

Chapter 6 – Voting Rights and the Ecological Inference Problem

The 15th Amendment began the century as little more than a parchment guarantee for African Americans. The wording seems clear enough: “The right of citizens of the United States to vote shall not be denied or abridged by the United States or by any state on account of race, color, or previous condition of servitude.” Even clearer, though, is that the Jim Crow South permitted very few African Americans to enter the polls—a disenfranchisement implemented not directly by the state as a matter of policy, but through the superficially constitutional delegation of authority to institutions that could discriminate on the state’s behalf.¹

Studying whites of the Solid South, therefore, was not difficult. The Southern electorate *was* the white Southern electorate, and race (at least in the Black Belt) was almost a single-issue obsession. V.O. Key could analyze 1928 voting returns for their political or sociological importance without the slightest concern that black voters were contaminating his sample. But progress toward racial justice has added minorities to the turnout mix, and progress toward racial tolerance has added alternate issue concerns to the voting mix. Social scientists no longer enjoy the luxury of letting election returns speak for themselves.

Contemporary racial politics, by which I mean the way political affairs (especially elections) mirror racial divisions, is submerged within the larger give and take of electoral politics. We have progressed enough as a society that race is difficult to study, but not enough that it is no longer worth the effort. This is a methodological barrier, not a theoretical one. What we wish to know is still how

¹ Not that defenders of the former slaves missed this sleight of hand. Supreme Court Justice Harlan, in his *Civil Rights Cases* dissent, warned that the greatest danger to civil equality came “not altogether from unfriendly state legislation, but from the hostile action of corporations and individuals in the states” (Yarbrough 1995, 150).

individual Southern whites respond to the racial mix of their community—psychologically, attitudinally, and politically. The difficulty is figuring out how to measure that reaction, how to focus on racially based variation, how to parse out groups whose behavior naturally differs.

Solving the confusion of issues is easier, although imperfect. Occasionally a candidate or an election comes along that polarizes voters. Perhaps one of the candidates is African American, and racially sensitive white voters are likely to defect. Or one candidate engages in blatant race baiting, attracting the racist vote and discouraging other potential supporters displeased with such a divisive message. Or the candidates split clearly on racial policies but otherwise appear fairly similar. Not every voter chooses how to vote based upon race, of course, even in these extraordinary cases, but the election should capture politically salient racial polarization. If we can figure out white preferences in these sorts of elections, we'll have in hand a proper variable for analysis. Repeated applications, over different sorts of polarized elections conducted at different times and in different places, should allow social scientists to triangulate on the truths of racial politics. It is, at any rate, as close as the contemporary political system gets.²

Electoral data are the obvious choice for studying racial politics. They are, after all, tangible expressions of political preference—the distillate of prejudices, attitudes and partisan inclinations rather than the raw material. Their geographical richness is unmatched, and the costs usually are limited to a small fee provided to a state agency. Unfortunately, analysts cannot break down electoral choice directly by race (or any other demographic characteristic of interest). The vote is protected by use of secret ballots; states cannot report electoral returns at the individual level. All we have are election returns aggregated over geographical units, and common techniques of aggregate-data analysis are unreliable.

² Another solution, of course, would be to study racial issue preferences directly. However, the electoral system is a complex one; issue preferences do not translate cleanly into political activity. To explore racial attitudes would shift the focus away from racial *politics* as I wish to study them (see Chapter 4). It also would require the use of survey data, which I wish to avoid for the bulk of my analysis (but see Part III).

The remainder of this chapter takes on two tasks, both intended to connect the general ecological inference problem to the substance of Southern politics. First, I offer a brief sketch of how emerging African-American voting rights derailed geographic studies of Southern racial voting. It is this project, temporarily abandoned by political science, that I wish to pick up again. Second, I present the ecological-inference problem as it appears in Southern voting data more formally, and explain why until recently existing methods could not make Southern election returns tractable for study. The following chapter, however, shows that Gary King's "solution to the ecological inference problem" (King 1997), as implemented by the software packages EI and EzI, allows more accurate estimates of racial voting than ever before. King's EI is extremely successful at overcoming difficulties in precisely the sort of data required for studying Southern elections.

The Rise and Fall of Southern Voting Studies

The popular image of Reconstruction is that Rutherford B. Hayes betrayed Southern blacks in exchange for a bartered presidential victory against Samuel Tilden. The "Compromise of 1877" certainly resulted in the collapse of Southern Republican governments, so the image contains some truth. On the other hand, black freedoms eroded much more slowly. As long as African Americans exercised the franchise, they could ally with sympathetic factions in Southern politics to defend their rights. The erosion of voting rights took more than a generation to accomplish, and even then operated within the bounds of American democratic institutions and laws.

The "white primary" illustrates how Southern states could produce discriminatory results without requiring overtly discriminatory laws. Political parties bore the responsibility to select their nominees for office. Since parties were not an arm of the government, they did not fall under the 15th Amendment. They could exclude from membership, and therefore from voting in party primaries, whomever they wished. The Democratic Party in the South, not surprisingly, chose to

exclude African Americans.

In no way did the white primary deny blacks their “voting rights” as commonly understood. They were free, in theory, to form or join political parties, to offer their own candidates, and to vote in the general election. Problem was, their candidates would never win, because the white electoral majority consistently voted Democratic nominees into office. It was this cohesion, combined with discriminatory party rules, that denied blacks political influence. The practice continued until the U.S. Supreme Court, in *Smith v. Allright* (1944), declared political parties an extension of the state when their primaries were “tantamount to election.”³

Other procedures chipped away at 15th Amendment rights as well. Many states charged a poll tax before granting the right to vote. Only someone with money, after all, possessed enough stake in society to exercise their vote responsibly. The indigent might sell their franchise too cheaply. The law applied to whites as well as blacks, and therefore was not overtly discriminatory, but everyone knew who would have the hardest time scraping together the cash. Black sharecroppers, even when financially secure, tended not to have liquid assets. Also, potential voters would have to keep track of their tax receipts, and proffer them on election day. Farm laborers were not accustomed to this sort of paperwork.

Many states required voters to be literate. Sometimes the literacy test would require interpreting portions of baroque Southern state constitutions. Only a literate citizen with an understanding of political affairs, after all, could read campaign literature or newspaper reports, could weigh the issues intelligently, and could read referenda on the ballot.⁴ The law applied to

³ *Smith v. Allright* 321 U.S. 649 (1944).

⁴ Note the peculiar mixture of tyranny and Progressive politics undergirding disenfranchisement. A politics based upon barbecues, parades and uniforms, as described artfully by McGerr (1986), would not lend itself so easily to the Jim Crow logic. But the condescending elitism that infused the Progressive movement fit snugly with Southern racism. Indeed, they cross-fertilized. The party primary—a linchpin of the era’s electoral reforms—was a Southern innovation (Price 1957).

whites as well as blacks, and therefore was not overtly discriminatory, but everyone knew who faced the greatest barriers getting access to educational resources. Anyway, someone had to administer the test. Invariably this would be a white registrar of voters, drawn from the local community and not particularly interested in fair application of the law (Rodgers and Bullock 1972). The behavior of individual officials, even when acting under their formal authority, could not amount to unconstitutional state *policy* then.

Many states instituted long residency requirements. Only a citizen who has spent some time in a community, after all, can evaluate candidates and local issues there. If strangers may show up soon before an election and still vote, fraud becomes too easy. The law applied to whites as well as blacks, and therefore was not overtly discriminatory, but everyone knew who moved frequently from county to county, following work as it became available.

Should these methods fail, such as in a majority-black county where the white primary would not guarantee victory in a general election, extralegal methods might come into play. Economic pressures, intimidation by law-enforcement officers at the polls, and the Ku Klux Klan's brand of terrorism were particularly successful. Many Black Belt counties consistently reported zero black registration. State governments did not countenance any of these activities—they looked the other way entirely.

In sum, the black vote was not particularly formidable at the turn of the 20th century.

Razing the Jim Crow System of Disenfranchisement

The structure of disenfranchisement was an intricate one. State government formed one major pillar, by enacting the laws or enforcing the constitutional provisions that allowed voting discrimination. But there were other pillars. One was the Southern white voter—especially the Southern elite—who remained faithful to the party of white supremacy. Few pressured politicians,

sheriffs or registrars to loosen the caste system, and then usually only as a preface to migrating North.

Another culprit was the white North. Northern elites—besieged by immigrants, hungry for colonies inconveniently populated by “lesser” races, and profiting from exploitation of the Southern labor force—permitted and often indirectly abetted the Jim Crow racial code.⁵ The U.S. Supreme Court upheld the constitutional loopholes upon which the system relied. Congress, meanwhile, did not pass national legislation that would put teeth into rights the Constitution did enumerate. GOP dominance was so complete that politicians had little use for, and little interest in pursuing, the votes of Southern black Republicans.

This pillar fell out with World War II. The U.S. replaced colonization with more subtle forms of imperialism, and Southern institutions became an international embarrassment. Black migration northward, to fill wartime industries, had produced an electoral faction that Northern politicians could attract with the relatively cheap promise to reform someone else’s institutions. Constitutional law also changed, laying the groundwork for national intervention in state laws and community mores. Instead of tolerating Southern practices, the Supreme Court began chipping away at other parts of the Jim Crow system (Seagull 1975, 3).

Meanwhile, social changes within the South corroded the foundations. Valiant black soldiers, returning from war against an evil enemy and expecting treatment worthy of their heroic patriotism, instead found their ambitions thwarted by a rather familiar-looking wickedness (Potter, Miles and Rosenblum 1992). Some were willing to continue battling the enemy, and thus empowered their communities (Terkel 1984, 340, 565). The Southern black population also became increasingly urbanized. Black political activity was harder to suppress in the city (Salamon and Van Evera 1973,

⁵ Richard Bensel’s recent work argues convincingly that, because capitalist exploitation of the agricultural sector was mediated through sectional politics, U.S. democratic institutions survived industrialization in relatively pristine shape (Bensel 1998, 25-29).

1294). African Americans gathered significant economic resources in a booming economy, and therefore enjoyed greater independence as well as greater muscle for asserting their rights. Some African-American professionals, because they served a uniformly black clientele, could afford to fight for justice without fear of economic retribution.

Cracks began to appear in the disenfranchisement system during the 1940's, grew during the 1950's, and reduced almost the entire system to rubble with the Voting Rights Act of 1965. Early in the process, African-American participation was more a function of white tolerance than of black interest in politics. Yet African Americans did vote in the 1940's and 1950's, especially in urban areas.⁶ By 1964, 2 million African Americans were registered to vote in the South (Salamon and Van Evera 1973, 1290-91). The level of activism responded to more than just variations in white tolerance; it also fluctuated with white power resources and the uneven motivation provided by local black leadership (Price 1957, 54). Thus even the earliest students of postwar Southern voting were troubled by the unmeasured black vote (Cosman 1966, 65; Heard 1952, 251-52; Pettigrew and Campbell 1960, 444), or at least should have been, especially if they wanted conclusions that could generalize to the South's growing urban areas (Strong 1977, 33-34).

The black franchise widened, sporadically and unsteadily, then surged after the Voting Rights Act of 1965 (Alt 1995). The main beneficiary of this expansion, ironically, was the party of white supremacy. Young black voters showed much less interest in the party of Lincoln than they did the party of Roosevelt and Truman, which occasionally worked to expand civil rights and also seemed

⁶ Some researchers link tolerance to the "cosmopolitanism" of an urban environment. To the extent the city represents diversity, their argument is an implicit rejection of the white backlash logic (ironic since awareness of this urban tolerance sometimes appears in work studying the backlash phenomenon). But social life differs so much from city to country—including occupational categories, consumption patterns, the structure of social relations, etc.—that many other sources might explain different levels of tolerance. Metropolitan areas may be so atomized, for example, that group distinctions are useless. Everyone is a stranger, not just members of other castes. This theme often appears in the comments of African Americans yearning for, or returning to, Southern communities after living in a Northern metropolis. Bluesman Son Seals sings, in "Going Back Home" (1976), "I didn't think a city could be so doggone mean. But this is the meanest place, Lord, I've ever seen." A city "too busy to hate" may be a city too busy to care as well.

a closer fit to their economic interests. Huckfeldt and Kohfeld (1989, 15) estimate that the Democrats, in 1964, mobilized a third of African Americans who had not voted in the previous presidential election. Therefore, after the VRA black votes and white votes were falling in behind some of the same candidates.

Once federal marshals began registering African Americans in even the most recalcitrant Southern counties, voting data in heavily black areas flipped in meaning entirely—from the best gauge of anti-black sentiment to the best gauge of black sentiment (Birdsall 1969, 58-59). Black and Black (1973) illustrate this methodological difficulty in their study of George Wallace's Alabama voting support, an analysis that includes elections from before and after the franchise expanded. They find that the effect of black density reverses around 1962; racially diverse counties that once provided Wallace his strongest support suddenly became his Achilles Heel.

Furthermore, attempts to parse out white voting behavior using ecological analysis fell prey to the white-backlash pattern itself.⁷ Because rising black power in mixed counties ran up against segregationist voting in those same places, they counteracted each other, such that statistical solutions for parsing out the white vote were not reliable. An analysis by Schoenberger and Segal (1971) implied that Wallace received heavy black support in his 1968 presidential bid, since election returns were especially favorable where blacks were most populous. They did not embrace this absurd conclusion, attributing the findings to Wallace's heavy white support in the Black Belt (1971, 585), but lacked any statistical basis for the claim or any real estimates of white behavior. Not even additional methodological refinements, intended to counteract the "ecological fallacy," could remove their reliance on prior knowledge, or provide reliable estimates of how each race actually voted (Wasserman and Segal 1973, 179). Only Gerald Wright's (1977, 500) switch to survey data successfully provided evidence that Wallace's strong support in diverse counties resulted from

⁷ See the next section for a formal explanation of this claim.

changing white behavior rather than unexpected black behavior.

The Ecological Inference Problem in Southern Data

Census data usually are reported in tables, with the figures aggregated over geographical units such as counties or towns. Election returns and voter registration statistics also are collected and reported for discrete areal units. This sacrifices the individual-level detail that is often the object of comparative inquiry. In particular, we cannot discern how Southern whites voted across different communities without first estimating racial behavior and removing black voters from the mix.

The hypothetical cross-tabulation in Table 6-1 presents a typical situation. One might wish to know the extent of white support for George Wallace's 1968 presidential campaign in Louisiana, yet election returns are not reported by race, and individual-level data to compute such a breakdown are unavailable. The empty cells in this table therefore reflect the missing information. The marginal totals, on the other hand, represent data typically available to a researcher: overall voter preferences, and the number of potential voters from each grouping of interest.⁸

The challenge of ecological inference, then, is to estimate Table 6-1's missing numbers. If we only possessed such figures for Louisiana as a whole, the internal cells of the table would be lost, but we have parallel data for all 64 parishes (i.e., counties) in the state. Presumably Wallace's vote will vary across parishes in response to shifting racial densities, reflecting the unequal rates of support the segregationist Alabama governor received from blacks and whites. Experts armed with both detailed contextual knowledge and adequate statistical techniques might be able to tease reliable racial voting estimates out of this common variation.

The problem is, until quite recently the statistical techniques used for this purpose were inadequate to the task. Those methods, discussed in more detail below (as well as in King 1997, 58-

⁸ For purposes of this analysis, the label "whites" refers to everyone who does not self-identify as an African American. Other races and ethnic groups are simply too few to support finer group distinctions.

Table 6-1: Wallace's 1968 Louisiana Vote

	Pro-Wallace	Anti-Wallace	No Vote	Total Voting-Age Population
Voting-Age Whites	?	?	?	1,508,009
Voting-Age Blacks	?	?	?	532,947
All Races	530,300	567,150	943,506	2,040,956

Note: Data on voting-age population come from the Census Bureau. Election results are from *America Votes*. "Whites" include all citizens who did not identify themselves as African Americans.

60), required assuming that voters of the same race behaved identically everywhere. This cannot be true, because voters are not aggregated together randomly; contextual variation is almost always present in real-life voting studies. Whites who live in wealthy suburbs will not share the same material interests or political priorities as whites in urban neighborhoods, or even in lower-status suburbs. Migration patterns, over time, may accentuate community differences. State legislatures mold political boundaries, sometimes even precincts, to embrace or exclude identifiable social groups.

Furthermore, once groupings are in place they may shape internal behavior. The most obvious case, to the extent it really exists, is the "white backlash" phenomenon being studied here: whites may react directly to the racial mix of their communities. But they also may respond indirectly to racial densities. A community with a large minority population may exhibit different social needs, may promote itself through different political alliances, and may spawn different political cleavages

than found in all-white communities. Whites with minority neighbors may not receive the same encouragement to turn out on election day, or they may be caught up in the excitement behind an African-American politician that they otherwise would have ignored.⁹

Neglecting contextual behavior patterns, which are partly hidden in data gathered by areal units, will produce systematic estimation error called “aggregation bias” (King 1997; Palmquist 1993; Palmquist and Voss 1996, 21-22). Aggregate-data analysis that ignores this bias frequently produces embarrassing results—including impossible estimates (Lublin and Voss 1997, 41; King 1997, 15-17; Voss and Lublin 1998) and naive findings that contradict all known substantive information (Freedman et al. 1991; Schoenberger and Segal 1971, 585; Wasserman and Segal 1973). The assumption also can throw off estimates when the mistakes are not detectable (Robinson 1950). Leaving aside the estimation error inherent in these traditional methods, furthermore, the aggregation bias they assume away is precisely the phenomenon for which I need to test.

Ecological Regression

Ecological regression (sometimes called Goodman’s method) is the most common approach for estimating how people voted. Although researchers have tried a host of refinements, the basic intuition is fairly simple. Wallace’s vote consists of four quantities, combined as such:

$$\begin{aligned} \text{Vote} &= \# \text{ of blacks} && * \text{ black rate of Wallace support} \\ &+ \# \text{ of whites} && * \text{ white rate of Wallace support} \end{aligned}$$

Two of these figures, the number of blacks and the number of whites, appear as marginals of table 6-1. Which is to say, we know them. We also know them for each parish. The other two are the

⁹ Huckfeldt (1983) finds, for example, that a radical Polish activist running in 1968 Buffalo won disproportionate support among all ethnics in heavily Polish neighborhoods, not just among Polish voters themselves.

quantities we would like to know, statewide if not in each parish.

We can express the same relationship in proportions:

$$V^w = \beta^b X + \beta^w (1-X)$$

where V^w is Wallace's proportion of the three-party vote,

β^b and β^w represent his rates of support among blacks and non-blacks, respectively, and X represents the proportion of the state population that is African American.

Wallace's vote in parish i will have the same components:

$$V_i^w = \beta_i^b X_i + \beta_i^w (1-X_i)$$

Goodman's insight was that, if we treat underlying racial voting preferences as constant (e.g., $\beta_i^b = \beta_{i+1}^b = \beta_{i+2}^b = \dots = \beta^b$), and add an error term e_i to make room for stochastic processes, the Wallace vote notation resembles a linear regression formula:¹⁰

$$\begin{aligned} V_i^w &= \beta^b X_i + \beta^w (1-X_i) + e_i && \text{Version with no constant} \\ &= \beta^b X_i + \beta^w - \beta^w X_i + e_i \\ &= \beta^w + (\beta^b - \beta^w) X_i + e_i && \text{Version with a constant} \end{aligned}$$

The researcher therefore has two choices: regressing the Wallace vote (V^w) on the black and white population proportions without a constant, or regressing the vote on black density with a constant and computing the quantities of interest from the parameter estimates. The two approaches to "naive" ecological regression are equivalent, and either way will capture the linear relationship between parish racial densities and Wallace's vote.

Any of the plausible contextual effects described in the last subsection would violate the assumption that β^w is constant across units, however, and do so in a particularly destructive way: by

¹⁰ A criticism I will not address, but is worth mentioning here, is that the error actually should appear in the coefficients, β^b and β^w , because once we know the black population proportion and the racial voting rates the vote cannot vary. The *coefficients* must be what fluctuate from one county to the next, causing any stochastic variation in the vote. Pretending the variation is an Ordinary Least Squares disturbance term causes heteroskedasticity even when the parameter variation is purely random (Hildreth and Houck 1968).

causing voting behavior to vary across communities in a manner correlated with their racial makeup. That is, they create aggregation bias, which distorts ecological regression. Consider a simple white backlash reaction, in which white support for Wallace increases linearly as parishes contain more blacks. The formula for his support now contains a shifting parameter, β_i^w , which is a function of racial density:

$$\begin{aligned} V_i^w &= \beta^b X_i + \beta_i^w (1 - X_i) + e_i \\ &= \beta_i^w + (\beta^b - \beta_i^w) X_i + e_i \\ &\text{where } \beta_i^w = \gamma^0 + \gamma^1 X_i \end{aligned}$$

Naive ecological regression, performed on data of this sort, will not estimate the desired quantities (i.e., β^b and the weighted average of β_i^w , presuming all other assumptions of the model were correct).

Instead, it would turn up the following:

$$\begin{aligned} V_i^w &= \beta_i^w + [\beta^b - \beta_i^w] X_i + e_i && \text{where } \beta_i^w = \gamma^0 + \gamma^1 X_i \\ &= (\gamma^0 + \gamma^1 X_i) + [\beta^b - (\gamma^0 + \gamma^1 X_i)] X_i + e_i \\ &= \gamma^0 + [\beta^b + \gamma^1 - (\gamma^0 + \gamma^1 X_i)] X_i + e_i \\ &= \gamma^0 + [\beta^b + \gamma^1 (1 - X_i) - \gamma^0] X_i + e_i \end{aligned}$$

The estimated intercept, assumed to represent white support for Wallace, will be γ^0 only. How does γ^0 differ from the real white rate, $\gamma^0 + \gamma^1 X_i$? If the contextual effect γ^1 is positive, as implied by the backlash logic, then the missing component is also positive because X_i will never fall below zero.

The estimate of white voting for Wallace will be too low. Similarly, the estimated black vote for Wallace will be too high.¹¹

¹¹ Adding the intercept to the slope provides not β^b but $\beta^b + \gamma^1 (1 - X_i)$, with the bias term positive under the backlash logic. One way to consider the problem is as a special case of omitted variable bias. The equation one should have estimated in this simple example is:

This same logic applies in cases of indirect aggregation bias. Let's say, for example, that whites who live in mixed neighborhoods are poorer than those who live in lily-white suburbs, and favor the Democratic Party more than suburbanites because of their greater interest in generous social-welfare policies. Even though this underlying behavior pattern has only an indirect connection to race, it will still bias racial voting estimates. None of the increasing Democratic support as black density rises will be attributed to white voters, since ecological regression assumes that whites all behave the same way. It will be attributed entirely to blacks.¹²

Given the negative consequences of wrongly assuming constant parameters, it is worth questioning the feasibility of assuming that voters of the same race register at uniform rates everywhere. The presumption is clearly false. Registered voters in two adjacent parishes, Jefferson and Plaquemines, were more than 90% white. Whereas Plaquemines voted overwhelmingly for Wallace (75%), fewer than half (46%) of voters did so in neighboring Jefferson. Clearly white preferences varied, a violation of the model. More complex contextual effects could undermine ecological regression models to an even greater degree, a flaw for which the models offer little warning and no solution. The most traditional method of aggregate-data analysis thus is incapable

$$\begin{aligned} V_i^w &= \gamma^0 + [\beta^b + \gamma^l(1 - X_i) - \gamma^0]X_i + e_i \\ &= \gamma^0 + (\beta^b - \gamma^0)X_i + \gamma^l(1 - X_i)X_i + e_i \end{aligned}$$

The exclusion of $(1 - X_i)X_i$ biases the coefficient on X_i . This bias is a product of two terms: γ^l , and the correlation between the two potential independent variables. Both are positive in this example, so the bias on estimated black support for Wallace would be upward.

Here, adding the missing variable would solve the bias difficulty. More generally, one way to lessen aggregation bias in ecological regression would be to allow interaction terms that permit a varying coefficient on X_i (Alt 1993). The white voting rate could shift according to a slew of precinct or county traits, when data on such are available. Unfortunately, this method loses usefulness when both races respond directly to the community's racial breakdown, as commonly occurs (Branton and Jones 1999; Bullock, Gaddie and Kuzenski 1995; Freedman et al. 1991; Gay 1996; but see Gaddie and Bullock 1997, 37). Then the model falls into indeterminacy (King 1997, 39-46).

¹² This is not just an hypothetical example. It explains why estimates of the black Democratic vote regularly fall outside the range of possible values (e.g., King 1997, 16). Minorities appear to give more than 100% of their vote to liberal candidates because their white neighbors are also more likely to be Democrats (Freedman et al. 1991, 690-91).

of testing for, or even operating in the presence of, the effect I wish to study.

Ecological regression, as commonly practiced, also exhibits a number of well-known deficiencies unrelated to aggregation bias (King 1997, chap. 4). Goodman's method:

- Allows impossible vote predictions—without artificial corrections, at least—whether caused by aggregation bias or not. For example, my application of ecological regression (detailed in the next chapter) implies that more people voted in 1968 than possibly could have given the observed turnout.
- Produces only a single voting estimate for any collection of states, counties or precincts. The precinct data for a voting district produce a single district estimate. The county data for a state produce a single state estimate. State data only produce one estimate for the entire country. The researcher loses the rich geographical variation present in aggregate data, one of the central motives for using them in the first place.
- Cannot handle when voters of one group are less consistent in their voting behavior than voters of another group (heteroskedasticity). Since white preferences varied widely from one parish to another, but minorities consistently opposed George Wallace's candidacy, that statistical problem clearly occurs here.

Homogeneous Precinct Analysis

An alternate approach to estimating racial behavior is to restrict the analysis to fully segregated units. This sort of research does not require computer-intensive statistical tools. The analyst merely looks at units that are all white, using their turnout and vote choice figures as the best guess for how whites behaved in general, and then similarly examines all-black units to guess how blacks voted in general.

This method has the virtue of certainty because we know almost exactly how blacks and

white voted within the small subset of segregated units. Yet it can be disastrously misleading when conclusions are then generalized to more diverse communities, which are often quite different. Both whites and blacks tend to turn out at higher rates where they dominate, for example (Gay 1996). Some researchers have found that mixed areas exhibit pathologies that neither all-white nor all-black communities do (Rieder 1985, 21). In other words, the approach is even more vulnerable to contextual effects than ecological regression. Nor does the method, being ad hoc, allow any measurement of the uncertainty contained in estimates.

The most common name for the approach—homogeneous *precinct* analysis—underscores other drawbacks. It is only useful when residential patterns are highly segregated, as one often finds in urban precincts. It is useless for county-level analysis, since counties are seldom almost entirely black. It is also unreliable when the bulk of aggregation units fall between the extremes, since that situation would require inferring the behavior of a large area from a handful of unrepresentative locations. Finally, and most importantly, homogeneous precinct analysis has nothing to offer the study of white backlash, since it leaves out any place where backlash is supposed to appear. It cannot operate in the face of aggregation bias any more than ecological regression can.

Informed Assumptions

The last approach is to fix some of the parameters otherwise found in an ecological regression, simplifying what the researcher must estimate. Consider again the Wallace vote formula:

$$V_i^w = \beta^b X_i + \beta^w (1 - X_i) + e_i$$

If the researcher can support a strong prior assumption that no blacks backed Wallace, for example, then $\beta^b = 0$, and fluctuations in Wallace's vote are entirely a function of his white rate of support. Similarly, in the case with an African-American Democrat running against a white Republican, the

researcher might assume that no blacks crossed over to the white candidate. Relying on informed assumptions is particularly tempting, because it provides more than just a statewide representation of white behavior; it means that researchers can assume they know what happened everywhere. This is why so much research falls back on the technique.

Most ecological studies of racial politics rely on assuming how blacks voted.¹³ Giles and Buckner (1993, 705n), for example, assume that no blacks voted for former Klansman David Duke in Louisiana, contrary to available survey evidence (*New York Times* Poll 1991). Orey (1997) made the same assumption for Louisiana's 1995 gubernatorial election, when the Republican candidate was State Sen. Mike Foster, again in the face of contrary exit polls. Roscigno and Tomaskovic-Devey (1994) assume that no blacks voted for North Carolina Sen. Jesse Helms, whether against white Democrat Jim Hunt or against black Democrat Harvey Gantt. In an edited volume chapter, Giles and Buckner (1995) assume that no Louisiana blacks voted for George Bush in 1988.

The most obvious limitation of this approach, of course, is that it's only feasible for severely polarized elections. Evidence indicates that, for whatever bizarre reason, even the most extreme candidates will draw some black supporters (*New York Times* Poll 1991; Palmquist and Voss 1996, 1997). Once the researcher's interest expands to more mainstream elections, such as the 1988 presidential contest (Giles and Buckner 1995), absolute priors of this sort become extremely difficult to defend.

More problematic is that researchers will receive no warning when their presumption is inappropriate, unless they happen to have other data flagging the problem. Incorrectly assuming that no blacks (or all blacks) supported a candidate will bias voting estimates, a bias that will not be "randomly distributed with regard to the variable of interest" (Giles and Buckner 1993, 705n). Rather, the bias will be proportionate to the size of the black (or white) population. If we treat λ

¹³ Salamon and Van Evera (1973, 1,292-93) provide a rare exception of assuming how whites voted.

as the Wallace vote proportion among the entire population, then the estimate of his white support (i.e., of λ_i^{ww}) will include whatever small proportion of his vote came from blacks (λ_i^{bw}):

$$\hat{\mathbf{I}}_i^{ww} = \mathbf{I}_i = \mathbf{I}_i^{ww} + \mathbf{I}_i^{bw}$$

Computed as the conditional frequency of *white* support, β^w , this becomes:

$$\hat{b}^w = \frac{\hat{\mathbf{I}}_i^{ww}}{(1 - \mathbf{X}_i)} = \frac{\mathbf{I}_i^{ww} + \mathbf{I}_i^{bw}}{(1 - \mathbf{X}_i)} = \frac{\mathbf{I}_i^{ww}}{(1 - \mathbf{X}_i)} + \frac{\mathbf{I}_i^{bw}}{(1 - \mathbf{X}_i)}$$

The last term, to the right of the addition sign, captures bias in any one parish (since the portion to the *left* of the addition sign is the quantity of interest). If 1% of the population in a 50% black parish consists of African-American Wallace voters, then bias added to the white support rate will be .01 divided by .5, or 2 percentage points. If 2.8% of the population in a 40% black parish consists of African-American Bush supporters, bias in the white support rate will be .028 / .6 = .047, or almost 5 percentage points!¹⁴ The assumption creates bias in favor of the white backlash phenomenon, because the white racial conservatism estimate is inflated most where blacks are most numerous.

One other pitfall can foul up either ecological regression or informed assumptions. Usually researchers lack more than just racial voting preferences; they also do not know racial turnout rates. Analysts applying ecological regression often get around this difficulty by conducting a two-stage analysis—one ecological regression to estimate turnout by race, then a second to gauge preferences among the voting population implied by the previous stage (Grofman 1993; Grofman and Migalski 1988; Kleppner 1985, 82-83; Kousser 1973). Each stage of double regression, of course, runs up

¹⁴ I chose these hypothetical figures because they represent plausible rates of real black voting: 2% of the black vote for Wallace, and 7% of the black vote for Bush.

against the difficulties with ecological regression discussed above—plus offers no means for computing uncertainty estimates (King 1997, 71-73). However, at least double regression acknowledges that the racial makeup of a voting population is not identical to the racial makeup of a parish. Homogeneous unit analysis also gets around this difficulty, since the researcher simply assumes that all votes were cast by people of the same race.

The method of applying informed assumptions, on the other hand, suggests no solution. Even in the rare case when informed assumptions about racial voting behavior are plausible, the analyst is unlikely to possess defensible prior assumptions about racial turnout rates.¹⁵ Dividing Wallace's vote by the entire pool of potential white voters, rather than by the whites who turned out, would treat all abstentions as votes against Wallace.¹⁶ Assuming equal turnout rates, by contrast, allows the estimated Wallace white vote to vary as a function of African-American mobilization!¹⁷ Neither of these implicit decisions strikes me as acceptable.

The “neighborhood model,” proposed by Freedman and associates (1991, 678), makes an equally implausible assumption: that whites and blacks in a given locale behave identically.¹⁸ Racial behavior might differ statewide, but only as a function of distinctive residential patterns. Whites and blacks in the same neighborhood face similar experiences, hold similar resources, and possess similar life chances. Thus, they suggest, in many cases the safest assumption is that every racial and ethnic group within an aggregation mobilizes uniformly. This argument may sound like sophistry, given how strongly the American electorate is polarized by race, but it actually underscores a critical

¹⁵ The obvious exception is when, as in Louisiana or South Carolina, racial turnout rates are available (Palmquist and Voss 1996).

¹⁶ Giles and Buckner (1993) select this approach in their study of David Duke's Louisiana campaigns.

¹⁷ Bositis (1997) implicitly adopts this assumption in his analysis of the 1996 Georgia congressional elections.

¹⁸ To be fair, Freedman et al. (1991) defend this assumption when applied to precincts or “neighborhoods”—for which it makes some sense—not to whole counties. But I include it in all levels of analysis, just to have an additional standard of comparison.

insight that ecological regression misses. If whites in mixed-race neighborhoods differ from their segregated counterparts, ecological regression will not capture the deviation. Otherwise, given the central role that race plays in shaping attitudes, this sort of “informed” assumption does not seem any more acceptable than the others.

Conclusion

Political scientists have struggled in vain to reconcile contextual hypotheses with the known shortcomings of aggregate data. Most simply do not try—they use voting intentions expressed in surveys, a method with its own attendant limitations, or they concentrate on racial attitudes, a worthy subject for analysis but not identical to racial politics. Others use the dominant ecological inference methods, apologetically or not, but operate under a veil of suspicion (among social scientists, if not in the courtroom).¹⁹ The political geography of Southern politics therefore mostly lies fallow, and the conventional wisdom drawn from that line of research survives, for the most part, unchallenged. Overcoming the ecological inference problem would open up numerous untapped possibilities. In the next chapter, I show that Gary King’s proposed solution (1997)—whatever its limitations in other applications (Cho 1998)—overcomes the most severe difficulties with studying Southern voting data, and even introduces new opportunities that previous methods did not permit.

¹⁹ Since the pitfalls of survey implementation and analysis are legion (see, for example, Brady and Orren’s relatively sympathetic 1992 discussion of media polls), and quite likely to produce erroneous or unstable conclusions (Smith 1995), the level of suspicion probably is not proportional to the merit.