EVALUATION REVIEW Vol. 15 No. 6 Dec 1991

Ecological regression analysis has successfully estimated minority and White voting behavior in jurisdictions throughout the United States. Critics have imposed assumptions on the technique that are far too restrictive and have misinterpreted relevant data from Los Angeles County. Ecological regression is more reliable than the "neighborhood model," devised by defendants' experts for the Los Angeles litigation. Unlike ecological regression, the neighborhood model generates results that are an arbitrary function of how individuals are grouped into precincts: the classic form of the "ecological fallacy." Ecological regression, not the neighborhood model, accounts for the actual results of elections in Los Angeles County.

PASSING THE TEST

Ecological Regression Analysis in the Los Angeles County Case and Beyond

ALLAN J. LICHTMAN
American University

Social science informs judicial decisions in cases ranging from antitrust suits to criminal prosecutions. But few analytic techniques have been subjected to as stern a test of predictive capacity as the methodology, namely, ecological regression and extreme case analysis, that federal courts typically rely on for estimating the behavior of voter groups. In hundreds of jurisdictions throughout America, these procedures have guided the creation of election districts intended to provide minority voters the opportunity to elect candidates of their choice.[1] Almost invariably, such districts have elected the preferred candidate of the minority group.

This experience has now been replicated in Los Angeles County. In 1981, an all-Anglo Board of Supervisors carved the county into districts for electing future Boards, splitting the large (about 28% of the county's 1980 population), residentially compact, and growing Hispanic community into three districts with non-Hispanic majorities. In striking down this plan, District Court Judge David Kenyon relied on ecological regression and extreme case analyses showing that county voters are polarized along ethnic lines. Candidates supported by a cohesive Hispanic electorate, he found, would not be able to gain enough non-Hispanic votes to win in the existing districts. The judge ordered that the Hispanic community be united into a single district that, according to ecological regression results, would enable Hispanics to elect a candidate of their choice. Judge Kenyon rejected an attack on ecological regression launched by three defendants' experts as well as their alternative methodology that purportedly showed that Hispanics and non-Hispanics in Los Angeles County voted approximately alike.[2]

The Ninth Circuit Judge of Appeals sustained Judge Kenyon's decision and the Supreme Court denied cert. In special primary and runoff elections held in the new district in early 1991, Gloria Molina became the first American of Hispanic origin to serve on the Los Angeles County Board of Supervisors.

This article briefly reviews the statistical analyses of racially polarized voting that Bernard Grofman and I performed as experts for the United States Department of Justice, one of several plaintiffs in *Garza v. Los Angeles County*. It then responds to the objections raised by the three statisticians hired by the defendant Board of Supervisors. Both at the trial and in their article, "Ecological Regression and Voting Rights" (Freedman et al. 1991, this issue), these experts argued that ecological regression could not reliably estimate the voting of Hispanics and non-Hispanics in Los Angeles County. They also devised especially for the *Garza* litigation a "neighborhood model," which posits that irrespective of ethnic identity, people vote according to political and economic traits they supposedly share with their neighbors.

Ecological regression, it will be shown, is a flexible, robust technique capable of penetrating the reality of ethnic group voting not only in Los Angeles County but in jurisdictions throughout the country. It will also be shown that political knowledge is needed to interpret the statistical analysis of voter behavior and assess its implications for minority opportunities. Experts for the defendants exaggerated limitations of the technique and misinterpreted relevant data from Los Angeles County. Their alternative model churned out results that are simply an arbitrary function of how ethnic groups are apportioned into precincts: a classic example of the "ecological fallacy" that ecological regression analysis was designed to circumvent.

AUTHOR'S NOTE: *This article is based, in part, on the author's testimony on behalf of the U.S. Department of Justice in* Garza v. County of Los Angeles *(C.D. Cal. June 4, 1990). The opinions expressed are solely the author's own.*

EVALUATION REVIEW, Vol. 15 No. 6, December 1991 770-799

A PRIMER ON ECOLOGICAL REGRESSION ANALYSIS

Grofman and I used ecological regression to infer the behavior of voter groups from data collected for political units such as wards or precincts. Political historians such as Kousser (1974), Lichtman (1979), Baum (1984), and Gienapp (1987) have relied on this technique in the absence of direct information on how individuals actually voted. The logic of ecological regression is simple. The procedure estimates the behavior of population groups such as Hispanics and non-Hispanics from prediction equations computed for all political units within a jurisdiction or district.

Unless turnout by ethnic group is available for a particular contest, the appropriate form of ecological regression is the "two-equation" method that adjusts for turnout differences in population groups. Experts for both plaintiffs and defendants used the two-equation method in this case.[3] The method requires estimating the parameters of two equations: one in which the percentage of the population voting for the Hispanic candidate(s) in each unit is the dependent variable and the other in which the percentage of the population voting for the non-Hispanic candidate(s) is the dependent variable. The population could be measured by total population, voting-age population, registered voters, or voters turning out on election day.[4] Both sides in the *Garza* case analyzed voter precincts and relied primarily on registration data for which computer matching identified registrants with Spanish surnames.[5]

For the two-equation method, both estimated regression equations have the same linear specification:

$$Y_h = a_h + b_h X + u_h; \quad (1)$$

$$Y_{nh} = a_{nh} + b_{nh} X + u_{nh} \quad (2)$$

For Equation 1, Y_h is the percentage of registrants voting for the Hispanic candidate(s). For Equation 2, Y_{nh} is the percentage of registrants voting for the non-Hispanic candidate(s). For both equations, X is the percentage of Hispanic registrants and u is the error term.

Consider next the parameter estimates. For Equation 1, the constant term a_h estimates the percentage of non-Hispanic registrants voting for the Hispanic candidate(s) (i.e., the equation would reduce to $Y_h = a_h$, if X, the proportion of Hispanics, had a value of zero). The regression coefficient b_h estimates the difference between the percentage of Hispanic and non-Hispanic registrants voting for the Hispanic candidate (i.e., the equation would reduce to $Y_h = a_h + b_h$, if X, the proportion of Hispanics, had a value of 1). Thus the

TABLE 1: Computing Bloc Voting Estimates: The Two-Equation Method

Percentage of Hispanic voters voting for Hispanic Candidate(s)
$(a_h + b_h)/(a_h + b_h) + (a_{nh} + b_{nh})$

Percentage of non-Hispanic voters voting for Hispanic Candidate(s)
$a_h/(a_h + a_{nh})$

percentage of Hispanic registrants voting for the Hispanic candidate is $a_h + b_h$. A positive slope means that the percentage of Hispanic registrants voting for the Hispanic candidate exceeds the percentage of non-Hispanic registrants voting for this candidate, a negative slope means that the percentage of Hispanic registrants voting for the Hispanic candidate is less than the non-Hispanic percentage, and a zero slope means that the Hispanic and non-Hispanic percentages are equal.

For Equation 2, the constant term a_{nh} estimates the percentage of non-Hispanic registrants voting for the non-Hispanic candidate(s). The regression coefficient b_{nh} estimates the difference between the percentage of Hispanic and non-Hispanic registrants voting for the non-Hispanic candidate. Thus the percentage of Hispanic registrants voting for the non-Hispanic candidate is $a_{nh} + b_{nh}$. A positive slope means that the percentage of Hispanic registrants voting for the non-Hispanic candidate exceeds the percentage of non-Hispanic registrants voting for this candidate, a negative slope means that the percentage of Hispanic registrants voting for the non-Hispanic candidate is less than the non-Hispanic percentage, and a zero slope means that the Hispanic and non-Hispanic percentages are equal.

As indicated in Table 1, the percentage of Hispanic and non-Hispanic *voters* supporting Hispanic candidate(s) are derived by simple algebra from these parameter estimates for *registrants*.

For example, in the 1982 Democratic primary that pitted Esteban Torres against two Anglo candidates for the congressional nomination in District 34, ecological regression analysis yields the following parameter estimates:

Equation 1 (Hispanic candidate): $Y_h = .12 + .24X$;
Equation 2 (Non-Hispanic candidate): $Y_{nh} = .31 - .27X$.

In turn, these parameter estimates generate the following polarization measures:

Percentage Hispanics for Hispanic candidate(s) = (.12 + .24)/(.12 + .24) + (.31 − .27) = .36/.40 = 90%;
Percentage non-Hispanics for Hispanic candidate(s) = .12/(.12 + .31) = .12/.43 = 28%.

The ecological regression analysis indicates that 40% of Hispanic registrants participated in this contest and 36% voted for Torres, for an Hispanic cohesion level of 90%. In contrast, 43% of non-Hispanic registrants participated and 12% voted for Torres, for a non-Hispanic crossover level of 28%.

These estimates were corroborated by computing the degree to which the ethnic composition of precincts predicts the vote for Torres (R^2) and by an extreme case analysis that looks at the actual vote cast for Torres in the most heavily Hispanic (80%+ Hispanic registrants) and non-Hispanic (90%+ non-Hispanic registrants) precincts. If voting is polarized along racial lines, the percentage of the vote cast for the minority candidate in the largely minority precincts should be higher than the percentage of the vote cast for the minority candidate in the largely nonminority precincts. The greater the degree of polarized voting, the greater the expected differences between the vote for the minority candidate in the extreme precincts. Warning bells should sound in the event of major discrepancies between ecological regression and extreme case results.

The defendants' experts argued at the trial that extreme case analysis provides no check on ecological regression because the regression line runs through the heavily Hispanic and the heavily non-Hispanic precincts. In fact, except in the highly improbable case that contextual effects are precisely linear across precincts, comparison of extreme case and ecological regression results will disclose problems in the regression technique (see, for example, the analysis of the 1986 gubernatorial primary, shown in Table 10). No major discrepancies emerged in the Hispanic versus non-Hispanic elections that Grofman and I analyzed for the *Garza* case. In the 1982 primary, for example, Torres garnered 28% of the vote in 90%+ non-Hispanic precincts, compared to 82% in 80%+ Hispanic precincts.[6]

Table 2 reports ecological regression estimates and R^2 values for all 20 elections that Grofman and I analyzed: all 8 nonpartisan Hispanic versus non-Hispanic contests in the 1980s and 12 open-seat, partisan contests.[7] Hispanic and non-Hispanic candidates are combined; the purpose of the analysis was to determine whether there were polarized responses to ethnically mixed elections, not which particular candidate Hispanic and non-Hispanic voters preferred.

The bloc voting estimates in Table 2 are corroborated by R^2 values showing that the Hispanic composition of a precinct generally predicts at least half the variation in the vote for the Hispanic candidate. The ecological regression results for each election were also consistent with the findings of extreme case analysis. As an additional check for potential anomalies, Grofman and I examined scatter plots for all elections.

TABLE 2: Hispanic Versus Non-Hispanic Nonpartisan and Open-Seat Partisan Elections in Los Angeles County, 1980s

Election	*Percentage of Hispanic Voters for Hispanic Candidates*	*Percentage of Non-Hispanic Voters for Hispanic Candidates*	R^2
Nonpartisan			
1978 primary			
Supervisor 3	56	20	.67
1982 primary			
Supervisor 3	44	5	.86
Supervisor 1	21	9	.08
Sheriff	85	21	.53
1983 primary			
City Council District 14	73	23	.71
1986 primary			
Assessor	38	10	.27
1988 primary			
Supervisor 5	45	1	.74
1989 primary			
City Council District 7	72	0	.79
Partisan			
1982 Democratic primary			
CD 30	78	34	.50
CD 34	90	28	.68
AD 59	83	31	.45
1982 general			
CD 30	100	37	.74
CD 34	100	42	.75
State Senate District 24	100	52	.82
AD 52	100	27	.65
AD 56	95	63	.72
1984 Democratic primary			
AD 63	90	18	.73
1986 Special runoff			
AD 55	87	47	.35
1986 general election			
AD 55	80	53	.24
1988 Republication primary			
CD 30	60	45	.01

NOTE: CD = congressional district; AD = assembly district.

The results of our analysis show that Hispanic voters generally voted for Hispanic candidates in much greater proportion than did non-Hispanic voters. Hispanic voters are least likely to unite behind Hispanic candidates in supervisorial contests—a consequence of the Board's fragmentation of Hispanics into non-Hispanic districts. Neither prominent nor adequately funded Hispanic candidates sought supervisorial posts.

In four instances (all in the 1982 general election), the estimate of the percentage of Hispanic voters voting for the Hispanic candidate exceeded 100%. Investigators sometimes encounter such "out of bounds" estimates when voting is extremely polarized and the slope of the regression line is highly pitched. Using the "method of bounds," these estimates were reduced to 100%, with a corresponding increase in the percentage of non-Hispanics voting for the Hispanic candidate. The method of bounds is consistent with works such as Duncan and Davis (1953) and Shively (1974) recommending that ecological regression studies draw on what is known a priori about voting behavior. Obviously less than 100% of the voting group may actually be supporting the minority candidate, even when the estimate falls out of bounds. But the method of bounds has been verified in numerous jurisdictions by the examination of nearly homogeneous precincts that included a majority of the population group in question. These analyses consistently show group voting to be close to the 100% level posited by the method of bounds.

The results in Table 2 also help show why Hispanics have failed to elect a supervisor of their choice under the 1981 plan. Candidates run for supervisor on a nonpartisan basis. Therefore, results for nonpartisan elections and party primaries (where partisanship is held constant) are most relevant. In every one of the 12 nonpartisan and primary elections listed in Table 2, the non-Hispanic bloc vote would be sufficient to defeat the Hispanic candidate in a district where 25.3% of registered voters were of Spanish origin (the most heavily Hispanic of the districts created in 1981, according to 1988 registration data). Among those 12 contests, the three partisan primaries held in 1982 have the highest relative levels of Hispanic cohesion and non-Hispanic crossover voting, and are located within the Hispanic "core," where any Hispanic district would be drawn.

As Table 3 shows, the application of these results to a 25.3% Spanish-origin district results in a projected defeat of the Hispanic candidate. For a district with a given Hispanic percentage of registrants, the projected vote for the Hispanic candidate is a function of three factors: the turnout rates of Hispanic and non-Hispanic registrants, the Hispanic vote for the Hispanic candidate (Hispanic cohesion), and the non-Hispanic vote for this candidate

TABLE 3: Projected Vote for Hispanic Candidate by 25.3% Spanish-Origin District, Based on 1982 Partisan Primary Results and Using Spanish-Origin Data (Hispanic turnout = 62.6%; non-Hispanic turnout = 67.5%)

	Hispanic Vote for Hispanic Candidate	*Non-Hispanic Vote for Hispanic Candidate*	*Total Vote for Hispanic Candidate*
U.S. Congress, CD 30 (78% Hispanic cohesion, 33% non-Hispanic crossover)	.78 x 23.9% = 18.6%	.33 x 76.1% = 25.1%	18.6% + 25.1% = 43.7%
U.S. Congress, CD 34 (88% Hispanic cohesion, 26% non-Hispanic crossover)	.88 x .23.9% = 21.0%	.26 x 76.1% = 19.8%	21.0% + 19.8% = 40.8%
State Assembly, SD 59 (83% Hispanic cohesion, 29% non-Hispanic crossover)	.83 x 23.9% = 19.8%	.29 x 76.1% = 22.1%	19.8% + 22.1% = 41.9%
Hispanic voters = 25.3% x .626/(25.3% x .626 + 74.7% x .675) = 23.9%			

(non-Hispanic crossover). The application of turnout rates to the proportion of Hispanic registrants in the district (taken from voter sign-in data for 1988) yields the percentage of Hispanics and non-Hispanics within the participating electorate. The percentage vote for the Hispanic candidate is then obtained from three simple calculations, included in Table 3. First, the proportion of Hispanics voting for the Hispanic candidate is multiplied by the percentage of Hispanics in the electorate. Second, the proportion of non-Hispanics voting for the Hispanic candidate is multiplied by the percentage of non-Hispanics in the electorate. Third, these two products are added together.

Even though Table 3 uses the three elections with cohesion and crossover levels most favorable to Hispanic candidates, the projected vote for Hispanic candidates in a district that is 25.3% Hispanic range from 40.8% to 43.7%. In contrast, the same turnout, cohesion, and crossover statistics show projected victories for the Hispanic candidate in the 50.2% Hispanic-origin district (1988 registration data) created under the remedial plan for Los Angeles County. As shown in Table 4, calculations identical to those performed in Table 3 project majorities for the Hispanic candidate in the remedial district ranging from 54.8% to 55.9%.

AVERAGES AND ASSUMPTIONS: WHERE CRITICS GO ASTRAY

At the trial and in their article, Freedman et al. (1991), the experts for the defendants, raised two related objections to the use of ecological regression and extreme case analysis in Los Angeles County. First, the data in Los Angeles County violates a so-called constancy assumption, which holds that apart from purely random variation, ethnic group turnout and voting is the same in all precincts. Second, bivariate regression analysis yields biased results because it omits variables such as party affiliation and income that in Los Angeles County are related both to voting behavior and to the aggregation of ethnic groups into precincts.

Before responding, it is important to note that this critique cannot necessarily be generalized beyond Los Angeles County. Freedman et al. relied on the fact that relatively few of Los Angeles County's Hispanic registrants are located in heavily Hispanic precincts. At the trial, Stephen Klein testified that "the ecological problem would not be serious in someplace in say the southeast, where most of minority groups – most of minority people belong to one great many; namely they're black and they live in a racially segregated area. . . . The only time it starts getting in trouble on the voting cases is when

TABLE 4: Projected Vote for Hispanic Candidate by 50.2% Spanish-Origin District, Based on 1982 Partisan Primary Results and Using Spanish-Origin Data (Hispanic turnout = 62.6%; non-Hispanic turnout = 67.5%)

	Hispanic Vote for Hispanic Candidate	*Non-Hispanic Vote for Hispanic Candidate*	*Total Vote for Hispanic Candidate*
U.S. Congress, CD 30 (78% Hispanic cohesion, 33% non-Hispanic crossover)	.78 × 48.3% = 37.7%	.33 × 51.7% = 17.1%	37.7% + 17.1% = 54.8%
U.S. Congress, CD 34 (88% Hispanic cohesion, 26% non-Hispanic crossover)	.88 × .48.3% = 42.5%	.26 × 51.7% = 13.4%	42.5% + 13.4% = 55.9%
State Assembly, SD 59 (83% Hispanic cohesion, 29% non-Hispanic crossover)	.83 × 48.3% = 40.1%	.29 × 51.7% = 15.0%	40.1% + 15.0% = 55.1%
Hispanic voters = 50.2% × .626/(50.2% × .626 + 49.8% × .675) = 48.3%			

you don't have that kind of racial segregation, when the groups are much more mixed up."[9] Jerome Sacks was even more specific, testifying that ecological regression will give reliable estimates of group voting if about 35% to 40% of the group reside within precincts that are 90% or more homogeneous.[10]

THE CONSTANCY ASSUMPTION

Freedman et al. are misguided in their assertion that ecological regression depends on the assumption that voting behavior is constant across precincts except for purely random variation. Ecological regression estimates represent only the average behavior of groups within the jurisdiction or district under analysis. In Los Angeles County, as elsewhere, substantial nonrandom variation in majority and minority voting across precincts need not subvert ecological regression estimates of their average behavior for all precincts combined. Contrary to the implications of the defendants' experts, Grofman and I never claimed that ecological regression is assumption free. Our concern was to set forth the correct assumption: that the technique produces accurate estimates in the presence of nonrandom variations, so long as such variations are not substantially related to the percentage of Hispanics within precincts. The implications of this assumption are explicated later in this article and discussed in depth in Langbein and Lichtman (1978).

In its ability to estimate average voter behavior correctly despite subgroup differences, ecological regression mirrors exit polls. In the 1986 gubernatorial election between Democrat Tom Bradley and Republican George Deukmejian, for example, the defendants' experts relied on a *Los Angeles Times* exit poll showing that 60% of Hispanics voted for Bradley. The poll also indicated that Bradley was supported by 33% of Anglos, 94% of Blacks, and 46% of others, for a weighted, non-Hispanic average of about 42%. This average may be perfectly accurate for non-Hispanics countywide even though it overestimates Bradley's support among Anglos in Pacific Palisades and underestimates his support among Blacks in Watts and, to a lesser extent, Asians in Monterey Park (see Figure 1 for a graphic representation of the exit poll).

Ecological regression closely replicates the exit poll (see Figure 2), estimating that 45% of non-Hispanic voters and 69% of Hispanic voters supported Bradley. The accuracy of these countywide estimates is not subverted by their underestimating support for Bradley in heavily Black Democratic precincts and their overestimating his support in heavily Anglo Republican precincts.

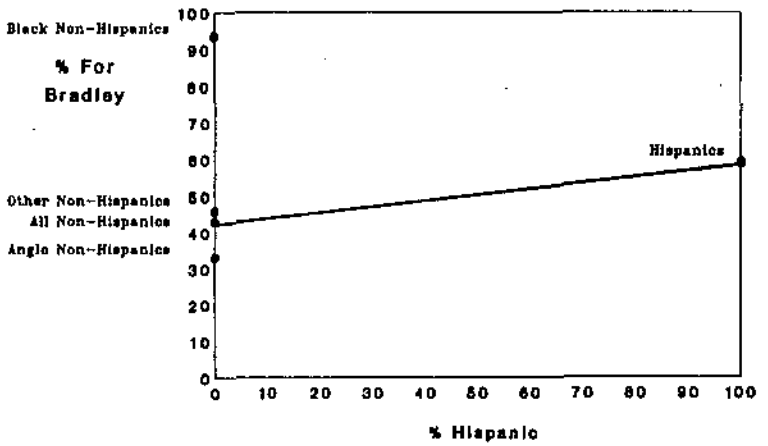

Figure 1: *Los Angeles Times* **Exit Poll Results, 1986 General Election Vote for Democratic Gubernatorial Candidate Tom Bradley, Los Angeles County**

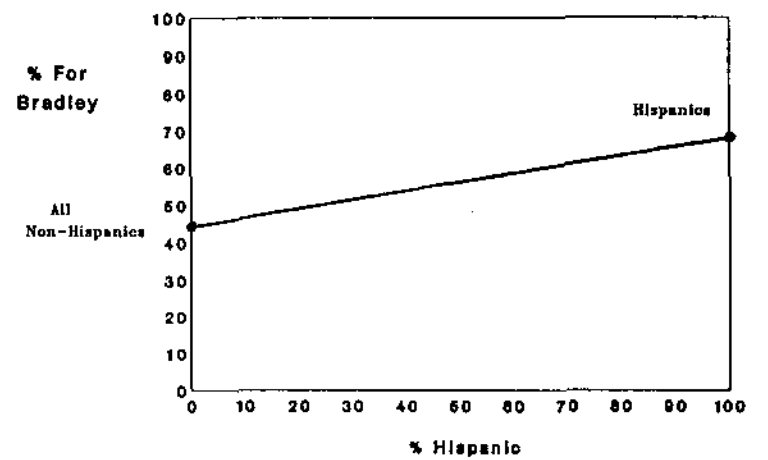

Figure 2: Ecological Regression Analysis, 1986 General Election Vote for Democratic Gubernatorial Candidate Tom Bradley, Los Angeles County

Ironically, Freedman et al. (1991) best demonstrated the robustness of ecological regression through their analysis of the Bradley vote. They showed that support for Bradley varied nonrandomly within the county by looking

TABLE 5: Countywide Ecological Regression and Exit Poll Results Used to Estimate Support for Democratic Gubernatorial Candidate Tom Bradley in Five Supervisorial Districts, 1986 General Election[a]

			Percentage for Bradley		
District	Hispanic	Percentage Democratic	Percentage Regression	Ecological Exit Poll	Actual
1	23	50	48	46	36
2	8	76	45	43	75
3	22	61	48	46	54
4	8	51	45	43	41
5	9	44	46	44	35

a. Countywide estimates of Hispanic and non-Hispanic voting from ecological regression and the *Los Angeles Times* exit poll are used to make predictions by ethnic group within each district. Votes are added over the groups to predict results for that district. The percentages for Bradley include both Hispanics and non-Hispanics.

at the five supervisorial districts. "Using the constancy assumption," they wrote, "ecological regression estimated that support for Bradley was about the same in all five supervisorial districts" (p. 697), but Bradley did best in the more heavily Democratic districts, irrespective of their Hispanic percentage. As the experts showed, countywide ecological regression results underestimated support for Bradley in the more heavily Democratic districts and overestimated support for him in the more heavily Republican districts.

Freedman et al. concluded from these results that "there is a real conflict between the constancy assumption and the data" (p. 699). Exactly. Experts for the defendants took great pains to prove what Grofman and I have been saying all along: *that ecological regression analysis does not make and does not depend on the so-called constancy assumption*. Despite violation of their version of the constancy assumption, the countywide ecological regression estimates for the Bradley race closely matched the exit poll results. The data show only that countywide ecological regression estimates may not apply to parts of the county – which is equally true of exit polls. As Table 5 indicates, exit polls and ecological regression predicted nearly identical percentages for Bradley in each supervisorial district.

It is inconsistent to rely on countywide exit polls and to reject countywide ecological regression results for making the same district-by-district errors as the polls.[11] Both ecological regression and exit polls show what they are supposed to show: the average voting behavior of all non-Hispanics in the jurisdiction.[12] This is also true for the 1986 reconfirmation of Supreme Court Justice Reynoldo Cruz-Reynoso and the vote that year on Proposition 63 (English as an official language), also cited by the defendants' experts for

violating the alleged constancy assumption. Exit polls showed that 57% of Hispanics and 42% of non-Hispanics voted to reconfirm Cruz-Reynoso, compared to estimates of 58% and 42% for ecological regression. Similarly, exit polls showed that 43% of Hispanics and 76% of non-Hispanics voted for Proposition 63, compared to estimates of 46% and 72% for ecological regression.

The experts confused how groups voted with the reasons why groups voted as they did. In the governor's race, for example, about two thirds of Hispanics voted for Bradley according to ecological regression, and about two thirds of Hispanics are Democrats. Neither regression nor poll estimates depend on whether party loyalty or other factor(s) explain Hispanic voting for a Black Democrat. Only Freedman et al.'s own misuse of ecological regression indicated that Bradley received solid Democratic support. We found major differences among Black and Anglo Democrats. Their results came from equations run for the five districts rather than the county's 6,000 precincts.

Poll results from within each district would obviously come closer to actual voting than would projections from countywide averages. But the same would be true for regressions run within districts. Still, if the experts had examined Hispanic versus Anglo elections that we analyzed, they would have found that countywide ecological regression estimates of Hispanic versus non-Hispanic voting accurately projected actual district-by-district voting. Instead, they focused on general elections involving a Black Democrat and a Supreme Court reconfirmation that were not part of our analysis. Moreover, we did not test the electability of candidates preferred by Hispanics through projections to a particular district from countywide general election results. We used only elections (a) within the county's Hispanic core and (b) without party labels.

THE "OMITTED VARIABLE" PROBLEM

Only under special circumstances will systematic variations in voting across precincts have the potential to bias ecological regression results. Variables omitted from a regression equation that are related both to the dependent and independent variables can produce variations in voting behavior that are related to the percentage of Hispanics in a precinct (for exploration of this problem, see Lichtman 1974; Hanushek, Jackson, and Kain 1974; Langbein and Lichtman 1978; Grofman and Migalski 1988). Even so, two conditions must still be satisfied if an omitted variable is to have the effect of biasing estimates of group behavior taken from ecological re-

gression results. First, there must be a strictly aggregate relationship between the omitted variable and the independent variable (in this case, the percentage of Hispanics in a precinct). That is, the relationship between the omitted variable and the independent variable must be different among precincts than among individuals. Among individuals, for example, there is a positive relationship between Hispanic affiliation and Democratic party registration in Los Angeles County: Individual Hispanics are more likely than individual non-Hispanics to be Democrats. A similar positive relationship would be expected among precincts: Heavily Hispanic precincts would have a greater percentage of Democrats than less heavily Hispanic precincts. Unless this aggregate-level relationship is different (weaker, stronger, or in the opposite direction) than the corresponding relationship among individuals, it will not bias the ecological regression estimates of Hispanic and non-Hispanic voting.

Second, the omitted variable must also be related to voters' decisions when controlling for the influence of the independent variable. In the Hispanic versus non-Hispanic elections analyzed for the *Garza* case, this would mean, for example, that irrespective of their ethnic identity, Democrats were more likely to vote for an Hispanic candidate.

The degree of potential bias depends on both the strength of the omitted variable's independent influence on voting behavior and the strength of its strictly aggregate relationship to the percentage of Hispanics in a precinct. Unless both relationships are quite strong, the potential bias will be low (see Langbein and Lichtman 1978, 19-23).

The data in Los Angeles County does show that there is an aggregate relationship between political and economic variables and the percentage of Hispanics in a precinct. But these relationships are relatively weak; unless party and economic status have a very strong pull on voting, independent of ethnic identity, potential bias will be negligible. Yet Freedman et al. (1991) provided no evidence that nonethnic variables independently influence ethnic group voting when Hispanics compete against non-Hispanics for public office. The statisticians did present information on the independent effect of party registration and income on voting, but only for the reconfirmation of Chief Justice Rose Bird—an Anglo Democrat—in the 1986 general election (p. 690). The Bird vote should be an ideal example for critics of ecological regression because Hispanic identity should not play an independent role. It is reasonable to conclude that voters responded positively or negatively to Bird because they are liberal or conservative, but not because they are Hispanic or non-Hispanic. Still, the relationships that Freedman et al. uncovered are modest. According to a statewide field poll, independent of ethnic identification, voters with household incomes below $20,000 were 7% more likely to support Bird than were voters with incomes above $20,000 (p. 690).

TABLE 6: Hispanic Electoral Success in Congressional Districts, Los Angeles County

District	Percentage Democrat	Median Household Income	Percentage Hispanic Democrat	Elected 1980?
25	63	$12,400	47	Yes
30	63	$15,700	44	Yes
34	65	$20,000	41	Yes
24	60	$13,200	6	No
28	76	$12,700	6	No
29	81	$11,700	9	No
31	74	$17,100	11	No
32	60	$17,100	13	No

Comparison with exit polls indicates that even for the critic's self-selected example, ecological regression did not inflate the difference between Hispanic and non-Hispanic support for Bird. Ecological regression estimated that 36% of Hispanics voted for Bird, compared to 39% of non-Hispanics. The *Los Angeles Times* exit poll estimated that 47% of Hispanics voted for her, compared to 35% of non-Hispanics.

There is, moreover, direct evidence on whether party and socioeconomic standing independently influence ethnic group response to Hispanic and non-Hispanic candidates. If non-Hispanic, low-status Democrats strongly supported Hispanic candidates competing against non-Hispanics, such candidates should be successful in heavily Democratic low-income districts regardless of the Hispanic component of these districts. Since the post-1980 redistrictings, however, no Hispanic has been elected to public office countywide nor in a nonpartisan district with a percentage of Spanish-origin registrants below 45% nor in a partisan district with a Democratic percentage of Spanish-origin registrants below 41%. Yet Hispanic candidates have prevailed in every nonpartisan district in which Spanish-origin registrants exceeded 45% of the total and in every partisan district in which Spanish-origin registrants exceeded 42% of Democratic registrants.

For congressional districts, Table 6 shows that Hispanics have been elected from three districts since 1981. More than 60% of the registrants in these districts were Democratic, and at least 41% of the Democrats were Hispanic. But Hispanics failed to win even a single election in any of the five

other 60%+ congressional districts, all of which had relatively low median incomes but lacked a substantial Hispanic component.

These five non-Hispanic districts include a mix of both Whites and Blacks. According to 1980 Census data, Congressional District (CD) 29 is 64% Black in its citizen voting-age population, CD 28 is 47% Black, CD 31 is 36% Black, CD 24 is 9% Black, and CD 32 is 8% Black. This same pattern of Hispanic success and failure applies to supervisorial, city council, state assembly, and state senate districts. Hispanic candidates almost invariably win in districts with substantial Hispanic components and invariably lose in all other districts regardless of their socioeconomic or racial composition.

In response to this experience of many hundreds of elections held in Los Angeles County since 1981, Freedman et al. (1991) were able to cite only the single example of Hispanic candidate Sarah Flores gaining a plurality vote in a district lacking a substantial Hispanic component. This occurred in a special primary held after the trial—but before Judge Kenyon's final ruling—in Supervisorial District 1.

Once again, however, the statisticians drew misleading political conclusions from an election result. First, contrary to their statement that she "won a contest for the County Board of Supervisors by a plurality vote" (p. 702), she actually qualified for a runoff election with Anglo candidate Gregory O'Brien. Second, the voting was polarized along ethnic lines, with Flores and other Hispanic candidates gaining virtually all of the Hispanic vote and non-Hispanic candidates gaining two thirds of the non-Hispanic vote. Third, it is common for a White, or Anglo, community to sponsor a favored minority candidate in a calculated effort to defeat a voting rights lawsuit. That is precisely what happened in Los Angeles County, as Flores had the support of incumbent Anglo Supervisors Dana and Antonovich. Judge Kenyon found credible the testimony that "these two supervisors decided to endorse the candidacy of Ms. Flores in the belief that their support of an Hispanic candidate would favorably influence the outcome of this lawsuit."[13] Fifth, in the special primary subsequently held in the remedial district (created by Judge Kenyon's order) with a 50%+ Hispanic registration, candidate Flores failed even to qualify for the runoff election. Elections in the heavily non-Hispanic supervisorial districts had attracted only marginal Hispanic candidates. But the contest in the newly formed district attracted two prominent Hispanic politicians: Gloria Molina and Esteban Torres.

For countywide elections, there is also sufficient variation in the political and ethnic composition of precincts to isolate precincts that are heavily Democratic but contain few Hispanics. According to ecological regression, the 1982 primary for sheriff was the most highly polarized of countywide

contests, with 85% of Hispanics and 21% of non-Hispanics voting for the Hispanic candidate. If the theory of the defendant's experts is correct, low Hispanic, heavily Democratic precincts should provide strong majorities for Hispanic candidates. Yet some 125,000 voters in precincts that are 80% or more Democratic and 20% or less Hispanic cast only 17% of their vote for Hispanic candidates—54 percentage points less than the vote cast for the Hispanic candidates in 80%+ Hispanic precincts (see Table 2).

THE "NEIGHBORHOOD MODEL": A CRITIQUE

The "neighborhood model" assumes that Hispanics and non-Hispanics vote identically in each precinct. The rationale is that within a precinct, members of different ethnic groups share traits such as party affiliation and socioeconomic standing that determine their voting. Freedman et al. (1991) did not defend the neighborhood model as a reliable means for estimating voting behavior in Los Angeles County or other jurisdictions. Rather, they presented it as an alternative procedure that purportedly is a priori "more plausible than ecological regression and . . . fits the facts better" (p. 678). In fact, the model has serious theoretical and practical deficiencies that make it an inherently unreliable instrument for probing group voting.

By definition, support for minority candidates can be explained simply by positing that voters in each precinct vote alike. That is why experts for the defendants could generate particular simulations consistent with the neighborhood model. For certain preselected characteristics of Hispanics and non-Hispanics, the neighborhood model can produce reasonably accurate results, but as a method for estimating ethnic polarization, the model suffers from deficiencies that make it far inferior to ecological regression. I know of no examples of political researchers actually using such a model to estimate voting behavior.

THE WRONG VARIABLES WRONGLY DISTRIBUTED

Despite asserting a priori credibility for their model, Freedman et al. accepted on faith that in Hispanic versus non-Hispanic contests, voters' party registration and socioeconomic standing determine their choice of candidates. This must be true not only in general elections but in nonpartisan primaries and in partisan primaries where party registration is obviously constant and socioeconomic variation is relatively low. Of 20 contests that Grofman and I analyzed for the *Garza* case, 8 were nonpartisan primaries

TABLE 7: Relationship Between Percentage of Vote for Hispanic Candidates in 1982 Primary for Sheriff and Percentage of Democrats and Hispanics in Los Angeles County Precincts, by Quintiles

	Percentage for Hispanic Candidates	
Quintiles	*Percentage Hispanic*	*Percentage Democratic*
Less than 10%	24	—
10%-20%	31	—
20%-30%	34	—
30%-40%	40	20
40%-50%	45	23
50%-60%	51	27
60%-70%	58	33
70%-80%	65	40
80% or more	71	26

NOTE: There are no precincts that are less than 30% Democratic.

and 6 were partisan primaries. As was already demonstrated, moreover, the available evidence undercuts the supposition that nonethnic variables determine how voters respond to Hispanic and non-Hispanic candidates.

Even if such variables did influence voter behavior, the variables would have to be distributed across precincts in a way that could explain what is known about voting in Los Angeles County. In Hispanic versus non-Hispanic elections, the vote for Hispanic candidates consistently and often substantially increases as the Hispanic component of a precinct increases (see the R^2 values in Table 2). To explain this correlation, the party registration and socioeconomic standing of Hispanics and non-Hispanics would have to change steadily and substantially as the Hispanic component of precincts increased.

The defendants' experts' own data, displayed in their Table 2, fail to show the required relationships. The correlations in their table should be squared, since R^2 values measure the extent to which variation in an independent variable accounts for variation in the dependent variable. The R^2s from their Table 2 range from nonexistent for home ownership (.001) to modest for education (.17). In contrast, Table 2 in this article shows that for the relationship between the percentage of Hispanics in a precinct and the vote for the Hispanic candidate, 18 of 20 elections have R^2 values of .24 or greater and 14 of 20 have R^2 values of .50 or greater.

The inability of nonethnic factors to account for variation across precincts in the vote for Hispanic candidates is also shown by the 1982 primary vote for sheriff. Table 7 shows that the percentage of the vote for Hispanic candidates for sheriff tracks the percentage of Hispanics in a precinct but not the percentage of Democrats.

Nonethnic variables must also be distributed among precincts to account for the presumption of the neighborhood model that Hispanics and non-Hispanics within a precinct vote alike. This means that within precincts, Hispanics and non-Hispanics must be roughly equivalent in their Democratic registration, income, and so on. Average differences on the order of 10% to 20%, the statisticians have indicated, would "create major problems for the neighborhood model."[14]

In fact, the percentage of Democrats among Hispanics and non-Hispanics are not similar, on average, within precincts. A charting of the Democratic percentages for Hispanics and non-Hispanics for each quintile of the Hispanic percentage in a precinct consistently shows differences well within the 10% to 20% danger range for the neighborhood model.

Analysis of this data points to yet another major problem for this model: 68% of Hispanics in precincts that are under 10% Hispanic are Democrats. Table 7 shows that for the 1982 sheriff's race, the Hispanic candidates garnered 24% of the vote in these under-10% Hispanic precincts. Accordingly, the neighborhood model estimates that 24% of the Hispanics in these precincts voted for the Hispanic candidates. Likewise, there is an identical 68% Democratic registration rate for non-Hispanics in precincts that were 60% to 70% Hispanic. Table 7 shows that the Hispanic candidates gained 58% of the vote in the 60% to 70% precincts. So the neighborhood model would estimate that 58% of the non-Hispanics in these precincts voted for the Hispanic candidates. This leads to a paradoxical conclusion. Despite equivalent 68% Democratic registration rates, the neighborhood model tells us that 58% of non-Hispanics in 60% to 70% Hispanic precincts voted for Hispanic sheriff candidates, compared to only 24% of Hispanics in under-10% Hispanic precincts, for a 34% difference.

As Table 8 shows, paradoxical results are also obtained for the three socioeconomic variables cited by experts for the defendants. According to the neighborhood model, the non-Hispanic percentage of persons lacking high school diplomas, of households with incomes under $20,000, and of renters in the 60% to 70% Hispanic precincts should be much higher than the corresponding Hispanic percentages in the under-10% Hispanic precincts, otherwise there would be nothing to account for the 34% difference in the model's estimated vote for Hispanic sheriff candidates between these groups of voters. Yet the non-Hispanic percentage of persons lacking high school diplomas in the 60% to 70% Hispanic precincts is only 4% higher than the Hispanic percentage in the under-10% Hispanic precincts. The non-Hispanic

TABLE 8: Vote for Sheriff, 1982 Primary, Los Angeles County (in percentages)

	Neighborhood Model Estimated Vote for Hispanic Candidates	Democratic	Did Not Graduate High School	Income < $20,000	Not Homeowners
Hispanic voters in Under-10% Hispanic precincts	24	68	30	45	47
Non-Hispanic voters in 60%-70% Hispanic precincts	58	68	34	55	33
Difference between non-Hispanics in 60%-70% Hispanic precincts and Hispanics in Under-10% Hispanic precincts[a]	+34	0	+4	+10	−14

NOTE: Neighborhood model estimates of vote for Hispanic candidates do not correspond to variables that the model presumes are responsible for voting.

a. According to the neighborhood model, all these differences should be strongly positive.

percentage of households with incomes below $20,000 in the 60% to 70% Hispanic precincts is only 10% higher than the Hispanic percentage in the under-10% Hispanic precincts, and the non-Hispanic percentage of renters in the 60% to 70% precincts is actually 14% *lower* than the Hispanic percentage in the under-10% Hispanic precincts—a relationship that cuts in the opposite direction of the model's presumption that renters should be more likely than homeowners to vote for Hispanic candidates.

MISTAKING PRECINCTS FOR NEIGHBORHOODS

Contrary to what its name implies, the neighborhood model actually analyzes precincts rather than neighborhoods. The problem is that precinct lines may arbitrarily combine or separate minority and Anglo neighborhoods, with varying results for the model. Assume, for example, that a 100% Hispanic and a 100% non-Hispanic neighborhood are adjacent to one another, divided cleanly by a north/south boundary. Assume also that 100% of the Hispanics and none of the non-Hispanics vote for the Hispanic candidate in an election. A precinct boundary that bisects the neighborhoods north to south, along the neighborhood lines, would create a 100% Hispanic precinct and a 100% non-Hispanic precinct. The neighborhood model would correctly estimate that 100% of Hispanic voters and none of the non-Hispanic voters voted for the Hispanic candidate. But a boundary that bisected the neighborhoods east to west would create a 50% Hispanic and a 50% non-Hispanic precinct. The neighborhood model would incorrectly estimate that 50% of the Hispanics and non-Hispanics voted for the Hispanic candidate. Thus the arbitrary location of precinct lines could result in estimates ranging from the mathematically maximum (100%) to the mathematically minimum (0%) polarization between Hispanics and non-Hispanics.

A GEIGER COUNTER THAT DOESN'T CLICK

Even if precincts and neighborhoods coincided, the neighborhood model would still fail to meet the basic objective of a diagnostic model: the ability to detect both the presence and absence of ethnically polarized voting in a jurisdiction or district. The model's results reflect how ethnic groups are distributed into precincts rather than how group members actually vote. The neighborhood model thus returns political analysis to the preecological regression era; its results are artifacts of the aggregated units being analyzed—the classic form of the "ecological fallacy."

TABLE 9: Neighborhood Model Results Versus Known Polarization Levels, Countywide (in percentages)

Known Polarization Level		Neighborhood Model Results	
Hispanic for Hispanic Candidate	Non-Hispanic for Hispanic Candidate	Hispanic for Hispanic Candidate	Non-Hispanic for Hispanic Candidate
100	0	31	8
90	10	35	16
80	20	39	25
70	30	43	33
60	40	46	42
50	50	50	50

For Los Angeles County overall and electoral districts in which most Hispanic registrants are included within majority White precincts, the neighborhood model cannot detect any significant degree of polarization between Hispanics and non-Hispanics. Given this precinct structure, a model positing that ethnic groups vote alike in each precinct will greatly reduce the actual spread between Hispanic and non-Hispanic voters by pulling the minority Hispanic vote toward the majority non-Hispanic vote. Table 9 shows that if 100% of Hispanics and 0% of non-Hispanics voted for the Hispanic candidate countywide, the neighborhood model would estimate absurdly that 31% of the Hispanics voted for the Hispanic candidate: an error of 69%. Because the neighborhood model skews the Hispanic vote toward the non-Hispanic vote, Table 9 shows that as the actual level of Hispanic cohesion declines to 90%, 80%, 70%, and so on, the neighborhood model churns out rising rather than falling cohesion estimates. This leads to the bizarre result that the model is better able to detect a slight departure from a 50/50 split than far more extreme polarization.

This analysis does not merely demonstrate that certain simulations confound the neighborhood model. It does not just demonstrate that the neighborhood model will be wrong when the so-called constancy assumptions of ecological regression holds. Significant polarization between Hispanics and non-Hispanics does not require constancy in the behavior of all subsets of these ethnic groups. The analysis shows something much more fundamental: that no matter what the reality of voting, the neighborhood model is inherently incapable of detecting significant polarization between Hispanics and non-Hispanics in Los Angeles County. This conclusion also applies to most electoral districts within the county, including all supervisory districts created in 1981.

TABLE 10: Neighborhood Model Results Versus Known Polarization Level, Los Angeles County and Selected Districts (in percentages)

	Neighborhood Model Results				Difference Between Actual and Estimated Polarization
	Hispanic Registrants	Hispanics for Hispanic Candidate(s)	Non-Hispanics for Hispanic Candidate(s)	Polarization Level	
Entire county	12	31	8	23	77
SD 1	21	32	14	18	82
CD 34	41	50	34	16	84
AD 56	63	77	45	32	68

NOTE: SD = supervisorial district; CD = congressional district; and AD = assembly district. Results are based on 100% Hispanics for Hispanic candidate(s) and 0% non-Hispanics for Hispanic candidate(s). All results use data from all registrants in 1982, with the exception of CD 34, which uses Democratic registration data for 1982.

The neighborhood model is like an Geiger counter that only clicks when radiation is weak or, more chillingly, an AIDS test that cannot detect AIDS. For certain low-risk populations, such a test can be shown to be correct nearly all of the time. For such populations, it may even be correct more often than a genuine AIDS test that generates a few false positives. But as a real-world diagnostic, the nontest for AIDS—like the nontest for polarization—has no value.

The evidence suggests that even at the 60% to 40% level, the neighborhood model cannot detect polarization in Los Angeles County. For both the Cruz-Reynoso reconfirmation and the vote for Democratic gubernatorial candidate Bradley in the 1986 general election, the *Los Angeles Times* exit poll showed approximately 60%-40% splits between Hispanic and non-Hispanic voters. Ecological regression results closely conform to the exit polls, but the neighborhood model finds an absence of polarization in both cases, showing a 47%-44% split for Cruz-Reynoso and a 50%-46% split for Bradley. These results are close to the 46%-42% split predicted by the simulations in Table 9 for the neighborhood model. The problem here cannot be that the results were arranged to make sure that the constancy assumption applies; on the contrary, the defendants' experts took great pains to show that the constancy assumption does not hold for either of these contests.

An additional deficiency of the model's dependence on precinct structure is that it generates different estimates of polarization in different districts, even when the known polarization level is the same. Table 10 shows that although the model does not come close to the true values in any district, the estimates vary widely across districts.

REAL AND IMAGINED TESTS OF MODELS

The lack of correspondence between neighborhood model results and group voting is also shown by comparing results of the *Los Angeles Times* exit poll for the 1986 general election to estimates produced by the neighborhood model and ecological regression. The 1986 general election is appropriate for comparison because it is by far the most reliable of available exit polls (with a sample size from 149 to 165, it surveyed more than twice the number of Hispanics included in any other available poll). The 1986 general election is also the only contest that provided a mix of both polarized and unpolarized contests according to exit polls. The analysis draws on a definition of polarization provided by Stephen Klein: that polarization occurs when a majority of Hispanics and non-Hispanics vote differently. The 1986 elections included contests with major differences between Blacks and Anglos (which should pose major problems for ecological regression, according to the defendants' experts). General elections should also maximize the influence of party affiliation and correlated factors that bias ecological regression, according to the defendants' experts.

The 1986 data include three polarized and three nonpolarized elections. Table 11 shows that ecological regression, extreme case analysis, and exit polls agreed on the identification of polarization in each case, but the neighborhood model failed to find polarization in any contest.

No matter what the exit poll indicated about voter choice, the neighborhood model grinds out the same answer: There is minimal difference between Hispanic and non-Hispanic voting. For the 1986 elections, the neighborhood model shows polarization that ranges from 0% to 4% compared to a range of 0% to 33% for the exit polls.

In their article, Freedman et al. (1991) added five more elections, chosen from among some 25 additional contests for which exit poll data are available. Unlike the choice of the 1986 general election, there is no rationale for their choices, other than that these present the neighborhood model in the best possible light. For Hispanics, the sample size from their selected polls was as low as 66 for 1982. Although the purpose of analysis was to estimate ethnic polarization, Freedman et al. told only half the story in their article, reporting results for Hispanic but not for non-Hispanic voters. So a reader cannot tell whether the chosen elections were polarized or not. In fact, as Table 11 indicates, of their five selections, only the voting for Proposition 51 in the 1986 primary was polarized, according to Klein's definition.

As Table 11 shows, when choices made by Freedman et al. were added to the 1986 general election contests, the neighborhood model missed three of four elections that the exit poll identified as polarized. And it barely got

TABLE 11: Ecological Regression and Neighborhood Model Results Compared to Exit Poll Results, Los Angeles County 1986 General Election

	Is the Election Polarized?[a]			
Case	Los Angeles Times *Exit Poll*	*Neighborhood Model*	*Ecological Regression*	*Extreme Analysis*
1986 general elections				
Governor (Bradley)	Yes	No	Yes	Yes
Senator (Cranston)	Yes	No	Yes	Yes
Reconfirm (Reynoso)	Yes	No	Yes	Yes
Reconfirm (Bird)	No	No	No	No
Proposition 63 (English)	No	No	No	No
Proposition 64 (AIDS)	No	No	No	No
Elections selected by defendants' experts				
1986 general election				
Proposition 65 (toxics)	No	No	No	No
1986 primary				
Proposition 51 (tort law)	Yes	Yes	Yes	Yes
1982 general elect				
Proposition 12 (nuclear freeze)	No	No	No	No
Proposition 15 (gun control)	No	No	No	No
1982 primary				
Governor (Bradley)	No	No	Yes	No

a. Based on definition provided by Stephen Klein: Polarized voting occurs when more than 50% of the Hispanics and more than 50% of the non-Hispanics vote differently for a candidate or proposition.

Proposition 51 right, showing a non-Hispanic/Hispanic split of but 51% to 44%. The correspondence between the neighborhood model and the nonpolarized elections proves nothing, since the model will always show minimal polarization in Los Angeles County elections. Ecological regression corresponded to exit poll results for all four polarized elections and in six of seven nonpolarized elections. Even for the remaining nonpolarized contest (1982 Democratic primary for governor), extreme case analysis failed to show polarization, providing the check on regression results that this technique is designed to achieve. No such disparity between ecological regression and extreme case analysis occurred in any of the 20 Hispanic versus non-Hispanic elections that Grofman and I analyzed for the *Garza* case.

There are some 30 elections for which exit polls are available from the 1980s. If ever there was a data set that could punch holes in ecological regression, this is it. There is not a single Hispanic versus Anglo contest in

the entire sample of elections, the vast majority of which showed no significant polarization between Hispanic and non-Hispanic voters. Yet with these 30 elections available, experts for the defendants failed to come up with a single example of where ecological regression and extreme case analysis reached a conclusion about ethnic polarization that is different from exit polls.

Freedman et al. also repeated in their article so-called reality tests of the neighborhood model and ecological regression presented at the trial. The experts presented the chimerical example of regressing votes for president on state-level percentages of Hispanics, when only the tiny state of New Mexico is more than one fifth Hispanic. They cited analyses of bloc voting in Stockton, California, where the problem was separating Black and Hispanic voting—a nonissue in Los Angeles County where these groups are not combined in the same precincts.

The remaining tests involved predictions of party registration, high school graduation rates, home ownership, and income for Hispanics and non-Hispanics. Again, the experts reported only the vote for Hispanics, so it is impossible to tell that none of these factors involved significant polarization between Hispanics and non-Hispanics. At most, the analysis shows what we already know: that there is an aggregate-level relationship between these factors and the Hispanic component of Los Angeles County precincts. The analysis says nothing about estimations of voter behavior.

Moreover, Freedman et al. omitted the one socioeconomic factor included in the census data set that does exhibit significant polarization: whether persons speak Spanish at home. There is an obvious reason for this omission. The census data show that none of the non-Hispanics and about 70% of the Hispanics speak Spanish at home. The neighborhood model gives the absurd result that 25% of Hispanics and 8% of non-Hispanics speak Spanish at home.

CONCLUSIONS

For multiple reasons, ecological regression has proved itself to be a reliable indicator of voting behavior, not just in Los Angeles County but in a great variety of political contexts. The technique is sensitive to actual voting behavior, not just the arrangement of voter groups into precincts. It yields accurate estimates of average group voting even when voting varies systematically from precinct to precinct. It will be affected by omitted variables only under special circumstances, and such bias is likely to be small.

In Los Angeles County and other jurisdictions throughout the United States, ecological regression and extreme case analysis have shown voting to be polarized along ethnic and racial lines. Experts hired by defendant governments have challenged these results, but they have developed neither a credible critique of the ecological regression and extreme case techniques nor reliable methods of their own for analyzing aggregate voting data. These defendants' experts have failed to find that voting is polarized along racial or ethnic lines not just in Los Angeles County but in the great cities of the North, the Delta of Mississippi, and the Black-belt counties of Alabama. Taken together, their results would paint a picture of an America that no one would recognize—because it does not exist.

As in Los Angeles County, experts' conclusions are generally contradicted not only by ecological regression and extreme case analysis but by the actual results of elections, the available exit poll data, and the results of elections in districts designed under the guidance of these methods. The vast majority of Black and Hispanic officeholders in the United States are elected in jurisdictions or districts with majority Black or Hispanic populations. Many of these districts were created as a result of voting rights litigation. The few Black victories in White districts or jurisdictions (e.g., Douglas Wilder's election as governor of Virginia) have received such publicity precisely because they are exceptions.

The findings of ecological regression are also generally corroborated by the available exit polls. In both 1984 and 1988, for example, exit polls in states throughout the country showed that the great majority of Black voters supported Jesse Jackson, whereas the great majority of White voters favored one of his White competitors in the Democratic presidential primaries. These findings are consistent with the results of ecological regression analyses within the states where Jackson competed.

Ecological regression has also proved itself to be remarkably accurate in predicting the ethnic composition of districts that provide minorities opportunities to elect candidates of their choice. Examples of jurisdictions where such predictions have been proven correct through the actual election of candidates preferred by minority groups include several—Pittsburgh, Pennsylvania; Springfield and Peoria, Illinois; and Watsonville and Los Angeles, California—in which a member of the *Garza* team of experts denied that ecological regression yielded any reliable information about voting behavior. Now Los Angeles County can be added to the list of examples in which ecological regression and extreme case analysis provided reliable guidance to a court of law.

NOTES

1. Relevant cases include the Supreme Court decision that endorsed ecological regression and extreme case analysis as the appropriate methods for racial bloc voting analysis: *Thornburg v. Gingles*, 478 U.S. 30 (1986). For other examples, see *United States v. Dallas County Commission*, 850 F.2nd 1433 (11th cir. 1988); *McNeil v. City of Springfield*, 658 F.Supp. 1015, App. C (C.D. Ill., 1987); and *Gomez v. City of Watsonville*, 863 F.2nd 1409 (9th Cir. 1988).

2. *Garza v. County of Los Angeles*, CV 88-5143 KN and CV 88-5435 KN (C.D. Cal. June 4, August 3, and August 6, 1990), *aff'd. in part, vacated in part and remanded* 918 F.2nd 763 (9th Cir. 1990), *cert. denied* 111 S.Ct. 691 (1991).

3. In other cases, experts for the defendants have raised frivolous arguments against ecological regression, for example, that weighting precincts according to population or using the two-equation method that adjusts for turnout somehow distorts results. No such arguments were made by the three statisticians testifying for the defendants in the *Garza* case. For an example of frivolous criticisms see Wildgen 1990. For persuasive responses, see Loewen and Grofman 1989, Engstrom 1990, and Loewen 1990.

4. If ethnically specific data are available for turnout in a particular contest, the two equations reduce to a single equation, with turnout as the denominator of both the independent and dependent variables.

5. Spanish-surname matching, however, may not fully capture the Hispanic component of a precinct because some registrants with a Spanish surname may not be of Spanish origin, and conversely, some registrants without a Spanish surname may be of Spanish origin. As a result, 1980 Census data were used to obtain the number of registrants with a Spanish origin in each precinct. The use of Spanish-surname versus Spanish-origin data does not, however, affect the substantive results. The commonality of results for the two sets of data provides additional corroboration of the reliability of the analyses of Hispanic and non-Hispanic voting. The analysis reported in Table 2 uses Spanish-surname data because Spanish-origin information was not available for every election studied.

6. The extreme case analysis uses 80%+ Hispanic precincts because of the paucity of more heavily Hispanic precincts in Congressional District 34 and other Los Angeles County districts.

7. The analysis excluded a few elections in which the losing Hispanic or non-Hispanic candidate(s) garnered less than 15% of the vote cast.

8. This analysis is based on methodology that I first developed in drawing a districting plan for Dallas County, Alabama (see Note 1).

9. *Garza* trial transcript, March 5, 1990, 115.

10. *Garza* trial transcript, March 13, 1990, 101-3, 111-12.

11. There is good reason, however, to be cautious about using exit polls for the analysis of racial bloc voting. Polls are expensive to conduct, are not generally available for past elections, and may understate polarization in elections where minority candidates are competing against Anglo candidates. See, for example, Loewen 1990.

12. Differences among supervisorial districts would not affect our use of ecological regression to project a defeat for Hispanic candidates in the districts created in 1981. First, we chose the supervisorial districts most favorable to Hispanic candidates (Districts 1 and 3). Second, from all elections in all types of districts studied, we chose the cohesion and crossover results most favorable to an Hispanic victory.

13. See also *Garza* trial transcript, August 3, 1990, 14.

14. See *Garza* deposition of Stephen Klein, February 26, 1990, 558-59.

15. We have avoided exact numerical comparisons with exit poll results for two reasons. First, the exit polls have relatively wide error bands, especially for elections other than the 1986 general election. Second, the great majority of the elections for which exit polls are available were not polarized between Hispanics and non-Hispanics. Given that it invariably finds minimal polarization in countywide elections, the neighborhood model will artificially appear to reflect the results of this sample of elections. What is critical, however, is the ability to distinguish between polarized and nonpolarized contests.

REFERENCES

Baum, D. 1984. *The Civil War party system: The case of Massachusetts*. Chapel Hill: University of North Carolina Press.

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

Engstrom, R. L. 1990. Getting the numbers right: A response to Wildgen. *Urban Lawyer* 22:495-502.

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

Gienapp, W. 1987. *The origins of the Republican Party, 1852-1856*. New York: Oxford University Press.

Grofman, B., and M. Migalski. 1988. Estimating the extent of racially polarized voting in multicandidate contests. *Sociological Methods & Research* 16:427-53.

Hanushek, E. A., J. E. Jackson, and J. F. Kain. 1974. Model specification and the ecological correlation fallacy. *Political Methodology* 1:87-106.

Kousser, J. M. 1974. *The shaping of Southern politics: Suffrage restriction and the establishment of the one-party South 1880-1910*. New Haven, CT: Yale University Press.

Langbein, L. I., and A. J. Lichtman. 1978. *Ecological inference*. Beverly Hills, CA: Sage.

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

———. 1979. *Prejudice and the old politics: The presidential election of 1928*. Chapel Hill: University of North Carolina Press.

Loewen, J. W. 1990. Sand in the bearings: Mistaken criticisms of ecological regression. *Urban Lawyer* 22:503-513.

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

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

Wildgen, J. K. 1990. Vote dilution and cold fusion technology. *Urban Lawyer* 22:489-502.

Allan J. Lichtman is Professor of History at American University. He was written six books and numerous articles, primarily addressing political history and quantitative methodology. He has been a consultant or expert witness for plaintiffs, defendants, and the Department of Justice in more than 40 voting rights cases.