DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

THE "NEW POLITICAL HISTORY:" A METHODOLOGICAL CRITIQUE

J. Morgan Kousser

SOCIAL SCIENCE WORKING PAPER 105

December 1975

THE "NEW POLITICAL HISTORY": A METHODOLOGICAL CRITIQUE*

In a recent review article, one of the leading figures of the "new political history," Samuel P. Hays, argued that "a preoccupation with technique" on the part of both critics and defenders of the genre has obscured the more important advances it had brought to the discipline -- the reformulation of historical concepts and the substitution of "systematic" for "intuitive" tests of hypotheses. "The social research movement," he concluded, "critically needs to take stock of itself, seriously debate where it is going, and move from its initial enthusiasm with techniques to a concern for methods and theory."

While I would be one of the last to condemn a call for better theory, I suggest that Professor Hays has mistaken enthusiasm for technical expertise and failed to appreciate the interrelationship between modeling and measurement, that he has therefore posed a misleading choice between good tools and good thoughts, and that his proposed concentration on "conceptual reconstruction" at the expense of technique is unlikely to assure that "historical social research" will have "a constructive and wide-ranging impact on the historical imagination."[1] In particular, I shall argue that three

*I wish to thank Allan G. Bogue of the University of Wisconsin and my Caltech colleagues, John Ferejohn, Mo Fiorina, David Grether, Dan Kevles, and Forrest Nelson for their most helpful comments, some of which I rejected, no doubt at my peril. They are not, therefore, responsible for any remaining blunders or infelicities.

well-received recent examples of the new history employ procedures so unreliable as to call their conclusions into serious question, that therefore the historical profession should put more rather than less emphasis on a sophisticated training in quantitative techniques, and that the adoption of better techniques might well help clarify theory.[2]

Though somewhat disparate in their aims, Ronald Formisano's The Birth of Mass Political Parties: Michigan, 1827-1861 (1971); F. Sheldon Hackney's Populism to Progressivism in Alabama (1969); and Paul Kleppner's The Cross of Culture: A Social Analysis of Midwestern Politics, 1850-1900 (1970) all devote much of their attention to analyses of mass voting behavior, and it is this facet of these rich works that I shall emphasize in this essay. Members of the so-called "ethnocultural school," which is headmastered by their respective dissertation directors, Lee Benson and Samuel P. Hays, Formisano and Kleppner concentrate on the political behavior of religious and ethnic groups in the nineteenth-century Midwest. Hackney, on the other hand, focuses on turn of the century Alabama in an attempt to contrast what he sees as the widely divergent social and psychological characters of populism and progressivism. Interested primarily in determining the relationship between various socioeconomic traits of voters and those voters' political behavior, all three scholars draw their conclusions chiefly from interval-level data aggregated by geographical or "ecological" units. Since they are not content merely to list the characteristics of the various coalitions favoring each party, but desire instead to ferret out the reasons why the voters

acted as they did, each scholar often attempts to "control," in some manner, the relations between two voter traits for a third. Formisano and Kleppner, for example, argue that apparent correlations between wealth and voting patterns merely reflect more basic clashes between ethnocultural groups.

Given their objectives and the nature of their data on mass voting behavior, they might well have employed a more sophisticated multivariate technique than they did. To understand the statistical deficiencies of their works, it is convenient to view them in the framework of one such technique -- multiple regression analysis. Almost universally available in computer software packages and given prominent place in virtually every introductory social science statistics text, multiple regression is also one of the most powerful statistical procedures.[3] Assume that a historian has collected information on the economic, ethnic, or religious composition, as well as the voting records, of counties, townships, wards, or other geographical entities. Then he may formulate and test competing theories about the determinents of mass political behavior by estimating the parameters of an equation of the following form:

(1) $$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon,$$

where Y = the percentage of the vote for some candidate (dependent variable),

$X_1 \ldots X_n$ = various socioeconomic variables (independent variables),

$\beta_0 \ldots \beta_n$ = the values of the intercept (β_0) and the coefficients (β_1 to β_n) or parameters of the independent variables, and

ε = an error term.

If he has data for a sufficient number of units, the historian can estimate all the βs, calculate standard errors to determine how reliable each β is, run significance tests to find out whether the βs differ significantly from zero or some other number, and compute an overall statistic (R^2) as a measure of how good a predictor his equation is as a whole. Nonlinear as well as linear (straight-line) relationships between independent and dependent variables can be accomodated within a regression framework, and the effects of each independent variable is "controlled" for those of every other independent variable. Yet despite its clear superiority to other simple modes of statistical analysis, this genuinely multivariate technique is almost never employed by Hackney, Kleppner, or Formisano.[4]

Consider, in the light of regression analysis, the use throughout Kleppner's and Formisano's books of data from homogeneous townships and wards to estimate the state- or region-wide voting percentages by ethnic, religious, and occupational groups. Assume for a moment that there are only two groups of voters in the population, one expected to be pro-Democratic and the other neutral. Then areas where the proportion of the first group is high should vote more heavily Democratic than areas containing smaller percentages of that sociological group. When we have data on voting units with widely varying proportions in each group, we can use least-squares techniques

to estimate how each votes, assuming that the relationships were approximately linear; check for nonlinearities by observing the scatterplots and residual plots; and recalculate the estimates using a different model if necessary.[5] By employing data only from homogeneous areas, on the other hand, Kleppner and Formisano ignore possible nonlinearities in the relationships they estimate.[6]

The situation is exhibited in the hypothetical graph in Figure 1. Focusing on data represented by the points in the upper right-hand corner of the graph, Kleppner and Formisano assume, in effect, that relation between the percentage of the population

(Figure 1 about here)

in group X and the percentage for party Y falls on a straight line (line b). But suppose that members of group X feel less group consciousness when they make up smaller proportions of the population, and therefore vote less solidly for party Y. Then the relationship between the two variables might resemble curve a more closely than line b, and the authors' method would overestimate the state- or region-wide percentage of group voting for party Y. Or suppose that the self-consciousness of group X peaks when they feel most in conflict with some other group, which might well come when the other group makes up a nearly equal proportion of the total population. Then the relationship might follow curve c, and the Kleppner-Formisano procedure might over- or underestimate the percentage of group X voting for party Y. By excluding the possibility of finding nonlinear

relationships, the homogeneous area method may, therefore, both fail to uncover subtleties in the social process and lead to misinterpretations of voting behavior.

Even if the relationships are all nicely linear, however, the homogeneous area procedure can still lead to misestimates of the relationships because of what statisticians refer to as "truncation." Suppose we collect voting percentages only for those geographical units which contain high percentages of some group ("truncate on X"). Then, though the resulting estimate of β (the slope of line Figure 1) will be unbiased, it will not be very reliable, since the standard error, computed according to equation (2) will often be quite high.

$$(2) \qquad \text{Standard Error of } \beta = \frac{\hat{\sigma}}{\Sigma(X_i - \bar{X})^2},$$

where $\hat{\sigma}$ = the estimated standard deviation of the error term in equation (1),

X_i = the percentage of group X in area i, and

$\bar{X}$ = the mean percentage of group X in areas represented in the sample.

If we had data for areas containing widely varying proportions of the group, the denominator in equation (2) might be a comparatively large number, and our confidence in the estimate strong. But if we restrict the range of our sample to a few areas with quite similar proportions of the group, the denominator will be small, the standard error large, and our confidence that we have correctly described the

relationship correspondingly small.[7]

The predicament is even worse if we choose to look at the socioeconomic composition of areas where party Y did very well or very poorly -- the "banner units," in the terms of Kleppner and Formisano. To see why this situation ("truncating on Y") will lead to biased estimates, let us look at some actual data drawn from page 29 of Formisano's work.

To determine the relationship between the rise of the Liberty party and the Whig decline in Michigan between the 1840 and 1844 presidential contests, Formisano presents the voting percentages only for those thirteen counties in which the Liberty party scored above its statewide mean percentage in 1844.[8] Such a procedure will usually bias upward the estimate slope of the regression line, since, to follow this example, the Whigs might also have declined in areas of Liberty weakness. Figure 2 plots the percentage of voters casting Liberty ballots in 1844 against the 1840-44 change in the Whig percentage of those voting. Lines e, f, and g are regression lines calculated on the basis of, respectively, the thirteen counties where the Libertyites ran ahead of their statewide average, the seventeen where they ran below average, and for all thirty counties.

(Figure 2 about here)

Line e, estimated from the data Formisano presents, had a slope of -0.30 and a standard error of 0.13, which means that the Liberty percentage in 1844 rose about 3 percent for every 10 percent

decline in the Whig percentage from 1840 to 1844. In the counties excluded from Formisano's table, however, the slope was slightly positive, and for all counties it was -0.28 with a standard error of 0.12. Now, one may draw four conclusions from this figure: First, considering only the strong Liberty counties inflated the estimate of the regression relationship between Whig decline and Liberty rise by 7 percent and the standard error by 9 percent. Any sampling scheme which focuses on "banner units" or party elites, when these are considered dependent variables and their socioeconomic traits are considered independent variables, will in general lead to a larger standard error and an overestimate of the strength of relationships.[9] Second, the relationship between the two variables for all counties is neither very strong nor very linear. Since the Whig decline explains only 17 percent of the variance in the Liberty percentages in the bivariate linear model, we should, third, change the specification of the model by trying to fit nonlinear functions of the Whig decline and by constructing a multivariate model through the addition of other variables which might be associated with Liberty strength. Fourth, none of the preceding points would have occurred to someone whose plan of analysis did not include an examination of the slopes and scatter diagrams.

In fact, none of these authors ever presents the reader with estimates of regression parameters, scatterplots, significance tests, or other indicators of the reliability of their estimates. Kleppner and Formisano tend, instead, to rely on what might be called "gestalt correlation" -- that is, they cite or tabulate the

voting percentages in several, say, "predominantly Pietist" units (or social composition of a few highly Republican or Democratic areas), and, in effect, invite the reader to construct a pattern in his head.[10] This technique can lead to misunderstanding, since, first, the authors usually exhibit only the percentage of those who voted who cast Democratic or Republican ballots, not the percentage of potential voters. If turnout varied from place to place or election to election, considering only the percentage of those who did participate may foster misconceptions.[11] If, for example, Irish Catholics voted 95 percent Democratic, but only fifty percent of them turned out in a certain election, they should be given less emphasis in analyses of the Democratic coalition in that election than if 90 percent had taken part. Because it can alert the analyst to possible variations in participation rates, a stress on votes as a percentage of the potential electorate may also lead to more dynamic explanations of voting behavior as well as modifications in the substantive explanation of the effects of different campaign strategies, varying organizational effort, or changes in the influence of local leaders.[12]

Gestält correlation may, second, lead to what might be called "proving correlation by intimidation." After wading through lengthy reports of similar voting patterns in six or eight townships described as, say "heavily New British," the reader is tempted to cry "Enough! I believe!" Yet the overpowered reader may have given in to mere bullying, for, as we have seen, the homogeneous area data may incorrectly portray the relationships, and the author has needlessly worn himself out reiterating statistics which could have been

presented both more quickly and correctly in one or two equations.[13]

A third problem with this mode of analysis is that, despite their enormous diligence, Formisano and Kleppner are often forced to characterize an area as "predominantly" or "mostly" composed of a certain group, instead of specifying the precise percentage. Sympathies for their difficulties in collecting data aside, we should be aware that variations in the percentages within a range of 50 to 100 percent might markedly change our conclusions about the strength of a relationship, as Figure 3, based on hypothetical data, demonstrates. Suppose that two sets of townships, imprecisely characterized as "largely" containing members of group X, recorded identical 70 percent margins for party Y. Assuming linear and error-free relationships, and that the percentage of other social groups favoring party Y was the same in both situations, the regression slopes in this example would vary from 0.83 to 0.56, depending on whether the exact percentage group X in the sets of townships was 60 percent or 90 percent.

(Figure 3 about here)

A fourth difficulty with gestält correlation is that it can lead easily to error even if the data appears to be relatively precise. In an attempt to show the connection between religion and politics, Formisano tabulates the Whig percentages and a proxy for evangelical strength in 1850.[14] As Table 1 makes clear, the apparent pattern in Formisano's tables fades when the data are subjected to regression

analysis. Judged by a standard t test for significance at the 0.05 level, none of the regression slopes is significantly different from zero. The nonsignificant slopes and the low percentages of variance explained (R^2s), as well as scatterplots not shown here, imply that on the basis of this data, at least, evangelicals were no more likely than nonevangelicals to vote Whig.[15]

(Table 1 about here)

Even if one went beyond these informal methods, it is not true, as some have claimed, that data at the township or ward level would always be superior to data aggregated by counties.[16] Suppose, to take the simplest case, that the percentage voting for some candidate was determined almost entirely by two independent variables, say, ethnicity (X_1) and class (X_2), and that the independent variables were also related to each other, as presented in equations (3):

$$(3a) \qquad Y = \beta_1 + \beta_{Y,X_1} X_1 + \beta_{Y,X_2} X_2 + \varepsilon_1$$

$$(3b) \qquad X_1 = \beta_2 + \beta_{X_1,X_2} X_2 + \varepsilon_2$$

$$(3c) \qquad \beta_{X_1,X_2} \neq 0.$$

Then the estimate of β_{Y,X_1} would be a function not only of the bivariate relationship between ethnicity and voting, but also of the relation between class and ethnicity (β_{X_1,X_2}).

Employing equations (3), we can illustrate conditions under which data aggregated at the county level would be preferable to data at the ward and township level. First, suppose X_2 was measurable on a county, but not on a sub-county basis. It might be, for example, that a national or state census contained measures of certain variables relevant to voting for the larger, but not the smaller geographical units. In this case, the estimate of β_{Y,X_1} would be biased at the township level because X_2 would be excluded from the equation, but β_{Y,X_1} would not be biased in the fully-specified county equation. Second, suppose that measures of class were not available for either counties or townships, and class and ethnicity happened to be correlated at the sub-county, but not at the county level. This might be the case if, for instance, the native-born whites held the best agricultural land within each county, but that a large proportion of the immigrants lived in cities which were, on the average, more prosperous than the rural areas. Thus, though wealth and ethnicity might be related at both the individual and township levels, the relationship at the county level might "wash out." Under these conditions, the estimate of the relation between ethnicity and voting would again be biased in the township, but not in the county data. The point in both these examples is not that county-level statistics are always preferable to those taken at lower levels of aggregation; indeed, the sub-county estimates may be better most of the time. Nevertheless, the township estimates are certainly not necessarily to be preferred, for under conditions which may well arise, estimates made with county data will contain less bias.

Nor is individual level data from pollbooks or county directories, as in Richard Jensen's The Winning of the Midwest (Chicago: University of Chicago Press, 1971): 59-62, 310-14, always preferable to aggregate data. First, the sample of surviving records is certainly not random, and may be too biased to permit confident generalization to the population of, in Jensen's case, "the entire Midwest" (p. 59). Only 8.8 percent of Jensen's Geneseo, Illinois sample, for example, were Roman Catholics, compared to about 17 percent in the entire midwest (pp. 61, 87). And since none of Jensen's directories covers a large urban area, conclusions based on them might distort relationships in this fairly heavily urbanized region. Second, because of the unrepresentative population mixes in these townships, individual voters might well have acted differently than they did elsewhere. The fact that only half of the Geneseo sample listed a church affiliation, for instance, might have increased the group consciousness of the township's pious, relative to that of churchgoers elsewhere in the region, and thereby produced unusually intense sectarian political cohesiveness in the township. Third, some variables which conceivably influenced voting, such as wealth, are either missing or perhaps imprecisely measured in Jensen's individual data. The occupational labels in the directories, for example, may have lumped together prosperous commercial farmers with poor subsistence farmers or referred to someone who spent part of his time tending a small store as a "businessman." In sum, while individual-level data may be quite informative, it is by no means an infallible guide to the truth.

Formisano, Hackney, and Kleppner do not, furthermore, escape difficulties when they employ more formal statistical procedures. All three resort to Pearson product-moment correlation coefficients based on data grouped by geographical units in order to estimate individual-level relationships. But as W. S. Robinson showed in a famous article in 1950, correlations computed from aggregate data may differ markedly from the (unknown) individual correlations. And long before publication of these three books, Leo Goodman had demonstrated superior regression analytic alternatives to ecological correlation.[17] Moreover, even if the individual correlations were correctly estimated, they would not yield much interesting information. The squared correlation coefficient (R^2) tells us the percentage of variance explained -- that is, how well our model predicts the independent variable. The regression parameters (the βs from equation [1], on the other hand, give us the actual prediction, a fact of much greater interest. What do we learn, for instance, from the table entry (Hackney, p. 336) which lists the correlation between the Negro percentage and the Populist percentage in the 1892 Alabama governor's race as -0.59? Is it not considerably more useful to an understanding of the effects of campaign strategies, the success of interracial radicalism, and the degree of ballot-box chicanery to be able to estimate from regression equations that, of the 64 percent of the blacks who voted, only about 22 percent of them had their ballots counted for the Populists?[18]

If it is difficult to comprehend why our three authors compute ecological correlation coefficients to estimate statewide

relationships, it is even harder to determine why Hackney and Kleppner often substitute rank order for Pearsonian correlations and why Formisano and Kleppner calculate respectively, rank order and Pearsonian coefficients within counties. For unstated reasons, nearly one-fifth of the correlations in Hackney's book (pp. 336-43) are rank-order correlations calculated for the full set or various subsets of the Alabama counties. Discarding knowledge of the amount of the difference between, for instance, the percentage of Negroes in two counties, Hackney, in effect, settles for the knowledge that one county scored higher than the other when all counties were ranked by their black percentages. In addition to squandering information, rank-order procedures eliminate the possibility of discovering nonlinear relationships, as well as the opportunity to employ powerful distributional assumptions to test whether coefficients differ significantly from zero.[19]

The other two author's use of within-county correlations is just as mysterious. Attempting to determine how French Canadians voted in the Midwest in the period from 1876 to 1892, Kleppner, for example, ran ecological rank-order correlations between the Democratic and French Canadian percentages separately for each of four Michigan counties and the city of Detroit (p. 58). But relationships within counties cannot logically be generalized past the county level for a single election unless we are willing to make the extreme assumption that, in this case, each of the counties is a microcosm of a static Midwest, or, to put it in regression terms, that the values of the parameters and error terms were the same within each county in these elections

as they were in equations predicting voting behavior for the Midwest as a whole in all elections for sixteen years. Running within-county correlations can also be misleading because, with only a small number of observations, chance factors can easily produce seemingly large, but actually statistically insignificant, correlations. In Kleppner's table on page 58, for instance, only four of ten coefficients are significantly different from zero at the 0.05 level. In fact, of fifty-five within-county correlations in the five tables in Kleppner's book containing such correlations (pp. 52, 58, 68, 334, 359), only eighteen to twenty of the coefficients are significant, and Formisano's box score in five similar tables (pp. 52, 55, 232, 292, 297) is only twenty-three or twenty-four of ninety-five.[20] The methods employed by these "new political historians," in short, do not inspire much confidence in their conclusions about mass voting behavior. They may be wholly correct, partly correct, or wholly incorrect -- on present evidence, we cannot tell. In light of the stress Formisano puts on the role of the "new British" immigrants, it is perhaps fitting to hand down a "Scotch verdict": not proven.

Yet if part of a critic's task is to point out deficiencies in existing treatments, another equally important part is to suggest improvements. Suppose election returns for all important elections in a certain place and time period are available on the county level and most, also, on the township and ward level, but the data on the socioeconomic composition of these areas must be painstakingly recovered from a variety of obscure sources. Then a logical research strategy would be to draw a random sample of townships and wards from the state or region, concentrate on discovering all relevant

socioeconomic facts about these units, and then apply a multivariate technique, such as regression analysis, to the township and county data.[21] Some of these results could be cross-checked against results of similar analyses on the more readily available and perhaps richer county data. To see how a regression analysis of such data might help an analyst to specify his concepts, as well as provide firmer answers to questions already raised, let us concentrate on the theories of the ethnocultural school.

A simple formulation of the ethnocultural thesis might be expressed in equations of the following form:

$$(4) \qquad Y = \beta_o + \beta_i X_i + \ldots + \beta_j X_j + \beta_k X_k + \ldots + \beta_l X_l + \beta_m X_m + \ldots + \beta_n X_n + \varepsilon ,$$

where the parameters are naturally defined, and

Y = the percentage for some party,

$X_i \ldots X_j$ = measures of wealth,

$X_k \ldots X_l$ = measures of ethnic composition, and

$X_m \ldots X_n$ = measures of religious affiliation.

Ethnoculturalists would expect all the coefficients for the wealth variables to be zero, many of the other coefficients to be nonzero, and the percentage of variance explained to be high. (It is not clear whether they would postulate that all the relationships would be linear.) If the regression was run with separate stages for each of the three sets of variables, ethnoculturalists would predict that the amount of variance explained by the ethnic and religious sets

would be much larger than that explained by the wealth factors. In either case, one might be able to disentangle, statistically, the effects of ethnicity from those of religious spirit, and thereby determine what weight ought to be attached to each in an interpretation of voting behavior. Inclusion of nonlinear functions of the independent variables might shed light on the process of group identity formation. Were evangelicals, for instance, Republicans wherever they lived, or did they have to compose a large percentage of the population before they became politically self-conscious? Was there possibly a curvilinear relationship between wealth and Republicanism, as Formisano seems to imply on page 292?

One could also use multiple regression techniques to test the very strong assumption that the clash of two unified world views (Kleppner's Pietists and Ritualists) underlies nearly all of nineteenth century American political history. A regression of temperance strength in statewide referenda or local option elections on support for the GOP should, if Kleppner is correct, produce a zero intercept and a regression coefficient close to 1.0. Antiliquor, anti-Catholic, antislavery, and pro-Negro rights sentiment expressed in referenda or in votes for candidates specially associated with those issues should also yield high positive coefficients when regressed on Republican strength and when regressed on each other. The possible impact of a transportation-poor area's desire for better canals or railroads could be evaluated by adding to equation (4) a variable measuring the distance of each county seat from the nearest railroad or canal, or a dummy

variable coded one or zero, depending on whether such transportation was or was not available. The role of a "swing group" in deciding a crucial election, such as Kleppner's German Lutherans in 1896 (pp. 323-326), could be delineated by regressing the appropriately normalized party percentages on the ethnic group percentages in two successive elections. If other determinants of voting behavior were uncorrelated with the ethnic percentages, this procedure would yield unbiased estimates of each ethnic group's voting in the pair of elections from which identification of the switchers and loyalists would follow. One could then test whether the regression coefficients for a group differed systematically in the two elections by running an "F" test of significance.[22]

Perhaps some member of the ethnocultural school might object to equation (4), claiming that the influences of nationality and theology were inherently inseparable. But that idea, too, could be tested by adding to equation (4) terms of the form

$$\beta_{km}X_kX_m + \ldots + \beta_{ln}X_lX_n. \tag{5}$$

The prediction might then be that only the coefficients of the interaction terms involving both ethnic and religious variables would be nonzero, while the coefficients of all the purely ethnic and religious, as well as the economic, variables would be nonsignificant. Whatever the final equation, the very process of stating the hypothesis formally and examining the evidence with a technique which forces one to be explicit

could lead to major changes in the interpretation of voting behavior, as well as to significant advances in theoretical clarity.

An Associate Professor of History at the California Institute of Technology, Kousser is the author of The Shaping of Southern Politics: Suffrage Restriction and the Establishment of the One-Party South, 1880-1910 (1974).

FOOTNOTES

1. Samuel P. Hays, "Historical Social Research: Concept, Method, and Technique," Journal of Interdisciplinary History 4 (1974): 475-82. By "method," Hays appears to refer not to statistical procedures but to "systematic rather than intuitive" tests of propositions drawn from various sources. In another review article, Richard Jensen has claimed that students with no background in statistics, computers, or, presumably, calculus or linear algebra can, in ten weeks, "learn enough technique to criticize intelligently most of the literature and to initiate their own quantitative projects." My experience, as well as the examples analyzed here, suggest that this promise of a short and easy road to statistical expertise is misleading, and may well result in a great deal of wasted effort or bad history. See Jensen's, "Quantitative American Studies: The State of the Art," American Quarterly, 26 (1974), 239.

2. For the favorable reception of these works, see reviews of Hackney in the Journal of American History 56 (1970): 928-29, and American Historical Review 75 (1970): 1537-38; Kleppner in Journal of American History 57 (1971): 723-24, American Historical Review 76 (1971): 198, and Journal of Interdisciplinary History 4 (1974): 505-7; and of Formisano in Journal of American History 59 (1973): 713-14, and American Historical Review 80 (1975): 134-85. Short critical statements on Hackney by Allan G. Bogue and Joel Silbey, respectively, appear in Journal of Interdisciplinary History 1 (1971): 366-67, and Historical Methods Newsletter, 3 (1969): 9-11. For essentially nonstatistical critiques and analyses of the ethnocultural school, see Richard L. McCormick, "Ethnocultural Interpretations of American Voting Behavior," Political Science Quarterly 89 (1974): 351-77; James E. Wright, "The Ethnocultural Model of Voting," in Emerging Theoretical Models in Social and Political History, ed. Allan G. Bogue (Beverly Hills, California: Sage, 1973), pp. 35-56; and earlier articles cited by McCormick and Wright.

3. Obviously, I can only sketch this technique here. A good practical introduction to regression analysis is N. R. Draper and H. Smith, Applied Regression Analysis (New York: John Wiley and Sons, 1966). For a more advanced and abstract treatment, see, for instance, Henri Theil, Principles of Econometrics (New York: John Wiley and Sons, 1971). This discussion will be confined to the relatively simple ordinary least squares technique, but the essential points about the works of Formisano, Hackney, and Kleppner could be developed similarly within the framework of such more complicated statistical techniques as probit or logit analysis.

4. Formisano apparently ran multiple regressions at two points in the book, but never reports any of the resulting regression coefficients. See Formisano, p. 54, note 71, and note to table XIV.5, p. 297. Despite a great deal of talk about "multivariate

analysis" or a "multivariate approach" in works of the ethnocultural school (see, for example, Lee Benson, The Concept of Jacksonian Democracy, paper ed., [Princeton: Princeton University Press, 1970], p. x, and Kleppner, p. 4), writers in that tradition instead employ a simple variety of "spurious correlation analysis." Typically, they show that townships with the same ethnocultural composition and different levels of mean wealth voted similarly, or that other units with the same wealth, but different ethnic groups, voted differently. In both cases they conclude that apparent correlations between wealth and voting were spurious. This sort of analysis may be misleading for two reasons: First, by selecting only a few units, their descriptions of the relationships may be biased or unreliable, for reasons discussed in the text. Second, they may have distorted the nature of the relationships between the included independent and dependent variable by controlling for only one or two, rather than for all other relevant variables. For example, differing access to agricultural markets may well have led to differing voting behavior in the antebellum era, and failing to control for this variable might well lead to misestimates of the actual relation between wealth and vote decision.

5. See my "Ecological Regression and The Analysis of Past Politics," Journal of Interdisciplinary History 4 (1973): 237-62; and Eric A. Hanushek et al., "Model Specification, Use of Aggregate Data, and the Ecological Correlation Fallacy," Political Methodology 1 (1974): 89-107. A scatterplot is a two-dimensional graph on

Cartesian coordinates where the points represent the values of the independent and dependent variables. A residual plot, also in two-dimensional space, shows the distance of each point from the estimated bivariate or multivariate regression line.

6. Though Formisano is clearly aware that his generalizations hold for homogeneous areas (see, for example, pp. 10, 194) he makes no effort to determine how members of each group who lived in integrated areas voted, and usually treats the inferences he draws from segregated areas as unqualified estimates of, for example, the "Irish Catholic" vote or the "German vote." For instance, after presenting only homogeneous area data, he comments on p. 184 that "most Germans in the period from 1835 to 1852 voted heavily Democratic."

7. Examples of truncating on X occur in Formisano, pp. 51, 143-44, 148, 170-73, 180-91, 291, 301-09, 313-18; Hackney, pp. 283, 338-42; and Kleppner, pp. 20-33, 40-68, 136-39, 161, 185-90, 195, 202, 228-31, 247-48, 289-93, 295-96, 325-36.

8. It is difficult to determine from Formisano's discussion, pp. 28-30, whether he actually considers the Liberty vote as dependent or independent. In many other cases, for example, pp. 51, 139, 153, 165, 190, 276, 291, he clearly truncates on Y. The example was chosen to illustrate the bias introduced by truncation solely because data on the election returns for all counties was easy to locate in Walter Dean Burnham, Presidential Ballots, 1836-1892 (Baltimore; Johns Hopkins University Press, 1955).

9. Had Formisano chosen a higher cut-off point than 6.5 percent Liberty, say 10 percent, he could have obtained a slope of approximately -1.0, which would imply that a 10 percent Whig decline led to a 10 percent Liberty rise, and which would have been much further than the thirteen county-estimates from the true -0.28 slope for all counties. The higher the truncation point on Y, in other words, the more the bias. Presenting data on the sociological composition of only the two or three most partisan townships in a county containing, say, twenty townships, will therefore probably lead to very biased estimates of the relationship between social and political groupings.

10. Much less quantitative than the other two in general, Hackney's book contains fewer examples of gestalt correlation. But see pp. 114, 283, and the maps on pp. 343-45.

11. A similar problem arises in considering changes from one election to the next, or correlations between them, because of massive population shifts. For instance, the white adult male population of Michigan grew by about a third from 1840 to 1844. All three authors ignore this grave difficulty in their statistical analyses -- Formisano offers the unsubstantiated assertion (p. 20) that despite high geographical mobility, the "social character" of Michigan townships persisted -- but the problem is not intractable. One solution is to divide the votes for each party in each election by the population at the time of the second election, thus obtaining, for example, this regression equation for the 1844 Michigan race:

$$Y = M + \beta_1 D + \beta_2 W + \beta_3 N + \varepsilon$$

where Y = the percentage for one of the parties or not voting in the 1844 Presidential race,

M = the percentage of migrants from 1840 to 1844,

D = the Democratic percentage in 1840,

W = the 1840 Whig percentage, and

N = the percentage present, but not voting in 1844.

Note that all the variables should be divided by the 1844 white adult male populations in the geographic units.

12. Despite the fact that Formisano and especially Kleppner are concerned with explaining alterations in voting behavior, their tabular presentation of guesses about the usual voting habits of various social groups (Formisano, p. 192, and Kleppner, p. 70) lend a static quality to their works which tends to stay with readers longer than their discussions of change.

13. For an example of a reviewer apparently overawed in this manner, see Rowland Berthoff's review of Kleppner in Journal of American History, 57 (1971), 723-24.

14. While he concentrated on the relationship between evangelicals and Whig voting, Formisano, pp. 145, 340-41 presents two separate independent variables: the proportion of total church seats which belonged to evangelical denominations, or what he calls "evangelical preference," and the proportion of total church seats to population,

which he calls "religiosity." Since his bifurcation treats as equally evangelical townships where evangelical churches had seats for widely varying proportions of the population, I used evangelical preference multiplied by religiosity (which equals the ratio of evangelical seats to population) as my single independent variable.

15. Although the present essay is primarily concerned with analyses of mass voting behavior, I should note that mistaken gestalt correlation occurs in these works in other contexts as well. In a chapter which assesses the legislative behavior of Michigan Whigs and Democrats in order to support his view that Whigs "leaned toward authoritarianism," while Democrats preferred "laissez faire," Formisano stresses the party split on an 1840 vote banning railroad travel on Sunday (pp. 123-24). This is the only roll call for which he gives a party breakdown in this crucial chapter, perhaps because, as he says in discussing another vote, "Party lines were even less clearly drawn on this question..." (p. 125). A simple chi-square test on the never-on-Sunday roll call, however, demonstrates that the actual party split would have occurred between fifty percent and seventy percent of the time when there was absolutely no relationship between an individual's partisan identification and his attitude toward Sabbatarianism.

16. Richard Jensen, for example, claims in "Quantitative American Studies," pp. 236-37, that "The ecological fallacy operates so devastatingly on units as large as counties that it is virtually impossible to use . . . census and election data in a social analysis of who voted for whom. For this purpose homogeneous ward and township returns are needed . . . "

For a more extensive development of the points in this and the succeeding paragraph, see Hanushek, "Model Specification."

17. W.S. Robinson, "Ecological Correlations and the Behavior of Individuals," American Sociological Review 15 (1950): 351-57; Leo Goodman, "Some Alternatives to Ecological Correlation," American Journal of Sociology 64 (1959): 610-625. As Hanushek et al., "Model Specification," have pointed out, the "ecological" correlation in a correctly specified (usually multivariate) equation will be an unbiased estimate of the individual correlation. The generally low percentages of variance explained in the bivariate ecological correlations Formisano, Hackney, and Kleppner present, however, imply fairly large specification errors. They present no correlations based on multivariate equations.

18. The regression estimates of Negro voting behavior are mine.

19. There are, of course, "nonparametric" significance tests available, but these are much less powerful than the more conventional F and t tests based on the assumption that the error terms are distributed approximately normally. Unless one has theoretical or empirical reasons to doubt this assumption, it is preferable to use the more powerful tests. On the basis of significance tests given for Spearman's Rho in W. J. Conover, Practical Nonparametric Statistics (New York: John Wiley and Sons, 1971), pp. 245-49, 390, only twelve of the thirty-seven rank order coefficients in Hackney, pp. 339-41 were significant at the 0.05 level.

20. Since neither Formisano nor Kleppner ever present the results of significance tests for their data, it was necessary to determine the number of townships in each county from the relevant U. S. census volume. The ambiguity in my text on the numbers of significant relationships arises because it is unclear how the authors treated words within small cities. For Kleppner, I used the procedures described in footnote 18. For Formisano, I mapped the values of the correlation coefficients into their corresponding F -test levels using the formula

$$F_{1,\ n-2} = (n-2)\ r^2/(1-r^2),$$

where

$F_{1,\ n-2}$ = the value of the F distribution for a bivariate relationship and n observations,

n = the number of townships in a county, and

r = the Pearsonian correlation coefficient.

The formula is given in Maurice G. Kendall and Alan Stuart, The Advanced Theory of Statistics vol. 2 (New York: Hafner, 1967), p. 300.

21. A researcher with deeper knowledge of the availability of data for a particular area might well be able to use some sort of stratified sampling technique and then weight the data to make them more nearly representative of the population. If he had election returns for all sub-county units, he might use them to estimate the turnover from election to election and compare these with the results based on the county-wide data and the sample from the sub-county units. Another possibly interesting approach to use if the county-level

independent variable set happened to be particularly rich would be to combine county-level independent variables in an equation with township-level dependent variables. Each township would be given its parent county's value on each independent variable. The point of this suggestion, which I owe to David Grether, would be to increase the range of variation in the dependent variable.

22. For a description of the procedures to follow in running an F-test on the differences between analogous regression parameters in separate equations, see John Johnston, Econometric Methods 2d ed., (New York: McGraw-Hill, 1972), pp. 204-07.

Figure 2

The Effect of Truncating on the Dependent Variable

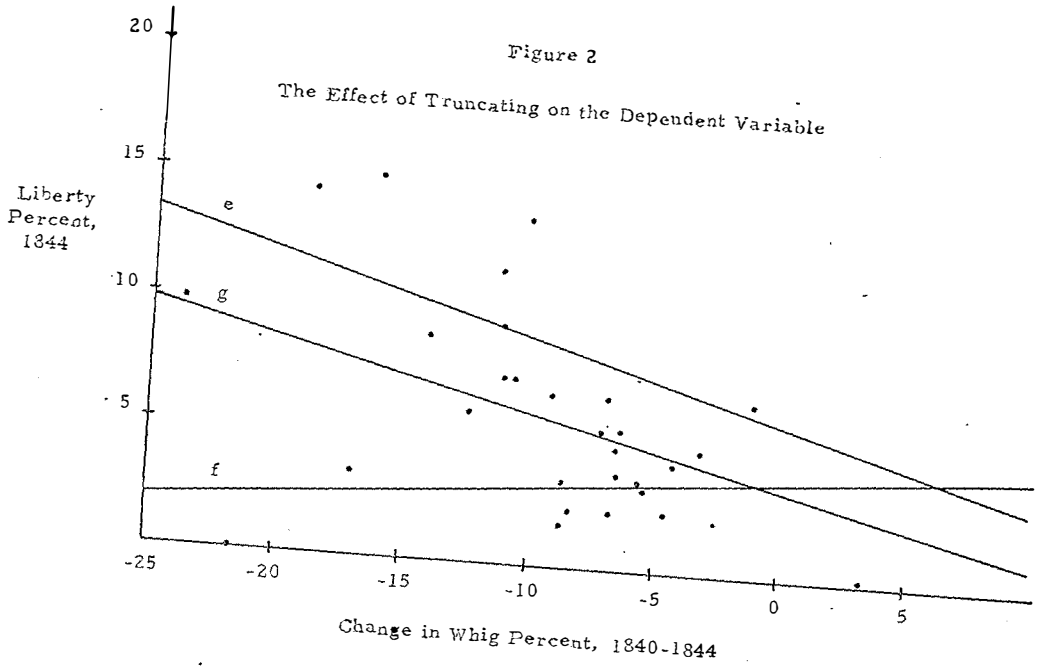

Figure 1

Hypothetical Graph Illustrating the Dangers of Generalizing from Homogeneous Areas

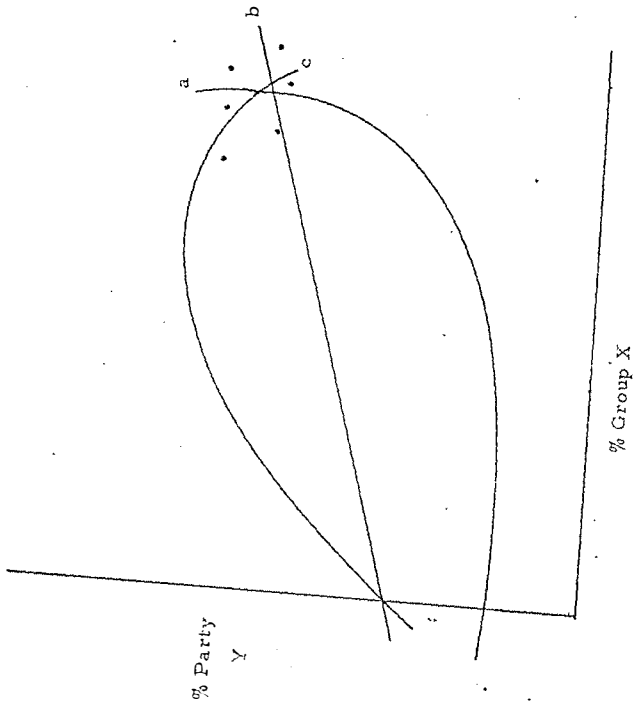

Figure 3

Hypothetical Data Illustrating Dangers of
Imprecise Characterizations of Electoral Units

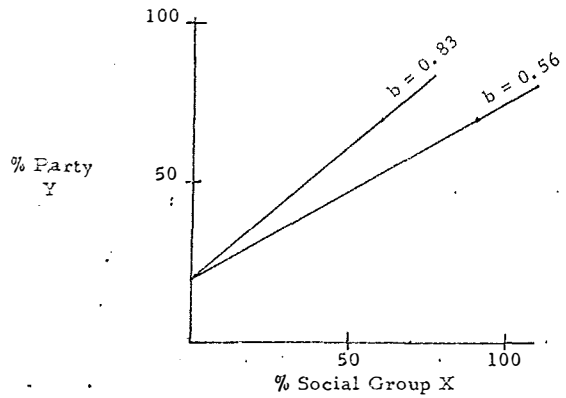

TABLE I: Evangelicals, Whigs, and Formisano

Data Source	Regression Slope	T statistics*	R^2
p. 145	0.19	1.25	0.06
Appendix C	0.31	1.44	0.06
Both p. 145 and Appendix C	0.12	1.29	0.03

*Given the number of observations (26, 35, and 61, respectively) for the three regressions run here, only values of t above 1.67 to 1.71 would be significant at the 0.05 level for a one-tailed test.