*When Politics and Models Collide: Estimating Models of Multiparty Elections**

R. Michael Alvarez, *California Institute of Technology*
Jonathan Nagler, *University of California, Riverside*

Theory: The spatial model of elections can better be represented by using conditional logit models which consider the position of the parties in issue spaces than by multinomial logit models which only consider the position of voters in the issue space. The spatial model, and random utility models in general, suffer from a failure to adequately consider the substitutability of parties sharing similar or identical issue positions.
Hypotheses: Multinomial logit is not necessarily 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 party 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 multiparty elections.
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. Estimation of multinomial probit with more than three choices is feasible.

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.

*This is one of many joint papers by the authors on multiparty elections, the ordering of their names reflects alphabetic convention. 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, Simon Jackman, John Jackson, Jonathan Katz, Gary King, Dean Lacy, Eric Lawrence, 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, and participants in the Second CIC Interactive Video Methods Seminar which was broadcast from the University of Minnesota on October 25, 1996. 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.

American Journal of Political Science, Vol. 42, No. 1, January 1998, Pp. 55–96

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. Furthermore, 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 multicandidate or multiparty elections and we seek to correct some common and widespread misconceptions about the discrete choice models which are best suited for studying multicandidate or multiparty elections. We first demonstrate that two simple econometric techniques commonly used in the literature, binomial logit and multinomial logit, produce almost identical estimates. Multinomial logit is a model of only pairwise 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.

We then demonstrate, however, 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 party characteristics. 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 should at least use conditional logit.[1]

[1]We use the term "conditional logit" to encompass models that combine both choice-specific and individual-specific variables. We use the following terminology throughout the paper. 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

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 theoretical questions that multinomial logit estimates cannot. This should be a critical criteria in choosing an estimator.

This raises an important question: what are the methodological implications of using the spatial model to specify empirical models of multicandidate or multiparty elections? Unfortunately, moving from a two candidate to a multicandidate 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 multiparty 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, but imposing the IIA condition on voters implies

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 call these *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.

[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 and the data for this article can be obtained from the ICPSR article replication archive.

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. Last we provide estimates using multinomial probit, a model which does not impose IIA, showing that IIA is indeed violated in examples taken from recent elections in both the United Kingdom and the Netherlands.

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 allows for choice-specific independent variables measuring characteristics of each party. 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 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 the 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

Table 1. Characteristics of Discrete Choice Models

	Multinomial Logit	Conditional Logit	Generalized Extreme-Value	Multinomial Probit
Alternative Specific Values	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

model is more limited than the multinomial probit model since the former allows only for a more restricted set of error correlations. 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:

$$U_{ij} = \beta_j X_i + u_{ij} \qquad [1]$$

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 from a set of J alternatives are given by:

$$P_{ij} = \frac{e^{\beta_j' X_i}}{1 + \sum_{l=1}^{J-1} e^{\beta_l' X_i}} \qquad [2]$$

and this implies:

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

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 IIA 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 and binomial logit are equivalent using three different approaches: 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 yield 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 which provide the foundation for both models. Multinomial logit and binomial logit both produce estimates of parameters of a random utility model.[3] 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:

$$U_{i1} = \beta_1 X_i + u_{i1}$$
$$U_{i2} = \beta_2 X_i + u_{i2}$$
$$U_{i3} = \beta_3 X_i + u_{i3}$$

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

[3]There are at least three ways in which discrete choice models can be motivated: from a random utility maximization framework, from a direct probabilistic model, or from threshold models (King 1989). In this paper we concentrate on the random utility maximization approach, since we feel that it produces a clear linkage between the theories of individual political choice which we are testing in our work and the statistical models we wish to specify and estimate. Space does not permit us to motivate all of the statistical models we discuss in this paper in each of these three different ways. Instead, we refer the reader to King's excellent discussion of these alternative motivations for binary choice models (1989, 98–115).

both multinomial logit and binomial logit one parameter is normalized to zero. Thus, when estimating binomial logit only one set of coefficients is produced since the other set has been normalized to zero and never appears. In this case, say β_3 is normalized to zero. Then if choice 1 is omitted, binomial logit will generate consistent estimates of β_2 using only individuals who pick choice 2 or choice 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. Of course 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, 1222–23).

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 and multinomial logit produce 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.

That binary logit and multinomial logit are estimating essentially the same coefficients is an existing result in both the statistical and econometric literature; this point was discussed by Amemiya (1976) and demonstrated by Hausman and McFadden (1984).[4] Unfortunately, these points are missed by most analyses in the political science literature. In one recent example, Whitten and Palmer (1996) urge empirical researchers examining multiparty elections to use multinomial logit instead of binomial logit, based on their

[4]The fact that binary logit and multinomial logit consistently estimate the same coefficients was exploited by Hausman and McFadden (1984) in their derivation of their statistical test for the independence of irrelevant alternatives. Under the assumption that the multinomial logit model is correctly specified, consistent estimates of the same subvector of parameters can be obtained from both a multinomial logit model estimated with a full choice set and from another multinomial logit or binary logit model estimated with a restricted choice set (Ben-Akiva and Lerman 1985). Thus, if we have a three choice problem and the model was correctly specified, both multinomial logit on the full choice set and binary logit on each restricted choice set (each pair of choices) would yield consistent estimates of the same model parameters (Hausman and McFadden 1984).

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 multiparty 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 U.K. 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 equivalence 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 multiparty elections. Here we focus on data taken from the 1987 British general election survey (Heath 1989).[5] 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

[5]*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.

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)
Nationalization	.16*	.16*	–.16*	–.20*
	(.03)	(.03)	(.03)	(.03)
Redistribution	.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)
Scotland	–.36	–.42**	.68*	.61*
	(.25)	(.26)	(.25)	(.26)
Union Member	–.50	–.49*	.37*	.35*
	(.16)	(.16)	(.16)	(.17)
Public Sector 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)
Unemployment	.30*	.30	.04	.08
	(.06)	(.06)	(.07)	(.08)
Number of Observations	2131	1494	2131	1172
Log Likelihood	–1500.8	–734.71	–1500.8	–588.27

Standard Errors in parentheses. * indicates significance at 95% level; ** indicates significance at 90% level.

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.[6]

Cursory inspection of Table 2 shows that the estimated coefficients for each respective pairwise 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, ocular 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.[7] 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 successive applications of binomial logit we offer a simulation. We specified the following spatial model:

$$U_{i1} = -2 * (x_i - C1_x)^2 - 3 * (y_i - C1_y)^2 + u_{i1} \qquad [4]$$

[6]Whitten and Palmer (1996) compare multinomial logit estimates to binomial logit estimates not of successive pairwise comparisons, but rather to binomial logit estimates of "incumbent" versus "nonincumbent." 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.

[7]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 in samples of these sizes.

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

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 decisionmaker *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 5,000 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

Table 3. Multinomial Logit and Binomial Logit Estimates of a Simulated Spatial Model

	MNL (Choice 2)/ (Choice 1)	BL (Choice 2)/ (Choice 1)	MNL (Choice 3)/ (Choice 1)	BL (Choice 3)/ (Choice 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				

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.

2.5 Conditional Logit

The conditional logit model (i.e., conditional on the choices) is a fundamentally different model specification than multinomial logit or binomial logit. Conditional logit 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:

$$U_{ij} = \beta X_{ij} + u_{ij} \qquad [7]$$

where X_{ij} indicates a variable measuring the characteristics of alternative j relative to individual i.[8] 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 (such as the ideological distance between the party and the respondent) could vary across individuals. 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):

$$U_{ij} = \beta X_{ij} + \psi_j a_i + u_{ij} \qquad [8]$$

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 ($\psi_1, \psi_2, \ldots, \psi_J$) for each individual-specific variable where J is the number of alternatives.[9] However, as with multinomial logit, one of the sets of ψ's is normalized (generally to 0, and generally for the first alternative), hence actually $J - 1$ sets of ψ's are estimated.

[8]Both multinomial logit and conditional logit assume the disturbances have independent extreme-value distributions, and the likelihood functions in both models can be expressed as functions of e^V, where V represents the systemic component of each model; but the underlying specifications are still fundamentally different from a substantive perspective since they include different information on the right-hand side.

[9]In the models we consider here, the multinomial logit model allows the issue parameters to vary across choices, whereas the issue parameters (β) in the conditional logit models presented here do not. However, the conditional logit model can be specified with the issue parameters varying across choices.

Probabilities are of the form:

$$P_{ij} = \frac{e^{\beta X_{ij} + \psi_j a_i}}{\sum_{k=1}^{J} e^{\beta X_{ik} + \psi_k a_i}} \qquad [9]$$

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

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 pairwise 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 that used in Table 2. Here, we specify the issue measures as the distance between the voter and the party on each issue.[10] Also, we estimate only one issue distance parameter for each issue (the first seven coefficients reported in Table 4), whereas in Table 2 there were two estimated coefficients for each issue (where each represented the position of the respondent, not the distance between the respondent and the party). Thus conditional logit permits a much better specification of the spatial model and 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.

With both the multinomial logit estimates from Table 2 and the conditional logit estimates we could compute tables of first differences—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 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. Conditional logit lets us examine what happens as parties change their positions in the issue space. Multinomial logit does not let us do this.

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

Table 4. Conditional Logit Estimates British Election—1987

	Conservative/Alliance		Labour/Alliance
Defense[a]		−.18*	
		(.02)	
Phillips Curve		−.11*	
		(.02)	
Taxation		−.16*	
		(.02)	
Nationalization		−.18*	
		(.02)	
Redistribution		−.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)
Scotland	−.41		.69*
	(.25)		(.25)
Union Member	−.50*		.37*
	(.16)		(.16)
Public Sector 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)
Unemployment	.28*		.01
	(.06)		(.07)
N		2131	
Log Likelihood		−1477.6	

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

Standard Errors in parentheses. * indicates significance at 95% level; ** indicates significance at 90% level.

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 one half standard deviation to the left, and one half 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.[11] 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 these are estimates aggregated across all respondents. The largest impact that 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.[12]

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.[13] This, not the precise magnitude of impacts of Labour's movement on the issues, is what we wish to

[11]This technique is similar to that employed by Wolfinger and Rosenstone (1980), and later by Nagler (1991, 1994) 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* voter, then aggregate over all voters to measure the total impact of the change.

[12]See Alvarez, Bowler, and Nagler (1996) for a demonstration of this approach to determining the optimal positions of parties on issues.

[13]The superiority of the conditional logit model over multinomial logit is not a new notion in the literature, either. This point has been made in a number of different places (c.f. Agresti 1990; Ben-Akiva and Lerman 1985; Hoffman and Duncan 1988; McFadden 1973). However, the differences between the conditional logit and multinomial logit model are not widely appreciated in political science. Our intention here is not to break new technical ground in the comparison of multinomial logit and conditional logit, but rather to make political scientists more cognizant of the implications of choosing between the two techniques.

Table 5. Conditional Logit Estimates of Effect of Movement by the Labour Party $\pm\frac{1}{2}$ Standard Deviation—British Election—1987

		Conservatives	Labour	Alliance
Baseline		45.2	29.5	25.3
Defense	$-\frac{1}{2}\sigma$	45.7	28.3	26.0
	$+\frac{1}{2}\sigma$	44.7	30.6	24.8
	Difference	–1.0	2.3	–1.3
Phillips	$-\frac{1}{2}\sigma$	45.2	29.7	25.2
	$+\frac{1}{2}\sigma$	45.3	29.0	25.7
	Difference	0.2	–0.7	0.6
Taxation	$-\frac{1}{2}\sigma$	45.6	28.6	25.8
	$+\frac{1}{2}\sigma$	45.1	29.4	25.5
	Difference	–0.5	0.8	–0.3
Nationalization	$-\frac{1}{2}\sigma$	45.9	27.7	26.4
	$+\frac{1}{2}\sigma$	44.6	30.8	24.6
	Difference	–1.3	3.1	–1.8
Redistribution	$-\frac{1}{2}\sigma$	45.3	29.2	25.5
	$+\frac{1}{2}\sigma$	45.2	29.4	25.4
	Difference	–0.1	0.2	–0.0
Crime	$-\frac{1}{2}\sigma$	45.6	28.4	26.0
	$+\frac{1}{2}\sigma$	44.9	29.9	25.1
	Difference	–0.7	1.5	0.8
Welfare	$-\frac{1}{2}\sigma$	45.4	29.3	25.4
	$+\frac{1}{2}\sigma$	45.2	29.3	25.6
	Difference	–0.2	0.0	0.2
All Issues	$-\frac{1}{2}\sigma$	47.1	24.9	28.0
	$+\frac{1}{2}\sigma$	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.

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.[14] 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:

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

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.[15] 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 its votes from the parties closest to it 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 that 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 two parties (i.e., $|S_s| = 2$). Say the two parties are a liberal

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

[15]It is very important to note here that the IIA property is an assumption we make about the behavior of individuals. It is not an aggregate property, since it is possible for IIA to hold for individuals but to be violated in the aggregate. We discuss this point fully in Appendix C.

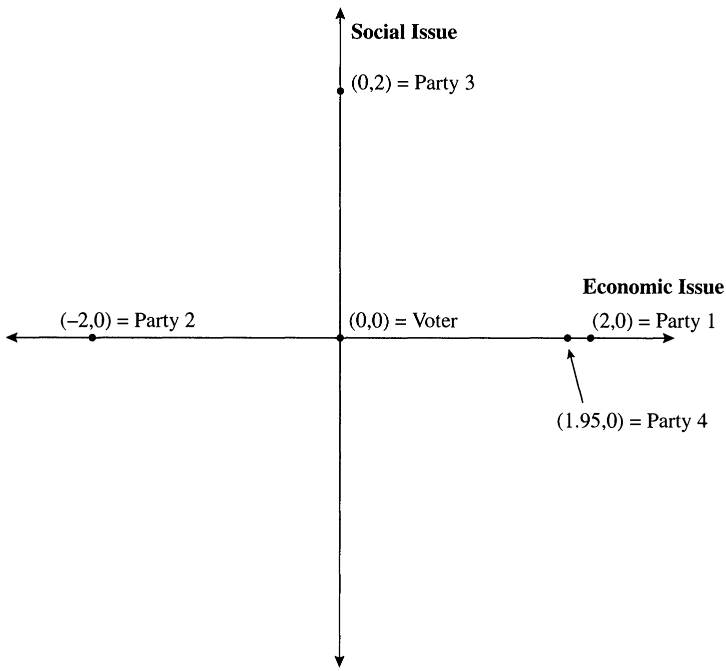
and a conservative (Parties numbers 1 and 2, respectively). And say voter *i* is a moderate who is indifferent between the two choices. Then: $P_{i1} = P_{i2} = 0.5$, and $P_{i1}/P_{i2} = 1$. Now add another conservative party to the set, one that is indistinguishable from Party 2. The voter might still have probability of .5 of voting liberal and .5 of voting conservative. After all, he or she 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 the 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-à-vis the voter on the economic issue and even only ".05 different" than Party 3 vis-à-vis the voter taking into account both issue dimensions. Yet most students of politics would say that nine 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.

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 parties which 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

Figure 1.



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 nonincumbent 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. {nonincumbents party}, chooses the first set with probability P_{is1} and the second set with probability P_{is2}, and if all nonincumbents 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 $P_{car}/P_{blue\ \mathrm{bus}}$ 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. This is a very strong belief about the importance of party as a label for candidates.

We want to emphasize that IIA is fundamentally an assumption *about how individuals make decisions*: it is not an assumption about what is on the right-hand side of a model nor the distribution of the disturbances. Notice that in our discussion of IIA in this section we have only talked about IIA in terms of relative probabilities of individual choice, and that we have introduced no particular model specification. This should reinforce the idea that IIA is about how individuals choose, and that it is not an assumption we make about how choice models are specified.

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

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

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.

Random utility models 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 decisionmaker. 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, random utility models, and our intuitions 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.

This leads to another source of confusion in the practical matter of econometric model specification. A necessary assumption for an econometric model to be consistent with the IIA axiom is that the disturbance terms for the individual's utility for each choice be uncorrelated. Thus, one of the consequences of the failure of the IIA assumption to hold for individual voters is that these disturbances terms will be correlated. If we use an econometric model which assumes that IIA holds, we will end up with inconsistent estimates.

However, while voters failing to abide by the IIA restrictions imply that the disturbance terms in a random utility model will be correlated, the corollary is not true—observing correlated disturbances when estimating a random utility model does not necessarily imply that the voters are violating the IIA restriction. Such correlations can be observed because of omitted variables. Consider a simple error in model specification. Assume that there are two candidates who share a common position on a single issue, for example they are both pro-choice. Assume the third candidate is pro-life. Now if neither the candidates' nor the voters' positions on abortion are included in the model, then the systemic component of the model will underestimate the tendency of pro-choice voters to prefer both of the first two candidates. This could result in estimating positively correlated disturbances for the two pro-choice candidates. However, the voters may be obeying the IIA restrictions. In this case, correcting the model specification by including abortion will lead to uncorrelated disturbances. Should the voters fail to obey the IIA condition, correlated errors will still be observed even with abortion included in the model. This brings us back to the point we made above: IIA is an assumption about voter behavior, but it is not an econometric assumption about the distribution of disturbance terms. Our substantive assumptions about behavior come first; then we adopt a set of econometric assumptions consistent with our behavioral assumptions.[16]

What the above example demonstrates is that *if* voters do not obey the IIA constraint, then kitchen sink solutions of adding right-hand side variables to the utility functions will not solve the problem and lead to consistent estimates. What would be much more satisfactory would be a theory which detailed the process which caused correlations among the disturbances and allowed us to parameterize and estimate those correlations.[17] Development of such models would allow for tests of competing theories of voter decision-making.[18]

4. Models Which Do Not Assume IIA

4.1 Multinomial Probit and British Politics

There are two models of discrete choice which do not assume IIA: the generalized extreme-value model (GEV) and the multinomial probit model (MNP).[19] 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.[20] The IIA assumption is removed because the error

[16]This is the fundamental point of King (1989): we should develop substantive models, then write down likelihood functions that accurately represent our assumptions of the real world.

[17]The Hausman-Wise formulation of MNP does allow for the stochastic component to be a function of characteristics of the alternatives—a backdoor way of letting substance in.

[18]An interesting development in this direction can be seen in the work on "quantal response equilibria" by McKelvey and Palfrey (1995a, 1995b). In this class of game theoretic models, players are assumed to make choices much in the same way econometricians assume people make decisions using discrete choice models. In the quantal response model, players have utilities defined over outcomes which are a function of conditional expected utilities for each action which are known to all players and a random component which is known only by each player. By assuming that the random component has a certain distribution (usually that it is type I extreme value), McKelvey and Palfrey are able to derive choice probabilities which follow the standard multinomial logit distribution, and which under certain additional assumptions produce predictions about the relative likelihoods about which strategies players will use. It is possible that future work on quantal response can relax the IIA assumption (which has been in the work to date) and could use utility functions which incorporate the aspects of political choice which are absent from the common specifications of the spatial model.

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

[20]The multinomial probit model we specify here follows the framework originally proposed by Hausman and Wise (1978), which has been widely discussed in the econometric literature (Bolduc 1992; Bunch 1991; Daganzo 1979; Dansie 1985; Keane 1992) and which has been seeing notice in the political science literature (Alvarez and Nagler 1995, 1997). Many of the recent papers on the

process of the multinomial probit model allows for correlations between the disturbances for the different choices.[21]

In Table 6 we present multinomial probit estimates of a model of the 1987 British election where the systemic component of the model is specified exactly as was our conditional logit model. 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 that we estimate the error correlations across the disturbances. And the two 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.[22] Thus we are at least 99% confident that IIA is violated.

We see that IIA was violated. What does this imply? 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

multinomial probit model have focused on identification and estimability of this model (Borsch-Supan 1989; Borsch-Supan and Hajivassiliou 1993; Geweke 1987; Keane 1994; Lerman and Manski 1981; McFadden 1989; Pakes and Pollard 1989). Following this literature, we estimate only two of the six possible parameters in the error covariance matrix; the standard result in the literature is that at most there are only $[K(K-1)/2]-1$ free parameters in the error covariance matrix after allowing for scaling (where K is the number of choices) (Bolduc 1992; Bunch 1991). In Alvarez and Nagler (1995) we estimated three correlations between disturbances in an MNP model. Such a specification is not identified (though the earlier work estimated two of the three correlations to be almost zero [–.07 and –.08] the results remain the same when only two correlations are estimated), and we thank Eric Lawrence for his many discussions with us about the identification of that particular specification. We avoid the "fragile identification" problem inherent in multinomial probit models by restricting the β parameters to be equal across parties. Additionally, the generalized extreme-value model has received widespread attention in the econometric literature since McFadden's seminal contributions (1974, 1978, 1981) [see Brownstone and Small (1989) for a review of the early literature]. For an extensive comparison of the multinomial probit and general extreme-value models with political science applications see Alvarez and Nagler (1997).

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

[22]There are three possible disturbance covariance matrices which can be considered if two correlations between disturbances are estimated. We estimated all three. One such matrix (with the correlation between the disturbance of Labour and Alliance constrained to be zero) is rejected since the estimated matrix was not positive definite. The other two produce results that are statistically indistinguishable and substantively similar. Since assuming that the disturbances between Labour and Conservative are uncorrelated seems to be the less plausible assumption, we report results from the model with the correlation between Conservative and Alliance constrained to be zero.

Table 6. Multinomial Probit Estimates, 1987 British Election

Independent Variable	Conservatives/Alliance		Labour/Alliance
Defense		–.14*	
		(.02)	
Unemployment/Inflation		–.09*	
		(.02)	
Taxation		–.13*	
		(.02)	
Nationalization		–.14*	
		(.01)	
Redistribution		–.07*	
		(.01)	
Crime		–.08*	
		(.03)	
Welfare		–.11*	
		(.02)	
Constant	.35		1.84*
	(.73)		(.63)
South	–.09		–.30*
	(.13)		(.14)
Midlands	–.23**		–.11
	(.14)		(.14)
North	–.12		.43*
	(.15)		(.13)
Wales	–.49**		.95*
	(.27)		(.22)
Scotland	–.42*		.48*
	(.20)		(.17)
Union Member	–.45*		.26*
	(.14)		(.11)
Public Sector Employee	.08		.01
	(.12)		(.11)
Blue Collar	.02		.47*
	(.12)		(.12)
Female	.21**		–.04
	(.11)		(.10)
Age	.03		–.16*
	(.04)		(.04)
Home Ownership	.37*		–.37*
	(.14)		(.12)
Family Income	.06*		–.05*
	(.02)		(.02)
Education	–.63**		–.46
	(.36)		(.31)
Inflation	.23*		–.01
	(.08)		(.08)
Unemployment	.23*		–.00
	(.05)		(.05)
Taxes	.02		–.07
	(.05)		(.04)
δ_{LA}		.32*	
		(.12)	
δ_{CL}		–.43**	
		(.24)	
Number of Obs		2131	
LL		–1476.5	

Standard Errors in parentheses. * indicates significance at 95% level; ** indicates significance at 90% level.

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	25.2
Two-Party Race		
Conservative	59.1	57.5
Labour	40.9	42.5

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

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 could be the difference between winning and losing.

This result demonstrates the critical nature of the presence of a center party 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.

4.2 Simulated Multinomial Probit and Multiparty Politics

The multinomial probit model is the preferred discrete choice model for examining voter choice in multiparty and multicandidate elections. But, with the exception of relatively simple applications where there is a limited choice set (like the 1987 British election in the previous section, or the 1992 American presidential election, both situations with three choices facing voters), the multinomial probit model has not seen much use in the political science literature.[23] This is due to the fact that estimation of even three choice multinomial probit models had, until recently, required the numerical integration of bivariate normal distributions, which is computationally difficult even

[23]Exceptions include the work of Alvarez and Nagler (1995) and Alvarez (1997) on U.S. presidential elections, Alvarez, Bowler, and Nagler (1996) on British politics, Dow (1997) on French presidential elections, Lawrence (1997) on U.S. presidential primary elections, Quinn, Martin, and Whitford (1996) on Dutch elections, and Schofield et al. (1997) on Dutch and German elections.

using the relative fast computers available to many political scientists. The computational problem had been so severe that most authors state that estimation of the multinomial probit model with more than four choices is impossible using direct numerical integration of multivariate normal densities in maximum-likelihood models. This has led some to assert that the multinomial probit model is of limited use in political science (e.g., Whitten and Palmer 1996).

But recent advances in estimation techniques (as opposed to simply the expected advances in brute force computing power) have brought significant reductions in the computational cost of estimating multinomial probit models. One advance has been in the use of the method of simulated moments (McFadden 1989; Pakes and Pollard 1989); others have advocated Gibbs sampling approaches using the innovative concept of data augmentation (Albert and Chib 1993; McCulloch and Rossi 1994). The third advance has come in the use of simulated maximum-likelihood to estimate multinomial probit models, especially with the development of the Geweke-Hajivassiliou-Keane (GHK) probability simulator for multivariate normal densities (Geweke 1991; Geweke, Keane, and Runkle 1994; Hajivassiliou and McFadden 1990; Keane 1994).

Here we discuss only the use of the GHK probability simulator for the estimation of multinomial probit models, since this approach is an extension of the maximum-likelihood estimation of multinomial probit models using direct numerical integration of multivariate normal densities. In fact, we still will use the maximum-likelihood approach for estimation of the multinomial probit model; however, instead of numerically evaluating multivariate normal densities we will recursively simulate these densities. In Appendix D we present a more technical description of the GHK probability simulation approach.

We begin with a simple, but compelling, replication of the multinomial probit model we presented in the previous section. In Table 8 we present multinomial probit estimates generated using the GHK simulation method of the same model of the 1987 British general election we estimated in the previous section using traditional maximum likelihood and numerical evaluation of bivariate integrals. Comparing the results in Tables 6 and 8 shows that the simulated multinomial probit estimates are generally almost identical to those presented in Table 6. In the few cases where estimated values are not identical between the two sets of estimates, they are statistically indistinguishable. The estimated standard errors are also almost identical across the two sets of estimates.

Next, we turn to an application which demonstrates the ability of the simulated maximum-likelihood to estimate multinomial probit models in a situation where voters can choose from five political parties—the 1994 parliamentary elections in the Netherlands. Elections in the Netherlands were

Table 8. Simulated Multinomial Probit Estimates, 1987 British Election

Independent Variables	Conservatives/Alliance		Labour/Alliance
Defense		–.14*	
		(.02)	
Unemployment/Inflation		–.09*	
		(.02)	
Taxation		–.13*	
		(.02)	
Nationalization		–.14*	
		(.01)	
Redistribution		–.07*	
		(.01)	
Crime		–.08*	
		(.04)	
Welfare		–.11*	
		(.02)	
Constant	.40		1.85*
	(.61)		(.49)
South	–.09		–.29*
	(.13)		(.14)
Midlands	–.22**		–.10
	(.13)		(.14)
North	–.10		.44*
	(.14)		(.14)
Wales	–.47**		.95*
	(.27)		(.22)
Scotland	–.39*		.49*
	(.20)		(.17)
Union Member	–.44*		.25*
	(.13)		(.11)
Public Sector Employee	.09		.01
	(.12)		(.11)
Blue Collar	.03		.48*
	(.12)		(.12)
Female	.20**		–.04
	(.11)		(.10)
Age	.03		–.16*
	(.04)		(.03)
Home Ownership	.35*		–.37*
	(.14)		(.12)
Family Income	.06*		–.04*
	(.02)		(.02)
Education	–.63*		–.47
	(.28)		(.23)
Inflation	.23*		–.00
	(.08)		(.08)
Unemployment	.23*		.00
	(.05)		(.05)
Taxes	.01		–.08*
	(.05)		(.04)
δ_{LA}		.30*	
		(.12)	
δ_{CL}		–.34**	
		(.21)	
Number of Obs		2131	
LL		–1486.2	

Standard Errors in parentheses. * indicates significance at 95% level; ** indicates significance at 90% level.

long thought to be quite predictable, given that voting seemed to revolve primarily around social and religious cleavages in the electorate (Daalder 1966; Lijphart 1968). However, many scholars have begun to reexamine electoral politics in the Netherlands given the sudden rise in electoral volatility in recent decades (e.g., Middendorp and Tanke 1990; Van Der Eijk and Niemoller 1987; Whitten and Palmer 1996). Unfortunately few of the existing studies on electoral politics in the Netherlands have utilized models which do not assume that IIA holds for voters.[24]

Furthermore, there are a number of important questions which need to be answered about political change in the Netherlands. Most immediate is determining what has produced the dramatic increase in electoral volatility seen in the Netherlands since the mid-1960s (Bartolini and Mair 1990). Many scholars attribute this to the breakdown of "consociationalism" (Van Der Eijk and Niemoeller 1983). But what is fueling this breakdown? What factors are driving voter choice in contemporary Dutch politics? While some have argued that ideology is now determining voter choice (Van Der Eijk and Niemoller 1987), others have asserted that retrospective economic voting is the key to understanding recent elections in the Netherlands (Middendorp and Tanke 1990), and others have found the explanation somewhere in between (Whitten and Palmer 1996).

We use the 1994 Dutch Parliamentary Election Study for our analysis. We are able to develop a set of independent variables which would allow for close examination of the factors which determined voting in this election (ideological positioning of the parties, views on materialist and post-materialist issues, retrospective economic views, as well as religious and social status). The survey data were rich enough to allow us to explore voting for five of the parties which received the greatest vote shares in the 1994 election: Christian Democratic Appeal (CDA, 22.2%), Labor Party (PvdA, 24.0%), Liberal Party (VVD, 19.9%), Democrats' 66 (D66, 15.5%), and Green Left (GL, 3.5%).[25]

[24]Quinn, Martin, and Whitford (1996) and Schofield et al. (1997) provide extensive analyses of the 1979 Dutch election using a different estimation technique than we utilize; their work employs the Gibbs sampling for estimating multinomial probit models (Albert and Chib 1993; McCulloch and Rossi 1994).

[25]The variables we used in our model of the 1994 Dutch election were taken from the *Dutch Parliamentary Election Study (DPES) 1994,* overseen by H. Ankers and E. V. Oppenhuis; this data is available from the ICPSR. The ideology variable we employ is coded as the absolute distance between the respondent and the mean ideological position of each party, with the latter estimated from the survey sample. We use variables measuring materialist and post-materialist values; each of these variables are factor scales, where positive values indicate strong materialist or post-materialist values, constructed from a two dimensional principal components analysis of responses to 17 questions included in the DPES (variables v497–v513). To measure economic perceptions, we use three variables, each of which is coded so that the high category expresses favorable responses about the

This means that we estimate a five choice multinomial probit model. Since only a limited number of parameters can be estimated in the error covariance matrix, we estimated a number of different specifications involving four parameters of the error covariance matrix (the estimated coefficients in the systemic part of our model were not influenced by which specification of the error covariance matrix we estimated). In Table 9 we report the results from one of these specifications (the coefficients for the Green Left party are normalized to zero).[26]

First, we have evidence that IIA is violated in these data, since here we have one estimated error correlation which is statistically significant ($\delta_{PVDA,CDA}$) while another verges on being statistically significant ($\delta_{CDA,VVD}$). This implies that inferences made from models which assume IIA will be suspect, and instead we should rely on inferences from less restrictive models of political behavior like that presented here.

But additionally, in Table 9 we see a number of important estimated effects. For example, we find that both the ideological positioning of the parties and retrospective economic evaluations were important in determining which party voters supported in this election. Both of these results clearly support previous work on recent Dutch elections which has found that ideology and economic evaluations are strong determinants of voter choice in the Netherlands (Middendorp and Tanke 1990; Van Der Eijk and Niemoller 1987; Whitten and Palmer 1996). Yet we also find that attitudes about materialist and post-materialist values were important determinants of voting, thus confirming the continued importance of these cleavages in the politics of advanced democratic nations (Inglehart 1977, 1990). Last, we estimate significant social effects, mainly for the roles of age, gender, and income—but we do not estimate much of a significant impact for religious affiliations. These results may shed light on the breakdown of old social cleavages in Dutch politics (e.g., Van Der Eijk and Niemoeller 1983). Clearly these questions need further examination, both in the context of Dutch politics and in

economy, employment, or the respondent's personal finances. We include three indicator variables for religious affiliations; atheists are the reference category. We control for the effects of age, education, gender, income, and whether the respondent lives in a urban area of the Netherlands. We also use indicator variables for whether the voter is a manual worker or a union member.

[26]This model took 15 hours to converge on a multiuser Hewlett-Packard Model 735/125 workstation with 228 megs of memory, using gaussx 3.5 and gauss 3.2.30. While truly reliable speed comparisons across different computers are impossible without running identical programs, existing benchmarks suggest that newer computers would be significantly faster. According to Hewlett-Packard's data, newer H-P workstation models are over two times as fast at floating point calculations as the 735/125. Existing benchmark comparisons indicate that newer Pentium-based computers now available may also enable researchers to run similar models faster (benchmarks can be found at **http://www.unifrankfurt.de/~stst/gaussst.html.**

Table 9. Simulated Multinomial Probit Estimates, 1994 Dutch Election

Independent Variables		PVDA/GL	CDA/GL	VVD/GL	D66/GL
Ideology	–.37*				
	(.06)				
Constant		.67	–1.1	1.0	2.6**
		(1.2)	(.76)	(1.3)	(1.3)
Materialism		–2.4*	–2.2*	–2.5*	–1.5
		(.96)	(1.1)	(1.0)	(.98)
Postmaterialism		.58*	1.1*	1.1*	.54*
		(.19)	(.20)	(.19)	(.18)
Economy		.51*	.72*	.28	.16
		(.22)	(.23)	(.25)	(.23)
Employment		.46*	.34*	.21	.27**
		(.16)	(.16)	(.16)	(.15)
Personal Finances		–.10	–.13	–.40**	–.52**
		(.20)	(.22)	(.21)	(.20)
Catholic		–.31	.86*	–.07	–.24
		(.29)	(.32)	(.86)	(.28)
Reform		.23	–1.1	.30	.02
		(.23)	(.76)	(.24)	(.28)
Calvinist		–.81	1.6**	–.07	–.05
		(.88)	(.82)	(.86)	(.81)
Age		1.50*	1.8*	1.0*	–.30
		(.41)	(.42)	(.40)	(.44)
Education		–.15	–.05	–.03	–.18
		(.17)	(.17)	(.17)	(.17)
Gender		–.40*	–.55*	–.41*	–.29
		(.18)	(.20)	(.18)	(.19)
Income		.40	.86*	1.1*	.66*
		(.26)	(.28)	(.28)	(.26)
Urban		.10	.16	.03	.08
		(.10)	(.10)	(.11)	(.10)
Manual workers		–.21	–.73*	–.38	–.29
		(.29)	(.33)	(.29)	(.31)
Union members		.25	–.01	–.19	–.00
		(.27)	(.30)	(.29)	(.28)
$\delta_{CDA,VVD}$	.47				
	(.31)				
$\delta_{PVDA,CDA}$	.54*				
	(.18)				
$\delta_{PVDA,VVD}$	.29				
	(.26)				
$\delta_{PVDA,D66}$	.007				
	(.29)				
Number of Obs	901				
LL	–931.0				

Standard Errors in parentheses. * indicates significance at 95% level; ** indicates significance at 90% level.

contrast to political change in other advanced democratic nations. Our purpose here is to demonstrate that the technology exists to examine these questions with flexible and appropriate models.

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 computing estimates of *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 multiparty and multicandidate 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 multicandidate elections. There are econometric models that allow for "grouping" of similar choices and thus remove the IIA restriction: in particular multinomial probit allows for this and we have shown that multinomial probit allows for more accurate predictions in real elections of the impact of removal of a third party.[27]

But these econometric models only relax the IIA assumption through the specification of the stochastic (random) component of the model. We think that if we have some theoretical reason to believe voters do not obey the IIA axiom, then it is important to correctly specify how voters perceive their choices in the systemic component of our models. For example, if a voter faced with three candidates—two conservative candidates and one liberal candidate—views the two conservative candidates as identical and makes a choice between two ideologies, not three candidates, then that is

[27]See Alvarez and Nagler (1997) for a comparison of conditional logit, GEV, and multinomial probit estimates.

what we need to model. We believe that this will be an important task in the future for better understanding voting situations where voters have many choices.

Manuscript submitted 19 March 1996.
Final manuscript received 10 March 1997.

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

$$1 = P_1 + P_2 + P_3 \qquad [12]$$

And since we assume that IIA holds,

$$\frac{P_1}{P_3} = \frac{P_{13}}{P_{31}} \qquad [13]$$

So,

$$P_3 = P_1\left(\frac{P_{31}}{P_{13}}\right) \qquad [14]$$

Similarly,

$$\frac{P_1}{P_2} = \frac{P_{12}}{P_{21}} \qquad [15]$$

And,

$$P_2 = P_1\left(\frac{P_{21}}{P_{12}}\right) \qquad [16]$$

Now we substitute,

$$1 = P_1 + P_1\left(\frac{P_{21}}{P_{12}}\right) + P_1\left(\frac{P_{31}}{P_{13}}\right) \quad [17]$$

[18]

$$= P_1\left(1 + \frac{P_{21}}{P_{12}} + \frac{P_{31}}{P_{13}}\right) \quad [19]$$

[20]

$$P_1 = 1 / \left(1 + \frac{P_{21}}{P_{12}} + \frac{P_{31}}{P_{13}}\right) \quad [21]$$

Since P_{12}, P_{21}, 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.

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.[28] 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

[28]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.

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

$$U_{ij} = -\beta_j * (X_i - C_j)^2 \qquad [22]$$

or,

$$U_{ij} = -\beta_j * D_{ij} \qquad [23]$$

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

$$U_{ij} = -\beta_j * D_{ij} + u_i \qquad [24]$$

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:

$$Pr(Y_I = 1) = f(\alpha_o + \beta_1(D_{i1} - D_{i2})) \qquad [25]$$

and α_0 and β_0 are identified then we can substitute $-2X_ia^* = b^*$ for $D_{i1} - D_{i2}$ and recover reduced form estimates:

$$\begin{aligned} Pr(Y_i = 1) &= f(\alpha_o + \beta_1(-2X_ia^*) + \beta_1 b^*) && [26] \\ &= f(\alpha_0 + \beta_1 b^*) + (-2a^*\beta_1 X_1) && [27] \end{aligned}$$

So one could recover:

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

from standard MNL estimates (i.e., binomial logit in this case). Again, note that this result holds assuming quadratic utility functions.

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

Table B1. Conditional Logit Estimates of a Spatial Model

	(Party 2)/ (Party 1)	(Party 3)/ (Party 1)
α_0	−.00	.05
	(.07)	(.08)
β_x	−.25	−.25
	(.01)	(.01)
β_y	−.36	−.36
	(.01)	(.01)
Observations	5000	

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 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,

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.

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!

APPENDIX C
IIA Does Not Aggregate

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:

$$U_{ij} = -(x_i - C_j)^2 + \varepsilon_{ij} \qquad [28]$$

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 five person electorate with voters at: –2, –.5, 0, .25, and .75. Table B3 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

Table B3. Voter's Shares and Individual Probabilities: IIA Does Not Aggregate

Voters Position	Probabilities From Two Cand Race $P_C\|\{C,L\}$	$P_L\|\{C,L\}$	Probabilities From Three Cand Race $P_C\|\{C,L,M\}$	$P_L\|\{C,L,M\}$	$P_M\|\{C,L,M\}$
–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

$P_{iL}|\{L,C\}/P_{iC}|\{L,C\}$ 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.

APPENDIX D
Simulation Methods for Estimating Multinomial Probit Models

A number of different methods for simulating probabilities for the multinomial probit model have been advanced: the frequency simulator (Lerman and Manski 1981), a kernel-smoothed frequency simulator (McFadden 1989), the method of simulated moments (McFadden 1989; Pakes and Pollard 1989), and simulated maximum likelihood (Albright, Lerman, and Manski 1977; Lerman and Manski 1981). However, these probability simulation techniques have been shown to be generally inferior to the recursive probability simulator (known in the literature as the GHK simulator) developed in a series of more recent papers (Geweke 1991; Geweke, Keane, and Runkle 1994; Hajivassiliou and McFadden 1990; Hajivassiliou, McFadden, and Rudd 1992; Keane 1994). Accordingly we used the GHK simulator to estimate the models presented in Tables 8 and 9, and in this appendix we describe how this recursive probability simulator works. Similar discussions of the GHK simulator appear in Geweke, Keane, and Runkle (1994) and Hajivassiliou, McFadden, and Rudd (1992).

We begin by defining a random utility function for voter i over each party j, where $j = 1, \ldots, J$:

$$U_{ij} = x_{ij}\beta + a_i\psi_j + \varepsilon_{ij} \qquad [29]$$

where x_{ij} is a vector of characteristics of party j relative to voter i, a_i is a vector of characteristics unique to the individual i, and ε is a random disturbance. This is identical to the setup described in the body of the paper for conditional logit and multinomial probit. The disturbance terms are assumed to have a multivariate normal distribution, with mean zero and covariance matrix, Σ.

If we stack the model by parties, Equation 29 can be rewritten as:

$$U_i = X_i\beta + A_i\psi + \varepsilon_i \qquad [30]$$

where $U_i = (U_{i1}, \ldots, U_{iJ})'$. Finally, instead of working directly with this model, we utilize a differenced model where all utilities are normalized relative to the last choice:

$$U_i^* = X_i^*\beta^* + A_i^*\psi^* + \varepsilon_i^* \qquad [31]$$

where U_i^* is a $J \times 1$ vector with $U_i^* = \dfrac{(U_{ij} - U_{iJ})}{\left(\sigma_j^2 - 2\sigma_{jJ} + \sigma_J^2\right)^{\frac{1}{2}}}$ and ε_i^* has a multivariate normal distribution with mean zero and covariance matrix Σ^*, and $U_{iJ}^* = 0$.

From here, the probabilities of choosing each alternative for the individual voter are easy to specify (see Alvarez and Nagler 1995 for derivation of these probabilities for the three choice case). However, in this form the multinomial probit model requires the numerical solution of $J - 1$ dimensional integrals to evaluate these choice probabilities; thus in the relatively simple three choice example, two dimensional integrals must be evaluated. Despite the fact that Hausman and Wise (1978) demonstrated that some additional transformations can reduce the dimensionality of the integrals to $J - 2$, the computational costs of evaluating high dimensional multivariate integrals have kept the multinomial probit model from being widely used in the literature.

This is exactly the problem solved by the various probability simulators. The probability simulator of interest to us here, the GHK simulator, simply simulates the choice probabilities rather than explicitly evaluating the integrals. The insight which drives the GHK probability simulator is that the choice probabilities of the multinomial probit model can be translated into a series of conditional probabilities which can be simulated recursively (Geweke 1991; Hajivassiliou and McFadden 1990; Keane 1992, 1994). Again, GHK is not a substitute for traditional maximum likelihood; rather GHK simply refers to a technique for computing the probabilities required to produce the maximum likelihood estimates. We show here how the GHK probability simulator works, by showing the details for computing the probability that the voter picks the j^{th} party, which is based on the random utility function from Equation 31.

Now we define some notation for the derivation of the probability simulator (we will suppress the subscript i for ease of exposition). First, we write:

$$\begin{aligned} \hat{U}_k^j &= U_k^* - U_j^* \\ \hat{\varepsilon}_k^j &= \varepsilon_k^* - \varepsilon_j^* \end{aligned} \qquad [32]$$

for $k = 1, \ldots, J$. Recall that the J^{th} $U_{iJ}^* = 0$, which implies $\varepsilon_J^* = 0$. Now it is useful to define two sets of $(J - 1) \times 1$ vectors:

$$\begin{aligned} \hat{U}^j &= (\hat{U}_1^j, \ldots, \hat{U}_{j-1}^j, \hat{U}_{j+1}^j, \ldots, \hat{U}_J^j)' \\ \hat{\varepsilon}^j &= \hat{\varepsilon}_1^j, \ldots, \hat{\varepsilon}_{j-1}^j, \hat{\varepsilon}_{j+1}^j, \ldots, \hat{\varepsilon}_J^j)' \end{aligned} \qquad [33]$$

For the voter to choose choice j, we require that all elements of $\hat{U}^j$ be less than or equal to zero, since party j is the choice of the voter $U_k^* - U_j^* \leq 0 \,\forall\, k$.

The distribution of $\hat{\varepsilon}^j$ will be IIDN(0, $\hat{\Sigma}^j$). Now if $\hat{A}^j$ is the lower triangular Cholesky factorization of $\hat{\Sigma}^j$, and we define $n = (\nu_1, \ldots, \nu_{j-1}, \nu_{j+1}, \ldots, \nu_J)'$ where n is a $(J - 1) \times 1$ vector of independent standard normal random variables, then $\hat{\varepsilon}^j$ can be expressed as $\hat{A}^j \nu$. Next, define $\hat{U}_k^j$ as a function of the ns, $\hat{U}_k^j(\nu_1^l, \ldots, \nu_p^l)$, when the random variables n_1 through n_p are fixed at a draw $(\nu_1^l, \ldots, \nu_p^l)$ for $p \leq k$. Notice that when $p = k$ this is a value, but when $p < k$ it is a random variable that depends upon $\hat{\Sigma}^j$. The l superscript is used to indicate that we repeat these draws of ν_j^l for M trials. Thus if we have five choices, and we are looking at the probability of choosing choice 3, and we want to do 50 trials, we draw four independent normal variables (n_1, n_2, n_4, n_5), 50 times.

Now GHK proceeds by drawing v, which is easy to do since they are independent and normally distributed. The v are drawn until the following criteria are satisfied:

$$
\begin{array}{ll}
(1) & \text{Draw } \nu_1^l \text{ such that } \hat{U}_1^1(\nu_1^l) < 0. \\
\vdots & \\
(j-1) & \text{Draw } \nu_{j-1}^l \text{ such that } \hat{U}_{j-1}^1(\nu_1^l, \ldots, \nu_{j+1}^l) < 0. \\
(j) & \text{Skip } \nu_j. \\
(j+1) & \text{Draw } \nu_{j+1}^l \text{ such that } \hat{U}_{j+1}^1(\nu_1^l, \ldots, \nu_{j-1}^l) < 0. \\
\vdots & \\
(J-1) & \text{Draw } \nu_{J-1}^l \text{ such that } \hat{U}_{j-1}^1(\nu_1^l, \ldots, \nu_{j+1}^l, \ldots, \nu_{J-1}^1) < 0.
\end{array}
\tag{34}
$$

Using these draws, we can now write down the probability that the voter chooses the j^{th} party:

$$
\begin{aligned}
\hat{P}(j|\beta^*, \psi^*, \Sigma^*, X^*, A*) = \frac{1}{M}\sum_{l=1}^{M} (P(\hat{U}_1^l < 0) \times \prod_{k=2}^{j-1} P[\hat{U}_k^l(\nu_1^l, \ldots, \nu_{k-1}^l) < 0] \\
\times P[\hat{U}_{j+1}^l(v_1^l, \ldots, v_{j-1}^l) < 0] \\
\times \prod_{k=j+2}^{J} P[\hat{U}_k^1(v_1^l, \ldots, v_{j-1}^l, v_{j+1}^l, \ldots, v_{k-1}^l) < 0]).
\end{aligned}
\tag{35}
$$

This is a straightforward calculation and enables us to estimate the MNP model for any reasonable number of choices. The GHK simulator has two important properties. First, it is unbiased and is smooth in the model parameters (Geweke, Keane, and Runkle 1994). But second, as is seen in the derivation we have just presented, it requires the evaluation of only univariate integrals and draws from univariate truncated normal distributions. This second property is where the GHK simulator obtains the computational advantages over other probability simulators and certainly over the direct numerical evaluation of multivariate integrals.

REFERENCES

Agresti, Alan. 1990. *Categorical Data Analysis.* New York: Wiley.

Albert, James, and Siddhartha Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88:669–679.

Albright, R. Steven Lerman, and Charles Manski. 1977. *Report on the Development of an Estimation Program for the Multinomial Probit Model,* Report prepared by Cambridge Systematics for the Federal Highway Administration.

Alvarez, R. Michael. 1997. *Information and Elections.* Ann Arbor: University of Michigan Press.
Alvarez, R. Michael, Shaun Bowler, and Jonathan Nagler. 1996. "Issues, Economics and the Dynamics of Multi-party Elections: The British 1987 General Election." Pasadena, CA: California Institute of Technology Social Science Working Paper 949.
Alvarez, R. Michael, and Jonathan Nagler. 1995. "Economics, Issues and the Perot Candidacy: Voter Choice in the 1992 Presidential Election." *American Journal of Political Science* 39:714–44.
Alvarez, R. Michael, and Jonathan Nagler. 1997. "Correlated Disturbances in Discrete Choice Models: A Comparison of Multinomial Probit Models and Logit Models." *Political Analysis.* Forthcoming.
Amemiya, Takeshi. 1976. "The Maximum Likelihood, the Minimum Chi-Square and the Nonlinear Weighted Least-Squares Estimator in the General Qualitative Response Model." *Journal of the American Statistical Association* 71:347–51.
Amemiya, Takeshi. 1985. *Advanced Econometrics.* Cambridge: Harvard University Press.
Bartolini, Stephan, and Peter Mair. 1990. *Identity, Competition, and Electoral Availability.* New York: Cambridge University Press.
Ben-Akiva, Moshe, and Steven R. Lerman. 1985. *Discrete Choice Analysis.* Cambridge: MIT Press.
Bolduc, Denis. 1992. "Generalized Autoregressive Errors in the Multinomial Probit Model." *Transportation Research B* 26:155–70.
Borsch-Supan, Axel. 1989. "Recent Developments in Flexible Discrete Choice Models: Nested Logit Analysis Versus Simulated Moments Probit Analysis." In *Behavioral Modelling of Spatial Decisions and Processes,* ed. M. M. Fischer et al. Amsterdam: North-Holland.
Borsch-Supan, Axel, and Vassilis A. Hajivassiliou. 1993. "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models." *Journal of Econometrics* 58:347–68.
Brownstone, David, and Kenneth A. Small. 1989. "Efficient Estimation of Nested Logit Models." *Journal of Business and Economic Statistics* 7:67–74.
Bunch, D. S. 1991. "Estimability in the Multinomial Probit Model." *Transportation Research B* 25:1–12.
Campbell, Angus, Phillip Converse, Warren Miller, and Donald Stokes. 1960. *The American Voter.* New York: Wiley and Sons.
Daalder, Hans. 1966. "The Netherlands, Opposition in a Segmented Society." In *Political Oppositions in Western Democracies,* ed. R. A. Dahl. New Haven: Yale University Press.
Daganzo, Carlos. 1979. *Multinomial Probit.* New York: Academic Press.
Dansie, B. 1985. "Parameter Estimability in the Multinomial Probit Model." *Transportation Research B* 19:526–8.
Davis, Otto, Mel Hinich, and Peter Ordeshook. 1970. "An Expository Development of a Mathematical Model of the Electoral Process." *American Political Science Review* 64:426–48.
Dow, Jay K. 1997. "Voter Choice and Strategies in French Presidential Elections." Presented at the annual meetings of the Midwest Political Science Association, Chicago.
Downs, Anthony. 1957. *An Economic Theory of Democracy.* New York: Harper and Row.
Enelow, James, and Mel Hinich. 1984. *The Spatial Theory of Voting.* New York: Cambridge University Press.
Geweke, John. 1987. "Bayesian Inference in Econometric Models Using Monte-Carlo Integration." *Econometrica* 57:1317–40.
Geweke, John. 1991. "Efficient Simulation from the Multivariate Normal and Student-*t* distributions Subject to Linear Constraints." *Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface.* Alexandria, VA: American Statistical Association, 571–8.
Geweke, John, Michael Keane, and David Runkle. 1994. "Alternative Computational Approaches to Inference in the Multinomial Probit Model." *Review of Economics and Statistics* 76:609–32.
Greene, William H. 1993. *Econometric Analysis.* 2nd ed. New York: Macmillan Publishing Company.

Hausman, Jerry, and Daniel McFadden. 1984. "Specification Tests for the Multinomial Logit Model." *Econometrica* 52:1219–40.

Hausman, Jerry A., and David A. Wise. 1978. "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." *Econometrica* 46: 403–26.

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

Hajivassiliou, Vassilis, Daniel McFadden, and Paul Rudd. 1992. "Simulation of Multivariate Normal Orthant Probabilities: Methods and Programs." Cowles Foundation Discussion Paper 1021, Yale University.

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

Hoffman, Saul D., and Greg J. Duncan. 1988. "Multinomial and Conditional Logit Discrete Choice Models in Demography." *Demography* 25:415–27.

Inglehart, Ronald. 1977. *The Silent Revolution: Changing Values and Political Styles Among Western Publics.* Princeton: Princeton University Press.

Inglehart, Ronald. 1990. *Culture Shift in Advanced Industrial Society.* Princeton: Princeton University Press.

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

Keane, Michael P. 1992. "A Note on Identification in the Multinomial Probit Model." *Journal of Business and Economic Statistics* 10:193–200.

Keane, Michael P. 1994. "A Computationally Practical Simulation Estimator for Panel Data." *Econometrica* 62:95–116.

King, Gary. 1989. *Unifying Political Methodology.* New York: Cambridge University Press.

Lawrence, Eric D. 1997. "A Simulated Maximum Likelihood Application to the 1988 Democratic Primary." Presented at the annual meetings of the Midwest Political Science Association, Chicago.

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

Lijphart, Arend. 1968. *The Politics of Accommodation: Pluralism and Democracy in the Netherlands.* Berkeley: University of California Press.

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

McCulloch, Robert, and Peter Rossi. 1994. "An Exact Likelihood Analysis of the Multinomial Probit Model." *Journal of Econometrics* 64:207–40.

McFadden, Daniel. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics,* ed. P. Zarembka. New York: Academic Press.

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

McFadden, Daniel. 1978. "Modelling the Choice of Residential Location." In *Spatial Interaction Theory and Residential Location,* ed. A. Karlquist et al. Amsterdam: North-Holland.

McFadden, Daniel. 1981. "Econometric Models of Probabilistic Choice." In *Structural Analysis of Discrete Data,* ed. C. F. Manski and D. McFadden. Cambridge: MIT Press.

McFadden, Daniel. 1989. "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration." *Econometrica* 57:995–1026.

McKelvey, Richard D., and Thomas R. Palfrey. 1995a. "Quantal Response Equilibria in Normal Form Games." *Games and Economic Behavior* 10:6–38.

McKelvey, Richard D., and Thomas R. Palfrey. 1995b. "Quantal Response Equilibria for Extensive Form Games." Pasadena, CA: California Institute of Technology Working Paper 947.

Middendorp, C. P., and P. R. K. Tanke. 1990. "Economic Voting in the Netherlands." *European Journal of Political Research* 18:535–55.

Nagler, Jonathan. 1991. "The Effects of Registration Laws and Education on Voter Turnout." *American Political Science Review* 85:1393–405.

Nagler, Jonathan. 1994. "Scobit: An Alternative Estimator to Logit and Probit." *American Journal of Political Science* 38:230–55.

Pakes, Ariel, and David Pollard. 1989. "Simulation and the Asymptotics of Optimization Estimators." *Econometrica* 57:1027–58.

Quinn, Kevin M., Andrew D. Martin, and Andrew B. Whitford. 1996. "Explaining Voter Choice in a Multi-Party Democracy: A Look at Data from the Netherlands." Presented at the annual meetings of the American Political Science Association, San Francisco.

Schofield, Norman, Andrew D. Martin, Kevin M. Quinn, and Andrew B. Whitford. 1997. "Multiparty Electoral Competition in the Netherlands and Germany: A Model Based on Multinomial Probit." *Public Choice.* Forthcoming.

Van Der Eijk, C., and B. Niemoller. 1983. *Electoral Change in the Netherlands.* Amsterdam: CT Press.

Van Der Eijk, C., and K. Niemoller. 1987. "Electoral Alignments in the Netherlands." *Electoral Studies* 6:17–30.

Whitten, Guy, and Harvey Palmer. 1996. "Heightening Comparativists Concern for Model Choice—Voting Behavior in Great Britain and the Netherlands." *American Journal of Political Science* 40:231–60.

Wolfinger, Raymond E., and Steven J. Rosenstone. 1980. *Who Votes?* New Haven: Yale University Press.