ECOLOGICAL REGRESSION AND VOTING RIGHTS

by

D A Freedman, S Klein, J Sacks
C Smyth, and C. Everett

Prepared for Evaluation Review

Technical Report No 248
August 1990

Department of Statistics
University of California
Berkeley, California

Working Paper 90-23

Voting research
Minorities x Suffrage 2 California 2 Los Angeles
Hispanic American & " County

IGS
INSTITUTE OF
GOVERNMENTAL
STUDIES

UNIVERSITY OF CALIFORNIA AT BERKELEY

ECOLOGICAL REGRESSION AND VOTER RIGHTS

by

D A Freedman, S Klein, J Sacks, C Smyth, and C Everett

Prepared for Evaluation Review

Technical Report No 248
August 1990

Department of Statistics
University of California
Berkeley, California

Working Paper 90-23

August 1990

ECOLOGICAL REGRESSION AND VOTING RIGHTS

by

D A Freedman
Department of Statistics, UC Berkeley

S Klein
The RAND Corporation, Santa Monica

J Sacks, C Smyth
Department of Statistics
University of Illinois at Urbana-Champaign

C Everett
Applied Automated Engineering Corporation, Pennington, NJ

Abstract

Ecological regression is a statistical mainstay in litigation brought under the Voting Rights Act of 1965. The technique will be discussed in the context of a suit against the County of Los Angeles, which came to trial in 1989. Ecological regression depends on very strong assumptions about political behavior. We identify these assumptions and show they are not supported by the data. We also describe an alternative "neighborhood model", which is *a priori* more plausible and fits the data better. The neighborhood model leads to quite different conclusions about voting behavior.

Author's note

Authors consulted for McDermott, Will & Emery, the law firm representing the County of Los Angeles in Garza et al vs the County of Los Angeles et al (Federal District Court, Los Angeles). Freedman, Klein and Sacks gave expert testimony at trial. Smyth handled the data processing, and Everett did the graphics. We thank Larry Bazel, Dick Berk, Jerome Horowitz, John McDermott, John Rolph and Dan Rubinfeld for useful comments.

This paper is about the use of ecological regression in voting rights cases. The context is a law suit brought against the County of Los Angeles, which came to trial in 1989. The legal background will be summarized. Then statistical arguments about ecological regression will be discussed.

The Voting Rights Act, passed in 1965 and amended in 1982, prohibits any "voting qualification or prerequisite to voting, or standard, practice, or procedure...which results in a denial or abridgement of the right of any citizen of the United States to vote on account of race or color...." Under section 2 of the Act, discriminatory intent does not have to be proved; only results matter.

The courts have interpreted the statute quite broadly. To illustrate, suppose a city council has 5 seats and 20% of the electorate is black; most of the blacks live in one part of the city; blacks vote almost unanimously for black candidates; whites vote overwhelmingly against black candidates. In at-large elections, or in systems with district lines drawn so that the blacks are a minority in each district, no black candidate will be elected; and courts would find that blacks are disenfranchised. (In an at-large election, you can vote for any candidate; in a district system, you only get to vote for candidates from your district.)

Under such circumstances, a law suit could be filed to compel redistricting. At-large systems would be found illegal; the court would require electoral boundaries in a district system to be drawn so the blacks form a clear majority in one district. Then the black community can elect a candidate to city council.

The leading case is Thornburg v Gingles (106 S Ct 2752, 1986). The Supreme Court ruled that plaintiffs must prove three things in order to win their case:

> First, the minority group must be able to demonstrate that it is sufficiently large and compact to constitute a majority in a single-member district.... Second, the minority group must be able to show that it is politically cohesive.... Third, the minority must be able to demonstrate that the white majority votes sufficiently as a bloc to enable it usually to defeat the minority's preferred candidate.... [our italics]

The italicized phrases have become terms of art in the wave of litigation which followed. Perhaps to clarify the meaning of these terms (at the expense of converting other words into legalese), the Supreme Court continued:

Factual background

The focus is the 1980s. Los Angeles County is governed by a board of supervisors with 5 members. Each supervisor represents a district with about 1.5 million residents. There are other elected county officers, including a Sheriff and an Assessor. About one third of the county's population is hispanic and 15% is black. The blacks are heavily concentrated in supervisorial district #2.

Reapportionment in 1981 creates within the County of Los Angeles a number of heavily democratic-- and hispanic-- districts for Congress and the California State Legislature. Hispanic candidates are regularly elected to these seats, but no hispanics are elected to county office.

Plaintiffs contend that the supervisorial districts are drawn to fragment the hispanic community. They try to prove that the district lines could be drawn so hispanics would constitute a majority in one district-- meeting the compactness test in Gingles.

Their proposed districts must also meet the "one person one vote" rule. And in Los Angeles, the distinction between residents and voting-age citizens matters, so the definition of "person" is a major issue. Indeed, according to the 1980 Census, the population of the County is 28% hispanic overall. But the voting-age citizen population is only 15% hispanic: hispanics are younger than average, and less likely to be citizens.

Furthermore, self-reported citizenship rates in the 1980 Census are seriously biased upward, especially among hispanics; see Warren & Passel (1987), Passel & Woodrow (1984). Passel (1988) estimates that the voting-age population of Los Angeles County is only 12% hispanic in 1980, after correcting for reporting errors. On this basis, no redistricting plan can satisfy both the compactness test and the "one person one vote" rule: the hispanic voting-age citizens are too dispersed. (If most of the voting-age hispanic citizens were concentrated in a district that held 20% of the voting-age citizens, the compactness test would be satisfied, because 12% is more than half of 20%; but see Figure 1b below.)

The statistical problem

To demonstrate ethnically polarized voting, plaintiffs have to show that hispanics are politically cohesive and are outvoted by non-hispanics. This requires estimates of the numbers of votes cast by the two groups for each candidate. The statistical problem is created by the secret ballot: election data do not show how ethnic groups vote.

The number of votes for each candidate in each precinct can be obtained from public records; so can the number of registrants in each ethnic group (details are in the technical appendix). However, the number of votes for each candidate by each ethnic group is unknown, as indicated by the question marks in Table 1. In principle, to fill in these question marks, we need data on individuals: How did a registrant vote, and what is that registrant's ethnic group? Individual-level data may be available from exit polls or other surveys. But in Los Angeles, and in many other cases, survey data were not collected for the relevant elections. The question marks can be filled in only by using some imputation procedure.

The balance of the paper can be summarized as follows: Plaintiffs use ecological regression to fill in the missing data and to demonstrate racially polarized bloc voting. However, as we show, ecological regression depends on the "constancy assumption": that within each ethnic group, apart from random variation, the probability of turning out and voting for a candidate is the same in all precincts. Thus, ecological regression explains differences in voting behavior among precincts solely on the basis of ethnic makeup.

To show the force of the constancy assumption, we introduce an alternative: the "neighborhood model". Our model assumes that within each precinct, there is no systematic difference between the voting behavior of hispanics and non-hispanics. (In the aggregate across precincts, hispanics and non-hispanics may vote differently, because hispanics are more concentrated in some precincts than others and voting patterns differ from one precinct to another.) The neighborhood model is more plausible than ecological regression, and it fits the data better. The two models reach opposite conclusions about polarization.

Of course, plaintiffs had rebuttal to our points, and we had answers to their answers; part of the exchange is reported. Finally, there is a brief literature review, and our conclusions.

SUMMARY OF PLAINTIFFS' STATISTICAL ARGUMENT

To demonstrate ethnically polarized bloc voting, plaintiffs' experts selected 8 non-partisan elections for analysis, and 12 partisan elections. They argued that the support rates for hispanic candidates in these contests were much higher among hispanics than among non-hispanics. By definition, the hispanic support rate for a candidate equals the number of votes cast by hispanics for that candidate, divided by the total number of votes cast by hispanics; the non-hispanic support rate is defined similarly.

Plaintiffs used "ecological regression" to infer the breakdown of vote totals between ethnic groups. We illustrate the technique with a concrete example: the 1982 primary for Sheriff, where Feliciano was the major hispanic candidate. Figure 2 gives a scatter plot for the voting data. Each dot represents one precinct. The horizontal axis shows the percentage of registrants in the precinct who are hispanic. The vertical axis shows the "turnout rate" for Feliciano-- the percentage of registrants in the precinct who came to the polls and voted for Feliciano. (Los Angeles has about 6,500 voting precincts; to make the scatter diagrams more readable, every tenth precinct is plotted.)

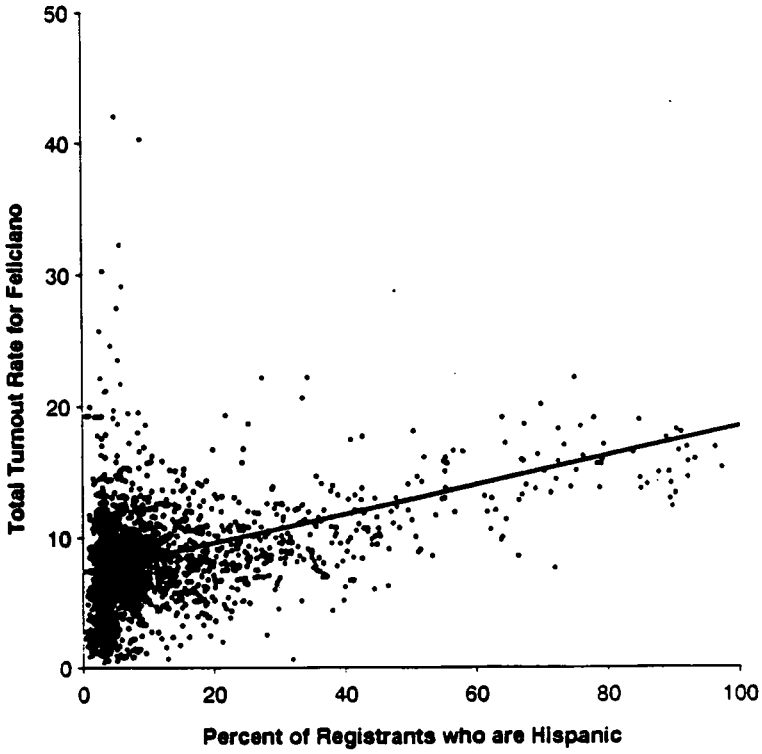

Figure 2: Turnout rates for Feliciano, the major hispanic candidate in the 1982 sheriff primary, Los Angeles County.

NOTE: The unit of analysis is the precinct. The regression line is shown.

The constancy assumption

In essence, ecological regression assumes that the turnout rate of 18.4% for Feliciano, predicted in a 100% hispanic precinct, applies to hispanics all across the board-- whether they are rich or poor, whether they are republican or democratic, whether they live in the suburbs or the barrio. Likewise, the turnout rate of 7.4% for Feliciano, predicted in a 100% non-hispanic precinct, applies to all non-hispanics no matter where they live. We dubbed this the "constancy assumption". Of course, the model does permit random variation around the predicted rates-- variation unrelated to the percentage of hispanics in the precinct (Figure 3).

In Figure 3a, the horizontal trend line for the turnout rates puts the constancy assumption for non-hispanics in graphical form. The line does not slope up or down because, by assumption, there is no systematic relationship between the non-hispanic turnout rate and the percentage of hispanic registrants in the precinct. Likewise for the hispanics in Figure 3b.

To state the ecological regression model in mathematical terms: There is an expected turnout rate by non-hispanics for Feliciano; call this α. Likewise, there is an expected turnout rate $\alpha+\beta$ by hispanics for Feliciano. Index the precincts by i. Let h_i be the number of hispanic registrants in precinct i, and n_i the number of non-hispanic registrants; so h_i+n_i is the total number of registrants in the precinct.

In precinct i, the turnout for Feliciano (from non-hispanics and hispanics combined) is expected to be

$$(2) \qquad \alpha n_i + (\alpha+\beta)h_i = \alpha(h_i+n_i) + \beta h_i$$

Let $f_i = h_i/(h_i+n_i)$, the fraction of registrants in precinct i who are hispanic. Divide (2) by h_i+n_i. The expected turnout rate for Feliciano in precinct i, among all registrants, is

$$(3) \qquad \alpha + \beta f_i$$

This expression is linear in f_i, taking the value α when $f_i=0$ and $\alpha+\beta$ when $f_i=1$.

Actual turnout rates vary from expectations, so random error must be added. The ecological regression model assumes that the turnout rate for Feliciano in precinct i is

$$(4) \qquad \alpha + \beta f_i + \epsilon_i$$

Let v_i be the total number of votes for Feliciano in precinct i; so v_i, h_i and n_i are known. If the constancy assumption holds, the parameters α and β in (4) can be estimated in a regression of $v_i/(h_i+n_i)$ on f_i, just as plaintiffs say. See (1), which is the regression equation expressed in percent.

Testing the constancy assumption

The constancy assumption is quite questionable, because hispanic registrants in highly hispanic precincts differ in many ways from hispanics in other precincts. The same is true for non-hispanics. The magnitude of these differences is large, and their direction will come as no surprise. In the heavily hispanic precincts, both hispanics and non-hispanics have lower incomes and educational levels; they are more likely to be renters rather than owners. Furthermore, as the precincts become more hispanic, hispanics and non-hispanics are less likely to vote; but they more likely to register as democrats. Data on party affiliation are shown in Figure 4, for hispanics and non-hispanics; the trend is very strong.

Table 2 shows how party affiliation, likelihood of voting, and demographic variables change with the percentage of hispanics in the precinct (or tract). Take party affiliation. The slope for hispanics is .358, so the percentage of hispanic registrants who are democrats increases by about 36 points-- from 54% to 90%-- as you go from precincts with few hispanics to precincts which are heavily hispanic. The change in the education variable is even more dramatic: among hispanic voting age citizens, the percentage of high school graduates drops from 70% to 24% as the census tract becomes more and more hispanic.

From polling data, demographic variables like income are known to be closely related to voting behavior: for results on California, see Table 3; and for national results, Gallup (1988). Party affiliation is also closely related to voting behavior, even in ostensibly non-partisan contests (Table 4). To sum up, hispanics in highly hispanic precincts are very different from hispanics in other precincts, along dimensions that are strongly related to voting behavior. Non-hispanics show the same pattern. The constancy assumption is not tenable.

Table 2: Regression statistics for characteristics of hispanics and non-hispanics, as a function of the percentage of hispanics in the precinct or tract.

Characteristic	Intercept	Slope	Correlation
Party affiliation			
hispanic	54.0	.358	.367
non-hispanic	50.4	.187	.174
Sign-ins			
hispanic	71.8	-.175	.278
non-hispanic	75.4	-.178	.330
Education			
hispanic	69.7	-.461	-.412
non-hispanic	80.3	-.333	-.399
Income			
hispanic	54.2	-.224	-.178
non-hispanic	55.0	-.288	-.249
Ownership			
hispanic	53.0	-.057	-.038
non-hispanic	57.2	-.052	-.036

NOTE: For the first two characteristics, the data apply to registrants in the 1988 general election in Los Angeles. The unit of analysis is the precinct. The independent variable is the percent of registrants in the precinct who are hispanic. The dependent variables are as follows:

percent of hispanic registrants who are democrats
percent of non-hispanic registrants who are democrats
percent of hispanic registrants who signed in to vote
percent of non-hispanic registrants who signed in to vote

For the last three characteristics, the data apply to voting-age citizens in the 1980 Census in Los Angeles. The unit of analysis is the tract. The independent variable is the percent of voting-age citizens in the tract who are hispanic. The dependent variables are as follows:

percent of hispanic voting-age citizens who have 12 or more years of education
percent of non-hispanic voting-age citizens who have 12 or more years of education
percent of hispanic voting-age citizens who have household incomes of $20,000 per year or more
percent of non-hispanic voting-age citizens who have household incomes of $20,000 per year or more
percent of hispanic voting-age citizens who live in owner-occupied housing
percent of non-hispanic voting-age citizens who live in owner-occupied housing

The neighborhood model

How much difference do the assumptions make? In an area with highly segregated housing, the constancy assumption makes little difference to the empirical results. In Los Angeles, most hispanic voters live in areas where they are a minority, so modeling assumptions have a lot of leverage. To make this vivid, we introduced an alternative assumption about voting behavior. Within a precinct, our "neighborhood model" assumes that there are no systematic differences between hispanic and non-hispanic voting behavior. Informally, people who live near each other are like each other with respect to income, education, housing--and political opinions.

To state our model in mathematical terms: Index the precincts by i, as before. Let h_i be the number of hispanic registrants in precinct i, and n_i the number of non-hispanic registrants. Within precinct i, the turnout rate ρ_i for a particular candidate is expected to be the same for the hispanics and non-hispanics (with random error terms for each precinct and ethnic group). Therefore, the expected number of hispanic votes for the candidate in precinct i is $\rho_i h_i$; the number of non-hispanic votes, $\rho_i n_i$. The total number of hispanic votes for the candidate is obtained by summing $\rho_i h_i$ across precincts; likewise for non-hispanics.

Two versions of the neighborhood model were considered, non-linear and linear. The non-linear model allows the turnout rate for a candidate to vary freely from precinct to precinct. Then ρ_i is estimated as $v_i/(h_i+n_i)$, where v_i is the total number of votes for the candidate in precinct i.

In the linear model, as the name suggests, ρ_i is assumed to be a linear function of $f_i = h_i/(h_i+n_i)$, allowing for random variation:

$$\rho_i = \alpha + \beta f_i + \delta_i \qquad (5)$$

The δ_is are assumed to be independent, with mean 0. Then $\hat{\alpha}$ and $\hat{\beta}$ are obtained by the regression of $v_i/(h_i+n_i)$ on f_i, and ρ_i is estimated by $\hat{\alpha}+\hat{\beta}f_i$, its fitted value. In particular, the linear neighborhood model fits the same line to the same data as ecological regression; but it uses the coefficients quite differently.

Table 5: Estimated number of votes for Feliciano by 3 models, in a precinct with 50 hispanic registrants and 450 non-hispanics.

	Neighborhood model		Ecological regression
	non-linear	linear	
By hispanics	4	4.25	9.20
By non-hispanics	36	38.25	33.30
Total for Feliciano	40	42.50	42.50

NOTE: The regression line of turnout rate for Feliciano on percent hispanic has intercept 7.4% and slope .11. The actual number of votes for Feliciano is 40. With the non-linear neighborhood model, the predicted total vote for Feliciano must equal the actual figure. The linear model and the ecological regression model agree on the predicted total for Feliciano; the difference from actual represents the deviation from the regression line.

Comparing the models

Ecological regression assumes that any systematic differences between precincts are explained by differences in ethnic composition; no other variable matters. The neighborhood model, on the other hand, assumes that differences in voting behavior are due to precinct effects. Of course, in total (across precincts), the neighborhood model allows the two ethnic groups to have quite different voting behavior.

Table 6 compares the estimates from the ecological regression model and the neighborhood model, for the partisan elections considered by plaintiffs (the results for the non-partisan elections were similar). As the table demonstrates, it is the modeling assumptions which drive the estimates. Generally, the neighborhood model indicates far less polarization than does the ecological regression model.

Table 6: Estimates of hispanic and non-hispanic support rates for major hispanic candidates in 12 partisan elections in Los Angeles County. (These support rates are not observable.)

	Ecological Regression		Neighborhood Model	
Contest	Hispanic	Non-Hispanic	Hispanic	Non-Hispanic
1982 Democratic Primary				
US Congress CD 30				
M Martinez	42	29	36	34
O Moreno	35	3	20	16
US Congress CD 34				
E Torres	88	27	57	47
State Assembly AD 59				
C Calderon	80	28	55	49
1982 General				
US Congress CD 30				
M Martinez	106	33	61	52
US Congress CD 34				
E Torres	104	38	65	55
State Senate SD 24				
A Torres	101	50	84	64
State Assembly AD 52				
M Reza	231	19	39	34
State Assembly AD 56				
G Molina	94	62	88	77
1984 Democratic Primary				
State Assembly AD 63				
L Escontrias	44	-5	8	3
C Fuentes	32	11	16	14
1986 State Assembly AD 55				
Special Election				
R Polanco	84	46	65	53
General Election				
R Polanco	78	53	66	58
1988 Republican Primary				
US Congress CD 30				
R Ramirez	59	44	46	45

NOTE: The 231% hispanic support rate for Reza is not a typographical error; ecological regression often produces physically impossible results.

The regression line rises from 50.1% at the left end to 80.5% at the right. So, ecological regression would estimate that 50.1% of the non-hispanics and 80.5% of the hispanics are democrats. The actual figures are 53.5% and 66.8%: the technique over-estimates the difference between hispanics and non-hispanics, by a substantial margin. The neighborhood model does much better; it estimates 54.5% and 61.0% for the non-hispanics and hispanics, respectively.

Results for other variables are reported in Table 7. For ownership, the slopes of the trend lines are nearly 0 (Table 2). However, these slopes are enough to cause an appreciable bias in the ecological regression estimates, as shown in Table 7. (The ecological regression model and the neighborhood model will also be compared using exit poll data, in Tables 9 and 10.)

In the ecological regression model, the only systematic differences between precincts are due to differences in ethnic makeup. This is contrary to fact. In the heavily hispanic precincts, the hispanics are much more likely to register as democrats-- and so are the non-hispanics. Differences within an ethnic group across precincts bias the ecological regression estimates. That is, the slopes in the trend lines of Figure 4 cause bias in the ecological regression estimates. Of course, the two trend lines are different. And that biases the estimates from the neighborhood model. However, at least in this application, the neighborhood model seems to be more robust.

The known association between voting behavior and variables such as party affiliation or income strongly suggests that the turnout rate for a candidate by each ethnic group is more like Figure 4 than Figure 3: voting behavior of each ethnic group in a precinct is strongly related to the percentage of hispanics in that precinct. If so, the neighborhood model will give more accurate support estimates than ecological regression (Table 7) and there is little difference between hispanic and non-hispanic voting patterns (Table 6).

SUMMARY AND CONCLUSIONS

Ecological regression may be useful, if the constancy assumption is close to right. However, this assumption is a priori quite unlikely; and the neighborhood model seems more plausible, at least in the present case.

How much do assumptions matter? In voting rights cases, they matter a lot. As Table 6 shows, the ecological regression model and the neighborhood model gave results with opposite policy conclusions. The first model said that voting was highly polarized; the second, that hispanics and non-hispanics had similar political preferences.

With aggregate data-- total votes for each candidate by precinct, and total registration by ethnic group and precinct (Table 1)-- neither model can be tested at all seriously. Therefore, such data cannot determine whether voting was polarized or not. No other data were available on the crucial elections in Los Angeles.

Other kinds of data were examined (Tables 2-4 and 7-9): party affiliation, census characteristics, exit polls from other elections, etc. On balance, we consider this evidence to weigh heavily against the constancy assumption: hispanics who live in highly hispanic areas are very different from other hispanics, with respect to many characteristics related to voting behavior. By comparison, the neighborhood model looks good: hispanics and non-hispanics who live near each other share characteristics related to voting behavior.

This evidence bears on the contests at issue in Los Angeles, but the connection is not direct. The data to make a direct assessment were not available. For contests where an evaluation was possible, the neighborhood model missed in some cases and ecological regression in others (Table 10). Even when ecological regression was on target for the county as a whole, it failed to detect large differences between supervisorial districts (Table 11).

In the end, we think surveys offer a better approach to estimating voting behavior than models, because surveys start from data on individuals, and the data can be validated. Indeed, without survey data, there may be no reliable answers to questions about ethnic voting behavior. Like anything else, modeling has limits.

REFERENCES

Broder, I. and A. Lichtman, 1983. Modeling the past: the specification of functional form. Interdisciplinary History 13:489-502.

Daily Appellate Report, 1990. Yolanda Garza et al v County of Los Angeles et al. 90 Daily Journal DAR 6141.

Duncan, O.D. and B. Davis, 1953. An alternative to ecological correlation. American Sociological Review 18:665-6.

Frisbie, W.P. and J.D. Kasarda, 1988. Spatial processes. In N.J. Smelser, ed. Handbook of Sociology. Sage. pp629-66.

Gallup, G. Jr. 1988. The Gallup Poll. Scholarly Resources, Wilmington.

Goodman, L. 1953. Ecological regression and the behavior of individuals. American Sociological Review 18:663-4.

Goodman, L. 1959. Some alternatives to ecological correlation. American Journal of Sociology 64:610-25.

Grofman, B. 1990. Multivariate methods and the analysis of racially polarized voting: pitfalls in the use of social science by the courts. Technical report, School of Social Sciences, UC Irvine.

Grofman, B. 1987. Models of voting. Research in Micropolitics 2:31-61.

Grofman, B., M. Migalski and N. Noviello, 1985. The "totality of circumstances test" in section 2 of the 1982 extension of the Voting Rights Act: A social science perspective. Law and Policy 7:199-223.

Heer, D.M. and J.S. Passel, 1986. Comparison of two methods for estimating the number of undocumented Mexican adults in Los Angeles County. Technical report, Population Research Laboratory, USC.

Irwin, L. and A. Lichtman, 1976. Across the great divide: inferring individual level behavior from aggregate data. Political Methodology 3:411-39.

Iversen, G.R. 1974. Recovering individual data in the presence of group and individual effects. American Journal of Sociology 79:420-34.

TECHNICAL APPENDIX

Data sources

The 1980 Census long form was given to one household in five, approximately. This form gives information on age, race, citizenship, hispanic origin, and spanish surname status. Generally, there are more persons of hispanic origin than of spanish surname.

Only spanish surname status can be determined from registration data. Plaintiffs imputed hispanic origin from spanish surname status, using conditional probabilities derived from tract-level 1980 Census data. This increased the number of hispanic persons-- which was vital for the compactness argument-- at the expense of reducing the polarization estimates.

Defendants contested the validity of the imputation; the arguments back and forth are not reported here. For Figure 1, defendants used similar imputation procedures to estimate the numbers of black registrants in each precinct; plaintiffs made no objections.

More specifically, the number of black registrants in a precinct was estimated as sp+tq. Here, s is the number of registrants in the precinct who have a spanish surname; p is the fraction of spanish surname voting age citizens who are black (so sp is the number of spanish surname blacks); t is the number who have a non-spanish surname; q is the fraction of non-spanish surname voting age citizens who are black (so tq is the number of non-spanish surname blacks). The numbers s and t come from 1988 registration data; p and q come from 1980 tract level Census data.

The models

It may be useful to give a compact, mathematical description of the three models and their estimation procedures (Tables (A1-2). As before, index the precincts by i. Let h_i be the number of hispanic registrants in precinct i, and n_i the number of non-hispanic registrants. So $f_i = h_i / (h_i + n_i)$ is the fraction of registrants in precinct i who are hispanic. Fix one candidate; let v_i be the number of votes for this candidate in precinct i. Then h_i, n_i, and v_i are known.

Let π_i be the hispanic turnout rate for the candidate in the precinct, and θ_i the non-hispanic turnout rate. (For example, π_i is the number of hispanic votes for the candidate, divided by the number of hispanic registrants.) These are not observable, although $h_i \pi_i + n_i \theta_i = v_i$. The three models make different assumptions about π_i and θ_i, expressed in Table A1.

Computer simulations

We continue with previous notation. Suppose there are two candidates, one being hispanic and the other non-hispanic. Suppose their respective turnout rates in precinct i are p_i and q_i .

A simulation in favor of ecological regression.

With a small change of notation in the ecological regression model, the hispanic turnout rate for the hispanic candidate in precinct i is $a+b+\zeta_i$; the non-hispanic turnout rate is $a+\xi_i$. The overall turnout rate for the hispanic candidate in precinct i is therefore

$$(A1) \qquad p_i = a+bf_i+U_i ,$$

where $U_i = f_i\zeta_i + (1-f_i)\xi_i$. We chose the ζ's and ξ's as independent normal random variables, with mean 0 and variances σ^2/h_i and σ^2/n_i respectively. Thus, var U_i = $\sigma^2/(h_i+n_i)$. Likewise, the overall turnout rate for the non-hispanic candidate in precinct i is

$$(A2) \qquad q_i = c+df_i+V_i$$

Here, V_i is a normal random variable with mean 0 and variance $\tau^2/(h_i+n_i)$. (Assuming constant variances would not change the results in any noticeable way.) We constrained the rates to fall between 0 and 1.

The total turnout rate in precinct i, for the two candidates combined, is

$$(A3) \qquad p_i+q_i = (a+c)+(b+d)f_i+U_i+V_i$$

In the ecological regression model, the hispanic support rate for the hispanic candidate is (a+b)/(a+b+c+d); the non-hispanic support rate for this candidate is a/(a+c).

We choose a,b,c,d to give a total turnout line which looks reasonable, the support rate for "the hispanic candidate" being 80% among the hispanics and 20% among the non-hispanics:

$$a = .11 \qquad b = .29$$
$$c = .44 \qquad d = -.34$$

We choose σ^2 and τ^2 to produce reasonable-looking spread around the regression lines:

$$\sigma^2 = 4 \qquad \tau^2 = 4$$

The "double-equation" method involves two regressions:

$$(A4) \qquad v_i / N_i = \hat{a} + \hat{b} f_i + e_i$$

$$(A5) \qquad V_i / N_i = \hat{A} + \hat{B} f_i + E_i$$

Here, e_i and E_i are residuals. The non-hispanic support rate is estimated as $\hat{a}/\hat{A}$. Similarly, the hispanic support rate is estimated as $(\hat{a}+\hat{b})/(\hat{A}+\hat{B})$.

The "single-equation" method just involves one equation, like (A4); but the denominator of the left-hand side variable is votes rather than registration:

$$(A6) \qquad v_i / V_i = \hat{c} + \hat{d} f_i + g_i$$

Again, g_i is a residual; $\hat{c}$ is interpreted as the non-hispanic support rate, and $\hat{c}+\hat{d}$ is the hispanic support rate.

If hispanics and non-hispanics turn out at the same rate ($\hat{B} = 0$), there is no difference between the two methods. Usually, hispanics turn out at a lower rate-- that is, a smaller percentage of the hispanic than of the non-hispanic registrants actually vote. Then, the two methods differ. Generally, the double-equation method will estimate more polarization, while the single-equation method will produce a higher R^2.

Plaintiffs' experts usually estimate polarization by the double-equation method, on the grounds that it adjusts for differences in turnout rates. (In Los Angeles, controlling for precinct of residence virtually eliminated the difference in turnout rates.) However, these experts also tend to report the R^2s from the single-equation method-- an unusual combination by ordinary statistical standards.

89-23 **Learning From Risk: Organizational Interaction Following the Armenian Earthquakes,** Louise K. Comfort

89-22 **The Elusiveness of Rural Development Theory and Practice: Domestic and Third World Perspectives Joined,** Ted K. Bradshaw

89-21 **Saints and Cardinals in Appropriations Subcommittees: Academic Pork Barreling and Distributive Politics in an Era of Redistributive Budgeting,** James D. Savage

89-20 **The Case for Experiential Knowledge,** Gene I. Rochlin

89-19 **Choice vs. Control: Increasing Organizational Effectiveness in Interdependent Environments,** Louise K. Comfort and Keun Namkoong

89-18 **Bingo! An Untapped Revenue for California Cities,** William B. Rumford, Jr. and Randy H. Hamilton

89-17 **The Research on Higher Education Program: An Appreciation of Eskil Bjorklund,** Martin Trow

89-16 **Santa Cruz County Planning Issues: Papers on Planning, Housing And Forestry,** Edward J. Blakely and Ted K. Bradshaw

89-15 **Why Pretend One Size Fits All: An Examination of Management Issues That Concern Small Federal Agencies,** Randy H. Hamilton

89-14 **Music of the Squares, a Lifetime of Study of Public Administration,** Herbert Kaufman

89-13 **Technology and the Future: Issues Before the Bush Administration,** Edward Wenk, Jr.

89-12 **How Much Does Law Matter? Labor Relations in Rotterdam and U.S. Ports,** Robert A. Kagan

89-11 **Party, State, and Ideology in the U.S. House of Representatives, 1967-76,** K. G. Armstrong

89-10 **Ernst Fraenkel Lecture, Free University of Berlin the American Election of 1988: Outcome, Process and Aftermath,** Nelson W. Polsby

89-9 **The Ambiguous Status of Science and Technology in Australia,** Anthony Pecotich and Kelvin Willoughby

89-8 **1992, European Integration and *The Times*,** David Morgan

89-7 **American Higher Education: "Exceptional" or Just Different?** Martin Trow

89-6 **American Higher Education: Past, Present and Future,** Martin Trow

89-5 **The Distribution of Academic Earmarks in the Federal Government's Appropriations Bills, FY 1980-1989,** James Savage

89-4 **London 2001,** Peter Hall

89-3 **The University of London: An American Perspective,** Eugene C. Lee, Frank M. Bowen

89-2 **Ukiah, 1904: A Modest Footnote to the History of the Council-Manager Form of Municipal Government in the United States,** Randy H. Hamilton

89-1 **American Identity and the Politics of Ethnic Change,** Jack Citrin, Beth Reingold, Donald P. Green

1988

88-27 **Locality and Custom: Non-Aboriginal Claims to Customary Usufructuary Rights as a Source of Rural Protest,** Louise Fortmann

88-26 **Bruce Keith's Almanac: Patterns of Voting in California,** Bruce Keith

88-25 **Cold Turkeys and Task Force: Pursuing High Reliability in California's Central Valley,** Todd R. La Porte and Ted Lasher

88-24 **Environmental Ethics in California,** Carolyn Merchant

88-23 **Crisis as Opportunity: Designing Networks of Organizational Action in Disaster Environments,** Louise K. Comfort

88-22 **The Logic of Uncertainty: Interorganizational Coordination in International Disaster Assistance,** Louise K. Comfort

88-21 **Information Networks in International Disaster Assistance,** Louise K. Comfort

88-20 **The Decay of Federal Theory,** S. Rufus Davis

88-19 **Inside Japan's Leviathan Decision-Making in the Government Bureaucracy,** Brian Woodall and Nobuhiro Hiwatari

88-18 **Technology and Adaptive Hierarchy: Formal and Informal Organization for Flight Operations in the U.S. Navy,** Gene I. Rochlin and Energy Resources Group

SEND ORDER TO:
Institute of Governmental Studies
102 Moses Hall
University of California
Berkeley, CA 94720
(415) 642-5537

PLEASE PRE-PAY ALL ORDERS
UNDER $30: checks payable to The Regents of the University of California.
SALES TAX: California residents add sales tax.
HANDLING AND SHIPPING: add 20% of sales price. Allow 4 weeks for delivery.

Please add my name to the *PAR* mailing list (free). ____

TITLE	QUANTITY/COST

SUBTOTAL _____
SALES TAX _____
HANDLING (20%) _____
TOTAL _____

NAME _____
ADDRESS _____
_____
CITY _____ STATE _____ ZIP _____