DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

Patterns of Voting on Ballot Propositions: A Mixture Model of Voter Types

Jeffrey A. Dubin

Elisabeth R. Gerber

SOCIAL SCIENCE WORKING PAPER 795

May 1992

Patterns of Voting on Ballot Propositions: A Mixture Model of Voter Types

Jeffrey A. Dubin Elisabeth R. Gerber

Abstract

In this paper we analyze the patterns of behavior voters exhibit over a set of votes. We explore a set of structural estimation problems that involve analyzing several votes at one time and develop estimation techniques for identifying and analyzing patterns. Using the information in these patterns, we introduce a method for studying voter heterogeneity based on a finite mixture model. Finally, we employ data containing actual micro-level vote returns to estimate the mixture model parameters.

Patterns of Voting on Ballot Propositions: A Mixture Model of Voter Types*

Jeffrey A. Dubin† Elisabeth R. Gerber‡

1 Introduction

The 1990 General Election ballot in California contained no less than 49 separate candidate and proposition races.[1] Voters cast votes for a range of elected officials from school board member and water district representative, appellate and state supreme court justices, to state Assembly member and Senator, Governor, and Member of Congress. In addition, they considered several county and local initiatives and referendums, plus twenty eight statewide ballot propositions. The official Ballot Pamphlet (California Secretary of State, 1990), which contained the language, description, analysis, and endorsements of the statewide propositions was 222 pages of legal jargon, in two volumes. Voters in several other American states also faced similarly lengthy ballots.

For voters who are busy with their private lives, jobs, and other interests, the implicit costs of obtaining information and learning about so many candidates and so many

---

*We thank Ken McCue for help in obtaining the data and Henry Brady, Bruce Cain, Skip Lupia, David Grether, Rod Kiewiet, and Doug Rivers for helpful comments.

†Associate Professor of Economics, Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125

‡Assistant Professor of Political Science, Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125

[1]The actual number of races on a ballot varies from precinct to precinct, reflecting differences in county and local elections.

measures are enormous. Political scientists, pundits and journalists have expressed concern about the ability of voters to make sense of such complicated ballots and cast well informed and responsible votes.[2]

Concerns about voter competence are corroborated by survey evidence and exit polls (for example, the National Election Studies (Center for Political Studies) and the California Poll (The Field Institute)) which suggest that voters often have very little content information about the candidates and measures they are either about to vote for or have just cast votes on. Yet, in spite of these difficulties, voters do vote, and empirical evidence that votes vary systematically across voters suggests that they do so in non-random ways. Research on the cognitive and psychological bases of voting suggest that voters use information short-cuts such as partisan cues and other endorsements to help them make sense of the process and to simplify their vote decisions. To the extent that voters rely on partial and potentially non-credible information short-cuts, arguments about the viability and performance of democratic government are called into question (Lupia, 1992; Gerber and Lupia, 1992).

We find it peculiar that, given the length and format of the American ballot and the difficulty of the voting task *as a whole,* most empirical research on voting behavior has focused on the determinants of a single vote and has all but ignored questions about how voters consider the set of votes they are asked to make. The current research directly addresses the question of how voters consider a set of votes by analyzing actual vote patterns.

The analysis of vote patterns should provide leverage for testing alternative voting theories. For instance, the theory of voting which identifies partisanship as the primary determinant of candidate choice leads to predictions about how partisans vote across candidates (see early works by Campbell, Gurin, and Miller, 1954; Campbell, Converse, Miller, and Stokes, 1960; Converse, 1976; Nie, Verba, and Petrocik, 1976; and Jackson, 1975; and more recent works by Franklin and Jackson, 1983; and Wattenberg 1984). Thus in candidate races in a given election, individuals who identify with the Democratic party should be more likely to vote for Democratic candidates and their vote patterns

[2]Historically, students of the American political system debated the very viability of democratic government and the appropriate role for citizen participation. Early work on voting behavior concluded that voters were incapable of performing an active role in government (Gosnell, 1948), and showed little consistency and constraint in their votes (Converse, 1964). More recently, active debate about voter competence has re-emerged in the context of direct legislation. Evidence on voter confusion and information about direct legislation measures is mixed. Magleby (1984) shows that voters have little information about (pp. 140-141) and have difficulty understanding (pp. 118-119) complicated ballot measures, although Cronin (1989) argues that voters can at least understand the meaning of ballot measures and act competently and Zisk (1987) finds that the length of the ballot itself does not seem to affect voting behavior.

should reflect this with a preponderance of Democratic votes. A second example is the theory of strategic voting which predicts that voters will vote against their sincere policy preferences on one proposition to facilitate their policy preferences on a second competing proposition (Dubin, Kiewiet, and Noussair, 1992). To the extent that we observe otherwise moderate voters voting "no" on a moderate proposition and "yes" on a competing extreme proposition, we can estimate the degree of strategic voting across these measures.[3]

We analyze patterns in votes across a set of initiatives and referendums. In the United States, twenty-six states and thousands of municipalities provide for direct citizen voting on public policy propositions. Most Western democracies also have provisions for national referendums.[4] In California, the direct legislation process has received considerable attention in recent years as an important policy vehicle and is viewed by many in other states as a possible solution to legislative stalemate and one-party dominance. Proponents and opponents of the process disagree about its relative merits and shortfalls, but agree for certain that direct legislation has become an important component of state and local politics in many American states.

Direct legislation data provides a unique source for studying voting behavior. Most importantly, we can think of direct legislation votes as direct revelations of policy preferences. Voters vote directly for or against policies, and so the inferences we make about their underlying policy preferences are much more direct than the inferences we can make by studying, for example, votes for candidates. Several works employ aggregate direct legislation data as measures of underlying public opinion toward policy. For example, Deacon and Shapiro (1975) use aggregate referendum returns as indicators of demand for public goods such as environmental protection and mass transit. Kuklinski (1978) and Snyder (1991) use direct legislation results, aggregated to the level of the state legislative district, as measures of constituency preferences in their studies of representation and the dimensionality of issue constraint.

Our analysis relies on individual level rather than aggregate level direct legislation results.[5] The advantage of using individual level vote data to study patterns and individ-

[3]The analysis of voting patterns could also be applied to theories of divided government and split-ticket voting which predict that some voters mix their votes across partisan lines.

[4]See Magleby (1984) for an extensive discussion of the legal requirements and details of comparative direct legislation provisions.

[5]To the extent that individual level votes have been analyzed in the past, it has been through analysis of survey data in which individuals are asked to report their vote intention (in pre-election surveys) or to recall votes they already cast (in exit polls and post-election surveys). Vote validation studies provide evidence that large amounts of error are associated with these reports (Parry and Crossley, 1950; Clausen,

ual vote decisions is simply that we avoid the problems of ecological inference associated with aggregate data (Robinson, 1950; Goodman, 1953; Shively, 1969; Kramer, 1983) and study individual votes directly. Individual level direct legislation data also provide leverage for studying other voting questions, such as strategic voting and vote switching, which to date have been primarily addressed using aggregate data.

The primary disadvantage of using actual direct ballot returns to study voting is that, for reasons of confidentiality, we must do without the personal demographic information about individual voters that we usually obtain in surveys. Measurement of such factors as age, education, income, or race can only be conducted at a higher level of aggregation, for example voting precincts or census tracts. When the model of voting behavior relates individual level characteristics to an individual's votes, using aggregate proxies of these characteristics introduces measurement error into the estimation. Alternatively, if the model relates contextual effects to votes then aggregate measures of these variables are appropriate.

In addition, accessing and analyzing actual ballot returns can be prohibitively costly. By law, most states must make the physical ballots available for inspection for some period after each election, but the ballot returns are rarely available in a user-friendly format.

2 Vote Patterns and Voter Heterogeneity

Studying vote patterns across propositions provides a powerful method for identifying and analyzing voter heterogeneity. We use the term heterogeneity to describe a situation in which voters have similar objective characteristics yet follow different decision rules and exhibit fundamentally different behavioral patterns. We can think of a heterogeneous population as consisting of a finite number of distinct voter types or conversely as consisting of a continuum of unique voters.

Several bodies of voting research are concerned explicitly with the question of voter heterogeneity, especially research on the group bases of voting behavior. For example, a

1968; Traugott and Katosh, 1979). Actual ballots, on the other hand, are free from problems of reporting error. To the extent that they are available, individual ballots provide a much more accurate and reliable record of voting behavior. One study in which actual ballots were analyzed is described in Mueller (1969). Mueller studied a small sample of actual ballots cast in the 1964 general election in California and analyzed patterns of proposition votes cast in that election. Since he analyzed the ballots manually, Mueller was limited by practical concerns to a small number of ballots. Nevertheless, his findings suggest that distinctive patterns of votes across propositions do exist.

great deal of recent research has been directed towards estimating the nature and extent of racially polarized voting (Engstrom and McDonald, 1987; Lupia and McCue, 1990; Freedman, Klein, Sachs, Smyth, and Everett, 1990). These studies are concerned with the degree to which members of minority groups vote homogeneously and in a way that is distinguishable from other identifiable groups. The existence of racial homogeneity has formed the basis for assessing claims for protection of minority groups and "communities of interest" under the Voting Rights Act. Since estimation of racially polarized voting from aggregate vote returns is fraught with statistical difficulties (see e.g. Lupia and McCue, 1990), comparing the micro-level vote patterns of minority group members with those of non-minorities may allow researchers to better estimate the degree of racially polarized voting.

From a statistical perspective, we must be sensitive to the existence of heterogeneity because any statistical model of voting must account for differences in underlying behavior. For example, an empirical study of voting behavior over several candidates or propositions would include explanatory factors that help explain why individuals make particular vote choices. If there are individuals in the population who follow different decision rules than those specified by the model, then an analysis which ignores these differences produces estimates which are likely to be biased for any of the individuals.

In general, voter heterogeneity creates a serious estimation problem because there is not enough information in the data to estimate different coefficients for each individual, and adding more information (in the form of more observations) only increases the scope of the problem by increasing the number of parameters to be estimated. In the political science literature, several works attempt to estimate and correct for heterogeneity across voters. Rivers (1988) and Jackson (1990) recognize that differences in underlying behavior may exist across individuals, and develop estimation techniques to accommodate those differences.[6] Rivers presents a model of candidate preference which estimates the importance of partisanship and issue positions across individuals. He assumes that a voter's utility is a weighted sum of the squared differences between the candidates' partisanship and ideology and her own. These weights vary across individuals and are estimated using additional information in the form of candidate rank orderings. Jackson frames voter heterogeneity as a variable coefficients problem in which parameter values vary across individuals. His approach imposes structure on the ways coefficients can vary across individuals.

Since vote decisions are usually discrete in nature, such as voting for or against

[6]Brady (1988) has identified a different form of heterogeneity. His analysis of survey responses reveals that individuals employ different scales when they rank candidates on thermometer scales. Thus, even if the underlying behavioral patterns across individuals are constant, the manifestation of that behavior can vary.

a candidate or proposition, methods for estimating models with discrete choice data are often appropriate. A recent contribution to the literature on heterogeneity which is particularly useful in the study of voting behavior is the heterogeneous logit model proposed by Dubin and Zeng (1991). Their model investigates extensions to the standard multinomial logit which allow for heterogeneity both across individuals and across choices. Gerber and Lupia (1992) apply a variant of the Dubin and Zeng technique to estimate a heteroscedastic logit model of policy preferences in which individual covariates are weighted by estimated levels of information.

Analyzing patterns of voting behavior across votes provides a new technique for estimating voter heterogeneity. We can think about the patterns of behavior individuals exhibit as being either "typical" or "atypical," and treat the atypical observations much as we would treat outlying observations in structural analysis. Detection of outliers in multivariate data, however, is known to be especially difficult. One reason is that multivariate outliers, unlike univariate outliers, do not stand out easily. In fact, in a multivariate context a point may be an outlier because of small errors in several components rather than a single large error in one component. Thus the standard univariate detection techniques which attempt to find points that "stick out" is not easily generalized to the multivariate context. The situation is clearly exacerbated in the context of discrete data. In the limiting case in which only one vote is under consideration it would be difficult to call someone's yea an outlier while calling someone else's yea a valid vote. This problem diminishes, however, as we increase the dimensionality of the voting space by taking into account the correlations in an individual's votes.

The following example illustrates this point. Suppose our structural voting model predicts that an individual will vote yes on a proposition with high probability, yet we observe her voting no. One explanation is that the observed behavior is an outlier and the individual is following a decision rule other than the one described in our model. A second explanation, however, is that the individual is following the model and simply selects the low probability choice (in this case, a no vote). Alternatively, suppose we extend the voting space to five votes, and our model of voting predicts that an individual will vote yes, no, yes, no, and yes on the five propositions, respectively. If we observe her casting an aberrant string of votes, say no, no, no, no, no, the probability of this pattern, given the model which predicts a mix of yes and no votes is very low, and we are in a better position to classify the observation as an outlier.

There are several reasons why one might hesitate to regard individual behavior that deviates from that of the average voter as "outlying" behavior. First, in real elections every valid vote counts no matter what the motivation. An individual who votes atypically may have cast a perfectly valid set of votes even if they do not happen to conform to the same patterns cast by the typical voter. Second, outlying observations in statistical analysis are usually regarded as a nuisance – something without intrinsic value to be removed from the analysis. Moreover, the observation that voters are segmentable into types and that within those types voters follow a particular strategy is of independent interest.

Thinking about heterogeneity in the context of outlier analysis suggests two methods for handling atypical observations: elimination and accommodation. The elimination method requires that outliers are identified and removed from the "base" model.[7] The accommodation principle, rather than eliminating outlying observations, requires that the model conform to all of the data. This may entail robust estimation methods or modeling the process that generates the outlying observations.

3 Qualitative Mixture Models

An alternative approach to accommodating outlying observations and to studying patterns of voting behavior more generally is to employ a model based on a finite mixture distribution. Generically, a finite mixture model is appropriate when observed phenomena are the consequence of two or more underlying probability distributions (usually unobserved) and when the component distributions arise probabilistically (usually with a frequency which is unknown as well).[8] The estimation problem is to use a sample of observations to decompose the mixture, that is, to estimate the parameters of the mixture distribution and those of the component distributions.[9] We assume that the population of voters can be characterized by a finite mixture of distinct voter types, and that each voter type behaves according to a unique voting rule which governs their vote choices.

It is useful to distinguish the mixture model from two other related but distinct approaches to classifying groups of observations. In the method of discriminant analysis (see e.g. Kendall and Stuart, 1976) an observation is assumed to have been sampled from one of a finite number of populations about which one has supplementary information. The purpose of discriminant analysis is to find a rule which assigns an observation to a member class.[10] In the related area of pattern recognition (Fukunaga (1990), Duda

[7]This technique is employed in the contingent valuation literature to eliminate survey responses reporting implausibly large values of willingness to pay for environmental resources. See e.g. Schultze, McClelland, and Waldman (1991).

[8]In finite mixture problems it is common to assume that the number of populations (or types) from which observations are drawn is known, as are the parametric distributions associated with each type.

[9]The statistical analysis of finite mixture distributions is well presented in Titterington, Smith and Makov (1985).

[10]Standard discriminant analysis which utilizes a linear classification rule is particularly inappropriate

and Hart (1973)), the classification rule is found using a known "training" set of observations.[11] In each case, significant *a priori* information must be employed to facilitate the classification problem. When training data sets or ancillary population statistics are not available, researchers have attempted to use various classification techniques such as clustering, principal components, or other geometric based procedures (Gordon, 1981). These classification methods are often ad hoc in that they lack established statistical properties and are generally ill suited in applications which have high dimensionality.[12] In the absence of auxiliary information that would permit the use of standard classification techniques, we employ a mixture model of qualitative choice to represent the observed voting outcomes.

Several examples of mixture models for continuous (rather than discrete) outcomes have appeared in the econometrics literature. The most well known of these models is the switching regression model of Goldfeld and Quandt (1973). Many of the statistical issues which arise in the context of switching regression models apply equally well in discrete choice models. For instance, the switching regression may not be identifiable under some circumstances[13] and the regimes themselves may be endogenous (i.e. the unobserved random effects which lead individuals to switch regimes may be correlated with the unobserved random effects which affect the individual's response).[14]

Applications of mixture models to qualitative choice situations are apparently new in

---

when applied to discrete variables (Goldstein and Dillon, 1978).

[11]The neural network approach to pattern recognition also requires the use "training" observations (see e.g. Pao, 1989).

[12]A common clustering technique, for example, relies on n(n-1)/2 dissimilarity measures computed between n distinct objects. If dissimilarity is measured using Euclidean distance, the number of calculations required to classify objects with large dimension can be excessive. Even in the case where the component observations are discrete, the storage requirement of order $O(n^2)$ will limit the size of the problem which can be handled by clustering methods.

Another geometric solution to the classification problem is multiple correspondence analysis (see e.g. Hoffman and de Leeuw (1990). This method suffers, however, from the same limitations as arise in cluster analysis. Finally, Christofferson (1975) and Muthen (1978) have developed a latent factor representation for multivariate dichotomous variables. Estimation of such models is constrained by the necessity to evaluate high dimensional multiple integrals.

[13]Mixtures of qualitative choice models are subject to identification constraints. Identification issues are fairly well understood for mixtures of simple univariate distributions (see e.g. Teicher, 1961, 1963); Yakowitz and Spragins, 1968. General identification results for mixtures of multivariate distributions, however, do not exist (see e.g. Rennie, 1974). In our case it should be clear that unless some component distribution exhibits dependence among the votes cast on individual propositions, then there will be no hope in achieving identification.

[14]A good exposition of these issues is given in Maddala (1983, Chapter 8).

the econometrics literature.[15] A family of models which bears some similarity to our proposed class of qualitative mixture models is the randomized response technique of Warner (1965).[16] This technique combines the survey response of an individual with a randomization mechanism so that individual's true response is masked to the interviewer.[17] Using maximum likelihood methods, the effect of covariates on the true response can be recovered from the randomized responses. The random response model differs from the qualitative mixture model in two important respects. First, in the random response model, the mixture is induced by the analyst to preserve respondent anonymity, while in the qualitative mixture model, nature makes the random assignments and thus becomes an inherent part of the choice process. Second, the rule that assigns individuals to groups in the random response model is known to the analyst whereas in the qualitative mixture model, the group assignment problem itself must be estimated.[18]

Another class of models sometimes employed in qualitative choice situations is the random parameter model. In these applications, the parameters of the probabilistic choice model are themselves assumed to be randomly distributed. Random parameter models are the infinite analogues of finite mixture models and are appropriate when the number of individual types is plausibly infinite. The most well known of the random parameter discrete choice models is the covariance probit model due to Hausman and Wise (1978).[19]

---

[15]Mixture models for quantal response have appeared in the statistics literature. One example is a mixture of probit distributions used to analyze arsenical responses by Ashford and Walker (1972). Mixtures of distributions in the logistic class have been considered by Anderson (1979). Related models include mixtures of binomial distributions as discussed in Blischke (1962, 1964) and mixtures of multinomial models as employed by Gordon and Prentice (1977). Issues of estimation are well covered by Titterington et. al (1985) but see especially Hasselblad (1966) for estimation of finite mixture models by maximum likelihood using gradient methods. A useful case study is presented by Do and McLachlan (1984).

[16]The randomized response model is discussed in Zdep, Rhodes, Schwartz, and Kilkenny (1989).

[17]This technique can be useful in situations in which the question content is sensitive such as sexual behavior, tax cheating, domestic violence, etc.

[18]The estimated mixture model can be used for purposes of observation classification. The optimal classification rule (Welch, 1939, Hoel and Peterson, 1949), assigns observation pattern x_i^0 to voter group k if $\pi_k f_k(x_i^0) \geq \pi_\ell f_\ell(x_i^0), \forall \ell \neq k$.

[19]For further examples and discussion see Fischer and Nagin (1981).

4 A Mixture Model of Voter Types

Consider a model of vote choice in which there are K types of voters. Let the probability that a voter is of type k be given by π_k where $\Sigma_{k=1}^{K}\pi_k = 1$. Denote the pattern of votes cast by individual i as x_i. The vector x_i has dimension L where L is the number of candidates or propositions which appear on the ballot. Each component of x_i represents the vote by individual i on a given candidate race or proposition. We employ the notation $x_i = (x_{i1}, x_{i2}, ..., x_{iL})$ to refer to the individual component votes. Let $f_k(x_i)$ be the joint density function for the pattern of votes cast by individual i who is of type k. The finite mixture model implies a sample likelihood function

$$L = \prod_{i=1}^{N} \sum_{k=1}^{K} \pi_k f_k(x_i) \quad (1)$$

with log-likelihood function

$$\mathcal{L} = \sum_{i=1}^{N} log(\sum_{k=1}^{K} \pi_k f_k(x_i)). \quad (2)$$

Unlike previous applications, we adopt a parametric probability distribution function for the unknown parameters π_k. Choosing to represent π_k in a probability family has two advantages. First, if π_k is estimated using equation (1) without restriction, it is possible to produce estimates of π_k which are not bounded between zero and one.[20] Second, we anticipate applications in which the mixture probabilities are themselves functions of individual covariates. Using a parametric probability distribution for the unknown parameters π_k facilitates this construction.[21] In the current analysis, we use a multinomial logit representation:

$$\pi_k = e^{V_k} / \sum_{k=1}^{K} e^{V_k} \quad (3)$$

[20]An alternative solution is to set $\hat{\pi_k} < 0$ equal to zero and to set $\hat{\pi_k} > 1$ equal to one.

[21]If the prior probability that an individual is of type k depends on her observed characteristics, we would follow standard practice and specify V_k to be a function of the covariates.

where the V_k are the unknown parameters to be estimated. We employ a maximum likelihood technique which uses analytic gradient information in our estimation.[22] Since

$$\frac{\partial \pi_j}{\partial V_k} = \begin{cases} -\pi_k \pi_j & j \neq k \\ \pi_k(1-\pi_k) & j = k \end{cases}$$

it follows that the derivative of the log likelihood with respect to V_k is:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial V_k} &= \sum_{i=1}^{N} \left\{ \frac{\sum_{j \neq k} f_j(x_i) \frac{\partial \pi_j}{\partial V_k} + f_k(x_i) \frac{\partial \pi_k}{\partial V_k}}{\sum_{j=1}^{K} \pi_j f_j(x_i)} \right\} \\ &= \sum_{i=1}^{N} \left\{ \frac{\pi_k (f_k(x_i) - \bar{w}_i)}{\bar{w}_i} \right\} \end{aligned}$$

where $\bar{w}_i = \sum_{k=1}^{K} f_k(x_i) \pi_k$.

Estimation of equation (1) requires that we specify some particular functional form for the joint density $f_k(x_i)$. Consider for the moment the simpler problem of estimating the joint density $f(x)$ associated with one of the K voter types.[23] Since the components of x are discrete, the problem of estimating the density $f(x)$ is equivalent to the problem of estimating $Prob[x = x^0]$. Conceptually, this problem is quite simple. One need only count the fraction of times x^0 occurs in a sample of N trials and rely on the law of large numbers for an estimate of the probability $Prob[x = x^0]$. This non-parametric approach has some practical limitations. Taking the simplest case in which the components of x_i are binary, there are 2^L possible vote patterns x^0.[24] If the components of x are independent, the problem becomes much simpler. Under independence,

$$Prob[x = x^0] = \prod_{\ell=1}^{L} Prob[x_\ell = x_\ell^0] \tag{4}$$

[22]This procedure is implemented using Statistical Software Tools (SST), Dubin-Rivers Research, 1510 Ontario Ave., Pasadena, Ca. 91103.

[23]We drop the subscript i to simplify notation.

[24]In the current analysis we consider voting over twenty eight propositions, and so in theory there are over 26 million distinct voting patterns. Therefore, a complete specification of the density function $f(x)$ would require estimation of an inordinately large number of joint probabilities.

In the binary case, equation (4) can be written as:

$$Prob[x = x^0] = \prod_{\ell=1}^{L} P_\ell^{x_\ell^0}(1 - P_\ell)^{1-x_\ell^0}$$

with $P_\ell = Prob[x_\ell = 1]$. Here, estimation of the joint density only requires the determination of $L - 1$ individual probabilities.

To avoid either estimating 2^L parameters or imposing the assumption of independence, several researchers have sought parametric representations of the density $f(x)$ using finite approximations. When the components of x are binary, it is known that a linear combination of 2^L independent basis functions can exactly represent the density $f(x)$ (Ito (1968)). The set of functions $\{\phi_i(x)\}$ forms a basis for the density $p(x)$ provided that it is possible to write $p(x) = \sum_{i=1}^{\infty} c_i\phi_i(x)$. If the basis functions satisfy $\int k(x)\phi_i(x)\phi_j(x)dx = \delta_{ij}$ where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise, then the $\phi_i(x)$ are said to be orthonormal with respect to the kernel $k(x)$. The coefficients c_i in the expansion for $p(x)$ can then be computed by $c_i = \int k(x)p(x)\phi_i(x)dx$.

Two expansions for the joint density $f(x)$ are well known. The first relies on the basis functions $\phi_0(x) = 1, \phi_1(x) = z_1, \ldots, \phi_n(x) = z_n, \phi_{n+1} = z_1z_n, \ldots$ where $z_j = 2x_j - 1$ and is known as the Rademacher-Walsh expansion. The second expansion, due to Bahadur (1961) and Lazarsfeld (1961), uses $z_j = (x_j - P_j)/(P_j(1 - P_j))^{\frac{1}{2}}$. In the first case the kernel function is simply $k(x) = 2^{-L}$ while in the second case the kernel satisfies $k(x) = \prod_{j=1}^{L} P_j^{x_j}(1 - P_j)^{1-x_j}$.

The Bahadur-Lazarsfeld representation for the density $f(x)$ is:

$$f(x) = \left\{\prod_{j=1}^{L} P_j^{x_j}(1 - P_j)^{1-x_j}\right\} \cdot \left\{1 + \sum_{j<k} \rho_{jk}z_jz_k + \sum_{j<k<\ell} \rho_{jk\ell}z_jz_kz_\ell + \ldots\right\}$$

where $\rho_{jk} = E(z_jz_k)$, $\rho_{jk\ell} = E(z_jz_kz_\ell), \ldots$ and so on. The Bahadur-Lazarsfeld expansion shows that every density in binary variables can be written as the product of its kernel (a simple function of the marginal probabilities embodying the independence assumption) and a correction factor which may be well approximated by only a few terms when higher product moments are nearly zero. In the present application, the discrete components of x are trichotomous variables (taking on values yes, no, and abstain) rather than binary variables. Thus the existing orthonormal basis expansions are not applicable. Instead we utilize an alternative approximation to the joint probability distribution based on recursive conditioning.[25]

The joint density $f(x)$ may be written as

$$f(x) = f(x_1)f(x_2|x_1)f(x_3|x_2,x_1)\dots f(x_L|x_{L-1},\dots,x_1).$$

Under independence, the product of the conditional densities reduces to the product of the marginal densities as before. If the component discrete variates are not independent but the conditional densities can be assumed to depend on only a few of the conditioning variables, then a significant savings in complexity can be achieved. Expansions using conditional densities were extensively explored by Chow (1962).[26]

To implement the conditional density expansion we use a multinomial logit specification for the conditional probability that an individual votes yes, votes no, or abstains on each proposition. The explanatory variables are the indicators of the individuals' votes on earlier propositions. This is equivalent to a log linear representation where the log-odds probability of voting yes to abstaining is linear in the previous state effects:

$$\log\left(\frac{\text{Prob}[x_\ell = \text{yes}]}{\text{Prob}[x_\ell = \text{abstain}]}\right) = \gamma_o^\ell + \alpha_1^\ell Y_{\ell-1} + \beta_1^\ell N_{\ell-1} + \alpha_2^\ell Y_{\ell-2} + \beta_2^\ell N_{\ell-2} + \dots$$

and

$$\log\left(\frac{\text{Prob}[x_\ell = \text{no}]}{\text{Prob}[x_\ell = \text{abstain}]}\right) = \psi_o^\ell + \lambda_1^\ell Y_{\ell-1} + \eta_1^\ell N_{\ell-1} + \lambda_2^\ell Y_{\ell-2} + \eta_2^\ell N_{\ell-2} + \dots$$

where $Y_{\ell-1}$ and $N_{\ell-1}$ are binary indicators for a yes and no vote, respectively, on proposition $\ell - 1$. A few comments are in order. First, we have omitted an indicator for abstentions among the explanatory variables since its effect is implicit in the specification. Second, we have used only the first-order effects and omitted higher-order interactions. This is analogous to using a first-order correction in the Bahadur-Lazarsfeld expansion, the properties of which are discussed in Solomon (1961) and Moore (1973).[27] Finally, the logistic representation can be easily extended for the case of higher order discrete variables.

---

[25]The Bahadur-Lazarsfeld technique can be generalized to cases where the discrete components take on more than two values (Bahadur, 1961, pp. 167-168).

[26]Chow applied his expansion to a problem in pattern recognition where he was able to exploit a natural spatial dependence between observations.

[27]Gilbert (1968) used a log linear model to represent the density $f(x)$ and also assumed that all second- and higher-order effects were zero.

5 Estimation

In proposition voting, most voters cast a mix of yes and no votes and abstentions over a set of propositions.[28] Propositions, placed on the ballot by either the legislature or by citizen's groups can represent a wide range of interests.[29] Therefore, a voter following one of several "typical" decision rules, such as voting all liberal or all conservative will vote yes on some measures and no on others. Other decision rules, such as following endorsements by the major parties or newspapers, also will result in a mix of yeas, nays, and abstentions.

In contrast to a pattern of votes mixing yeas, nays, and abstentions, some of the most distinctive patterns are strings of all yes votes, all no votes, or all abstentions. We can infer that a voter casting a string of all no votes, for example, is following a different decision process than a voter mixing yeas and nays, perhaps using his vote as a way of signaling disaffection to politicians. Similar inferences can be made regarding all abstainers and all yes voters.

Our mixture model captures these differences between "typical" voters and "pure type" voters by positing four types of voters in the sample. Three represent the all yes, all no, and all abstain pure types, and the fourth base category represents voters who mix their votes. Included in the base category are voters who exhibit a wide range of vote patterns across propositions, including those we would expect to observe if the individual was voting according to issue evaluations, elite or newspaper endorsements, or ideology. Subsequent analyses may specify some of these other patterns as distinct types as well. The current implementation of the finite mixture model compares pure type voters to the rest of the sample.

To simplify the estimation of our base voting model, we employ the logistic probability expansion discussed in the previous section.[30] This representation of the base voting model provides an approximation of the underlying vote patterns in the data based on

[28] As discussed below, over 96% of voters in Los Angeles county cast a pattern of votes with some mixing (compared with all yes, all no, or all abstain patterns). This mixed pattern is also overwhelmingly present in similar data from other years and seems to be the general rule.

[29] In the period between 1974 and 1990, propositions were placed on the California ballot by such diverse groups as farmers, the insurance industry, the tobacco and alcoholic beverages industries, environmental groups, and a wide range of citizen groups. Some of the measures were pro-citizen, pro-consumer, or pro-taxpayer, while others were clearly pro-business in nature.

[30] A full structural specification would require a deep analysis of the issue content of each proposition and a series of hypotheses about how the propositions (and votes on them) are linked. Such an analysis is beyond the scope of this paper.

the first-order correlations between each of the propositions. Structural interpretation of the estimates generated by this model, however, must recognize the conditional nature of this particular expansion.

For pure type voters, we assume that the individual has a large but unknown probability of voting identically on every proposition. If we assume that this probability is exactly equal to one, this is equivalent to specifying *a priori* that a pure type voter must vote identically on every proposition. Such a characterization seems too extreme, however, and we prefer a model which is less stringent and accepts small deviations from the pure type model. This will allow us to include voters in the pure type model who make a small number of mistakes or mispunches in their votes, or who deviate from the pure type model by only a few votes.[31]

Our pure type voting models are special cases of the logistic expansion which we use to represent the base voting model. In the pure type models, we assume that propositions are considered independently.[32] The joint density for the pure yes type, for example, has the form:

$$f_Y(x_i) = P_H^{Y_{i1}} P_L^{N_{i1}} P_L^{A_{i1}} P_H^{Y_{i2}} P_L^{N_{i2}} P_L^{A_{i2}} \dots \tag{5}$$

where Y_{ij}, N_{ij}, and A_{ij} are indicator variables for a yes, no, or abstention on proposition j by individual i. P_H is the probability that a pure type voter votes yes on a given proposition, and P_L is the probability that she votes no or abstains. By construction, $2 \cdot P_L + P_H = 1$. Collecting terms in equation (5), we see that:

$$f_Y(x_i) = P_H^{Y_i} P_L^{L-Y_i} \tag{6}$$

where $Y_i = Y_{i1} + Y_{i2} \cdots + Y_{iL}$ is the number of yes votes by individual i among all L propositions. Similar specifications for the all no and all abstain pure type models are:

$$\begin{aligned} f_N(x_i) &= P_H^{N_i} P_L^{L-N_i} && (7) \\ f_A(x_i) &= P_H^{A_i} P_L^{L-A_i}. && (8) \end{aligned}$$

[31]Whether an observation is categorized as a pure type depends both on the number of deviations from the pure pattern and the location of the deviations.

[32]If pure type voters are making their vote decisions independent of the issue content of the proposition and according to some other motivation, this independence assumption is well justified.

As a final simplification, we estimate the mixture model in two steps. First we estimate the parameters of the base model, f_0, for a subsample of non-pure type voters.[33] From the first stage estimates, we predict f_0 in the full sample and denote these predictions by $\hat{f}_0$. The second stage estimates the conditional maximum likelihood:

$$\mathcal{L} = \sum_{i=1}^{N} \log\left(\pi_0 \hat{f}_0(x_i) + \pi_N f_N(x_i) + \pi_Y f_Y(x_i) + \pi_A f_A(x_i)\right) \tag{9}$$

where f_N, f_Y, and f_A are given above. As with the mixture probabilities, π_k, we introduce an unknown parameter, α_H, to measure the probability P_H via a logistic transformation:

$$P_H = 1/(1 + e^{-\alpha_H}). \tag{10}$$

Finally, by using equations (6), (7), and (8) and imposing the restriction $2P_L + P_H = 1$, we obtain the last analytic derivative required to estimate $\mathcal{L}$:

$$\frac{\partial \mathcal{L}}{\partial \alpha_H} = \sum_{i=1}^{N} \left\{ \frac{[f_N(x_i)(N_i - P_H L) + f_Y(x_i)(Y_i - P_H L) + f_A(x_i)(A_i - P_H L)]}{\bar{w}_i} \right\}.$$

The remaining unknown parameters to be estimated are V_N, V_Y, V_A from equation (3), and α_H from equation (10).

6 Data

We employ a unique data set which consists of images from the actual ballots cast by approximately 1.8 million voters in Los Angeles county in the November 1990 General Election. LA county has a unique method for tabulating and recording votes. The county uses punch-card style ballots on which the voter creates physical perforations to indicate his vote. The cards also contain precinct identification information which allows

[33]To estimate the parameters of f_0, we include only those observations in our sample whose vote patterns deviate from pure type behavior by at least five deviations. We then predict f_0 in the full sample using the subsample parameter estimates. Our results are not sensitive to the number of deviations we use to identify the base model and very similar estimates resulted when we used a larger number of deviations to select the base model sample.

us to determine the voter's residence and an exact list of candidates the voter considers. These cards are transported from the precincts to a central location and fed into a machine similar to the card readers used in the early days of computer data analysis. The machine reads the information on the cards and tabulates the votes for each race and measure. As a by-product of the counting process, a binary image of each ballot is written to a magnetic tape. This tape, after extensive manipulation, provides the data for our analysis.

Several factors complicate our analysis of the ballot image tape. First, the information on the tape is not used for official counting purposes, and so the procedures used to generate the tape are often ad hoc. For example, there are several circumstances in which ballots from a given precinct are fed into the reader, written to the tape, and then at a later time erased from the official count (but not the tape) and re-read. This shows up as a duplication on the tapes, and our procedures are designed to identify and eliminate these duplications. Second, ballots from different precincts have information contained at different places on the cards. Punch positions for a given candidate race or proposition vary across precincts because voters in different areas in the county vote for different local officials and measures, and so the sequence of votes differ across these areas. Punch positions within candidate races also vary, because LA county rotates the sequence of candidates on the ballots so no individual candidate can unfairly benefit from her position on the ballot. In total, there are 320 different ballot "groups," representing each of the possible combinations of races and rotation sequences. Since the images on the tape are simply binary representations of the votes, we must match each ballot to one of 320 different ballot groups to correctly match punches with votes. Third, ballot cards may be read into one of several card readers simultaneously and their images are also written to the tape simultaneously. This means that images of cards from different precincts may be "shuffled" together on the tape, and so must be "unshuffled" in order to facilitate the matching described above.

Once we have identified, read, and recoded all of the individual ballot images, we can aggregate voters within each precinct and verify our counts by comparing precinct totals to the official vote record reported in the Statement of the Vote (California Secretary of State, 1990). Since the rules that define the exact collection of votes used in creating the SOV are not fully understood (by us or by the county) and since the procedures for creating the tapes are even less rigorous, it is impossible for us to exactly replicate the SOV totals. However, our procedures and implementation of the rules as we understand them identifies 1,839,960 of the 1,925,811 records (96%) that contribute to the totals reported in the SOV. We discard the 85,851 records that we can not match exactly to avoid the possibility of double counting. These records come from approximately 300 of the 6614 precincts in the county, and the discarded precincts appear to be distributed randomly across the county.

For each of the approximately 1.8 million valid cases, the data shows the actual vote cast on each of the twenty eight statewide ballot propositions, which we analyze in this study, plus votes for state, county, and municipal candidates and county and city measures for each individual ballot. For the purposes of the present analysis, we have taken a 0.5 percent random sample from the universe of valid records. This random sample contains a total of 9148 valid records.

The statewide propositions on the 1990 ballot consisted of five constitutional amendments or statutes proposed by the legislature, ten bond measures (also proposed by the legislature), and thirteen constitutional amendments or statutes proposed by citizens.[34] The measures ranged in substance from taxation and government reform to fisheries and environmental regulation, veterans and education bonds and drug treatment programs. Appendix A lists the number and title of each of the twenty eight propositions.

Compared to previous years, the 1990 California ballot had several distinguishing features. First, with twenty eight measures on the ballot, voters were faced with an unusually high number of propositions.[35] Second, the passage rate was historically low, with only 6 of the twenty eight measures passing. Together, these two trends have been popularly interpreted as evidence of a high degree of voter alienation and a general rejection of the direct legislation process. The results of subsequent elections will show whether November 1990 was the beginning of a trend in direct legislation politics or rather an aberrant election. In addition, the November 1990 election was characterized by extremely high levels of campaign spending, especially on several measures such as the alcohol tax and the environmental regulation measures. Also, several of the measures on the ballot were "competing measures," in which an extreme measure was countered with a similar, more moderate measure. This strategy was introduced (with mixed success) with respect to the insurance reform measures in the 1988 general election and again in the current election on the alcohol tax, term limits, forestries, and environmental measures.

Table 1 shows the aggregate vote results from our sample.[36] Several interesting patterns are evident. First, we find that the percent of voters abstaining increases on later propositions. This finding is consistent with the common lore that voters tire as they work their way down the ballot and "drop off." Note, however, that this trend is not monotonic, and that there are both positive and negative variations in the percent of voters abstaining on later propositions. From the aggregate data alone, we are unable

[34]An additional four of the citizen initiatives provided for the approval of general obligation bonds as well.

[35]Only the November 1988 ballot had more statewide measures, with twenty-nine.

[36]In Table 1, invalid ballots were those which contained multiple punches for a given proposition.

to determine whether some individuals abstain from all measures or whether individuals pick and choose which propositions they will vote on or not.

(Insert Table 1 Here)

Second, the returns in Los Angeles were quite different from the statewide returns. In particular, more propositions gained the required majority of votes to pass in the county than statewide, and several of the propositions that did pass failed to gain a majority vote in Los Angeles county. These differences reflect the nature of the Los Angeles electorate which is more urban, ethnic and liberal than the rest of the state, on average.

7 Results

Table 2 presents the actual counts and frequencies of voters who voted all yes, all no, and all abstain, respectively, on the twenty-eight ballot propositions, plus the residual category containing all other vote patterns. The observations included in the first three categories represent "pure type" voters in an absolute sense – they cast pure vote patterns without any exceptions. We find that these absolute pure types occur very rarely in the sample, with only 0.2% of Los Angeles county voters casting all yes votes, 1.72% casting all no votes, and 1.75% completely abstaining on the propositions. The vast majority, 96.31%, mixed their votes over the propositions with some yeas, some nays, and some abstentions.[37]

(Insert Table 2 Here)

As we have discussed, this characterization of pure type voters is excessively restrictive and excludes many voters who cast only one or two errant votes and who ought to be

[37] The all abstain cases in our sample are particularly interesting. All of the voters in our sample were actual voters - that is, they turned out to vote on election day (or mailed in absentee ballots before election day) and cast at least one vote. Therefore, we know that voters who abstained on all of the propositions must have cast at least one vote on the candidate portion of the ballot, otherwise they would not be included in our sample. This is a different type of abstention, then, than a voter not turning out and voting at all. Subsequent analyses will examine the all abstention cases to determine exactly what types of voters incur the costs of turning out but fail to consider many of the measures on the ballot. Similar analyses of the geographic bases for all no voting should provide insight into this particular form of "protest" voting.

categorized as pure types. Our use of the mixture model, which assigns observations to each type probabilistically should be more tolerant of small deviations from the pure type patterns and should therefore better capture voter heterogeneity in the data.

Tables 3 and 4 present the results of the conditional logit probability expansions used to estimate our base model. Recall that a structural interpretation of these results is complicated by the conditional nature of the logistic specification. Nevertheless, the patterns that emerge from the estimated models display some interesting relationships among the sample votes and provide insight into the viability of our base model for capturing these relationships. Table 3 presents the log-odds of an individual voting yes on each proposition compared with that individual abstaining. Table 4, which follows later, compares no votes against abstentions.

(Insert Table 3 Here)

Along the vertical axis of Table 3, we list each of the twenty-eight propositions. Along the horizontal axis, we place a series of dichotomous variables indicating a yes vote and a no vote, respectively, for propositions 124 through 150. These dichotomous variables constitute the explanatory variables in the series of twenty-eight conditional logit estimations. In this manner, we report the relationship between each explanatory variable and the log-odds of a yes vote to an abstention for each proposition. Table 4 reports the relationship between each explanatory variable and the log-odds of a no vote to an abstention. Given that the conditional logit models have a large number of coefficients and that the substantive interpretations of the estimated coefficients are ambiguous, we notate in the tables only whether there was a statistically significant positive or negative estimated relationship. Blank positions in the lower triangle of these tables indicate that the estimated relationship was not significant.

Two points about Table 3 are of primary interest. First, we observe that as the number of right hand side variables increases (that is, for higher numbered propositions), levels of significance fall off dramatically and we observe fewer and fewer significant relationships. This suggests that there is probably a high degree of collinearity between the vote variables so that it is difficult to separate out the independent effects of each individual variable as we add more and more. Second, we find that the coefficients on the variables closest to the diagonal are significant and positive in every case. For example, in the equation for votes on Proposition 148, the variables for vote yes and vote no on Proposition 147, as well as the variables for vote yes and vote no on Proposition 146 are positive and significant. This suggests that the order of the propositions matter - voters exhibit a "smoothness" across their votes on contiguous propositions. What is peculiar is that both the yes vote and the no vote are positively associated with the log-odds of a yes vote versus an abstention. This implies that if an individual casts a vote of either type on a proposition, she is likely to cast a vote (rather than abstain) on the next proposition as well. The magnitude of these effects drops for later propositions–consistent with the observation that voters experience "voter fatigue" and "drop-off" (i.e. abstain) with a higher probability later on the ballot.

The results contained in Table 4 are very similar to those contained in Table 3. Again, we find that as we increase the number of votes on the right hand side of the conditional logit models, the number (and percentage) of significant coefficient estimates falls. We also continue to observe the strong pattern of positive, significant relationships on contiguous propositions, suggesting a spatial correlation in an individual's votes.[38]

(Insert Table 4 Here)

Table 5 presents the maximum likelihood results of the estimation of the qualitative mixture model. The estimates for V_j correspond to the unknown parameters in equation (3). These estimates are shown in column 2, and the corresponding t-statistics are reported in column 3. In the fourth column, these estimates are converted back into the mixture probabilities π_2 (all yes), π_3 (all no), and π_4 (all abstain).

(Insert Table 5 Here)

We observe that the mixture model estimates 0.73% of the cases as pure type all yeas, 3.91% as pure type all nays, and 1.95% as pure type all abstains. By construction this implies a mixture probability for the base model, π_1, of 93.44%. Comparing these estimates with the absolute pure type frequencies in Table 1, we find that, as expected, the mixture model analysis identifies more cases as pure types (in fact, about twice as many).

Further examination of our estimates for the mixture probabilities show that nearly 4% of the cases are identified as pure type all nays. This suggests that in the data, there are a substantial number of voters who cast "nearly pure" no vote patterns over the twenty-eight propositions. This is in contrast with our estimate for, π_4, the all abstain pure type probability. We know from our initial counts that there were about the same number of absolute pure type all no voters and absolute pure type all abstainers. However,

[38]We might alternatively think of voting on a series of propositions as a sequential process, and thus cast the correlation between votes as a matter of serial, rather than spatial correlation.

the relatively smaller value of π_4 indicates that there were many fewer near-pure type all abstainers. Voters were less likely to abstain on all but one or two propositions–if they abstained on most propositions, they abstained on all. In contrast, voters who were pre-disposed to vote all no were more likely to abstain or vote yes on at least a few propositions.

Finally, we estimate α_H from equation (10) as 3.2978 (t-statistic: 73.04). This implies a value of 96.44% for P_H. As expected the value of P_H is very close to one and corresponds to our hypothesized notion of a pure type voter.

8 Discussion and Conclusion

Our use of a finite mixture of qualitative choice models has several unique features. First, the mixture approach allows us to identify and analyze voter heterogeneity in a powerful new way. By using the finite mixture class we rigorously adopt a well known statistical methodology which may provide new insights into voting and voter decision making. Second, our model develops a mechanism for *ex-post* classification. Thus, a new observation with a given vote pattern may be classified into a distinct voting type based on the *a priori* probability that an observation arises from a given joint probability distribution, and on the likelihood that a specific pattern would arise if drawn from that distribution. This paradigm makes it unlikely that we classify a pattern that consists of many yes or no votes as a pure type if the prior probability of finding a pure type voter is small. Third, our development of the joint probability of a vote pattern allows position to matter. Indeed, some dependence of the vote pattern is an essential condition for identification. Alternative techniques which simply count the number of yes or no votes cast by an individual do not allow for the dependence in voting behavior across propositions. Fourth, when we do assume independence across propositions (in the pure type voting models) we implicitly employ a form of probabilistic counting. Thus, the model does not require that a pure type voter "get it right" on every proposition (i.e., vote yes, no, or abstain identically for all propositions). Instead, pure type voters are allowed to deviate but with a small probability. Finally, we believe the mixture paradigm will prove valuable in other settings. For example, a mixture model could used to identify alternative types of voters who are more difficult to identify or count, such as voters who follow newspaper recommendations or other endorsements. Alternatively, a mixture model may prove useful for analyzing patterns in other types of political data. For example, these techniques could easily be applied to legislative roll call data to identify and analyze distinct patterns in legislators' voting behavior.

9 References

Anderson, J. 1979. "Multivariate Logistic Compounds," *Biometrika,* 66:17-26.

Ashford, J. and P. Walker. 1972. "Quantal Response for a Mixture of Populations," *Biometrics,* 28:981-988.

Bahadur, R. 1961. "A Representation of the Joint Distribution to N Dichotomous Items," in *Studies in Item Analysis and Prediction*, H. Solomon (Ed.), Palo Alto, Calif.: Stanford University Press, pp. 158-168.

Blischke, W. 1962. "Moment Estimation for the Parameters of Two Binomial Distributions," *Ann. Math. Statistics*, 33:444-454.

Blischke, W. 1964. "Estimating the Parameters of Mixtures of Binomial Distributions," *Journal of the American Statistical Association*, 59:510-528.

Brady, H. 1988. "The Perils of Survey Research: Inter-Personally Incomparable Responses," *Political Methodology*, 11:269-291.

California Secretary of State. 1990. *Ballot Pamphlet.*

Campbell, A., G. Gurin, and W. E. Miller. 1954. *The Voter Decides,* Row, Peterson and Co.: Evanston, IL.

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

Center for Political Studies. *The National Election Studies.* Ann Arbor, MI.

Chow, C.K. 1962. "A Recognition Method Using Neighbor Dependence," *IRE Trans. Elec. Comp.*, EC-11, 683-690.

Christofferson, A. 1975. "Factor Analysis of Dichotomized Variables," *Psychometrica,* 40:5-32.

Clausen, A. 1968. "Response Validity: Vote Report," *Public Opinion Quarterly,* 41:56-64.

Converse, P. E. 1964. "The Nature of Belief Systems in Mass Publics," in *Ideology and Discontent,* David E. Apter (ed.), Free Press: New York.

Converse, P. 1976. *The Dynamics of Party Support: Cohort-Analyzing Party Identification,* Sage Publication: Beverly Hills, CA.

Cronin, T. E. 1989. *Direct Democracy: The Politics of Initiative, Referendum, and Recall,* Harvard Univ. Press: Cambridge, MA.

Deacon, R. and P. Shapiro. 1975. "Private Preference for Collective Goods Revealed Through Voting on Referenda," *The American Economic Review*, 65:943-955.

Do, K. and G. McLachlan. 1984. "Estimation of Mixing Proportions: A Case Study," *Applied Statistics*, 33:134-140.

Dubin, J. and L. Zeng. 1991. "The Heterogeneous Logit Model," California Institute of Technology, Social Science Working Paper No. 759.

Dubin, J., D. Kiewiet, and C. Noussair. 1992. "Voting on Growth Control Measures: Preferences and Strategies." Forthcoming in *Economics and Politics.*

Duda, R. and P. Hart. 1973. *Pattern Classification and Scene Analysis*, John Wiley and Sons: New York, NY.

Engstrom, R. and M. McDonald. 1987. "Quantitative Evidence in Vote Dilution Litigation, Part II: Minority Coalitions and Multivariate Analysis," *Urban Lawyer*, 19:65-75.

The Field Institute. *The California Poll.* San Francisco, CA.

Fischer, G. and D. Nagin. 1981. "Random Versus Fixed Coefficient Quantal Choice Models," in *Structural Analysis of Discrete Data with Econometric Applications,* edited by C. Manski and D. McFadden, The MIT Press: Cambridge, MA.

Franklin, C. and J. E. Jackson. 1983. "The Dynamics of Party Identification," *American Political Science Review*, 77:957-973.

Freedman, D. A., S. Klein, J. Sacks, C. Smyth, and C. Everett. 1990. "Ecological Regression and Voting Rights," Technical Report No. 248, Department of Statistics, University of California, Berkeley, CA.

Fukunaga, K. 1990. *Statistical Pattern Recognition*, 2nd Edition, Academic Press: San Diego, CA.

Gerber, E. and A. Lupia. 1992. "Competitive Campaigns and the Responsiveness of Direct Legislation: Theory and Evidence." Paper presented to the 1992 Annual Meeting of the Midwest Political Science Association, April 9-11, Chicago, IL.

Gilbert, E. 1968. "On Discrimination Using Qualitative Variables," *Journal of the American Statistical Association*, 63:1399.

Goldfeld S. and R. Quandt. 1973. "The Estimation of Structural Shifts by Switching Regressions," *Annals of Economic and Social Measurement*, 2:475-85.

Goldstein, M. and W. Dillon. 1978. *Discrete Discriminant Analysis,* John Wiley and Sons: New York, NY.

Goodman, L. 1953. "Some Alternatives to Ecological Correlation," *American Journal of Sociology*, 64:610-625.

Gordon, A. 1981. *Classification*, Chapman and Hall: University Press, Cambridge.

Gordon, A. and I. Prentice. 1977. "Numerical Methods in Quaternary Palaeoecology. IV. Separating Mixtures of Morphologically Similar Pollen Taxa," *Review Palaeobotany and Palynology*, 23:359-372.

Gosnell, H. 1948. *Democracy: The Threshold of Freedom,* Ronald Press: New York, NY.

Hasselblad, V. 1966. "Estimation of Parameters for a Mixture of Normal Distributions," *Technometrics*, 8:431-444.

Hasselblad, V. 1969. "Estimation of Finite Mixtures of Distributions From the Exponential Family," *Journal of the American Statistical Association*, 64:1459-1471.

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

Hoel, P. and R. Peterson. 1949. "A Solution to the Problem of Optimal Classification," *Annals of Mathematical Statistics*, 20:433-438.

Hoffman, D. and J. de Leeuw. 1990. "Geometrical Aspects of Multiple Correspondence Analysis: Implications for the Coordinate Scaling Debate," University of California, Los Angeles, Department of Statistics Working Paper No. 49.

Ito, T. 1968. "A Note on a General Expansion of Functions of Binary Variables," *Inform. and Control*, 12:206-211.

Jackson, J. E. 1975. "Issues, Party Choices, and Presidential Voting," *American Journal of Political Science*, 19:161-186.

Jackson, J. E. 1990. "Estimation of Models with Variable Coefficients," Paper presented to the Annual Meeting of the Midwest Political Science Association, Chicago, IL.

Kendall, M. and A. Stuart. 1976. *The Advanced Theory of Statistics*, Vol. 3, 3rd edition, Griffin: London.

Kramer, G. 1983. "The Ecological Fallacy Revisited: Aggregate- versus Individual-Level Findings on Economics and Elections, and Sociotropic Voting," *American Political Science Review*, 77:92-111.

Kuklinski, J. 1978. "Representativeness and Election: A Policy Analysis," *American Political Science Review*, 72:165-177.

Lazarsfeld, P. 1961. "The Algebra of Dichotomous Systems," in *Studies in Item Analysis and Prediction*, H. Solomon (Ed.), Stanford University Press: Palo Alto, CA, 111-157.

Lupia, A. 1992. "Credibility and the Responsiveness of Direct Legislation," forthcoming in *International Symposia on Economic Theory and Econometrics*, Vol. 8, Cambridge University Press: New York.

Lupia, A. and K. McCue. 1990. "Why the 1980s Measures of Racially Polarized Voting are Inadequate for the 1990s," *Law and Policy*, 12:353-387.

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

Magleby, D. 1984. *Direct Legislation: Voting on Ballot Propositions in the United States*, Johns Hopkins Univ. Press: Baltimore, MD.

Moore, D.H., "Evaluation of Five Discrimination Procedures for Binary Variables," *J. Am. Statistical Association,* 68:339-404.

Mueller, J. E. 1969. "Voting on the Proposition: Ballot Patterns and Historical Trends in California," *American Political Science Review*, 63:1197-1212

Muthen, B. 1978. "Contributions to Factor Analysis of Dichotomous Variables," *Psychometrica*, 43:551-560.

Nie, N., S. Verba, and J. Petrocik. 1976. *The Changing American Voter,* Harvard University Press: Cambridge, MA.

Pao, Y. 1989. *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley: Reading, MA.

Parry, H. and H. Crossley. 1950. "Validity of Responses to Survey Question," *Public Opinion Quarterly*, 14:61-80.

Rennie, R. 1974. "On the Interdependence of the Identifiability of Finite Multivariable Mixtures and the Identifiability of Marginal Mixtures," Sankya, A34, 449-452.

Rivers, D. 1988. "Heterogeneity in Models of Electoral Choice," *American Journal of Political Science*, 32.3:737-757.

Robinson, W. S. 1950. "Ecological Correlations and the Behavior of Individuals," *American Sociological Review*, 15:351-357.

Schultze, W., G. McClelland, and D. Waldman. 1991. "Sources of Errors in Contingent Valuation," Working Paper, University of Colorado, Department of Economics.

Shively, W. P. 1969. "Ecological Inference: The Use of Aggregate Data to Study Individuals," *American Political Science Review*, 63:1183-1196.

Snyder, J. 1991. "The Dimension of Constituent Preferences: Voting on California Ballot Propositions, 1974-1990," University of Chicago Business School. Mimeo.

Solomon, H. 1961. "Classification Procedures Based on Dichotomous Response Vectors," in *Studies in Item Analysis and Prediction*, H. Solomon (ed.), pp. 177-186, Stanford University Press: Stanford, California.

Teicher, H. 1961. "Identifiability of Mixtures," *Annals of Mathematical Statistics,* 32:244-248.

Teicher, H. 1963. "Identifiability of Finite Mixtures," *Annals of Mathematical Statistics,* 34:1265-1269.

Titterington, D., A. Smith and U. Makov. 1985. *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons Ltd: Great Britain.

Traugott, M. W. and J. Katosh. 1979. "Response Validity in Surveys of Voting Behavior," *Public Opinion Quarterly*, 43:359-377.

Warner, S. 1965. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American Statistical Association*, 60:63-9.

Wattenberg, M. 1984. *The Decline of American Political Parties 1952-1980,* Harvard University Press: Cambridge, MA.

Welch, B. 1939. "Note on Discriminant Functions," *Biometrika*, 31:218-226.

Yakowitz, S. and Spragins, J. 1968. "On the Identifiability of Finite Mixtures," *Annals of Mathematical Statistics,* 39:209-214.

Zisk, B. 1987. *Money, Media, and the Grass Roots: State Ballot Issues and the Electoral Process,* Sage Publication: Newbury Park, CA.

Zdep, S. M., I. Rhodes, R. Schwartz, and M. Kilkenny. 1989. "The Validity of the Randomized Response Technique," in *Survey Research Methods,* E. Singer and S. Presser (eds.), University of Chicago Press: Chicago, IL.

Appendix A

California Statewide Propositions
General Election, November 1990

Prop	Title
124	Local Hospital Districts
125	Motor Vehicle Fuels Tax. Rail Transit Funding
126	Alcoholic Beverages. Taxes
127	Earthquake Safety. Property Tax Exclusion
128	Environment. Public Health. Bonds
129	Drug Enforcement, Prevention, Treatment, Prisons. Bonds
130	Forest Acquisition. Timber Harvesting Practices. Bond Act
131	Limits on Terms of Office. Ethics. Campaign Financing
132	Marine Resources
133	Drug Enforcement and Prevention. Taxes. Prison Terms
134	Alcohol Surtax
135	Pesticide Regulation
136	State, Local Taxation
137	Initiative and Referendum Process
138	Forestry Programs. Timber Harvesting Practices. Bond Act
139	Prison Inmate Labor. Tax Credit
140	Limits on Terms of Office, Legislators' Retirement, Legislative Operating Costs
141	Toxic Chemical Discharge. Public Agencies
142	Veterans' Bond Act of 1990
143	Higher Education Facilities Bond Act of November 1990
144	New Prison Construction Bond Act of 1990-B
145	California Housing Bond Act of 1990
146	School Facilities Bond Act of 1990
147	County Correctional Facility Capital Expenditure and Juvenile Facilities Bond Act of 1990
148	Water Resources Bond Act of 1990
149	California Park, Recreation, and Wildlife Enhancement Act
150	County Courthouse Facilities Capital Expenditure Bond Act
151	Child Care Facilities Financing Act of 1990

Proposition	% Abstain	% Yes	% No	% Invalid
124	12.36	43.56	43.92	0.15
125	11.36	43.68	44.81	0.15
126	9.55	37.14	53.06	0.24
127	10.67	59.87	29.16	0.30
128	7.07	38.40	53.95	0.58
129	9.66	28.75	61.18	0.40
130	7.54	47.87	43.95	0.63
131	9.97	33.66	56.06	0.32
132	12.24	48.66	38.90	0.20
133	10.43	33.76	55.62	0.20
134	8.33	31.02	60.46	0.19
135	10.18	27.03	62.58	0.21
136	10.60	42.12	47.05	0.23
137	12.43	41.77	45.64	0.16
138	9.47	26.27	64.07	0.20
139	10.60	44.62	44.61	0.16
140	9.91	44.02	45.71	0.35
141	14.22	46.23	39.95	0.20
142	12.52	55.04	32.29	0.15
143	12.83	45.64	41.36	0.16
144	11.47	39.53	48.92	0.09
145	13.63	44.83	41.36	0.17
146	12.06	47.43	40.31	0.20
147	15.28	35.17	49.45	0.14
148	14.16	41.47	44.01	0.10
149	12.09	45.84	41.91	0.16
150	14.71	27.16	57.95	0.17
151	12.69	46.99	40.14	0.17

Table 1: Aggregate Vote Returns

Type	Count	Frequency
All Yes	3863	.0022
All No	30,248	.0172
All Abstain	30,676	.0175
Other	1,690,693	.9631
Total	1,755,480	1.0000

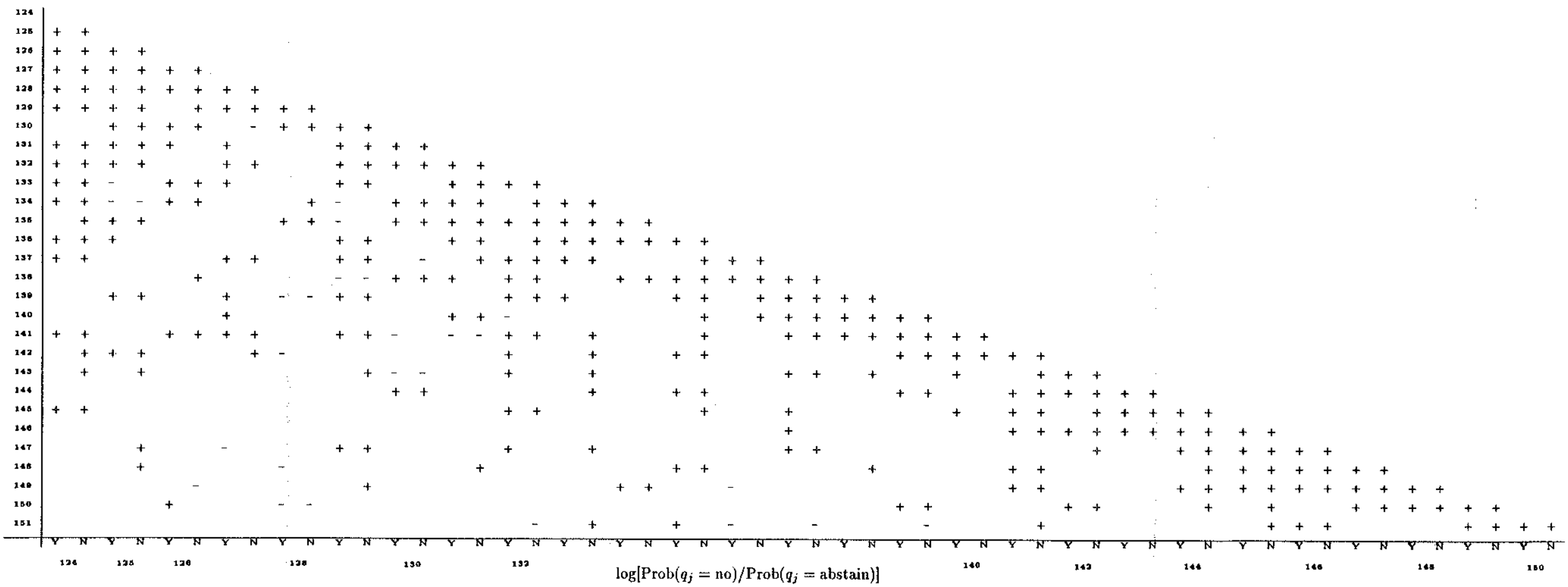
Table 2: Counts and Frequencies of Voter Types

Table 3
Conditional Logit Estimates

	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
124																																																						
125	+	+																																																				
126	+	+	+	+																																																		
127	+	+	+	+	+	+																																																
128			+	+	+	+	+	+																																														
129	+	+	+	+			+	+	+	+																																												
130			+	+	+	+		-	+	+	+	+																																										
131	+	+	+	+	+	+	+	+			+	+	+	+																																						
132	+	+	+	+			+	+	+		+	+	+	+	+	+																																				
133	+	+			+		+				+	+			+	+	+	+																																				
134	+	+		-	+	+			+	+	-		+		+	+			+	+																																		
135	+	+	+	+							-		+	+	+	+	+	+	+	+	+	+																																
136		+		+			+			+		+			+	+			+	+	+	+	+	+																														
137	+	+								+	+	+			+	+		+					+	+	+	+																												
138					+	+				+			+	+			+	+			+	+	+	+	+	+	+	+																										
139			+	+			+		-		+	+			+	+	+	+	+	+			+	+			+	+	+	+																								
140		+											-		+	+			+	+			+	+	+		+	+		+	+	+																						
141	+	+				+	+				+	+		-	-	-	+	+	+	+	+	+				+	+	+		+	+	+	+																					
142	+	+					+		-						+	+	+	+					+	+							+	+	+	+	+	+																		
143	+	+						+					-	-					+	+							+	+							+	+	+	+																
144								+	-		+								+				+								+	+			+	+	+	+	+	+														
145	+																	+				+	+	+			+								+	+	+		+	+	+	+												
146	+						+								+						+	+									+	+			+		+	+	+	+	+	+	+	+										
147			+								+																+	+							+								+	+	+	+	+	+						
148															+	+							+	+									-		+	+	+				+	+	+	+	+	+	+	+						
149			+																		+	+			-				+							+	+							+	+	+	+	+	+	+	+			
150									-									+										+				+					+	+			+	+					+	+	+	+	+	+		
151																											-	-									+						+		+	+					+	+	+	+
	124		125		126				128				130				132																140				142				144				146				148				150	

log[Prob(q_j = yes)/Prob(q_j = abstain)]

Table 4
Conditional Logit Estimates



Type	V_j	t-statistic	$\hat{\pi}_j$
All Yes	-4.8585	-34.82	.0073
All No	-3.1731	-53.66	.0391
All Abstain	-3.8666	-39.58	.0195

Table 5: Mixture Parameter (π_j) Estimates