DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

A STATISTICAL MODEL FOR MULTIPARTY ELECTORAL DATA

Jonathan N. Katz
California Institute of Technology

Gary King
Harvard University

SOCIAL SCIENCE WORKING PAPER 1005

August 1997
Revised March 1998

A Statistical Model for Multiparty Electoral Data*

Jonathan N. Katz†   Gary King‡

Abstract

We propose an internally consistent and comprehensive statistical model for analyzing multiparty, district-level aggregate election data. This model can be used to explain or predict how the geographic distribution of electoral results depends upon economic conditions, neighborhood ethnic compositions, campaign spending, and other features of the election campaign or characteristics of the aggregate areas. We also provide several new graphical representations for help in data exploration, model evaluation, and substantive interpretation.

Although the model applies more generally, we use it to help resolve an important controversy over the size of and trend in the electoral advantage of incumbency in Great Britain. Contrary to previous analyses, which are all based on measures now known to be biased, we demonstrate that the incumbency advantage is small but politically meaningful. We also find that it differs substantially across the parties, about half a percent for the Conservatives, 1% for the Labor Party, and 3% for the Liberal party and its successors. Also contrary to previous research, we show that these effects have not grown in recent years. Finally, we are able to estimate from which party each party's incumbency advantage is predominantly drawn.

*An earlier version of this paper was presented at the annual meetings of the Midwest Political Science Association, Chicago, Illinois, April 1997. Our thanks to Jim Alt, Larry Bartels, Neal Beck, M.F. Fuller, Dave Grether, Mike Herron, James Honaker, Ken Scheve, Ken Shepsle, Bob Sherman for helpful suggestions; Josh Tucker for useful suggestions and research assistance; Gary Cox for providing his British election data; and Selina Chen for help in collecting additional British data. Burt Monroe saw the virtues of using the compositional data analysis literature at essentially the same time as we did, and we appreciate his comments. For research support, Jonathan N. Katz thanks the Haynes Foundation, and Gary King thanks the U.S. National Science Foundation (grant SBR-932121).

†Division of the Humanities and Social Sciences, Pasadena, CA 91125; JKatz@golem.Caltech.Edu, http://JKatz.Caltech.Edu, (626) 395-4401.

‡Department of Government, Littauer Center North Yard, Cambridge MA 02138; King@Harvard.Edu, http://GKing.Harvard.Edu, (617) 495-2027.

1 Introduction

We propose the first internally consistent and comprehensive statistical model for analyzing multiparty, district-level aggregate election data. Our model can be applied directly to explain or predict how the geographic distribution of electoral results depends upon economic conditions, neighborhood ethnic compositions, campaign spending, or other features of the election campaign or characteristics of the aggregate areas. We also provide several new graphical representations for help in data exploration, model evaluation, and substantive interpretation.[1]

Our general model is intended to address three serious lacunae in the study of comparative politics. First, most literatures focusing on non-American elections are dominated by survey data alone rather than also including studies of real election results.[2] Survey research has enormous advantages for studying individual-level preferences but, as analyses of random selections of isolated individuals from unknown geographical locations, they necessarily miss much of electoral politics. As such, they are often best complemented with studies of aggregate electoral returns.

Second, with surprisingly few exceptions (e.g., Bellucci, 1984, 1991; Rattinger, 1991; Slider, 1994; and Conford et al., 1995), electoral analyses in comparative politics based on real election returns use national rather than regional, district, or precinct-level electoral data.[3] This approach has the advantage of allowing more countries to be included in the analysis without much data collection effort, but it also has serious disadvantages. National-level studies prevent researchers from learning where votes come from and why, and they generally result in studies based on small numbers of observations and little variation on many relevant dimensions. Studies of post-war OECD countries usually contain only about a dozen observations (see Paldam, 1991: 18, Table 1), and analyses of former communist countries could include only three or four elections. This is often insufficient information with which to parse out many of the interesting effects, and it ignores the substantial information content in the often vast differences across different regions of a country.

Finally, the vast majority of electoral studies in multiparty democracies artificially dichotomize the electoral system into a pseudo-two party contest. Researchers analyze the

[1] Our model can also be used to evaluate features of electoral systems, such as whether the districting system is fair to all the political parties and electorally responsive, although we leave details of this task to a future paper.

[2] A few examples of the good survey analyses conducted in multiparty democracies include those in England (e.g., Goodhart and Bhansali, 1970), Mexico (Dominguez and McCann, 1996), Poland (Przeworski, 1996), Peru (Stokes, 1996), Russia (White et al., 1997), Denmark (Miller and Listhaug, 1985), Italy (Bellucci, 1985), West Germany (Frey and Schneider, 1980), multi-country studies (Lewis-Beck, 1988), among many others.

[3] Such national-level studies have used data from France (Rosa and Amson, 1976), Japan (Inoguchi, 1980), England (Whiteley, 1980), Italy (Bellucci, 1984), time series of cross-sectional data from multiple countries (Powell and Whitten, 1993; Paldam, 1986; Host and Paldam, 1990; Lewis-Beck and Mitchell, 1990; Paldam, 1991), single cross-sections of multiple countries (Lewis-Beck and Bellucci, 1982), among many others.

vote for the incumbent party vs. all others grouped together, or the vote for a particular group, such as left-wing parties, vs. the combination of all the others. This procedure has the advantage of enabling the use of standard statistical methods, but since these methods were designed for the study of two-party systems (largely in U.S. data), two serious problems result: bias and information loss. The procedure is biased whenever all parties do not run candidates in all election districts. For example, even when the governing party contests every election, different numbers of parties composing the "other" category will generally have large effects on a variable like the "percent of the vote for the governing party." Since the vote a party expects to receive will normally be related to whether they run a candidate, the observed variable will systematically overstate the true underlying support for the governing party when this support is highest. This, and other similar problems, can combine to severely bias inferences based on such data.

Moreover, even if partially contested elections happen to cause no bias in a particular case, important information, critical to comparative politics, is always lost by these methods. For example, when the economic pain caused by pro-market reforms in post-communist countries causes the reformers to be thrown out of office (Przeworski, 1991; Haggard and Webb, 1994; Haggard and Kaufmann, 1995), or when the increasing salience of ethnic divisions upsets the political order (Offe, 1992; Horowitz, 1985, 1993; Tucker, 1996), which parties benefit? How do the electoral fortunes of each of the parties depend on the degree of economic hardship or ethnic divisions? Answering these questions about multiparty systems require statistical models that permit multiparty outcomes. Shoehorning a complex multiparty democracy into a fake "two-party" system in order to perform an analysis that looks like those conducted in American politics takes the wrong lessons from that subfield. Making methodological decisions merely to accommodate the requirements of familiar statistical methods risks missing the most distinctive and interesting aspects of the electoral system under analysis. The bottom line is that multiparty systems require the development of multiparty statistical models. It would appear that much substantive knowledge can be gained, and bias reduced, by designing models of electoral systems with the special features of these systems in mind.

Although we intend our model to be applicable to a wide variety of multiparty electoral data, we apply it here to resolve one important scholarly controversy: the size of and trend in the electoral advantage of incumbency in the United Kingdom. For decades, the conventional wisdom has been that U.K. incumbency advantage is small to nonexistent and not increasing. However, this conclusion has come under strong attack recently by researchers whose results seem to show that incumbency advantage is moderate to large, and growing fast. Unfortunately, it turns out that all estimates given in the literature are based on measures now known to be biased. In partial agreement and disagreement with the substantive results from both sides, we demonstrate that the incumbency advantage is small but markedly different for each of Britain's three major parties. Our methods also provide information that others have not attempted to estimate, such as from precisely which party each incumbent's advantage comes at the expense of. In part because the study of incumbency advantage in the United States was so greatly enhanced by a quarter century of scholarly work increasing the precision and accuracy of estimates in this two-party system, we hope a similar gain will result in electoral studies of multiparty democracies, such as the United Kingdom.

From a methodological perspective, analyses of aggregate electoral data fall into two fundamental categories that should be carefully distinguished — contextual effects and ecological inferences. Research questions about the relationships among aggregate variables require a model of contextual effects, such as that offered here. In contrast, research questions about the characteristics of the individuals who make up aggregate electoral data require ecological inferences and the special models designed for this purpose (see King, 1997). For example, a study of the effect of having a college in town on the vote for more liberal parties is a contextual effect, for which the model we propose is directly useful. In contrast, the question of whether individual college students vote for more liberal political parties is an ecological inference. Our statistical model applies to questions at the district level, such as incumbency advantage estimates or predicting which candidate will win. Formal models of individual behavior might be of interest for some purposes, but they are not always necessary in cases like these. For much of our discussion, we make the common assumption that candidates are strategic and well informed. Our working (testable) assumption about voters is that, conditional on the candidates, their voting behavior follows regular patterns of some sort.

In what follows, we briefly describe the substantive problem (Section 2), discuss the characteristics of multiparty electoral data (Section 3), summarize the problems that need to be fixed in order to develop a general statistical model (Section 4), introduce a simple version of the model for cases when all parties contest (Section 5), discuss assumptions to deal with partially contested elections (Section 6), and show how to estimate (Section 7) the model and compute quantities of interest (Section 8) from these estimates. We present substantive results as the model is developed.

2 Incumbency Advantage in Great Britain

Until the 1980s, scholars generally agreed that British elections were decided by national and not local forces. The electoral advantage of incumbency — what is known in the U.S. and the methodology literature as "incumbency advantage" and in the U.K. as "the personal vote" — was thought to be essentially nonexistent in Britain.[4] For example, Butler and Stokes (1969: 6) repeatedly emphasize "the importance of national political issues and events as opposed to more local influences on the choice of the individual elector." They even went so far as to conclude (p. 8) that "So important are the [national] parties in giving meaning to contests in the individual parliamentary constituencies in Britain that for many voters candidates have no identity other then their partisan one" (see also Butler and Kavanagh, 1980: 292).

[4]The concept of "the personal vote" in Britain includes any local, candidate-specific effect, but for empirical analyses, it is treated synonymously with incumbency advantage. In the U.S. "the personal vote" is often considered to be the fraction of the incumbent's advantage attributable to the person rather than the party.

The few numerical estimates of incumbency advantage in Britain come from a newer literature that contradicts this conventional wisdom. For example, Curtice and Steed (1980, 1983) find the "sophomore surge" for Labour — the average difference in the vote for Labour in open seats it wins and the vote in the subsequent election for the now-incumbent Labour party candidate — is about 1500 votes (about 3.8%) in the 1979 and 1983 elections. Norton and Wood (1990) modify sophomore surge by correcting for regional swings and find a surge of 1.6% for the Conservatives and a remarkable 7.8% for Labour. Finally, Wood and Norton (1992), along with the prominent survey-based analyses of Cain, Ferejohn, and Fiorina (1987), also strongly argue for the proposition that the personal vote is increasing.

Although the idea of sophomore surge, on which most measures are based, is very intuitive, Gelman and King (1990) proved that it gives biased estimates of the causal effect of incumbency. In fact, in the U.K., the problem may actually be worse than in two-party systems: since most British elections have very few sophomores (usually only about two dozen), the measure discards more than 95% of the district observations and is therefore exceptionally inefficient. Moreover, the measure is usually applied without controls and without any feature of the statistical model that recognizes that the system being analyzed has more than two parties.

Thus, we have on one side the conventional wisdom, based on many years of traditional analyses, that incumbents have no electoral advantage. On the other side, we have a growing systematic quantitative literature which argues that the incumbency advantage is moderate to large and steadily growing. We hope to resolve this scholarly dispute.

Our data for this paper includes constituency-level election results from England for 1959 to 1992. We also have more limited data from the 1955 elections, which we use whenever possible. Geographic districts in England are called "constituencies," but we use the two words interchangeably because our model applies more generally. For convenience, we usually refer to the Liberal Party and the Liberal Party alliance with the Social Democratic Party more simply as the Alliance.[5]

3 Characteristics of Multiparty Data

Let V_{ij} denote the proportion of the <u>v</u>ote (the underline denotes our mnemonic labeling convention) in district i ($i = 1, \dots, n$), for party j ($j = 1, \dots, J$). Two fundamental features of multiparty voting data are that each proportion falls within the unit interval

$$V_{ij} \in [0, 1] \qquad \text{for all } i \text{ and } j \tag{1}$$

[5]We began with the dataset "British General Elections, 1955–1992 (version 8 August 1993), constructed by D.F.L. Dorling of University of Newcastle, extracted data from England, updated it, and added information on incumbency status. The number of observations in our data for each election year is 1955: 460, 1959: 460, 1964: 455, 1966: 488, 1970: 471, 1974(Feb): 463, 1974(Oct): 491, 1979: 467, 1983: 491, 1987: 522, and 1992: 521.

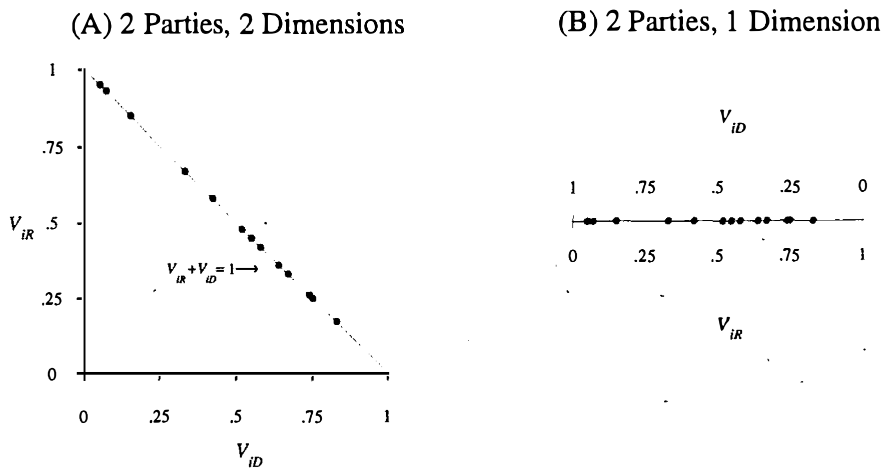

Figure 1: *The Simplex for Two Parties: This figure explains graphically how to reduce two vote variables to one dimension. Graph (A) portrays the familiar relationship between the vote for the Republicans (V_{iR}) and Democrats (V_{iD}) in the U.S. two-party system; because of the constraints of Equations 1 and 2, all points fall on a line segment and can be portrayed more simply as Graph (B).*

and the set of vote proportions for all the parties in a district sum to one:

$$\sum_{j=1}^{J} V_{ij} = 1 \qquad \text{for all } i \tag{2}$$

Thus, an important criterion of a good (and logically possible) statistical model of multiparty voting data is that it satisfy the constraints in Equations 1 and 2. Variables that meet these constraints fall in a region generally referred to as the *simplex*.

We now illustrate this simplex sample space graphically for the two and then the three party case. For each case, we apply a simple trick to reduce the number of dimensions required, making the graphical presentation more manageable, and ultimately informative, without losing information. The graphic version of these relationships will also be useful for exploring the data, and understanding the model fit.

For the two party example, we use V_{iD} for the Democratic, and V_{iR} for the Republican, shares of the vote in district i for candidates for the U.S. House of Representatives. Obviously, we can easily represent both variables by just one, say V_{iD}, since the other is merely $V_{iR} = 1 - V_{iD}$. Graph (A) in Figure 1 plots V_{iD} by V_{iR}. Because of the constraint in Equation 2, all district vote fractions fall on a single line segment and, due to the constraint in Equation 1, the line ends at the axes. Thus, all the points in the two-dimensional plane in Graph (A) of Figure 1 fall on a simpler one-dimensional line segment. Presenting this line segment in Graph (B) reduces the problem from two to one dimension without losing any information.

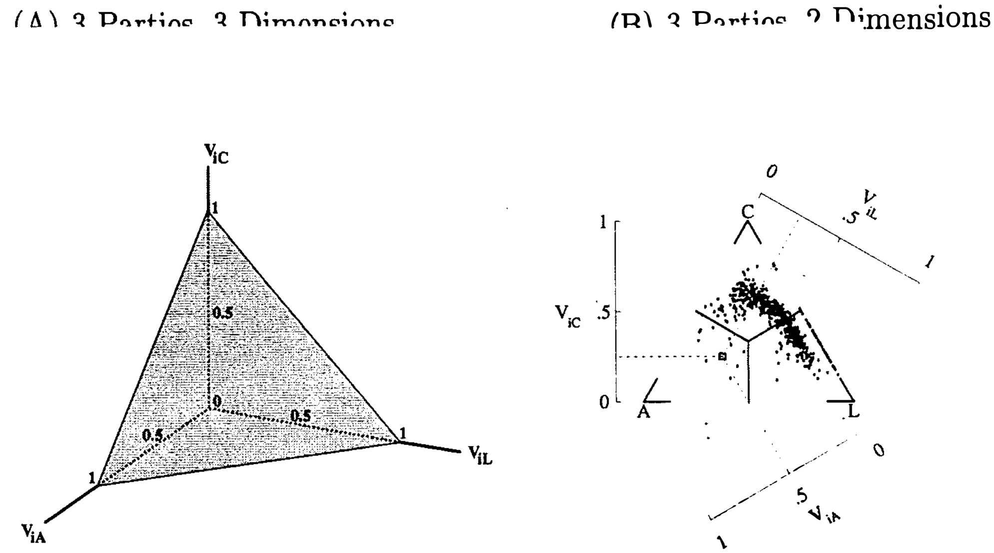

Figure 2: *The Simplex For Three Parties: In a manner analogous to Figure 1, this figure reduces three vote variables to two dimensions. Graph (A) portrays the relationship among the votes for the Conservative (V_{iC}), Labour (V_{iL}), and Alliance (V_{iA}) parties; because of the constraints of Equations 1 and 2, all points fall on an equilateral triangle that is the intersection of a plane with the three dimensional figure. Graph (B) portrays this more simply in two dimensions in a version of what is known as a "ternary diagram." Values of the three variables can be read off by where the dots fall perpendicular to the three numbered axes. The little square point (with dotted lines referencing the axes) is the example discussed in the text.*

Figure 2 provides analogous information for three parties as Figure 1 did for two. Graph (A) in Figure 2 plots in three dimensions the three variables from the British electoral system, V_{iC}, V_{iL}, and V_{iA}, for the Conservative, Labour, and Alliance vote proportions, respectively. The constraints in Equation 2 imply that valid points must fall on the plane cutting through the three-dimensional space. The constraints in Equation 1 require this plane to end at each of the three axes. The resulting area that satisfies both constraints is the equilateral triangle we have shaded in Graph (A).

Because all the points in three dimensions fall in a two-dimensional area, we can save space by presenting the triangle alone in two dimensions, which we do in Figure 2, Graph (B). This graph is a version of what is known in the literature as a *ternary diagram*, although we have added several new features. In this triangle, each dot fully characterizes a single constituency result from the 1979 British election. Roughly speaking, the closer a dot is to a vertex (with a party's label, C, L, or A), the higher the vote total for that party; more precisely, a vote total for a party equals the perpendicular distance from the side of the triangle opposite to the labeled vertex, as calibrated on the scales we have added. That is, the vertical position of the dots in the figure indicate the value of V_{iC}, as indicated on the scale on the left. As a dot falls farther from the side opposite the "L" vertex, the larger the V_{iL} variable (as can be seen by comparing the point to the scale at the top).

For example, in addition to the real data, we have added one (hypothetical) election result as a small box in the bottom left of the graph. In order to clarify how to read the voting results in this district, we added dashed lines connecting this point to the three axes. In this district, the Conservatives received 25% of the vote, as can be seen by following the dashed line from the box to the left axis that calibrates V_{iC}. The dashed line traces the shortest distance from the axis to the point, or in other words a line that is perpendicular to the axis. The same district also gave 25% of its vote to the Labour party (see the dashed line that heads northward to the V_{iL} axis) and half of its votes to the Alliance (as can be determined from the dashed line that heads down to the right to meet the V_{iA} axis). The electoral results for all the real districts, represented by dots, can also be read by tracing out perpendicular lines to each axis. With all this precision available when needed, it is still worth remembering the easier rough way to interpret these ternary diagrams: the closer a point is to a vertex of the triangle, the more votes that district gave to the party whose label appears there.

We have removed most of the sides of the triangle in order to make visible districts with zero votes for one of the parties. For example, districts uncontested by the Alliance fall on the right side of the triangle, where the bottom axis reveals that $V_{iA} = 0$. Substantively, these partially contested districts appear to be generated by a different process than the mass of (fully contested) districts that fall inside the triangle. This can be seen since the distribution of points does not gradually get smaller (or larger) as it approaches the side of the triangle; instead, there appears to be an area without dots, indicating every party that merely appears on the ballot receives at least 15-20% of the vote.

We have also added lines that divide the triangle into thirds. We call these *win lines*, since they indicate which party wins, depending on which region a point falls in. For example, if a point falls in the region at the top of the graph, the Conservatives win a plurality of the votes and (by the electoral rules) the seat for that constituency. Points that fall within the left region are wins for the Alliance, and the right go to the Labour Party. The same logic applies to multiparty elections with $J > 3$ parties, even though graphical displays become more unwieldy.

4 Problems to Fix

Standard regression-type models applied to multiparty electoral data usually generate nonsensical results. For example, one common approach is to use the vote for each party as a dependent variable (fraction for the Conservatives, fraction for the Labour party, etc.), and to regress each on a set of explanatory variables. These J regressions are run separately, or via a "seemingly unrelated" system of equations. Since neither constraint from Equations 1 and 2 is satisfied, this approach generally fails to give results that are sensible. That is, the model often implies that some parties will get fewer than zero votes, or that the sum of votes for all parties will be greater or less than 100%. Moreover, even when point predictions happen to fall within the constraints of the simplex, the full probabilistic implications of the model are virtually always logically impossible, as some of the predictive density always falls outside the simplex. Some of the few who recognize this problem transform V_{ij} to an unbounded scale (separately for each party j), such as with a logistic function, and then apply separate or seeming unrelated regressions, but this too is insufficient: the results will satisfy Equation 1 but not Equation 2. Similarly, running only $J-1$ regressions and computing the predictions for V_{iJ} from the others satisfies Equation 2 but not Equation 1. Various other ad hoc approaches can be taken to fix different parts of the problem but, especially because computing most quantities of interest require the full probabilistic model, we decided to pursue a more general approach.

The model we develop can be considered a generalization of two independent lines of statistical research. The first line of research includes models for "compositional data" (Aitchison, 1986), a term that describes data sets with multiple outcome variables that sum to unity for each observation. A few of the many examples of compositional data from other fields include soil samples in geology (with measurements of the fractions of sand, silt, and clay), rock samples in geochemistry (with fractions of alkali, Fe_2O_3, and MgO), and blood measurements in biology (proportions of white blood cell types measured for each patient). Compositional data are also common in political science and economics (as in multiparty voting data, the allocation of ministerial portfolios among political parties, trade flows or international conflict directed from each nation to several others, or proportions of budget expenditures in each of several categories), but researchers in these fields have not taken advantage of the connection to this more general statistical approach. This is unfortunate since compositional data would seem to be closely related to the raison d'etre of political science research: if politics is the authoritative allocation of resources, then fractions of resources received by each group is exactly compositional data.

The key contribution of this literature is statistical models that allow only possible outcomes to occur. That is, predictions or simulations from such a model satisfy Equations 1 and 2 or, in other words, have positive density only over the simplex. The most influential models of compositional data are due primarily to Aitchison (1986), who criticized the earlier models based on Dirchlet distributions, since those models require the "compositions" (votes for each party in our application) to be independent. Aitchison avoided this unrealistic assumption by applying the normal distribution to the log-ratios of the individual components. This procedure starts with the multivariate normal fit to the unconstrained real plane and then maps it into the simplex via the multivariate logistic transformation. This works in the same way as, for example, the log-normal maps the real number line onto the positive real numbers.

The second statistically similar, but independent, line of research that we generalize are models of votes and seats for two-party systems (King and Browning, 1987; King, 1989b; King and Gelman, 1991; and Gelman and King, 1991), and for multiparty systems (by King 1990). Like compositional models, some of these votes and seats models also transform votes (in different ways) using logistic transformations and then stochastically model the transformed variables. The resulting statistical models differ in a variety of ways, but they also constrain the result to the proper sample space so that Equations 1 and 2 are satisfied.

Unfortunately, the models from neither line of research will work without modification for multiparty voting data. For one, our extensive evaluations of the assumptions of normality underlying the models proposed for compositional data indicate that they do not fit real election data. For another, these models also do not capture a fundamental feature of voting data: the pattern of "missing data" that occurs when one or more parties receive zero votes in a district, as occurs when no candidate from a party stands for election in a district. Zero vote totals in electoral data constitute politically crucial information and must therefore be treated very differently from examples where zeros are considered to be missing due to slight measurement error (as when instruments for measuring the compositions of soil samples miss the always-present traces of some elements). The models discussed above for analyses of seats and votes fit two-party systems well and (for King, 1990) they can fit multiparty data on seats given votes, and some of these include special features for uncontested districts, but they are not directly applicable to explaining or predicting multiparty electoral data.

Thus, a proper model of multiparty voting data must have the following special features. It must have positive density only over the simplex and must use a distribution more flexible than the multivariate normal, which does not fit real voting data. It must also allow covariates (explanatory variables). Section 5 provides this basic model for fully contested district elections. The model must also provide special features to deal with uncontested and partially contested seats, which we do in Section 6. The complete likelihood function is given in Section 7.

Finally, a proper model must allow for estimates of precisely the quantities of scholarly interest, and these may differ across applications. That is, we should not have to teach readers to interpret the arcane results of statistical models; rather, the models should be modified to produce results in the form of most natural interest to substantively-oriented political scientists. For example, the raw coefficients estimated by models of compositional data are not the quantities of interest for any political, or indeed for virtually any non-political, application. We use methods of simulation, described in Section 8, to compute estimates of a wide range of theoretically interesting quantities. These methods of computing quantities of interest are critical to political science applications of this new model. We also believe the methods will enable those in other scholarly disciplines, who use somewhat related models for very different purposes, to compute numerical quantities of more interest to their research than the usual results of compositional data models.

5 The Basic Model for Fully Contested Elections

In this section, we only consider district elections that all parties contest and in which every party gets at least one vote: $V_{ij} \in (0,1)$ for all i and j. Since, in practice, no officially registered candidate that appears on the ballot ever gets fewer than 15–20% of the vote in our data, the only real assumption here is that all parties contest all district elections. We generalize this model to deal with partially contested elections in Section 6.

Let $V_i = (V_{i1}, \ldots, V_{i(J-1)})$ be a $(J-1) \times 1$ vector for each district i $(i = 1, \ldots, n)$. This vector contains all the information in the individual vote fractions since the votes for party J can be computed deterministically from the others:

$$V_{iJ} = 1 - \sum_{j=1}^{J-1} V_{ij} \tag{3}$$

The model we are about to propose is "symmetric," in the sense that changing the party labeled J does not affect anything of substantive importance.

Aitchison (1986) proposes that compositional data like V_i be modeled with his additive logistic normal distribution. This distribution can be formed as follows. First let Y_i be the vector of $J-1$ log-ratios $Y_{ij} = \ln(V_{ij}/V_{iJ})$, for party j $(j = 1 \ldots, J-1)$ relative to party J. Then assume that the $(J-1) \times 1$ vector $Y_i = (Y_{i1}, \ldots, Y_{i(J-1)})$ is multivariate normal with mean vector μ and variance matrix Σ. To get to the observed votes, use the multivariate logistic transformation:

$$V_{ij} = \frac{\exp(Y_{ij})}{1 + \sum_{j=1}^{J-1} \exp(Y_{ij})} \tag{4}$$

Although compositional data analysts have found this specification to be useful for their applications, we demonstrate below that it is inappropriate for multiparty voting data. In our data, a majority of districts tend to be more highly clustered, and a minority are much more widely dispersed, than the multivariate normal implies.

Political scientists modeling seats and votes have avoided this distributional problem by a combination of mixtures of independent normals and appropriately chosen covariates, but these fixes are insufficient for a general approach to multiparty voting data.

We now derive a new model that solves these problems. We label the distribution the *additive logistic Student t* (LT) distribution, which we demonstrate is superior when used to fit political data than the additive logistic normal, which it includes as a limiting special case. To derive the LT distribution, first let Y_i be multivariate Student t (Johnson and Kotz, 1972) and then apply the transformation in Equation 4:

$$\begin{aligned}
\mathrm{P}(V_i|\mu, \Sigma, \nu) =& \mathrm{LT}(V_i|\mu, \Sigma, \nu) \\
=& \mathrm{T}\left[\ln(V_i/V_{iJ})|\mu_i, \Sigma\right] / \textstyle\prod_{j=1}^{J-1} V_{ij} \\
=& \frac{\Gamma[(\nu + J - 1)/2]|\Sigma|^{-1/2}}{\Gamma(\nu/2)\nu^{(J-1)/2}\pi^{(J-1)/2}\left(\prod_{j=1}^{J-1} V_{ij}\right)} \\
& \times \left(1 + \frac{1}{\nu}(Y_i - \mu_i)'\Sigma^{-1}(Y_i - \mu_i)\right)^{-(\nu+J-1)/2}
\end{aligned} \tag{5}$$

where the extra factor in the denominator is the Jacobian of the transformation (required when creating a new distribution from an existing one through a deterministic transformation), the expected value and variance of Y_{ij} are μ_{ij} and $\Sigma\nu/(\nu - 2)$, and ν ($\nu > 0$) is the "degrees of freedom" parameter. The $(J - 1) \times (J - 1)$ parameter Σ is known as the "scatter" matrix. This distribution happens to be equivalent to the predictive distribution, under certain conditions, when using the additive logistic normal (see Aitchison, 1986: 174).

This model differs from the additive logistic normal when $\nu < \infty$, and it differs more the smaller ν is. We find in practice that our estimates of ν are fairly small and thus the LT distribution differs significantly from the additive logistic normal.[6] For example, Figure 3 gives two ternary diagrams with normal (for the left graph) and t (for the right graph) confidence regions fit to real electoral data (from 1970). For both, the inner loop is the 50% confidence region, and, if the model is appropriate, 50% of the points should fall within it. In fact, 66.7% fall within the normal-based region, whereas 48.5% of the points fall within the t-based region. A similar situation, but slightly less extreme, holds for the 95% confidence region, which is the outer loop in both graphs. (Because of the large number of constituencies, the figure is more useful for understanding the differences in how the two models fit these data, and the nature of confidence regions on the simplex, rather than making it easy to count points within each region.) This demonstrates clearly the advantage of the t distribution for British electoral data.[7]

When ν is sufficiently large, the normal and t distributions are identical. This means that our generalization has great potential benefits, because it fits a much wider range of data more common in multiparty democracies, and it is also essentially costless (i.e., except for the trivial efficiency loss caused by estimating the extra degrees of freedom parameter). Given this risk profile, there seems little reason to not use this more general model.

[6]So that we consider only cases where the moments exist on the logistic scale, we impose the technical restriction that $\nu > 2$. This assumption, while not necessary for our model or estimator (since the moments of the additive logistic t are always finite), does make estimation and simulation simpler. Given that our estimates of ν stay far from the boundary (even when permitted to do otherwise) this technical assumption is unambiguously supported by our data.

[7]The fit of the model could be closer to a normal after conditioning on explanatory variables, but our studies indicate that this is not usually the case with our multiparty data. We present the simple case in Figure 3 without covariates for ease of presentation.

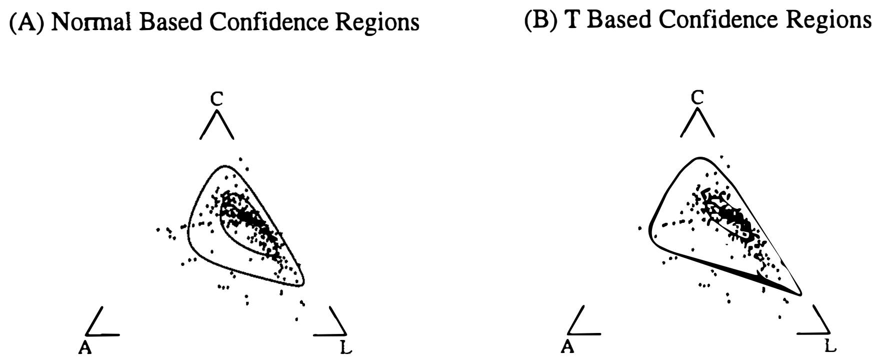

Figure 3: *The Fit of the Logistic Normal and Logistic t distributions: Both graphs give a ternary diagram for 1970 British House of Commons electoral data, with uncontested districts deleted, and 50% and 95% confidence regions based on the additive logistic normal in (A) and additive logistic t in (B). The better fit to the t distribution is indicated by the approximately 50% and 95% of the constituencies that fall within the 50% and 95% confidence region, respectively, for the t distribution, but only 64% and 91% for the normal. Note also how the inner region is much narrower, and the outer region is wider, for the t than the normal.*

The generalization that the additive logistic t provides would be traditionally described (by using the general textbook description for t-based distributions) as allowing for "fatter tails" — a small number of constituencies surprisingly (according to the normal) far from the center of the distribution. This description is accurate, but a perhaps more informative characterization is the other half of the story: when ν is small, most of the constituencies are surprisingly (according to the normal) heavily clustered together (compare the inner confidence region in the two graphs in Figure 3). That is, what the traditional description of t-based distributions miss is that when ν is small, a t distribution with the same variance as a normal has both fat tails and heavier cluster around the mode. The two features must exist simultaneously to counterbalance each other, in order that the result is a proper distribution. Our reason for also emphasizing the heavy cluster is that this describes more of the points than focusing on the relatively small number of "outliers" in the tails.

As can be seen by the counts of districts within the confidence regions in Figure 3, the additive logistic t model fits the data better than the additive logistic normal. The substantive reason is that most constituencies in England have vote fractions that are very similar to one another, but a smaller set of constituencies are quite far from this main cluster.

We also present, in Figure 4, summaries of the fit of the two distributions for all the elections in our data. For each election year, the figure gives the percentage of districts that fall within the 50%, 80%, and 95% confidence regions. Graph (A) shows that for the

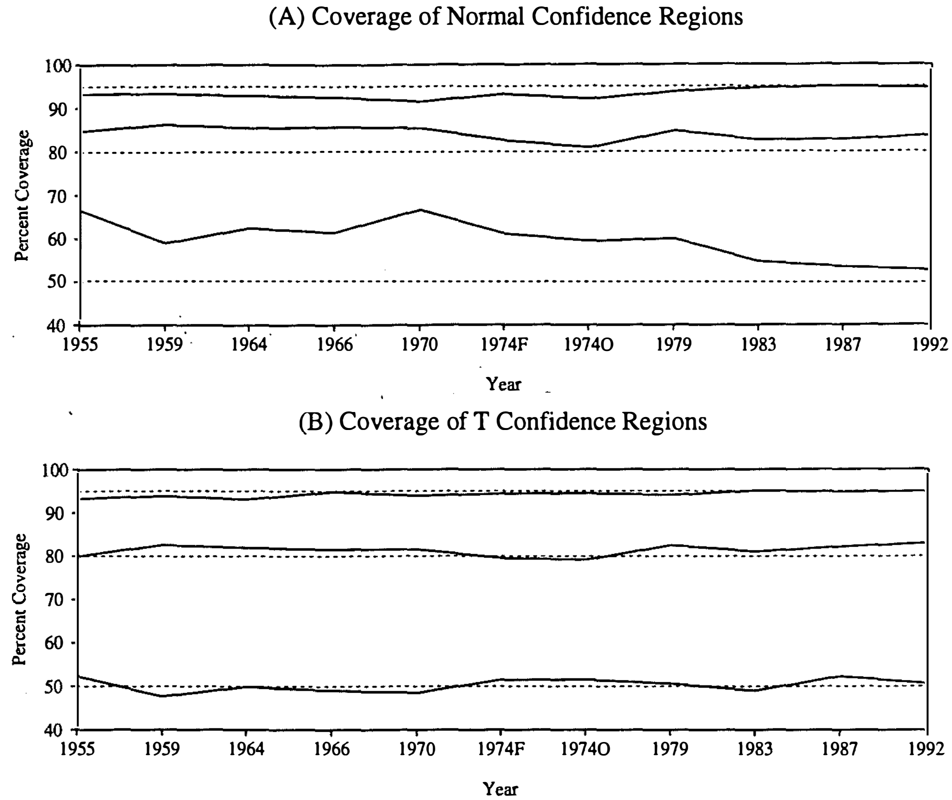

Figure 4: *Confidence Region Coverage: These graphs summarize the fit of the additive logistic normal (A) and additive logistic t (B) distributions for all U.K. elections in our data set. For each election, the solid lines mark the percent coverage for the 50%, 80%, and 95% confidence regions (where dotted lines are drawn). The better fit of the t distribution is indicated by the actual number of constituencies within each region (indicated by the solid line) staying much closer to the dotted line for the t than for the normal.*

additive logistic normal model, the actual fraction of points within each of these regions (indicated by solid lines) vary quite a distance from the theoretically correct (straight dotted) lines. (For visual clarity, the solid lines connect the points at the elections, where the estimation was actually conducted.) In contrast, the actual and theoretical values are very close for the additive logistic t, portrayed in graph (B). The normal seems to fit better for more recent elections than it once did, but there is no reason to think that this trend will continue.

For applications, we let the means of the log-ratios be linear functions of vectors of explanatory variables:

$$\mu_{ij} = X_{ij}\beta_j \tag{6}$$

where X_{ij} is a $p_j \times 1$ vector of explanatory variables, and β_j are parameters to be estimated. For most applications the explantory variables will be the same for all j, but this is not required. The parameters β_j, Σ, and ν are of little direct interest, but we show in Section 8 how to compute quantities of interest from them.

6 Assumptions for Partly Contested Districts

We now introduce methods of generalizing the basic model in Section 5 to allow for districts where some parties do not contest the outcome.

We follow King and Gelman (1991) by setting as the goal of estimation the *effective vote* — values of V_{ij} that we would observe if all J parties contested the election in district i. In districts with all parties contesting, the effective vote is the observed vote. In partly contested districts, the effective vote for all parties is unobserved but can be estimated. (That is, in districts where any party chooses not to contest, we lose information about the effective vote proportion for all parties since those which contest might get different vote fractions if they faced more competition.)

The effective vote concept covers all "national" political parties, even if they do not contest all elections. We distinguish regional parties, and do not try to estimate strained counterfactuals such as what would happen if, for example, the Scottish nationalists ran in English constituencies. Regional parties are easy to include in our model, although for expository purposes we skip this issue here.

In order to analyze the effective vote in not-fully-contested districts, some assumptions must be made. We introduce here several assumptions designed for electoral systems where the candidates and parties decide for themselves whether to contest a district election.[8] A reasonable assumption under these circumstances is that a party which chooses not to contest would not have won if it had nominated a candidate. After all, if they would have won, they probably would have nominated someone in the first place. Even if this assumption were false, it is unlikely that any statistical analyst could come up with a more realistic assumption than the non-contesting party is effectively able to do for us.

It also seems highly likely that the party not contesting the election would have received fewer votes, if it had contested, than the parties that did nominate candidates, and so we expand the assumption to this this more encompassing version. We recognize that this more encompassing assumption may occasionally be wrong. In other words, it is conceivable that the non-contesting party, if it ran, might get more votes (and yet still lose) than one of the parties that chose to run. However, even if this assumption were violated, the degree of violation would very rarely be large enough to make a noticeable

[8]Other assumptions would be necessary when, for example, a pro-government election commission prevents opposition parties from entering a race because they might win, as occurs in some of the new Eastern European "democracies". Similarly, when a small party, in a deal with a larger party, agrees to not contest in certain areas (as in the recent New Zealand elections), these assumptions would not hold. In these cases, our model could be modified accordingly. In all cases, scholars should tune the model assumptions to what we know about the details of party politics.

substantive difference. Moreover, the alternative possible assumptions are more arbitrary and would be difficult to justify.

Other assumptions could be chosen based on models of candidate entry and exit for different electoral systems, or by focusing on different features of the British electoral system (such as the requirement that candidates lose a monetary deposit if they do not receive a certain fraction of the vote). Our methods for deriving the model below under our chosen assumptions can be easily modified to handle these alternatives.

If covariates that predict which parties contest in each constituency are available, they can be used in interactions with indicator variables that code for the patterns of uncontestedness to avoid assumptions about parameter equivalence between district elections that are fully and partly contested. In most cases, these variables will be useful but not necessary. An alternative approach would be to develop full-blown models that predict which parties contest in each district as separate equations. Although future researchers may wish to consider this alternative approach, we do not pursue it because it is unnecessary, would make the model less robust, and, would require data that are very hard to come by in most applications.

We make no assumption analogous to "independence of irrelevant alternatives," as is sometimes necessary for individual-level, survey-based statistical models of multiparty voter choice (see Alvarez and Nagler, forthcoming). That is, our assumptions, and the model built from them, allow the entry or exit of a party into a district election contest to affect the relative vote totals of the parties already in the race.

7 Estimation

In this section, we propose methods of estimating the parameters of the model, $\beta \equiv \{\beta_j\}$, Σ, and ν. Section 8 explains how to compute quantities of interest given these results.

If all districts were contested by all parties, we could estimate the parameters by maximum likelihood. That is, we would maximize the log of the additive logistic t distribution in Equation 5, summed over all observations, with respect to the parameters. However, complications arise in partly contested districts. One attractive approach for "missing data" problems such as this is to use "Markov Chain Monte Carlo" (MCMC) methods (see Tanner, 1997). For example, we could impute the missing data given a guess for the parameters; then estimate the parameters given these "completed" data via maximum likelihood; then use these better parameter estimates to impute more realistic values for the missing data; and so on until (stochastic) convergence.

We have implemented a version of this MCMC approach, but in our experience and with our three-party data, this procedure is relatively slow, primarily because each of the two steps in every iteration is itself iterative. We therefore offer an alternative direct likelihood approach that is approximately twenty times faster. Our studies indicate that the direct likelihood approach is faster for smaller numbers of parties, but the MCMC approach will be more computationally efficient for larger numbers of parties.

In the remainder of this section, we describe our direct likelihood approach. We begin by denoting the set of parties contesting the election in district i as P_i, a set that can take on seven patterns: $\{1,2,3\}$, $\{2,3\}$, $\{1,3\}$, $\{1,2\}$, $\{1\}$, $\{2\}$, and $\{3\}$.[9]

When the effective vote is observed for all parties, the likelihood is the probability density of the observed variables. For simplicity, we write the likelihood as a function of Y_i rather than V_i, although the two give equivalent results. For districts with fully contested elections, the observed vote $(V_{i1}^*, V_{i2}^*, V_{i3}^*)$ equals the effective vote (V_{i1}, V_{i2}, V_{i3}). Thus, $Y_{i1} = \ln(V_{i1}/V_{i3})$ and $Y_{i2} = \ln(V_{i2}/V_{i3})$ are both observed and the likelihood function is the bivariate t probability density:

$$L^{\{123\}} = \prod_{P_i=\{1,2,3\}} \mathrm{T}(Y_{i1}, Y_{i2}|\psi) \tag{7}$$

with parameters

$$\psi = \{\mu_{i1}, \mu_{i2}, \sigma_1, \sigma_2, \rho, \nu\} \tag{8}$$

(and where σ_1, σ_2, and ρ make up the scatter matrix). This density differs from the additive logistic t for V_i in Equation 5 by a constant factor (the Jacobian of the transformation), which thus establishes the equivalence of writing the likelihood as a function of either V_i or Y_i.

When some of the effective votes are not observed (due to political parties not contesting an election), our assumptions designate a region in which the vote variable falls, in which case the likelihood is the area (or volume) under the probability density corresponding to this known region. Appendix A derives the likelihood function for these cases.

The complete likelihood function is the product of the likelihood for the fully contested case in Equation 7, and the partially contested cases derived in Appendix A and given in Equations 12, 15, 16, 17, 18, and 19:

$$L(\psi|V) = L^{123}L^{23}L^{13}L^{12}L^{1}L^{2}L^{3} \tag{9}$$

where we define the product over a null set (when no district elections of the type exists) as equaling unity. We also substitute $\mu_{i1} = X_{i1}\beta_1$ and $\mu_{i2} = X_{i2}\beta_2$ to introduce (overlapping, identical, or different sets of) covariates X_{i1} and X_{i2}. For our present application, we define X_{i1} and X_{i2} to include a lag of Y_{i1}, a lag of Y_{i2}, and three indicator variables to represent incumbency status for each party. We have conducted many other runs with demographics and other variables included, but, as is consistent with the results from

[9]Suppose they held an election and nobody ran? We ignore this amusing eighth possible pattern, despite its occasional appearance in some very low visibility local U.S. elections. In general, the number of patterns of missing data is 2^J.

analyses in the U.S. and other democracies, these tend to have only minor effects on our estimates of incumbency advantage.

To facilitate maximization, and to make the asymptotic normal approximations we use below feasible with fewer observations, it is helpful to reparameterize so that all parameters are unbounded (ranging between $-\infty$ and ∞), as β_1 and β_2 already are. Thus, our full set of transformations are as follows:

$$\begin{aligned} \mu_{i1} &= X_{i1}\beta_1, & \sigma_1 &= e^{\phi_1}, & \rho &= \frac{e^{2\phi_3}-1}{e^{2\phi_3}+1} \\ \mu_{i2} &= X_{i2}\beta_2, & \sigma_2 &= e^{\phi_2}, & \nu &= 2+e^{\phi_4} \end{aligned} \tag{10}$$

where the form of the equation for ϕ_3 is the inverse of Fisher's (1915) "Z transformation" (keeping ρ between -1 and 1 no matter what value ϕ_3 takes) and ϕ_4 constrains ν to be greater than two in order to guarantee that the moments of the distribution on the logistic scale exist.

To summarize all our knowledge about and uncertainty in the parameter vector

$$\phi = \{\beta_1, \beta_2, \phi_1, \phi_2, \phi_3, \phi_4\} \tag{11}$$

we maximize the likelihood function. This gives us an asymptotic normal posterior distribution with the maximum likelihood point estimates as a mean vector and, as usual, the inverse of the negative of the hessian as the variance matrix.[10]

Because maximum likelihood is invariant to reparameterization (see King, 1989), estimating ϕ and transforming to get ψ, or estimating ψ directly, give the same point estimates. We also use the standard "empirical Bayes" approach to specify normal priors (i.e., the hyperparameters are "estimated" from the means and variances of the maximum likelihood estimates across our ten election years). Our empirical results below are qualitatively the same as with straightforward maximum likelihood, but, as usual, empirical Bayes helps to reduce the across and within year random variability.

8 Computing Quantities of Interest

Since the point of developing a model of voting in a multiparty democracy is to explain and predict election results, our model ought to be capable of computing quantities on the

[10] We verified this asymptotic normal approximation by comparisons with the exact (i.e., finite sample) posterior distribution, thus avoiding the large-n assumption altogether. We did this with the technique of "importance sampling," an iterative simulation method based on a probabilistic rejection algorithm (see Tanner, 1997). Our experiments with this procedure indicate that the point estimates we report below are correct, and the standard errors are if anything conservative (i.e., somewhat larger than they should be). Because importance sampling is very computationally intensive and hence would be more difficult for others to apply, and since we found that the two approaches did not suggest any real substantive differences in our data, the analyses below are based on the asymptotic normal approximation.

scale of reported votes. That is, the estimated ϕ parameters that result from maximizing the likelihood are important, but they are of little *direct* interest. For starters, they are reparameterized for estimation via Equations 10. But even transforming back to the original ψ scale, by inverting these equations, is not that helpful since the estimation was done on the additive logistic scale rather than on the scale of substantive interest — the votes. The quantities of direct substantive interest are complicated functions of these parameters.

Computing some of these quantities of interest is possible analytically (through Taylor series approximations and the like), but it would be difficult. Computing many others is impossible. We simplify these problems by substituting computer time for human effort via a technique called random simulation (also called stochastic simulation, or Monte Carlo simulation, etc.). This is an increasingly popular technique in statistical analysis (see Jackman, 1996; Tanner, 1997; Gelman et al., 1995). Because simulation can generate results with *any* desired degree of precision, the technique entails no compromises (given a sufficiently powerful computer).

We describe the calculation of three quantities of interest in this section, a predicted vote, an expected vote, and a causal effect. With each, we use a combination of classical and Bayesian techniques. We save the calculation of other quantities, such as bias and responsiveness, for a future paper.

8.1 Predicted Vote

Our quantity of interest in this section is the probability distribution describing the predicted allocation of votes in a district conditional on a fixed value for each of the explanatory variables. The prediction is therefore a probability distribution over the simplex.

Our first requirement is a method of drawing one random election result from the approximate posterior distribution given the estimated model, which we label $(\tilde{V}_{p1}, \tilde{V}_{p2}, \tilde{V}_{p3})$, where the p subscript is mnemonic for prediction and the tilde indicates that the values have been simulated. We draw this simulated district election result given a set of values for the explanatory variables X_{p1} and X_{p2} (each being row vectors). To accomplish this we follow this algorithm:

1. Maximize the likelihood function in Section 7 (with the empirical Bayes priors), and record the vector of maximum likelihood estimates, $\hat{\phi}$, and the variance matrix, $\hat{V}(\hat{\phi})$.

2. Take one random draw of ϕ, which we designate as $\tilde{\phi}$, from a multivariate normal distribution with mean $\hat{\phi}$ and variance $\hat{V}(\hat{\phi})$.

3. Reparameterize from $\tilde{\phi}$ to $\tilde{\psi}$ by using Equations 10, where we use X_{p1} and X_{p2} in computing μ_{p1} and μ_{p2}.

4. Draw $\tilde{Y}_{p1}$ and $\tilde{Y}_{p2}$ randomly from a bivariate t distribution with parameters $\tilde{\psi}$.

5. Compute $\tilde{V}_{p1}$, $\tilde{V}_{p2}$, and $\tilde{V}_{p3}$ deterministically from $\tilde{Y}_{p1}$ and $\tilde{Y}_{p2}$ by the multivariate logistic transformation in Equation 4.

To compute the distribution of election results given X_p, we repeat Steps 2–5 of this algorithm M times (we find that $M = 1000$ is sufficient for most purposes). Then our approximate posterior distribution of V_{p1} is merely a histogram of the simulated values. A point estimate of the three party vote results can be computed by taking the numerical average of the simulations for each party. A standard error can be computed by taking the standard deviation of the simulations for each party. Similarly a (say) 80% confidence interval can be computed by sorting the values in numerical order and taking the values at the 10^{th} and 90^{th} percentiles. The full approximate posterior distribution may be calculated by a two-dimensional histogram over the simplex.

For an example of simulating predictive quantities of interest, we give an inference about a predicted value from a typical open seat. To be specific, we first estimated the model for 1987 with lags of Y_{i1} and Y_{i2} and two indicators for incumbency status. We included all variables in both equations. We then set the explanatory variables (X_{p1} amd X_{p2}) such that none of the candidates are presently members of the House of Commons, and the previous vote (i.e., in 1983) is equal to the average vote across constituencies ($V_{pC} = 0.46$, $V_{pL} = 0.28$, and $V_{pA} = .26$). We then applied the algorithm above to yield 500 simulations of the three vote vectors.

We use two graphical methods for portraying the results from this prediction, both appearing in Figure 5. Graph (A) in this figure plots the 500 simulations in one ternary diagram. The simulations are all predictions for a single district and thus vary only due to uncertainty in the prediction; the collection of dots in this figure then portrays the full nature of the probabilistic prediction about where the point (given the values of the explanatory variables) is likely to be. That is, we have higher confidence that the actual district vote in the average open seat district will be where the heavy cluster of the dots fall, and the result will fall with smaller probability where there are fewer dots. Substantively, the result shows that this typical district is very likely to be won by the Conservatives, since most of the mass of points fall into the upper third of the triangle. (The actual probability that this district will be won by the Conservatives equals the fraction of simulated dots that fall in this top region defined by the win lines.)

Graph (B) in Figure 5 gives density estimates (smooth versions of histograms) for each of the three vote variables. This graph helps emphasize the separate, but still obviously related, nature of the three variables. Each density estimate portrayed in the graph is an approximate posterior distribution of that quantity (i.e., it can be thought of as a pile of predictions or simulations), indicating where the future value of that vote is likely to be. Judging from the very little overlap in the distributions, it is highly likely that the Conservatives will out poll the Labour and Alliance parties. The Labour party will likely do better than the Alliance in this constituency, but because of the heavy overlap in these two distributions, this inference is less certain. Note that all information in Graph

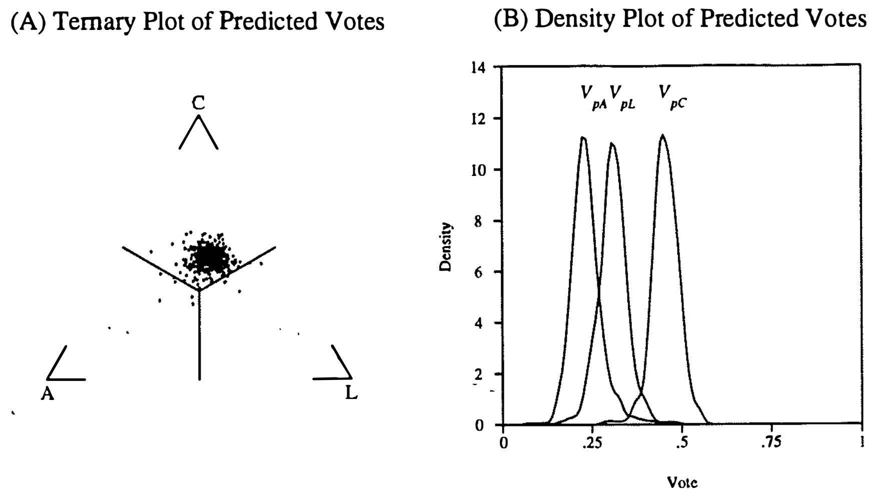

Figure 5: *Simulations of a Predicted Value: This figure interprets the results of a model by computing the distributional implications of a single prediction (for an open seat in the constituency with the average vote). Graph (A) plots 500 simulations from this prediction on a ternary diagram; Graph (B) gives density estimates of the simulations from the same three vote variables. According to the prediction, the district's vote heavily favors the Conservatives.*

(B) can also be found in Graph (A), although the images emphasize different aspects of the data.

In an application with a variety of explanatory variables, we would normally do many different computations like this. This would enable the researcher to understand the many substantive implications of models like this. To do this, we would set the explanatory variables at many different sets of values (low income and heavily minority; high income and rural, etc.). In this situation, we may wish more parsimonious summaries of the simulations such as point estimates and standard errors, or confidence intervals.

8.2 Expected Vote

Our knowledge of all real world random processes is affected by both *fundamental variability* and *estimation variability.* Estimation variability results from the limited number of observations we were able to collect (or the limited number of districts we are analyzing). If n were very large, estimation variability would vanish. In contrast, even if we had an infinite number of observations, the fundamental variability in the real world would still prevent our vote predictions from being perfect. Estimation variability is introduced because of the investigators' "failings," whereas fundamental variability affects our results because the world we are studying is intrinsically variable.

Our procedure for computing the predicted vote in Section 8.1 reflects both sources of variability. In the algorithm of that section, Step 2 simulates estimation variability (by drawing ϕ from its distribution) and Step 4 simulates fundamental variability (by drawing the logit of the vote variables from a t distribution). Since we wished the simulations to reflect our knowledge of the distribution of votes, both sources of variability were essential.

Closely related to computing the predicted vote is estimating the expected vote in a district: $[\mathrm{E}(V_{p1}), \mathrm{E}(V_{p2}), \mathrm{E}(V_{p3})]$. Like the predicted vote, these expected votes are also conditional on chosen values of the explanatory variables, X_p. Although fundamental variability affects our estimate of the expected value, we need to average over it to produce the expectation. In other words, the expected vote is fixed, and only our estimation of it is imperfect: if n were sufficiently large, the expected vote simulations would be constant. In practice, of course, our estimation procedure will produce uncertain estimates of the (fixed) expected vote.[11]

To compute *one* simulation of the expected vote, which we denote $[\tilde{\mathrm{E}}(V_{p1}), \tilde{\mathrm{E}}(V_{p2}), \tilde{\mathrm{E}}(V_{p3})]$, we follow this algorithm.

1. Maximize the likelihood function in Section 7 (with the empirical Bayes priors), and keep the vector of maximum likelihood estimates, $\hat{\phi}$ and the variance matrix, $\hat{V}(\hat{\phi})$.

2. Take one random draw of ϕ, which we designate as $\tilde{\phi}$, from a multivariate normal distribution with mean $\hat{\phi}$ and variance $\hat{V}(\hat{\phi})$.

3. Reparameterize from $\tilde{\phi}$ to $\tilde{\psi}$ by using Equations 10, where we use X_{p1} and X_{p2} in computing μ_{p1} and μ_{p2}, respectively.

4. Draw m values of $\tilde{Y}_{p1}$ and $\tilde{Y}_{p2}$ randomly from a bivariate t distribution with parameters $\tilde{\psi}$. ($m = 100$ is usually sufficient.)

5. Compute m simulations of $\tilde{V}_{p1}$ and $\tilde{V}_{p2}$ deterministically from each of the m simulations of $\tilde{Y}_{p1}$ and $\tilde{Y}_{p2}$ by using the multivariate logistic transformation in Equation 4.

6. Calculate the numerical average of the m simulations of $\tilde{V}_{p1}$ and $\tilde{V}_{p2}$ to yield one simulation of the expected votes, $\tilde{\mathrm{E}}(V_{p1})$ and $\tilde{\mathrm{E}}(V_{p2})$, respectively. Compute the simulation of the expected vote for party $J = 3$ by subtraction: $\tilde{\mathrm{E}}(V_{p3}) = 1 - \tilde{\mathrm{E}}(V_{p1}) - \tilde{\mathrm{E}}(V_{p2})$

We repeat steps 2–6 of this algorithm M times to produce M simulations of the expected vote, the mean of which is our point estimate, the standard deviation is the standard error, and a histogram of each component is the full probability density ($M = 1000$ is usually sufficient).

[11]The difference between the expected vote, and the predicted vote in Section 8.1, is primarily in the variability around the mean. If the model were linear, the average of the simulations of predicted votes and expected votes would be identical; in our case, the two are close.

8.3 Causal Effects, Including Incumbency Advantage

A causal effect is the difference between two expected votes given a change in the value of only one explanatory variable. For example, the incumbency advantage is the difference in the expected vote in a district with an incumbent running and the expected vote in the same district at the same time when the incumbent's party decides to nominate the best available non-incumbent willing to run (Gelman and King, 1990). That is, under this thought experiment, everything is held constant up to the start of the general election campaign, at which point either the incumbent runs for reelection or does not.

The causal effect of incumbency status in multiparty democracies is of course somewhat more complicated than two-party systems, since the effect on the expected vote of, for example, a Conservative incumbent seeking reelection may be of a different magnitude than for an Alliance or Labour incumbent. Such a partisan differential would also seem more likely in legislatures with more parties. In multiparty democracies, we could estimate the incumbency advantage averaged over the parties, but we prefer to estimate them separately in order to highlight several interesting substantive differences in our data.

Computing a causal effect thus requires two sets of expected votes, one with an incumbent and the other with no incumbent. To draw simulations of the causal effect of incumbency, we take the difference between a simulation of the expected vote when the incumbency status variable in X_p indicates a particular party's incumbent is running for reelection, and when X_p indicates an open seat, with all other variables held constant at (say) their means. That is, we maximize the likelihood and then run the algorithm in Section 8.2 twice, with a change only in the incumbency status variable.

We estimated the advantage due to three types of changes in incumbency status — open seat to a Conservative incumbent, an open seat to a Labour incumbent, and an open seat to an Alliance incumbent — in each of the 10 election years from 1959 to 1992. The additional complication is that for one year, and for one type of incumbency effect, we need to record changes to all three vote variables. To display all this information succinctly, we have devised a new graphical display. For example, Figure 6 presents the raw results for each year and type of effect. The top panel shows the effect of a Conservative incumbent, and each arrow in the large left portion of this graphic is the change from an open seat — which we construct so that it begins at the point on the line at the year indicated — to where the expected vote would be with a Conservative incumbent, as if each were part of a ternary diagram. Hence, the higher each arrow extends vertically (i.e., not the length of the arrow, although the two are obviously related), the larger is the incumbency advantage to the Conservative party. The scale on the left of the graph is in percentage points of incumbency advantage.

The direction each arrow leans indicates from which party the conservatives draw votes when they run an incumbent versus running another candidate in an open contest, with the vertices of the implicit ternary diagram indicated around the standard errors at the right. (Note also that there is a standard error in each direction, indicated by the

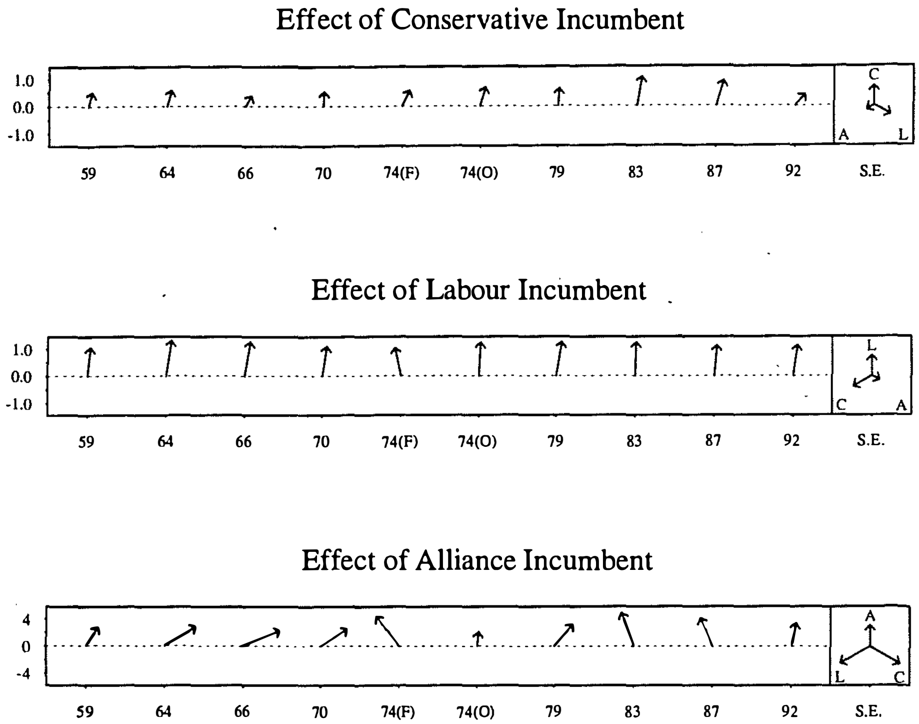

Figure 6: *Incumbency Advantage: The vertical distance each arrow is above the line indicates the advantage of running an incumbent, as compared to a nonincumbent, to the party indicated. The direction the arrow comes from indicates from which of the other parties support is being drawn (as indicated by the ends of the standard error bars, at the right). Note that the scale, in percentage points for all three parties, is larger for the Alliance graph than for the other two.*

length of each of the standard error arrows at the right.[12]) Thus, since all the arrows lean at least somewhat to the right in the top panel, Conservative incumbents draw more from the Alliance than from the Labor party to form their personal vote. To our knowledge, this type of effect has not been in the estimated elections literature of any country.

The most striking effect in Figure 6 is the unambiguous positive effect of incumbency, for all ten elections and incumbents of all three parties over half a century of British politics. This is indicated by every one of the 30 arrows in the figure pointing above the zero line. Most of the individual effects in Figure 6 are larger than their standard errors. When pooled, the average effects for each party's incumbent are 2–5 times their standard errors, and hence by any relevant statistical standard clearly greater than zero. Figure

[12]Because the Alliance receives many fewer votes on average than the other parties, the maximum range of votes that could be drawn from the Alliance to form a personal vote for any party's incumbents is quite small. As a result, the Alliance standard errors in Figures 6 and 7 are smaller than for the Conservatives or Labour.

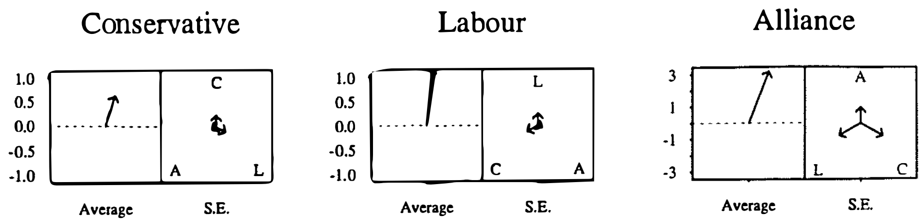

Figure 7: *Average Incumbency Advantage: This figure gives the incumbency advantage averaged over all the years portrayed in Figure 6, for which the intepretation is analogous. Note the scale for the Alliance graph, which is larger than the others.*

7 gives these average effects for each party. (As always, the standard error of an average is smaller than the standard error of its component parts.)

The summary in Figure 7 indicates that the average incumbency effect for the Conservatives is about half a percentage point, and for the Labour party it is twice as large, a full percentage point. In fact, in every one of the ten elections in Figure 6, the Labour incumbency effect is larger than the Conservative effect. Our interpretation of this disparity (which is necessarily more speculative than our results) is that Labour incumbents serve working class constituents and so have more constituency service to dole out. They have the discression to influence the position of people on various types of lists for social services, such as to get into, or to rennovate, council housing. Conservative members have few of these opportunities when serving their relatively more wealthy constituents with government services, and so their personal vote should not be as large.

Alliance incumbents receive an advantage of three percentage points (thrice the Labour advantage). This is not quite the size of the 8–9 percentage point incumbency advantage in the U.S. House of Representatives, but it is only slightly smaller than the size of the average advantage due to incumbency in the House in the mid-1960s or in most American state legislatures today. This effect is in part because the counterfactual involved in an open seat vs. an Alliance incumbent is more of a stretch in parts of the country than the analogous counterfactual involved for estimating major party incumbency effects. Having an Alliance incumbent in office implies that the major party hold on the political system has been broken, and voters take this cue to reevaluate their votes. Another way to look at it is that the collective action problem of moderate voters preferring the Alliance but wanting to avoid wasting their vote has been solved with an Alliance incumbent in office. Voters who prefer the Alliance Party when an Alliance incumbent is in office have less reason to vote strategically ("tactically" as it is called in Britain) and instead cast their vote for their sincere preference. Presumably in response to this idea, the Alliance, much more than the major parties, run local, almost U.S.-style, candidate-centered campaigns (often nominating well-known nonpolitical personalities), rather than national, party-oriented campaigns. The response to this, we believe, is the much larger Alliance incumbency advantage.

Taken together, these results support the claims of the new quantitative literature on the existence of the incumbency effect in Britain, but the size of the effect for all three parties is substantially smaller than previous biased methods had indicated. That is, our method is not biased and is also sufficiently powerful to be able to distinguish a small (but politically meaningful) effect from none at all. Our method is also able to discern distinctly different incumbency effects across the three parties.

There appears to be a hint in the top panel of Figure 6 that the Conservative incumbency advantage may be growing slightly, but formal statistical tests clearly reject this possible trend. Moreover, Labour or Alliance incumbency effects display no such apparent pattern. Thus, we find no support for the other claim of the new quantitative literature that argues, against the conventional wisdom, that the personal vote has been dramatically increasing in recent years.

When the Conservatives run an incumbent (as compared to no incumbent running), they pull most of their extra personal vote from the Alliance. As explained, this can be seen because the arrow in the first panel of Figure 7, and all 10 arrows in the first panel of Figure 6, leans to the right — away from the "A" in the standard error diagram. Our interpretation of this clear result is due to the Alliance being a transitional party for voters on the way to supporting one of the major parties; only 20% of voters stick with the Alliance for more than one election (Butler and Stokes, 1969: 315–338). Presumably because the Labour personal vote stems more from constituency service than the Conservatives (and because there are usually as many Conservative voters who could benefit from a Labour party member's constituency service as voters for the much smaller Alliance party who could benefit), Labour Party incumbents pull approximately equally from the Conservatives and Alliance parties (see Cain et al., 1987).[13]

9 Concluding Remarks

Our paper may help resolve an ongoing debate in the British elections literature. The conventional position is that incumbency advantage is nonexistent and has not changed. In contrast, a newer literature holds that incumbency advantage is moderate to large, and it is growing. With our approach, we find that both sides are right to an extent. That is, the incumbency advantage in England is small but meaningfully above zero. However, there is no evidence that these advantages are trending in any direction. In addition, our model is able to detect important differences among the parties, with the incumbency advantage for Labour being about twice that for the Conservatives, and the advantage for the Alliance triple that of Labour.

[13]The Alliance party does not predictably draw more from either major party. However, there is one other feature of the last panel of Figure 6 that may reflect a systematic pattern: The Alliance incumbents drew their personal vote almost exclusively from the Labour party between 1959 and 1970, but then abruptly changed and started to draw votes in different ways from Labour and the Conservatives over the subsequent six elections. Since this "pattern" is based on only four elections, further research will be required before drawing firm substantive conclusions as to its cause.

We have developed and presented our statistical model for three parties and applied it to the British electoral system. The model and estimation procedures are all directly applicable to electoral systems with any number of parties. (And we plan to turn our computer code into an easy-to-use software program and to distribute it freely.) Further research will be necessary to implement the more general version of our model. For one, our analytical approach, while much faster for three parties, does not scale up as well as a more general MCMC approach. Our experiments with MCMC applied to this problem convince us that it will scale fairly well. Priors may be needed given the large number of parameters relative to a smaller number of districts, but there is substantial information in multiparty electoral systems that would be extraordinarily valuable to researchers in comparative politics and so this extension seems well worth the effort. Our graphical displays obviously do not generalize directly to more than four parties, but with judicious use of color, shading, and perspective, we have found it possible to display the results from between five and eight parties, depending on how complicated and empirically clear the relationships are in the data.

Many opportunities for future research remain in this area, three others of which we now describe. First, we can easily extend the interpretation of the model to include other quantities of interest, such as bias and responsiveness of the electoral system, important and controversial issues in the U.K. and elsewhere. This can be done by clearly defining the quantity to be computed and then making slight modifications of the algorithms described in Section 8.

Second, some two-party models of seats and votes have in recent incarnations included random effects terms, more highly modeled versions of what our empirical Bayes approach accomplishes for the multiparty case. These terms help keep the estimates reasonable even if certain types of explanatory variables are not observed and included in the model. They are especially well-suited to models of legislatures, since the structure of the random effects model can reflect the panel structure of the data.

Finally, electoral systems vary considerably across countries. All but two countries have districts of some kind, and most elect legislators to more than one seat in a district. Modifications to fit all the different types of electoral systems will require some detailed work, but should be achievable by straightforward extensions of the model we present here.

Appendix A: The Likelihood Function for Partially Contested Districts

A.1 Districts with Two of Three Parties Contesting

First consider elections where the parties contesting in district i include $P_i = \{2,3\}$. For these elections, the effective vote for all three parties, V_{i1}, V_{i2} and V_{i3}, are unobserved. Nevertheless, our assumptions for less-than-fully contested elections introduced in Section 6 imply that we have some information about these quantities. In particular, we know that $V_{i1} < \min(V_{i2}, V_{i3})$ and therefore $Y_{1i} < \min(0, Y_{i2})$ or, equivalently, $Y_1 < 0$ and $Y_1 < Y_2$. Thus, the likelihood function for this case is as follows:

$$\begin{aligned} L^{23} &= \Pr(Y_{i1} < 0, Y_{i2} > Y_{i1}|\psi) \\ &= \prod_{P_i=\{2,3\}} \int_{-\infty}^{0} \int_{Y_{i1}}^{\infty} \mathrm{T}(Y_{i1}, Y_{i2}|\psi) dY_{i2} dY_{i1} \\ &= \prod_{P_i=\{2,3\}} \int_{-\infty}^{0} \mathrm{T}(Y_{i1}|\mu_{i1}, \sigma_1, \nu)[1 - \mathrm{F_T}(Y_{i1}|\mu_{2|1}, \sigma_{2|1}, \nu+1)] dY_{i1} \end{aligned} \quad (12)$$

where $\mathrm{F_T}$ is the cumulative distribution function of the (univariate) t,

$$\mu_{2|1} = \mu_{i2} + (\rho\sigma_2/\sigma_1)(Y_{i1} - \mu_{i1}) \quad (13)$$

is the conditional mean of Y_{i2} given Y_{i1} and

$$\sigma^2_{2|1} = \sigma^2_2(1-\rho^2)\left(\frac{\nu + (Y_{i1} - \mu_{i1})^2/\sigma^2_1}{\nu+1}\right) \quad (14)$$

is the conditional variance. These conditional t distributions are analogous and mathematically similar to the more commonly known conditional normals (see Liu, 1994). The function in Equation 12 is easily calculated by one-dimensional numerical integration.

By a parallel logic, the likelihood function for district elections where parties 1 and 3 contest is:

$$L^{13} = \prod_{P_i=\{1,3\}} \int_{-\infty}^{0} \mathrm{T}(Y_{i2}|\mu_{i2}, \sigma_2, \nu)[1 - \mathrm{F_T}(Y_{i2}|\mu_{1|2}, \sigma_{1|2}, \nu+1)] dY_{i2} \quad (15)$$

When parties 1 and 2 contest, but party 3 is missing, a slight computational difference occurs because V_{i3} appears in the denominator of both Y_{i1} and Y_{i2}. As a result, we know, by our assumption, that $V_{i3} < \min(V_{i1}, V_{i2})$, which translates into $Y_{i1} > 0$ and $Y_{i2} > 0$. Hence:

$$L^{12} = \prod_{P_i=\{1,2\}} [1 - \mathrm{F_T}(0,0|\psi)] \quad (16)$$

where $\mathrm{F_T}$ is the cumulative distribution function of (in this context) the bivariate t. To compute this function, we follow the standard procedure of applying one-dimensional numerical integration after factoring the joint t distribution into a marginal and conditional — directly analogous to Equation 12.

A.2 Districts with One of Three Parties Contesting

For example, when only party 1 contests (and hence automatically wins), we have no information about Y_{i2}, but we know that $Y_{i1} > 0$. We use this information to form the likelihood function:

$$\begin{aligned} L^1 &= \prod_{P_i=\{1\}} \int_0^{\infty} \int_{-\infty}^{\infty} \mathrm{T}(Y_{i1}, Y_{i2}|\psi) dY_{i1} dY_{i2} \\ &= \prod_{P_i=\{1\}} [1 - \mathrm{F_T}(0|\mu_{i1}, \sigma_1, \nu)] \end{aligned} \tag{17}$$

Similarly, the likelihood function for the remaining two cases is

$$L^2 = \prod_{P_i=\{2\}} [1 - \mathrm{F_T}(0|\mu_{i2}, \sigma_2, \nu)] \tag{18}$$

and

$$\begin{aligned} L^3 &= \prod_{P_i=\{3\}} \int_{-\infty}^{0} \int_{-\infty}^{0} \mathrm{T}(Y_{i1}, Y_{i2}|\psi) dY_{i1} dY_{i2} \\ &= \prod_{P_i=\{3\}} \mathrm{F_T}(0, 0|\psi) \end{aligned} \tag{19}$$

References

Alvarez, R. Michael and Jonathan Nagler. forthcoming. "When Politics and Models Collide: Estimating Models of Multi-Party Elections," *American Journal of Political Science.*

Aitchison, J. 1986. *The Statistical Analysis of Compositional Data*, New York: John Wiley.

Bellucci, P. 1991. "Italian Economic Voting: A Deviant Case or Making a Better Theory?" *Economics and Politics: The Calculus of Support.* Pp. 63-81, in H. Norpoth, M. S. Lewis-Beck and J.-D. Lafay, eds., Ann Arbor, Michigan, University of Michigan Press.

Bellucci, P. 1984. "The Effect of Aggregate Economic Conditions on the Political Preferences of the Italian Electorate, 1953-1979." *European Journal of Political Research* 12, 4: 387-401.

Butler, David E. and Dennis Kavanagh. 1980. *The British General Election of 1979.* London: Macmillian.

Butler, David E. and Donald Stokes. 1969. *Political Change in Britian: Forces Shaping Electoral Choice.* New York: St. Martin's Press.

Cain, Bruce; John Ferejohn; and Morris Fiorina. 1987. *The Personal Vote: Constituency Service and Electoral Independence*, Cambridge: Harvard University Press.

Cornford, J.; Dorling, D.; and Tether, B. 1995. "Historical precedent and British electoral prospects," *Electoral Studies*, 14, 2: 123–142.

Curtice, J. and M. Steed. 1983. "The Voting Analyzed," in D.E. Butler and D. Kavanagh, eds., *The British Gernaral Election of 1983.* London: Macmillian.

Curtice, J. and M. Steed. 1980. "An Analysis of Voting," in D.E. Butler and D. Kavanagh, eds., *The British General Election of 1979.* London: Macmillian.

Domínguez, Jorge and James McCann 1996. *Democratizing Mexico: Public Opinion and Electoral Choices.* Baltimore, Johns Hopkins University Press.

Frey, B. and F. Schneider. 1980. "Popularity Functions: The Case of the US and West Germany. Models of Political Economy." Pp. 47–84 in Paul Whiteley. London: Sage.

Gelman, Andrew et al., 1997. *Bayesian Data Analysis*, New York: Chapman and Hall.

Gelman, Andrew and Gary King. 1994a. "Enhancing Democracy Through Legislative Redistricting," *American Political Science Review*, 88, 3 (September): 541–559.

Gelman, Andrew and Gary King. 1994b. "A Unified Method of Evaluating Electoral Systems and Redistricting Plans," *American Journal of Political Science*, 38, 2 (May): 514–554.

Gelman, Andrew and Gary King. 1990. "Estimating Incumbency Advantage Without Bias," *American Journal of Political Science*, 34, 4 (November, 1990): 1142–1164.

Goodhart, C. A. E. and R. J. Bhansali. 1970. "Political Economy." *Political Studies.* 18: 43-106.

Horowitz, Donald L. 1985. *Ethnic Groups in Conflict.* Berkeley: University of California Press.

Horowitz, Donald L. 1993. "Democracy in Divided Societies." *Journal of Democracy* 4, 4: 18-38.

Host, V. and M. Paldam 1990. "An International Element in the Vote?" *European Journal of Political Research* 18: 221–239.

Inoguchi, T. 1980. "Economic Conditions and Mass Support in Japan, 1960–76" *Models of Political Economy.* Pp. 121-151 in Paul Whiteley, ed., London: Sage.

Jackman, Simon. 1996. "Bayesian Tools for Social Scientists," paper presented to the Annual Meetings of the American Political Science Association.

Johnson, Norman Lloyd and Samuel Kotz. 1972. *Distributions in Statistics: Continuous Multivariate Distributions.* New York, Wiley.

King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*, Princeton: Princeton University Press.

King, Gary. 1990. "Electoral Responsiveness and Partisan Bias in Multiparty Democracies," *Legislative Studies Quarterly*, XV, 2 (May): 159–181.

King, Gary. 1989a. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* New York: Cambridge University Press.

King, Gary. 1989b. "Representation Through Legislative Redistricting: A Stochastic Model," *American Journal of Political Science*, 33, 4 (November, 1989): 787–824.

King, Gary and Robert X. Browning. 1987. "Democratic Representation and Partisan Bias in Congressional Elections," *American Political Science Review*, 81, 4 (December, 1987): 1251–1273.

King, Gary and Andrew Gelman. 1991. "Systemic Consequences of Incumbency Advantage in the U.S. House," *American Journal of Political Science*, 35, 1 (February): 110–138..

Lewis-Beck, M. S. 1988. *Economics and Elections: The Major Western Democracies.* Ann Arbor: University of Michigan Press.

Lewis-Beck, M. S. and P. Bellucci. 1982. "Economic Influences on legislative elections in Multi-party Systems: France and Italy." *Political Behavior* 4: 93-107.

Lewis-Beck, M. S. and G. Mitchell (1990). "Modelos transnacionales de voto econmico: estudio de un conjunto de paises europeos." *Estudios Economicos: Elecciones y Economia* 4: 66–81.

Liu, Chuanhai. 1994. "Statistical Analysis Using the Multivariate t Distribution," Ph.D. Dissertation, Department of Statistics, Harvard University.

Miller, A. H. and O. Listhaug. 1985. "Economic Effects and the Voting in Norway. Pp. 125–43 in M. S. Lewis-Beck and H. Eulau. *Economic Conditions and Electoral Outcomes* New York, Agathon Press, Inc..

Norton, P. and David Wood. 1990. "Constituency Service By Members of Parliament: Does it Contribute to a Personal Vote?" *Parliamentary Affairs*, 43:196–208.

Offe, C. 1992. "Strong Causes, Weak Cures." *East European Constitutional Review* 1, 1: 21–23.

Paldam, M. 1991. "How Robust is the Vote Function?: A Study of Seventeen Nations over Four Decades." Pp. 9–31 in H. Norpoth, M. S. Lewis-Beck and J.-D. Lafay, eds., Ann Arbor, Michigan, University of Michigan Press.

Paldam, M. 1986. "The Distribution of Election Results and the Two Explanations of the Cost of Ruling." *Europaische Zeitschrift fur Politische Okonomie / European Journal of Political Economy* 2(1): 5-24.

Powell, G. B. and G. D. Whitten. 1993. "A Cross-National Analysis of Economic Voting: Taking Account of the Political Context." *American Journal of Political Science* 37, 2: 391–414.

Przeworski, Adam. 1991. *Democracy and the Market.* New York: Cambridge University Press.

Przeworski, Adam. 1996. "Public Support for Economic Reforms in Poland." *Comparative Political Studies.* 29, 5: 520–543.

Rosa, J.-J. and D. Amson. 1976. "Conditions Economiques et Elections." *Revue Francaise de Science Politique* 26, 6: 1101–1119.

Rattinger, H. 1991. "Unemployment and Elections in West Germany." Pp. 49–62 in H. Norpoth, M. S. Lewis-Beck and J.-D. Lafay, eds., Ann Arbor, Michigan, University of Michigan Press.

Slider, D. 1994. "Poltical Tendencies in Russia's Regions: Evidence from the 1993 Parliamentary Elections." *Slavic Review* 53, 3: 711-32.

Stokes, S.C. 1996. "Economic Reform and Public Opinion in Peru, 1990-95," *Comparative Political Studies*, 29(5), 544-65.

Tanner, Martin A. 1997. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, third edition, New York: Springer-Verlag.

Tucker, Joshua. 1996. "The White Fear Hypothesis Meets Vladimir Zhironovsky: Ethnic Composition of Regions and Voting Patterns in the 1993 Russian Duma Elections." paper prepared for the annual meetings of the Northeastern Political Science Association 14–16 November.

White, S., R. Rose, et al. 1997. *How Russia Votes.* Chatham, NJ: Chatham House Publishers.

Whiteley, Paul 1980. "Politico-Econometric Estimation in Britain: An Alternative Explanation. Models of Political Economy." Pp. 85–100 in Paul Whiteley. London: Sage.

Wood, D. and Philip Norton. 1992. "Do Candidates Matter? Constiuency-Specific Vote Changes for Incumbent MPs, 1983–1987," *Political Studies*, XL: 227–238.