

Rich state, poor state, red state, blue state: What’s the matter with Connecticut?*

Andrew Gelman,[†] Boris Shor,[‡] Joseph Bafumi,[§] David Park[¶]

Abstract

We find that income matters more in “red America” than in “blue America.” In poor states, rich people are much more likely than poor people to vote for the Republican presidential candidate, but in rich states (such as Connecticut), income has a very low correlation with vote preference. In addition to finding this pattern and studying its changes over time, we use the concepts of *typicality* and *availability* from cognitive psychology to explain how these patterns can be commonly misunderstood. Our results can be viewed either as a debunking of the journalistic image of rich “latte” Democrats and poor “Nascar” Republicans, or as support for the journalistic images of political and cultural differences between red and blue states—differences which are not explained by differences in individuals’ incomes.

For decades, the Democrats have been viewed as the party of the poor, with the Republicans representing the rich. Recent presidential elections, however, have shown a reverse pattern, with Democrats performing well in the richer “blue” states in the northeast and west coast, and Republicans dominating in the “red” states in the middle of the country. Through multilevel modeling of individual-level survey data and county- and state-level demographic and electoral data, we reconcile these patterns.

Key methods used in this research are: (1) plots of repeated cross-sectional analyses, (2) varying-intercept, varying-slope multilevel models, and (3) a graph that simultaneously shows within-group and between-group patterns in a multilevel model. These statistical tools help us understand patterns of variation within and between states in a way that would not be possible from classical regressions or by looking at tables of coefficient estimates.

Keywords: availability heuristic, ecological fallacy, hierarchical model, income and voting, multilevel model, presidential elections, public opinion, secret weapon, varying-slope model.

*We thank the members of the Quantitative Political Science Discussion Group at Columbia University, participants at the Midwest Political Science Association meeting, Greg Wawro, Robert Erikson, Robert Shapiro, Elke Weber, Eric Johnson, Craig Newmark, Shigeo Hirano, David Epstein, Stephen Ansolabehere, Stuart Jordan, Dave Krantz, and participants at seminars at the George Mason University economics department and the Columbia University psychology department for helpful comments; Maryann Fiebach, Boliang Zhu, Brian Fogarty, Amanda Czerniawski, and Andres Centeno for research assistance; David Weaver and Eunseong Kim for data from the American Journalist Survey; David Leip for election data; the Inter-university Consortium for Political and Social Research for poll data; and the National Science Foundation for financial support.

[†]Department of Statistics and Department of Political Science, Columbia University, New York, gelman@stat.columbia.edu, www.stat.columbia.edu/~gelman

[‡]Harris School of Public Policy, University of Chicago

[§]Department of Political Science, Dartmouth College

[¶]Department of Political Science, Washington University, St. Louis

I never said all Democrats are saloon-keepers. What I said is that all saloon-keepers are Democrats. — Horace Greeley, 1860

Pat doesn't have a mink coat. But she does have a respectable Republican cloth coat. — Richard Nixon, 1952

Like upscale areas everywhere, from Silicon Valley to Chicago's North Shore to suburban Connecticut, Montgomery County supported the Democratic ticket in last year's presidential election, by a margin of 63 percent to 34 percent. — David Brooks, 2001

There is, for example, this large class of affluent professionals who are solidly Democratic. DataQuick Information Systems recently put out a list of 100 ZIP code areas where the median home price was above \$500,000. By my count, at least 90 of these places—from the Upper West Side to Santa Monica—elect liberal Democrats. — David Brooks, 2004

A lot of Bush's red zones can be traced to wealthy enclaves or sun-belt suburbs where tax cuts are king. — Matt Bai, 2001

But in the Ipsos-Reid surveys, 38 percent of voters in "strong Bush" counties said that they had household incomes below \$30,000, while 7 percent said that their families earned at least \$100,000. In "strong Gore" counties, by contrast, only 29 percent of voters pegged their household income below \$30,000, while 14 percent said that it was above \$100,000. — James Barnes, 2002

One of the more glaring differences between the two groups is income. The average per capita income in the Blue states was \$27,899 in 1998, versus \$23,722 for the Red states, per the Bureau of Economic Analysis. Of the top 10 metropolitan areas with the highest per capita income, eight were from Blue states; the other two were from the highly contested state of Florida. Similarly, eight of the top 10 metro areas with the lowest income levels were from Red states. — Todd Wasserman, 2002

1 Democrats and Republicans, rich, and poor: two perspectives

Throughout the twentieth century and even before, the Democratic Party in the United States has been viewed as representing the party of the lower classes and thus, by extension, the "average American." More recently, however, a different perspective has taken hold, in which the Democrats represent the elites rather than the masses. The view of Democrats as elitists began perhaps with the party's control over government planning during the New Deal era and was developed as a key theme by populist Republicans in the postwar period, including Nixon, Goldwater, and Reagan.¹ The Democratic party too has had many prominent populists, from Huey Long to Harry Truman to Jimmy Carter and Al Gore ("the people versus the powerful"), but this

¹See, e.g., Buckley and Bozell (1954), Rovere (1959), McGirr (2001), Perlstein (2002), and Greenberg (2003).

is less remarkable given its majority status during this period. Throughout, populism has been associated with the attitudes and interests of lower and middle-income Americans.

What is happening now? Do richer voters still support Republicans? If so, how can we understand the pattern that the Democrats do best in the richer “blue states” of the Northeast and West, while the Republicans dominate in the poorer “red states” in the South and between the coasts? And does living in a poor or rich state change individual vote preferences in some fashion? In other words, does context matter for individual voting behavior and, if so, how? We explore these questions by studying the relation between income and presidential vote preference, at the individual, county, and state levels. It turns out that the connections between income and voting in the United States are not simple; we find that rich and poor states differ in the relation between individual income and partisan preferences.

1.1 Perspectives from social science and the news media

Census and opinion poll data since 1952 reveal that higher-income voters continue to support the Republicans in presidential elections.² However, higher-income states have in recent years favored the Democrats. The Republicans have the support of the richer voters within any given state but have more overall support in the poorer states. Thus, the identification of rich states with rich voters, or more generally, the “personification” of so-called red and blue states, is misleading. For example, in the context of the Brooks quotes above, within an “upscale” area that supports the Democrats, the more “upscale” voters are still likely to vote Republican.

The connection between income and support for conservative parties has long been noted and has attracted interest from political scientists and sociologists studying ideological polarization. McCarty, Poole, and Rosenthal (2005) argue that partisanship and presidential vote choice have become more stratified by income over the past fifty years. Comparing survey respondents in the highest and lowest quintiles of income, they find that in 1956 and 1960, the proportion of Republican identifiers was only slightly higher in the highest than in the lowest quintile, but in 1992–2000, respondents in the highest quintile were more than twice as likely to identify as a Republican than were those in the lowest. Stonecash (2000, 2005) finds a growth of support for the Democrats since the 1970s among poor persons and in high-poverty areas. In contrast, Fiorina, Abrams, and Pope (2005) find polarization of the political class but not of the general voting population, with only small differences in issue preferences when comparing voters in “red” and “blue” states. Fiorina et al., however, do not discuss voting in relation to income, so our analysis supplements theirs by considering this variable. In an extensive analysis of opinion poll data, Ansolabehere, Rodden,

²For example, from 2004 exit polls, Bush received 36% of the support of voters with incomes under \$15,000, 41% with incomes between \$15,000 and \$30,000, monotonically increasing to 62% of those with incomes over \$200,000.

and Snyder (2005) find voters to be most strongly motivated by economic issues, but they note that the connection between income and economic views can be weak. Brooks and Brady (1999) and Bartels (2005) find that income continues to be predictive of partisanship,³ and Filer, Kenny, and Morton (1993) studied the connection between income and voter turnout.

In contrast, media attention has focused much more on comparisons of states (and, to a lesser extent, counties), as illustrated by many of the quotations that lead off this article. Our aim is to provide a framework to simultaneously understand Republican strength among richer voters and in poorer states, and to study these trends over time. The journalists who see patterns on the red-and-blue map and the political scientists who analyze polls are talking past each other because they are looking at different levels of aggregation. Public perceptions of the two parties are important, and after setting the record straight on what is actually happening with income and voting, we consider some explanations from cognitive psychology for why misunderstandings about the correlations between income and vote preference could persist among otherwise well-informed observers.

1.2 Studying patterns at the state level

Comparing to previous studies of income and voting, our key contribution is to study patterns both within and between states, with both individual income and state-level income as predictors, using survey data on individuals and election and Census data for states and counties. The pattern that richer states support the Democrats is *not* a simple aggregation of rich voters supporting the Democrats. This can be viewed either as a debunking of the journalistic image of rich “latte” Democrats and poor “Nascar” Republicans—or as support for the journalistic images of political and cultural differences between red and blue states—differences which are not explained by differences in individuals’ incomes.

We find that income matters more in “red America” than in “blue America.” In poor states, rich people are much more likely than poor people to vote for the Republican presidential candidate, but in rich states (such as Connecticut), income has almost no correlation with vote preference. The United States has red and blue voters, and red and blue states, but income cuts across them in different ways (a point noted by Alford (1963) in his study of social class and voting by region of the U.S.). As we shall see in Sections 2–4, the statistical technique of multilevel modeling allows us to understand the relation between income and vote among individuals, counties, and states.

³Manza and Brooks (1998, chapter 3) show that the consistent correlation of high income with Republican vote masks changes in particular social and occupational groups (for example, professionals have moved toward the Democrats and self-employed persons toward the Republicans); here we focus on income, partly because of its relevance for government policy but especially because of its salience in current political discourse, an issue we return to at the end of this article.

Section 5 considers reasons for these patterns, along with psychological reasons why certain misunderstandings have persisted, and a discussion of the relevance of income/voting patterns to political perceptions.

The patterns of income and voting by state are politically important, and the common misperceptions of these patterns are also important. This article attempts to make sense of the data and also the misperceptions.

2 Studying the relation between income and vote preferences

Survey data show a small but persistent correlation between income and support for the Republican party, but at the aggregate level, it is the Democrats who do better in the richer states. Our strategy to understand these patterns is to study the relation between income and voting in four ways:

- Aggregate, by state: to what extent do richer states favor the Democrats?
- Nationally, at the level of the individual voter: to what extent do richer voters support the Republicans?
- Individual voters within states: to what extent do richer voters support the Republicans, within any given state? In other words, how much does context matter?
- Counties within states: to what extent do richer counties favor the Democrats, within any given state?

Patterns at these four levels have much different political interpretations from those supposed by confused political commentators. Most notably, the support for the Democrats in the richer states had led observers to view the typical Democrat as an upper-middle-class resident of a coastal metropolitan area, and the typical Republican as lower-middle-class and rural (see Brooks, 2001). That these claims have been overstated (see Frank, 2004, and Issenberg, 2004) does not seem to lessen their appeal.

A multilevel strategy in understanding voting behavior is useful because we are interested in election outcomes as well as individual decisions. Elections are not simple cumulations of voter decisions (because of institutional features such as electoral rules and geographic boundaries, and the political decisions of parties and candidates), and so aggregate analysis should not be discarded simply by citing the ecological fallacy problem (Wright, 1989). Trends of economic voting at multiple levels of analysis may or may not be similar, and their causes may or may not be similar.

As in Wright (1989, p. 390), we consider the variation at each level of analysis. Income varies far more *within* states than average income does *between* states. Consequently, it is the within-state rather than the between-state effect of income that dominates the national patterns. In particular, a positive correlation of income and Republican voting within states, plus a negative correlation between states, combine to form a positive correlation among all voters.

We have both individual and aggregate data on income and votes. Thus, the statistical analysis is relatively straightforward, without the well-known problems that can arise when only aggregate data are available (Robinson, 1950, Kramer, 1983). For individual vote preferences, we use the National Election Study (NES) polls taken during the month before each election. For aggregate patterns, we use presidential election returns and Bureau of Economic Affairs data on average income by state and county. We estimate the relations between income and vote preferences for each presidential election from 1968 to 2000, with a particular goal of studying trends including any changes over time in the support for particular political parties.

2.1 Analysis of state averages

We begin by fitting a state-level linear regression for each election year, predicting state support for the Republican candidate in the election from the average income in the state. Positive coefficients imply that richer states are supporting the Republicans more. We also examine the coefficient of average income after controlling for percent African-American in the state. To allow comparability over time, we adjust incomes in each year to 1996 dollars.

2.2 National analysis of individual voters

Our first individual-level analysis is a simple logistic regression modeling vote preference from the NES polls (1 = Republican, 0 = Democrat, excluding respondents who were undecided or supported third-party candidates) on income. We summarize family income with a five-point quantile-based scale⁴ used by the NES, which allows the results to be comparable over time in terms of relative incomes. We fit a separate logistic regression for each election year; if the coefficients are positive, this implies that Republicans were differentially supported by richer voters.

We also examine the coefficient of income when additional predictors are added into the model, including ethnicity (African-American or other), sex, age (18–29, 30–44, 45–64, or 65+), education, party identification, and ideology. However, our primary analyses use only income

⁴The National Election Study uses 1 = 0–16 percentile, 2 = 17–33 percentile, 3 = 34–67 percentile, 4 = 68–95 percentile, 5 = 96–100 percentile. We label these as –2, –1, 0, 1, 2, centering at zero so that we can more easily interpret the intercept terms of regressions that include income as a predictor.

as a predictor, because our goal is to study differences between richer and poorer voters. Even if income effects were “explained” by other predictors, the correlations would still be real.

2.3 Analysis of individuals within states

To study the relation of income to individual vote preferences, controlling for state, we fit from each election year’s NES data a multilevel logistic regression of vote preference on income (using the five-point scale noted in Footnote 3) and state:

$$\Pr(y_i = 1) = \text{logit}^{-1}(\alpha_{s[i]} + \beta x_i), \text{ for } i = 1, \dots, n, \quad (1)$$

where $s[i]$ indexes the state (from 1 to 50) corresponding to respondent i , x_i is the person’s household income (on the five-point scale), and n is the number of respondents in a given poll. This varying-intercept model gives us 50 state-level coefficients allowing geographic variation in support for the Republican candidate in each election. The coefficient for income then represents the extent to which Republicans are differentially supported by richer voters, within any given state. As before, we fit the model separately from the NES in each presidential election year and then display estimated coefficient of income, β , over time. For the year 2000, we also fit using the National Annenberg Election Survey, which, with over 100,000 respondents, allows accurate estimation of the patterns in individual states. We also check our results by fitting to the exit polls from 2000 and 2004.⁵

Because we are interested in comparing states of different income levels, we include the average income within each state as a group-level predictor. Including this predictor also increases the precision of the multilevel model fit, by reducing the residual error at the state level. As in the national poll analyses, we also examine the coefficients for income when ethnicity, sex, and age are included in the model. In addition, we consider models including the state-level Gini index to account for income inequality within states. Any of these models can be written as,

$$\alpha_s = U_s \gamma + \epsilon_s, \text{ for } s = 1, \dots, 50,$$

where U is the matrix of state-level predictors, and the errors ϵ_s have mean 0 and variance σ_α^2 , which can be estimated from the data when combined with the individual-level model (1). Multilevel modeling is the appropriate tool for this set of analyses because it allows for variation among states, even when sample sizes are insufficient for accurate estimation of the individual state-level

⁵We used the well-known (if controversial) news media exit polls: Voter News Service in 2000 and National Election Pool in 2004. These polls have disadvantage of a messier sampling scheme and use different income categories than the Annenberg and NES surveys. However, the exit polls have large sample sizes (even in small states) and provide an independent source of data with which to check our results.

effects, as in the NES.⁶ We fit the multilevel models using the Bayesian software package Bugs (Spiegelhalter et al., 1994, 2002) as linked from the open-source statistical software package R (R Project, 2000, Gelman, 2003).

We shall also fit varying-intercept, varying-slope models for individual income, but we defer these to Section 4.2, following a thorough exploration of the models described here.

2.4 Analysis of county averages within states

We can study aggregate patterns at a more local level using vote and income data from counties, thus giving us 3000+ data points in each year rather than 50. We perform a county-level analysis controlling for states (that is, a hierarchical model of counties within states, using income as a county-level predictor), to see the extent to which poorer or richer counties favor the Republicans, within any given state. For each election year, we fit a varying-intercept, varying-slope model:

$$y_c = \alpha_{s[c]} + \beta_{s[c]}x_c + \epsilon_c, \text{ for } c = 1, \dots, C,$$

where y_c is the Republican proportion of the two-party presidential vote in county c , $s[c]$ is the state corresponding to county c (thus, not the same $s[]$ index of the individual-level model (1)), x_c is the average income of the county, and the error terms ϵ_c have mean 0 and variance σ_y^2 , which is estimated from data. The state-level coefficients are modeled hierarchically:

$$\begin{aligned} \alpha_s &= a^\alpha + b^\alpha u_s + \epsilon_s^\alpha, \text{ for } s = 1, \dots, 50 \\ \beta_s &= a^\beta + b^\beta u_s + \epsilon_s^\beta, \text{ for } s = 1, \dots, 50, \end{aligned}$$

with errors $\epsilon_s^\alpha, \epsilon_s^\beta$ having mean 0, variances $\sigma_\alpha^2, \sigma_\beta^2$, and correlation ρ , all estimated from data.

We are particularly interested in the within-state coefficients β_s for county income, which can be interpreted as the amount by which richer counties support Republicans in any given state.

3 Results

3.1 Richer states now support the Democrats

We first present the comparison of red and blue states—more formally, regressions of Republican share of the two-party vote on state average per-capita income (in tens of thousands of 1996 dollars). Figure 1a shows that, since the 1976 election, there has been a steady downward trend in the income coefficient over time. As time has gone on, richer states have increasingly favored the

⁶See Datta et al. (1999) for a similar analysis and Kreft and De Leeuw (1998), Snijders and Bosker (1999), and Raubenbush and Bryk (2002) for further discussion of these methods; and Gelman and Little (1997) and Park, Gelman, and Bafumi (2004) for multilevel modeling of vote-preference data.

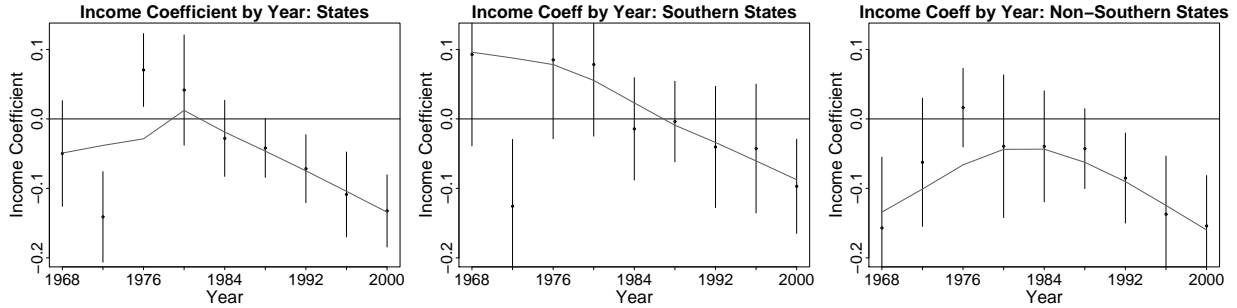


Figure 1: (a) Coefficients for average state income in regression predicting Republican vote share by state. The model was fit separately for each election year. Estimates and 95% error bars are shown. (b, c) Same model but fit separately to Southern and non-Southern states each year. Republicans do better in poor states than rich states, especially in recent years. The curves on this and the other graphs in this article (except for Figures 11 and 12) show lowess (Cleveland, 1979) fits to the time series.

Democrats. So far, this fits with the “David Brooks” story of increasing elite support for the left, rather than the “Horace Greeley” story of elite support for the right. Rich, “blue” states such as California and New York are voting for Democratic presidential candidates, while poorer, “red” states like Alabama and Mississippi are voting Republican. For the past twenty years, the same patterns appear when fitting Southern and non-Southern states separately (Figure 1b,c).

There has been a trend of richer states supporting the Democrats. It makes sense that the “red/blue” issue has been more widely discussed in recent years, as this pattern has become increasingly clear.

We hypothesized that some of this variation could be explained by inequality. However, after refitting the models including the state-level Gini index of income inequality, we found the coefficients for the Gini index to be essentially zero, and there was little change in the coefficients for state income.

3.2 Richer voters continue to support the Republicans overall

We fit a logistic regression to the reported Republican presidential vote preference on personal income, fit separately to each presidential election since 1952. Figure 2 shows that higher-income people have been consistently more likely to vote Republican. Next, we add ethnicity, sex, education, and age into the model. After controlling for other individual-level predictors, the coefficient of income is still consistently positive. Figure 3 shows that the same substantive interpretation and trends remain.

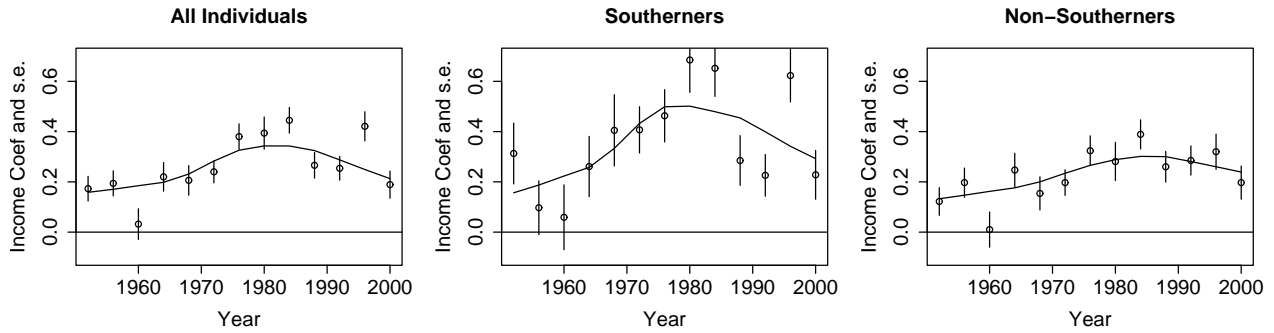


Figure 2: Coefficients for income in logistic regressions of Republican vote, fit to NES data from each year. The positive coefficients indicate that higher-income voters have consistently supported the Republicans, a pattern that holds both within and outside the South.

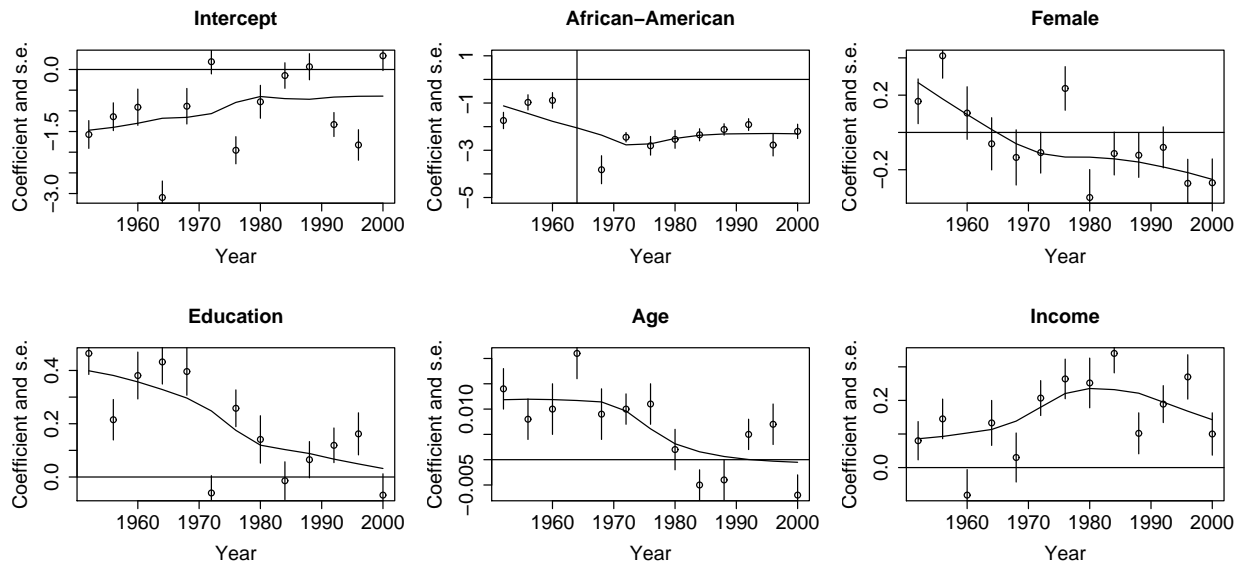


Figure 3: Logistic regression estimates predicting Republican vote given income, ethnicity, sex, education, and age. The coefficients for income are still generally positive, implying that higher-income voters support the Republicans even after controlling for the other predictors in the model.

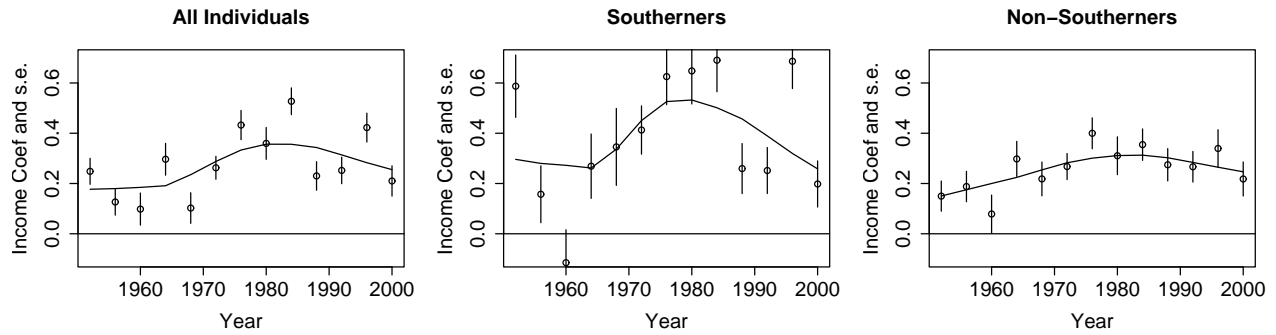


Figure 4: Coefficients for individual income in multilevel model predicting Republican vote by income and state, fit separately to NES data in each election year. The positive coefficients indicate that higher-income voters within states have consistently supported the Republicans.

3.3 Richer voters continue to support the Republicans within states

We fit a multilevel logistic regression to each election year’s NES data, with the 5-point income scale as an individual-level predictor and states as the groups.⁷ Figure 4 shows the estimated coefficient for individual income (and its standard error) over time. The coefficient is consistently positive, implying that, on average, richer voters within states support the Republicans. We also fit the multilevel logistic regression including demographic controls as well as income and state. The resulting coefficient estimates (not shown here) are similar to those in Figure 3, and the estimated coefficients for income remain consistently positive. In recent decades, richer voters have supported the Republicans within states, even after controlling for these other variables.

3.4 Richer counties support the Republicans in some states and the Democrats in others

National trends of county votes

Richer counties used to support the Republicans, but this pattern has steadily declined to zero during the past forty years, a pattern we can see by regressing the Republican proportion of the two-party presidential vote in each county on average county income for each election year; see Figure 5a. In the 2000 presidential election, richer counties tended to support Democrats, for the first time since 1972. The second and third panels of Figure 5 show the patterns in the South—where Republicans still do better in the richer counties, although less so than before—and the rest of the country, where the richer counties in recent years have favored the Democrats after

⁷The NES uses cluster sampling and so, strictly speaking, the states in this analysis actually represent collections of sampled clusters. By ignoring the cluster sampling in the analysis, we may be understating standard errors. We are not so worried about this issue here because the results show a consistent pattern over time.

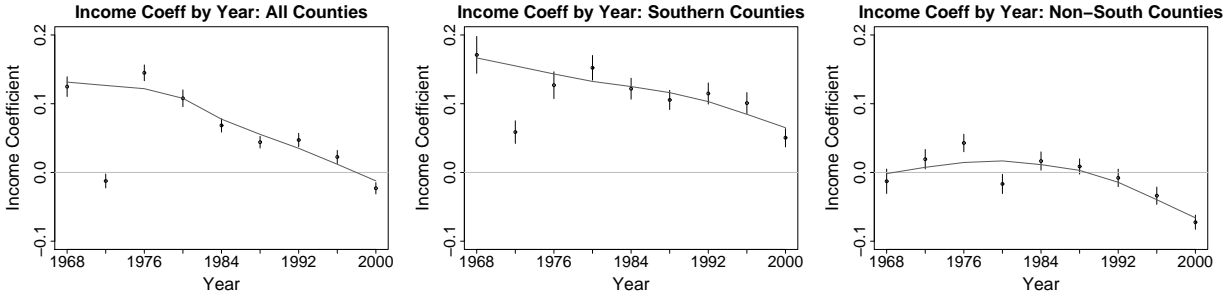


Figure 5: For each presidential election year, coefficient of the regression of Republican vote proportion on county income: (a) model fit to the whole country each year, (b) South only, (c) non-South only. Richer counties used to support the Republicans, but this pattern has declined, to the extent that in 2000, for the first time in almost thirty years, the Republicans did better in poorer counties.

previously supporting the Republicans.

Trends of counties within states

We next perform a multilevel analysis with state as the group-level units, allowing the coefficient for county income to vary by state. In other words, we allow the relation between voting and income to differ in different states. We fit the hierarchical model to all 50 states at once, but for convenience we display the four regions of the country separately.

Figures 6–9 show regional patterns in the trends in the county income coefficient. In most Southern states, rich counties voted for Republicans in the past and continue to do so. An additional pattern is revealed by the ordering of these plots by the percentage of the vote received by George W. Bush in 2000. The Southern states that most supported the Republicans in 2000 show the highest coefficients—that is, the strongest relations between county income and Republican vote share. In contrast, in Western states, richer counties once tended to vote Republican, but now increasingly vote for Democrats (that is, the coefficients in the plots for these states are now negative). Trends in the Midwest and Northeast are more mixed.

Another way to understand these patterns is to compare counties within richer “blue” Democratic-leaning states and poorer “red” Republican-leaning states. In deep-red southern states such as Oklahoma, Texas, and Mississippi, the richer counties support the Republicans and poorer counties support the Democrats. In contrast, consider the states nearest major national media: New York, Maryland, Virginia, and California. In these particular states, the richer counties showed a slight tendency to support the Democrats.

Thus, amusingly, the journalists noticed a pattern (richer counties supporting the Democrats) that is concentrated in the states where the journalists live. For example, Brooks (2001) compared

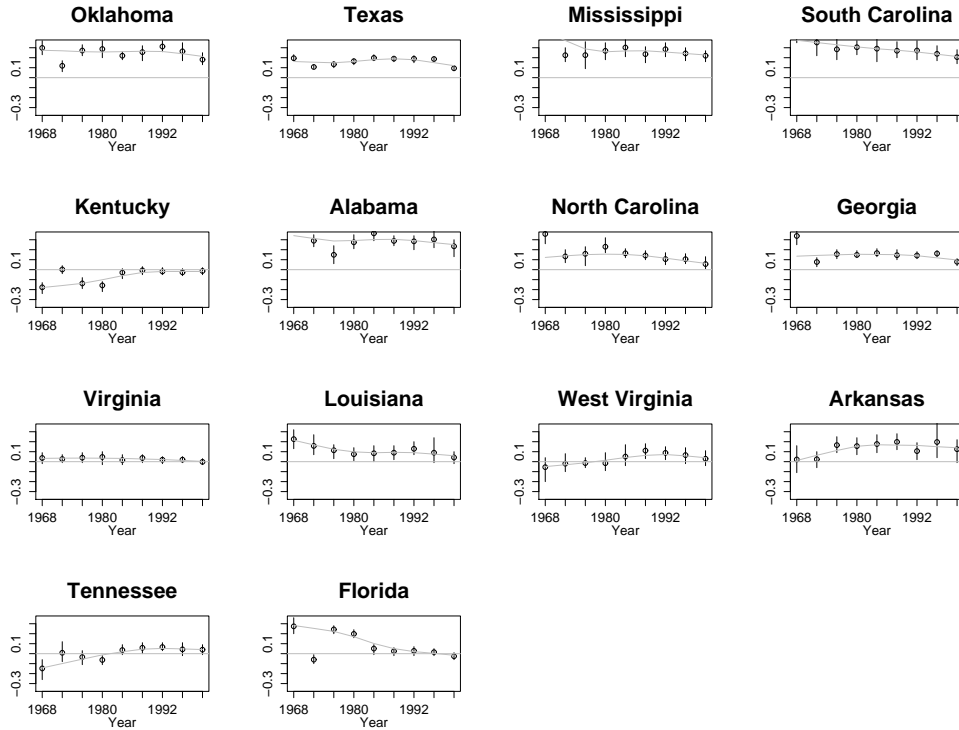


Figure 6: For the Southern states: coefficients of county-level income from multilevel varying-slope model predicting the county-level Republican share of the presidential vote within states, 1968–2000. Within the region, the states are in decreasing order of support for Bush in 2000. Richer counties tend to support the Republicans (that is, they have positive coefficients), especially in the most strongly Republican states.

a rich county in Maryland to a poor county in Pennsylvania. Had he compared counties within states such as Oklahoma, he would likely have noticed an opposite pattern.

We can look at these patterns another way by comparing all 50 states within each election year. Figure 10 shows the coefficient of county income within each state (that is, the varying slopes in the multilevel regression), plotted against average state incomes, for each election year. In most years (especially the early ones), the income effect was smaller for richer states than it was for poorer states. In the poorer states, the richer counties used to strongly support the Republicans but this has mostly disappeared by 2000.

4 Modeling state-level differences in individual-level patterns

4.1 Reconciling the individual and aggregate results

Many observers have been misled by the seemingly contradictory pattern of richer states supporting Democrats but richer voters supporting Republicans. As we have seen with our hierarchical

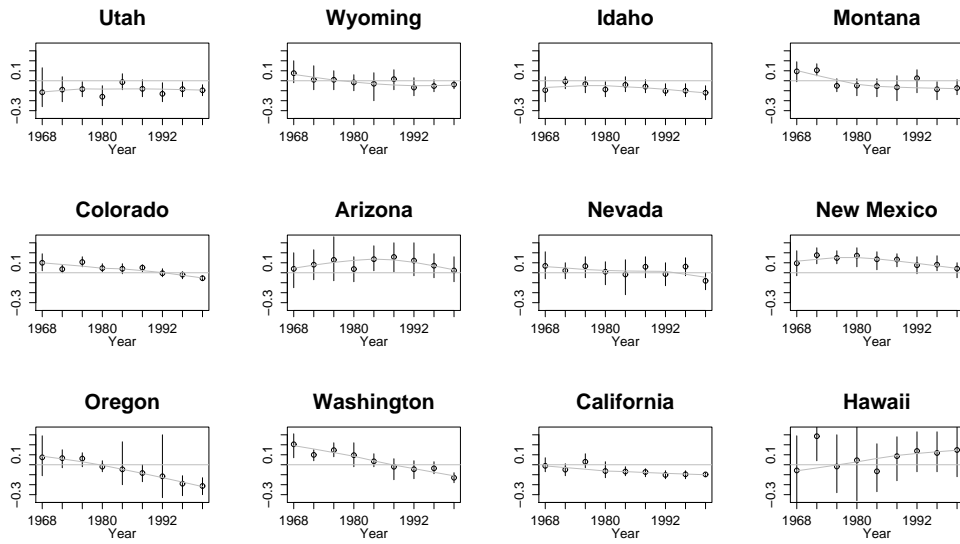


Figure 7: For the Western states: coefficients of county-level income from multilevel varying-slope model predicting the county-level Republican share of the presidential vote within states, 1968–2000. Within the region, the states are in decreasing order of support for Bush in 2000. There is no consistent pattern; however we notice that richer counties tend to support the Democrats in California, a state that gets much media attention.

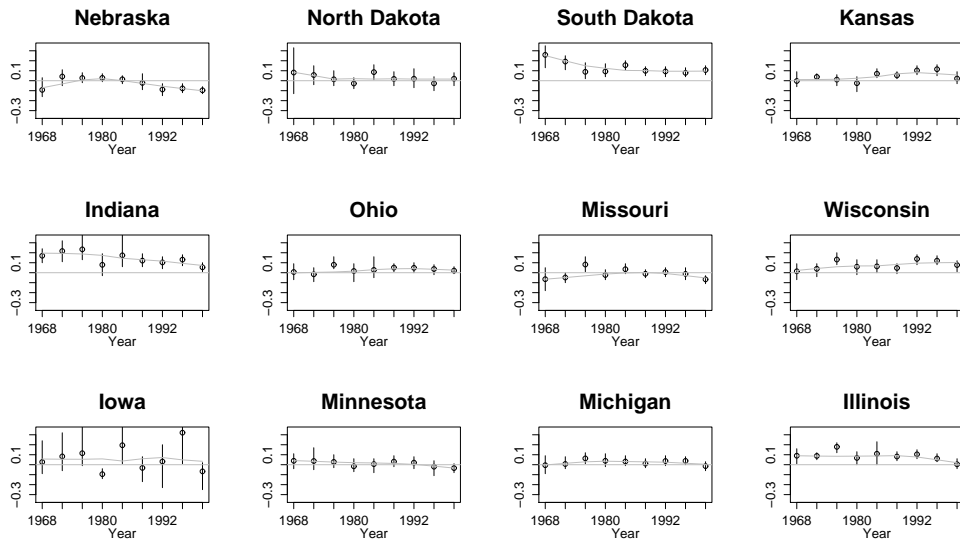


Figure 8: For the Midwestern states: coefficients of county-level income from multilevel varying-slope model predicting the county-level Republican share of the presidential vote within states, 1968–2000. Within the region, the states are in decreasing order of support for Bush in 2000. There is no consistent pattern.

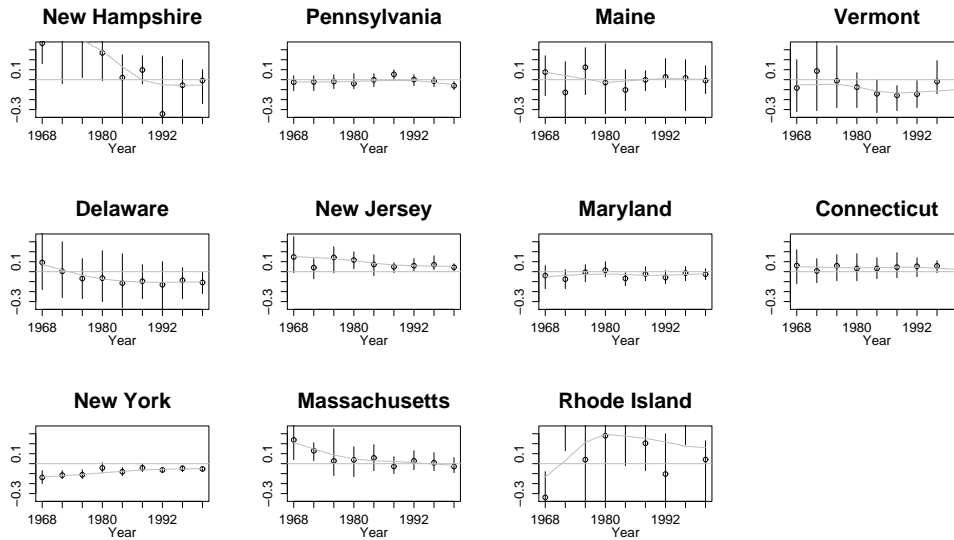


Figure 9: *For the Northeastern states: coefficients of county-level income from multilevel varying-slope model predicting the county-level Republican share of the presidential vote within states, 1968–2000. Within the region, the states are in decreasing order of support for Bush in 2000. The patterns are not strong, but in the states nearest the major media and political centers—New York and Maryland—the richer counties tend slightly to support the Democrats. (Because the analysis are at the county level, standard errors are larger in states such as Rhode Island and Delaware that have few counties.)*

model, richer voters support Republicans within states as well as overall; thus direct comparisons of voters for the two parties do not fit the “red-blue” stereotype. However, the income and voting differences between “red” and “blue” states are real.

To better visualize this puzzling pattern, we construct a graph that simultaneously displays variation within and between states. Figure 11 shows three lines, representing the probability of support for Bush in 2000 for each of the five income categories in each of three states—Connecticut (the richest state, which supported Gore), Ohio (an intermediate state, which was closely contested), and Mississippi (the poorest state, which supported Bush). The three lines show the estimated probability from the multilevel logistic regression (the lines are, in fact, portions of logistic curves, shifted by different amounts corresponding to the varying intercept in the model).

Figure 11 shows a statistical resolution of the red-blue paradox. Within each state, income is positively correlated with Republican vote choice, but average income varies by state. For each of the three states in the plot, the open circles show the relative proportion of households in each income category (as compared to national averages), and the solid circle shows the average income level and estimated average support for Bush in the state. The Bush-supporting states have more lower-income people, and as a result there is a negative correlation between average state income

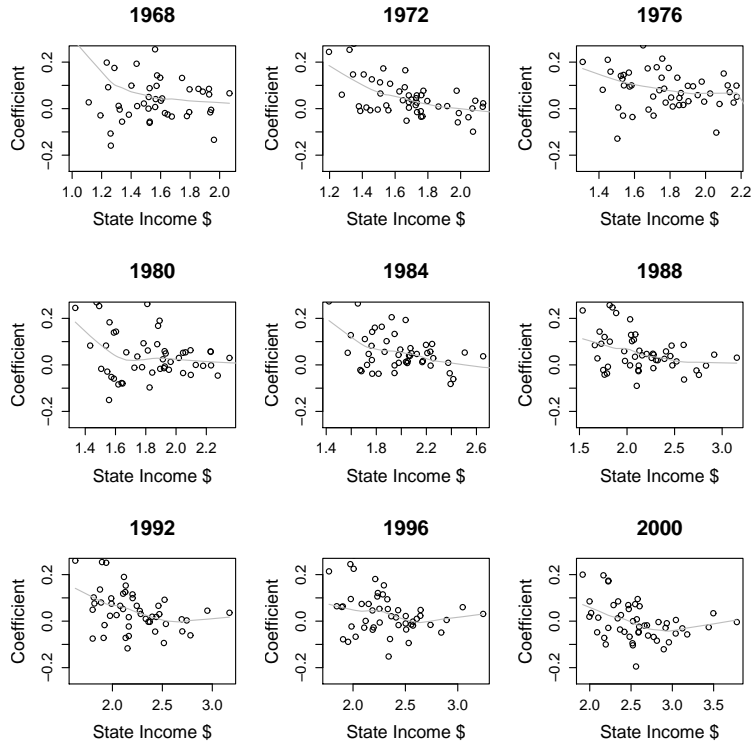


Figure 10: *Coefficient of county income within state, plotted vs. average state income, in each election year from 1968–2000. For poor states, the coefficient of county income tends to be positive (that is, the Republicans doing better in richer counties within these states), but for richer states the average coefficient of county income tends to be near zero.*

and state support for Bush, even amid the positive slope for each state. The poor people in “red” (Republican-leaning) states tend to be Democrats; the rich people in “blue” (Democrat-leaning) states tend to be Republicans. Income matters, but geography also matters. Individual income is a positive predictor, and state average income is a negative predictor, of Republican presidential vote support.

4.2 Varying-intercept, varying-slope model

As we have just seen, the varying-intercept multilevel model allows us to understand the positive correlation of individual income with Republican support, in the context of countervailing patterns between states. Our next step is to allow the relation between income and voting to vary by state. We fit a multilevel varying-intercept, varying-slope logistic regression:

$$\Pr(y_i = 1) = \text{logit}^{-1}(\alpha_{s[i]} + \beta_{s[i]}x_i), \text{ for } i = 1, \dots, n, \quad (2)$$

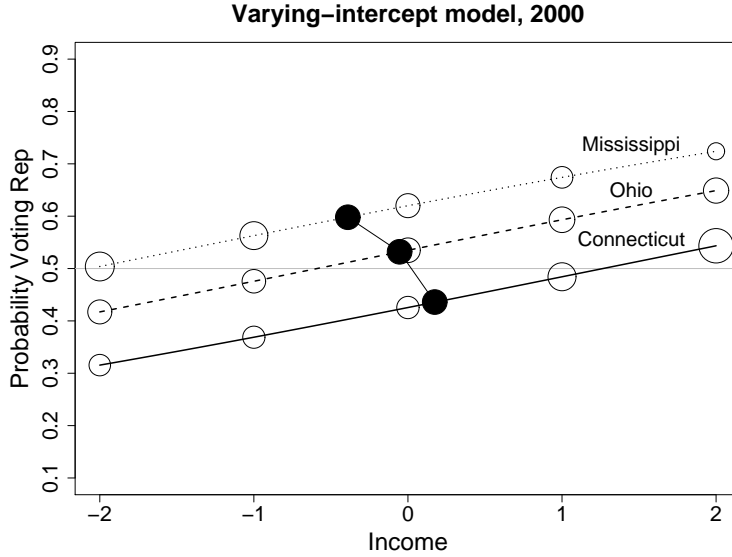


Figure 11: *The paradox is no paradox.* From the multilevel logistic regression model for the 2000 election: probability of supporting Bush as a function of income category, for a rich state (Connecticut), a medium state (Ohio), and a poor state (Mississippi). The open circles show the relative proportion (as compared to national averages) of households in each income category in each of the three states, and the solid circles show the average income level and estimated average support for Bush for each state. Within each state, richer people are more likely to vote Republican, but the states with higher income give more support to the Democrats.

where, as in (1), x_i is respondent i 's income (on the -2 to $+2$ scale). The state-level intercepts and slopes that are themselves modeled given average state incomes u_s :

$$\begin{aligned}\alpha_s &= a^\alpha + b^\alpha u_s + \epsilon_s^\alpha, \text{ for } s = 1, \dots, 50 \\ \beta_s &= a^\beta + b^\beta u_s + \epsilon_s^\beta, \text{ for } s = 1, \dots, 50,\end{aligned}\quad (3)$$

with errors $\epsilon_s^\alpha, \epsilon_s^\beta$ having mean 0, variances $\sigma_\alpha^2, \sigma_\beta^2$, and correlation ρ , all estimated from data. By including average income as a state-level predictor, we are not requiring the intercepts and slopes to vary linearly with income—the error terms ϵ_s allow for deviation from the model—but rather are allowing the model to find such linear relations to the extent they are supported by the data. We fit the model to the 2000 Annenberg Election Survey, as its huge sample size allows precise estimates of slopes in individual states.

From this new model, we indeed find strong variation among states in the role of income in predicting vote preferences. Figure 12 recreates Figure 11 with the estimated varying intercepts and slopes. As before, we see generally-positive slopes within states and a negative slope between states. What is new, though, is a systematic pattern of the within-state slopes, with the

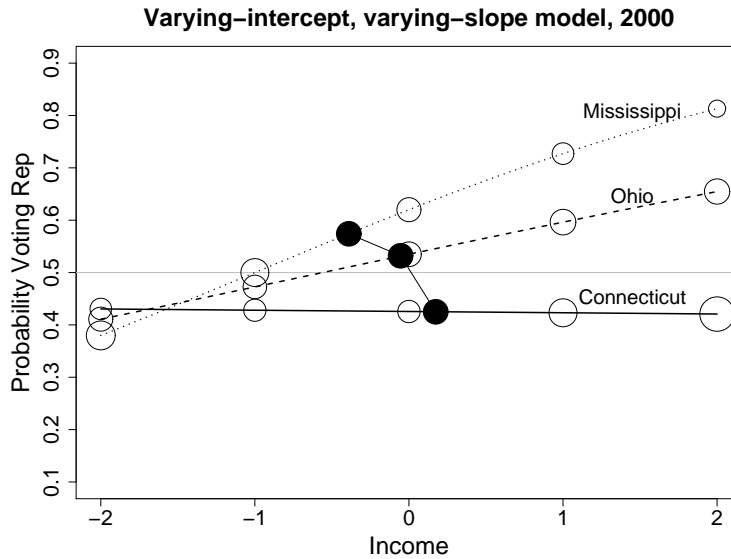


Figure 12: From the multilevel logistic regression model with varying intercepts and slopes for the 2000 election: probability of supporting Bush as a function of income category, for a rich state (Connecticut), a medium state (Ohio), and a poor state (Mississippi). The open circles show the relative proportion (as compared to national averages) of households in each income category in each of the three states, and the solid circles show the average income level and estimated average support for Bush for each state. Income is a very strong predictor of vote preference in Mississippi, a weaker predictor in Ohio, and does not predict vote choice at all in Connecticut. See Figure 13 for estimated slopes in all 50 states, and compare to Figure 11, in which the state slopes are constrained to be equal.

steepest slope in the poorest state—Mississippi—and the shallowest slope in the richest state—Connecticut.

Figure 13 shows the estimated slopes for all fifty states and reveals a clear pattern, with high coefficients—steep slopes—in poor states and low coefficients in rich states. Income matters more in “red America” than in “blue America.” Or, to put it another way, being in a “red” or “blue” state matters most for rich voters.

The large sample size of the Annenberg survey makes it easy to estimate a varying-slope model. However, the survey is new and does not go back into history. To see how varying state income effects have changed over time, we turn back to the NES. In Figure 14, we replicate Figure 12 for the years 1968–2000. The results for any given election are less stable (compare, for example, the lower-right graph in Figure 14 to the Annenberg estimates in Figure 12) but the trends over time are clear. The generally positive slopes within states have persisted for decades, but only since 1992, and especially since 1996, have systematic differences between rich and poor states

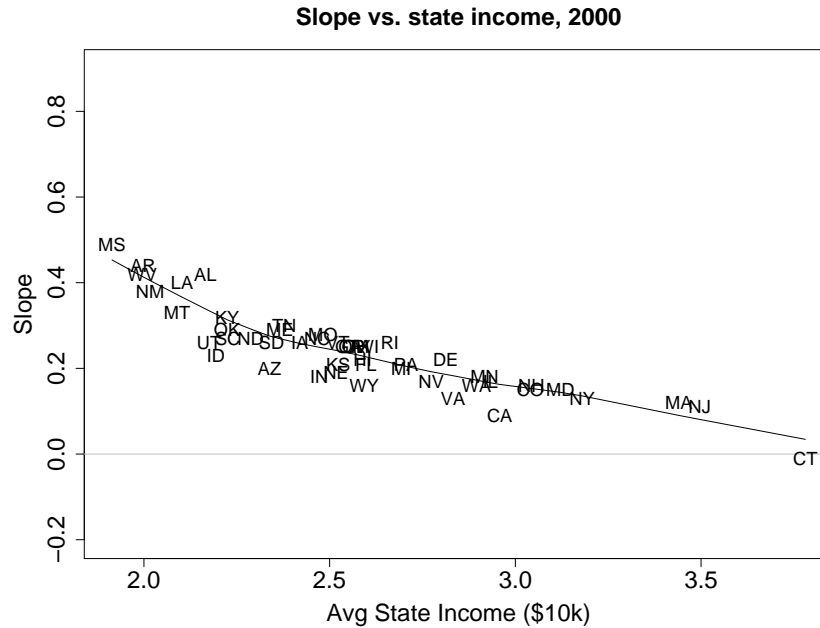


Figure 13: *Estimated coefficient for income within state plotted vs. average state income, for the varying-intercept, varying-slope multilevel model fit to the Annenberg survey data from 2000.*

become so clear.⁸

Figures 15 and 16 show the estimated intercepts α_s and slopes β_s in (2) as a function of average state income for each presidential election year since 1968. The estimates vary from year to year, but we again see the strongest patterns in 1996 and 2000. Since 1996, poor states have been consistently more Republican, and the coefficients for income have been higher in these states.⁹

Finally, we get similar results when replicating the 2000 Annenberg results—Figures 12 and 13—with exit polls from the 2000 and 2004 elections. Income and Republican voting are positively correlated within states, negatively correlated between states, and the slope for income is higher

⁸A problem with fitting state-by-state models here is that the NES uses cluster sampling (see the footnote on page 11), and so technically model 2, which assumes independence among data within a state, is inappropriate. Essentially, we must interpret the resulting estimates for each state as applying to the selected clusters rather than to the entire state. This cluster sampling is one reason that we emphasize, in the caption to Figure 14, that the state names in the plot represent rich, middle-income, and poor states, rather than specifically applying to Mississippi, Ohio, and Connecticut in each year. We trust the general patterns, however, because we are interested in the general patterns of income and vote preference comparing rich and poor states. NES samples clusters at random, so we are still getting representative selections of these states.

⁹The multilevel model shrinks the state estimates toward the estimated group-level regression lines (3). In a year such as 2000 where intercepts and slopes are shrunk very strongly towards the fitted lines (see the lower-right plot in Figures 15 and 16), this does not mean that we are certain that all 50 states fall along these lines, but rather that the data are *consistent* with the fitted lines, and the multilevel model finds this pattern. That is, we believe there is a strong (negative correlation) between intercept and average state income, and between slope and average state income, even though any of the particular states might not fall exactly on these fitted lines.

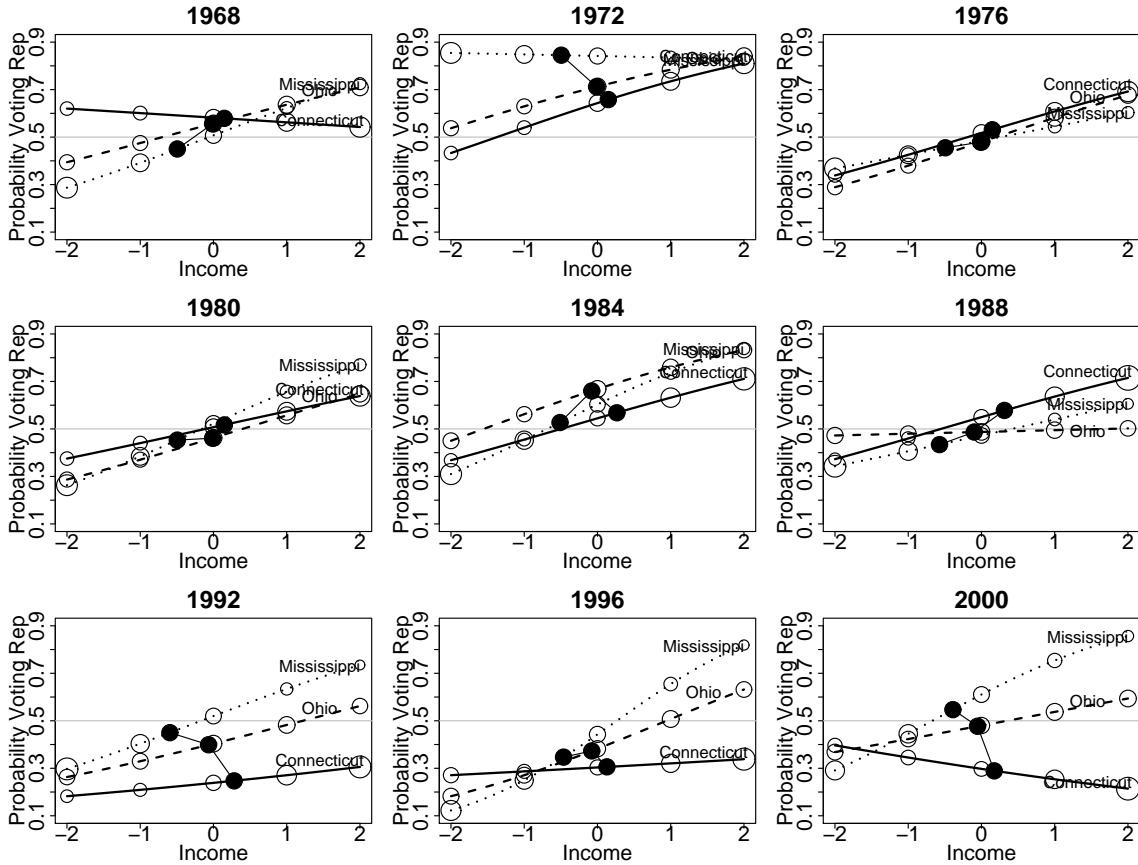


Figure 14: Using NES data: 1968–2000 results for the varying-intercept, varying-slope multilevel logistic regression: probability of supporting Bush as a function of income category. In each year, “Mississippi,” “Ohio,” and “Connecticut” represent rich, middle-income, and poor states, respectively. Differences between rich and poor states became much more consistent during the 1990s.

in poor states and lower in rich states. The exit poll analysis for 2004 gave similar results to 2000, which was no surprise given that 2004 was in many ways a replay of 2000.

We performed some model checking with both the Annenberg and exit polls, comparing individual states to the fitted models. A natural concern is nonlinearity or even nonmonotonicity in the relation between income and Republican voting, either in aggregate or within states. We did not find any serious departures from approximate linearity, and binned residual plots (Gelman et al., 2000) did not reveal any problems with the fitted logistic regression model. (In contrast, had we stopped at the varying-intercept model shown in Figure 11, we would have found big problems with the model fit, most notably in the extreme income categories in the richest and poorest states.)

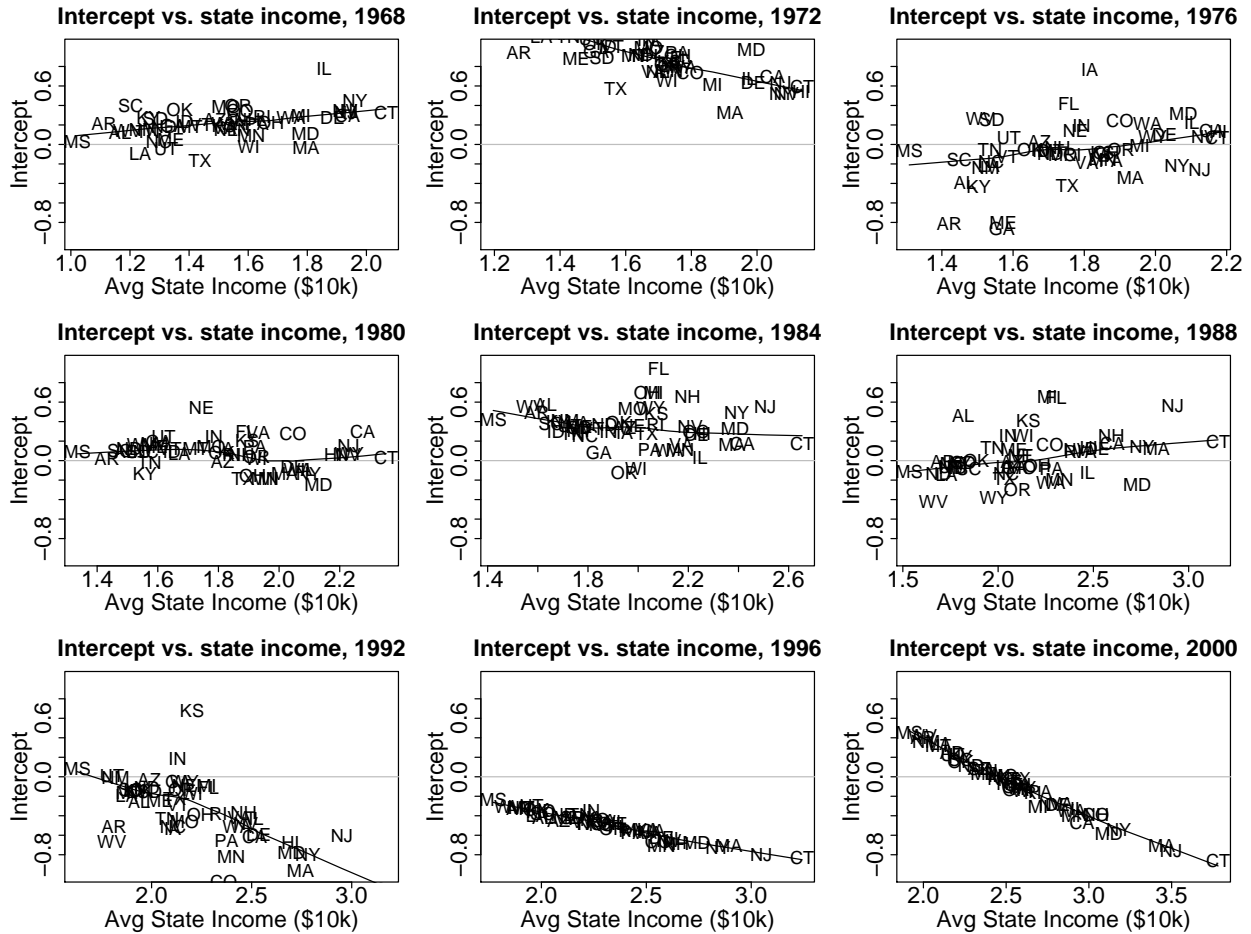


Figure 15: For the varying-intercept, varying-slope logistic regressions of Republican presidential vote preference on individual income: estimated state intercepts plotted vs. average state income. Models fit separately to NES data from each election year. Because of the small sample sizes in individual states, the multilevel model tends to pull the states toward the estimated regression line. Thus, in 2000, for example, we do not believe that the true values for all the states fall so close to a line, but rather that with the available data, these values are consistent with that estimated line.

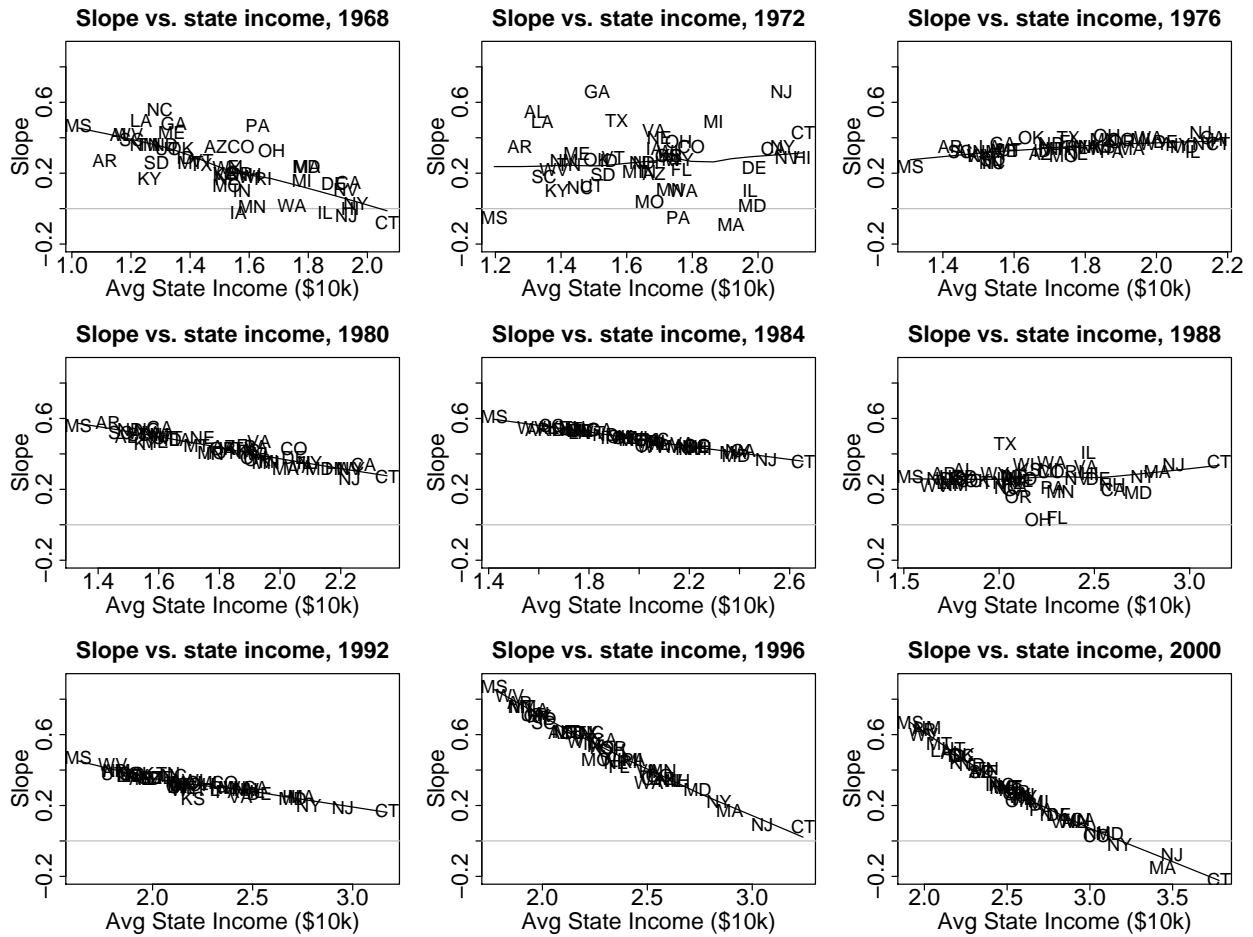


Figure 16: For the varying-intercept, varying-slope logistic regressions of Republican presidential vote preference on individual income: estimated state slopes plotted vs. average state income. Models fit separately to NES data from each election year. See caption of Figure 15 for more discussion.

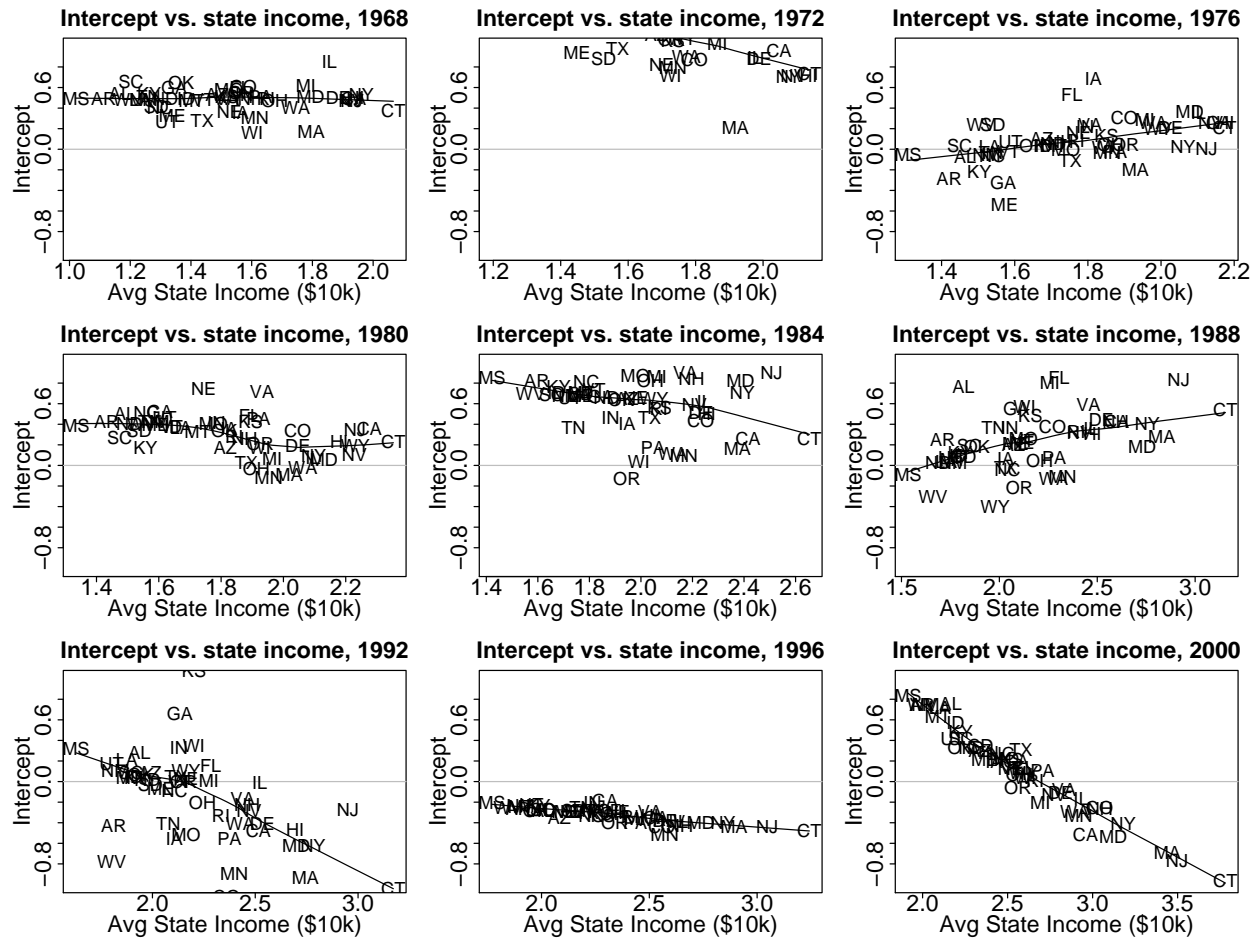


Figure 17: Intercepts for the varying-intercept, varying-slope logistic regressions fit to NES data for each year, excluding African-Americans. The results are similar to the estimates in Figure 15 from the model fit to all the respondents.

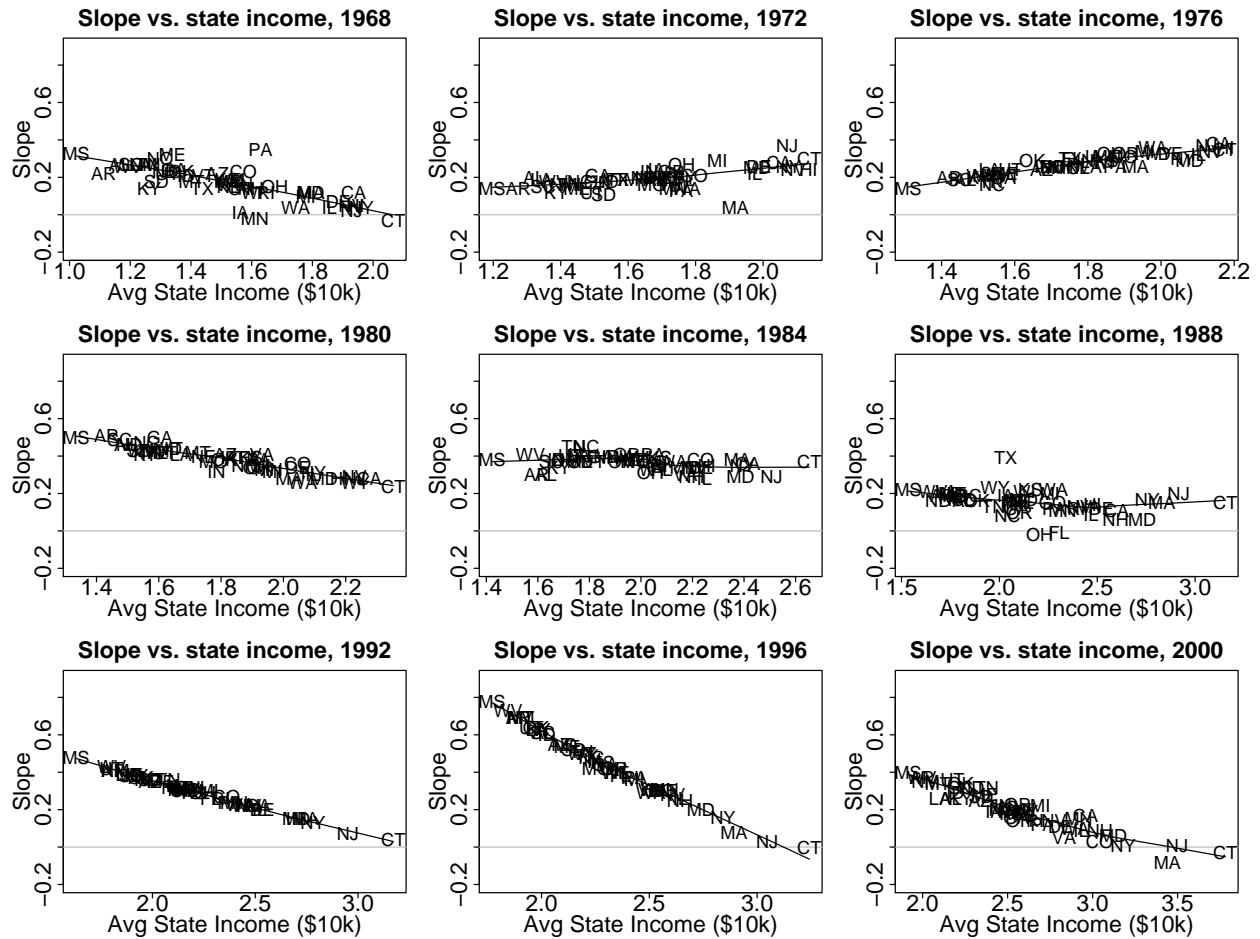


Figure 18: Slopes for the varying-intercept, varying-slope logistic regressions fit to NES data for each year, excluding African-Americans. The results are similar to the estimates in Figure 16 fit to all the respondents.

4.3 Ethnicity and other demographic variables

Could the varying income effects we have shown be merely a proxy for race? This is a potentially plausible story. Perhaps the high slope in Mississippi reflects poor black Democrats and rich white Republicans, while Connecticut's flatter slope arises from its more racially homogeneous population. To test this, we replicate our analysis, dropping all African-American respondents. The results (displayed as Figures 17 and 18) are similar to those of Figure 14, implying that the income and state patterns since 1992 can *not* be explained by differential racial mixes in rich and poor states.

To see if the income effects could be explained by other demographic variables, we went back to the full dataset for the Annenberg surveys in 2000 and added individual-level predictors for female, black, four age categories, and four education categories; and group-level predictors for percent black and average education in each state. After controlling for all these, the patterns for income remained: within states, the coefficient for individual income on probability of Republican vote was positive, with steeper slopes in poorer states; after controlling for the individual and group-level predictors, richer states supported the Democrats.

Our varying-intercept, varying-slope model has thus redefined the puzzle: in asking why the patterns within states differ from those between states, we are specifically interested in why slopes have become so shallow in rich states—that is, what's the matter with Connecticut? We have found that the differences between rich and poor states have become much more prominent in the past ten years, that they cannot simply be explained by race, and that they cannot be explained by the set of demographic variables that are typically used in adjusting survey respondents.

This is not to say that income is “causing” support for Republicans (or that such a causal relation is stronger in Mississippi than in Connecticut), but rather that richer voters within any state are more likely to support the Republicans, even after controlling for basic demographic variables—and this pattern is strong in the poor states but weak in rich states.

5 Discussion

5.1 Explaining the differences between states

As summarized in Figures 12–16, our multilevel analysis reveals three patterns that we would like to understand:

1. Voters in richer states support the Democrats—even though, within any given state, richer voters tend to support the Republicans.

2. The slope within a state—the pattern that richer voters support the Republicans—is strongest in poor, rural, Republican-leaning “red” states and weakest in rich, urban, Democrat-leaning “blue” states.
3. These patterns have largely arisen in the past ten or fifteen years.

We have no conclusive explanations for these patterns—our contribution is to discover and highlight them—but we can consider some ideas. First, the positive slopes within states are no surprise—given both the history and the policies of the two parties, it makes sense that the Democrats would do better among the poor and the Republicans among the rich, a pattern that has persisted for decades. At the same time, votes are far from being determined economically—even in Mississippi, which is the state with strongest correlation between income and voting, over 30% of voters in the lowest income category support the Republicans. Income is one of the many factors contributing to voters’ ideological and partisan worldviews, and one could, for example, use detailed survey data to try to understand individual-level positive correlation of income and Republican vote choice as coming from differential attitudes toward redistribution, as discussed by McCarty, Poole, and Rosenthal (2005).

Also interesting are the recent differences between rich and poor *states* that have gone in the other direction. Having ruled out the most obvious explanation—that rich and poor states represent the preferences of rich and poor voters—it makes sense to consider systematic differences between states, which are particularly interesting given the increasing mobility of Americans, the possibilities of self-stratification in exposure to news media and choosing where to live, and the increasing polarization of states and counties (Klinkner, 2004).

For example, state average income could be a proxy for secularism or some kind of cosmopolitanism. In other words, the cultural or social conservatism of states may be increasingly becoming negatively correlated with average income. At the same time, if these social issues are increasingly important to voters (perhaps made more salient by Clinton’s scandals, as suggested by Fiorina, Abrams, and Pope, 2005), this would induce changes in the relation between state income and individual vote. It would be interesting to study the relation between income and factors such as church attendance in different states.

Or, to put it another way, economic issues might well be more salient in poorer states such as Mississippi, and so one would expect voting to be more income-based. Conversely, in richer states like Connecticut, voters are more likely to follow non-economic cues. (These issues are raised by Ansolabehere, Rodden, and Snyder, 2005, although without the focus on comparing rich to poor states.) In any case, a challenge for explanations of this sort is to understand why they become more relevant in the 1990s. Journalists have also picked up on this time period as pivotal in voters’

changing perceptions of the two parties (see, for example, Marlantes, 2004, and Bishop, 2004), and these perceptions are increasingly important as the lens through which voters view political and economic issues (Bafumi, 2004). As discussed by Fiorina, Abrams, and Pope (2005), diverging ideological positions of the parties can lead to diverging attitudes about the parties among voters, even if the voters themselves remain largely centrist.

5.2 Methodological issues

We briefly review the statistical methods that made this work possible. Multilevel regression allowed us to estimate separate patterns within each of the 50 states, even for NES data with few respondents per state in any given year. The multilevel models also allowed us to include state average income as a predictor while also allowing unexplained state-level variation. Bayesian inference allowed us to simultaneously estimate the varying intercepts and slopes along with group-level regression coefficients, variances, and error correlations. For the varying-intercept, varying-slope models, the transformation of income on to a -2 to $+2$ scale allowed a more direct interpretation of state intercepts.

For all the analyses, both simple and complex, we gained important insights by replicating over several election years, making use of the availability of NES data. Although obvious, this sort of replication is not always done, and when it is done, the resulting pile of analyses can seem too overwhelming to display. Using time-series plots of data summaries and parameter estimates (as in many of the figures here) and repeated graphs (also called “small multiples”; see Bertin, 1967/1983 and Tufte, 1990), we were able to make these patterns clear, in a way that would be difficult using tabular representations (see Gelman, Pasarica, and Dodhia, 2002).¹⁰

Finally, Figures 11, 12, and 14 (which are related to plots developed for examining variation in medical statistics; see Baker and Kramer, 2001, and Wainer, 2002) simultaneously display variation within and between states that would be difficult to see simply by studying regression coefficients. These plots, along with the multilevel models themselves, allowed us to complement demographic and regional analyses such as in McCarty, Poole, and Rosenthal (2002, 2005), Ansolabehere, Rodden, and Snyder (2005), Bartels (2005), by examining state-level patterns.

¹⁰The method of repeated modeling, followed by time-series plots of estimates, is sometimes called the “secret weapon” because it is so easy and powerful but yet is rarely used as a data-analytic tool. We suspect that one reason for its rarity of use is that, once one acknowledges the time-series structure of a dataset, it is natural to want to take the next step and model that directly. In practice, however, there is a broad range of problems for which a cross-sectional analysis is informative, and for which a time-series display is appropriate to give a sense of trends. In our example, the secret weapon allows us to see how cross-sectional estimates for individual states and the entire U.S. vary over time. Expanding our multilevel models to include time series would be a major research undertaking that would require evaluation of additional time-series assumptions that are peripheral to our substantive research goals here.

5.3 The perils of summarizing categories by “typical” members

As a result of the electoral college system and also, perhaps, because of the appeal of colorful maps, state-level election results are widely presented and studied. After seeing the pattern of richer coastal states supporting the Democrats and poorer states in the south and middle of the country supporting the Republicans—a pattern that has intensified in recent years (see Figure 1)—it is natural to “personify” the states and assume that the Democrats have the support of richer voters too. Psychologists have studied the human tendency to think of categories in terms of their “typical” members; for example, a robin and a penguin are both birds, but robins are perceived of as “typical” members of the bird category and penguins are not (Rosch, 1975, Rosch and Mervis, 1975). When looking at the electoral map, commentators are misled by the patterns in red and blue *states* into thinking of typical Republican and Democratic *voters* as having the characteristics of these states.¹¹

If we had to pick a “typical Republican voter,” he or she would be an upper-income resident of a poor state, and the “typical Democratic voter” would conversely be a lower-income resident of a rich state. But these are more subtle concepts, not directly readable off the red-blue map—and, in any case, we would argue that given the diversity among supporters of either party, choosing typical members is misleading.

The mistake of “typicality” occurs in many political variables; for example, voters in Democrat-leaning states have higher average education levels than voters in Republican-leaning states—but more educated voters tend to favor the Republicans. Income, however, is a key attribute because of its direct connection to policy outcomes such as taxing and spending, and also because the middle class, including those on the lower end of the economic scale, are part of the “mom and apple pie” cluster that is perceived as virtuous (an issue we discuss further in Section 5.5).

5.4 First-order and second-order availability biases

I can’t believe Nixon won. I don’t know anybody who voted for him. — attributed to¹² Pauline Kael, 1972

It evidently irritates many liberals to point out that their party gets heavy support from superaffluent “people of fashion” and does not run very well among “the common people.” — Michael Barone, 2005

¹¹Political scientists have also made the point that the division into “red” and “blue” is somewhat unnatural, considering that distributions of votes and issue preferences tend to be unimodal, with most voters, and most states, falling in the middle of the distribution (Ansolabehere, Rodden, and Snyder, 2005, Fiorina, Abrams, and Pope, 2005). Here we are making a slightly different point, which is that a “typical” Republican (or Democratic) state does not look like an aggregation of “typical” Republican (or Democratic) voters.

¹²According to Rubio (2004), Kael never actually said this.

In addition to the challenge of trying to summarize diverse groups by their “typical” members, journalists who compare Democrats and Republicans are subject to another cognitive illusion—the *availability heuristic*, which is the pattern of making judgments based on easily-remembered experiences rather than population data (Tversky and Kahneman, 1974).

In this case, we could speak of first-order and second-order availability biases. A national survey of journalists found that about twice as many are Democrats as Republicans (see Poynter Online, 2003, summarizing the work of Weaver et al., 2003). Presumably their friends and acquaintances are also more likely to support the Democrats, and a *first-order availability bias* would lead a journalist to overestimate the Democrats’ support in the population—as in the above quote that has been attributed to Pauline Kael.¹³

However, political journalists are well aware of the latest polls and election forecasts and are unlikely to make such an elementary mistake. However, they can well make the *second-order* error of assuming that the correlations they see of income and voting are representative of the population.¹⁴ Journalists are predominantly college graduates and have moderately high incomes (median salary in 2001 of \$44,000, compared to a national average of \$36,000; see Weaver et al., 2003)—so it is natural for them to think that higher-income voters such as themselves tend to be Democrats, and that lower-income voters whom they do not know are Republicans. Michael Barone, for example, although no liberal himself, probably knows many affluent liberal Democrats and then, from a second-order availability bias, imputes an incorrect correlation of income and Democratic voting to the general population.

In fact, a national survey of journalists finds a correlation between high income and support for the Democrats,¹⁵ which is consistent with the “latte Democrat, Nascar Republican” storyline although not representative of the country as a whole, where Republicans are, on average, richer than Democrats.

Another form of availability bias is geographic. As already noted in Section 3.4, the centers of national journalistic activity are relatively rich states including New York, California, Maryland, and Virginia. Once again, the journalists—and, for that matter, academics—avoid the first-order

¹³This sort of bias—unrealistically believing others to be like oneself—has been much studied in psychology is also called the *false consensus effect*; see Lee, Green, and House (1977) and Krueger and Clement (1994).

¹⁴We use the term “second-order” because this bias does not involve inference about a frequency (that is what we refer to as first-order, availability bias, e.g., thinking that muggings are more likely if you have been mugged, or thinking that cancer is rare because you don’t know anyone with cancer), but rather inference about a correlation (e.g., that richer people are more likely to vote for the Democrats). Correlation, or more precisely covariance, is a second moment in statistical terms ($E((x - \mu_x)(y - \mu_y))$), as compared with simple frequencies ($E(x)$) which are first moments. What we have termed the “second-order availability bias” is related to the systematic errors in estimation of covariation that have been found by cognitive psychologists (see, e.g., chapter 5 of Nisbett and Ross, 1980).

¹⁵For example, in the Weaver et al. (2003) survey, 37% of Democratic journalists reported incomes exceeding \$50,000, compared to only 24% of Republican journalists.

availability bias: they are not surprised that the country as a whole votes differently from the residents of big cities. But they make the second-order error of too quickly generalizing from the *correlations* in their states. As shown in Figures 6–9, richer counties tend to support the Democrats within the “media center” states but not, in general, elsewhere. And as shown in Figure 13, richer voters support the Republicans just about everywhere, but this pattern is much weaker—and thus easier to miss—within these states.

Much has been written in the national press about the perils of ignoring “red America” but these second-order availability biases may have done just that, in a more subtle way. At this point, our hypotheses about journalistic biases are purely speculative; however, we hope these ideas can lead to a clearer picture, not only of the correlations between income, voting, and other variables, but of public understanding of these correlations. Future work in this area could include further analysis of journalists’ beliefs and attitudes, along with studies of average citizens’ perceptions of Democrats and Republicans, and how these perceptions differ by state.

5.5 Representation, ideology, and authenticity

I come from Huntington, a small farming community in Indiana. I had an upbringing like many in my generation—a life built around family, public school, Little League, basketball and church on Sunday. My brother and I shared a room in our two-bedroom house. — Dan Quayle, 1992

Clinton displays almost every trope of blackness: single-parent household, born poor, working-class, saxophone-playing, McDonald’s-and-junk-food-loving boy from Arkansas. — Toni Morrison, 1998

Income is not the driving factor in politics in the United States. However, income is important in political perceptions and is also clearly relevant to a wide range of policies including minimum wage regulation, tax rates, Social Security, etc., and is also correlated with many measures of political participation (Verba, Schlozman, and Brady, 2005). Similarly, geography is not an all-important factor in politics: red/blue maps of elections are appealing, but most of the states are not far from evenly divided. But, once again, geography is highly relevant to decisions on government spending, among other policies.

As the above quotations illustrate, both income and geography are relevant to politicians’ claims of authenticity, just as the income and geography of a candidates’ supporters are used to signify political legitimacy. In the 2000 presidential election, richer states voted for the Democrat. The recognition of this fact, and especially this long-term trend, was correctly noted by prominent journalists and pundits like David Brooks. But they went a step too far by attributing properties of “red” and “blue” states to “red” and “blue” voters and constructing inappropriate pictures of typical Republicans and Democrats. The psychological notion of typicality and the second-order

availability bias discussed in Section 5.4 give us insight as to how journalists could make this error, and the ongoing issues of authenticity and legitimacy explain why this error can have political consequences.

Sociologists and political scientists such as Brooks and Brady (1999), Stonecash (2000), McCarty, Poole, and Rosenthal (2002), and Bartels (2005) have recognized that higher-income voters continue to support Republican candidates, and lower-income voters support Democrats (in fact, this trend has been increasing since the 1950s). They have shown less interest in state-level differences in preferences (with notable exceptions being the Erikson, Wright, and McIver, 1993, study of state opinions and state policies, and the comparison of party identification among rich and poor voters within states in McCarty, Poole, and Rosenthal, 2005). As we have seen, state income is an important predictor of voting behavior in presidential elections, especially for people on the higher end of the income scale. Journalists' focus on red/blue maps has been somewhat misguided, but the differences between states are real, and indeed have changed in recent decades.

Geography matters politically. States are not merely organizational entities—mere folders that divide individuals for convenience. Nor are the differences cosmetic: a y'all here, a Hahvahd Yahd there. No—states have real, significant cultural and political differences. And despite the centripetal tendencies of a national media, drastically lower transportation costs, and a consumer economy frequently indistinguishable along regional lines (Starbucks everywhere)—they seem, if anything, to be getting more pronounced in the last decade or two.

To the extent political scientists want to understand political behavior in a federal system, we must recognize these differences. From a politician's perspective, given policies will be received differently in various states, even though those states are internally diverse. Therefore, an incentive to target policy geographically exists and has only gotten stronger. For policy analysts, then, increased attention to geography is also warranted.

The technique of multilevel modeling has allowed us to understand these patterns together. Individual income is positively correlated with Republican voting preference, but average state income is negatively correlated with aggregate state presidential voting for Republicans. The apparent paradox is *no paradox at all*, because Figure 12 clearly shows that these are not mutually exclusive relationships.

We can understand the state average income effect as one of context. The Mississippi electorate is more Republican than that of Connecticut; so much so that the richest segment of Connecticutians is only barely more likely to vote Republican than the poorest Mississippians. In poor states, rich people are very different from poor people in their political preferences. But in rich states, they are not.

References

- Alford, R. R. (1963). The role of social class in American voting behavior. *Western Political Quarterly* **16**, 180–194.
- Ansorablehere, S., Rodden, J., and Snyder, J. M. (2005). Purple America. Technical report, Department of Political Science, Massachusetts Institute of Technology.
- Bafumi, J. (2004). The stubborn American voter. Paper presented at the annual meeting of the American Political Science Association, Chicago.
- Baker, S. G., and Kramer, B. S. (2001). Good for women, good for men, bad for people: Simpson's paradox and the importance of sex-specific analysis in observational studies. *Journal of Women's Health and Gender-Based Medicine* **2001**, **10**, 867–872.
- Bartels, L. M. (2005). What's the matter with *What's the Matter with Kansas?* Unpublished manuscript, Department of Politics, Princeton University.
- Bertin, J. (1967, 1983). *Semiology of Graphics* (translation by W. J. Berg). University of Wisconsin Press.
- Bishop, B., ed. (2004). The great divide. *Austin American-Statesman*.
www.statesman.com/specialreports/content/specialreports/greatdivide
- Brooks, C., and Brady, D. (1999). Income, economic voting, and long-term political change in the U.S., 1952–1996. *Social Forces* **77**, 1339–1374.
- Brooks, D. (2001). One nation, slightly divisible. *Atlantic Monthly* **288** (5), 53–65.
- Buckley, W. F., and Bozell, L. B. (1954). *McCarthy and His Enemies*. Chicago: Regnery.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* **74**, 829–836.
- Datta, G. S., Lahiri, P., Maiti, T., and Lu, K. L. (1999). Hierarchical Bayes estimation of unemployment rates for the states of the U.S. *Journal of the American Statistical Association* **94**, 1074–1082.
- Erikson, R. S., Wright, G. C., and McIver, J. P. (1993). *Statehouse Democracy: Public Opinion and Policy in the American States*. Cambridge University Press.
- Filer, J. E., Kenny, L. W., and Morton, R. B. (1993). Redistribution, income, and voting. *American Journal of Political Science* **37**, 63–87.
- Fiorina, M. P., with Abrams, S. J., and Pope, J. C. (2005). *Culture War? The Myth of a Polarized America*. New York: Longman.
- Frank, T. (2004). Lie down for America. *Harper's Magazine* **308** (April).

- Gelman, A. (2003). Bugs.R: functions for running WinBugs from R.
www.stat.columbia.edu/~gelman/bugsR/
- Gelman, A., Goegebeur, Y., Tuerlinckx, F., and Van Mechelen, I. (2000). Diagnostic checks for discrete-data regression models using posterior predictive simulations. *Applied Statistics* **49**, 247–268.
- Gelman, A., and King, G. (1990). Estimating the electoral consequences of legislative redistricting. *Journal of the American Statistical Association* **85**, 274–282.
- Gelman, A., and Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology* **23**, 127–135.
- Gelman, A., Pasarica, C., and Dodhia, R. (2002). Let’s practice what we preach: using graphs instead of tables. *American Statistician* **56**, 121–130.
- Greenberg, D. (2003). *Nixon’s Shadow: The History of an Image*. New York: Norton.
- Issenberg, S. (2004). Boo-boos in paradise. *Philadelphia Magazine*, April.
- Klinkner, P. A. (2004). Red and blue scare: the continuing diversity of the American electoral landscape. *The Forum* **2** (2), article 2. Berkeley Electronic Press.
- Kramer, G. H. (1983). The ecological fallacy revisited: aggregate- versus individual-level findings on economics and elections, and sociotropic voting. *American Political Science Review* **77**, 92–111.
- Kreft, I., and De Leeuw, J. (1998). *Introducing Multilevel Modeling*. London: Sage.
- Krueger, J., and Clement, R. W. (1994). The truly false consensus effect: an ineradicable and ego-centric bias in social perception. *Journal of Personality and Social Psychology*, **67**, 596–610.
- Manza, J., and Brooks, C. (1999). *Social Cleavages and Political Change*. Oxford University Press.
- Marlantes, L. (2004). Inside red-and-blue America. *Christian Science Monitor*, July 14.
- McCarty, N., Poole, K. T., and Rosenthal, H. (2002). Political polarization and income inequality. Unpublished manuscript, Department of Politics, Princeton University.
- McCarty, N., Poole, K. T., and Rosenthal, H. (2005). *Polarized America: The Dance of Political Ideology and Unequal Riches*. Cambridge, Mass.: MIT Press, to appear.
- McGirr, L. (2001). *Suburban Warriors: The Origins of the New American Right*. Princeton University Press.
- Nisbett, R., and Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, N.J.: Prentice-Hall.
- Park, D, K., Gelman, A., and Bafumi, J. (2004). Bayesian multilevel estimation with poststratifica-

- tion: state-level opinions from national polls. *Political Analysis* **12**, 375–385.
- Perlstein, R. (2002). *Before the Storm: Barry Goldwater and the Unmaking of the American Consensus*. New York: Hill and Wang.
- Poyner Online (2003). The face and mind of the American journalist. Quicklink A28235, 10 April.
- R Project (2000). The R project for statistical computing. www.r-project.org
- Raudenbush, S. W., and Bryk, A. S. (2002). *Hierarchical Linear Models*, second edition. Thousand Oaks, Calif.: Sage.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review* **15**, 351–357.
- Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology* **7**, 532–547.
- Rosch, E., and Mervis, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology* **7**, 573–605.
- Ross, L., Green, D., and House, P. (1977). The false consensus effect: an egocentric bias in social perception and attribution process. *Journal of Experimental Social Psychology* **13**, 57–70.
- Rovere, R. H. (1959). *Senator Joe McCarthy*. New York: Harper and Row.
- Rubio, S. (2004). Kael/Nixon update.
begonias.typepad.com/srubio/2004/12/kaelnixon_updat.html
- Snijders, T. A. B., and Bosker, R. J. (1999). *Multilevel Analysis*. London: Sage.
- Spiegelhalter, D., Thomas, A., Best, N., Gilks, W., and Lunn, D. (1994, 2002). BUGS: Bayesian inference using Gibbs sampling. MRC Biostatistics Unit, Cambridge, England.
www.mrc-bsu.cam.ac.uk/bugs/
- Stonecash, J. M. (2000). *Class and Party in American Politics*. Boulder, Colo.: Westview Press.
- Stonecash, J. M. (2005). Scaring the Democrats: what's the matter with Thomas Frank's argument? *The Forum* **3** (3), article 4. Berkeley Electronic Press.
- Tufte, E. R. (1990). *Envisioning Information*. Cheshire, Conn.: Graphics Press.
- Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* **185**, 1124–1130.
- Verba, S., Schlozman, K. L., and Brady, H. E. (1995). *Voice and Equality*. Harvard University Press.
- Wainer, H. (2002). The BK-plot: making Simpson's paradox clear to the masses. *Chance* **15** (3), 60–62.
- Wattenberg, M. P. (1998). *The Decline of American Political Parties 1952–1996*. Cambridge, Mass.: Harvard University Press.

Weaver, D., Beam, R., Brownlee, B., Voakes, P. S., and Wilhoit, G. C. (2003). The American Journalist Survey. Indiana University School of Journalism.

Wright, G. C. (1989). Level-of-analysis effects on explanations of voting: the case of the 1982 U.S. Senate elections. *British Journal of Political Science* **19**, 381–398.