DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

HETEROGENEITY IN MODELS OF ELECTORAL CHOICE

Douglas Rivers

SOCIAL SCIENCE WORKING PAPER 505

December 1983

ABSTRACT

Heterogeneity or the presence of a variety of decision rules in a population has usually been ignored in voting research. A method for handling heterogeneous preferences using rank order data is developed and applied to a simple issue-voting model. The estimated average effect of partisanship is substantially higher when the assumption of homogeneity is relaxed, though many self-identified partisans also use ideological criteria to evaluate candidates and many independents rely on partisan criteria.

HETEROGENEITY IN MODELS OF ELECTORAL CHOICE*

Douglas Rivers

California Institute of Technology

1. INTRODUCTION

Understanding why voters prefer one candidate over another has been a fundamental concern of voting research. Various models of individual decision making have been proposed and tested using aggregate and survey data on public opinion and voting behavior. Most of these tests have involved predicting vote choice, candidate preference, or satisfaction with incumbent performance on the basis of voter characteristics (demographic variables, policy preferences, party identification, etc.) or candidate characteristics (party, issue positions, performance variables) or some combination of these. The best predictors of vote (or some similar variable) are identified as "important" and the remaining variables are usually discarded as potential explanatory factors of voting behavior.

While such procedures might be justified if one's goal was to explain <u>election</u> outcomes, they turn out to be unsatisfactory if one's goal is to characterize <u>voter</u> decision processes. Their shortcoming stems from the fact that taste variations are allowed to influence vote or candidate preference only in a very restrictive way. For

example, two voters may prefer different candidates because the voters have different policy preferences, but not because they possess different decision rules. In fact, if two voters have identical policy preferences and demographic characteristics, then any of the standard methods of analyzing voting behavior would predict that the two would cast identical votes. Yet it is easy to think of situations where such a prediction would be unwarranted. If issues have different levels of salience to voters, then identical policy preferences do not necessarily imply identical (or even similar) voting patterns for voters.

The assumption that voters share the same decision rule except for some measured attributes such as policy preferences, occupation, and social characteristics is common in voting research. Its violation, which will be called heterogeneity, is the topic of this paper. Some research areas where it is plausible to expect heterogeneity may serve to illustrate the concept:

<u>Retrospective and Sociotropic Voting</u> (Fiorina, 1981; Kinder and Kiewiet, 1978). Are voters self-interested or sociotropic? Do they vote more on the basis of prospective or retrospective considerations? These questions are probably ill-posed since some voters may be more self-interested or more retrospective than others and any universal generalizations would be inappropriate.

<u>Economic Influences on Voting</u> (Kramer, 1971; Hibbs et al., 1981). Time series regressions of presidential popularity and aggregate electoral

returns on various macroeconomic variables have attempted to assess the relative weight voters place on unemployment, inflation, and real income growth. Again there is good reason to suspect that voters do not share common evaluations of the importance of each of these variables (Hibbs, 1979).

Issue voting (Pomper, 1975; Popkin et al., 1976). Regressions of vote on party identification and ideology measures have been used to assess the importance and extent of "issue voting." While there are many serious conceptual problems involved in this procedure, one that has been neglected is the "homogeneity assumption," i.e., that all voters put equal weight on the variables. This example is discussed below in Section 3.

Two methods of coping with the heterogeneity problem have been employed in the past. The traditional approach in political science (as well as in market research) has been to ask individuals what is important to them (Repass, 1971; Rabinowitz et al., 1982). Unfortunately, people are notoriously bad reporters of their own decision processes. The Gallup poll, for example, periodically asks its respondents what they think is the most important problem facing the country. While this series fairly accurately mirrors what is in the newspapers at any time, it does not seem to be very closely related to the factors which influence individual voting (see Markus and Converse, 1979, p. 1065). The other approach to the heterogeneity problem is to estimate separate regressions for various groups in the

population (Hibbs et al., 1981). The main drawback of this technique is that one must choose which groups will be allowed to take different coefficients and homogeneity is imposed within each group. What is needed then is an exploratory technique for analyzing heterogeneous decision rules that does not rely on individual reports of preference weights or arbitrary restrictions of parameters within some groups.

With the usual type of preference data — where only an individual's most preferred alternative is reported — there is little hope of learning much about the structure of individual decision processes. However, in some cases a complete preference order is elicited from each individual. The technique proposed here exploits information about the nature of individual preferences that can be obtained from complete preference orders. For example, suppose each individual is asked to rank m alternatives, denoted $\theta_1,\ldots,\theta_m$. Then a complete preference order is a ranking:

$$\theta_{j_1} \; R_i \; \theta_{j_2} \; R_i \; \cdots \; R_i \; \theta_{j_m}$$

where the indices $j_1,\ldots,j_m$ are distinct with $1 \leq j_\ell \leq m$ for $\ell = 1,\ldots,m$ and R_i is a weak binary preference relation (i.e., $\theta_j \; R_i \; \theta_k$ if and only if voter i prefers θ_j to θ_k or is indifferent between them). In most cases we do not have a complete preference order for each individual, but the individual's most preferred alternative, say θ_{j_1}, is elicited. For θ_{j_1} to be the voter's most preferred alternative, it follows that:

$$\theta_{j_1} \; R_i \; \theta_\ell \quad \text{for} \quad \ell = 1,\ldots,m$$

which involves m − 1 preference comparisons (excluding the trivial relation $\theta_{j_1} R_i \theta_{j_1}$). For a complete preference order, transitivity of the preference relation implies:

$$\theta_{j_1} \; R_i \; \theta_{j_\ell} \qquad \text{for} \quad \ell = 2,\ldots,m$$

$$\theta_{j_2} \; R_i \; \theta_{j_\ell} \qquad \text{for} \quad \ell = 3,\ldots,m$$

$$\vdots$$

$$\theta_{j_{m-1}} \; R_i \; \theta_{j_\ell} \qquad \text{for} \quad \ell = m$$

or $(m - 1)m/2$ nontrivial preference comparisons.[1] If the number of alternatives J is moderately large, there is a great deal of information in a complete preference order that is lost if only an individual's most preferred alternative is elicited. In principle, this information can be used to study individual decision processes. An econometric specification that exploits this possibility is developed in the next section. The last section presents an empirical application using data from the 1980 National Election Study. General characterizations of voters as either partisan or ideological are shown to be misleading and some doubt is cast on the meaning of the standard party identification item.

2. THE ECONOMETRIC PROBLEM AND A POSSIBLE SOLUTION

If, as in common practice, heterogeneity is ignored, then parameter estimates based on data where the homogeneity assumption is violated can be seriously misleading. Consider, for example, a linear regression[2] with random coefficients:

$$y_i = x_i'\beta_i + \varepsilon_i \qquad (i = 1,\ldots,n) \tag{2.1}$$

where $E(\varepsilon_i | x_i,\beta_i) = 0$ and the observations are assumed to be independent identically distributed. Then a regression of y_i on x_i will not, in general, identify $E(\beta_i)$:

$$\hat{\beta} = (\sum_{i=1}^{n} x_i x_i')^{-1} \sum_{i=1}^{n} x_i y_i \tag{2.2}$$

so that:

$$\operatorname*{plim}_{n\to\infty} \hat{\beta} = E(\beta_i) + V(x_i)^{-1}\mathrm{Cov}(x_i,\beta_i) \tag{2.3}$$

where $V(\cdot)$ and $\mathrm{Cov}(\cdot)$ denote variance and covariance matrices, respectively. Unless x_i and β_i are uncorrelated, the least squares estimator (2.2) will provide estimates of $E(\beta_i)$ which depend on both the "corresponding" and "noncorresponding" microparameters. (The terminology is borrowed from Theil, 1954, pp. 13-14.) For example, if the explanatory variables in (2.1) include party and ideology variables, the estimated party coefficient will depend not only on the individual party coefficients, but also on the individual ideology coefficients. In other words, the estimated coefficients do not even provide accurate estimates of the average individual effects, but confound the effects of all the variables.

Beyond possible coefficient biases that it might cause, heterogeneity itself might be of interest and it would be desirable to discover what structure — if any -- there is to the heterogeneity and how it influences model predictions. This possibility is particularly important in discrete choice situations where the response function is nonlinear. In political business cycle models, for instance, the support of voters with very low or high probabilities of favoring the incumbent will be difficult to capture or lose. If the policymaker faces some type of Phillips curve constraint, his optimal policy (from the standpoint of vote maximization) will depend on the relative elasticities of substitution between the control variables for the neutral voters (see Kernell and Hibbs, 1981). In this case, it is the nature and extent of heterogeneous preferences that are of central interest, but previously there has been no satisfactory way of studying this phenomenon.

Following McFadden (1974), suppose that we have a random sample of n individual with measured attributes x_i. Individuals are asked to rank m alternatives described by measured attributes θ_j. The utility of alternative j to individual i is given by:

$$u_{ij} = V(x_i, \theta_j) + \varepsilon_{ij} \quad (i = 1,\dots,n;\ j = 1,\dots,m) \qquad (2.4)$$

where $V(\cdot,\cdot)$ is called the strict utility of an alternative and captures the effects of measured attributes of the decision maker and the alternative while unmeasured and possibly idiosyncratic elements

of choice are captured by ε_{ij}. Utilities u_{ij} are unobserved, but individuals may report a ranking $r_i = (r_{i1},\dots,r_{im})$ of alternatives reflecting declining perceived utility, i.e.:

$$u_{ir_{i1}} \geq u_{ir_{i2}} \geq \cdots \geq u_{ir_{im}} \quad (i = 1,\dots,n). \qquad (2.5)$$

Here the indices $r_{i1},\dots,r_{im}$ are assumed to be distinct with $1 \leq r_{ij} \leq m$ for $i = 1,\dots,n$ and $j = 1,\dots,m$. A convenient specification for the distribution of the errors, which we will adopt, is that the ε_{ij} be independently and identically distributed type I extreme value random variables (see Johnson and Kotz, 1970, ch. 21; McFadden, 1978, proposes a generalized extreme value distribution of which this is a special case).

The framework outlined above can be applied in modelling electoral choice. In voting models, it is usually assumed that candidates compete in a multidimensional policy space X which is a compact subset of $\mathbb{R}^p$. Each voter has an ideal point $x_i \in X$ representing his or her most preferred policy. A platform $\theta_j \in X$ is associated with each of the m candidates. In nonstochastic electoral models, the utility of a candidate with platform θ to a voter with ideal point x is usually taken to be a quasi-concave function of θ which is maximized at $\theta = x$. For empirical work, a specific functional form must be selected and an obvious choice (also used in some early theoretical work; see Davis and Hinich, 1966) is the quadratic form:

$$u_{ij} = (x_i - \theta_j)'A_i(x_i - \theta_j) + \varepsilon_{ij} \qquad (i = 1,\ldots,n; \; j = 1,\ldots,m) \tag{2.6}$$

where A_i is a symmetric negative semi-definite matrix.

The specification (2.6) differs from that used by Davis and Hinich (1966) in two respects. First, the addition of the unobservable stochastic term ε_{ij} makes utilities stochastic and accommodates idiosyncracies of preference that are commonly observed in empirical work. The second difference between (2.6) and the original Davis-Hinich setup, is that the matrix of preference weights A_i is allowed to vary between voters, while Davis and Hinich assumed that all voters share an identical matrix A of preference weights. Davis and Hinich (1966, pp. 178–179) commented:

> It should be pointed out that the matrix A . . . is not given a subscript. The reason for this omission is a rather strong assumption. Although the components of the vector x_i can assume any values which the ith individual might desire, it is presumed that the tastes of the voters are such that the matrix A enters the loss function of each individual. The population of voters is assumed to have a certain "homogeneity." In other words, although voters desire differing values of the indices of policy, all voters assign the same relative "weight" (or "importance" to any given issue. This (admittedly unrealistic) presumption is made solely for the reason of analytical convenience.

In later work by Davis and Hinich and others, the homogeneity assumption implied by a constant matrix of preference weights was

discarded, but it still appears occasionally (e.g., Kramer, 1977). In empirical work, as stressed in the introductory section of this paper, a similar type of homogeneity assumption is almost always invoked.

In a voting model with random utilities (2.6) and extreme value errors, the candidate choice probabilities are given by the following result:

<u>Proposition 1</u>: The probability a voter with ideal point x_i and preference weights A_i most prefers candidate j among the candidates in some choice set $B \subseteq \{1,\ldots,m\}$ is given by:

$$P(j|x_i,A_i,B) = e^{V_{ij}} / \sum_{k \varepsilon B} e^{V_{ik}} \qquad (j \varepsilon B) \tag{2.7}$$

where $V_{ik} = (x_i - \theta_k)'A_i(x_i - \theta_k)$ is the strict utility of candidate k to voter i.

<u>Proof</u>: See McFadden (1974, p. 111, lemma 1).

The multinomial logit model, which has been successfully employed for analyzing voting data, generates choice probabilities of the form (2.7). Since with the usual sort of voting data (where only an individual's most preferred candidate is recorded) it is not possible to estimate individual preference weights, a homogeneity assumption is required. If, however, complete rank-order information is available, the following result allows estimation of individual preference weights:

<u>Proposition 2</u>: Under the assumptions of Proposition 1, the

probability of an individual having a ranking $(r_{i1},\ldots,r_{im})$ of the candidates is given by:

$$\Pr(u_{ir_{i1}} \geq \cdots \geq u_{ir_{im}}) = \prod_{j=1}^{m-1} \left(e^{V_{i,r_{ij}}} \Big/ \sum_{k=j+1}^{m} e^{V_{i,r_{ik}}}\right) \tag{2.8}$$

Proof: See appendix.

For estimation purposes, we form the log-likelihood function:

$$\log L(A_1,\ldots,A_n) = \sum_{i=1}^{n} \sum_{j=1}^{m-1} (V_{i,r_{ij}} - \log(\sum_{k=j+1}^{m} e^{V_{i,r_{ik}}})) \tag{2.9}$$

which can be maximized by standard numerical algorithms. The resulting estimates will be called COLOGIT (Completely Ordered Logit) estimates.

Two limitations of the COLOGIT procedure should be mentioned, one statistical and the other conceptual. An undesirable feature of choice probabilities of the form (2.7) is the independence of irrelevant alternatives (IIA). This property, also known in the psychological literature as Luce's choice axiom, requires that the odds of picking one candidate over another not depend on the presence of other candidates in the available choice set:

$$\frac{P(i|x_i,A_i,\{i,j\})}{P(j|x_i,A_i,\{i,j\})} = \frac{P(i|x_i,A_i,B)}{P(j|x_i,A_i,B)} \tag{2.10}$$

where $\{i,j\} \subseteq B$. If the other alternatives are really irrelevant, then (2.10) poses no real problem, but if the other alternatives have

similar unobserved attributes (e.g., $E\varepsilon_{ij}\varepsilon_{ik} \neq 0$), then IIA implies implausible preference patterns. For example, suppose initially the choice-set includes candidate 1 (a liberal), candidate 2 (a moderate), and candidate 3 (a conservative) and voter i is indifferent between the alternatives:

$$\begin{aligned} P(1|x_i,A_i,\{1,2,3\}) &= P(2|x_i,A_i,\{1,2,3\}) \\ &= P(3|x_i,A_i,\{1,2,3\}) = 1/3 \end{aligned} \tag{2.11}$$

Now suppose the field of candidates is enlarged to include two more conservatives (candidates 4 and 5) whose measured characteristics are identical to candidate 3.[3] It seems plausible that the conservative candidates (3,4, and 5) might split candidate 3's vote in the three-way race and, in particular, that the probability of voter i choosing candidate 1 or 2 should be unaffected. Unfortunately, from (2.7) it is clear that the present model would predict:

$$\begin{aligned} P(1|x_i,A_i,\{1,2,3,4,5\}) &= P(2|x_i,A_i,\{1,2,3,4,5\}) = P(3|x_i,A_i,\{1,2,3,4,5\}) \\ &= P(4|x_i,A_i,\{1,2,3,4,5\}) = P(5|x_i,A_i,\{1,2,3,4,5\}) = 1/5 \end{aligned} \tag{2.12}$$

Some specifications that permit departures from IIA have been proposed (McFadden, 1978; Hausman and Wise, 1978; Tversky, 1972; Small, 1981), but these require further parameterizations of the model which are unworkable in this context. In defense of the present specification, it requires no additional assumptions other than those used in all extent logit analyses of candidate preference, while it does dispense with the crucial homogeneity assumption.

The main statistical limitation of the COLOGIT procedure is that with a fixed number m of candidates to be ranked, maximization of the likelihood function (2.9) does not yield consistent estimates of the A_i matrices since the number of parameters is increasing with sample size n. This, however, is not a serious deficiency since we are not generally interested in the parameters for a specific individual. The following theorem states that the COLOGIT estimates can be used to consistently estimate the average parameter for a group as the group's size increases. Alternatively, regressions of the parameter estimates on any set of exogenous variables will be consistent (subject to standard regularity conditions). Let $\theta_i = \mathrm{vech}A_i$ denote the free parameters in A_i.

Proposition 3: Suppose $\theta_i \in \Theta$, a compact convex subset of $\mathbb{R}^q$ with the true value of θ_i, denoted θ_i^*, an interior point of Θ for $i = 1,\ldots,n$. Let:

$$\ell_i(\theta_i) = \sum_{j=1}^{m-1} V_{i,r_{ij}} - \log\left(\sum_{j=1}^{m} e^{V_{i,r_{ik}}}\right) \qquad (2.13)$$

denote the log likelihood of the ith observation and assume:

(i) $E(\partial\ell_i(\theta_i^*)/\partial\theta_i) = 0$ and $V(\partial\ell_i(\theta_i^*)/\partial\theta_i)$ is uniformly bounded for $i = 1,\ldots,n$.

(ii) $\partial^2\ell_i(\theta)/\partial\theta_i\partial\theta_i'$ is positive definite for each $\theta \in \Theta$ for each i and is uniformly bounded by a constant M.

Then

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}_i - \theta^{*}{}_i) \rightarrow 0 \text{ almost surely.}$$

Proof: See appendix.

The theorem shows that the average COLOGIT estimates are strongly consistent estimates of the average group parameters as group size increases. Assumption (i) above requires that the choice probabilities (2.7) be correctly specified while assumption (ii) is a regularity condition that is satisfied if the x_i are bounded (e.g., $x_i \in X$ is sufficient). These assumptions can be weakened somewhat further, but do not seem unduly stringent for the proposed applications.

3. APPLICATION TO ISSUE VOTING

The debate over issue voting has been a contentious affair with emotionally charged exchanges between, on one side, those who doubt voters' "competence" to make informed policy choices and, on the other side, those who find voters "surprisingly sophisticated." The argument has centered upon the relative importance of voters' policy preferences and partisanship on their voting decisions. The following analysis by Nie et al. (1979, p. 166) is not atypical and if anything, a good deal more careful in its choice of language and drawing of conclusions than the average paper on issue voting:

> The contrast between the attitude/vote correlations and the party/vote correlations is dramatic. The relationship between

> issues and the voting decision rises sharply. In 1956, the correlation between issue positions and the vote was 0.18. The correlation goes up dramatically in 1964 and remains substantially above the earlier years. In the pre-1964 period a citizen's position on the issues, as measured by our summary attitude scale, had little or no impact on the way in which he voted. Citizens on the left side of the scale were almost as likely to vote for a Republican as for a Democrat. Citizens on the right were almost as likely to vote Democratic as Republican. There was a small increase in the relationship between issue position and the vote in 1960, but the major shift seems to have come between 1960 and 1964. From 1964 on, there is a considerable association between left-right issue position and direction of the presidential vote. . . . The data suggest that the American public has been entering the electoral arena since 1964 with quite a different mental set than was the case in the late 1950s and early 1960s. They have become more concerned with issues and less tied to their parties.

While the present paper offers no evidence to contradict these conclusions, it is doubtful that correlations between voting and party identification can support such sweeping generalizations about "the American public." As Nie et al. (1979, pp. 227-229) and others have pointed out, blacks as a group have resisted the aggregate trend toward increasing independence. Is it safe to assume the rest of the population has behaved homogeneously?

That not all voters view politics in the same terms has long been a staple of public opinion research. The American Voter classified people into categories based on whether they mentioned abstract ideological criteria, group benefits, or just the "nature of the times" in evaluating parties and candidates (Campbell et al., 1960, chap. 9). Certainly few analyses of public opinion have been as influential in spawning further research. But, in another sense, this message of The American Voter seems to have had almost no impact at all since the heterogeneity of voter decision processes has been ignored in almost all subsequent analyses of issue voting.

One surmises that the neglect of heterogeneity in voting research stems from data analysts' inability to handle the problem. The COLOGIT procedure described in Section 2 provides a feasible way for exploring the heterogeneity problem with standard survey data. The National Election Studies do not ask respondents to rank order a set of candidates, but a rank order can be constructed from the candidate thermometer scores. The most severe practical problem with constructing a rank order in this way is the tendency of respondents to place candidates about whom they have no information at 50° (the midpoint of the thermometer scale) despite the presence of a filter question designed to eliminate evaluations of candidates the respondent doesn't know much about. (If a respondent cannot evaluate a candidate, that candidate is just removed from the individual's choice set but remains in the choice set of other voters.) Note that the COLOGIT procedure uses only the ordinal properties of the thermometer item.

The measured attributes of voters and candidates consist of placement on a three category party identification scale and a seven point left/right scale. All data are from the pre-election (September) wave of the 1980 National Election Study. Voters are classified as Democrats (scored +1), Republicans (-1), or independents (0) based on responses to the first part of the standard party identification item:

> Generally speaking, do you usually think of yourself as a Republican, a Democrat, an independent, or what?

The scoring is, of necessity, somewhat arbitrary with no party preference or minor party preference counted as independent. The second half of the party identification item is not used since selection of the "leaning independent" and "weak partisan" categories appears quite sensitive to current candidate preference (Brody, 1977; Petrocik, 1974). If stronger partisans are in fact more partisan in their behavior, this should be reflected in a larger coefficient for the party term in those voters' strict utility function. The classification of candidates by party is straightforward except for John Anderson who was classified as an independent because at the time of the survey he was conducting a well publicized independent candidacy for the presidency.

Voters are placed on a left/right scale based on their responses to the following question:

> We hear a lot of talk these days about liberals and conservatives. Here is a seven point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative. Where would you place yourself on this scale, or haven't you thought much about this?

Responses were scored from -3 for extremely conservative to +3 for extremely liberal. Voters were also asked to place candidates on this scale. The mean candidate placements, reported in Table 1, were used to locate the candidates. The mean candidate placement was used in place of the individual's placement of the candidates in the hope of avoiding the rationalization problem identified by Brody and Page (1972).

[Table 1 about here]

In the present case, the specification (2.6) of voters' utility function reduces to:

$$u_{ij} = a_{1i}(x_{1i} - \theta_{1j})^2 + a_{2i}(x_{2i} - \theta_{2j})^2 + \varepsilon_{ij} \qquad (3.1)$$

where we have assumed that A_i is diagonal and:

x_{1i} = self placement of voter i on left/right scale

x_{2i} = party identification of voter i

θ_{1j} = placement of candidate j on left/right scale

θ_{2j} = party of candidate j

Due to computational costs, a 15 percent subsample of 238 persons was randomly selected from the 1980 NES survey. Respondents who were not interviewed in the pre-election wave, who declined to place themselves on the left-right scale, or who were unable to rank at least five candidates were deleted from the sample. In addition the numerical algorithm used to compute the COLOGIT estimates failed to converge to a solution in 6 cases. After these deletions, 122 cases were left for analysis.

Is heterogeneity a problem in the NES data? Table 2 compares the standard multinomial logit estimates of (3.1) which impose the homogeneity assumption of equal party and ideology weights for each voter with the average COLOGIT estimates. The discrepancies between the MNL and average COLOGIT estimates are striking. The average party weight estimated by the COLOGIT procedure is more than twice as large as the MNL estimate while the average ideology weight is 50 percent less than the MNL estimate. In neither case is the average COLOGIT estimate within two standard errors of the MNL estimates.

[Table 2 about here]

To obtain a better sense of the variety of voter decision rules found in the sample, consider table 3. Relatively few voters have large party and ideology weights. The predominant pattern is for a voter to weight one of the candidate characteristics--party or left/right placement--fairly heavily and apparently to ignore the other characteristic (with the estimated weight sometimes being perversely signed). A description of the "average" voter based on the MNL equation disguises population heterogeneity in much the same way

census averages describing the "average" family as having 2.5 children do: one has trouble finding an average family.

[Table 3 about here]

Even though the COLOGIT and MNL estimates provide rather different pictures of voter behavior, they do not differ significantly in their implications for a hypothetical election. Tables 3 and 4 simulate the Democratic vote share based on which strategies are adopted by the Democratic and Republican candidates. In both the COLOGIT and MNL simulations, the optimal Democratic and Republican strategies are identical with the parties adopting middle-of-the road positions and the Democrats winning about 52 percent of the vote.[4] This estimate, which ignores all candidate characteristics except party and left/right placement is very similar to Converse's estimate of the "normal vote" (Converse, 1966, p. 27). Since 1964 no Democratic presidential nominee has attained the "normal vote" which may be explained in part by the nomination of left-of-center candidates by the Democrats. Note that a moderately conservative Republican (θ_2 = 1) defeats a moderately liberal Democrat (θ_2 = -1) easily (53-47 approximately) and defeats a truly liberal Democrat (θ_2 = -2) in a landslide (61-39). While these simulations ignore other important factors, such as the performance of the incumbent and the personalities of the candidates, electoral outcomes do appear quite sensitive to the ideological strategies adopted by candidates.

[Tables 4 and 5 about here]

That ideological factors can swing elections but not all voters appear to evaluate candidates in terms of their left-right placement leads us to ask who are the ideological voters? More specifically, does partisanship as it is traditionally measured discriminate between voters who base their candidate preference more on the candidate's party affiliation than on his ideological placement. Table 6 provides a partial answer. Strong and weak partisans differ very little (if at all) in their distribution of ideology weights. A smaller fraction of independents have large ideology weights than either partisan group though a larger fraction of independents fall into the middle category of ideology weights ($-0.2 \leq \hat{a}_1 < 0$). Roughly the same fraction of each group (32 to 38 percent) place essentially no weight on candidate left-right placement in reaching their candidate evaluations. In short, strength of partisanship does not seem to be very closely related to use of ideological criteria in evaluating candidates. If anything, the more highly ideological voters are to be found among the supposed partisans rather than the independents though these differences should not be exaggerated.

[Table 6 about here]

More surprising is the evidence in Table 7 that partisanship, as we usually measure it, does not accurately reflect the degree to which voters employ partisan criteria in evaluating candidates. Strong partisans are somewhat more likely to place greater weight on a candidate's partisan affiliation than weak partisans, but even many self-identified strong partisans place little or no weight on

partisanship in evaluating candidates. Though the quadratic form of the specification (3.1) makes comparison of the independent and partisan categories hazardous in this case, it is quite surprising to find over half of the supposed independents possessing large partisan weights. As evidence by their revealed preferences, the fact that someone declares himself or herself to be an "independent" or a "weak partisan" should not be interpreted to mean that that individual does not employ partisan criteria in evaluating candidates.

To summarize the empirical results in this section, heterogeneity does appear to be present in a simple issue-voting model. Moreover, the failure of the homogeneity assumption does bias one's estimate of the "average" decision rule as indicated by the departure of the average COLOGIT estimates from the MNL estimates. Finally, the relationship between self-reported partisanship and estimated party weights illustrates the unreliability of self-classification as a method of identifying heterogeneity and may explain the limited success of weighing respondent's policy preferences by their own assessments of the policy's importance.[5]

4. CONCLUSIONS

Heterogeneous preferences complicate analyses of voting and other types of individual choice behavior because when models of the "average" decisionmaker are applied to heterogeneous populations, estimates requiring a homogeneity assumption will not, in general, accurately describe "average" behavior. How serious a problem heterogeneity poses depends, of course, on the purposes of the particular analysis in question. For example, if one's interest in analyzing issue voting is to determine whether candidate strategies are (or should be) sensitive to changes in voters' issue preferences, then standard estimation techniques (such as MNL) appear satisfactory since their predictions do not differ substantially from those of the COLOGIT procedure. If, however, one's interest is the psychology of voting then the results of this paper should be quite unsettling. As shown in Section 3, standard estimates of an issue voting model provide a very misleading picture of voter decisionmaking since, when heterogeneity is taken into account, few voters resemble the "average" voter supposedly described by the MNL estimates.

Recent efforts in voting research, best exemplified by the important paper of Markus and Converse (1979) and the influential book of Fiorina (1981), have attempted to capture the complexity of voter decisionmaking. The present paper suggests that perhaps we have placed too much emphasis on developing complex models and ignored the diversity of voting behavior. Recognition of such diversity has sometimes been used as an excuse for avoiding systematic analysis of political behavior and, on occasion, even used to argue that such analysis is impossible. The methods described here allow one to explore the structure of heterogeneous preferences with a minimum of additional assumptions than are required for standard techniques.

APPENDIX

Lemma: $\Pr(U_{ij} \le \xi \mid U_{ij} \ge U_{ik} \text{ for all } k \in B) = \exp[-e^{-(\xi - t_{iB})}]$ where $t_{iB} = \log(\sum_{k \in B} e^{V_{ik}})$ and $j \in B \subset \{1,\ldots,m\}$.

Proof: First, using (2.4) and Proposition 1, observe that:

$$\Pr(U_{ij} \le \xi \mid U_{ij} \ge U_{ik} \text{ for all } k \in B)$$

$$= \frac{\Pr(\varepsilon_{ij} \le \xi - V_{ij} \text{ and } \varepsilon_{ik} \le \varepsilon_{ij} + V_{ij} - V_{ik} \text{ for all } k \in B)}{\Pr(U_{ij} \ge U_{ik} \text{ for all } k \in B)}$$

$$= \left[\frac{e^{V_{ij}}}{\sum_{k \in B} e^{V_{ik}}}\right]^{-1} \int_{-\infty}^{\xi - V_{ij}} \{\prod_{k \in B\setminus\{j\}} \exp[-e^{-(\varepsilon + V_{ij} - V_{ik})}]\} \exp[-e^{-\varepsilon} - \varepsilon]d\varepsilon$$

Making the change of variables $u = e^{-\varepsilon}$ yields:

$$\Pr(U_{ij} \le \xi \mid U_{ij} \ge U_{ik} \text{ for all } k \in B)$$

$$= \frac{\sum_{k \in B} e^{V_{ik}}}{e^{V_{ij}}} \int_{e^{V_{ij} - \xi}}^{\infty} \exp(-u \sum_{k \in B} e^{V_{ik} - V_{ij}})du$$

$$= \exp[-e^{V_{ij} - \xi} \sum_{k \in B} e^{V_{ik} - V_{ij}}] = \exp[-e^{-(\xi - t_{iB})}]$$

Q.E.D.

Proof of Proposition 2: Without loss of generality (reordering indices if necessary), suppose $r_{ij} = m - j + 1$ so $u_{i1} \le u_{i2} \le \cdots \le u_{im}$. First we show that:

$$\Pr(u_{im} \le \xi | u_{i1} \le \cdots \le u_{im}) = \Pr(u_{im} \le \xi | u_{im} \ge u_{ik} \text{ for } k = 1,\ldots,m)$$

The proof is by induction on m. Using the notation of the preceding lemma, define $B_j = \{1,\ldots,j\}$. For $m = 1$, the result holds trivially. Since u_{im} is independent of $u_{i1},\ldots,u_{i,m-1}$, it follows from the induction hypothesis that:

$$\Pr(u_{im} \le \xi \text{ and } u_{im} \ge u_{ik} \text{ for } k = 1,\ldots,m | u_{i1} \le \cdots \le u_{i,m-1})$$

$$= \int_{-\infty}^{\xi - V_{im}} \exp[-e^{-(\varepsilon + v_{im} - t_{iB_{m-1}})}] \exp[-e^{-\varepsilon} - \varepsilon] d\varepsilon$$

Making the change of variable $u = e^{-\varepsilon}$ and noting the identity

$$1 + e^{-(V_{im} - t_{iB_{m-1}})} = e^{t_{iB_m} - V_{im}} = 1/\Pr(u_{im} \ge u_{ik} \text{ for } k = 1,\ldots,m)$$

simplifies the integral:

$$\Pr(u_{im} \le \xi \text{ and } u_{im} \ge u_{ik} \text{ for } k = 1,\ldots,m | u_{i1} \le \cdots \le u_{i,m-1})$$

$$= \int_{e^{-(\xi - v_{im})}}^{\infty} \exp\{-[1 + e^{-(v_{im} - t_{iB_{m-1}})}]u\} du$$

$$= \exp[-e^{-(\xi - t_{iB_m})}] \Pr(u_{im} \ge u_{ik} \text{ for } k = 1,\ldots,m)$$

This establishes that the probability of choosing one alternative from any choice set is independent of the ordering of less preferred alternatives in the set. Therefore:

$$\Pr(u_{ir_{i1}} \ge \cdots \ge u_{ir_{im}}) = \prod_{j=1}^{m-1} \Pr(u_{ir_{ij}} \ge u_{ir_{ik}} \text{ for } k = j + 1,\ldots,m)$$

$$= \prod_{j=1}^{m-1} (e^{V_{ij}} / \sum_{k=j+1}^{m} e^{v_{ij}})$$

Q.E.D.

Proof of Proposition 3: From the first order conditions for $\hat{\theta}_i$ we obtain:

$$0 = \frac{\partial \ell_i}{\partial \theta_i}(\hat{\theta}_i) = \frac{\partial \ell_i}{\partial \theta_i}(\theta_i^*) + \frac{\partial^2 \ell_i}{\partial \theta_i \partial \theta_i'}(\tilde{\theta}_i)(\hat{\theta}_i - \theta_i^*)$$

where $\tilde{\theta}_i$ is a point between $\hat{\theta}_i$ and θ_i^*. Therefore:

$$\left|\left|\frac{1}{n} \sum_{i=1}^{n} (\hat{\theta}_i - \theta_i^*)\right|\right| = \left|\left|\frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial^2 \ell_i}{\partial \theta_i \partial \theta_i'}(\tilde{\theta}_i)\right)^{-1} \frac{\partial \ell_i}{\partial \theta_i}(\theta_i^*)\right|\right|$$

$$\le M \left|\left|\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \theta_i}(\theta_i^*)\right|\right|$$

by assumption (ii). Since $\frac{\partial \ell_i}{\partial \theta_i}(\theta_i^*)$ is independently distributed and, by assumption (i):

$$\sum_{i=1}^{\infty} V(\partial \ell_i(\theta_1^*)/\partial \theta_1)/i^2 < \infty$$

the desired result follows directly from the Kolmogorov strong law of large numbers (Rao, 1973, p. 114).

Q.E.D.

FOOTNOTES

* This research was supported in part by a grant from the National Science Foundation (SES 83-0994).

1. With only a rank order but without the complete set of binary comparisons, transitivity is not testable.

2. Similar results are worked out for limited dependent variable model by Rivers and Vuong (1983). In the linear case, the usual regression estimates can still be justified as the optimal linear predictors (see Rao, 1973, pp. 266-267).

3. This example is a variant of the well-known red bus/blue bus problem.

4. In the simulations, all positive coefficient estimates were set to zero to preserve single-peakedness of individual utility functions.

5. In a recent paper, Rabinowitz et al. (1982), using an analysis of covariance scheme, find somewhat stronger salience effects than earlier studies. The effects are still quite modest (the coefficient of the issue variable increases by roughly 35 percent if the issue is reported salient by the votes) so it is unclear what impact varying preference intensities might have on aggregate electoral outcomes. Markus and Converse (1979, p. 1065) claim,

citing some psychological literature, "that the choice of weights tends to make little difference in the ultimate predictions."

REFERENCES

Brody, R. A. "Stability and Change in Party Identification: Presidential to Off-Years." Paper presented at the 1977 annual meeting of the American Political Science Association, Washington, D.C., September 1977.

Brody, R. A. and Page, B. I. "Comment: The Assessment of Policy Voting." American Political Science Review 66 (1972):450-458.

Campbell, A. et al. The American Voter. New York: John Wiley, 1960.

Converse, P. E. "The Concept of a Normal Vote." In Elections and the Political Order, edited by A. Campbell. New York: John Wiley, 1966.

Davis, O. A. and Hinich, M. "A Mathematical Model of Policy Formation in a Democratic Society." In Mathematical Applications in Political Science II, edited by J. L. Bernd. Dallas: Southern Methodist University Press, 1966.

Fiorina, M. P. Retrospective Voting in American National Elections. New Haven: Yale University Press, 1981.

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

Hibbs, D. A., Jr. "The Mass Public and Macroeconomic Policy: The Dynamics of Public Opinion Toward Unemployment and Inflation." American Journal of Political Science 23 (1979):705-731.

Hibbs, D. A., Jr. et al. "The Dynamics of Political Support for American Presidents Among Occupational and Partisan Groups." American Journal of Political Science 26 (1982):312-332.

Johnson, N. I. and Kotz, S. Distributions in Statistics: Continuous Univariate Distributions - 2. New York: John Wiley, 1970.

Kernell, S. and Hibbs, D. A., Jr. "A Critical Threshold Model of Presidential Popularity." In Contemporary Political Economy, edited by D. A. Hibbs et al. Amsterdam: North-Holland, 1981.

Kinder, D. R. and Kiewiet, D. R. "Economic Discontent and Political Behavior: The Role of Personal Grievances and Collective Economic Judgments in Congressional Voting." American Journal of Political Science 23 (1979):495-517.

Kramer, G. H. "Short-Term Fluctuations in U.S. Voting Behavior, 1896-1964." American Political Science Review 65 (1971):131-143.

__________. "A Dynamical Model of Political Equilibrium." Journal of Economic Theory 16 (1977):310-334.

Markus, G. B. and Converse, P. E. "A Dynamic Simultaneous Equation Model of Electoral Choice." American Political Science Review 73 (1979):1055-1070.

McFadden, D. "Conditional Logit Analysis of Qualitative Choice Behavior." In Frontiers of Econometrics, edited by P. Zarembka. New York: Academic Press, 1974.

__________. "Modelling the Choice of Residential Location." In Spatial Interaction Theory and Residential Location, edited by A. Karlgvist. Amsterdam: North-Holland, 1978.

Nie, N. H., et al. The Changing American Voter, enlarged ed. Cambridge: Harvard University Press, 1979.

Petrocik, J. R. "An Analysis of Intransitivities in the Index of Party Identification." Political Methodology 1 (1974):31-48.

Pomper, G. M. Voters' Choice: Varieties of American Electoral Behavior. New York: Dodd, Mead, 1975.

Popkin, S., et al. "Comment: What Have You Done for Me Lately? Toward an Investment Theory of Voting." American Political Science Review 70 (1976):779-805.

Rabinowitz, G., et al. "Salience as a Factor in the Impact of Issues on Candidate Evaluation." Journal of Politics 44 (1982):41-63.

Rao, C. R. Linear Statistical Inference and its Applications, 2nd ed. New York: John Wiley, 1973.

RePass, D. E. "Issue Salience and Party Choice." American Political Science Review 65 (1971):389-400.

Rivers, D. and Vuong, Q. H. "A Note on Misspecification in Limited Dependent Variable Models." Unpublished paper. California Institute of Technology, December 1983.

Small, K. A. "Ordered Logit: A Discrete Choice Model with Proximate Covariance Among Alternatives." Unpublished paper. Princeton University, December 1981.

Theil, H. Linear Aggregation of Economic Relations. Amsterdam: North-Holland, 1954.

Tversky, A. "Elimination-by-Aspects: A Theory of Choice." Psychological Review 79 (1972):281-299.

TABLE 1

LIBERAL/CONSERVATIVE CANDIDATE PLACEMENT

Candidate	Mean	N
Ted Kennedy	-1.446	892
Jerry Brown	-1.185	644
John Anderson	-0.432	748
Walter Mondale	-0.430	744
Jimmy Carter	-0.261	925
John Connally	0.486	556
George Bush	0.571	744
Ronald Reagan	1.212	921

TABLE 2

RANDOM UTILITY FUNCTION ESTIMATES

Coefficient	MNL[a]	COLOGIT[b]
$\hat{a}_1$ (Ideology)	-0.214)	-0.112
	(0.051)	(0.774)
$\hat{a}_2$ (Party)	-0.313	-0.730
	(0.085)	(2.995)
N	122	122
ln L	-224.4	-1257

[a] Asymptotic standard errors in parentheses.

[b] Mean and standard deviation of estimated coefficients.

TABLE 3

JOINT DISTRIBUTION OF PARTY AND IDEOLOGY WEIGHTS

	$\hat{a}_2 \geq 0$	$-0.5 \leq \hat{a}_2 < 0$	$\hat{a}_2 < -0.5$
$\hat{a}_1 \geq 0$	9.0% (11)	8.2% (10)	18.0% (22)
$-0.2 \leq \hat{a}_1 < 0$	2.5% (3)	14.8% (18)	8.2% (10)
$\hat{a}_1 < -0.2$	18.9% (23)	9.0% (11)	11.5% (14)
$\gamma = -0.32$	N = 122		

TABLE 4

ESTIMATED DEMOCRATIC VOTE SHARES (COLOGIT ESTIMATES)

Democratic Strategy	Liberal	Republican Strategy					Conservative
Liberal	0.5271	0.3468	0.3236	0.3276	0.3403	0.3609	0.4266
	0.6835	0.5271	0.3897	0.3723	0.3856	0.4392	0.4934
	0.7063	0.6443	0.5271	0.4537	0.4654	0.5128	0.5643
	0.7025	0.6609	0.5914	0.5271	0.5405	0.5784	0.6278
	0.6899	0.6515	0.5774	0.5133	0.5271	0.6021	0.6566
	0.6706	0.6068	0.5198	0.4598	0.4412	0.5271	0.6499
Conservative	0.6210	0.5316	0.4607	0.3990	0.3693	0.3818	0.5271

TABLE 5

ESTIMATED DEMOCRATIC VOTE SHARES (MNL ESTIMATES)

Democratic Strategy \ Republican Strategy	Liberal						Conservative
Liberal	0.5205	0.3112	0.2395	0.2421	0.2805	0.3424	0.4273
	0.7391	0.5205	0.3835	0.3496	0.3782	0.4415	0.5295
	0.8179	0.6629	0.5205	0.4600	0.4748	0.5360	0.6263
	0.8110	0.6962	0.5791	0.5205	0.5377	0.6054	0.7007
	0.7593	0.6534	0.5479	0.4918	0.5205	0.6140	0.7291
	0.6840	0.5719	0.4631	0.3942	0.4055	0.5205	0.6862
Conservative	0.5884	0.4639	0.3529	0.2775	0.2575	0.3244	0.5205

TABLE 6

PARTISANSHIP AND ESTIMATED IDEOLOGY WEIGHTS

	Independent	Weak	Strong
$\hat{a}_1 \geq 0$	36.7%	31.9%	38.5%
$-0.2 \leq \hat{a}_1 < 0$	30.6%	23.4%	19.2%
$\hat{a}_1 < -0.2$	32.7%	44.7%	42.3%
	99.9% (49)	100.0% (47)	100.0% (26)

TABLE 7

PARTISANSHIP AND ESTIMATED PARTY WEIGHTS

	Independent	Weak	Strong
$\hat{a}_2 \geq 0$	22.4%	36.2%	34.6%
$-0.5 \leq \hat{a}_2 < 0$	22.4%	44.7%	26.9%
$\hat{a}_2 < -0.5$	55.1%	19.1%	38.5%
	99.9% (49)	100.0% (47)	100.0% (26)