Sociological Methods & Research

http://smr.sagepub.com

Estimating the Extent of Racially Polarized Voting in Multicandidate Contests

BERNARD GROFMAN and MICHAEL MIGALSKI

Sociological Methods Research 1988; 16; 427

DOI: 10.1177/0049124188016004001

We extend the standard ecological regression methodology for analysis of racial bloc voting in single-member districts in which there is one seat to be filled to cover multimember districts without a numbered-place system. In such multimember districts voters may vote for up to k candidates and not all voters need to cast the same number of votes. These complications render the usual single-equation OLS technique inappropriate. The methodology we propose involves a two-step estimating procedure. The variable that is our principal concern, the proportion of white voters who vote for the black (or other minority) candidate, has its value determined by combining information from each of our two separate estimating equations. Thus we cannot directly obtain a confidence limit for our estimated value of this variable from either of the two equations alone. We derive a value of the standard error of our estimate of this variable by combining formula for the variance of the quotients and sums of two variables of known variance and covariance. We verify that our results are robust by comparing our estimates with those generated by the "seemingly unrelated regression" (SURE) model that is designed to cope with correlated errors in regressions that share in the same independent variables. The methodology we develop has an important application in voting rights litigation challenging the constitutionality of the multimember districts and at-large elections, which are the most common form of election mechanism in U.S. cities.

Estimating the Extent of Racially Polarized Voting in Multicandidate Contests

BERNARD GROFMAN
MICHAEL MIGALSKI
School of Social Sciences
University of California, Irvine

The 1982 extension of the Voting Rights Act of 1965 amended the Act in several ways with the key change occurring in Section 2. The intent of this change was to allow a

AUTHORS' NOTE: *This research was partially supported by NSF Grant SES#84-21050 to the first-named author and by NSF Grant BNS#80-11994 to the Center*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 16 No. 4, May 1988 427-453

violation to be established by demonstration of a discriminatory effect without the necessity of proving any kind of discriminatory intent (Report of the Committee on the Judiciary of the U.S. Senate on S1992: 28). One critical factor that needs to be proved in order to establish a discriminatory effect is the extent to which voting in a political area is racially polarized (Report of the Committee on the Judiciary of the U.S. Senate on S1992: 28-30). Racially or linguistically polarized voting is said to occur when the voting patterns of minority voters differ from those of the majority community in a statistically and substantively significant fashion (*Gingles v. Edmisten,* D.C. North Carolina, 1984). Although the existence of racially polarized voting has come to be recognized by civil rights attorneys as a necessary, if not sufficient, condition for establishing a violation, it has resulted in some of the greatest conceptual problems in the post-1982 Section 2 litigation, because of the dispute over threshold values to establish a voting rights violation and also because of disputes over the appropriate regression techniques to use (Grofman, 1985).

Further complicating determination of the existence of polarized voting are the differences in appropriate methodology for different types of election systems. There are two basic election types in the United States, single-member districts and multimember districts. In the United States, multimember districts (or their polar type, at-large elections) are in general quite common. For example, over 60% of all U.S. cities use multimember districts in whole or part. In this article, we shall be concerned with multimember districts in which voters may vote for up to k candidates and not all voters need cast the same number of votes.

for Advanced Study in the Behavioral Sciences, Stanford, where he was a fellow in 1985 through 1986. We are indebted to the staff of the Word Processing Center, School of Social Sciences, University of California, Irvine, for manuscript typing and to Kenneth Small for helpful suggestions as to the econometrics. We are also grateful to the anonymous reviewers of this manuscript who provided invaluable suggestions for improving the earlier draft. Errors remaining are solely the responsibility of the authors.

EDITOR'S NOTE: *This article was accepted under the editorship of George W. Bohrnstedt.*

This type of multimember election, known as multimember districts without numbered places,[1] is used by most U.S. jurisdictions that use multimember districts, most commonly in at-large elections to city councils and other local offices (Grofman, 1982). In areas in which there is a reasonably sized minority community, many such multimember districts are currently coming under challenge under Section 2 of the Voting Rights Act as being dilutive of minority voting strength if minorities or minority-supported candidates fail to be elected in reasonable numbers and there is a pattern of racially polarized voting in which the majority of white voters are by and large unwilling to vote for minority candidates.

Ecological regression methodology to detect differences in voting patterns between minority and majority voters for single-member districts involving at least one minority candidate and at least one majority candidate is relatively well developed (see, e.g., Grofman et al., 1985; Loewen, 1982; Kousser, 1974; Goodman, 1953). Regression techniques and/or homogeneous case analysis of area units with high population proportion of one group are used to estimate the proportion of the votes cast by nonminority voters that goes to minority candidates and the proportion of the votes cast by minority voters that goes to white (or non-Hispanic) candidates.

Our methodology, such as that used for racial bloc voting analysis in single-member districts, requires that registration data or population data or voting-age population data by race by precinct be available. Such data are directly available in some southern states for blacks. It may be generated for Hispanic voters by means of Spanish surname count. Even when precinct-level data by race/ethnicity are not directly available, because voting precincts commonly use census blocks or census tracts as their building blocks, census data on voting-age racial characteristics may be merged with the precinct-level information on registration. Even when the match up between precinct and census data units is not perfect, detailed matching of census and precinct geography can be done using map overlays or by matching voter addresses in a precinct (by block) with the census units in which the addresses are located. Both of these methodologies have been made use of by one of the present authors. Using census data instead of registration data requires the assumption that there have been no large shifts in the minority population since the last census that would change the location of areas of minority concentration. The accuracy of this assumption would need to be verified by local experts and need not be considered here.

The purpose of this article is to extend the methodology commonly used for ecological regression of racial bloc voting so as to be able to deal with the complexities posed by voting outcomes in multimember district systems where voters can vote for up to k candidates and there are not simple head-on-head contests. We then provide examples of the application of our method to data on state legislative elections (in multimember districts without numbered places) in the state of North Carolina. We confine our analysis to the case where there are only two mutually exclusive population groups, black and white, but our methodology can readily be extended to multiple groups by applying it in a sequential pairwise fashion. (For example, with three groups—non-Hispanic blacks, non-Hispanic whites, and Hispanics—we would look at black versus nonblack patterns, Hispanic versus non-Hispanic patterns, and white versus nonwhite patterns and then combine the information from each pair of oppositions.)

We shall provide a two-equation technique to estimate the voting behavior of white and black voters (i.e., the proportion of each group that votes for any given candidate, and the proportion of each group's vote that goes to candidates of the same race). This two-equation technique is necessitated by the fact that white voters and black voters commonly differ in their levels of registration and turnout (Brace et al., 1986; Grofman et al., 1986), and also in the average number of candidates for whom they vote in situations where each voter may vote for up to k candidates.

Our two-equation technique poses a variety of methodology problems in model specification and estimation, and we provide ways to cope with each of the difficulties it poses. First, our technique requires us to combine information from two separate equations to develop an estimate of the key parameter of interest, proportion of the white (black) electorate voting for a given white (or black) candidate, and thus we cannot directly find a confidence boundary on its reliability. We derived a formula for the standard error of this estimate that makes use of a well-known result on the variance of a ratio of variables (see, for example, Kish, 1965: 206-208; Theil, 1971: 373-374). Second, our technique requires us to estimate a series of equations that share a common independent variable. Thus we must be sensitive to the possibility of correlated errors. We can check for this by making use of a standard econometric technique, "seemingly unrelated regression" (SURE; see Zellner, 1962) that was designed for exactly this situation. Finally, an ecological regression technique has as a key assumption that either (a) group voting behavior is consistent across all areal units, or (b) the differences in group voting behavior across areal units are uncorrelated with the independent variable, the proportion of the areal unit that the group makes up. While this assumption is never perfectly satisfied, it is often closely approximated. We provide several different tests of the plausibility of this assumption for our data. Since we are concerned with the *average* behavior of groups, some degree of variation in group behavior across areal units does not cause insuperable difficulties if proper caution is taken in analysis.

Our conclusions are that the two-equation technique provides reliable estimates of group voting behavior: Our estimates are unbiased, given the assumptions of the model, standard errors are low, correlations are high and highly significant, and variance estimates generally match to two decimal places those obtained from the method of seemingly unrelated regression. Thus the well-known problem of ecological fallacy (Robinson, 1950) does not arise with these data. Later in the article we consider a number of ways (including a variant of the Duncan and Davis [1953] homogeneous case analysis) to verify in general that no ecological fallacy has occurred in interpreting aggregate-level statistics to provide information about individual behavior.

BASIC TWO-EQUATION METHODOLOGY

We shall look at multimember district situations without a numbered place system where each voter may cast from 0 to k votes, and where there is at least one candidate from each of two identifiable linguistic or ethnic groups. We shall provide our analysis in terms of voter registration data,[2] but the model is essentially identical for population or voting-age population data. Strictly speaking, the analysis that follows should have been done using weighted regression, with each precinct weighted by its population. We have done this with several other data sets and found no difference (to two decimal places) between estimates from weighted and unweighted regression equations. Robinson (1950: 355) also found little numerical difference between weighted and unweighted ecological correlations, even using data with few observations. Thus, while the discussion below is presented as if weighted regression had been used, the results presented are based upon unweighted regression estimates.

Let

x = the proportion of total registration which is white

$1 - x$ = the proportion of total registration which is black

P_{WW} = the proportion of the total vote eligible to be cast by white registered voters which goes to the white candidate(s)

P_{BW} = the proportion of the total vote eligible to be cast by black registered voters which goes to the white candidate(s)

P_{BB} = the proportion of the total vote eligible to be cast by black registered voters which goes to the black candidate(s)

P_{WB} = the proportion of the total votes eligible to be cast by white registered voters which goes to the black candidate(s)

P_W = the proportion of the total votes eligible to be cast by registered voters which goes to the white candidate(s)

P_B = the proportion of the total votes eligible to be cast by registered voters which goes to the black candidate(s)

$\bar{n}_W$ = mean number of candidates voted for by white registered voters

$\bar{n}_B$ = mean number of candidates voted for by black registered voters

$\bar{n}$ = mean number of candidates voted for $= \dfrac{\text{Total Votes Cast}}{\text{Registered Voters}}$

T = the proportion of total registered voters who turn out to vote

It is apparent that $P_W + P_B \leq 1$, $P_{WW} + P_{WB} \leq 1$, and $P_{BW} + P_{BB} \leq 1$, since some registered voters will either not vote at all or not vote for a full slate of candidates. (Note that some registered voters, those who don't vote, vote for zero candidates.)

It is apparent that

$$\bar{n} = \bar{n}_W x + \bar{n}_B (1 - x) \quad [1]$$

so

$$\bar{n} = (\bar{n}_W - \bar{n}_B) x + \bar{n}_B. \quad [2]$$

Let $\bar{n}^{(i)}$ = the value of $\bar{n}$ in the i^{th} voting precinct. Thus we may express the value of $\bar{n}$ in the i^{th} precincts as a simple linear function of the proportion of white registered voters, that is,

$$\bar{n}^{(i)} = m_1 x^{(i)} + b_1 + \mu_1^{(i)} \quad [3]$$

where $\mu_1^{(i)}$ is an error term with mean over all precincts assumed to be zero.

If equation 3 holds in each precinct with the same parameters m_1 and b_1 (at least approximately), then if we regress $\bar{n}$ on x, the slope and intercept have a natural interpretation in terms of the mean number of candidates voted for by white (black) registered voters. This will be the case if $\bar{n}_W$ and $\bar{n}_B$ are consistent across precincts in the sense that while the observed values will vary from precinct to precinct, we assume that there is one underlying value for each parameter, $\bar{n}_W$ and $\bar{n}_B$, so that the observed differences are due to sampling variation and not to the fact that there are distinct subpopulations within the black or white community that are behaving differently. This assumption would be violated by the existence of certain types of contextual effects that are discussed in more detail below.

In particular,

$$\bar{n}_W = m_1 + b_1 \qquad [4a]$$

$$\bar{n}_B = b_1 \cdot \qquad [4b]$$

We see that $\bar{n}$ for multimember districts without numbered places is analogous to voter turnout, T, for single-member districts. The key equations for this multimember district case are identities of the form

$$\bar{n}P_W = \bar{n}_W(x)P_{WW} + \bar{n}_B(1-x)P_{BW} \qquad [5]$$

Rewriting, we obtain

$$P_W = \frac{\bar{n}_W P_{WW} - \bar{n}_B P_{BW}}{\bar{n}}(x) + \frac{\bar{n}_B}{\bar{n}} P_{BW}. \qquad [6]$$

If we regress P_W on x, that is, if equation 6′ holds in each precinct (at least approximately) with the same parameters, m_2 and b_2, we get

$$P_W^{(i)} = m_2(x^{(i)}) + b_2 + \mu_2^{(i)}. \qquad [6']$$

Then we obtain, after some algebra:

$$P_{BW} = \frac{\bar{n}}{\bar{n}_B} b_2 \qquad [7]$$

and

$$P_{WW} = \frac{\bar{n}}{\bar{n}_W}(m_2 + b_2), \qquad [8]$$

where m_2 is the slope and b_2 the intercept of the regression equation of P_W on x. Of course, if $\bar{n}_B \simeq \bar{n}_W$, then $P_{BW} \simeq b_2$ and $P_{WW} \simeq m_2 + b_2$.

In like manner, by using the identity

$$\bar{n}P_B = \bar{n}_W(x)P_{WB} + \bar{n}_B(1-x)P_{BB}, \qquad [9]$$

by regressing P_B on x, we can obtain

$$P_{BB} = \frac{\bar{n}}{\bar{n}_B} b_3. \qquad [10]$$

$$P_{WB} = \frac{\bar{n}}{\bar{n}_W} (m_3 + b_3). \qquad [11]$$

The plausibility of the assumption that equations 3 and 6′ and similar equations we present later are approximately satisfied in all precincts, along with some conditions under which our estimating techniques will yield misleading results, are discussed in our concluding discussion. This is a question that has been extensively discussed in the literature (see, e.g., Goodman, 1959: 612-614), but we provide several approaches, some of which build on the fact that the estimates we derive can be directly checked against other observed data.

ESTIMATING RACIAL BLOC VOTING IN THE ELECTORATE

Let P'_{WW} and P'_{BB} be the white (black) vote for the white/black candidates as a proportion of the total *votes cast* by whites (blacks).

Thus

$$P'_{WW} = \frac{P_{WW}}{P_{WW} + P_{WB}} = \frac{m_2 + b_2}{m_2 + b_2 + m_3 + b_3} \qquad [12]$$

and

$$P'_{BB} = \frac{P_{BB}}{P_{BB} + P_{BW}} = \frac{b_3}{b_3 + b_2}. \qquad [13]$$

Although for single-member district elections P'_{WW} and P'_{BB} are the variables of most interest (Loewen, 1982), for multimember districts without a numbered place system the coefficients P'_{WW} and P'_{BB} must be interpreted with great care because of the confounding effects of the relative proportion of white and black candidates. If, for example, there are 5 white candidates and 1 black candidate in a race in which voters may cast up to four votes, if white voters were "colorblind" in the way they cast their ballots (i.e., voted with equal (2/3) probability for each of the six candidates), the black candidates would receive exactly 1/6 of the ballots of white voters. Thus we would find a P'_{WW} of .83 (5/6ths), which would on the surface appear to be evidence of highly polarized voting. For multimember district races without a numbered place system, comparing the obtained P'_{WW} and P'_{BB} values with what would have been obtained had all candidates been voted for with equiprobability may help us to judge the actual degree of racial polarization, but even the assumption that all candidates of the *same* race are voted for with equal probability is suspect; and thus comparisons with the equiprobability model will normally be *quite* misleading. For multimember district races without a place system, it is preferable to focus on racial bloc voting *as measured by* the relative proportion of whites/blacks who vote for any *given* black/white candidate.[3] For multimember district elections without a place system, this will provide us a better measure of racial bloc voting than P'_{WW} or P'_{BB}.

CANDIDATE-BY-CANDIDATE ANALYSIS

For multimember district elections without a place system, we shall develop a general technique for estimating the proportion of white/black voters who vote/do not vote for any *given* candidate. We can then use this technique to consider separately differences in the support bases of the black and white candidate(s). In particular, for the case where there is only a single black candidate, this method will yield estimates of the proportion of whites who vote for that candidate vs. the proportion of blacks who vote for that candidate, the difference between which provides a direct measure of racial bloc voting.

We specify the analysis for the case of a black candidate. The model is identical for a white candidate.

Let

C_j = the j^{th} candidate, a black candidate

T = proportion of registered voters who turned out to vote

P_{VC_j} = proportion of registered voters who vote for candidate C_j

P_{VNC_j} = proportion of registered voters who did not vote for candidate C_j which equals $T - P_{VC_j}$

P_{WVC_j} = proportion of white registered voters who voted for candidate C_j

P_{BVC_j} = proportion of black registered voters who voted for candidate C_j

P_{WVNC_j} = proportion of white registered voters who did *not* vote for candidate C_j

P_{BVNC_j} = proportion of black registered voters who did *not* vote for candidate C_j

By definition

$$P_{VC_j}{}^{(i)} = P_{WVC_j}(x^{(i)}) + P_{BVC_j}(1 - x^{(i)}). \quad [14]$$

$$P_{VC_j}{}^{(i)} = (P_{WVC_j} - P_{BVC_j})(x^{(i)}) + P_{BVC_j}. \quad [14']$$

Thus by regressing P_{VC_j} on x, we obtain the OLS estimates for the slope (m_4) and the intercept (b_4) of the linear model

$$P_{VC_j}{}^{(i)} = m_4 x^{(i)} + b_4 + \mu_4^{(i)}.$$

A natural interpretation of the estimated slope and intercept is

$$P_{WVC_j} = m_4 + b_4 \quad [15]$$

and

$$P_{BVC_j} = b_4. \qquad [16]$$

Similarly $P_{VNC_j} = P_{WVNC_j}(x) + P_{BVNC_j}(1 - x)$, and so by regressing P_{VNC_j} on x we obtain the OLS estimates m_5 and b_5 of

$$P_{VNC_j}{}^{(i)} = M_5 x^{(i)} + B_5 + \mu_5{}^{(i)} \qquad [17]$$

that may be interpreted as

$$P_{WVNC_j} = m_5 + b_5 \qquad [18a]$$

and

$$P_{BVNC_j} = b_5. \qquad [18b]$$

So, for any black candidate C_j

$$P'_{BVC_j} = \frac{P_{BVC_j}}{P_{BVC_j} + P_{BVNC_j}} = \frac{b_4}{b_4 + b_5} \qquad [19]$$

$$P'_{WVC_j} = \frac{P_{WVC_j}}{P_{WVC_j} + P_{WVNC_j}} = \frac{m_4 + b_4}{m_4 + b_4 + m_5 + b_5} \qquad [20]$$

where P'_{WVC_j} and P'_{BVC_j} are the proportion of white (black) voters who vote for the black candidate C_j.[4] It is these primed variables that are of principal interest in the analysis of multimember districts without a numbered place system, since they allow us to compare the proportion of the vote from each race that goes to each candidate and also enable us to determine the ranking of the candidates of each race based upon the proportion of the black (white) vote received by each candidate. These rankings then may be used to compare the sets of candidates that would have been elected by each race and thus to determine whether racially polarized voting has occurred.[5]

AN EMPIRICAL ILLUSTRATION OF THE TWO-EQUATION METHODOLOGY

An example from Mecklenburg and Cabarrus counties in North Carolina will be used to illustrate this procedure. The example is from a general election for the state senate, held in November of 1982, in which four positions were up for election with seven candidates in the running.

We first regress $\bar{n}^{(i)}$ on $x^{(i)}$ in order to estimate $\bar{n}_W$ and $\bar{n}_B$. The slope, m, of the best fitting regression line is equal to .37 (.06) with an intercept of .96 (.05), with the standard errors of the estimates in parentheses (for purposes of this example all values will be rounded to the second decimal place). Thus

$$\bar{n}_W = m_1 + b_1 = \underset{(.02)}{1.33}$$

and

$$\bar{n}_B = b_1 = \underset{(.05)}{.96}$$

The number shown in parentheses is the standard error of the estimate. These values are obtained separately for the slope and intercept, and combined as appropriate.

We next compute $\bar{n}$ by dividing the total number of votes cast in the two counties (286,488) by the registration total for both counties (230,457) yielding $\bar{n} = 1.24$.

We now regress P_W on x resulting in the estimates $m_2 = .66$ (.05) and $b_2 = .52$ (.05), and thus

$$P_{WW} = \frac{\bar{n}}{\bar{n}_W}(m_2 + b_2) = \frac{1.24}{1.33}(1.18) = \underset{(.02)}{1.10}$$

and

$$P_{BW} = \frac{\bar{n}}{\bar{n}_B}(b_2) = \frac{1.24}{.96}(.52) = \underset{(.07)}{.67}\ .$$

A similar regression of P_B on x results in $m_3 = -.29$ (.01) and $b_3 =$ .44 (.01), so that

$$P_{WB} = \frac{\bar{n}}{\bar{n}_W}(m_3 + b_3) = \frac{1.24}{1.33}(.15) = \underset{(.004)}{.14}$$

and

$$P_{BB} = \frac{\bar{n}}{\bar{n}_B}(b_3) = \frac{1.24}{.96}(.44) = \underset{(.03)}{.57}\,.$$

We are now able to compute P'_{WW} and P'_{BB} using equations 12 and 13:

$$P'_{WW} = \frac{P_{WW}}{P_{WW} + P_{WB}} = \frac{.57}{1.10 + .14} = \underset{(.003)}{.89}$$

and

$$P'_{BB} = \frac{P_{BB}}{P_{BB} + P_{BW}} = \frac{.57}{.57 + .67} = \underset{(.02)}{.46}\,.$$

To find the standard errors above, we make use of the formulae derived from the variance of the sum and the variance of the ratio of two variables.[6]

Notice that with seven candidates, one of whom was black, if we assume that the white voters were color (and party) blind and assume that they distributed their votes equally among all of the candidates, then we would expect

$$P'_{WW} = \frac{6}{7} = .86,$$

a value very close to our estimate of .89.

Even if we assume that most voters are Democrats (in actuality only 66% of the registered voters were) and the votes were distributed evenly among the four Democratic candidates (three white and one black), we would expect $P'_{WW} = .75$, a value that is still very close to our estimate. If we refined our assumption to include approximately 1/3 of the voters voting for the two Republican candidates and one Independent candidate, our expected P'_{WW} would be .83. In any case, the key point is that a comparison of our estimated P'_{WW} based upon our equiprobability model of voter choice reveals little or nothing about polarized voting patterns. Thus we must turn to estimates of the proportions of voters of each race who voted for *each* candidate if we wish to understand racially polarized voting in Mecklenburg's multimember district senate election in 1982.

We will show how this is done for two of the seven candidates, C_3, the black candidate, and C_4, one of the white Democratic candidates. We begin with candidate C_3 by first regressing P_{VC_3} on x giving $m_4 = -.29$ (.01) and $b_4 = .44$ (.01). Thus

$$P_{WVC_3} = m_4 + b_4 = \underset{(.003)}{.15}$$

and

$$P_{BVC_3} = b_4 = \underset{(.01)}{.44}\ .$$

Note that, since there is only one black candidate in this race, this is the same as regressing P_B on x. We now regress P_{VNC_3} on x giving $m_5 = .28$ (.02) and $b_5 = .03$ (.01). Thus

$$P_{WVNC_3} = m_5 + b_5 = \underset{(.006)}{.31}$$

and

$$P_{BVNC_3} = b_5 = \underset{(.01)}{.03}\ .$$

In order to determine the proportions of those who actually voted who voted for C_3 we compute

$$P'_{WVC_3} = \frac{P_{WVC_3}}{P_{WVC_3} + P_{WVNC_3}} = \frac{.15}{.15 + .31} = \underset{(.006)}{.33}$$

and

$$P'_{BVC_3} = \frac{P_{BVC_3}}{P_{BVC_3} + P_{BVNC_3}} = \frac{.44}{.44 + .03} = \underset{(.03)}{.94}\,.$$

Thus approximately 33% of the whites who voted cast a vote for C_3, while 94% of the blacks who voted cast a vote for this candidate. Again, the standard errors for the ratio formulae used to calculate the proportions of black and white voters (rather than registered voters) who voted for C_3 are calculated using formulae derived from those for the variance of the sum of two variables and the estimate of the variance of the ratio of two variables.

These differences by race in the vote for the black candidate are striking and show clear evidence of racially polarized voting. Moreover, the results are statistically significant and the standard errors are tiny.

White candidates are handled in exactly the same manner. By regressing P_{VC_4} on x we find that $m'_4 = .07\ (.01)$ and $b'_4 = .18\ (.01)$, and by regressing P_{VNC_4} on x we find that $m'_5 = .07\ (.02)$ and $b'_5 = .29\ (.01)$. Skipping the intermediate steps of calculating P_{BVC_4}, P_{WVC_4}, P_{BVNC_4}, and P_{WVNC_4} we use the slopes and intercepts directly to calculate

$$P'_{WVC_4} = \frac{m'_4 + b'_4}{m'_4 + b'_4 + m'_5 + b'_5} = \underset{(.008)}{.54}$$

and

$$P'_{BVC_4} = \frac{b'_4}{b'_4 + b'_5} = \underset{(.02)}{.39}\,.$$

For this white candidate, differences between black and white levels of support were not very large; 54% of the white voters cast a vote for candidate 4 while 39% of the black voters voted for this candidate.

It is possible, as Goodman (1953, 1959) pointed out, to obtain estimates outside the interpretable range, in this case between 0 and 1, even when the assumptions of the model are met. Occasionally when racial polarization is extreme and there is some turnout variation between the groups of interest, estimates of the proportion of own-race voting may exceed 1 or estimates of cross-race voting may be less than 0. In such cases we set the value equal to 1 or 0, respectively. In North Carolina we have never found gross deviations of this kind, and when this does occur, extreme case analysis, discussed below, may be used to justify the adjustment. For example, in *Thornburg v. Gingles* (1986) the most negative estimates obtained were –.02. In every case the negative regression estimate corresponded to homogeneous case estimates of less than .05. Thus we believe it plausible, if other checks prove out, to treat estimates outside the (0, 1) range as falling at extreme values in that range.[7]

It is necessary to examine these proportions for all of the candidates. Analogous information is summarized in Table 1 for all seven candidates, each identified by party and by race. Table 1 shows the rankings of each candidate according to the proportion of voters of each race who voted for the candidate.

While we have assigned ranks to candidates as if the vote proportions from each group were without error, this is certainly not the case. The rankings, based upon estimates of proportions that have sampling errors, will also be subject to variability. However, the definition of racially polarized voting (see note 5) does not involve a precise ranking of candidates, which we have provided in Table 1 only to clarify discussion, but rather a demonstration that the sets of candidates that would be elected by each group differ.

Recall that four positions were up for election. Notice that according to the proportion of the black vote for each candidate the top four candidates in order are C_3(BD), C_1(WD), and C_2(WD) and C_4(WD). Thus if only black voters had voted in this election, these four candidates would have won. According to the ranking based upon the proportion of the white voters voting for each candidate, the top four contenders are C_6(WR), C_4(WD), C_2(WD), and C_1(WD). This is not the same set of candidates that would have been elected by the black voters since C_3(BD) has been replaced by C_6(WR). It is at this point that we have demonstrated the existence of racially polarized voting that is substantively significant (Grofman, 1985). As a point of information, the

TABLE 1

Candidate Vote Proportions by Race from the General Election for the State Senate in Mecklenburg and Cabarrus Counties, North Carolina, November 1982. Four Seats Up for Election

			P'_{BVC_j}				P'_{WVC_j}			
Candidate	Race	Party	From Regression	Regression Estimate of Ranking	From Extreme Case	Extreme Case Estimate of Ranking	From Regression	Regression Estimate of Ranking	From Extreme Case	Extreme Case Estimate of Ranking
C1	W	D	.41 (.02)	2	.31	4	.47 (.009)	4	.42	4
C2	W	D	.39 (.02)	3-4	.33	2-3	.54 (.01)	3	.52	3
C3	B	D	.94 (.03)	1	.86	1	.33 (.006)	6	.31	6
C4	W	D	.39 (.02)	3-4	.33	2-3	.54 (.008)	2	.53	2
C5	W	R	0 (.02)	5-7	.02	6	.40 (.007)	5	.41	5
C6	W	R	0 (.04)	5-7	.05	5	.59 (.008)	1	.61	1
C7	W	I	0 (.004)	5-7	.005	7	.02 (.001)	7	.02	7

NOTES: Numbers in parentheses are standard errors of the estimate. W = white candidate; B = black candidate; D = Democrat; R = Republican; I = Independent.

actual winners in this election were C_1, C_2, C_4, and C_6: reflecting the preferences of the white voters.

An examination of the standard errors of the vote proportion estimates is necessary at this point to assure ourselves that the variability of the rankings does not affect these results.[8] In this case, the standard errors of the vote proportion estimates are so small relative to the differences between the estimated proportions of interest, those of candidates 3 and 6, that our conclusion would not change. Examining the estimated proportion of the vote from white voters, it is quite unlikely that the black candidate, C_3, with an estimated proportion of .33 and a standard error of .006 would have been among the top four vote getters when the lowest estimated proportion of four declared winners of the white vote is .47 with a standard error of .009. Similarly, looking at the estimated black vote proportions, given the standard errors, it is very unlikely that any but candidates 1, 2, 3, and 4 would have been elected if only black voters had voted. As noted above, the sets of candidates that would have been elected are important, not the precise rankings of candidates.

VERIFYING THE PLAUSIBILITY OF OUR TWO-EQUATION BASED ESTIMATES OF THE MAGNITUDE OF RACIALLY POLARIZED VOTING

We now consider the plausibility that the regressions equations presented above, such as equations 3 and 6′, are approximately satisfied in all precincts (of course the identities, such as 1 and 2, which are aggregate-level statements, must be satisfied by definition).

By approximately satisfied we mean only that the nature of the error is uncorrelated with the independent variable, the proportion white in the electorate, since we are trying only to estimate the *average* behavior of white (and black) voters.[9] The same racial bloc can have somewhat different voting behavior in different precincts as long as our OLS estimates of $\bar{n}_W$ and $\bar{n}_B$, P_{BB}, P_{BW}, and so on, are unbiased; but these estimates will be unbiased *unless* we have correlated error, that is, unless the location of the errors from the best fitting regression line is a function of the independent variable.

There are three indirect ways to check the plausibility of our assumption that the voting behavior of voters of each race is roughly constant across precinct units. One way is to run a polynomial regression. If there is correlated error, this will be expressed in terms of a curvilinearity effect wherein parameters such as P_{WW} and P_{BB} will be functions not simply of x but also of x^2 and perhaps higher-order powers of x (Miller, 1977; Grofman, 1987). We have run such regressions for these data and find that introducing higher-order powers of x does not add more than one or two percentage points to the explained variance of the vote proportion going to the minority candidate(s), as compared with the simple linear model that yields r^2 values typically in the range of .6 to .8.[10]

A second way to check for the presence of possible biases in estimates caused by contextual effects is to make use of information derived from homogeneous precincts—precincts that are overwhelmingly of one race. If regression-based estimates coincide with those from the extreme case precincts, contextual effects of any magnitude are virtually impossible.

All of the above parameters were also estimated using a homogeneous case analysis (Duncan and Davis, 1953; Loewen, 1982). This was done by looking at precincts that are nearly all black or all white in their voting registration. In this example, precincts with white registration greater than or equal to 95% of the total registration (74 out of 143 precincts) were considered all white and precincts with black registration greater than or equal to 95% of the total registration (6 out of 143) were considered all black. The parameters previously calculated from our regression methodology were calculated directly from the vote totals for each set of homogeneous precincts.

For example, C_3 received a total of 17,622 votes in the 74 all-white precincts from a total of 56,282 voters. Thus the extreme case estimate of P'_{WVC_3} would be

$$P'_{WVC_3} = \frac{17622}{56282} = .31.$$

In the six all black precincts C_3 received 4,525 votes from 5,273 voters so that

$$P'_{BVC_3} = \frac{4525}{5273} = .86.$$

The proportions for the other candidates are calculated similarly and are reported in Table 1. Except in one case, the two sets of estimates are close enough (within 10%) so as to give us confidence that contextual effects are minimal.

Extreme case estimates of P'_{WW} and P'_{BB} are also possible and yield values of $P'_{WW} = .89$ and $P'_{BB} = .46$—virtually identical to our earlier values—further increasing our confidence in our regression-based two-equation methodology.[11]

A third way to check the plausibility of our parameter estimates is to compare them with an external check on their validity—namely, the total votes received in the election by each candidate. We have estimates of the proportion of votes cast by members of each race for candidates of each race. We also know the total population/registration figures by race. Multiplying these two factors gives us the total votes cast by members of each race for each candidate, the sum of which should simply be the actual vote received by each candidate. By comparing the estimated vote results with the actual vote results we have an indirect check on the plausibility of the assumptions used to generate our estimates from the two-equation model. When we do so we find that our estimates of total votes received are always within 10% of the actual total, and usually they are much closer.[12] Thus once again we have reason for confidence in the plausibility of the estimates generated by our two-equation method.[13]

Of course, the presence of contextual effects is likely to show up in low r^2 and in high standard errors. Thus low explained variance or estimates with high variance suggest the need for great care. For these data, however, no such problems arose.

However, although this methodology has been shown to be quite reliable given the assumptions, it must be noted that these assumptions are potentially crucial to obtaining unbiased estimates. That is, these methods cannot be used blindly. The most common problem is the existence of contextual effects. It is possible that such correlated error might occur; for example, blacks in highly black areas might be more likely to vote for a single candidate (most likely, the black candidate) than blacks in more integrated areas who might be more familiar with the white candidates and better able to differentiate among them. Similarly, whites in more nearly integrated areas might be more (or less) race prejudiced than whites living in all-white areas and thus might be different from these whites in terms of their propensity toward voting for candidates of the opposite race.

But even if there are contextual effects their biasing effects may not be serious, especially if one's chief concern is simply to classify candidates as to whether or not they have received enough votes from a group to have been elected if that group were the only group voting. We have identified a series of checks on the plausibility of our estimates, the inventory of which is original with the present authors. If there are potential problems with the data, some or all of these may be used in conjunction with the two-equation methodology to set bounds on the accuracy of the derived estimate and to determine if the methodology is inappropriate. We have made use of this methodology to measure the extent of racially polarized voting in well over a hundred elections in over a dozen jurisdictions, and we have considerable confidence in its overall reliability.

SEEMINGLY UNRELATED REGRESSION

Yet another check on the reliability of our estimates is derived from using the method of seemingly unrelated regressions (SURE) (Zellner, 1962). This methodology is used to deal with the case where we have a series of equations each of which uses the same independent variable.[14] We show in Table 2 variance and covariance estimates from OLS and SURE for the variables $\bar{n}$, P_W, P_B, and P_{VC_i} and P_{VNC_i} for i = 1, 7—a total of seventeen equations in all.

It is apparent from inspection of Table 2 that the variance and covariance estimates of OLS and SURE are essentially identical (to three significant digits) for our North Carolina data set. Thus we gain further confidence in the reliability of our parameter estimates.[15]

TABLE 2
Comparison of Variance and Covariance of Regression Coefficients Under OLS and SURE

	OLS		SURE	
	Variance	covariance	variance	covariance
	m_1: .003522		m_1: .003473	
$\bar{n}^{(i)} = m_1x^{(i)}+b_1$	b_1: .002678	-.002909	b_1: .002641	-.002868
	m_2: .002975		m_2: .002933	
$D_W^{(i)} = m_2x^{(i)}+b_2$	b_2: .002262	-.002457	b_2: .002423	-.002231
	m_3: .0001105		m_3: .0001089	
$P_B^{(i)} = m_3x^{(i)}+b_3$	b_3: .00008400	-.00009125	b_3: .00008282	-.00008997
	m: .0002698		m: .0002661	
$C1^{(i)} = mx^{(i)}+b$	b: .0002052	-.0002052	b: .0002023	-.0002198
	m: .0003793		m: .0003740	
$CN1^{(i)} = mx^{(i)}+b$	b: .0002884	-.0003133	b: .0002844	-.0003089
	m: .0001778		m: .0001753	
$C2^{(i)} = mx^{(i)}+b$	b: .0001352	-.0001468	b: .0001333	-.0001448
	m: .0002471		m: .0002437	
$CN2^{(i)} = mx^{(i)}+b$	b: .0001879	-.0002041	b: .0001853	-.0002013
	m: .0001105		m: .0001089	
$C3^{(i)} = mx^{(i)}+b$	b: .00008400	-.00009125	b: .00008282	-.00008997
	m: .0002878		m: .0002837	
$CN3^{(i)} = mx^{(i)}+b$	b: .0002188	-.0002377	b: .0002157	-.0002344
	m: .0002160		m: .0002130	
$C4^{(i)} = mx^{(i)}+b$	b: .0001642	.0001784	b: .0001619	-.0001759
	m: .0002436		m: .0002403	
$CN4^{(i)} = mx^{(i)}+b$	b: .0001853	-.0002012	b: .0001827	-.0001985
	m: .0001520		m: .0001498	
$C5^{(i)} = mx^{(i)}+b$	b: .0001156	-.0001255	b: .0001238	-.0001139
	m: .0002769		m: .0002731	
$CN5^{(i)} = mx^{(i)}+b$	b: .0002106	-.0002106	b: .0002076	-.0002255
	m: .0003281		m: .0003235	
$C6^{(i)} = mx^{(i)}+b$	b: .0002495	-.0002710	b: .0002460	-.0002672
	m: .0002003		m: .0001975	
$CN6^{(i)} = mx^{(i)}+b$	b: .0001523	-.0001654	b: .0001502	-.0001631
	m: .000003586		m: .000003535	
$C7^{(i)} = mx^{(i)}+b$	b: .000002727	-.000002962	b: .000002688	-.000002920
	m: .0004084		m: .0004027	
$CN7^{(i)} = mx^{(i)}+b$	b: .0003106	-.0003374	b: .0003062	-.0003327

NOTE: C1 and so on = vote for candidate 1/RT; CN1 and so on = vote not for candidate 1/RT.

CONCLUSION

We have shown how it is possible to develop reliable methods to estimate black and white voting behavior using aggregate census and electoral data even in situations in which blacks and whites differ in their turnout and/or in the number of candidates for whom they cast ballots. This methodology should prove to be of considerable importance in voting rights litigation in any of the more than half of our nation's cities that use some form of multimember district election. Moreover, the two-equation ecological regression technique we offer should have a number of other applications, for example, in the analysis of historical data where the aim is to identify how two or more (mutually exclusive) *groups* differed in their (voting or other) behavior in situations where it is possible to match behavioral outcomes to group percentages across areal (or other) units.

NOTES

1. In many multimember district elections, where there are k seats open, each voter has k votes, but there exist k numbered districts (not necessarily geographic in nature) such that in each district there is a single seat to be filled. Such numbered place systems give rise to the equivalent of head-on-head contests in single-member district systems, except that the electorate is that of the entire multimember district. In this article we shall be concerned with multimember districts without such a numbered place system. For analysis of racial bloc voting in head-on-head contests, see Grofman and Noviello (1984), and Grofman et al. (1986).

2. In the case of North Carolina data we make use of, for illustrative purposes, information on the race of each registered voter is part of the registration rolls.

3. While it is possible to apply standard statistical tests to determine the likelihood that the observed P'_{WW} and P'_{BB} values could have been obtained by chance from a racially neutral pattern of voting, such a test will determine only the existence of racial polarization and not its magnitude.

4. Alternatively, we can regress T on x and estimate T_B, the proportion of black registered voters who turn out to vote, as the intercept, and similarly regress T on $1 - x$ to estimate T_W. We can then calculate the proportion of voters who are black from the $BVP^{(i)} = (1 - x^{(i)})T_B/[x^{(i)}T_w + (1 - x^{(i)})T_B]$ and WVP, the proportion of the voters who are white equals, $1 - BVP$. If we then regress CVS_j, the proportion of the *voters* who vote for the black candidate, C_j, on WVP, the intercept gives us P'_{BVC_j} and similarly the intercept from the regression of CVS_j on BVP gives us P'_{WVC_j}. We have used this alternative technique on other data sets. Differences between the estimates obtained and those

derived from the method in the set were minimal, and we shall not discuss it further.

5. In the case of multimember districts without a numbered place system, voting would be said to exhibit substantively significant racial polarization if the *set* of candidates that would be elected by voters of one race differs from the *set* of candidates that would be elected by voters of the other race. *Gingles v. Edmisten*, 1984; *Thornburg v. Gingles*, 1986). For example, if there was one black candidate and if white voters (on average) ranked the black candidate fifth and black voters (on average) ranked the black candidate first, then substantively significant racially polarized voting as defined above would occur only if four or fewer candidates were being chosen. If there were more than one black candidate, substantively significant racial polarization would occur if majority voters vote for some minority candidates as long as they would not elect as many minority candidates as would the minority community.

6. If $r = y/x$ where x and y are themselves estimates with variability, then the variance of r may be estimated by $1/x^2\,[\mathrm{var}(y) + r^2\mathrm{var}(x) - 2r\,\mathrm{cov}(y, x)]$. However, this estimate is biased, but the appropriateness of this estimate can be evaluated by examining the coefficient of variation of the denominator of the ratio that is estimated by the ratio $Sx/\bar{x}$, where Sx is the estimated standard deviation of x and $\bar{x}$ is the estimated mean of x. Rules for evaluation vary, but a rather strict cutoff would seem to be .10 (see Kish, 1965: 209, which cites a value .1 to .2). That is, if the coefficient of variation of the denominator is less than .10 then this approximation will be satisfactory. For the vote proportion we estimate each candidate received from each race, for example (presented in Table 1), this coefficient was never greater than .05, with most falling in the range of .01 to .03, indicating that our estimates are only trivially biased at worst and certainly not enough to make any practical difference.

7. Goodman (1953) suggested redoing the regression estimates, forcing the equation through the origin when negative estimates were obtained. We have done this with some elections and found no substantial difference in the vote proportion estimates from those given by the two-equation methodology.

8. An examination of these standard errors will consider only the variance in vote proportions for the given ordering. This is likely to underestimate the actual variance since there is also a component of variance across different orderings. The estimate of this component is beyond the scope of this article. However, our view is that the difference is likely to be so trivial that it can be safely neglected. Further checks below would indicate whether there was likely to be a problem.

9. Unlike many other situations in which the nature (and causes) of variability in a population is the prime focus of interest, we do not care how much individual level of variety in behavior there is within groups as long as we can obtain a reliable estimate of between-group mean differences, since that is what racial bloc voting is all about. In particular, if we can obtain reliable estimates of a group's average behavior, we can be confident that the mean differences we find are real and not artifactual.

10. Polynomial runs with other data sets have also failed to find strong contextual effects for bloc voting patterns.

11. Some differences between regression-based estimates and those obtained from the average values of homogeneous districts are to be expected due to both contextual effects and the effect of class differences between districts that are almost entirely white and those that contain some numbers of blacks. The regression-based estimates will normally be slightly superior to those obtained from extreme case analysis since the regression estimate is based upon all of the data and thus provides a better estimate for the behavior of the average white/black voter (Grofman et al., 1985; Grofman, 1987).

12. The actual vote totals for each of the seven candidates from the empirical illustration discussed later are given along with the estimates (in parentheses) produced by using equation 14, which must be used instead of equation 6′ when performing the candidate-by-candidate analysis necessary for multimember district situations.

c_1 = 46412 (48267)	c_4 = 54048 (54497)	c_7 = 2071 (2183)
c_2 = 53860 (54653)	c_5 = 34446 (35598)	
c_3 = 45115 (46012)	c_6 = 50536 (51292)	

Our worst estimate is for candidate number 7, for which we were off by 5.43%. The rest of the estimates were off by 4.00% or less with 4 of the 7 estimates off by less than 2.00%.

13. A fourth technique, complementary homogeneous case analysis, is described in Grofman et al. (1986). This technique is a variant of the Duncan and Davis (1953) overlapping percentages method.

14. Seemingly unrelated regressions (SURE) is also known as Zellner estimation and iterative Zellner estimation. SURE is used when there is a number of equations (such as for different candidates) considered as a single model. In our situation it is likely that errors for the different candidates in the same precinct will be correlated—this leads to a violation of the scalar identity covariance matrix assumption (i.e., the off diagonal entries of the residual covariance matrix are not equal to zero). If OLS is used in this situation, the estimates will be unbiased, but they will not be minimum variance estimators. SURE uses the generalized least-squares estimator that is best linear unbiased.

15. We are not proposing that all of these checks should be done as a matter of course. All of the checks need be carried out only when problems with the data seem to be present. In most cases when severe racially polarized voting is in fact present, the patterns will be clear and these checks will not be necessary except to satisfy any doubts the expert himself or herself may have.

REFERENCES

BRACE, K., L. HANDLEY, B. GROFMAN, and R. NIEMI (1986) "What's special about 65%? The mathematics of minority voting equality." Presented at the annual meeting of the Midwest Political Science Association, Chicago, April.

BULLOCK, C. S. III. (1983) "The effects of redistricting on black representation in southern state legislatures." Presented at the annual meeting of the American Political Science Association, Chicago, September 1-4.

DUNCAN, O. D. and B. DAVIS (1953) "An alternative to ecological correlation." Amer. Soc. Rev. 18, 6: 665-666.

GINGLES v. EDMISTEN (1984) 590 F. Supp. 345 (E.D. N.C.).

GOODMAN, L. A. (1953) "Ecological regression and the behavior of individuals." Amer. Soc. Rev. 18, 6: 663-664.

GOODMAN, L. A. (1959) "Some alternatives to ecological correlation." Amer. J. of Sociology 64, 6: 610-625.

GROFMAN, B. (1982) "Alternatives to single-member plurality districts: legal and empirical issues." Policy Studies J. 9, 3: 875-898.

GROFMAN, B. (1985) "Criteria for districting: a social science perspective." UCLA Law Rev. (October): 77-184.
GROFMAN, B. (1987) "Models of voting," pp. 31-61 in S. Long (ed.) Micropolitics Annual, Vol. 2. Greenwich, CT: JAI Press.
GROFMAN, B., J. LOEWEN, and M. MIGALSKI (1986) "Improving ecological regression." (unpublished manuscript)
GROFMAN, B., M. MIGALSKI, and N. NOVIELLO (1985) "The totality of circumstances test in Section 2 of the Voting Rights Act: a social science perspective." Law and Policy 7, 2: 199-223.
GROFMAN, B. and N. NOVIELLO (1984) "An outline for the analysis of racial bloc voting." Prepared testimony in *Gingles v. Edmisten*, North Carolina.
KISH, L. (1965) Survey Sampling. New York: John Wiley.
KOUSSER, J. M. (1974) The Shaping of Southern Politics: Suffrage Restriction and the Establishment of the One-Party South, 1880-1910. New Haven: Yale Univ. Press.
LOEWEN, J. (1982) Social Science in the Courtroom. Lexington, MA: Lexington Books.
MILLER, W. L. (1977) Electoral Dynamics in Britain since 1918. London: Macmillan.
NIEMI, R., J. HILL, and B. GROFMAN (1985) "The impact of multimember districts on party representation." Legislative Studies Q. 10, 4: 441-455.
Report of the Committee on the Judiciary of the U.S. Senate on S1992 (1982) Government Printing Office.
ROBINSON, W. S. (1950) "Ecological correlations and the behavior of individuals." Amer. Soc. Rev. 5, 3: 351-357.
THEIL, H. (1971) Principles of Econometrics. New York: John Wiley.
THORNBURG v. GINGLES (1986) 106 S. Ct. 2752.
ZELLNER, A. (1962) "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias." J. of the Amer. Stat. Assn. 57: 348-368.

Bernard Grofman is Professor of Political Science and Social Psychology, School of Social Sciences, University of California, Irvine. He is a specialist in mathematical models of collective decision making and the political consequences of electoral law, with over seventy published articles on topics such as jury verdict choice, reapportionment and voter turnout, and coalition formation models. During the past five years he has been involved in eleven states as an expert witness in redistricting litigation or as a court-appointed reapportionment expert.

Michael Migalski is a Ph.D. candidate in the mathematical social sciences program at the University of California, Irvine. He served as a Research Assistant in testimony prepared for Gingles v. Edmisten.