DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

WHEN POLITICS AND MODELS COLLIDE: ESTIMATING MODELS OF MULTI-PARTY ELECTIONS

R. Michael Alvarez

Jonathan Nagler, University of California, Riverside

SOCIAL SCIENCE WORKING PAPER 959

March 1996

When Politics and Models Collide: Estimating Models of Multi-Party Elections*

R. Michael Alvarez　　Jonathan Nagler, University of California, Riverside

Abstract

Theory: The spatial model of elections can better be represented by using conditional logit than by multinomial logit. The spatial model, and random utility models in general, suffer from a failure to adequately consider the substitutability of candidates sharing similar or identical issue positions.

Hypotheses: Multinomial logit is not much better than successive applications of binomial logit. Conditional logit allows for considering more interesting political questions than does multinomial logit. The spatial model may not correspond to voter decision-making in multiple-candidate settings. Multinomial probit allows for a relaxation of the IIA condition and this should improve estimates of the effect of adding or removing parties.

Methods: Comparisons of binomial logit, multinomial logit, conditional logit, and multinomial probit on simulated data and survey data from a three-party election.

Results: Multinomial logit offers almost no benefits over binomial logit. Conditional logit is capable of examining movements by parties, whereas multinomial logit is not. Multinomial probit performs better than conditional logit when considering the effects of altering the set of choices available to voters.

*Earlier versions of this research were presented at the Annual Meetings of the American Political Science Association, Chicago, IL., September 1995 and at the Annual Political Methodology Summer Conference, Indianapolis, July, 1995. We thank John Aldrich, Nathaniel Beck, Mitch Sanders, Jonathan Katz, Simon Jackman, John Jackson, Dean Lacy, Jan Leighley, Will Moore, Mitch Sanders and Guy Whitten for their comments on earlier versions of this research, and Methodology Conference participants for their input. We also thank participants of the Southern California Political Economy Group for their discussion of this research on November 17, 1995 at the University of California-Irvine. Alvarez thanks the John M. Olin Foundation for support of his research. Nagler thanks the NSF for grant SBR-9413939. Comments may be directed to the authors at: DHSS 228-77, California Institute of Technology, Pasadena, Ca 91125, Internet: rma@crunch.caltech.edu; and Department of Political Science, University of California, Riverside, Ca 92521-0118, Internet: nagler@wizard.ucr.edu, respectively.

1 The Theory and the Practice of Issue Voting Models

The spatial model of voting has been a dominant paradigm in the voting literature over the past 25 years (Davis, Hinich and Ordeshook 1970; Downs 1957; Enelow and Hinich 1984), supplanting the 'funnel of causality' (Campbell et al., 1960) which had a brief reign beginning around 1960. The spatial model is scientifically appealing because of its elegance. It is easy to state: a person votes for the party nearest to him or her on the issues. Further, the spatial model is intuitively appealing to those political scientists who believe that politics is about policy — it states succinctly that issues matter.

Most quantitative analyses of elections based on the spatial model have involved two-party elections; however most elections involve more than two candidates or parties. Elections dominated by two parties are the rule in the United States, but not the rest of the world. In this paper we clarify the methodological implications of using the spatial model to understand multi-candidate or multi-party elections, and we seek to correct some common and widespread misconceptions about the discrete choice models which are best suited for studying multi-candidate or multi-party elections. We first demonstrate that two simple econometric techniques commonly used in the literature, binomial logit and multinomial logit, are in fact almost identical. Multinomial logit is a model of only pair-wise comparisons, and the only real difference between the two techniques is that multinomial logit produces more efficient estimates (i.e., all other things being equal, multinomial logit estimates will converge to the true model parameters more quickly than will binomial logit estimates) since it uses more sample information than binomial logit.

However, we then demonstrate that multinomial logit is a very limited technique since it represents a very limited substantive view of politics. The multinomial logit model includes only information about the individual voters, but does not include the issue positions of the parties and the candidates. Since issue positions of parties and candidates are fundamental to both the spatial theory and our intuitions about the political world, multinomial logit is not the most useful discrete choice model.

We argue that in most electoral settings multinomial logit is likely to represent the *wrong* model. We strongly advocate conditional logit as an alternative to multinomial logit for estimating models of elections. Conditional logit is 'conditional on the characteristics of the choices;' thus it explicitly allows for measures of characteristics of the parties. At a minimum, the spatial model *requires* conditional logit since the spatial model is based on positions of voters relative to parties. Thus, if you care about questions of strategy of candidates or parties, you must use conditional logit.[1] The conditional logit estimator is available for use in many econometrics packages, and is no less robust or harder to estimate than multinomial logit.[2] We are simply making the argument that conditional logit is a better match than multinomial logit for common political science theories of elections. Conditional logit is capable of answering questions that multinomial logit estimates cannot. This should be a critical criteria in choosing an estimator.

This raises an important question: what are the methodoligical implications of using the spatial model to specify empirical models of multi-candidate or multi-party elections? Unfortunately, moving from a two-candidate to a multi-candidate setting suggests a problem for both the spatial model and most common econometric choice models, since the spatial model and all commonly used discrete choice models impose the property of independence of irrelevant alternatives (IIA) on individual voters. IIA implies that the ratio of the probability of choosing one party to the probability of choosing a second party is unchanged for individual voters if a third party enters the race. In simple terms, this implies that in a contest between a liberal and a conservative party, the entry of a second conservative party would not alter the relative probability of an individual voter choosing between the two initial parties. However, because the two conservative parties are close together in the issue space, and hence are likely to be viewed as substitutes by voters, our intuition suggests that these relative probabilities will change.

There are at least three reasons to search for a model that does not impose the IIA condition. First, assuming IIA could lead to incorrect estimates of the model parameters. Second, assuming IIA to be true when it is not will be particularly troubling with regards to one of the more interesting questions regarding multi-party elections: what happens when one party is removed from the choice set? If IIA is violated, then the voters who had been choosing a removed alternative may shift their votes in an unanticipated manner. Third, we would like some substantive insight into the choice process used by individuals, imposing the IIA condition on voters implies that we are starting our research with a very restrictive assumption about that process.

Unfortunately, both conditional logit and multinomial logit suffer from the limitation that they assume IIA. The final point we make in this paper is that there are models available (multinomial probit and the general-extreme value models) which do not impose IIA. However, they avoid imposing IIA through the disturbance term rather than the systemic component. While this may not be an entirely satisfactory approach to resolving the IIA problem in models of voting decisions, it is better than ignoring the problem of IIA.

We begin by describing multinomial logit and show that it is the same as successive applications of binomial logit. We then describe conditional logit and illustrate its advantages over multinomial logit by demonstrating that we can measure the impact of changes in party issue-positions on aggregate vote-shares: a task which is impossible with multinomial logit. We then describe why a general class of models will impose IIA. And we provide estimates using multinomial probit, a model which does not impose IIA, showing that IIA is indeed violated in some real world examples.

2 Logit Models

We begin with a discussion of two logit models which have been used widely in economics, but more sparingly in political science. The first model we consider is *multinomial logit*, which is characterized by a systemic component that is a linear function of characteristics of the individual, as opposed to characteristics of the alternative (i.e., of the party). The second model we consider is the *conditional logit* model, which does allow for choice-specific independent variables which measure characteristics of the party.[3] We also discuss the common assumptions these two logit models make about the distribution of the disturbances. We characterize the basic properties of several discrete choice models in Table 1.

[Table 1 Here]

Table 1 shows that multinomial logit is the most restrictive discrete choice model we discuss in this paper; it models the choice probabilities as functions only of characteristics of the individual voter, it does not allow the error terms to be correlated across choices, and it provides few answers to important political questions. Conditional logit, however, is less restrictive since it allows for choice probabilities to be functions of the characteristics of both the individual and the alternatives. Conditional logit does not allow for correlations between the errors. Both the generalized-extreme value model and the multinomial probit model allow for choice-specific right-hand side variables, for the relaxation of the assumption of independent error terms, and can shed light on what might happen were parties to move in the issue space or drop out of elections. The generalized-extreme value model is more restrictive than the multinomial probit model since the former allows for only certain error correlations, while the multinomial probit model allows for a more flexible error correlation specification. In the following sections of the paper we discuss each model in more detail.

2.1 Multinomial Logit

Multinomial logit specifies the i^{th} individual's utility of the j^{th} choice as:

$$U_{ij} = \beta_j X_i \; + \; u_{ij} \tag{1}$$

where X_i is a vector of characteristics of the i^{th} individual. Note that this model estimates a set of coefficients for each choice: β_j is subscripted based on the alternatives. For one of the choices the coefficients are normalized to be zero.

The probabilities of the i^{th} individual choosing the j^{th} alternative are given by:

$$P_{ij} = \frac{e^{\beta_j' X_i}}{1 \; + \; \sum_{l=1}^{m-1} e^{\beta_l' X_i}} \tag{2}$$

and this implies:

$$\frac{P_{ij}}{P_{ik}} = \frac{e^{\beta_j' X_i}}{e^{\beta_k' X_i}} \tag{3}$$

Equation (3) implies that the ratio of the probabilities of choosing alternative j to alternative k for individual i is independent of the probability of choosing the other alternatives. This is the "independence of irrelevant alternatives" property.

The first point we emphasize about multinomial logit is that each set of coefficients are identical to the coefficients of a binomial logit model using only individuals who choose either alternative j or k, and ignoring all other individuals. This is a key part of the multinomial logit model: it is identical to comparing two choices and ignoring all the other choices. Thus if an 'ignored' choice affects the relative probabilities of choosing the two included choices differently, then the model will perform badly. An important corollary to this is that multinomial logit cannot produce richer empirical models of politics than binomial logit, since they are equivalent models. We demonstrate that multinomial logit and binomial logit are identical by three separate techniques below: 1) we describe the econometric intuition for what the two models are actually estimating; 2) we offer binomial logit and multinomial logit estimates from survey data; and 3) we offer a simulation demonstrating that multinomial logit and binomial logit give equivalent results.

2.2 Multinomial Logit is Equivalent to Binomial Logit

To see that multinomial logit and binomial logit are estimating the exact same thing we need to first describe in some detail the underlying assumptions that both models are based on. Multinomial logit and binomial logit both produce estimates of parameters of a random utility model. Random utility models assume that while individuals maximize their expected utility, these utilities are not known to researchers and must be assumed to be random variables. This allows us to assume that utility can be partitioned into an observed or systemic component, and an unobserved or random component. Consider such a model:

$$\begin{aligned} U_{i1} &= \beta_1 X_i + u_{i1} \\ U_{i2} &= \beta_2 X_i + u_{i2} \\ U_{i3} &= \beta_3 X_i + u_{i3} \end{aligned}$$

Here U_{ij} represents the utility to the i^{th} individual for the j^{th} party, and X_i is measuring characteristics of the i^{th} voter. β_j represents a parameter vector determining the contribution of voter characteristics to utility for choice j. In both multinomial logit and binomial logit one parameter is normalized to zero. Say β_3 is normalized to 0. Then if choice 1 is omitted, binomial logit will generate consistent estimates of β_2 using only individuals who choose 2 or 3. Omitting choice 2, binomial logit will generate consistent estimates of β_1.

A common point of confusion is the claim that binomial logit will not produce consistent estimates because it ignores the presence of a third choice. However, if IIA holds (which both binomial logit and multinomial logit posit), then this has no effect: binomial logit will produce consistent estimates of the parameters because the maintained model implies that the presence or absence of the third choice has no impact on the relative probabilities of choosing either of the other two choices. There is a loss of efficiency because some information is discarded. But binomial logit still produces consistent estimates of the true model parameters, a point proven by Hausman and McFadden in their article describing a test for IIA (1984, p. 1222-1223).

Now consider estimating the above model with multinomial logit. Again, we normalize β_3 to 0. Now multinomial logit gives us consistent estimates of the true parameters β_1 and β_2; consistent estimates of *the exact same parameters* we estimated with binomial logit! This is the central point: once IIA is assumed, binomial logit and multinomial logit are producing estimates of the same parameters. Thus using multinomial logit rather than binomial logit does not give estimates of a richer model positing a complex three-party choice process.

Unfortunately, these points are missed by most "sophisticated" analyses in the literature. In one prominent example, Whitten and Palmer (1996) urge empirical researchers examining multi-party elections to use multinomial logit instead of binomial logit, based on their comparison of the two techniques in data taken from British and Dutch elections: "Comparisons of the parameter estimates produced by each procedure and the substantive inferences derived from those estimates demonstrate the superiority of multinomial logit over BNL (binomial logit) as a means of modeling multi-party vote choice" (1996: 236). Whitten and Palmer reach this conclusion by comparing two fundamentally different specifications of the dependent variable of their models. The binomial choice models they estimate examine the probability of voting for one party, relative to the remaining two major parties (the Conservatives in the 1987 UK election relative to Labour and the Alliance). Their multinomial logit models examine the likelihood of voting for one party from a pair of parties (Conservatives vs. Labour, and the Alliance vs. Labour). Had they estimated successive binomial models for Conservative versus Labour and then Alliance versus Labour, no doubt Whitten and Palmer would have noticed the equivilance of the estimated effects of these models. Whitten and Palmer are essentially comparing estimates across models with completely different dependent variables. That the results differ is no surprise. But it should be looked at as a symptom of measurement error on the dependent variable (incorrectly recoding a trichotomous variable as a dichotomous variable), not as evidence that multinomial logit estimates contain information that binomial logit estimates do not.

2.3 Equivalence of Multinomial logit and Binomial Logit — An Example

That multinomial logit and binomial logit simply reproduce coefficients representing the same pairwise comparisons of choices can be shown in actual data from multi-party elections. Here we focus on data taken from the 1987 British general election survey (Heath 1989).[4] The specification of the models we estimate is identical to that used by Alvarez, Bowler and Nagler (1996). The model specification highlights the importance of issues, economic factors, and party politics, and departs in important ways from other models of British elections. We include a series of seven issue placement variables (defense spending, position on the "phillips curve", taxation, nationalization of industries, redistribution of wealth, crime, and welfare) as well as the voter's beliefs about recent changes in national inflation, unemployment and taxation. To control for factors cited in other works on British elections, we include variables which control for region, class, and other demographic effects.

In Table 2 we present estimates of this model specification modified for multinomial logit and binomial logit. The issue positions are operationalized as individual-specific: they are the voter's stated position on each issue, rather than the distance from the voter to the party. We treat Alliance as the base, or reference, category. Thus we report in the first two columns the estimates for Conservative relative to Alliance, for multinomial logit and binomial logit, respectively; and we report in columns 3 and 4 estimates for Labour relative to Alliance, again for multinomial logit and binomial logit, respectively. The multinomial logit estimates come from full information maximum likelihood utilizing the entire sample; the binomial logit estimates come from two separate binomial logit estimates. The first binomial logit omits voters who chose Labour; the second binomial logit omits voters who chose Conservative. [5]

[Table 2 Here]

Cursory inspection of Table 2 shows that the estimated coefficients for each respective pair-wise comparison are, though not actually identical, statistically indistinguishable across the multinomial logit and binomial logit estimates. Multinomial logit and binomial logit do not produce *identical estimates* of the coefficients in the samples; the multinomial logit estimator is working with more data than either of the two separate sets of binomial logit estimates. But the point is that they produce *consistent* estimates of *the same parameters.* Thus while the multinomial and binomial logit estimates in Table 2 are not identical to each other, occular examination of them is convincing evidence that they are awfully close to each other. As an estimation technique multinomial logit should be preferred to binomial logit because it is more efficient, but this simply means that it will approach the true parameters more quickly than will binomial logit.[6] However, multinomial logit is a model of pairwise comparisons and as such it posits *the same choice process* as binomial logit models do.

2.4 Equivalence of Multinomial logit and Binomial Logit — A Simulation

To further illustrate the equivalence of multinomial logit and succesive applications of binomial logit we offer a simulation. We specified the following spatial model:

$$U_{i1} = -2*(x_i - C1_x)^2 - 3*(y_i - C1_y)^2 + u_{i1} \quad (4)$$
$$U_{i2} = -2*(x_i - C2_x)^2 - 3*(y_i - C2_y)^2 + u_{i2} \quad (5)$$
$$U_{i3} = -2*(x_i - C3_x)^2 - 3*(y_i - C3_y)^2 + u_{i3} \quad (6)$$

where x_i and y_i indicate the i^{th} respondent's ideal point on the X and Y axes, respectively; and Cj_x and Cj_y indicate the position of the j^{th} party on the X and Y axes, respectively. The utility of the i^{th} individual for the j^{th} party is a function of the distance between the party and the individual on the X and Y axes; and the individual has separable preferences. Obviously since this model depends upon the position of the decision-maker *relative* to the position of the parties, the characteristics of the parties (alternatives) are relevant. We placed three parties in a two dimensional issue-space: Party 1 at (-2,0), Party 2 at (0,2), and Party 3 at (2,0). The positions of 5000 voters were drawn from two independent normal distributions, both with mean 0 and variance 2, determining their placement on the x and y axes. Consistent with both multinomial logit and binomial logit models, the disturbances are independent and identically distributed with type I extreme value distributions.

We first estimated a model with only the respondent's characteristics on the right-hand side. This is our naive model, which we estimated with both binomial logit and multinomial logit (later we estimate a correct systemic specification using conditional logit). The parameter estimates for multinomial logit and binomial logit are reported in Table 3. Columns 1 and 2 report the probability of choosing choice 2 relative to choice 1 for multinomial logit and binomial logit respectively. Columns 3 and 4 report estimates for the probability of choosing choice 3 relative to choice 1, again for multinomial logit and binomial logit respectively. The estimates are almost identical because they are estimates of the same parameters. As our sample size increases, the estimates would become indistinguishable. Again, this illustrates that multinomial logit cannot tell us more about politics than a simple set of successive binomial logit estimates, because the multinomial logit estimates are estimates of the exact same phenomena. But notice that both sets of parameters specify the wrong model. To estimate the right model, which would include the distance from the respondents to the parties on the issues, we need conditional logit.

[Table 3 Here]

2.5 Conditional Logit

The conditional logit model (i.e., conditional on the choices) is a fundamentally different model than multinomial logit or binomial logit. Conditional logit employs the same maximum likelihood estimation technique as multinomial logit, but allows for an individual's utility of an alternative to be based upon the characteristics of the alternative. Thus the i^{th} individual's utility of the j^{th} alternative will be given by:

$$U_{ij} = \beta X_{ij} + u_{ij} \tag{7}$$

where X_{ij} indicates a variable measuring the characteristics of alternative j relative to individual i. Multinomial logit did not include characteristics of the alternative on the right-hand side. The characteristics of the alternative are subscripted with respect to the individual because these characteristics could vary across individuals: such as the ideological distance between the party and the respondent. The model can be extended to include individual specific characteristics as multinomial logit does (there is no separate name for the combined model in the literature, so we continue to refer to this model as conditional logit):

$$U_{ij} = \beta X_{ij} + \psi_j a_i + u_{ij} \tag{8}$$

where a_i is a vector of characteristics of the i^{th} individual. Thus this model will yield one coefficient (β) for each alternative-specific variable, and J coefficients (ψ_1, ψ_2, ..., ψ_J) for each individual-specific variable where J is the number of alternatives. However, as with multinomial logit, one of the sets of ψs is normalized (generally to 0, and generally for the first alternative), hence actually $J-1$ sets of ψ's are estimated.

Probabilities will be of the form:

$$P_{ij} = \frac{X_{ij} \ + \ \psi_j a_i}{\sum_{k=1}^{m} e^{\beta} X_{ik} \ + \ \psi_k a_i} \tag{9}$$

Both conditional logit and multinomial logit models assume that the disturbances, u_{ij}, are independent across alternatives.[7]

2.6 Conditional Logit - An Example

To demonstrate what conditional logit estimates look like we again turn to the data from the 1987 British general election, and here we contrast conditional logit estimates to the multinomial logit and pair-wise binomial logit results presented earlier. We present the conditional logit estimates in Table 4. We do not have a great deal to say about these estimates here, except to note two points. First, the specification of the issue distance variables differs from those in Table 2. Here, we specify the issue effects as the distance between the voter and the party on each issue.[8] We estimate the issue distance parameters as choice-specific variables; that is, we estimate only one issue distance parameter for each issue, whereas in Table 2 there were two estimated coefficients for each issue: representing the position of the respondent, not the distance from the respondent to party. Thus conditional logit permits a much better specification of the relationship between issues, parties, and voters than does multinomial logit. Second, we include the other variables representing economic perceptions, class, and demographic status as individual-specific variables. Thus this conditional logit model has both choice-specific and individual-specific coefficients. The first column of individual specific coefficients is for Conservative relative to Alliance, the second column of individual specific coefficients is for Labour relative to Alliance.

[Table 4 Here]

With both the multinomial logit estimates from Table 2 and the conditional logit estimates we could compute tables of first differences – i.e., the effect of a change in the independent variable on the probability of choosing each party – based on the individual's characteristics. These first differences could provide the answers to a set of questions regarding the impact of voters' characteristics on vote-choice. However, an interesting substantive question is not only what is the effect of changes in the characteristics of the voter; but what is the effect of changes in the characteristics of the parties or parties. Conditional logit lets us examine what happens as one, two or three, parties change their positions in the issue space. Multinomial logit does not let us do this.

For instance, a major question regarding British elections is to what extent the extremity of Labour's left-wing positions hurt the party. This cannot be answered by considering differences caused by moving *voters*; rather to answer this we need to see what would happen if Labour moved on the issues (or, to be more precise, if respondents' perceptions of Labour's position changed systematically). To test this we reset Labour's mean perceived issue position to be $\frac{1}{2}$ standard deviation to the left, and $\frac{1}{2}$ standard deviation to the right of its actual mean perceived issue position on each of the seven issues. We then computed the distance Labour would be from each voter at these new positions, and computed the probability of each voter voting for each of the three parties under these two hypothetical scenarios.[9] The difference in predicted aggregate vote-share at the two hypothetical positions for Labour is the impact of a shift on that issue. We report the estimated aggregate vote-shares for each of the three parties with Labour at both hypothetical positions on each issue, as well as the difference, in Table 5. Note again that this is an estimate aggregated across all respondents. The largest impact a one standard-deviation change in Labour's position on any single issue would have is on nationalization of industry: where Labour moving one standard deviation would yield them a 3.1% increase in aggregate vote share. The last row of the table indicates that were Labour to move one standard deviation on all seven issues simultaneously they would increase their vote-share by 6.8%; with 3.6% of that coming at the Conservative party's expense and 3.2% coming at the Alliance's expense. Using this technique we could also determine the optimal placement of Labour on each of the seven issues.[10]

[Table 5 Here]

We present no similar table for multinomial logit because it is impossible to do so. The multinomial logit estimates cannot be used to make any inferences about the effect of moving the parties because the position of the party is not part of the multinomial logit model. This, not the precise magnitude of impacts of Labour's movement on the issues, is what we wish to emphasize with Table 5. The impact of a change by the party on the issues is a major question regarding elections. Yet multinomial logit can supply absolutely no information about this.[11] This is what we feel is the major reason for using conditional logit rather than multinomial logit. We are political scientists. We should analyze politics.

3 Independence of Irrelevant Alternatives (IIA)

While conditional logit is good, it is not perfect. A major characteristic of both multinomial logit and conditional logit is that they impose the "Irrelevance of Independent Alternatives" (IIA) property. As we described earlier, IIA holds when the ratio of the probability of choosing alternative j to the probability of choosing alternative k is not changed if more choices are added to or subtracted from the choice set, or:

$$\frac{P_{ij}|S_s}{P_{ik}|S_s} = \frac{P_{ij}|S_p}{P_{ik}|S_p} \ \forall \ j,k,s,p \tag{10}$$

where S_s and S_p denote sets of alternatives, $j,k \in S_p$, and $j,k \in S_s$, and $P_{ij}|S_s$ denotes the probability of the i^{th} individual choosing alternative j from choice-set S_s. To maintain the IIA condition is troubling when viewed from the perspective of several prominent political science theories of voter decision-making in elections. First, consider a spatial model of voting where individuals vote for the party closest to their ideal point in an issue-space. If we imagine a new party entering an election, our intuition is that the new party would take most of his/her votes from the parties closest to him/her in the issue space. This is not consistent with IIA at the individual level (though IIA may hold at the individual level and *not* preclude such a result at the aggregate level).

In simple terms, IIA implies that in a contest between a liberal and a conservative party, the entry of a second conservative party will not alter the relative probability than an individual voter chooses between the initial two parties. However, because the two conservative parties are close together in the issue space, and hence are likely to be viewed as substitutes by voters, our intuition suggests that the relative probabilities will change.

Consider an extreme case of an election in a single-dimensional space that initially has 2 parties (i.e., $|S_s| = 2$). Say the two parties are a liberal and conservative (parties numbers 1 and 2, respectively). And say voter i is a moderate who is indifferent between the two choices. Then: $P_{i1} = P_{i2} = 0.5$, and $P_{i1}/P_{i2} = 1$. Now add another conservative party to the set, one that is indistinguishable from party 2. The voter might still have probability of .5 of voting liberal and .5 of voting conservative. After all, s/he really still only has two **unique** choices: vote liberal or vote conservative. Choosing between the two identical conservative parties would presumably be done by the flip of a coin. This would yield: $P_{i1} = .5$, $P_{i2} = P_{i3} = .25$; which would mean that $P_{i1}/P_{i2} = 2$. This is a violation of the IIA condition. It is important to bear in mind that the set of probabilities presented here based on the entry of the third party here are derived from a choice process we *assume* for the voter.

Now consider a more complex case. Figure 1 portrays the three parties from our earlier simulations, and a voter at (0,0). Here we have a two-dimensional issue space, with one dimension for economic issues and the other for social issues. Notice that Parties 1, 2, and 3 are viewed as 'equivalent' by the voter: they are each two units away from the voter. In fact in the two-dimensional space it is easy to see that there are potentially an unlimited number of parties that would be viewed by the voter as equivalent: all parties on a circle of radius 2 centered at origin would appear as identical to this voter according to the spatial model. However, politically the three parties depicted clearly represent very distinct choices: Party 1 is moderate on the social issue, but to the right on the economic issue; Party 2 is also socially moderate, but to the left on the economic issue; and Party 3 is to the right on the social issue, but moderate on the economic issue. Now we add Party 4 at (1.95,0), which is viewed by the spatial model as being '.05 different' than Party 1 on the economic issue, and also only '.05 different' than Party 2 vis-a-vis the voter on the economic issue and even only '.05 different' than Party 3 vis-a-vis the voter taking into account both issue-dimensions. Yet most students of politics would say that 9 out of 10 voters would tell us that Parties 1 and 4 are very similar, and that Party 4 is very different from Party 2 or Party 3.

[Figure 1 Here]

Because the spatial model abstracts away the closeness of the parties *to each other* by only measuring the distance from parties *to the voter* the spatial model has no notion of parties as substitutes. The spatial model no more views identical parties to be substitutes for one another than it does for parties who are located at diametrically opposed positions on the issue scale - as long as the parties are equidistant from the voter they are treated the same. The spatial model cannot pick up closeness of parties to each other.

In addition to posing a challenge for spatial models, IIA poses a challenge for retrospective voting models. Retrospective voting models posit individuals' choices to be functions of their evaluations of the incumbent party. Such models are rarely explicit on how multiple alternatives to the incumbent party should be treated; but it seems safe to suppose that the non-incumbent alternatives would be grouped by voters and treated as having some inherent similarity. In fact the choice-process would presumably look exactly like the choice process posited by a nested-logit model: where a voter first chooses between two sets – incumbent party and all other parties – and if the chosen set has more than one choice within it then the voter would then choose from among those choices.

Again, consider the extreme choice. Say the probability of choosing the incumbent party is P_{iI} and the probability of choosing the j^{th} challenger's party is P_{iC_j}. If a voter first chooses between {incumbent party} vs. {non-incumbents party}, and chooses the first set with probability P_{is_1} and the second set with probability P_{is2}, and if all non-incumbents parties are treated equally, then P_{iI} will be independent of the number of challengers and P_{iC_j} is determined by the number of challengers. Obviously the ratio P_{iI}/P_{iC_j} is not independent of the number of alternatives, and IIA is again violated.

The above examples are political science analogues of the classic red-bus/blue-bus problem in econometrics. Imagine an individual who can choose between two modes of transportation: a red-bus or a car. If the individual is indifferent between these modes then the respective probabilities will be .5. Now if a blue-bus is added to the choice set there is no reason to think this alters a person's probability of choosing to travel by car (since the buses differ only in color and on no other relevant dimension), it seems apparent that the probability of choosing a car for most individuals will remain .5. Yet the probability of choosing the blue bus will now be .25, and the probability of choosing the red-bus will be .25 (assuming people are indifferent as to the color of the bus they ride in). Thus the ratio of $P_{car}/P_{blue-bus}$ will change and IIA is violated.

In the spatial model analog to the red-bus/blue-bus problem, we have two parties with (almost) identical positions on the ideological dimension. If a voter chooses on the basis of ideology, and not parties, it is irrelevant how many parties occupy a particular ideological position. Yet models which impose IIA insist that the voters choose among parties, not among issue positions, since parties cannot be substitutes in these models.

Random utility models more generally do not allow for any relationship between the choices in the systemic component of the model, as each systemic component of utility is based on the relationship between a single alternative and the decision maker. So again, two parties equidistant from the voter are treated identically: whether they are diametrically opposed to each other or at identical issue positions. This is where a bad fit between the spatial model and random utility models and our intuition about politics occurs. Currently only random utility models with disturbances correlated across choices allow for the sort of direct comparison of choices which matches our intuition about voter behavior. We believe it is a weakness of random utility models that they do not account for substitutability (or similarity) of alternatives in the systemic component.[12] Development of such models would allow for tests of competing theories of voter decision-making.

In general, models that give the probability of choosing parties as:

$$\frac{P_{ij}}{P_{ik}} = \frac{f(X_{ij}, A_i, Z_j)}{f(X_{ik}, A_i, Z_k)} \tag{11}$$

where:

- X_{ij} = characteristics of the i^{th} voter relative to the j^{th} party
- A_i = characteristics of the i^{th} voter
- Z_j = characteristics of the j^{th} party

will always impose IIA. This is because if we look at this equation, we see that the only things involved in determining P_{ij}/P_{ik} are the characteristics of the voter and the j^{th} and k^{th} parties; neither the existence nor characteristics of any other parties come into play. Thus IIA is guaranteed to hold: there is no way that the inclusion of additional choices could alter P_{ij}/P_{ik}. This is the crux of the problem. To avoid assuming or imposing IIA we must have a model where P_{ij}/P_{ik} incorporates properties of choices other than k and j; and we would prefer that these be incorporated in the systemic component of the model. But allowing for correlated disturbances does not pick this up; it just picks up omitted variables that are not present for *both* identical choices.

4 Models Which Do Not Assume IIA

There are two models of discrete choice which do not assume IIA: the generalized extreme-value model (GEV) and the multinomial probit model (MNP).[13] The GEV model imposes the constraint that the researcher must a priori specify a grouping of choices. The multinomial probit model is more flexible: it imposes no a priori constraint on how respondents view the choices. Multinomial probit allows for both individual specific and alternative-specific variables. The IIA assumption is removed because the error process of the multinomial probit model allow for correlations between the disturbances for the different choices.[14]

In Table 6 we present multinomial probit estimates of a model of the 1987 British election specified exactly as our conditional logit model was. The structure of the coefficients is the same as for conditional logit: we have one set of coefficients for the issue-distance variables, and two sets of coefficients for the individual-level variables. What is different here is the estimate of the error correlations across the disturbances. Two of the estimated error correlations are statistically significant: the correlation between the disturbances for Labour and Alliance is .34, and the correlation between the disturbances for Conservative and Labour is -.39. Thus we are at least 95% confident that IIA is violated. And the grouping of (Labour, Alliance) and non-grouping of (Conservative, Labour) is revealed.

[Table 6 Here]

We see that IIA was violated. So what? As we stated earlier, this suggests that inferences made of a hypothetical two-party race will be particularly suspect. To see the extent of the possible error, we computed predicted vote-shares for Labour and

Conservative in a two-party race with Alliance omitted using both the conditional logit estimates reported in Table 4, and the multinomial probit estimates reported in Table 6. Table 7 gives the estimated vote shares in two and three party races using both models. The conditional logit and multinomial probit estimates of three party vote shares are very close: the Conservative and Labour shares differ by only .3% across the two models. However, the two-party shares differs by 1.7% across the two models. A difference of 1.7% of the vote may not seem very large, but in close elections this is could be the difference between winning and losing.

[Table 7 Here]

This result demonstrates the critical nature of the presence of the Alliance in British politics for the electoral fortunes of Labour. With the Alliance a viable force in British politics in 1987, Labour is clearly disadvantaged (Alvarez, Bowler and Nagler 1996). Additionally, this shows the way in which the multinomial probit model can help answer important political questions when the IIA condition is not met in electoral situations.

4.1 IIA Does Not Aggregate

However, we want to be very clear on one potential point of confusion. All statements about IIA refer to relative probabilities of **individual** voters choosing parties. It is possible for IIA to hold at the individual level, and for the aggregate claim that "a second conservative party will 'take' voters from an existing conservative party" to be true. This may appear to be a paradox, but it is really just a matter of arithmetic. We illustrate this with a simple example. Consider a spatial model where voters choose between parties based on the positions of the parties on one issue. The i^{th} voter's utility for the j^{th} party is simply given by:

$$U_{ij} = -(x_i - C_j)^2 + \epsilon_{ij} \tag{12}$$

where x_i is the i^{th} voter's position, C_j is the j^{th} party's position, and ϵ_{ij} is a random disturbance term with an extreme value distribution. Assume we initially have two parties: Liberal (L) at -1 and Conservative (C) at 1. Now assume we have a 5-person electorate with voters at: -2, -.5, 0, .25, and .75. Table 8 gives the probability they will vote for each of the two parties in a two-party race in columns 2 and 3. The estimate of the aggregate vote-share for each party would be computed by taking the mean of the probabilities over all five voters. This gives relative vote-shares of .46 and .54, for the Conservatives and the Liberals respectively. Now add a third party: the Right-Moderates (M) at .5. The probabilities of each voter choosing any of the three parties in a three-way race are given in columns 4, 5, and 6 of Table 6. If one looks at the ratio of $P_{iL}|\{L,C\}/P_{iC}|\{L,C\}$ and compares it to $P_{iL}|\{L,C,M\}/P_{iC}|\{L,C,M\}$ for any respondent, they are equal. However, if one looks at the ratio of the *means* of P_{iL} to P_{iC} across the two different hypothetical elections they are different. In the first race the Conservatives are predicted to have 46% of the two-way vote. In the three-way race the Conservatives would only have 37% (i.e., .23 / (.23 + .40)) of the two-way vote between the Conservatives and Liberals. Thus, consistent with our intuition, the entry of the

Right-Moderates – a second right-party – takes more votes from the Conservatives than from the Liberals.

[Table 8 Here]

5 Conclusions

We have demonstrated three points in this paper. First, multinomial logit is no magic estimator compared to binomial logit. It offers efficiency gains; but it is estimating *precisely* the same parameters as is binomial logit. Thus any claims that multinomial logit embodies an individual choice process any more complex than two-party comparisons are false.

Second, if one is interested in more strategic questions about politics, such as what would happen if parties or candidates moved in the issue space, and what would be the effect of additional parties entering the race (i.e., questions that seem to come to mind every Presidential Primary season in the United States), then multinomial logit is the wrong model to use and researchers should utilize conditional logit. Multinomial logit simply ignores what is interesting in elections. Conditional logit utilizes the vital information of where parties are located in the issue space, and therefore is a better technique for multi-party and multi-candidate elections than multinomial logit.

Third, binomial logit, multinomial logit, and conditional logit are all quite limited in that they impose the IIA restriction upon voters. Since conditional logit is representing the classic spatial model quite faithfully, this identifies a limitation in the spatial model. The failure of the spatial model to consider the closeness of the parties to each other, as well as to the voter, may present problems in multi-candidate elections. There are estimation techniques that allow for 'grouping' of similar choices and thus remove the IIA restriction: both GEV and multinomial probit allow for this.[15] And we have shown that utilizing multinomial probit allows for more accurate predictions in real elections of the impact of removal of a third party. But both multinomial probit and GEV allow for this grouping in the stochastic (random) component of the model. We think that if we have some theoretical reason to believe parties are grouped by voters, then it is important to try to model that grouping in the systemic component of the model. We believe that this will be an important task in the future for better understanding voting situations where voters have many choices.

6 Appendix A: Computing Unconditional Probabilities from Binomial logit Probabilities

Estimating a model via a series of three binomial logits gives one the ability to estimate three sets of probabilities:

- $P_{12} = Prob(Y_i = 1 \mid \{1,2\})$
- $P_{21} = Prob(Y_i = 2 \mid \{1,2\})$
- $P_{13} = Prob(Y_i = 1 \mid \{1,3\})$
- $P_{31} = Prob(Y_i = 3 \mid \{1,3\})$
- $P_{23} = Prob(Y_i = 2 \mid \{2,3\})$
- $P_{32} = Prob(Y_i = 3 \mid \{2,3\})$

What we would like to recover are the unconditional probabilities:

- $P_1 = Prob(Y_i = 1 \mid \{1,2,3\})$
- $P_2 = Prob(Y_i = 2 \mid \{1,2,3\})$
- $P_3 = Prob(Y_i = 3 \mid \{1,2,3\})$

We know

$$1 = P_1 + P_2 + P_3 \tag{13}$$

And since IIA holds,

$$\frac{P_1}{P_3} = \frac{P_{13}}{P_{31}} \tag{14}$$

So,

$$P_3 = P_1(\frac{P_{31}}{P_{13}}) \tag{15}$$

Similarly,

$$\frac{P_1}{P_2} = \frac{P_{12}}{P_{21}} \tag{16}$$

And,

$$P_2 = P_1(\frac{P_{21}}{P_{12}}) \tag{17}$$

Now we substitute,

$$
\begin{aligned}
1 &= P_1 + P_1(\frac{P_{21}}{P_{12}}) + P_1(\frac{P_{31}}{P_{13}}) && (18)\\
& && (19)\\
&= P_1(1 + \frac{P_{21}}{P_{12}} + \frac{P_{31}}{P_{13}}) && (20)\\
& && (21)\\
P_1 &= 1/(1 + \frac{P_{21}}{P_{12}} + \frac{P_{31}}{P_{13}}) && (22)
\end{aligned}
$$

Since P_{12}, P_{21}, P_{13}, and P_{31} are all observed from the binomial logit estimates; we can compute the unconditional probability P_1. And similar calculations give the unconditional probabilities P_2 and P_3.

7 Appendix B: Multinomial Logit Gives Reduced Form Parameters of the Conditional Logit Model

The difference between conditional logit and multinomial logit is that conditional logit includes another piece of information: the position of the party. It would appear that since the conditional logit model makes use of more information than the multinomial logit model, it should perform better. It should more accurately mirror the truth (the spatial model); and hence give better predictions of individual behavior. In fact, we demonstrate here that when the true model is the common spatial model based on a quadratic utility function, multinomial logit recovers reduced form estimates of the spatial model. Hence the probabilities estimated with multinomial logit will be identical to the conditional logit probabilities. And any estimates of effects of changes in X_i from multinomial logit will be identical to such estimates from conditional logit. But, again, one cannot know a principle point of interest here: the effect of changes in *party* characteristics on voting.

First, we offer an analytical result for two parties showing that MNL recovers reduced form estimates of a common specification of the spatial model.[16] We first define several things:

- X_i = i^{th} Voter's Position in the issue space
- C_j = j^{th} Party's Position in the Issue Space
- D_{ij} = Distance from i^{th} Voter to j^{th} Party

Now according to the classic spatial model; the i^{th} individual's utility of the j^{th} choice is:

$$U_{ij} = -\beta_j * (X_i - C_j)^2 \tag{23}$$

or,

$$U_{ij} = -\beta_j * D_{ij} \tag{24}$$

This fits nicely into the conditional logit random utility model (RUM) setup:

$$U_{ij} = -\beta_j * D_{ij} + u_i \tag{25}$$

Now notice that:

$$\begin{aligned} D_{i1} &= X_i^2 - 2X_iC_1 + C_1^2 \\ D_{i2} &= X_i^2 - 2X_iC_2 + C_2^2 \\ D_{i1} - D_{i2} &= -2X_i(C_1 - C_2) + (C_1^2 - C_2^2) \\ &= -2xa^* + b^* \end{aligned}$$

where

$$\begin{aligned} a^* &= (C_1 - C_2) \\ b^* &= (C_1^2 - C_2^2) \end{aligned}$$

Notice that the difference between the two distances is a linear function of the voter's position. This suggests that if:

$$Pr(Y_i = 1) = f(\alpha_o + \beta_1(D_{i1} - D_{i2})) \tag{26}$$

and α_0 and β_0 are identified then you could substitute $-2X_i a^* + b^*$ for $D_{i1} - D_{i2}$ and recover reduced form estimates:

$$\begin{aligned} Pr(Y_i = 1) &= f(\alpha_o + \beta_1(-2X_i a^*) + \beta_1 b^*) && (27) \\ &= f((\alpha_0 + \beta_1 b^*) + (-2a^*\beta_1)X_i) && (28) \end{aligned}$$

So one could recover:

$$\begin{aligned} \tilde{\alpha} &= \alpha_0 + \beta_1 b^* \\ \tilde{\beta} &= -2a^*\beta_1 \end{aligned}$$

from standard MNL estimates (i.e.; binomial logit in this case).

To explicate these points, we used conditional logit to estimate the simulation model we reported earlier in Table 3. Since the simulation model is a spatial model that does include the position of the party, the conditional logit estimates faithfully correspond to the true model. The conditional logit estimates are reported in Table B1. What is of interest here are predicted probabilities from conditional logit and multinomial logit. We report predicted probabilities for a voter who we move from -2 to 2 along the x-axis in table B2. Notice that the multinomial logit model predicts probability estimates *identical* to the conditional logit model, despite using less information. This suggests that if we were to compute the first-difference to estimate the impact of a change in *respondents'* characteristics on the probability of of voting for any party these would be identical across the conditional logit and multinomial logit estimates. For instance, both conditional logit and multinomial logit predict that a respondent moving from (-2,0) to (-1.5,0) would cause a .07 change in the probability of voting for party 1. However, multinomial logit cannot produce an estimate of the change in predicted probability relative to changes in positions of the parties because the position of the party is not included in the multinomial logit model!

[Table B2 Here]

Notes

1. We use the term 'conditional logit' to encompass models that combine both choice-specific and individual specific variables.
2. Conditional logit estimates presented here were computed with SST. Limdep and Stata are other commonly used packages which allow for estimation of the conditional logit model. Estimates for appropriate syntax in these packages are archived along with the data for this article.
3. We use the following terminology. We refer to a logit model where the dependent variable can take more than two values and the independent variables are individual-specific as *multinomial logit*; it may be referred to elsewhere as *Polychotomous logit*. Our use is consistent with Maddala's (1983) use of this term. This should not be confused with *conditional logit*, as developed by McFadden (1974). Conditional logit is defined as a logit model where the dependent variable takes more than two values and the independent variables are choice-specific. An additional complexity emerges when we consider a logit model which includes *both* individual- and choice-specific independent variables: this we consider a generalization of the conditional logit model, and we will also term these models *conditional logit* models. Later in the paper, we turn to a probit model with both individual- and choice-specific variables. To maintain consistency with most existing literature, we will call this the *multinomial probit model*. Hausman and Wise (1978) use the term conditional probit; but Maddala (1983), Amemiya (1985), and Greene (1993) in subsequent texts have consistently used the term multinomial probit.
4. The British Election Study, 1987, was collected by A. Heath, R. Jowell, J.K. Curtice, and Social and Community Planning Research. The data is distributed by the ESRC Data Archive and the ICPSR.
5. Whitten and Palmer compare multinomial logit estimates to binomial logit estimates not of successive pair-wise comparisons, but rather to binomial logit estimates of 'incumbent' versus 'non-incumbent'. This coding scheme of the dependent variables is really simply an induced measurement error: the resulting comparisons say nothing about multinomial logit vs binomial logit. Of course Whitten and Palmer are absolutely correct in suggesting that researchers avoid recoding a trichotomous variable into a dichotomous variable: but they should not translate that advice into unduly broad claims regarding multinomial logit.
6. However, the standard errors in Table 2 are almost identical across the multinomial logit and binomial logit estimates, suggesting that multinomial logit will have no more statistical power than successive binomial logit estimates.
7. In the models we consider here, the multinomial logit model allows the issue parameters to vary across choices, whereas the issue parameters in the conditional logit models presented here do not. However, the conditional logit model can be

specified with the issue parameters varying across choices.

8. We measure the party's position as the sample average placement of the party by all respondents.

9. This technique is similar to that employed by Wolfinger and Rosenstone (1980), and later by Nagler (1991, 1992) to estimate the impact of an institutional change in voting rules on turnout. In both cases the key is to change a variable of interest and then estimate new predicted probabilities for *each* voters, then aggregate over all voters to measure the total impact of the change.

10. See Alvarez, Bowler, and Nagler (1996) for a demonstration of this.

11. We discuss more fully the technical relationship between multinomial logit and conditional logit estimates in Appendix B.

12. Though the Hausman-Wise formulation allows for the stochastic component to be a function of characteristics of the alternatives - a backdoor way of letting substance in.

13. GEV is equivalent to nested logit when the coefficient of inclusive value is not constrained to be 1.

14. See Alvarez and Nagler (1995), Appendix I for a description of the model.

15. See Alvarez and Nagler (1994) for a comparison of conditional logit, GEV, and multinomial probit estimates.

16. By common specification of the spatial model we mean the specification in which voters are assumed to evaluate their utility for each party through quadratic (squared) issue distances. The simple proof we provide here is dependent upon this particular functional form for voter utilities.

8 References

Aldrich, J. H. and R. M. Alvarez. 1994. "Issues and the Presidential Primary Voter". *Political Behavior* 16, 3: 289-317.

Aldrich, J. H. and F. D. Nelson. 1984. *Linear Probability, Logit, and probit Models.* Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-045. Beverly Hills and London: Sage Publications.

Alvarez, R. M. and J. Nagler. 1995. "Economics, Issues and the Perot Candidacy: Voter Choice in the 1992 Presidential Election." *American Journal of Political Science*, forthcoming in August.

Alvarez, R. M. and J. Nagler. 1994. "Correlated Disturbances in Discrete Choice Models: A Comparison of Multinomial Probit Models and Logit Models." Caltech Social Science Working Paper.

Alvarez, R. M., S. Bowler, and J. Nagler. 1996. "Issues, Economics and the Dynamics of Three-Party Elections: The British 1987 General Election." California Institute of Technology, Social Science Working Paper 949. *http://wizard.ucr.edu/polmeth/working_papers96/alvar96.html*

Amemiya, Takeshi. 1985. *Advanced Econometrics*, Cambridge: Harvard University Press.

Campbell, A., P. Converse, W. Miller and D. Stokes. 1960. *The American Voter.* New York: Wiley and Sons.

Coughlin, P. J. 1992. *Probabilistic Voting Theory.* Cambridge: Cambridge University Press.

Davis, O. A., M. J. Hinich and P. C. Ordeshook. 1970. "An Expository Development of a Mathematical Model of the Electoral Process." *American Political Science Review* 64: 426-448.

Downs, A. 1957. *An Economic Theory of Democracy.* New York: Harper and Row.

Enelow, J. M. and M. J. Hinich. 1984. *The Spatial Theory of Voting.* Cambridge: Cambridge University Press.

Greene, William H. 1993. *Econometric Analysis*, 2nd edition, New York: Macmillan Publishing Company.

Hausman, J. A. and D. McFadden. "Specification Tests for the Multinomial Logit Model", *Econometrica*, Vol 52: 1219-1240.

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

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

Key, V. O., 1966. *The Responsible Electorate.* New York: Vintage.

Kinder, D. R. and Kiewiet D. R. "Sociotropic Politics: The American Case." *American Journal of Political Science* Vol 11, 129-161.

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

Manski, C. F. and S. R. Lerman 1977. "The Estimation of Choice Probabilities from Choice-Based Samples." *Econometrica* Vol 45, 1977-1988.

McFadden, D. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior", in *Economic Theory and Mathematical Economics*, Zarembka, P. (ed), New York: Academic Press.

McFadden, D. 1981. "Econometric Models of Probabilistic Choice", in *Structural Analysis of Discrete Data*, Manski, C.F. and D. McFadden (ed), MIT Press.

Nownes, Anthony J. 1992. "Primaries, General Elections, and Voter Turnout: A Multinomial Logit Model of the Decision to Vote." *American Politics Quarterly* Vol 20:205-226.

Pomper, G. M. 1993. "The Presidential Election," in *The Election of 1992*, Pomper, G. M. (ed), Chatham House.

White, H. "A Heteroskedastic-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* Vol 48, 817-838.

Whitten, G.D. and H.D. Palmer. 1996. "Heightening Comparativists Concern for Model Choice - Voting Behavior in Great Britain and the Netherlands." *American Journal of Political Science* Vol 40, 231-260.

Table 1

Characteristics of Discrete Choice Models

	Multinomial Logit	Conditional Logit	Generalized Extreme Value	Multinomial Probit
Alternative Specific Variables	No	Yes	Yes	Yes
Correlated Disturbances	No	No	Some	Yes
Includes Position of Party	No	Yes	Yes	Yes
Can Correctly Measure Movement by Parties	No	Yes	Yes	Yes
Assumes IIA	Yes	Yes	No	No
Can Correctly Measure Omission of a Party	No	No	Sometimes	Yes

Table 2

Multinomial Logit and Binomial Logit Estimates
British Election - 1987

	Conservative/Alliance		Labour/Alliance	
	MNL	BL	MNL	BL
Intercept	-4.33*	-4.40*	4.55*	5.26*
	(.74)	(.76)	(.81)	(.86)
Defense	.14*	.17*	-.17*	-.19*
	(.03)	(.03)	(.03)	(.03)
Phillips Curve	.08*	.10*	-.03	-.05
	(.02)	(.03)	(.03)	(.03)
Taxation	.13*	.14*	-.06**	-.08*
	(.03)	(.03)	(.03)	(.04)
National.	.16*	.16*	-.16*	-.20*
	(.03)	(.03)	(.03)	(.03)
Redist.	.07*	.06*	-.08*	-.09*
	(.02)	(.02)	(.03)	(.03)
Crime	.08*	.08*	.02	.02
	(.03)	(.03)	(.02)	(.02)
Welfare	.11*	.12*	-.11*	-.10*
	(.02)	(.02)	(.03)	(.03)
South	-.12	-.06	-.41*	-.45*
	(.16)	(.17)	(.21)	(.22)
Midlands	-.26	-.26	-.12	-.15
	(.17)	(.17)	(.21)	(.21)
North	-.03	.03	.66*	.61*
	(.17)	(.18)	(.19)	(.20)
Wales	-.40	-.41	1.41*	1.46*
	(.35)	(.36)	(.31)	(.33)
Scot	-.36	-.42**	.68*	.61*
	(.25)	(.26)	(.25)	(.26)
Union	-.50	-.49*	.37*	.35*
	(.16)	(.16)	(.16)	(.17)
Public Employee	.04	.03	-.05	.03
	(.15	(.15	(.16	(.16
Blue Collar	.09	.14	.70*	.80*
	(.15)	(.16)	(.17)	(.17)
Gender	.29*	.33*	.04	-.03
	(.14)	(.14)	(.15)	(.16)
Age	.03	.03	-.21*	-.24*
	(.05)	(.05)	(.05)	(.05)
Homeowner	.31**	.26	-.55*	-52*
	(.18)	(.18)	(.17)	(.17)
Income	.07*	.07*	-.05	-.07*
	(.03)	(.03)	(.03)	(.03)
Education	-.81*	-.92*	-.54	-.65**
	(.31)	(.31)	(.35)	(.36)
Inflation	.28*	.31*	-.00	.05
	(.10)	(.11)	(.12)	(.12)
Taxes	.02	-.04	-.11	-.15*
	(.06)	(.07)	(.07)	(.07)
Unempl.	.30*	.30	.04	.08
	(.06)	(.06)	(.07)	(.08)
Number of Observations	2131	1494	2131	1172

Table 3

Multinomial Logit and Binomial Logit Estimates of a Simulated Spatial Model

	MNL Prob(Y=2)/ Prob(Y=1)/	BL Prob(Y=2)/ Prob(Y=1)/	MNL Prob(Y=3)/ Prob(Y=1)	BL Prob(Y 3)/ Prob(Y=1)
α_0	-.42	-.43	-.04	-.04
	(.11)	(.12)	(.10)	(.11)
β_x	.98	.94	1.99	2.07
	(.04)	(.05)	(.06)	(.10)
β_y	1.44	1.40	-.01	-.01
	(.05)	(.06)	(.03)	(.03)
Observations	5000			

Table 4
Conditional Logit Estimates
British Election - 1987

	Conservative/Alliance		Labour/Alliance
Defense[a]		-.18*	
		(.02)	
Phillips Curve		-.11*	
		(.02)	
Taxation		-.16*	
		(.02)	
National.		-.18*	
		(.02)	
Redist.		-.08*	
		(.02)	
Crime		-.10*	
		(.05)	
Welfare		-.14*	
		(.02)	
Intercept	.82		2.53*
	(.69)		(.75)
South	-.15		-.44*
	(.17)		(.21)
Midlands	-.29**		.19
	(.17)		(.20)
North	-.06		.64*
	(.18)		(.19
Wales	.48		1.3*
	(.36)		(.31)
Scot	-.41		.69*
	(.25)		(.25)
Union	-.50*		.37*
	(.16)		(.16)
Public Employee	.09		-.02
	(.15)		(.16)
Blue Collar	.11		.70*
	(.15)		(.16)
Gender	.28*		.00
	(.14)		(.15)
Age	.02		-.22*
	(.05)		(.05)
Homeowner	.37*		-.54*
	(.18)		(.16)
Income	.07*		-.06
	(.03)		(.03)
Education	-.82*		-.61**
	(.32)		(.35)
Inflation	.28*		-.03
	(.10)		(.11)
Taxes	.01		-.10
	(.07)		(.07)
Unempl.	.28*		(.01
	(.06)		(.07)
% pred.		70.3	
n		2131	

[a]The seven issues represent distance – absolute value – from the respondent to the mean of the party position.

Table 5
Conditional Logit Estimates of Effect of Movement
By the Labour Party +/- 1/2 Standard Deviation - British Election - 1987

		Conservatives	Labour	Alliance
Baseline		.xx	.xx	.xx
Defense	+ 1/2 σ	45.7	28.3	26.0
	= 1/2 σ	44.7	30.6	24.8
	Difference	-1.0	2.3	-1.3
Phillips	+ 1/2 σ	45.2	29.7	25.2
	- 1/2 σ	45.3	29.0	25.7
	Difference	0.2	-0.7	0.6
Taxation	+ 1/2 σ	45.6	28.6	25.8
	- 1/2 σ	45.1	29.4	25.5
	Difference	-0.5	0.8	-0.3
Nationalization	+ 1/2 σ	45.9	27.7	26.4
	- 1/2 σ	44.6	30.8	24.6
	Difference	-1.3	3.1	-1.8
Redistribution	+ 1/2 σ	45.3	29.2	25.5
	- 1/2 σ	45.2	29.4	25.4
	Difference	-0.1	0.2	-0.0
Crime	+ 1/2 σ	45.6	28.4	26.0
	- 1/2 σ	44.9	29.9	25.1
	Difference	-0.7	1.5	0.8
Welfare	+ 1/2 σ	45.4	29.3	25.4
	- 1/2 σ	45.2	29.3	25.6
	Difference	-0.2	0.0	0.2
All Issues	+ 1/2 σ	47.1	24.9	28.0
	- 1/2 σ	43.5	31.8	24.7
	Difference	-3.6	6.8	-3.2

Note: Estimated impact of the Labour party moving from one-half a standard deviation to the left of its mean perceived position to one-half a standard-deviation to the right of its mean perceived position on each of seven issues. The final row simulates Labour moving simultaneously on all seven issues. Column entries are estimated aggregate vote-shares.

Table 6:
Multinomial Probit Estimates, 1987 British Election - (Alliance Coefficient Normalized to Zero)

Independent Variables	Conservatives		Labour
Defense		-.14*	
		(.01	
Unemployment/Inflation		-.09*	
		(.02	
Taxation		-.13*	
		(.02	
Nationalization		-.14*	
		(.01)	
Redistribution		-.07*	
		(.01)	
Crime		-.08*	
		(.03)	
Welfare		-.11*	
		(.01)	
Constant	.35		1.82*
	(.51)		(.45)
South	-.09		-.29*
	(.07)		(.09)
Midlands	-.23*		-.11
	(.08)		(.09)
North	-.12		.43*
	(.10)		(.11)
Wales	-.48*		.94*
	(.24)		(.18)
Scotland	-.41*		.47*
	(.13)		(.14)
Union Member	-.44*		.26*
	(.07)		(.07)
Public Sector Employee	.08		.01
	(.06)		(.08)
Blue Collar	.02		.46*
	(.09)		(.08)
Female	.21*		-.04
	(.07)		(.07)
Age	.03		-.16*
	(.03)		(.03)
Home Ownership	.36*		-.37*
	(.08)		(.08)
Family Income	.06*		-.05*
	(.02)		(.02)
Education	-.62*		-.45*
	(.21)		(.20)
Inflation	.23*		-.01
	(.07)		(.06)
Unemployment	.23*		-.00
	(.04)		(.04)
Taxes	.02		-.07
	(.04)		(.04)
δ_{CA}		.02	
		(.06)	
δ_{LA}		.34*	
		(.08)	
δ_{CL}		-.39*	
		(.07)	
Number of Obs		2131	
LL		1476.5	

Table 7

Estimated Aggregate Vote Shares: Three Party and Two Party Races

	Conditional Logit	Multinomial Probit
Three Party Race		
Conservative	45.2	44.9
Labour	29.5	29.8
Alliance	25.3	25.3
Two Party Race		
Conservative	59.1	57.4
Labour	40.9	42.6

Column entries are predicted aggregate vote shares by Conditional Logit and Multinomial Probit, for three-party races and two-party races.

Table 8

Vote Shares and Individual Probabilities:
IIA Does Not Aggregate

	Probabilities From Two Cand Race		Probabilities From Three Cand Race		
Voter	$P_C \mid \{C, L\}$	$P_L \mid \{C, L\}$	$P_C \mid \{C, L, M\}$	$P_L \mid \{C, L, M\}$	$P_M \mid \{C, L, M\}$
-2	.00	1.00	.00	.99	.01
-.5	.12	.88	.08	.62	.29
0	.50	.50	.24	.24	.51
.25	.73	.27	.33	.12	.55
.75	.95	.05	.49	.02	.49
Mean	.46	.54	.23	.40	.37

Table B1

Conditional Logit Estimates of a Spatial Model

	CL Prob(Y=2)/ Prob(Y=1)	CL Prob(Y=3)/ Prob(Y=1)/
α_0	-.00	.05
	(.07)	(.08)
β_x	-.25	-.25
	(.01)	(.01)
β_y	-.36	-.36
	(.01)	(.01)
Observations	5000	

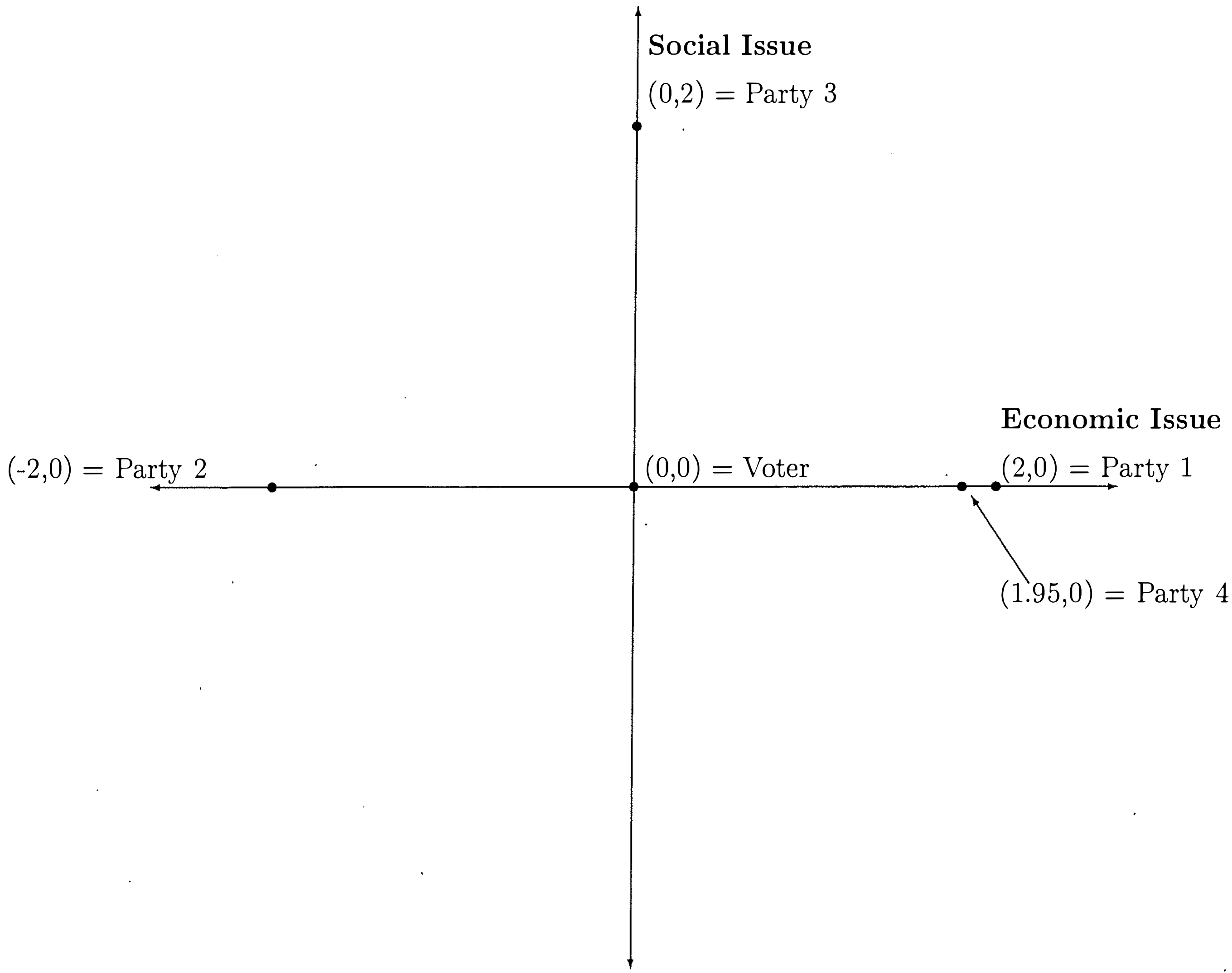
Table B2

Multinomial Logit Estimates and Conditional Logit Estimates of Probabilities for an Individual in the Issue Space

		Probability Party 1:		Probability Party 2:		Probability Party 3	
		$\hat{P}_1(MNL)$	$\hat{P}_1(CL)$	$\hat{P}_2(MNL)$	$\hat{P}_2(CL)$	$\hat{P}_3(MNL)$	$\hat{P}_3(CL)$
x	y						
-2	0	.90	.90	.08	.08	.02	.02
-1.5	0	.83	.83	.12	.12	.04	.04
-1	0	.73	.73	.18	.18	.10	.10
-.5	0	.57	.56	.23	.23	.20	.21
0	0	.38	.38	.25	.25	.37	.38
.5	0	.21	.21	.23	.23	.56	.56
1	0	.10	.10	.18	.18	.72	.72
1.5	0	.04	.04	.12	.12	.83	.83
2	0	.02	.02	.08	.08	.90	.90

Estimated probabilities are from a multinomial logit model including the respondents' position on the X and Y axes as independent variables, and from a conditional logit model estimated with the actual distance between the voter and party.

Figure 1


Social Issue
(0,2) = Party 3
Economic Issue
(-2,0) = Party 2
(0,0) = Voter
(2,0) = Party 1
(1.95,0) = Party 4