HM ECOLOGICAL INFERENCE

Ecological Inference from Goodman to King

J. MORGAN KOUSSER
Division of the Humanities and Social Sciences
California Institute of Technology

Since it was introduced to historians nearly three decades ago, a statistical technique known as *ecological*[1] *regression* has been widely used to analyze aggregate election returns and similar data in history, political science, and law, and methodologists have discussed problems with, extensions of, and alternatives to the technique. The literature has become so vast and complicated, and recent contributions to it are of such importance, that it is time for a comprehensive review. This article provides that review, starting at an elementary level, sorting through the major arguments and evidence, and explaining the nature of the most complicated as well as the simplest methods. Taking an intuitive, rather than a statistically rigorous approach, the article is aimed at historians and political scientists, particularly graduate students, who already have some statistical knowledge. Lawyers and expert witnesses in voting-rights cases may also find the article useful.

The Fundamental Problem of Ecological Inference

The fundamental problem of ecological inference is easy to explain. Often observers would like to know, for example, how members of certain political, social, or economic groups voted in some past election, but the only data that have survived give the total votes and the total number of members of the groups in precincts, townships, counties, or states. (Voting provides a convenient example, one that I will use throughout, but any other measured behavior or characteristic, such as buying a product, attaining literacy, or holding a certain amount of wealth, could be analyzed analogously.) We need some means not only of estimating how individuals with certain characteristics voted but also of assessing the reliability of those estimates. But as William Robinson (1950) pointed out in a famous article, it may be misleading to infer individual behavior, in the simplest and most obvious ways, from aggregate data. Introducing a term that frightened many scholars away from using aggregate data for years, Robinson called this the *ecological fallacy*.[2]

Since the 1950s, there have been three basic approaches to overcoming the ecological fallacy—(1) the *method of bounds*, (2) *homogeneous areas*, and (3) *ecological regression*. Recently, Gary King (1997) has combined the method of bounds with simulation procedures to produce a fourth technique, which has come to be known as EI (an acronym for *ecological inference*). The best way to understand the basic concepts of all four is to start with the method of bounds and to use one example consistently throughout. For this purpose, I have chosen the hotly contested North Carolina gubernatorial election of 1896. I will assume here that the smallest unit for which voting and census information is available is the county.

Table 1, a *contingency table*, relates the percentage Republican to the percentage black in a hypothetical county in North Carolina in 1896. U.S. census figures tell us that the county is 73 percent white and 27 percent black, whereas election returns show that the two-party vote divided 60:40 in favor of the Democrats. To simplify exposition, we assume for the present that every adult male in the county turned out to vote and that the racial percentages are the same for the eligible voters as for the population as a whole. The assumptions will be relaxed later. We denote the known *marginal* entries (those on the margins of the table), which

TABLE 1
Hypothetical Relationship between Party and Race in North Carolina, 1896

Race	Party: Republican	Party: Democratic	
White	P_{11}	P_{12}	$X_1 = .73$
Black	P_{21}	P_{22}	$X_2 = .27$
	$Y_1 = .40$	$Y_2 = .60$	

sum to 100 percent by definition, by *X*s and *Y*s for generality. Because no one recorded how the members of each race voted, we fill the internal or *partial* cell entries with letters:

P_{11} = proportion of whites who voted Republican,
P_{12} = proportion of whites who voted Democratic,
P_{21} = proportion of blacks who voted Republican,
and
P_{22} = proportion of blacks who voted Democratic. (1)

Note that because we have defined the partials in this way, each row adds to 100 percent. That is,

$$P_{11} + P_{12} = P_{21} + P_{22} = 1.0. \quad (2)$$

This method is different from defining the partials to sum to 100 percent across the whole table. In that case,

$$P_{11} + P_{12} + P_{21} + P_{22} = 1.0. \quad (3)$$

For any particular county, the marginal entries put some constraints on the values that the partials can attain—but usually not enough constraints, in such a simple table at least, to be very useful. For instance, a little arithmetic shows that the percentage of whites who voted Republican could not be over 55 percent, even if no blacks voted Republican (see table 1). By definition,

% Republican = (% whites who voted Republican
* % white) + (% blacks who voted Republican
* % black), (4)

where an asterisk signifies multiplication and parentheses have been inserted for clarity. Substituting the symbols from table 1 for the words in equation (4), we have

$$Y_1 = (P_{11} * X_1) + (P_{21} * X_2). \quad (5)$$

Because the maximum number of white Republican votes would be tallied if all Republicans were whites and there were no black Republicans, we just set $P_{21} = 0$, substitute in the remaining numbers for the particular county, and solve the resulting equation to compute the maximum:

$$.40 = (P_{11} * .73) + (0 * .27) \quad (6)$$

$$P_{11} = .40/.73 = .55. \quad (7)$$

To find the minimum number of whites who must have voted Republican, we assume that all the blacks did, which amounts to assuming that $P_{21} = 1.0$ in equation (5). We then fill in the other numbers that we know and solve:

$$.40 = (P_{11} * .73) + (1.0 * .27) \quad (8)$$

$$P_{11} = (.40 - .27)/.73 = .18. \quad (9)$$

Therefore, from the marginal totals alone we know that, in this county, between 18 percent and 55 percent of the whites voted Republican. But note that the county is so heavily white and so evenly balanced between the parties that the marginals do not constrain estimates of black voting at all. That is, from these marginal percentages, all we can say is that between 0 and 100 percent of the blacks voted Republican, because if we substitute 0 and 1.0 for P_{11} in equation (5), we get 148 percent and –122 percent, respectively. We have to set these numbers to 100 percent and 0 percent, because no fewer than zero and no more than all the persons in a group can vote. This shows that at least some whites must have voted Republican. Unless we have other information or substantive or statistical theory, we cannot narrow our estimate of racial voting behavior in this county.

It is useful to portray data like those in table 1 in what I will call a *bounds plot*.[3] Figure 1 is a *unit square*, a square in which each side is assigned values from 0 to 1. In this instance, the vertical axes give the range of possible percentages of African Americans who voted Republican, and the horizontal axes give the range of possible percentages of whites who voted Republican. The line crossing the unit square in figure 1 reflects the fact that the marginal percentages for the hypothetical county in table 1 constrain the percentage of whites who voted Republican to lie between 18 percent and 55 percent. For every point on the line connecting 18 percent on the top to 55 percent on the bottom, there corresponds an estimate of the proportion of African Americans voting Republican.

The values of the pair of points are the projections on the vertical and horizontal axes, respectively. For instance, figure 2 shows that if we somehow knew that 30 percent of the whites in this county voted Republican, we would also know for certain that 67 percent of the blacks did. Figure 2 also illustrates the fact that if we had other information that convinced us that at least two-thirds of the blacks voted Republican, then we would need to consider only that part of the line connecting 18 percent and 55 percent that lies

FIGURE 1
Bounds Plot for Data in Table 1

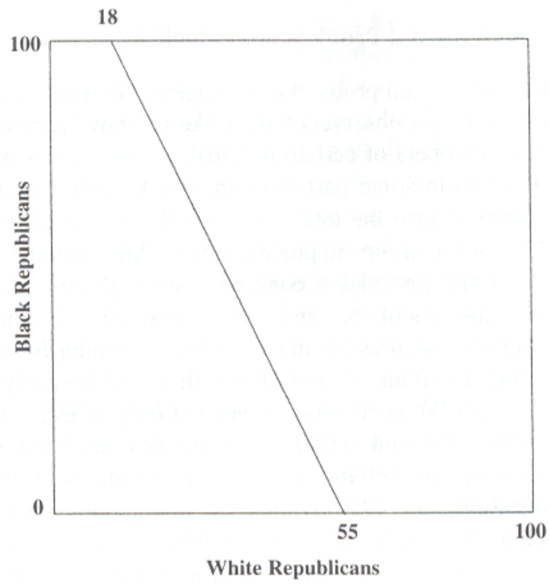
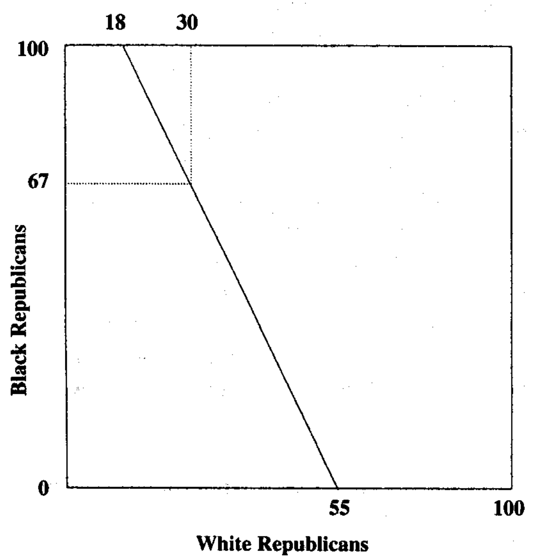


FIGURE 2
Bounds Plot for Data in Table 2, with Added Assumption that at Least Two-Thirds of African Americans Voted Republican



above the point where the two dashed lines meet. That part of the line corresponds to white Republican percentages of 18–30 percent and black Republican percentages of 66.7–100 percent. Similar lines would be dictated by the marginal percentages for each county, some intersecting both the horizontal and vertical axes and perhaps others intersecting only the horizontal axis. If we knew that we had to consider only parts of each line, we might obtain narrower estimates of voting behavior than if we had to consider every whole line.

The Homogeneous Areas Method

The method of bounds not only provides a convenient introduction to the fundamental problem but also is centrally involved in solutions. For what three of the general approaches do is to narrow the bounds by making assumptions. Ecological regression does not take the bounds explicitly into account. The *homogeneous areas* assumption is the simplest. Suppose that everyone in North Carolina in 1896 lived in counties that were either all white or all black, as illustrated in the two entirely hypothetical counties in table 2. In the 100 percent white county in table 2, panel A, 35 percent of the people voted Republican, 65 percent voted Democratic, and everyone was white; so we can be quite certain how whites in that county voted, as the partials in italics stress. Similarly, in table 2, panel B, which represents an all-black county, 90 percent of the people voted Republican, and 10 percent voted Democratic, so again we can say absolutely that the African American vote went 90:10 Republican there. The bounds in such counties are thus as narrow as possible. In a bounds plot, a county that was completely homogeneous would not be represented by a line, but by a single point on an axis.

TABLE 2
The Homogeneous Areas Method: Two Hypothetical Homogeneous Counties in North Carolina, 1896

Race	Party: Republican	Party: Democratic	
	A. An all-white county		
White	$P_{11} = .35$	$P_{12} = .65$	$X_1 = 1.0$
Black			$X_2 = 0$
	$Y_1 = .35$	$Y_2 = .65$	
	B. An all-black county		
White			$X_1 = 0$
Black	$P_{21} = .90$	$P_{22} = .10$	$X_2 = 1.0$
	$Y_1 = .90$	$Y_2 = .10$	

Note: Partial cell entries are italicized to stress that these hypothetical counties are completely homogeneous.

The obvious problem for observers is that voters are not completely geographically segregated, at least in surviving election data.[4] As a consequence, those who rely on the homogeneous-areas approach must assume that every other voter in the state (or other large area of interest) voted the same way as the voters in the most segregated counties and that areas that were not quite homogeneous were close enough so that the bounds are quite narrow. Substantively, the social processes in counties, wards, or townships that are very heterogeneous may differ considerably from those in which one group is overwhelmingly dominant. Statistically, in states like North Carolina in 1896, where even the most heavily black county was only 60 percent black, the bounds are insufficiently constricted to yield very reliable estimates, even in the most homogeneous counties (Kousser 1976; Lichtman and Langbein 1978; Kousser and Lichtman 1983).

The Method of Bounds

To estimate the behavior of voters throughout a state, the *method of bounds* calculates the logical bounds for each county—as in the previous discussion of table 1—weights them by the population in each county, adds them for all the counties in the state, and divides by the total state population (Duncan and Davis 1953; Shively 1974, 1991; Flanigan and Zingale 1985). In the 1896 governor's race in North Carolina, calculations for the bounds in every county produces the estimate that between 30 percent and 71 percent of the whites and between 7 percent and 100 percent of the blacks voted Republican, ranges that are much too inexact to be useful.[5] To overcome the problem that the resulting logical bounds may be too wide to be of much substantive interest, W. Phillips Shively suggested that the bounds for some cells in certain kinds of contingency tables might be shrunk by making two kinds of assumptions: (1) assumptions about the direction and strength of relationships among the internal cell entries (such as the *P*s in table 1) and (2) assumptions about the unknown probability distributions of the values within the range of estimates of those cells.[6] The discussion of figure 2 illustrated the first kind of assumption. If at least two-thirds of the blacks voted Republican, then the bounds for the white Republican percentage shrink from a maximum of 55 percent to a maximum of 30 percent. As an illustration of the second assumption, suppose that one is further willing to assume that the midpoint between the bounds is the best guess of the likely true value. Then one might say that the likely percentage that the Republicans got among whites was 24 percent [18 + (30 – 18)/2] = 24). If 24 percent of whites were Republican, then so were 83 percent of blacks, because if we substitute the values into equation (3), .40 = (.24 * .73) + (P_{21} * .27); then P_{21} = .83. For the state as a whole, calculations not given here show that if at least two-thirds of the African Americans voted Republican, then between 30 percent and 45 percent of the whites did. If the two assumptions are combined, and one posits that at least 67 percent of blacks voted Republican and that the midpoint of the truncated lines is the likeliest percentage, then 37 percent of the whites and 83 percent of the African Americans would have voted Republican. If, on the other hand, one makes no assumption about the proportion of African Americans who voted Republican, but instead computes the midpoint of the bounds lines for all counties and then sums the results, the statewide estimate of Republican voting is 51 percent among whites and 54 percent among blacks. Note that there is no substantive or statistical reason to focus on the midpoint here. The appropriate point might be close to one edge of the unit square, or it could differ from county to county. As the examples demonstrate, assumptions can have powerful effects on estimates.

Different assumptions, some far more complex and specific, could be made in other cases. In an article illustrated by the major-party vote in the 1968 and 1972 U.S. presidential elections, for instance, Wiliam Flanigan and Nancy Zingale (1985, 86) make three separate assumptions about the relationships between the cells of a four-celled table. In multicelled tables, not all entries can be estimated exactly with just these two types of assumptions. Weak tests can be made by observing whether the calculations produce logically impossible totals (i.e., numbers implying that over 100 percent or less than 0 percent of a group voted in a certain way) for some or many counties in a state. We can vary assumptions to see how robust the estimates are to changes in guesses about the relationships. Often, they are not very robust at all.

Simple Ecological Regression

The wide range of estimates usually produced by the method of bounds and the unrepresentativeness of the homogeneous areas method led the statistician Leo Goodman to develop ecological regression (ER) in 1959. Goodman's method incorporates information from all counties but with different assumptions from those of the two approaches considered previously. Unlike the homogeneous areas method, ER does not necessarily assume that groups voted exactly the same way in every county. Unlike the method of bounds, it does not take the logical bounds explicitly into account; and it does not look at each county separately, as the method of bounds does, but rather it uses data from all the counties together to produce estimates. A bit of first-year algebra is all that is needed to grasp the fundamentals of ER. Referring to table 1 again, we know that

$$X_1 + X_2 = 100\%, \tag{10}$$

because everyone was either white or black. The simplest algebra tells us that $X_1 = 1 - X_2$. Substituting this into (5), we obtain

$$(P_{21} * X_2) + [P_{11} * (1 - X_2)] = Y_1. \tag{11}$$

Multiplying out, removing the asterisks for multiplication, and rearranging terms, we have

$$Y_1 = P_{11} + (P_{21} - P_{11})X_2. \tag{12}$$

This equation has the same form as a simple two-variable ordinary least squares (OLS) regression equation, which may be found in any elementary statistics textbook:

$$Y = a + bX + e, \tag{13}$$

where e is an error term added in general, in OLS, because our estimate will not usually be perfect. Numerous computer software packages and many hand-held calculators will spew out estimates of the parameters (a, b, etc., the unknowns to be estimated from the empirical data) and many associated statistics in an instant, and statistics texts will give formulas and further explanations.

The crucial assumption when OLS is used to estimate the partials is that the error is random—that people in every county voted in the same pattern, except for small idiosyncratic differences in behavior that basically counterbalance each other. If both blacks and whites who lived in the eastern North Carolina counties that produced a lot of tobacco, for instance, voted quite differently from those in the Piedmont and the mountains, the assumption would be violated and the estimates would be unreliable.

It is useful to visualize ER and the homogeneous areas method in a scatter plot, such as figure 3, in which the values for the percentage black in each county are plotted on the horizontal or X axis, whereas those for the percentage Republican are on the vertical or Y axis, so that each point represents the intersection of a county's percentage of

FIGURE 3
Racial Polarization in the North Carolina Governor's Race, 1896

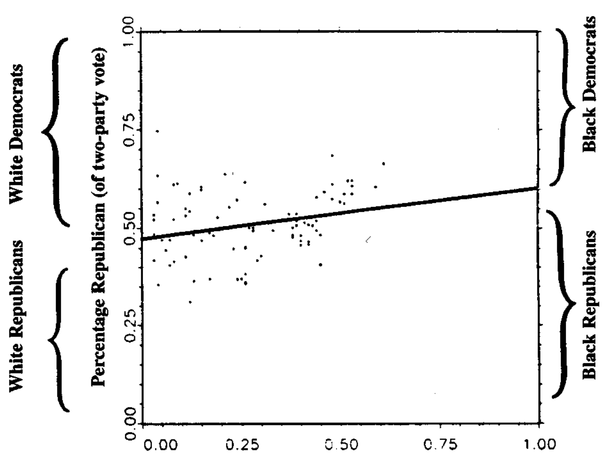
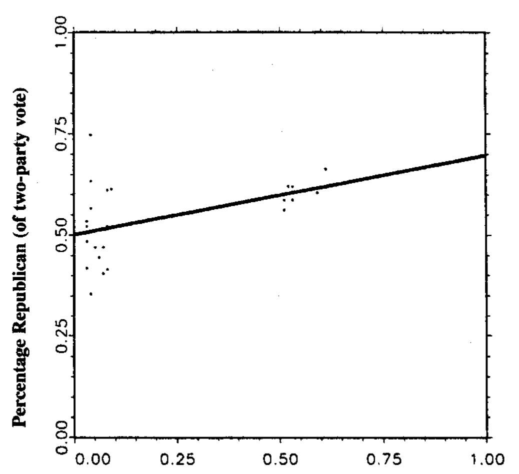


FIGURE 4
AU Plot of the Most Racially Homogeneous Counties in the North Carolina Governor's Race, 1896



blacks with its percentage of Republicans.[7] I will refer to this as an *aggregate unit* (*AU*) plot. Superimposed on the points is a *least-squares regression line,* which is the line that, in a particular statistical sense, best fits all the data points. The point at which the line crosses the Y axis (the Y *intercept*) is the one at which the percentage of blacks in the population is 0, which tells us, in effect, what the vote would have been in a county if all its citizens were white, whereas the point at which the regression line crosses a line drawn perpendicular to the X axis at 100 percent black tells us how the vote would have gone in an all-black county. The increase in the Y value as the line crosses the graph is referred to as the *slope* of the regression line. Because in the simplest Goodman model we assume that all whites and all blacks voted either Republican or Democratic, and that members of each racial group voted the same way throughout the state, we can find the percentage of whites who voted Democratic by subtracting the percentage who voted Republican from 100 percent, and we can do the same for blacks. This is illustrated by the curled brackets in figure 3.

Equation (12) is very easily related to figure 3. P_{11} is the Y intercept of the regression line, or the proportion of whites who are estimated to have voted Republican in the whole county, which is equal to about 47 percent in this instance. P_{21} is the point at which the regression line intersects a vertical line drawn perpendicular to the point at which X is 100 percent, or about 60 percent in this case. Thus, P_{21} is found if we begin with the Y intercept and add or subtract the amount that the regression line rises or falls as X changes from 0 percent to 100 percent.

The homogeneous areas assumption can be illustrated for an AU plot in figure 4, which looks at subsets of the North Carolina counties that have the most or the least African Americans. Note that in this real-world example, only one county falls on the left vertical axis, meaning that it is all white, and no county even approaches the right vertical axis, the line signifying a 100 percent black county. The world is usually not as homogeneous as required by the homogeneous areas assumption. Note also that there is a good deal of variation in the political behavior of the almost entirely white counties and that there are relatively few counties in the plot at all. We should be chary about generalizing under these conditions.

Attacks on Simple Ecological Regression: Neighborhood Models

Figures 3 and 4 suggest some of the difficulties with ER that arise because in practice some or all units are heterogeneous and because the points do not all fall on one line. Difficulties caused by possible variations in behavior within and across counties—violations of the *constancy assumption* of simple ER (hereinafter referrred to as SER) are best illustrated by reviewing a model that was invented in a nihilistic effort to undermine confidence in any method of inferring individual behavior from aggregate data: the "neighborhood model."[8] Because neighborhood models raise fundamental questions about the use of every ecological inference technique not only in legal cases but also in normal academic scholarship, because they are virtually unknown in the historical literature, and because they may not be fully understood even by political scientists and those connected with voting-rights litigation, it is important to treat them in some detail.

One of the crucial questions in minority voting-rights cases in federal courts is whether the voting patterns of Anglos, African Americans, and Latinos are "racially polarized," and the principal technique for answering the question has been ER.[9] Would regression estimates based on equation (12), for example, show that majorities of whites and blacks voted for opposing parties if it were applied to the 1896 North Carolina data? Was the relationship between race and party statistically significant—for example, can we be sure that the regression line in figure 3 should not actually be horizontal? Because plaintiffs bear the burden of proof in such cases, it occurred to expert-witness statisticians for defendants in the Los Angeles County Supervisors' case, *Garza v. Los Angeles County Board of Supervisors* (1991), that if they could show that SER did not necessarily produce accurate or reliable estimates of racially polarized voting, then the plaintiffs would be unable to prove their case and the defendants would win by default.

Table 3 compares assumptions in two varieties of the neighborhood model with those of the SER model. The nonlinear neighborhood (NLN) model outlined in panel A assumes that there is no relationship at all between race and voting behavior among individuals, but that race and other socioeconomic variables may, for unspecified reasons, be correlated with voting behavior from county to county. Within each county, therefore, everyone is assumed to have voted in the very same way. Because individual electoral behavior is assumed not to be at all correlated with the racial percentages within each county or from county to county, this rather formless model is termed *nonlinear*. Based on one extremely strong assumption, the NLN model disregards any available data about the socioeconomic characteristics of counties and does not even informally incorporate information about individual voting behavior, as the method of bounds may. It "proves" that no correlation exists between ethnicity and politics (or any other variables, such as voting for the candidates of the same party in successive elections) by assuming that none exists and by leaving no possible place in the model for any data that might test or modify that assumption.

The bivariate linear neighborhood (BLN) model depicted in panel B is slightly more complicated. As in the NLN version, members of both races are assumed to vote in the same way in each county, but now the percentage of both racial groups that vote for each party depends on the ethnic percentages in the individual county. In other words, in the NLN model the correlates of the social and political composition of counties were entirely unspecified, whereas in the BLN model we assume that one of those correlates is ethnicity. Within each county, people act the same way in the relevant elections either because they become similar through interacting with each other or because people who want to interact with the sorts of people in the county move there, while people who want to interact with other sorts leave. For instance, those whites who reside in a multiracial community are assumed to have preferences identical to their black, Latino, or Asian neighbors,whereas whites who flee to all-Anglo suburbs are assumed to share the preferences of their new neighbors. Furthermore, county lines and lines in these behaviorally uniform communities are assumed to coincide (Lichtman 1991, 791). Of course, if different counties have different percentages of blacks and whites, they will vote differently, and the summation of these differences across an electoral universe such as the State of North Carolina may imply that blacks voted differently from whites in the state as a whole. But all "neighbors" are presumed to have voted identically.

The SER model encapsulated in panel C assumes that

TABLE 3
A Comparison of the Assumptions in Three Models in One Hypothetical County in North Carolina, 1896

Race	% Republican	% Democratic	Total
	A. Nonlinear neighborhood model (NLN)		
White	$P_{11} = .40 + e_1$	$P_{12} = 1 - P_{11} = .60$	$X_1 = .73$
Black	$P_{21} = .40 + e_1$	$P_{22} = 1 - P_{21} = .60$	$X_2 = .27$
	$Y_1 = .40$	$Y_2 = .60$	
	B. Bivariate linear neighborhood model (BLN)		
White	$P_{11} = a + bX_{2j} + e_2 = .47 + (.13) * (.27) = .51$	$P_{12} = 1 - P_{11} = .49$	$X_1 = .73$
Black	$P_{21} = a + bX_{2j} + e_2 = .47 + (.13) * (.27) = .51$	$P_{22} = 1 - P_{21} = .49$	$X_2 = .27$
	$Y_1 = .40$	$Y_2 = .60$	
	C. Simple ecological regression model (SER)		
White	$P_{11} = a + e_3 = .47$	$P_{12} = 1 - P_{11} = .53$	$X_1 = .73$
Black	$P_{21} = a + b + e_3 = .47 + .13 = .60$	$P_{22} = 1 - P_{21} = .40$	$X_2 = .27$
	$Y_1 = .40$	$Y_2 = .60$	

Note: In panels B and C, the parameters *a* and *b* are taken from the same statewide ecological regression of the Republican percentage of the two-party vote on the black percentage of the population, namely, $a = .472$ and $b = .129$.

whatever the relation between ethnicity and voting behavior, a regression based entirely on data for counties will yield the same results as one would if data were available for every individual in the state. Moreover, the probability that an African American would vote Republican, for instance, would not be systematically different whatever county he lived in. In the neighborhood models, everyone in the county votes in the same way, regardless of race. By contrast, in the SER model, the voting patterns of members of particular ethnic groups vary only randomly both within counties and across the state.

Note that if everyone lived in ethnically homogeneous counties, NLN, BLN, SER, the method of bounds, and any other conceivable model would yield exactly the same estimates of ethnic voting (but not necessarily of the votes of those of different classes or political parties, etc.). The bounds would be exact and narrow to a series of points; the regression line would go through the mean for each political choice for each set of counties comprising members of separate ethnic groups; the neighborhood models would merely tally the numbers for each party in each set of segregated counties. Thus, total segregation would make calculation simple and statisticians superfluous. The example argues for putting considerable emphasis on the most homogeneous counties or precincts because, in them, estimates from different models converge. It suggests, as well, that a test for the proportion of people who live in relatively ethnically homogeneous counties might indicate how likely we are to find useful estimates with any model.

A comparison of panels A–C of table 3 for our hypothetical county makes this discussion more concrete. In the NLN model, the percentage of every group voting Republican in the county is the same, 40 percent. (We include error terms in models A–C to indicate that the relationships are not presumed to be deterministic but represent estimates instead.) If the citizens of another county voted 55 percent Republican, all groups in that county would be assumed to have voted 55 percent Republican. To get the total estimated number of blacks who voted Republican in the state, one simply multiplies each county's percentage Republican by the number of blacks and adds up the results. We then obtain the percentage of the state's blacks who voted Republican by dividing by the number of blacks in the state. Analogous figures for other subgroups are found similarly. If everyone is assumed to have voted Republican or Democratic in the 1896 governor's contest in North Carolina, the NLN estimate for the state as a whole is that 52.0 percent of the blacks and 50.4 percent of the whites cast ballots for the GOP.

As explained above in the discussion accompanying equations (12) and (13), one obtains partial-cell entries for the SER model in panel C by setting $P_{11} = a$ from the regression and setting $P_{21} = a + (b * 100\%)$, or simply $a + b$. The crucial point is that in this model, the partials are assumed to be the same in every county in the state, except for random variation. As is often the case, the SER estimates show more racial polarization than the NLN or BLN estimates. With the same two-party data from North Carolina, the SER model estimates that 47 percent of the whites and 60 percent of the African Americans favored the Republicans.

Panel B is the most interesting in table 3. To calculate the partials for the BLN model, first run the same bivariate linear regression as for panel C, but then apply it separately to each county. In the SER model, we obtain cell P_{21} by adding a to $(b * 100\%)$ for every county. For the 1896 gubernatorial election in North Carolina, the parameters are $.47 + (.13 * 1) = .60$ for every county. That is, 60 percent of the blacks are estimated to have voted Republican. In the BLN model for our hypothetical county that is 27 percent black, we obtain cell P_{21} by starting with the same base, .47, but then we add the product of .13 and .27, which is about .04. Thus, for this county, we estimate that 51 percent of the blacks voted Republican; because we assume that blacks and whites voted the same way in each county, 51 percent of the whites must have voted Republican as well. Suppose another county is 57 percent black. Then $.47 + (.57 * .13) =$ 54 percent of the members of each race would be expected to vote Republican. We then calculate statewide totals by multiplying through, county by county, by the numbers of whites and blacks, adding up the totals for all the counties, and then dividing by the total numbers in each state, just as in the NLN model. For our 1896 example, BLN estimates that for the whole state, 52.2 percent of the blacks and 50.8 percent of the whites backed the party of Lincoln.

The troublesome property of the BLN model—and the reason that it was proposed as a *reductio ad absurdum* for aggregate estimates of individual behavior—is that with purely aggregate data and only weak theory, one cannot reject it in favor of any other bivariate model, such as SER. In fact, its authors show that when SER estimates are obtained by weighting each county by the relevant population, SER predicts the same number of black votes as BLN does for each candidate in the state as a whole. The predictions of white votes for each candidate are also equal in both models.[10] Furthermore, if exit-poll survey data are taken to be unequivocally true,[11] then in certain instances the BLN model's predictions are closer to this standard than are those of the SER (Freedman et al. 1991a).[12]

Conceptual Foundations of Ecological Inference

What is to be done? Should historians and political scientists who want to study the presurvey era or elections or other events for which there are no surveys abjure statistics altogether and fall back on the shreds of evidence from newspapers and collections of private letters, whose representativeness cannot be assessed? Are courts best advised to give up the search for racially polarized voting, thus condemning blacks and Latinos to await a permanent liberalization of Anglo racial opinion, the arrival of a thoroughly colorblind society, before their preferences can be fully represented in office? Are we condemned to follow the dictates of the priest-statisticians who wish to prohibit others from predicting and to not engage in it themselves (Freedman et al. 1999, 357)?

We can best answer such questions and determine our proper course by first deepening our understanding of the conceptual foundations of ecological inference.[13] An outcome such as the percentage of votes that the Republican candidate for governor of North Carolina, Daniel L. Russell, received in 1896 may be viewed abstractly as the sum of three types of factors:

$$\text{Outcome} = \text{individual factors} + \text{group factors} + \text{interactions of individual and group factors.} \quad (14)$$

For convenience, substitute capital letters for phrases that represent variables, and small letters for parameters to be estimated:

$$Y = aI + bG + cIG. \quad (15)$$

In any particular instance, we may be able to measure some variables at the individual level, some at the group level, or some at both. Thus, observers of very recent events may have access both to relevant survey data and to information about the social environments in which each survey participant lives, such as data about individual political views and the overall party registration of the respondent's precinct.

But the data will almost never be complete. That is, we will be able to measure every influence on the voters' decisions only in the rarest instances. Consider equation (16), where we add i subscripts to make clear that we are predicting individual voters' decisions partly from individual voters' characteristics, we add j subscripts to indicate that there may be several individual and several group variables, and we append an error term to show that our prediction is not expected to be perfect.

$$Y_i = a_j I_{ij} + b_j G_{ij} + c_j I_{ij} G_{ij} + e_i. \quad (16)$$

This equation reveals the *fundamental condition for avoiding "aggregation bias" in ecological inference*: Do not exclude an independent variable that is correlated with both the dependent variable and the independent variable or variables in which we are most interested. If you do, then the coefficients of the variables of interest will be biased. That is, they will not be the same as they would be in the more complete equation. More specifically, in the case of ecological inference, *aggregation bias* occurs when the coefficients based on aggregate data differ from those we would find if we had access to individual-level data.[14] Conceptually, we are taking one or more of the I or G variables, throwing it (them) into the error term, and attributing its (their) influence incorrectly to one of the variables that we kept in the equation.

There are two ideal types of cases of such a mistake. First, if we omit all the individual-level variables, we run the risk of committing the "contextual fallacy" of assigning to a group-level variable (e.g., residence in a suburb, central city, or rural area) influence that should really be attributed to an individual-level variable (e.g., a person's party identification). Second, if we leave out all the group-level variables—or interpret variables measured at the county level as individual-level variables—then we are in danger of committing the ecological fallacy of assigning to an individual-level variable significance that really should be attributed to a group-level variable. For instance, much of the discussion of racial inequality in scores on IQ tests administered during World War I ignored stark regional differences and made no attempt to measure the quality of schooling each individual received (Gould 1981, 219–20).

For the purposes of this article, the ecological fallacy is much more important than the contextual fallacy, so let us concentrate on it. Equation (17) subdivides equation (16) into what is known and unknown in our typical case:

$$Y_i = a_{1...q} I_{1...q} + b_{1...m} G_{1...m} + c_{1...q} I_{1...q} G_{1...q} + b_{n...z} G_{n...z} + c_{r...t} I_{r...t} G_{r...t} + e_i. \quad (17)$$

where the first three sets of terms are unknown and the last two sets of terms are known.

Equation (17) signifies that we do not measure any variables at the individual level, nor do we measure all the possibly relevant variables at the group level or all the conceivable interactions between group and individual variables.[15] We do know the values of certain variables that are measured at the group level, which we want to interpret either as individual-level variables or as interactions between group-level and individual-level variables.

Equation (17) allows us to compare the NLN, BLN, and SER models succinctly:

NLN assumes the following:
$a_{1...q} = c_{1...q} = 0$;
$b_{n...z}$ and $c_{r...t}$ are not investigated.
BLN assumes: $a_{1...q} = b_{1...m} = 0$;
$c_{1...q}$ = a constant;
G_1 varies from county to county;
$b_{n...z}$ and $c_{r...t}$ are not investigated.
SER assumes: $a_{1...q} = b_{n...z}$;
$b_{1...m} = c_{1...q} = c_{r...t} = 0$. (18)

The formulations in equations (17) and (18) suggest three conclusions and a strategy for approaching the ecological fallacy: First, in any specific case, we are interested in a particular dependent variable and a particular set of independent variables. A particular arrangement of state, county, township, or precinct lines may produce biased or imprecise estimates of the parameters of one particular dependent variable's relation with one particular set of independent variables using one particular estimating equation, but a change in variables, models, or units of aggregation may alleviate or eliminate the bias or imprecision (or make them worse).[16] Thus, it is completely irrelevant to the question of which model better gauges racially polarized voting in Los

Angeles County that the NLN model can predict precinct-to-precinct variations in education, income, or home ownership more accurately than the SER model can. Nor are elections that no one would expect to split Latinos and Anglos good guides to the degree of ethnic division in elections that directly raise obvious ethnic issues. It is similarly immaterial that, on the basis of state-level election and ethnic data only, SER predictions of Latino voting for president match national exit-poll results less well than those of the NLN model, because the correlates of state boundaries are almost certainly completely different from those of precinct boundaries.[17]

Second, all three models make strong assumptions. It is a bit curious to "prove" that there is no racially polarized voting by making such an assumption and then examining no data that might possibly undermine the assumption. Even though the SER constancy assumption is just as strong as assuming some parameters to be 0, SER does allow tests that throw at least some light on the reasonableness of that assumption, as we shall later see in more detail. Third, by contrast, the failure of the NLN and BLN models to allow for the investigation of relationships between the dependent variable and other potentially measurable variables or functions besides the simplest ones seems a distinct failing. This forcible omission of exogenous variables reflects the wholly negative purpose of Freedman et al. in proposing these models. They did not want a more reasonable model in a particular instance. As they freely admitted, they simply wanted to rule out all aggregate models.[18] Finally, these criticisms suggest that relaxing some of the assumptions for certain variables may be a promising way to choose between the three (and other) models and to obtain better estimates of the coefficients of particular interest.

"Groups," Aggregations, and Social Theories

Although it is a convention in the literature, it is doubly misleading to refer to certain variables as group factors. First, the word *group* may seem to connote meaningful social groupings: families, fraternal groups, occupational groups, classes in school, even ethnic groups. By contrast, county, township, ward, and sometimes precinct boundaries may have been set long ago, and even if these lines were once politically relevant, their social, economic, and political correlates may have shifted markedly. In other words, the aggregations for which political data are available may have little or nothing to do with other social demarcations, and it is those correlations that determine, as a matter of statistical theory, whether it is valid to make inferences about individual behavior. Second, whether in a particular instance it is statistically correct to draw conclusions about individuals from aggregate data, interpreting a parameter value as a group, individual, or interactive effect is still a matter of common sense or substantive, nonstatistical theory. Thus, although Freedman and his coauthors did not distinguish between group and interactive effects, I interpreted their view to be that people took on the political cast of their surroundings or, anticipating it, migrated to live among like-minded people, which are both effects of the interactions between individuals and larger groups.

Certainly, there are some political variables that operate at the level of political boundaries. For example, a political machine bent on fraud may so dominate a county or precinct that individual "votes" depend almost entirely on where the individual lives. Indeed, Lawrence N. Powell (1989) and Dale Baum (1991) have used SER to pinpoint and estimate election fraud by searching scattergrams or plots of the residuals in regression equations for strikingly deviant cases, thus using aggregate data to separate aggregate from individual effects.[19] Another example of a truly aggregate variable is the case in which winning a precinct, county, district, or state has some special significance, such as in winner-take-all elections or when rewards are given to political workers on the basis of how hefty the majorities are in their bailiwicks. Thus, a political party may starve some states of campaign resources to carry majorities in another set of states, or ward bosses may vary widely in competence. In these cases, the units of aggregation themselves may be crucial to the pattern of outcomes.[20] In societies where most political, economic, and social boundaries are coterminous, there may also be genuine contextual effects.

But in many other instances, it makes no sense to treat political variables as the result of group effects. Some scholars have attempted to finesse the ecological fallacy verbally by speaking of the voting patterns of "German counties" or "Dutch units," and so on, which are defined as political aggregations that have some arbitrary percentage of members of the relevant population, such as 10 percent German-born[21] (Formisano 1971; Kleppner 1979). Other people in such areas are assumed to be disproportionately American-born of German or part-German stock, or whatever the characteristic. But this assumption is merely a transparent circumlocution. Counties or townships don't vote, and the scholars in these cases are clearly thinking of individual behavior, not that of aggregate units. Furthermore, by placing all the "Dutch units" and so on into the same category and analyzing their voting behavior by comparing their mean percentage for some political party with the party's mean in other types of counties, the analysts are discarding information and preventing themselves from using the more powerful and precise ER technique. In effect, these historians are treating each "type" of aggregate unit as a *dummy variable* (i.e., a variable that can only take on the values 0 and 1) in a multiple ecological regression but leaving out all the other potential economic or social influences on voting.

As we shall later see in more detail, it is often possible in ER to test for interactions between group and individual effects by inserting dummy variables for regions or threshold effects of certain variables or by transforming the values of variables—for example, through logarithmic or logit functions or by including multiplicative interaction terms or powers of variables higher than 1. These analytical moves are probably not so much driven by substantive theory as they are by observing empirical patterns; although they may help one to sort out some individual from aggregate effects, they ought not to be adopted without some clear substantive rationale.

Indeed, every interpretive choice in ecological analysis between individual, group, and interactive effects implicitly or explicitly implies a theory of social or political influence. One of the most obvious deficiencies of the neighborhood models is that they contain no such theory, refer to none, and seem to adopt a contradictory position. NLN is wide open: it just assumes that political behavior does not vary because of an individual's ethnicity but does not specify what factors lead to conformity within a precinct. BLN assumes that the ethnicity of a precinct determines the votes of all the citizens in the precinct, but it does not explain why that should be or how ethnic factors can be so determinative on individuals within precincts but have no effect on individuals in general. After all, many or most of the social and political influences on a person's attitudes and behavior originate outside her precinct or county, and this was true in the nineteenth century as well as the twentieth. By contrast, SER assumes, for instance, that African Americans or Latinos—having been subjected to discrimination as a group, sharing many economic, social, and cultural traits with other members of their groups, and often attacked or appealed to on the basis of their ethnic identities—will react similarly in political contests in which ethnic issues are implicitly or explicitly raised. The theory may be wrong in particular instances, but it is partially testable and at least logically coherent.

Testing Ecological Regression Models

Besides looking at patterns across units of aggregation in particular elections, another way of testing the assumptions underlying ER is to consider a sequence of elections. Suppose that two elections take place at different times in the same counties and involve different candidates and issues. Then, if the correlations between the election results and the independent variables of interest, such as race, across the counties are very different from election to election, the group or interactive effects of this pattern of aggregation must, at the least, not be constant. This fact in itself may lend support to one model rather than another. If, on the other hand, regression analyses of sequences of elections turn up very similar patterns, the implication may be that aggregation bias is likely to be small. Consider the case of *transition matrices*—multicelled tables charting how voters behaved in two elections—which is the other principal example of the use of ER besides the evaluation of ethnic voting (Achen and Shively 1995).

Table 4 contains estimates based on unweighted OLS regressions of the way in which voters and nonvoters in the 1892 North Carolina gubernatorial contest behaved in 1896. The most notable characteristic of the table for present purposes is the high degree of party loyalty among Democrats and Republicans. Whatever the social, economic, or other traits that caused people to vote one way in 1892, and whatever the correlations of those traits with county boundaries, men voted overwhelmingly the same way in 1896. Thus, estimates of entries in transition matrices that indicate solid party or candidate loyalty are unlikely to be very much affected by aggregation bias because, as explained above, that bias is present only when there is a relation between some excluded variable and the dependent variable, controlling for the included variables of interest. Here, the 1892 voting percentages, the included variables of interest, clearly absorb the vast amount of the effects of other variables on the dependent variables, the 1896 voting percentages. The obvious substantive theory, that like causes produce like effects, and that people act consistently to minimize the cost of information and social dislocation, likewise supports an individual rather than a group interpretation of the high coefficients. It is no wonder, then, that ER has been widely used to estimate transition matrices: the practice is often justified on grounds of both statistical and substantive theory.

TABLE 4
Transition Matrix for the North Carolina Governor's Races, 1892, 1896

	1896 election				
1892 election	Democratic	Republican	Populist	No vote	(County means)
Democratic	75.7	13.9	0.4	10.1	(38.7)
Republican	18.7	87.0	–8.2	2.3	(27.9)
Populist	9.5	21.0	51.3	18.8	(12.1)
No vote	12.5	37.5	16.4	33.6	(20.7)
(County means)	(38.3)	(40.2)	(7.6)	(13.9)	

Note: First four columns add to 100% across rows. Data were calculated from unweighted statewide ER.

Large defection rates and/or logically impossible entries, as in the column for the 1896 Populist percentage, suggest that the prediction equation needs to be modified or variables need to be added. But which of the many potential independent variables or other functional forms should be used, how do we gauge whether they improve the estimates, and how do we properly calculate the estimates of transition matrices or ethnic voting for more complex equations?

Most proposed solutions to theoretical or practical deficiencies in ER may be viewed as attempts to reduce or control for heterogeneity within geographical units or to increase the predictability of the relationships measured between these units. More advanced methods begin with criticisms of the SER model given in equation (12), and some obvious improvements have already been introduced to produce table 4. First, although the vast majority of the population in North Carolina was either white or black, only adult males could vote, and the age and gender structure of the two racial populations might have been different. To allow for that possibility directly, the racial or partisan percentages ought to be those of adult males, not of the total population.[22] Second, not everyone voted, although turnout was quite high—86.1 percent of the adult males—by more recent standards. Different proportions of whites and blacks might have participated in the election, and turnout might have varied from county to county in nonrandom ways. The denominator of the party percentage variables, therefore, ought to be the number of adult males, not of the total vote. Third, there was a serious minor party, the Populists, that contested the election, receiving 7 percent of the vote, and attracting quite uneven support, ranging from 0 percent to 30 percent, in counties around the state. Introducing an indication of Populist strength might not only be necessary to calculate a full transition matrix, but it might also improve estimates of the race/political choice relationship. Fourth, not all counties had the same population, and if they varied considerably in the number of adult males, estimates from SER might not, in effect, add up to the statewide totals. An obvious means of attacking this problem is to weight the observations by a factor such as their population.[23]

A fifth observation is suggested by figure 3. All the counties are between 30 percent and 75 percent Republican and between 0 percent and 60 percent black. If we look beyond 60 percent African American on the right of figure 3, the regression line just projects out into empty space. Even if there were no other problems, we might not be justified in claiming too much precision for our estimate of how a 100 percent black county would have voted. It would be comforting to find election data from more homogeneously black areas. Similarly, if we plotted the percentage Populist in the 1892 gubernatorial election against the percentage for that party in 1896, the highest point would correspond to 37 percent in the first election and 30 percent in the second. Our regression line would project out well beyond the data, which should always give one pause. Similar conditions explain why it is difficult to determine correlates of minor, but not heavily geographically concentrated, political parties.[24]

Sixth, there is a good deal of scatter around the regression line in figure 3, and two major patterns immediately appear: If one ignores the counties that are less than about 30 percent black, the relationship between race and party seems much more positive than the regression line calculated for the whole state, and there is more variation around the line in nearly all-white counties than elsewhere in the state. It is possible that another regression equation, including other variables such as region or a variable to take account of the apparently different behavior of counties above and below 30 percent black, might fit the data better, reducing the bias and increasing the *precision* (the amount of confidence we have in the estimates of the slope and intercept). Conceptually, the simplest way to deal with such a problem would be to calculate separate regression estimates for two or more sets of counties (say, those above 30 percent black and those below 30 percent black), weight the two by their respective populations, and add them to get a statewide estimate.

The trouble with running separate regression equations on subsets of counties is that the number of counties in a subset may be small and the estimates consequently less reliable than an estimate based on more counties. Fortunately, it is mathematically equivalent to insert into the regression equation dummy variables for counties with levels of black population above and below 30 percent. If observations suggest that the intercepts are different in two groups of counties, then a dummy variable can be inserted to take that into account, as in equation (19).

$$Y = a_1L + a_2H + b_1X + e, \tag{19}$$

where H (high) is a dummy variable with the value 1 for counties that were at least 30 percent black and 0 for those below 30 percent, L (low) is a dummy variable that takes the value 1 for counties below 30 percent African American and 0 for those above that level, and a_1, a_2, and b_1 are different parameters to be estimated.[25] In equation (19), a_1 is the intercept for the counties below 30 percent African American, a_2 is the intercept for counties above 30 percent black, and b_1 is the slope for the whole state. If information indicates that the slopes, not the intercepts, differ, then we have equation (20):

$$Y = a + b_1LX + b_2HX + e, \tag{20}$$

where b_1 gives the slope for the less heavily black region and b_2 for the more heavily black region.

If both vary, then equation (21) is appropriate.

$$Y = a_1L + a_2H + b_1LX + b_2HX + e. \tag{21}$$

The most obvious way to determine which of equations (12), (19), (20), and (21) best fits the data is to compare the *percentage of variance explained,* a measure of the explanatory power of an independent variable or series of variables, for each equation. This is referred to in the literature as R^2.

To deal with the fact that there is more scatter around one part of the regression line in figure 3—in the overwhelmingly white counties—than around the rest of the line, one might try conventional methods that statisticians have devised for the unequal scatter or heteroscedasticity problem. One standard procedure is to weight the relevant independent or right-hand-side variable (one on the right-hand side of the equals sign in the regression equation) by 1 divided by the square root of the population, as explained above.

Table 5 compares estimates of voting by race in that contest calculated with equations (12) and (21). There are considerable and substantively important differences in the estimates for Republicans and Populists and less stark differences for the Democrats. According to the SER equation, blacks were only a bit more likely to vote Republican than were whites, and they were three times as likely as whites to cast a vote for the Populists. According to the equation with dummy variables, on the other hand, the probability that an African American would vote Republican was twice that of a white, whereas the likelihood that he would vote Populist was less than half that of a white.[26]

TABLE 5
Racial Polarization in the North Carolina Governor's Race, 1896: A Comparison of Two Sets of Estimates

	Party			
Race	Democratic	Republican	Populist	No vote
A. Whole state				
White (%)	42.3	38.7	4.6	14.4
Black (%)	27.8	44.1	15.5	12.9
B. State divided at 30% black and then recombined				
White (%)	44.4	30.0	10.1	15.6
Black (%)	23.1	61.7	4.0	11.2

Note: Each row adds to 100%, except for rounding error.

Table 6 shows that the panel B, equation (21), estimates are much better than those of panel A, equation (13). On all four rows, the percentages of variance explained (R^2) are higher, and in three of four instances they are much higher. Most of the coefficients in panel B are much further than two standard errors away from the analogous entries in panel A. For instance, the coefficient for the percentage black term in the Republican equation in panel A is .054; those for the percentage black terms in the two sets of counties were much more than two standard errors distant. The difference between –.558 and .054 is more than four times the standard error for the parameter in the more complex equation (.13). Although they are not given here for reasons of space, estimates derived from equations that split the state at 25 percent or at 35 percent black are quite similar to those in tables 5 and 6.

What model of voting behavior might explain why individuals in the two contexts (counties above and below 30 percent black) voted as they did? Although southern whites in areas that contained only a relatively few African Americans may have accepted racist shibboleths, it has been repeatedly noted that they were not so transfixed by racial matters as were denizens of the Black Belt—see, for example, Key (1949); Matthews and Prothro (1966); Kousser (1974). Politically, whites in overwhelmingly white areas did not have to face the possibility of dealing with black elected officials, freeing them to a much greater degree than their Black Belt counterparts to vote on the basis of other issues and interests. As the black proportion rose past some

TABLE 6
Testing the Whole- and Divided-State Models

A. Whole-state statistics ($Y = a + bX + e$)

Dependent variable	Intercepts a Coeff.	a *SE*	% black b Coeff.	b *SE*	R^2
Democratic	.423	.012	–.146	.040	.135
Republican	.387	.020	.054	.064	.008
Populist	.046	.013	.109	.041	.074
No vote	.144	.012	–.015	.036	.002

B. State divided at 30% black ($Y = a_1L + a_2H + b_1LX + b_2HX + e$)

Dependent variable	a_1 Coeff.	a_1 *SE*	a_2 Coeff.	a_2 *SE*	b_1 Coeff.	b_1 *SE*	b_2 Coeff.	b_2 *SE*	R^2
Democratic	.433	.017	.459	.063	–.231	.098	–.217	.142	.148
Republican	.474	.023	.045	.084	–.558	.130	.854	.189	.333
Populist	–.002	.016	.253	.058	.441	.090	–.372	.130	.303
No vote	.096	.014	.245	.050	.347	.078	–.266	.114	.255

threshold, so might white fears, which would cause whites to vote increasingly against the political party that was most associated with blacks. At a higher threshold of the black percentage, whites might decide that blacks were likely to win anyway, and they might therefore cut whatever deal they could with the blacks, which might include supporting some of the same political candidates. Such a model would explain the phenomenon of "fusion" arrangements between whites and blacks in the Black Belt in the post-Reconstruction South (Wharton 1947). This theory is quite consistent with the information presented in table 6, which shows that although support for the Democratic party (known in North Carolina at the time as "the party of white supremacy") was roughly constant across the two contexts, that for the Republicans and Populists was not, and neither was overall turnout. Republican support diminished dramatically as one moved from counties that contained almost no blacks to those in which the black proportion was a quarter or a third, rising even more dramatically as the proportion increased further. Populist strength reversed the Republican pattern, though less overwhelmingly, whereas turnout (the opposite of not voting) was highest in the areas with the fewest or the most blacks.[27]

The improvement in the estimates of racial polarization obtained by dividing the state suggest that analogous equations for the party transition matrices ought to be examined. Equation (22), which relates the Democratic percentage in 1896 to the party percentages in 1892, is directly parallel to equation (21), but it contains more variables:

$$D_{96} = a_1L + a_2H + b_1LD_{92} + b_2HD_{92} + b_3LR_{92} + b_4HR_{92} + b_5LP_{92} + b_6HP_{92} + e, \qquad (22)$$

where the dummy variables and *a*s and *b*s are as in equation (21) and the variables with the subscripts 92 and 96 are the Democratic, Republican, and Populist percentages for governor in 1892 and 1896. Equations for the Republicans, Populists, and nonvoters in 1896 substitute those dependent variables for the D_{96}. To compute statewide estimates from this equation, multiply a_1 by the number of the state's nonvoters in 1892 who resided in counties below 30 percent black, and a_2 by the number in counties above 30 percent black, then divide by the total number of nonvoters. Then multiply b_1 by the number of Democrats in 1892 in counties below 30 percent black and b_2 by the number of Democrats in 1892 in counties above 30 percent black, divide by the total number of Democrats, and add the statewide estimate of nonvoters. Do the same for the Republicans and Populists as for the Democrats, and we have the first column of panel A of table 7. The other columns may be computed analogously. Compared with table 4, table 7 implies that more of the support for the Republican gubernatorial candidate in 1896 switched from the Democrats in 1892 and less came from 1892 Populists and nonvoters. The estimates outside the 0–100 percent logical bounds are smaller, but there are two of them, compared with one in table 4.

TABLE 7
Transition Matrix for Governor's Races in North Carolina, 1892, 1896

	1896 election			
1892 election	Democratic	Republican	Populist	No vote
A. Without wealth controlled				
Democratic	71.8	26.4	1.5	0.2
Republican	21.3	88.3	−5.8	−4.1
Populist	16.5	10.5	51.2	22.1
No vote	12.9	28.6	13.8	44.8
B. With wealth controlled				
Democratic	77.5	15.2	2.3	5.3
Republican	16.9	97.6	−7.1	−7.2
Populist	9.4	32.6	46.4	12.1
No vote	14.8	26.5	15.3	43.9
C. Percentages of variance explained[a]				
Table 4	.491	.468	.659	.205
Panel A (above)	.556	.657	.693	.418
Panel B (above)	.609	.717	.707	.426

Note: Data were calculated by splitting the state into counties above and below 30% black.
[a]Corrected for degrees of freedom.

Table 7 also shows estimates derived from equations just like (22), but containing two additional variables—white wealth per white male adult for counties above and below 30 percent black.[28] The entries in panel B were computed by the same method as with equation (22), except that to each cell entry was added the appropriate coefficient for the white wealth variable for the counties under 30 percent black times the mean white wealth per white male adult in that set of counties, plus the analogous figure for the counties over 30 percent black.[29] Panel B shows a much more marked shift from the 1892 Populists to the 1896 Republicans than panel A indicates, with corresponding declines in crossover voting in the two major parties. Panel C demonstrates that the additional variables explain increasing percentages of the variation, even allowing for their additional number.

The rationale for including wealth in the equation is that the Republicans and Populists offered much more active governmental programs than the Democrats did, especially expanding education for the masses, black and white, whereas the Democrats were the traditional party of the planter and Piedmont elites. Contrary to that rationale, however, the equations containing wealth (not given here) show a positive and statistically significant coefficient between Republicanism and white wealth in the counties over 30 percent black, without a significant coefficient in counties below that level. These coefficients suggest that rather than representing an effect of a white class struggle, wealth proxies something else in this instance, and race is the obvious

TABLE 8
Race Set the Context for Voting Transitions in North Carolina between 1892 and 1896

1892 election	1896 election							
	Democratic		Republican		Populist		No vote	
Party/% black	Coeff.	*SE*	Coeff.	*SE*	Coeff.	*SE*	Coeff.	*SE*
No vote < 30	.137	.094	.136	.104	.074	.082	.655	.088
No vote > 30	.301	.081	–.111	.090	.341	.071	.473	.076
Democratic < 30	.763	.126	–.046	.139	–.058	.109	–.657	.118
Democratic > 30	.363	.093	.076	.103	–.212	.081	–.231	.087
Republican < 30	–.052	.120	.930	.133	–.156	.105	–.729	.113
Republican > 30	.159	.078	.327	.086	–.274	.068	–.216	.073
Populist < 30	.051	.148	–.027	.164	.586	.129	–.601	.139
Populist > 30	–.027	.111	.185	.123	.087	.097	–.244	.104
Black < 30	–.239	.081	–.023	.089	.033	.070	.218	.076
Black > 30	–.244	.112	.894	.124	–.268	.097	–.382	.105
Corrected R^2	.612		.789		.714		.532	

Note: Coefficients are for counties above and below 30% African American.

candidate. In the state as a whole in 1896, the correlation between white wealth per white male adult and the percentage of male adults who were black was .686; it was .641 and .497 in the counties above and below 30 percent black, respectively. If the percentage black, separated into the same two groups, is substituted for the white wealth variables, the percentages of variance explained are even higher, and the signs of the coefficients accord better with expectations.[30]

Instead of multiplying out and adding up to get statewide estimates, I have presented all parameters of the relevant equations in table 8 to emphasize the differences in the patterns in the two sets of counties. According to this model, party preferences were rather stable in the whiter counties, but there was considerable shifting in the counties that contained more blacks. In the more heavily African American counties, 1892 Populists, white and black, shifted to the Republicans in 1896, and blacks who had not voted in 1892 turned out in large numbers in 1896 and overwhelmingly supported the GOP. The table suggests that although it may be interesting descriptively to tally the numbers divorced from their context, it may be useful to disaggregate the totals when constructing an explanation of why some voters shifted and others stayed loyal. When contextual variables matter and can be measured, the usual way in which ER estimates are presented is often not the most informative.

But the differences between estimates derived from the various ER models should not be exaggerated. There are different ways of presenting such estimates. In tables 4, 5, and 7, they are arranged so that each row adds to 100 percent. If the row and column marginal totals are approximately equal to each other—in the current example, if each party gets roughly the same percentage of the vote in each election—this form of table, which directs the reader's attention equally to each cell, is perfectly appropriate. But

TABLE 9
Range of Transition Matrix Estimates from Tables 4 and 7

1892 election	1896 election			
	Democratic	Republican	Populist	No vote
Democratic	27.8 to 30.0	5.4 to 10.2	0.2 to 0.9	0.1 to 3.9
Republican	4.7 to 5.9	24.3 to 27.2	–2.2 to –1.6	–2.0 to +0.6
Populist	1.1 to 2.0	1.3 to 3.9	5.6 to 6.2	1.5 to 2.7
No vote	2.6 to 3.1	5.5 to 7.8	2.9 to 3.4	7.0 to 9.3

Note: Data were normalized to make cells add to 100% across the whole table, instead of across each row.

in cases, as here, where the Populist percentages were so much lower than those for the Democrats and Republicans, the table may encourage the analyst to overemphasize differences that reflect the behavior of a very small number of people. Table 9 shows the range of estimates in tables 4 and 7 normalized to add to 100 percent across the whole table, rather than across each row. The estimates are easily calculated. To obtain the entry in the bottom right-hand corner of the table, for instance, observe that the smallest of the three entries in table 4 and panels A and B of table 7 is that from table 4—33.6 percent. Multiply 33.6 by 20.7, which is the percentage of people in the average county not voting in 1892, to obtain 7.0 percent of the whole population as the lower-bound estimate. The upper-bound estimate is 44.8, from panel A, table 7, multiplied by 20.7, or 9.3 percent. Other entries are calculated analogously. The obvious conclusion is that the estimates of individual voting behavior from the SER and two separate multiple ecological regression (MER) models are not very different. The ranges in only two of the sixteen cells exceed 3 percent of the population, and the average range in each cell is only 1.9 percent.

Overall, tables 4–9 and the accompanying discussion demonstrate the following: (1) the crucial constancy assumption of SER can be tested—unlike most of the assumptions of NLN, BLN, or Shively's variant of the method of bounds; (2) in some cases, the differences between the SER and MER models can be substantively as well as statistically significant; (3) even when there are pronounced and interesting contextual effects, the resulting estimates of individual behavior may not vary much between models; and (4) the choice between various SER and MER equations is a matter of substantive theory as well as of empirical data. An implausible or nonsensical theory should be rejected, no matter how good the apparent fit of the data.

This last observation is key to understanding the fundamental flaw in the logic of the attempt made by expert witnesses on behalf of defendants in voting rights cases to use MER to undermine estimates of racially polarized voting based on SER.[31] Before the neighborhood model strategy, there was the multiple regression strategy. Political scientists Charles Bullock (1991), Harold Stanley (1987), and others contended that it was not enough to show through SER that elections were in fact racially polarized, but that plaintiffs had to demonstrate that whites voted against blacks purely out of racial animosity, and that the way to separate out that racial animus was to control for all other possible influences on voting behavior. For instance, black candidates are generally Democrats, whereas southern white candidates are increasingly Republicans, so whites might have opposed blacks because of their party affiliations; black candidates in the early 1980s often had to face white incumbents, so white voters might have merely favored the better-known incumbent; African American candidates often have fewer financial resources available to fund their campaigns, so perhaps whites might not have received their messages. According to this view, one would therefore obtain the "true" racial effect by regressing the vote for each candidate against the percentage of voters in each county who registered Democratic, a dummy variable for incumbency, a measure of campaign spending, and perhaps other variables, as well as the percentage black. If the coefficient for the percentage black was statistically insignificant, voting was not racially polarized.[32] If all the measurable correlates of racial differences still did not reduce the coefficient of racial percentages to insignificance, the experts would try to convince the judge that unfortunately unmeasured variables such as "candidate quality" would account in principle for the apparent remaining effect.[33]

The best way to expose the error in this correlation strategem is to consider a ludicrous example. Most whites in America have straighter hair than most blacks. Suppose someone proposed to introduce the "curliness" of each candidate's hair as an independent variable, arguing that whites might have a preference for elected officials who had straight hair.[34] Why would this suggestion be ridiculed? At least for recent elections, the data could be gathered, the model would no doubt fit quite well, and such a highly multicollinear variable would unquestionably reduce the size of the coefficient of the racial percentage variable. There are two grounds for rejecting the proposal: First, the behavioral theory is nonsensical—hair texture is surely irrelevant to voter choice. Second, it is impossible to distinguish methodologically between a model in which all the influence is attributed to race and one in which all the influence is attributed to hair texture. Likewise, in the Bullock/Stanley models, it is impossible to determine on either substantive or methodological grounds whether whites vote against blacks because they are Democrats or whites are Republicans because of a white backlash against blacks; whether whites vote for white incumbents because they are incumbents, or the candidates got to be incumbents because whites voted against or did not consider black or potential black candidates because of race; whether whites vote for candidates who raise the most money, or white candidates have the most money because they do not suffer from a legacy of slavery and discrimination.[35] As a consequence, even if adding such variables as controls in MER improved the statistical fit, it would be incorrect to determine racial polarization by looking only at the signs of coefficients for racial variables in such tables as table 8. The proper effects could be gauged only by asking the question of how an average white and an average black would have voted for an average white or average black candidate—effects that require adding to the direct racial impact the indirect effects of the correlations between race and party, incumbency, campaign resources, and so on. The simplest way to do that is to follow the procedure used in panel B of table 7: to each cell entry, add the product of the appropriate value of the control variable and its coefficient. If one controlled for party in estimating racial polarization, for example, one would have to add in the coefficient for the Democrats times the proportion of blacks who are Democrats and, if non–major-party voters are included in the registration measure, the coefficient for the Republicans times the proportion of blacks who are Republicans, and likewise for the whites and for other variables. The result would be a descriptive measure of how members of each race voted, it would disregard as unanswerable with either aggregate or survey data the question of whether they voted the way they did because of racial animus or for some other reason, and, as table 9 suggests, it would probably not differ very much from the answer obtained much more easily with SER. Justice William Brennan reached essentially the same conclusion in *Thornburg v. Gingles* (1986).

Logically Impossible Estimates and Logical Bounds

An often-repeated criticism of SER and especially of MER is that they sometimes produce logically impossible estimates. How can you trust a method that tells you that, say, less than 0 percent of blacks voted Democratic during Reconstruction or that more than 100 percent of those who voted for one small party in one election supported it in the next? Statistical fixes for this embarrassing problem have involved three methods of ensuring that estimates stay within bounds: quadratic programming (Irwin and Meeter 1969), logit (Kousser 1980), and King's (1997) EI. The first two use the same aggregate data as SER or MER do, but fit different functional forms of equations that do not allow estimates to fall below 0 or above 100 percent. The application of quadratic programming to solve this problem has attracted almost no attention. Logit is nonlinear and produces estimates that may reflect only the portions of the data near the logical bounds. Only EI incorporates the extra information contained in the bounds for individual counties.

When is the problem of logically impossible estimates most likely to occur, and how serious is it? There are two common cases when ER is likely to generate estimates outside the 0–100 percent boundaries. The first, which should not be considered very serious, takes place when two variables of interest are very highly correlated, but some unmeasured independent variable affects their relationship, forcing the regression line or surface to intersect the relevant axes above 100 percent or below 0 percent, as in figure 5.

Generally, such estimates miss the bounds by a few percentage points, the confounding variables are difficult to specify, and one appropriate response is to set the estimates at their boundary points and adjust all the other estimates accordingly. For instance, in a bivariate relationship, if the regression estimate is that 105 percent of the African Americans voted Republican, set that at 100 percent. Then all the Democratic votes and the excess of Republican votes after subtracting all the blacks who turned out must have come from whites. Such difficulties are especially likely to occur when the amount of variation in the independent variable or variables is limited, such as in North Carolina in 1896, when the county with the highest percentage of blacks was only 60 percent black. It is very difficult to wring any more information out of such data, and setting the estimates at their logical limits is just a humble recognition of our ignorance.

The other instance arises when a group, such as a political party or ethnic group, is small but not very geographically concentrated. Then an AU plot would show, as noted before, that one must project the regression line far past the point where one has any observations. But in the case of small groups, a bounds plot will show that the county bounds do not contain much constraining information either. Consider figure 6, which is based on the most Populist county in North Carolina in 1892 and 1896, one in which the Populist gubernatorial candidate received 37 percent in the first election and 30 percent in the second. If in each election, we divide the voters into two groups—those who voted Populist and those who either voted for other parties or did not vote at all—then the entire Populist vote in 1896 (30 percent) might have come from those who had voted Populist in 1892 (37 percent). The maximum percentage of Populist repeaters was therefore 30/37 or 81 percent, so the bounds line crosses the vertical axis in figure 6 at 81 percent. On the other hand, the Populists in 1896 need not have received any votes at all from the 1892 Populist voters, because there were more than enough in the 1892 electorate who had not voted Populist (63 percent) to fill the ranks of 1896 Populists. The bounds line therefore crosses the horizontal axis at 30/63 or 48 percent. This bounds line does not constrain the estimates of stability and change in the two elections by very much, for Populist repeating could

FIGURE 5
AU Plot of a Logically Impossible Estimate

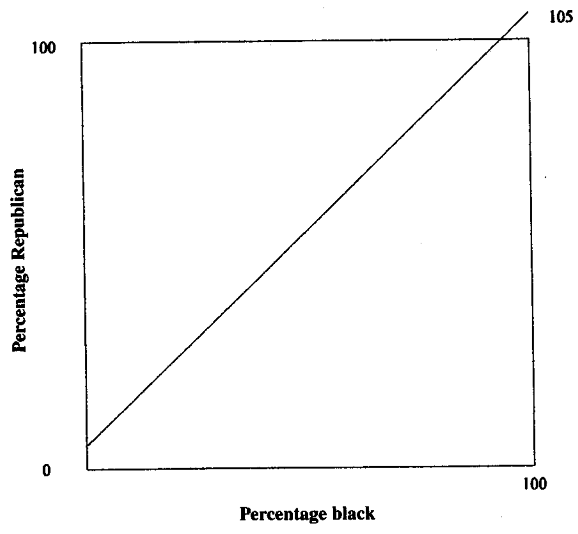
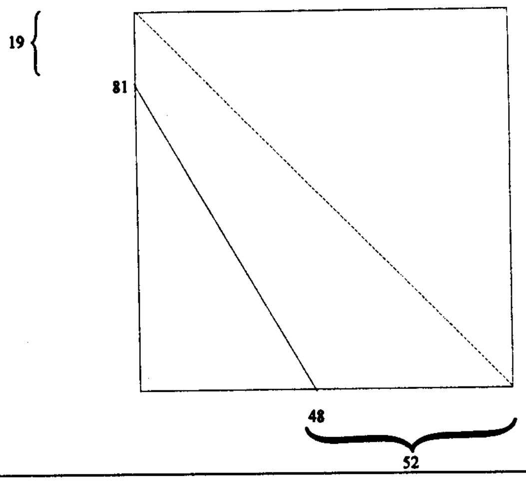


FIGURE 6
Bounds Plot of Most Populous County in North Carolina Governor's Races, 1892, 1896: How Much Do the Bounds Constrain?



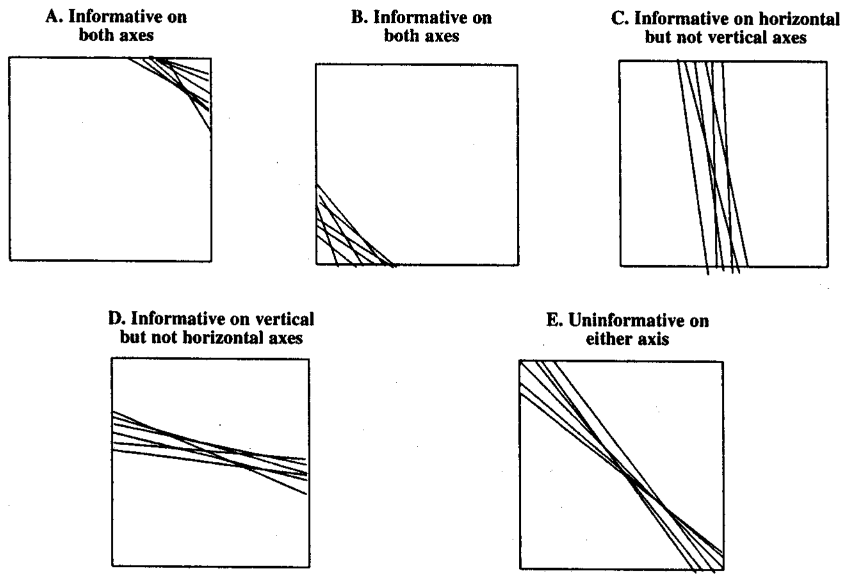
vary from 0 to 81 percent, and other voters' conversions could vary from 0 to 48 percent.

Figure 6 also suggests a simple measure of how much constraint inheres in one or more bounds lines. The notion is illustrated in the figure by the portions of the axes corresponding to the curly brackets. For the Populist repeaters, the bounds allow us to know for certain that at least 19 percent (7/37) of the Populists could not have been repeaters; for the converts, the bounds tell us that at least 52 percent (33/63) could not have been converts. As the example and the tilt of the line demonstrate, we learn more about the converts than about the repeaters from the bounds for this county. In general, if we wanted an index of the amount of information about each coefficient that was available from the bounds for a group of counties, perhaps for all those in a state, we could take an average over all the counties of all the analogous percentages for each coefficient. If, for instance, another county voted 20 percent Populist in 1892 and 15 percent in 1896, then the nonrepeater percentage would be at least 25 percent (1 – 15/20) and the nonconvert percentage would be at least 81 percent (1 – 15/80). On average in the two counties, we learn something about 22 percent of the repeaters [(19 + 25)/2] and about 67 percent of the nonconverts [(52 + 81)/2]. An overall average of all the units would tell us whether the bounds were likely to add much to our ability to make ecological inferences. We might also weight each bound by the population in the county that fell into that category—Populist voters in 1892 and 1896, in this example—and then add them up to determine the overall bounds for the state.[36]

Figure 7, based on hypothetical data, illustrates several general facets of bounds plots, which will be useful in our discussion of King's EI. In general, the shorter the bounds line, the more fruitful the information it contains. The most instructive sorts of lines are those in the northeast or southwest corners of a bounds plot, as in figure 7(a) and (b), for they constrain the estimates of both the coefficients, such as the Populist-repeater and non–Populist-convert estimates of the previous example. Closely spaced vertical or nearly vertical lines, as in figure 7(c), give us very definite information about the value on the horizontal axis, because their values on the horizontal axis fall in a very small range, but very little precise information about the value on the vertical axis, because they span every or nearly every value from 0 to 100 percent on that axis. The informational usefulness of flat or nearly flat bounds lines, as in figure 7(d), is just the reverse—they yield a great deal of information about the coefficient on the vertical axis and very little about the horizontal axis. (All bounds lines have negative or zero slopes—from northwest to southeast.) The least instructive lines are the longest, and the worst are diagonal lines going from the northwest corner to the southeast corner, as in figure 7(e), for they correspond to estimates that could vary from 0 to 100 percent on both scales.

How serious is the problem of SER producing estimates outside the logical boundaries? Whereas King (1997) treats it as extremely common, other experiences differ. In 399 SER or MER regressions that produced the tables based on ecological regressions in *The Shaping of Southern Politics* (Kousser 1974), for example, only 16 (4 percent) of the esti-

FIGURE 7
Information Content of Bounds Plots: Five Ideal-Type Patterns from Hypothetical Data



mates were below 0 or more than 100 percent, and only 3 of those 16 were more than ten percentage points outside the range. We should therefore approach King's vaunted solution with a touch of skepticism. The incidence of the out-of-bounds examples that provide much of the emotional force for his proposed solution may be overstated, and the additional information gained by including the bounds in the estimation procedure may, as figures 6 and 7 suggest, be exaggerated.

EI Made Simple

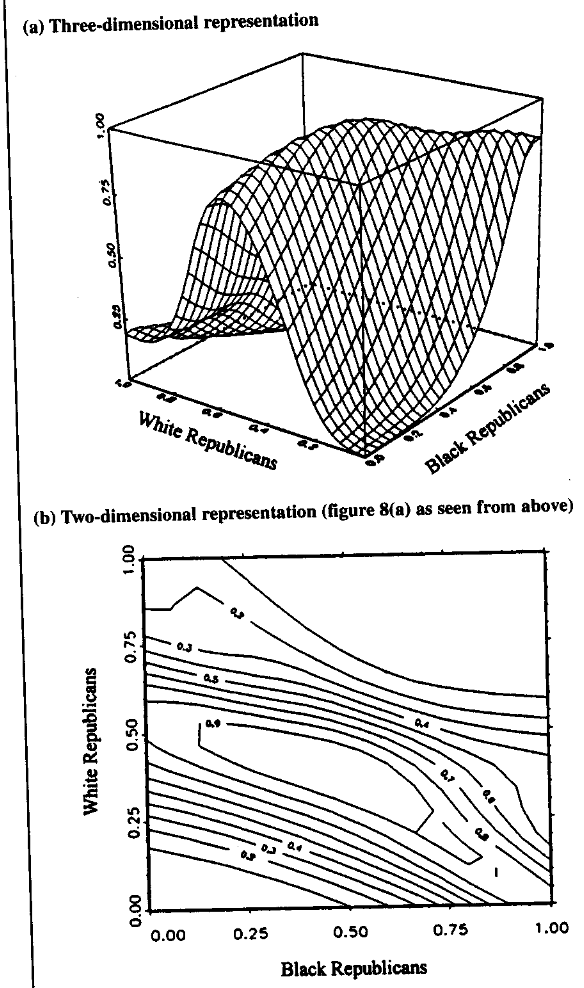
EI is essentially a two-step procedure that takes advantage of present high-speed computing power to combine information about the variation between counties (or precincts or any other level of aggregation) with information about the bounds on each county. In effect, King combines an AU plot and a bounds plot. In the first step, the AU step, King assumes that the relationship between counties is not just a linear regression, but rather that it is drawn from a complete truncated bivariate normal distribution. SER estimates a straight line, so that the proportion of African Americans voting Republican, for instance, is assumed to be roughly the same in every county. As one recalls from elementary geometry, only two points are needed to specify a line in any plane. The SER estimate of these two points and the scatter in each direction around the regression line can be generated by a simple series of arithmetic operations—addition, subtraction, multiplication, and division—in one pass through the data. EI, on the other hand, estimates a whole distribution, a surface that needs five quantities to be estimated to determine it.[37] King assumes that this surface is a *truncated normal distribution,* such as the familiar bell curve with a third dimension added, where all the parts that do not fit over the unit square (the area of a bounds plot) are cut off and disregarded. The five quantities or parameters must be approximated (to any desired degree of exactitude) by computer iterations that determine just what values maximize a complicated formula, which will not be presented here.

Figure 8(a) shows a pseudo-three-dimensional truncated normal distribution hovering over a unit square. Note that the cross-checked curved surface, which resembles a picnic tablecloth, is rather irregular. It has one mode or single highest point, but it does not diminish evenly in every direction from that mode. Note, too, that the walls and floor of the unit cube cut off—truncate—the tablecloth at different points on each side. Not all such graphs are so asymmetric. Figure 8(b) is the projection of the tablecloth onto the unit square, where it appears as contour lines, as on a topographic map. The contour lines are *quantile densities,* indicating the estimated amount of variance in the estimate or, to continue the metaphor, how much of the volume under the tablecloth is close to the mode. The mode is a bit below and to the left of the center of the square, and the lines show how quickly the tablecloth, folded as it is in 8(a), drops away from that mode.

King's step two is to take the distribution with its parameters estimated in step one and apply it to the lines in the bounds plot. For each bounds line, he draws, say, one hundred random samples from the slice of the particular truncated normal distribution estimated in step one that lies just above the line. Each of the samples appears as a single point on the bounds line. Of course, these need not all come from the center of each bounds line, and each overall sample need not look like the conventional bell-shaped curve. It might be skewed in one direction or the other or even partially cut off by a boundary. If the mode of the normal distribution estimated in step one lies over the middle of a particular bounds line, then most points sampled will lie in the center. If the mode lies toward one end of a particular line, then most points sampled on that line will come from that end.

After the samples are drawn for each county, King's program computes the average of all the points along each

FIGURE 8
Truncated Normal Distribution in King's EI



bounds line. Each of these averages corresponds to a value on the horizontal and vertical axes of the bounds plot. In our running example, they would be the percentage of blacks who voted Republican in the 1896 North Carolina governor's race, which King refers to as β_i^B, where the *i* subscript shows that he is referring to an individual county, and the percentage of whites who voted Republican, which he calls β_i^W. The superscripts ^{B} and ^{W}, of course, refer to the race of the voters. To arrive at an overall estimate for the state, he multiplies each β_i^B by the number of blacks in the county, sums that product for all the counties, and divides by the number of blacks in the state. He does the same thing for whites. As King emphasizes, this procedure gives each county's white and black communities the proper weight in the estimate, unlike weighting each county's importance by the total population, as is often advised, because King's procedure separates out the populations by race. It is also superior to the conventional SER procedure of weighting each county the same, regardless of population or the composition of population.

There is another way in which King's county-by-county estimates are superior to SER's. It would be possible to estimate β_i^B and β_i^W by using SER for the first step and then in the second step, determining the point on each bounds line that the regression line would predict. The problem is that the SER prediction would be the same for every bounds line, because SER predicts, for example, that the proportion of blacks who voted Republican (β_i^B in this case) would be the same in every county, except for random differences. It should also be noted that an SER estimate that was above 100 percent or below 0 percent for the whole state would not produce estimates on any bounds line, for by definition all the bounds lines must terminate at 0 and 100 percent. In contrast to SER, EI estimates may fall at different points on each bounds line.[38] King's county estimates are thus preferable to any that could be made with SER because they are at the same time constrained and unconstrained: they are constrained to fall on the bounds lines, but they are unconstrained to fall at the same point on each.

If certain forms of MER models fit the county-by-county variations better than an SER model, and if the particular MER model in question produced estimates within the 0–100 percent limits, however, it could be used in a second stage analogous to King's second stage. If it were, the county estimates would not all be the same, and they could be properly weighted before being added up to produce statewide estimates. This model would match some of the appealing characteristics of King's EI while retaining the ease and flexibility of multiple regression. So far, I know of no one who has attempted to estimate such a model.

EI, ER, and Aggregation Bias

Goodman showed in 1959 that if there is no aggregation bias—no correlation between the independent variable (e.g., the percentage black) and the estimate of the quantities of interest (e.g., the statewide percentage of whites and blacks who voted Republican)—then SER gives accurate estimates of the relationships that we could measure directly if we knew how each individual voted. The problem with all the techniques of ecological inference is that we usually do not know the degree or direction of aggregation bias. Both ER and EI allow one to try to gauge and correct for that bias. We have already seen that ER allows one to examine scatter plots, add independent variables (which may include dummy variables, such as in equations 19–21), and test competing models by comparing their percentages of explained variance. How does EI approach the same problem?

King (1997) begins by making sweeping claims for the adequacy of what will hereinafter be referred to as simple EI (SEI) in the face of aggregation bias. However, he almost immediately qualifies those claims, thus asserting on p. 19 that SEI is "robust to aggregation bias." By p. 21, he has drawn back: EI "is *almost* invariant to the configuration of district lines. If precinct boundaries were redrawn, even in some random fashion, inferences about" individual behavior "would not *drastically* change in *most* cases. . . . A researcher will *almost always* do far better by using [EI] than those previously proposed" methods (my italics). After comparing the results of EI with known individual-level relationships, King becomes even more restrained. On p. 218, for instance, he asserts that the estimates generated by SEI are "not always . . . not necessarily" biased because of aggregation. By the end of the book (p. 284), he has become very circumspect: "Even if there is a large amount of aggregation bias in the data, the basic model [SEI] can still be run if the bounds are sufficiently narrow or the truncated bivariate normal looks as if it will be heavily truncated."

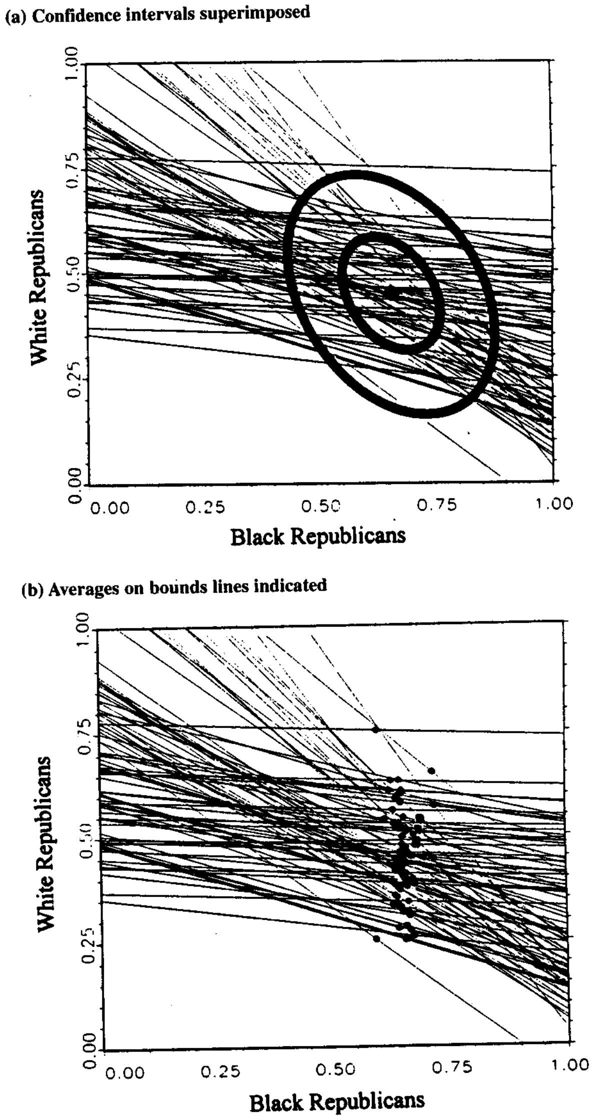
King is also confident of the power of SEI to overcome what might be termed *compression bias*. If we could observe each individual's vote, or if everyone lived in entirely homogeneous areas, then all the lines on the bounds plot would reduce to points on the periphery of the unit square and all the points on the AU plot would lie on one or the other vertical axes. But because we only observe the votes of people after they have been mixed together in precincts or counties, the observable variation in behavior is considerably less—that is, it is compressed. Thus, counties may vary only from 0 to 60 percent black, 0 to 30 percent Populist, or, in elections after 1920, from 48 to 52 percent male. These compressions of the data, which severely undermine the ability of ER to produce reliable estimates, can be gauged by Bradley Palmquist's "inflation factor" (which can be easily obtained with King's software), or they can be indexed by the variance of the independent variables in SER or MER. Despite what others have seen as the extreme difficulty of overcoming compression bias, King (1997, 24) asserts that SEI "usually" produces accurate estimates, even "when the process of aggregation eliminates most of the variation in one of the aggregate variables, and

TABLE 10
Republican Racial Composition in the North Carolina Governor's Race, 1896: A Comparison of ER and EI Estimates

Race	ER estimate % Republican	ER estimate *SE*	EI estimate % Republican	EI estimate *SE*
	A. SER versus SEI			
Black	44.1[a]	4.8	53.5	14.7
White	38.8	2.0	33.9	6.6
	B. Counties below 30% black			
Black	28.6	12.4	39.7	24.9
White	52.4	2.6	50.8	5.4
	C. Counties above 30% black			
Black	89.1	6.2	61.8	18.2
White	25.4	4.9	46.1	14.4
	D. Weighted averages of panels B and C			
Black	73.6		56.2	
White	41.6		48.9	

[a]Entries are percentages of each race that voted Republican (as a percentage of the Democratic and Republican votes).

FIGURE 9
Bounds Plots for EI Estimates of the Percentages of Blacks and Whites Voting Republican (of the Two–Major-Party Vote) in the North Carolina Governor's Race, 1896



when extrapolations far from the range of observed data are necessary." He presents no theoretical argument to back this assertion.

How does EI deal with a case in which the way people were arranged into counties obviously mattered? Does this case support King's claims of EI's ability to overcome aggregation bias? For simplicity, I consider the relationship between the percentage of the Democratic and Republican votes that went to the Republicans in the 1896 North Carolina gubernatorial election with the proportion of adult males in each county who were black.[39] Table 10, panel A, shows that SER and SEI produce quite different estimates of the proportions of blacks and whites who voted Republican, instead of Democratic. According to the SER estimate, only about 5 percent more blacks than whites voted Republican, whereas according to SEI, the margin was nearly 20 percent. Even though we have no survey or poll-book evidence of the way individuals voted, the SEI estimate is much closer than the SER estimate is to everything we know about the election from qualitative sources.

Plots produced by King's EZI software provide further insights into the way SEI works in practice. Figure 9(a) shows the bounds plot for this data with the mode and the 50 and 95 percent confidence intervals of the particular bivariate normal distribution that was generated in step one of SEI. The figure contains two signs favorable to an adequate estimate: (1) there is only one place where lots of bounds lines approximately cross and the mode falls in its center, and (2) the confidence intervals are rather tightly packed around the mode, indicating a small variance in the estimates. On the other hand, the fact that none of the bounds lines is in the southwest or northeast corners of the figure is unfavorable, for it indicates that the bounds do not constrain the estimates very much. As the shape of the contour ellipses and the tilt of the lines demonstrate, the bounds constrain the estimate of black Republicanism much less than they do that of white Republicanism. Figure 9(b) depicts the means on each bounds line of the numbers randomly chosen in step two from the truncated bivariate normal that had been estimated in step one. They form nearly a straight line drawn vertically through the mode shown in

FIGURE 10
Bias Plots of Percentage African American versus EI Estimates of Percentages of Blacks and Whites Who Voted Republican in the North Carolina Governor's Race, 1896

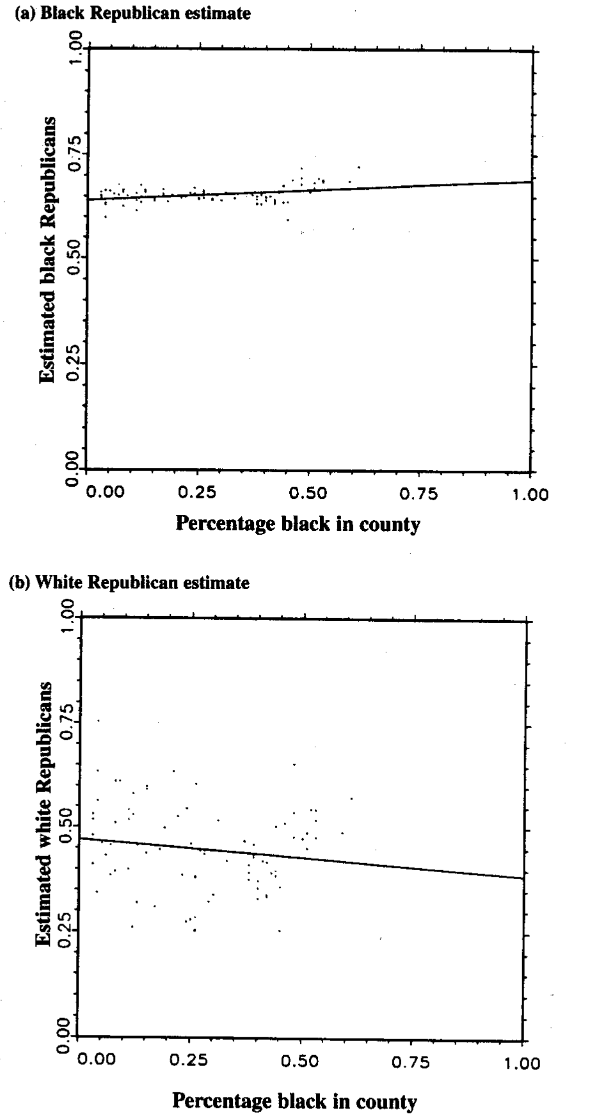


FIGURE 11
Bounds Lines for Black and White Republicans, Separately, versus Black Percentages in Each County in the North Carolina Governor's Race, 1896

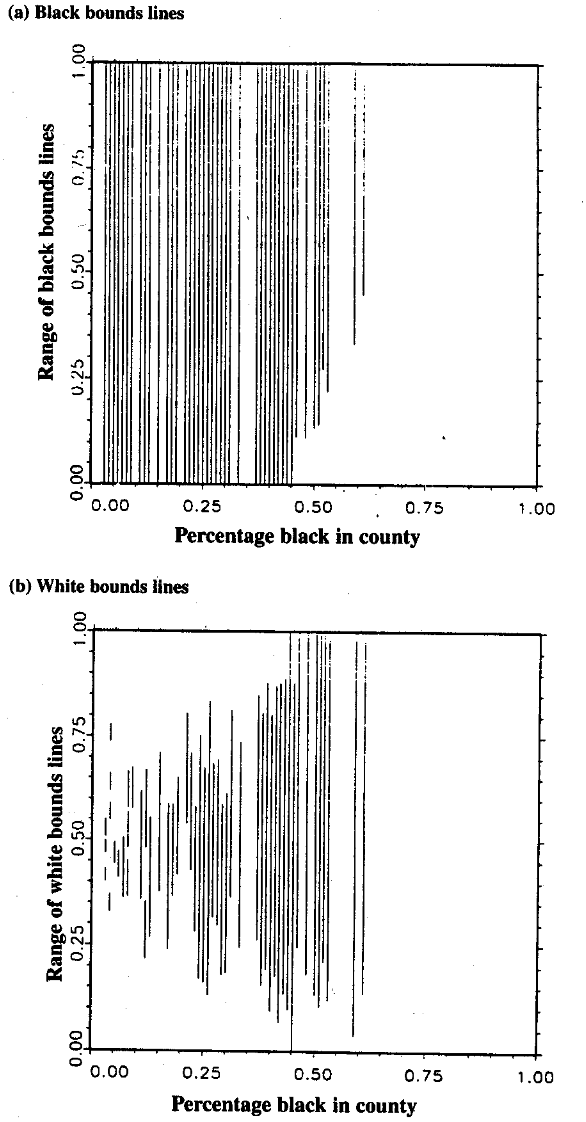


figure 9(a), which is another good sign. By contrast, a pattern of wide scattering would be evidence of imprecise estimates, whereas a pattern of clustering in two or more areas of the unit square would indicate aggregation bias or at least that the assumption of bivariate normality is incorrect.

Two of King's other indexes also provide ambiguous indications of aggregation bias. Figure 10 shows the bias plots of the percentage black against the estimates for each county of the proportion of blacks and of whites who voted Republican. This is an AU plot of the known percentage black and the estimates that EI generates for each county of the percentages of each race who voted Republican. Both bias plots have nearly flat slopes, which is a sign of no aggregation bias because it suggests that there is no correlation between the independent variable and the estimated quantities. The next graph, however, clouds the picture. Figure 11 depicts separately the length of the bounds lines for white and black Republicans that correspond to each county's value of the percentage black. The length of each line is the distance on the relevant side of the unit square spanned by the county bounds line. In a previous example in this article, in a hypothetical county that was 27 percent black, the proportion of white Republicans could have varied from 18 percent to 55 percent. Its line in figure 11 would therefore appear at 27 percent on the horizontal axis, and it would stretch from 18 to 55 percent on the vertical axis. The line for black Republicans in this county would also appear at 27 percent on the horizontal axis on its graph, but it would extend from 0 to 100 percent vertically. On the whole, figure 11 reveals no striking pattern in the graph for African American Republicans, but a consistent one for white Republicans: as the proportion black increases, so do the lengths of the bounds lines. The significance of this finding is that for this particular election in this particular state, any method such as SEI that relies on the bounds will make less-reliable estimates for white Republicans in counties with relatively large percentages of blacks than it will for those in counties with relatively small percentages of blacks. In sum, the diagnostics offered by the bounds plots and the estimates from steps one and two provide mixed evidence of the efficacy of the procedures in this case: The county-to-county estimates (step one) seem likely to be adequate, except possibly for white Republicans in heavily black counties, but the bounds lines do not provide much constraint.

Why is the SER estimate so contrary to our understanding of this much-studied election, what can be done about it, and how would parallel moves affect SEI? As one recalls from figure 3 and table 5 and their accompanying discussions, the relationship between Republican support and the percentage black appears to differ dramatically in counties above and below about 30 percent African American.[40] To produce table 5, I added both intercept and slope dummy variables for counties that fell above and below 30 percent black. For table 10, I simply calculated different regressions for the two sets of counties; weighted the resulting estimates by the populations of blacks and whites, respectively, in each set; and divided by the statewide populations of each race. Similarly, I calculated different SEI estimates for each set of counties and then weighted and divided as with the SER estimates.

The results are rather surprising, for correcting the specification of the statewide equation by dividing the state into two groups of counties produced a much more realistic estimate when I used ER rather than EI. The ER estimates in panels B and C are more extreme, finding more white Republicans than EI in overwhelmingly white counties and more black Republicans than EI in more heavily black counties. But because nearly three-quarters of North Carolina's black population in 1896 lived in counties in which the African American population exceeded 30 percent, the much larger ER estimate of black Republicanism in these counties swells the recombined statewide ER estimate of blacks who voted Republican to 73.6 percent, compared with 56.2 percent for EI. The gap between whites and blacks of the recombined ER estimate is 32 percent, compared with only 7.3 percent for EI. If the EI estimate were correct, the restriction of the suffrage in North Carolina in a 1900 referendum would have been neither necessary nor possible for the Democrats—unnecessary because the Democrats would have attracted enough black votes to avoid repeated electoral losses even without disfranchising African Americans, and impossible because they would have had to disfranchise too many whites to preserve "the party of white supremacy," as the southern Democratic party then called itself. Plots like those in figures 10 and 11 for the counties over 30 percent black indicate that the EI estimates of the voting behavior of both races in these counties are subject to aggregation bias and that the bounds for white Republicans there grow looser as the proportion of blacks increases. Likewise, the EI confidence intervals for both the black and white estimates are very much larger for the counties over 30 percent black than for the state as a whole, providing another indication of the poor quality of the EI estimate.

One may draw three conclusions from this example. First, integrating the information in the bounds with statewide estimates is useful only if the bounds narrow the estimates. Here, they provide little constraint. Second, several of King's indicators of aggregation bias and other worries about the EI estimates did suggest concerns about the bias and precision of the EI estimates, especially for the subset of counties where the black percentage exceeded 30. Third, the respecification of the relationship between race and party led to better estimates under ER, but worse estimates under EI. It is easier to investigate and evaluate different specifications under ER than under EI, and to the extent that better specifications can solve the problem of aggregation bias—as Hanushek, Jackson, and Kain argued in 1974—ER still has an important role to play in ecological inference.

The Past and Future of Ecological Inference

Historians, political scientists, sociologists, and statisticians have been making ecological inferences for more than a century. In the 1890s, Orin Grant Libby compared maps of federalist and antifederalist counties and minor civil divisions in the elections for delegates to the conventions that ratified the U.S. Constitution to maps of the areas' socioeconomic traits (Bogue 1998, 170). Unknown to Americans, a German political scientist, F. Bernstein, discovered ecological regression a generation before Goodman did (Lohmöller and Falter 1986). Since Robinson's critique in 1950, there have been repeated cycles of methodological innovations, followed by critiques, which were eventually followed by yet more innovations and more critiques. In the latest round, Gary King offered a bold and sophisticated combination of previous methods, coupled with a critique of Goodman and a set of new tests to assess the degree of aggregation bias and the success of the method in overcoming it in particular cases. Now EI, too, has begun to accumulate critics (Cho 1998, 2001; Freedman et al. 1999; McCue 2001; Rivers 1998). Their basic contentions are that the originality of EI and King's claims for its efficacy have been exaggerated, its assumptions are often violated, its software is imperfect, other existing methods are better, or new methods are needed to deal with aggregation bias under frequently occurring conditions.

King's assertions about overcoming aggregation bias, including what I have called compression bias, are based not on statistical theory but primarily on comparing estimates derived from SEI and SER with known facts. Thus, the articles in this theme issue of *Historical Methods* significantly increase the evidence on his claims and extend the cases back in time. Beyond the multiplication of this sort of empirical evidence, what can we now say about the advantages and disadvantages of ER and EI, taking into account critical comments already made by others?

First, in incorporating the bounds explicitly into a more comprehensive estimation procedure, EI represents a large advance. When the bounds are restrictive for one or more quantities to be estimated, SEI estimates are to be preferred to SER estimates.

Second, ER has often been caricatured in the literature as an inflexible method, mindlessly applied, which nearly always leads to nonsensical results. Bitterly antagonistic on other issues, Freedman, the developer of the neighborhood models, and King, the developer of EI, have agreed on that (Klein and Freedman 1993; King 1997, 1999; Freedman et al., 1999). But the characterization is careless and incorrect. At least as far back as 1973, historians and political scientists have been advised to explore and test different ER models, often with different combinations of independent variables, different functional forms, or different treatments of regional or special cases, as well as to incorporate insights from qualitative sources in their search for more accurate estimates (Kousser 1973; Lichtman 1974). And they have often done so.

Third, we must never forget that the principal aim of ecological inference is to overcome aggregation bias. Freedman's neighborhood models make no effort and afford no opportunity to move toward this goal. They assume that every manner of drawing precinct or district or county lines provides an equally bad way to examine any relationship between individual-level attributes and that there is no possibility of using any other information to alleviate the problem. The purpose of this statistical nihilism was to deny the possibility of knowledge, not to advance it. Neighborhood models were designed as graveyards for ecological inference, and one should pass them by with only a quick glance.

King's EI is much more serious, although in spite of all its advances, in some respects it is currently inferior to ER in its potential for attacking aggregation bias. EI has not yet successfully been extended to cases of multiple groups, its tests for the importance of independent or control variables are much less well developed than those of ER, and it does not seem to be as easily able to accommodate different functional forms. EI can produce tables such as tables 7–9 only by considering the four variables (Republicans, Democrats, Populists, and nonvoters) two at a time. Thus, one might estimate the origins of 1896 Republican voters by first dividing the election returns into 1892 Republicans and all others, including nonvoters, in 1892, and running EI to estimate the proportion of 1892 Republicans who voted Republican four years later and the proportion of all other members of the 1892 potential electorate who voted for the Republicans in 1896. Then, one would substitute the 1892 Democrats for the 1892 Republicans and run EI again to get analogous results. This procedure would be followed by 1892 Populists and then by 1892 nonvoters. Then one would start all over again, substituting 1896 Democrats for 1896 Republicans, and so on.

Such a procedure would not only be tedious, but it might also lead to more aggregation bias than the simpler MER procedure. Clumping all the non-Republicans into one group would increase the heterogeneity of that group, compared with that of the separate partisan groups, and it might induce a correlation between the proportion of the consolidated group and the proportion, for instance, of Populist repeaters. Suppose that the percentage of Populist repeaters varied with the strength of the Republicans in 1892, as Populists who had formed a coalition with Republicans on legislative tickets in 1894 eventually converted to the GOP. Then, the clumping procedure in EI would contain a source of bias that ER would escape. Furthermore, such clumping would lead to relatively long bounds lines in the predictions for the two major parties because many counties would be almost evenly divided between partisans of one large party and all other potential voters. Thus, one of the two chief advantages of EI over ER would be largely nullified. To take another perhaps more familiar example, if one were trying to estimate the influence of three or more significant ethnic groups—such as African Americans, Latinos, and whites—then similar problems might arise. If the proportion of African Americans who voted for Latino candidates, for example, varied with the proportion of Latinos in the electorate, then forming an artificial group of whites and Latinos, as one would have to do in EI, would increase the degree of aggregation bias, compared with running the less-aggregated MER equations. Attempts to correct this problem by extending EI have so far not been very satisfactory.[41]

Although, as in the case of ER, EI can incorporate the influence of control variables to reduce aggregation bias, the tests for the necessity or efficacy of adding such variables to EI are underdeveloped. As an alternative, Cho (2001) has recently shown that "switching regression" models with sophisticated tests for the usefulness of adding particular control variables to the predictive equations, can be applied to ecological data. Of course, ER has standard, well-formulated tests, and analysts have long been used to poring over scatter plots and plots of the residuals from equations in the search for patterns that could lead to better models. While King and others work on this problem, historians and others might use regression diagnostics to search for improvements in models before estimating what might be called multiple EI (MEI) models. If the best models uncovered in such tests turn out to be nonlinear, it will probably be preferable to use ER to estimate them.

The problems for which ecological inference is useful, even necessary, are of central importance in political science, voting-rights litigation, and, especially, in history. Goodman's initial 1959 article stimulated a generation of historians to study mass voting behavior systematically. The advances of Achen and Shively (1995) and especially of King (1997) have been widely adopted in political science, and they have also attracted often-trenchant criticisms. Extensions and reformulations of EI and more sophisticated variants of ER are being developed and used by political scientists. Cheap, high-speed computing and readily available, easy-to-use software make it possible for many more historians to rejoin the quest to understand past social and political behavior on which only aggregated data are available. All that is necessary is the will and some understanding of the methods, their weaknesses, and their great potential. The articles in this issue should provide an introduction to the means, as well as a stimulus, for nothing short of the restoration of the vitality of systematic political history.

Keywords: ecological inference, ecological regression, neighborhood models, voting rights

NOTES

Micah Altman, Wendy Tam Cho, and Gary King saved me from errors but should not be blamed for any remaining ones.

1. In this context, *ecological* has nothing to do with the environment. Rather, it refers to the units, such as counties or precincts, for which data on individuals are available.

2. Actually, Robinson was not the first to point this out. For a good short summary of early work with aggregate political data, see Achen and Shively (1995, 5–10).

3. When it contains many lines, King (1997) refers to such a plot as a *tomography plot. Bounds plot*, however, more clearly emphasizes the substance of the information.

4. The same criticisms hold for the poll-book data (those that show which candidates individuals actually voted for) that are available for a few very scattered areas in the United States and other countries: The data are very interesting—and yield relationships quite close to those made by ecological regression—but no means exist for assessing their representativeness, and there are not enough of them. See Bourke and DeBats (1979, 1980); Elklit (1985).

5. To produce statewide results from the method of bounds and simple variations of it, we can program the calculations in a few steps, using any statistical software package that allows one to sum numbers.

6. Similarly, Loewen and Grofman (1989) apply the method of bounds to relatively homogeneous precincts as a test for the validity of ecological regression estimates of racial polarization in elections. The ecological regression estimates are based on these precincts, as well as more heterogeneous ones. Like Shively, they then try to narrow the bounds by making reasonable assumptions about the relationships between cell entries. But the stronger the assumptions, the weaker the test. In other words, if they simply assume the result, it is not surprising that they find it. They refer to the method of bounds as "overlapping percentages" and to the restrictions on cell values as "simple percentages" and "complementary percentages."

7. The population figure for 1896 was computed by the conventional straight-line interpolation of those for 1890 and 1900.

8. The model was proposed during the case of *Garza v. Los Angeles County Board of Supervisors* (1991). An entire issue of the journal *Evaluation Review* 15, #6 (Dec. 1991) was devoted to articles and rebuttals based on the testimony during the *Garza* trial. For a further riposte, see Grofman (1993).

9. The U.S. Supreme Court explicitly endorsed the use of ecological regression for this purpose in *Thornburg v. Gingles* (1986).

10. The estimates of voting by race were based on unweighted regression.

11. On the grave but understudied problem of nonresponse in surveys, see Brehm (1993, 54). Latinos are especially likely to refuse to take part in surveys.

12. For criticisms of the BLN model and the "tests" that Freedman and his coauthors perform, see Grofman (1991c); Lichtman (1991); Loewen et al. (1993). For responses, see Freedman et al. (1991b); Rubinfeld (1991); Klein, Sacks, and Freedman (1991).

13. The following discussion draws extensively on Iversen (1991) and Stipak and Hensler (1982).

14. For more formal discussions of this matter, see Hannan and Burstein (1974); Firebaugh (1978); Burstein (1978).

15. The subscripts are meant only to indicate that there may be several variables of this kind, not how many there are or that some are the same in different parts of the equation.

16. This critique draws on Lichtman (1991) and Grofman (1991c).

17. Surely Freedman and his colleagues would not make such obviously fallacious arguments in their normal scholarship. See Freedman et al. (1991b, 692–97). Making AIDS a reportable disease, requiring warnings for carcinogenic hazards, reforming tort law, mandating a nuclear freeze, and handgun control were among the referendum issues when the NLN model more closely resembled the exit-poll results among Latino voters. Making English the official state language and voting to retain State Supreme Court Justice Cruz Reynoso, a Latino, were among the referendum issues and personal electoral campaigns when the SER model "won."

18. Most telling is the cavalier way in which they dismiss multiple regression as an alternative to NLN or SER in twenty-two lines of text. See Freedman et al. (1991a, 696–98). Rather than proposing alternative models, they merely raise vague cautions about the difficulties of model specification and then quit.

19. Kousser (1974) originally suggested several tables and much analysis of fraud and intimidation by observing deviant cases on scatter plots (for example, see pp. 107–9, 154, 166).

20. Sociological examples of group effects may be more common than those connected with electoral politics. Educational data may be available for classes, schools, or districts. There may be information at the level of religious institutions, social or occupational organizations, or workplaces. In such instances, the likelihood of interactions between group members that might affect variables of interest may be high.

21. By contrast, Kleppner (1987) minimizes such locutions.

22. The 1896 election took place before North Carolina adopted stringent disfranchising measures, so blacks could vote pretty freely. At that time, citizens did not have to register to vote in the state.

23. The conventional weight that is applied is 1 divided by the square root of the population. See Kousser (1980, 425–26). For a more sophisticated weighting scheme, see Achen and Shively (1995, 54–55).

24. With highly geographically concentrated parties, one might concentrate on a few nearly homogeneous units.

25. Note that there is no general intercept term in equations (19)–(21).

26. The algorithm for calculating the estimates in panel B is best shown by an example. Consider the estimate that 44.4 percent of the whites voted Democratic. First, take the intercept for whites in under 30 percent black counties from table 6 (.433). Second, multiply this by the number of whites in these counties (157,954) and conclude that Democrats received 68,405 votes from whites in counties under 30 percent black. Third, carry out an analogous operation for counties over 30 percent black (.459 * 107,843 = 49,498). Fourth, add the two vote totals and divide by the number of adult males in the state [(68,405 + 49, 498)/265,798 = 44.4%].

27. Equations of the form $Y = a + b_1X + b_2X^2 + e$, where Y is a percentage for a party and X is the percentage black, have lower R^2 values than the equations with dummy variables in the intercept and slope terms in this instance, and some of the racial percentages for the parties and not voting implied by the equations containing quadratic terms fall well outside the logical 0–1 boundaries. So the more conventional contextual variable model (i.e., the equation with the quadratic term above) may be rejected in this case on both substantive theoretical and empirical grounds.

28. The segregated data are available because influential white North Carolinians wanted to determine what proportion of taxes whites paid, so they tallied property and poll taxes by race.

29. This equation has a higher R^2 than one containing a wealth variable for the state as a whole, rather than being separated into two groups of counties.

30. Equations containing both white wealth and black percentage variables, each separated into counties above and below 30 percent black, explain about the same percentage of variance as those that exclude the wealth variables. There is one exception, which lends some support to the rationale for the predicted relationship between white wealth and political choice given in the text: There was a positive and statistically significant regression coefficient for white wealth in the counties under 30 percent black when the dependent variable was the percentage Democratic, and the percentage of variance explained by that equation, corrected for the additional degrees of freedom, was 4.7 percent higher than that in an equation that did not include wealth.

31. Scholarly articles in the debate include Grofman (1991a, 1991b, and 1993); Bullock (1991).

32. In extreme versions of the ploy, all the other variables were entered first in a stepwise regression, and the test was whether the R^2 rose significantly when the black percentage was added.

33. The height (or depth) of the strategy in legal cases was perhaps reached in *Equal Employment Opportunity Commission v. Sears* (1986), in which every conceivable variable that could reduce the apparent pattern of gender discrimination in hiring and promotion at Sears was entered into various regression equations or, if unmeasured, was assumed to correlate highly with gender and the dependent variable; the judge arbitrarily set an extraordinarily high level of statistical significance that had to be reached for him to find discrimination; and any remaining significant coefficients were dismissed on the testimony of a historian who said men and women have different (unmeasurable) goals and tastes.

34. Hair texture at an early age would have to be determined for currently bald candidates, and candidates who customarily wore toupees or wigs might have to be considered separately. The possibility of artificially altered hair (perms) opens the way for more richly textured analyses.

35. Mass-survey methods are probably as useless here as aggregate data, because whites are unlikely to give politically disapproved answers, especially when racial polarization is understood to be an issue in a legal case.

36. The areawide bounds for each variable are calculated in King's EZI or EI programs.

37. The fifth quantity is the correlation between the estimates of the two principal quantities of interest—for example, between the percentage of African Americans voting Republican and the percentage of whites voting Republican.

38. This is referred to in the statistical literature as a "random coefficients model." See Anselin and Cho (2001).

39. I chose this dependent variable instead of the preferred percentage Republican of the adult male population because the fit of the bivariate truncated normal for the preferred dependent variable was worse—that is, the confidence contours were much wider. The example in the text is thus more favorable to EI than the dependent variable used in table 5, Republicans as a percentage of adult males, would have been.

40. Although both figure 3 and table 5 use the percentage Republican of all adult males, instead of the percentage Republican of the Republican and Democratic votes, the scatter plot of the percentage black with that latter percentage follows the same pattern as in figure 3.

41. Thus, a student of King's, Karen Feree (1999, 2000), has proposed to use a multivariate normal model to estimate the voting behavior of more than two groups, but to do so she must assume that their behvior is uncorrelated—for instance, that the voting choices of African Americans are unaffected by the proportion of whites in the counties where they reside, and vice versa. This would eliminate models like that which produced table 10 in this article, which seems a high price to pay for attacking one possible source of aggregation bias.

REFERENCES

Achen, C., and W. P. Shively. 1995. *Cross-level inference*. Chicago: University of Chicago Press.

Anselin, L. and W. T. Cho. 2001. Spatial effects and ecological inference. Unpublished paper, available at <*http://cho.pol.uiuc.edu/~Wendy*>.

Baum, D. 1991. Pinpointing electoral fraud in the 1861 Texas secession referendum. *Journal of Interdisciplinary History* 22: 201–22.

Bogue, A. G. 1998. *Frederick Jackson Turner: Strange roads going down*. Norman: University of Oklahoma Press.

Bourke, P. F., and D. A. DeBats. 1979. Identifiable voting in nineteenth-century America: Toward a comparison of Britain and the United States before the secret ballot. *Perspectives in American History* 11: 259–88.

———. 1980. Individuals and aggregates: A note on historical data and assumptions. *Social Science History* 4: 229–49.

Brehm, J. 1993. *The phantom respondents: Opinion surveys and political representation*. Ann Arbor: University of Michigan Press.

Bullock, C. S. 1991. Misinformation and misperceptions: A little knowledge can be dangerous. *Social Science Quarterly* 72 (4): 834–39.

Burstein, L. 1978. Assessing differences between grouped and individual-level regression coefficients: Alternative approaches. *Sociological Methods and Research* 7: 5–28.

Cho, W. T. 1998. Iff the assumption fits. . . : A comment on the King ecological inference solution. *Political Analysis* 7: 143–63.

———. 2001. Structural shifts and deterministic regime switching in aggregate data analysis.

Duncan, O. D., and B. Davis. 1953. An alternative to ecological correlation. *American Sociological Review* 18: 665–66.

Elklit, J. 1985. Nominal record linkage and the study of non-secret voting: A Danish case. *The Journal of Interdisciplinary History* 15: 419–43.

Equal Employment Opportunity Commission v. Sears. 1986. 628 F. Supp. 1277.

Feree, K. 1999. Iterative approaches to R × C ecological inference problems: Where they can go wrong. Text available at <*http://Web.Polmeth.UFL.Edu/Working99*>.

———. 2000. Beyond 2 × 2 tables: An R × C ecological inference model.

Firebaugh, G. 1978. A rule for inferring individual-level relationships from aggregate data. *American Sociological Review* 43: 557–72.

Flanigan, W. H., and N. H. Zingale. 1985. Alchemist's gold: Inferring individual relationships from aggregate data. *Social Science History* 9: 71–91.

Formisano, R. P. 1971. *The birth of mass political parties: Michigan, 1827–1861*. Princeton, N.J.: Princeton University Press.

Freedman, D. A, S. P. Klein, J. Sacks, C. A. Smyth, and C. G. Everett. 1991a. Ecological regression and voting rights. *Evaluation Review* 15: 673–711.

———. 1991b. Rejoinder. *Evaluation Review* 15: 800–16.

Freedman, D. A., M. Ostland, M. R. Roberts, and S. P. Klein. 1999. Response to King's comment. *Journal of the American Statistical Association* 94: 355–57.

Garza v. Los Angeles County Board of Supervisors. 1991. 756 F. Supp. 1298 (C.D. Cal. 1990), aff'd, 918 F.2d 763 (9th Cir. 1990), *cert. denied*, 111 S. Ct. 681.

Goodman, L. A. 1959. Some alternatives to ecological correlation. *American Journal of Sociology* 64: 610–24.

Gould, S. J. 1981. *The mismeasure of man*. New York: Norton.

Grofman, B. 1991a. Multivariate methods and the analysis of racially polarized voting: Pitfalls in the use of social science by the courts. *Social Science Quarterly* 72: 826–33.

———. 1991b. Straw men and stray bullets: A reply to Bullock. *Social Science Quarterly* 72: 840–43.

———. 1991c. Statistics without substance: A critique of Freedman et al. and Clark and Morrison. *Evaluation Review* 15: 746–69.

———. 1993. The use of ecological regression to estimate racial bloc voting. *University of San Francisco Law Review* 27: 593–625.

Hannan, M. T., and L. Burstein. 1974. Estimation from grouped observations. *American Sociological Review* 39: 374–92.

Hanushek, E. A., J. E. Jackson, and J. F. Kain. 1974. Model specification, use of aggregate data, and the ecological correlation fallacy. *Political Methodology* 1: 89–107.

Irwin, G. A., and D. A. Meeter. 1969. Building voter transition models from aggregate data. *Midwest Journal of Political Science* 13: 545–66.

Iversen, G. R. 1991. *Contextual analysis*. Newbury Park, Calif.: Sage Publications.

Key, V. O., Jr. 1949. *Southern politics in state and nation*. New York: Vintage Books.

King, G. 1997. *A solution to the ecological inference problem: Recovering individual behavior from aggregate data*. Princeton, N.J.: Princeton University Press.

———. 1999. The future of ecological inference research: A comment on Freedman et al. *Journal of the American Statistical Association* 94: 352–54.

Klein, S. P., and D. A. Freedman. 1993. Ecological regression in voting rights cases. *Chance* 6:38–43.

Klein, S. P., J. Sacks, and D. A. Freedman. 1991. Ecological regression versus the secret ballot. *Jurimetrics Journal* 31: 393–413.

Kleppner, P. 1979. *The third electoral system, 1853–1892.* Chapel Hill: University of North Carolina Press.

———. 1987. *Continuity and change in electoral politics, 1893–1928.* New York: Greenwood Press.

Kousser, J. M. 1973. Ecological regression and the analysis of past politics. *Journal of Interdisciplinary History* 4: 237–62.

———. 1974. *The shaping of Southern politics: Suffrage restriction and the establishment of the one-party South, 1880–1910.* New Haven, Conn.: Yale University Press.

———. 1976. The "new political history": A methodological critique. *Reviews in American History* 4: 1–14.

———. 1980. Making separate equal: Integration of black and white school funds in Kentucky. *Journal of Interdisciplinary History* 10: 399–428.

Kousser, J. M., and A. J. Lichtman. 1983. "New political history:" Some statistical questions answered. *Social Science History* 7: 321–44.

Lichtman, A. J. 1974. Correlation, regression, and the ecological fallacy: A critique. *Journal of Interdisciplinary History* 4: 417–33.

———. 1991. Passing the test: Ecological regression analysis in the Los Angeles case and beyond. *Evaluation Review* 15: 770–99.

Lichtman, A. J., and L. I. Langbein. 1978. Ecological regression versus homogeneous units: A specification analysis. *Social Science History* 2: 172–93.

Loewen, J. W., O. V. Burton, R. R. Brichetto, and T. Finnegan. 1993. It ain't broke, so don't fix it: The legal and factual importance of recent attacks on methods used in vote dilution litigation. *University of San Francisco Law Review* 27: 737–80.

Loewen, J. W., and B. Grofman. 1989. Recent developments in methods used in vote dilution litigation. *Urban Lawyer* 21: 589–604.

Lohmöller, J-B., and J. W. Falter. 1986. Some further aspects of ecological regression analysis. *Quality and Quantity* 20: 109–25.

McCue, K. 2001. The statistical foundations of the EI method. *The American Statistician* 55: 106–10.

Matthews, D. R., and J. R. Prothro. 1966. *Negroes and the new Southern politics.* New York: Harcourt, Brace & World.

Powell, L. N. 1989. Correcting for fraud: A quantitative reassessment of the Mississippi ratification election of 1868. *Journal of Southern History* 55: 633–58.

Rivers, D. 1998. A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data. *American Political Science Review* 92: 442–43.

Robinson, W. S. 1950. Ecological correlations and the behavior of individuals. *American Sociological Review* 15: 351–57.

Rubinfeld, D. L. 1991. Statistical and demographic issues underlying voting rights cases. *Evaluation Review* 15: 659–72.

Shively, W. P. 1974. Utilizing external evidence in cross-level inference. *Political Methodology* 1:61–74.

———. 1991. A general extension of the method of bounds, with special application to studies of electoral transition. *Historical Methods* 24: 81–94.

Stanley, H. W. 1987. Runoff primaries and black political influence. In *Blacks in southern politics*, edited by L. W. Moreland, R. P. Steed, and T. A. Baker, 259–76. New York: Praeger.

Stipak, B., and C. Hensler. 1982. Statistical inference in contextual analysis. *American Journal of Political Science* 26: 151–75.

Thornburg v. Gingles. 1986. 478 U.S. 30.

Wharton, V. L. 1947. *The Negro in Mississippi, 1865–1890.* Chapel Hill: University of North Carolina Press.

United States Postal Service

Statement of Ownership, Management, and Circulation

1. Publication Title: HISTORICAL METHODS ISSN: 0161-5440
2. Publication Number: 5 1 7 5 - 3 0
3. Filing Date: October 1, 2001
4. Issue Frequency: Quarterly
5. Number of Issues Published Annually: 4
6. Annual Subscription Price: Institutions: $111 Individuals: $51
7. Complete Mailing Address of Known Office of Publication (Not printer): 1319 Eighteenth Street NW, Washington DC 20036-1802. Contact Person: Fred Huber. Telephone: (202) 296-6267
8. Complete Mailing Address of Headquarters or General Business Office of Publisher (Not printer): 1319 Eighteenth Street NW, Washington DC 20036-1802
9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor (Do not leave blank)

Publisher (Name and complete mailing address): Helen Dwight Reid Educational Foundation, 1319 Eighteenth Street NW, Washington DC 20036-1802

Editor (Name and complete mailing address): J. Morgan Kousser, California Institute of Technology Pasadena CA 91125

Managing Editor (Name and complete mailing address): Barbara Kahn, 1319 Eighteenth Street NW, Washington DC 20036-1802

10. Owner (Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)

Full Name	Complete Mailing Address
Helen Dwight Reid Educational Foundation	1319 Eighteenth St NW, Washington DC 20036-1802

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box → ☒ None

12. Tax Status (For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes:
☒ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months (Publisher must submit explanation of change with this statement)

PS Form 3526, October 1999 (See Instructions on Reverse)

13. Publication Title: Historical Methods
14. Issue Date for Circulation Data Below: Spring 2001

15. Extent and Nature of Circulation		Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies (Net press run)		750	775
b. Paid and/or Requested Circulation	(1) Paid/Requested Outside-County Mail Subscriptions Stated on Form 3541. (Include advertiser's proof and exchange copies)	357	354
	(2) Paid In-County Subscriptions Stated on Form 3541 (Include advertiser's proof and exchange copies)	0	0
	(3) Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Non-USPS Paid Distribution	0	0
	(4) Other Classes Mailed Through the USPS	119	115
c. Total Paid and/or Requested Circulation (Sum of 15b. (1), (2), (3), and (4))		476	469
d. Free Distribution by Mail (Samples, complimentary, and other free)	(1) Outside-County as Stated on Form 3541	52	52
	(2) In-County as Stated on Form 3541	0	0
	(3) Other Classes Mailed Through the USPS	7	7
e. Free Distribution Outside the Mail (Carriers or other means)		0	0
f. Total Free Distribution (Sum of 15d. and 15e.)		59	59
g. Total Distribution (Sum of 15c. and 15f)		535	528
h. Copies not Distributed		215	247
i. Total (Sum of 15g. and h.)		750	775
j. Percent Paid and/or Requested Circulation (15c. divided by 15g. times 100)		89	89

16. Publication of Statement of Ownership: ☒ Publication required. Will be printed in the Summer 2001 issue of this publication. ☐ Publication not required.

17. Signature and Title of Editor, Publisher, Business Manager, or Owner: Susan O. Lamluck (signature), Deputy Director. Date: September 28, 2001

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

Instructions to Publishers

1. Complete and file one copy of this form with your postmaster annually on or before October 1. Keep a copy of the completed form for your records.
2. In cases where the stockholder or security holder is a trustee, include in items 10 and 11 the name of the person or corporation for whom the trustee is acting. Also include the names and addresses of individuals who are stockholders who own or hold 1 percent or more of the total amount of bonds, mortgages, or other securities of the publishing corporation. In item 11, if none, check the box. Use blank sheets if more space is required.
3. Be sure to furnish all circulation information called for in item 15. Free circulation must be shown in items 15d, e, and f.
4. Item 15h., Copies not Distributed, must include (1) newsstand copies originally stated on Form 3541, and returned to the publisher, (2) estimated returns from news agents, and (3), copies for office use, leftovers, spoiled, and all other copies not distributed.
5. If the publication had Periodicals authorization as a general or requester publication, this Statement of Ownership, Management, and Circulation must be published; it must be printed in any issue in October or, if the publication is not published during October, the first issue printed after October.
6. In item 16, indicate the date of the issue in which this Statement of Ownership will be published.
7. Item 17 must be signed.

Failure to file or publish a statement of ownership may lead to suspension of Periodicals authorization.

PS Form 3526, October 1999 (Reverse)