that can be estimated may be less than  $K^*$  depending on the empirical error structure (Bunch, 1991, 1992).

That is, the ratio of the choice probabilities is simply a function of the difference in the alternative specific variable  $X$  for parties D and R. For example, if  $X_{iD}$  and  $X_{iR}$  are voter  $i$ 's respective spatial distances to parties D and R, then the probability ratio is the exponential of  $\beta$  times the difference in spatial proximity. The IIA property arises because all information about party G—such as whether it is similar or dissimilar to party D or party R—cancels out of the expression. Among other things, this means that the logistic model does not estimate substitution patterns across choices well. If, for example, one sought to infer how the D and R party vote shares would change if party G were removed from the election, these estimates would be inaccurate to the extent that voters see party G as a close substitute for either of the remaining parties.

While imposition of the IIA is restrictive for behavioral choice models, we should not exaggerate its importance in applied settings. First, IIA is a logical property of decision-making, not a statistical property such as consistency and unbiasedness. As a logical property, one expects some version of IIA to hold in both theoretical models of voter choice and observed choice behavior. Second, any model that specifies choice probabilities that violate the IIA property may be expressed in logit form with appropriate choice of right-hand-side variables. The estimated logit probabilities will be consistent with violation of IIA (Train, 1993: pp. 18–24).<sup>4</sup> Typically, this involves little more than including choice specific intercepts. That is, if one estimates a properly specified logit model with choice specific intercepts, the returned probabilities capture the observed choice frequencies regardless of whether the underlying choice process is consistent with IIA.<sup>5</sup>

The substantive questions relevant to IIA center on party and candidate entry and exit from elections. Scholars seeking to estimate the substitution patterns among candidates and parties—how party vote shares change if other parties enter or exit the election—cannot easily obtain this information from MNL and, appropriately, rely instead on MNP (Lacey and Burden, 1999) or more general logistic models (see, for example, Glasgow, 2001). Substitution patterns inform important questions in the study of voting and elections, and empirical researchers have paid insufficient attention to the relative substitutability of parties in democratic elections. However, for most studies these are not central concerns. Many democratic elections are contested by a fixed, stable pool of parties. The question of “how would Israeli voters choose if”, for example, Likud dropped out of Knesset elections is seldom of inter-

---

<sup>4</sup> These variables must typically measure characteristics of the alternatives other than the alternative for which the utility in question is defined. Of course, if the choice process admits violations of IIA, the estimation of choice specific intercepts in a logistic model will return the correct (relative frequency) aggregate probabilities.

<sup>5</sup> Our argument in this paragraph is conceptually similar to Alvarez and Nagler's (1998, Appendix C) observation that “IIA does not aggregate”. Their derivations show that even if individual choices reflect IIA, the introduction of a new alternative will produce mean probabilities that reflect violations of IIA. Our more general, empirical, point is that regardless of whether individual choice reflect, or do not reflect, IIA, the estimated model parameters will return the correct aggregate probabilities.

est.<sup>6</sup> Concern over the IIA property is most appropriate in candidate-centered elections—especially those where candidates enjoy ease of entry and exit and are seen as close substitutes—and one seeks to estimate the consequences of hypothetical changes in choice sets. US primary elections and nonpartisan contests for local offices present such opportunities.<sup>7</sup>

### 3. Application to data

#### 3.1. Overview

The relative merits of MNP and MNL primarily depend on which model better represents the stochastic process that generated the observed data. To answer this question, we are aware of only two studies that compare the MNP and MNL in political science applications: Alvarez and Nagler (2001) and Quinn et al. (1999). The former study argues for the use of MNP in electoral studies while the latter provides a more balanced assessment. Our view is that neither study provides convincing evidence that the probit specification is preferable under most circumstances.

Alvarez and Nagler (2001) use Monte Carlo analysis to compare MNP and the independent MNP (IMNP) in small samples in which the data generating process is MNP. The independent, MNP is simply the MNP with the off-diagonal error covariances constrained to equal zero. When the true model error structure is correlated, the independent MNP often fails to return reliable estimates of the probit coefficients, but the model probabilities are accurate. This seems counterintuitive. However, since the estimation method is maximum likelihood, the presence of modest mis-specification, such as constraining the off-diagonal elements of the MNP error covariance matrix to zero, produces structural parameters that adjust to maximize the likelihood of reproducing the sample frequencies. We question the extent to which this study greatly informs the comparison of MNP and MNL in application. Alvarez and Nagler use the IMNP as a surrogate for the MNL, but the MNP and IMNP log-likelihoods

---

<sup>6</sup> Of course, if concerned about IIA, we may directly test for violations of the assumption. Hausman and McFadden (1984) provide a relatively simple test statistic for determining whether the odds ratios for any two alternatives are affected by changes in the set of available alternatives. As with any other empirical question, even if in principle imposition of IIA is a concern, *in practice*—given data and specification—the IIA may not be a problem. See Long (1997: pp. 183–184) and Powers and Xie (2000: p. 247), for presentations of this Wald type test.

<sup>7</sup> When presented with choices that are, at least from most perspectives, close substitutes such as voter choice from a set that includes candidates from the French PCF and the Trotskyist parties, one might reasonably collapse these choices and code the respondents who voted for either of these parties' candidates as voting "far left". Combining substitutable parties in proportional representation systems might be better analyzed as choices among coalition partners rather than as independent, competitive parties. Collapsing categories with few observations also makes sense since there is often little to be gained by attempting to predict low probability events. Within the limits of typical data, neither the MNP nor MNL are capable of predicting low probability events at acceptable levels of precision. (But see King and Zeng (2001), who propose a limited dependent variable estimator for this purpose.)

have the same functional form. Consequently, their simulations provide limited information on the relative advantages of MNL that obtain from the logit's simpler likelihood as reflected in optimization, sample identification and parameter variation. Quinn et al. (1999) apply Bayesian methods to assess the fit of MNP and MNL to relatively simple models of British and Dutch voter choice. They report that the MNP better fits the Dutch data, while there is no difference in MNP and MNL fit to the British data.

### 3.2. Research design

We read these studies, as well as related studies by Horowitz (1980), Keane (1992) and Geweke et al. (1994), as providing little evidence that there are significant, tangible, gains from the MNP specification in typical political science application.<sup>8</sup> Transportation engineers might gain increases in statistical performance in estimating models of the tens of thousands of commuters making daily choices among cars, trains and buses on the basis of chooser and vehicle attributes. In political science application, however, MNP and MNL are simply too similar to allow one to gain purchase when the dependent variable consists of approximately 1000 observations from a National Election Study. Modest differences in distributional assumptions and the addition of a very few covariates in a model in which the systematic component is otherwise properly specified will rarely allow an empirical researcher to distinguish similar data generation processes given such limited data.

To test this argument we apply each model to data from the 1992 US presidential election and the 1995 French presidential election. The data are obtained from the 1992 American National Election Study and the 1995 French National Election Study. Alvarez and Nagler (1995) use the former in their study of US voter choice. Dow (1999) uses the latter data in a study of French voter decisions. For the analysis of US and French voter choice, we estimate the parameter values and the variance in the MNP and MNL estimates. We also calculate the MNP and MNL probabilities for changes in several variables to assess whether either method would lead to different interpretations of the data. Finally, we evaluate the numerical accuracy of the MNP and MNL optimization procedures. Our objective is to determine whether there are differences in the performance of the estimation methods according to standards for evaluating limited dependent variable models in applied research.

### 3.3. Application to US election data

Our data and specification for the 1992 US presidential election is nearly identical to Alvarez and Nagler's (1995) MNP analysis of voter choice in the same election.<sup>9</sup>

---

<sup>8</sup> For example, in a study relevant to transportation research, Horowitz (1980) reports that an average sample size of approximately 2000 is required to distinguish relatively simple logistic and MNP models. Geweke et al. (1994: p. 626) report significant optimization problems when estimating parameters using both classical and Bayesian methods in a complex (seven alternative) model using 1000 observations.

<sup>9</sup> There are two relatively minor differences between our estimation procedure and that of Alvarez and Nagler. First, we use the replication data set provided by Alvarez and Nagler that corrects some minor coding errors in the original work. Second, Alvarez and Nagler seek to estimate three covariance



We model voter choice among Clinton, Bush and Perot as a function of voter-candidate ideological proximity, voter issue preferences and voter demographic characteristics with Perot as the reference candidate. Table 1 presents the MNP and MNL parameter estimates and measures of fit.

Our MNP estimates largely replicate those of Alvarez and Nagler. The probit and

Table 1  
MNP and MNL estimates for voter choice in 1992 US presidential election

	MNP		MNL	
	Bush/Perot	Clinton/Perot	Bush/Perot	Clinton/Perot
Intercept	0.48 (0.55)	−0.44 (0.72)	0.67 (0.80)	−0.41 (0.71)
Ideological distance		−0.08** (0.02)		−0.12** (0.02)
Personal finances	−0.04 (0.05)	0.02 (0.05)	−0.07 (0.07)	0.02 (0.07)
National economy	−0.13 (0.10)	0.21* (0.09)	−0.21* (0.09)	0.27 (0.10)
Government jobs	0.07 (0.06)	−0.01 (0.05)	0.11 (0.08)	−0.01 (0.07)
Government health care	0.10 (0.06)	0.06 (0.05)	0.18** (0.06)	0.11 (0.06)
Minority assistance	0.01 (0.05)	−0.17** (0.05)	0.01 (0.07)	−0.24* (0.07)
Abortion	−0.33 (0.19)	0.01 (0.12)	−0.52** (0.11)	−0.02 (0.11)
Region—East	−0.15 (0.22)	0.31 (0.23)	−0.27 (0.27)	0.45 (0.25)
Region—South	0.24 (0.21)	0.49* (0.20)	0.36 (0.25)	0.70 (0.27)
Region—West	−0.11 (0.20)	−0.03 (0.22)	−0.17 (0.25)	−0.03 (0.25)
New/returning voter	0.27 (0.22)	−0.22 (0.21)	0.49** (0.16)	−0.25 (0.17)
Term limits	0.06 (0.20)	0.08 (0.20)	0.08 (0.22)	0.12 (0.22)
Deficit a problem	−0.56 (0.30)	−0.01 (0.20)	−0.90** (0.19)	−0.02 (0.19)
Democrat	−0.18 (0.38)	1.29** (0.24)	−0.36 (0.32)	1.57** (0.23)
Republican	0.95 (0.57)	−0.72 (0.46)	1.48** (0.27)	−0.86** (0.27)

(continued on next page)

terms in their MNP model. This means their model is technically unidentified. This appears to have minor implications for the coefficient estimates, but likely accounts for some differences in the estimated standard errors.

Table 1 (continued)

	MNP		MNL	
	Bush/Perot	Clinton/Perot	Bush/Perot	Clinton/Perot
Female	0.37 (0.23)	0.21 (0.16)	<i>0.54**</i> (0.19)	0.30 (0.19)
Education	0.14 (0.09)	0.01 (0.06)	<i>0.22**</i> (0.08)	0.01 (0.07)
Age: 18–29	–0.82 (0.52)	–0.55* (0.33)	<i>–1.38**</i> (0.36)	–0.87** (0.33)
Age: 30–44	–0.61 (0.38)	–0.52 (0.26)	<i>–1.00**</i> (0.301)	–0.77** (0.29)
Age: 45–59	–0.49 (0.31)	–0.10 (0.26)	<i>–0.82**</i> (0.31)	–0.19 (0.29)
$\sigma_1$	0.01 (0.50)		–	
$\sigma_2$	0.34 (0.70)		–	
Log <sub>10</sub> condition hessian	8.01		4.08	
Likelihood ratio statistic	767.54		769.08	
Likelihood ratio statistic for $\Sigma$	0.16		–	
Percent correctly predicted	74.06		74.95	

$n = 909$ .

\*  $p < 0.10$ .

\*\*  $p < 0.05$ .

logit likelihood ratio statistics indicate both specifications fit the data reasonably well. Both estimation methods correctly classify about the same proportion of observed choices. Finally, the likelihood ratio test for the MNP estimated error covariance shows that estimating a general error structure does not improve the fit of the model. This is further confirmed by the failure of the two estimated error correlations to obtain statistical significance.

Since the error correlations are estimated at zero, it is surprising that there are several important differences between the MNP and MNL structural coefficient estimates and the resulting inferences one would make about voter choices in the 1992 election. Using, as Alvarez and Nagler do, a generous 0.10 level of statistical significance, approximately twice as many MNL coefficients reach statistical significance as MNP coefficients. Using the standard 0.05 level, 13 MNL coefficients obtain statistical significance when the corresponding MNP coefficients do not. We use italics in the table to highlight the MNL coefficients that obtain statistical significance where the corresponding MNP coefficient does not. The largest discrepancies are found in the Bush/Perot comparison. Here, the MNP coefficients for the national economy, government health care, abortion, new or returning voter and views toward the deficit fail to obtain statistical significance whereas the corresponding MNL coefficients are statistically significant and in the expected direction. In the probit speci-

fication, even Republican Party identification is only a weak predictor of the Bush vote.

Why are there such major discrepancies between the probit and logit estimates? They are not due to MNL mis-specification resulting from off-diagonal elements in the candidate error structure. We believe they result from problems in the probit numerical optimization that produces larger coefficient standard errors. While there is no simple measure of statistical “precision” beyond estimated standard errors and confidence bands, we can gain insight on the extent to which the shape of the log-likelihoods and estimation methods produces these differences by calculating the condition numbers of the MNP and MNL hessian matrices (i.e. the parameter covariance matrix). The condition number (Greene, 1993: pp. 39–40) is the square root of the ratio of the largest and smallest roots of a matrix. In optimizations of this type, the  $\log_{10}$  of the hessian condition number provides a measure of numerical precision. As a general rule of thumb (Judd, 1998: pp. 67–70), if the  $\log_{10}$  of the hessian condition number is less than three or four then the optimization results are likely stable and accurate to several decimal places. A  $\log_{10}$  hessian condition greater than 10 indicates potential instability and lack of numerical precision in calculations such as matrix inversion.<sup>10</sup> For US data, the  $\log_{10}$  condition for the MNP hessian is 8.01, while the  $\log_{10}$  condition for the MNL hessian is 4.07. This indicates that the probit likelihood is less numerically stable than the logit likelihood, and this may account for the differences in the statistical significance of several key variables in the models.

We next determine if there are differences between MNP and MNL in the vote choice probabilities predicted by each estimation method. We do this by calculating in turn the expected vote probabilities for changes in several independent variables holding all other variables fixed at their mean values. For comparison purposes, several of these calculations ignore the fact that corresponding probit coefficients are not statistically significant and one would not ordinarily calculate marginal effects for these variables.<sup>11</sup> We do not present these marginal effects in the interest of conserving space, but there are virtually no differences in the estimated MNP and MNL probabilities. This is not unexpected given the similarities in the shapes of the

---

<sup>10</sup> King (1989: pp. 103–104) briefly makes this same point by noting that one may use the hessian matrix to measure of the shape of the log-likelihood function at its maximum. The relationship between the degree of numerical accuracy and the hessian condition number depends in part on the type of computer that one uses to estimate the model. A standard Pentium class computer carries approximately 12 significant digits without the use of an analytical hessian matrix. In this case, a  $\log_{10}$  hessian condition greater than 10 pushes the boundary of numerical accuracy upon which one might comfortably make statistical inferences. However, other types of computers carry many more significant digits, and a  $\log_{10}$  hessian condition significantly greater than 10 does not present cause for concern.

<sup>11</sup> The marginal change in the probability of the dependent variable for a given change in an independent variable is the estimated coefficient multiplied by the cumulative density function evaluated at the value of interest. If the coefficient is not statistically distinguishable from zero, then this product equals zero times some positive number. That is, the marginal effect is zero. Although we do not present tables detailing the estimated marginal effects for the US and French elections, this information is available from the authors on request.

logit and probit probability distributions. Major differences would arise only if there were significant differences in the estimated coefficients relative to their respective distributions. Still, the smaller MNL confidence bands mean that the corresponding high and low ranges for the estimated marginal effects are considerably smaller than the corresponding ranges for the MNP estimates. The substantive implication of these figures is that there are few differences in the expected vote or in the expected vote responses (marginal effects) suggested by either estimation method, except that we are more confident in the ranges of these responses when we use MNL.

### 3.4. *Application to French election data*

For the 1995 French presidential election, we present the MNP and MNL coefficients for voter choice on the first ballot in [Table 2](#). We model first ballot voter choice among Socialist Lionel Jospin, Gaullists Jacques Chirac and Edward Balladur, and National Front candidate Jean Marie le Pen. This listing corresponds to the left–right order of the candidates, with le Pen representing the resurgent French far right. Our model is simple, but captures the [Lewis-Beck and Skalaban \(1992\)](#) argument that ideology, religion and class are the three major predictors of French voter choice. Unlike the US specification, we estimate separate coefficients for the ideological distance variable for each candidate utility. This allows for the possibility that voter responsiveness to ideological distance may differ across candidates. We also include two issue variables relevant to the 1995 campaign: views toward European integration and immigration. Jospin is the reference candidate.

As the model likelihood ratio statistics indicate, both the MNP and MNL estimates fit the data well. The inclusion of the off-diagonal candidate error terms improves the fit of the model as measured by the likelihood ratio test for the error covariance matrix. This is substantiated by the two positive estimated error covariances. As in the US case, the logit correctly classifies a few more observations than the probit. The  $\log_{10}$  hessian condition numbers for the MNL and MNP are approximately 9 and 10, respectively, indicating that while both optimizations are likely less stable than one would hope for, the MNL is again more stable than the MNP.

The most notable characteristic of the French analysis is that despite the statistically significant off-diagonal errors, both estimation methods produce virtually identical information. Each estimation method produces the same number of statistically significant coefficients, and the predicted vote and marginal effects derived from MNP and MNL coefficients are again nearly identical. For example, the marginal effects reveal that both MNP and MNL pick up about the same religiosity effects, with Jospin and Balladur largely splitting the secular—non-secular ballot. Both record the effects of voter views toward the Maastricht treaty, with le Pen drawing support of opponents in about equal proportion from all candidates. The greatest MNP and MNL differences in marginal effects are in the ideological proximity variable, but these are modest and likely result from allowing the candidate specific ideological proximity coefficients.

Table 2  
MNP and MNL estimates for voter choice in the 1995 French presidential election

Independent variables	MNP			MNL		
	Chirac/ Jospin	Balladur/ Jospin	le Pen/ Jospin	Chirac/ Jospin	Balladur/ Jospin	le Pen/ Jospin
Intercept	−0.93** (0.35)	−1.88** (0.67)	−2.16** (0.63)	−2.04** (0.62)	−3.77** (0.70)	−3.28** (0.88)
Ideological distance	−0.35** (0.09)	−0.29** (0.08)	−0.28** (0.04)	−0.76** (0.04)	−0.76** (0.04)	−0.53** (0.04)
Gender	−0.01 (0.09)	0.02 (0.09)	−0.68** (0.14)	−0.01 (0.16)	0.06 (0.17)	−0.88** (0.20)
Age	0.06 (0.14)	−0.13 (0.12)	0.01 (0.24)	0.02 (0.17)	−0.21 (0.19)	−0.20 (0.27)
Age <sup>2</sup>	−0.03 (0.02)	0.03 (0.02)	−0.05 (0.04)	−0.04 (0.03)	0.05 (0.03)	−0.03 (0.04)
Education	0.10 (0.07)	0.22* (0.09)	−0.14 (0.12)	0.26* (0.13)	0.42** (0.14)	−0.01 (0.16)
Religiosity	0.06 (0.04)	0.16** (0.06)	−0.14 (0.08)	0.31** (0.08)	0.48** (0.08)	0.01 (0.11)
Masstrich vote	−0.04 (0.10)	0.09 (0.11)	−0.93** (0.15)	−0.20 (0.19)	−0.08 (0.20)	−1.30** (0.21)
Immigration	0.14* (0.06)	0.15* (0.06)	0.97** (0.12)	0.41** (0.10)	0.43** (0.10)	1.52** (0.17)
$\sigma_1$ 0.55* (0.25)	$\sigma_2$ 0.69** (0.19)	$\sigma_3$ −0.22 (0.64)				
$\sigma_4$ −0.35 (0.33)	$\sigma_5$ −0.28 (0.32)					
Log <sub>10</sub> condition		10.02				9.00
−2(LLF−LLR)		1283.48				1211.26
LRT for $\Sigma$		92.06**				−
Percent correct		58.9				59.2

$n = 1564$ .

\*  $p < 0.05$ .

\*\*  $p < 0.01$ .

## 4. Conclusion

The questions surrounding the estimation of multinomial, qualitative choice models is not an esoteric debate with interest limited to a few practitioners in the methods community. Our impression is that there is considerable uncertainty about the appropriate use of qualitative choice models in the study of voter choice and related applications, and recent studies focusing on these models have not clarified matters. We explain and compare the estimation methods, assuage unfounded concerns regarding the IIA problem, and propose disciplinary standards for estimating and reporting the results of complex maximum likelihood procedures such as the MNP.

With this in mind, we reiterate that the appropriate statistical method is one that corresponds to the stochastic process that generates observed data and informs theoretical questions of interest. This means that one must be able to distinguish MNP and MNL as models of data processes. In the study of mass politics, this is not typically possible given standard data, measures of model fit and diagnostics. If one has a sample size of several thousands, then one may profitably use the MNP if it is appropriate and extracts more useful information than MNL. One likely cannot do this with a sample of 1500 observations on voter choice among a few candidates or parties. Furthermore, the primary theoretical motivation that in disciplinary folk wisdom motivates the MNP over the simpler MNL—the IIA problem—is rarely relevant.

Beyond questions of statistical specification and estimation, we believe the purpose of empirical estimation of behavioral choice models is not curve-fitting or correcting disturbance terms, but model-building and hypothesis testing. In this spirit, we must not underestimate the importance of the logit's simplicity. An important advantage of MNL relative to MNP in electoral studies is that the logit offers more intuitive answers to our theoretical questions. Unlike logit, the probit cannot be easily solved by manual calculation or otherwise be easily manipulated. The fruitful interplay between theory and method depends much on these types of interpretive “hands on” exercises. In logit analysis, to determine the value of “ $x$ ” that produces a 50% probability of observing a given behavior, set the MNL equation to 0.50 and solves for “ $x$ .” To make the same determination using the MNP specification, one writes computer code. For many researchers, certainly for us, intuition and ideas come from efficiency and simplicity rather than from writing computer programs.<sup>12</sup>

This is not to say that there are not good reasons to employ MNP estimation under certain circumstances. If one has appropriate data and theoretical or empirical reasons to believe that the MNP is the proper statistical method, then by all means it should be used. If it is the correct estimation method, MNP may elucidate aspects of choice behavior that may be obscured by other methods. Still, in these cases, it is particularly important that researchers pay particular attention to model specification and diagnostics. At a minimum, log-likelihoods that fail to converge cleanly within a reasonable number of iterations, error covariance matrices that are nearly singular and other signs of sample under-identification, mis-specification and similar problems should motivate model re-specification or re-estimation using logit methods. We strongly encourage standard reporting for MNP estimation that includes the hessian condition (or  $\log_{10}$  condition), likelihood ratios for the estimated error covariance and the details of the optimization procedures. Presenting this information, however, is still an exception in political science. We suspect this is because doing so is not always necessary in the more familiar context of least squares or simpler maximum likelihood problems. However, as our methods become more complex, more attention

---

<sup>12</sup> Adams and Merrill III (2002, esp. pp. 279–282 and Appendix 1) provide a nice illustration of this point by taking advantage of the simplicity of the MNL/CL model to derive equilibrium candidate locations in multicandidate elections.

must be paid to the estimation details as they greatly inform the confidence we place in our findings.

## Acknowledgements

We thank Neal Beck and two anonymous reviewers for their helpful comments on this study. All errors are the responsibility of the authors.

## References

- Adams, J., Merrill, S. III, 2000. Spatial models of candidate competition and the 1998 French presidential election: are presidential candidates vote-maximizers? *Journal of Politics* 62, 729–756.
- Adams, J., Merrill, S. III, 2002. Centrifugal incentives in multicandidate elections. *Journal of Theoretical Politics* 14, 275–300.
- Alvarez, R.M., Nagler, J., 1995. Economics, issue and the Perot candidacy: voter choice in the 1992 presidential election. *American Journal of Political Science* 39, 714–744.
- Alvarez, R.M., Nagler, J., 1998. When politics and models collide: estimating models of multiparty competition. *American Journal of Political Science* 42, 55–96.
- Alvarez, R.M., Nagler, J., 2001. Correlated disturbances in discrete choice models: a comparison of multinomial probit and logit models. *Political Analysis*. Forthcoming.
- Alvarez, R.M., Nagler, J., Bowler, S., 2000. Issues, economics, and the dynamics of multiparty elections: the British 1987 general election. *American Political Science Review* 94, 131–149.
- Arrow, K.J., 1951. *Social Choice and Individual Values*, Second ed. Yale University Press, New Haven, CT.
- Bolduc, D., 1992. Generalized autoregressive errors in the multinomial probit model. *Transportation Research B* 26, 155–170.
- Bunch, D.S., 1991. Estimability in the multinomial probit model. *Transportation Research B* 25, 1–12.
- Dow, J.K., 1999. Voter choice in the 1995 French presidential election. *Political Behavior* 21, 305–324.
- Geweke, J., Keane, M., Runkle, D., 1994. Alternative computational approaches to inference in the multinomial probit model. *Review of Economics and Statistics* 76 (4), 609–632.
- Glasgow, G., 2001. Mixed logit models for multiparty elections. *Political Analysis* 9 (1), 116–136.
- Greene, W.H., 1993. *Econometric Analysis*, third ed. Prentice-Hall, Upper Saddle River, NJ.
- Hausman, J.A., McFadden, D., 1984. Specification tests for the multinomial logit model. *Econometrica* 52, 1219–1240.
- Horowitz, J., 1980. The accuracy of the multinomial logit model as an approximation to the multinomial probit model of travel demand. *Transportation Research B* 14, 331–341.
- Horowitz, J., 1991. Reconsidering the multinomial probit model. *Transportation Research B* 25, 433–438.
- Judd, K.L., 1998. *Numerical Methods in Economics*. MIT Press, Cambridge, MA.
- Keane, M.P., 1992. A note on identification in the multinomial probit model. *Journal of Business and Economic Statistics* 10, 193–200.
- King, G., 1989. *Unifying Political Methodology*. Cambridge University Press, New York.
- King, G., Zeng, L., 2001. Logistic regression for rare events data. *Political Analysis* 9, 137–163.
- Lacey, D., Burden, B.C., 1999. The vote stealing and turnout effects of Ross Perot in the 1992 US presidential election. *American Journal of Political Science* 43, 233–255.
- Lewis-Beck, M.S., Skalaban, A., 1992. France. In: Mark, F., Mackie, T., Valen, H. (Eds.), *Electoral Change*. Cambridge University Press, New York, pp. 167–178.
- Lijphart, A., 1984. *Democracies*. Yale University Press, New Haven, CT.
- Long, J.S., 1997. *Regression Models for Categorical and Limited Dependent Variables*. Sage, Thousand Oaks, CA.

- Luce, R.D., 1959. *Individual Choice Behavior: A Theoretical Analysis*. John Wiley, New York.
- Luce, R.D., Raiffa, H., 1957. *Games and Decisions*. John Wiley, New York.
- Maddala, G.S., 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, New York.
- Powers, D.A., Xie, Y., 2000. *Statistical Methods for Categorical Data Analysis*. Academic Press, New York.
- Quinn, K.M., Martin, A.D., Whitford, A.B., 1999. Voter choice in multi-party democracies: a test of competing theories and models. *American Journal of Political Science* 43, 1231–1247.
- Schofield, N., Martin, A.D., Quinn, K.M., Whitford, A.B., 1998. Multiparty electoral competition in the Netherlands and Germany: a model based on the multinomial probit. *Public Choice* 97, 39–76.
- Stokes, D.E., 1963. Spatial models of party competition. *American Political Science Review* 57, 368–377.
- Taagepera, R., Shugart, M.S., 1989. *Seats and Votes*. Yale University Press, New Haven, CT.
- Train, K., 1993. *Qualitative Choice Analysis*. MIT Press, Cambridge, MA.
- Whitten, G., Palmer, H., 1996. Heightening comparativist concerns for model choice: voting behavior in Great Britain and the Netherlands. *American Journal of Political Science* 40, 231–260.