The Primacy of Race in the Geography of Income-Based Voting: New Evidence from Public Voting Records

Eitan D. Hersh   Yale University
Clayton Nall   Stanford University

Why does the relationship between income and partisanship vary across U.S. regions? Some answers to this question have focused on economic context (in poorer environments, economics is more salient), whereas others have focused on racial context (in racially diverse areas, richer voters oppose the party favoring redistribution). Using 73 million geocoded registration records and 185,000 geocoded precinct returns, we examine income-based voting across local areas. We show that the political geography of income-based voting is inextricably tied to racial context, and only marginally explained by economic context. Within homogeneously nonblack localities, contextual income has minimal bearing on the income-party relationship. The correlation between income and partisanship is strong in heavily black areas of the Old South and other areas with a history of racialized poverty, but weaker elsewhere, including in urbanized areas of the South. The results demonstrate that the geography of income-based voting is inseparable from racial context.

Why are people in some places more likely to vote according to their income class? What characteristics of U.S. regions explain variation in the relationship between income and partisanship? While citizens' alleged "false consciousness" has been a topic of intrigue at least since Marx, recent work in behavioral social science has sought to explain deviations from income-based voting in states and counties. These explanations fall primarily under two schools of thought. The first, exemplified by Gelman et al. (2008), has focused attention on how the wealth and cosmopolitanism of a state corresponds to the income-partisanship relationship. For example, the relationship between income and partisanship is weaker in richer states than in poorer states. The second school has focused attention on race and its consequences for redistribution. Alesina and Glaeser (2004), for example, show that regional differences in support for redistributive social welfare policies and income redistribution are related to overall racial hetero-geneity, and other research shows that support for public goods is impeded by ethnic and racial fractionalization (Alesina, Baqir, and Easterly 1999). These two paradigms yield different hypotheses about the microlevel factors that affect the voting patterns of rich and poor citizens. Under the economic context paradigm, we might look to how the influence of nonmaterialist (or postmaterialist) preferences over issues such as religion, sexual rights, or environmental attitudes is responsible for observed differences. Under the racial context paradigm, phenomena such as racial threat, sorting, and historical migration patterns appear at the heart of the income-party relationship.

In this article, we sort out these paradigms by showing that racial context trumps economic context in structuring the geographic patterns of income and partisanship in the United States. Whatever microlevel processes underlie local differences in the income-party relationship, they are closely tied to racial context and relate to economic

context mostly because economics and race are tightly correlated. By narrowing the set of viable explanations for regional variation in income-based voting to those bearing on racial context, we explain substantial heterogeneity in the varying political affiliations of the rich and poor.

Capitalizing on new data sources that permit the study of voter behavior at much lower geographic levels than could previously be achieved using either survey data or aggregate-level election returns, we find support for the racial context paradigm. Just as opposition to redistribution has been predicted by white reaction to local racial heterogeneity (Alesina and Glaeser 2004), we expect that the relationship between income and support for the party of economic redistribution (the Democrats) will vary with local racial composition. We test this hypothesis using full-population data sets of all 73 million voters who register with one of the two parties in party registration states (aggregated in block groups), as well as a geographic file containing data on the 185,000 precincts from all 49 states that release precinct-level returns.

We find that the income-party relationship is explained, specifically, by the racial composition of *local* areas, which we operationalize using state house districts. In the predominantly white districts where most Americans live, the relationship between income and partisanship varies little with the wealth of the locality or the wealth of the state, nor does the relationship vary substantially by region. Regional differences in the relationship emerge only across the racially heterogeneous portions of states. In the predominantly urban districts in the Northeast with large proportions of African Americans, voters in affluent block groups are only slightly more Republican than voters in the least affluent block groups. However, in specific rural areas with high concentrations of minority poverty, particularly in the southern "Black Belt," the Rio Grande Valley, and California's Central Valley, the relationship between income and partisanship is stronger than would be expected otherwise. This holds not only because poor blacks and Latinos—who are overwhelmingly Democratic—are a large share of the population, but also because affluent whites living in these areas tend to register and vote for Republicans at higher rates. Alternative explanations for this pattern, including district-level indicators of religiosity and social conservatism, provide only minimal additional predictive value. Interregional differences in the income-party relationship are explained by differences among higher-income whites living in close proximity to poor racial minorities.

Tobler's Law, the Modifiable Areal Unit Problem, and Political Geography

Three related issues addressed by geographers and social scientists make it necessary to study the income-party relationship below state lines: the geographic clustering of behavior, the arbitrariness of political geography, and, combined, their influence on redistribution preferences.

First, Tobler's Law, an empirical regularity of human geography, tells us to expect local clustering of income-based voting. The widely quoted maxim states, "Everything is related to everything else, but near things are more related than distant things" (Tobler 1970, 236). Tobler's Law is consistent with much of the behavioral contextual effects literature, which has emphasized the importance of voter behavior within areas that are typically much smaller than states (Gimpel and Schuknecht 2004). While much of that literature relies heavily on the most convenient aggregate-level data (metro areas and counties), decades of scholarship have confirmed that political behavior is clustered in space. Tingsten (1937) shows that class-based voting was stronger where the Swedish working class was geographically concentrated. British political geographers of the 1960s and 1970s found more support for the Labour Party in high-density working-class neighborhoods than elsewhere (Butler and Stokes 1969, 146). Analyses that incorporate highly localized contextual data (at the tract or precinct level) often better predict behavior than those based on individual predictors alone (see, e.g., Gay 2006; Huckfeldt and Sprague 1995; Katz and Lazarsfeld 1955; Lazarsfeld 1948; Oliver 1999; Putnam 2007; Sampson, Raudenbush, and Earls 1997). When applied to contextual effects research, Tobler's Law suggests that more potent effects will appear when "place" is defined not as a state, or even as a county with hundreds of thousands or millions of people, but as much smaller contextual units.

Second, even as political scientists have engaged in contextual-effects research, little of this work has attended to an important methodological concern, the modifiable areal unit problem (MAUP; Fotheringham and Wong 1991), which explains that statistics are sensitive to how much individual units are aggregated (scale), and what lines are drawn to aggregate the units (zoning). Statistics obtained at the state or Census region level, or in urban and rural areas arbitrarily coded in the American National Election Studies and similar surveys, need not persist at other levels. Models that

assume the independent and identical distribution of the partisanship of demographically similar voters within states neglect the importance of substate geography and local clustering of voter behavior (Cho and Rudolph 2008). Indeed, caricatures of affluent, secular, and liberal blue-state voters and affluent, religious red-state voters only apply in some areas of states. The Black Belt, a rural area with a history of cotton cultivation by black slaves and sharecroppers, continues to have both high rates of overall Democratic voting with strong Republican voting among whites, but other areas of the "New South," including the Atlanta metropolitan area, look much like the rest of the country. Similarly, in the Northeast, affluent social liberals are concentrated in the metropolitan portions of the blue states, whereas voters elsewhere (e.g., upstate New York) look much like voters in other predominantly white, rural areas. The same can be said of areas in which the Republican Party holds sway, such as the Mormon Corridor, which covers only portions of the mountain states (Gimpel and Chinni 2011). In short, voter types are pertinent only within substate geographies, and analyses should account for this.

Tobler's Law and the MAUP suggest that directing attention to more local areas may resolve ongoing theoretical disputes over the geographic mechanisms underlying income-based voting. Gelman et al. (2007), building on work by Huber and Stanig (2007) and Inglehart (1990), suggest that "economic issues might well be more salient in poorer states" (74). In areas with above-average income, economic issues may be of lower salience, and politics in richer areas may be oriented toward nonmaterialist or "postmaterialist" issues. Related research on income inequality (Meltzer and Richard 1981; Romer 1975) and class bias in the electorate (Hill and Leighley 1992) demonstrates that political attitudes vary with *economic context*. Typically, in such studies, the chosen context is the state or nation-state, not one's more immediate political geography. The alternative view, that income-based voting is shaped by out-group threat and racial attitudes (e.g., Alesina and Glaeser 2004; Gilens 1999; Luttmer 2001), similarly has tended to examine variation within states or contextual effects within small areas, without analyzing variation across political geographies. Alesina and Glaeser (2004), for example, show that the extent of racial heterogeneity in a state predicts redistribution preferences.

Examining lower-scale geographic units enables us to decouple racial and economic context. Based on existing observational research on substate politics, we expect that districts' racial composition explains important differences in income-based voting. We expect few partisan differences between high- and low-income citizens in homogeneously white areas. We suspect

that particular groups of racially diverse places, not homogeneously white places, will exhibit the most variation in the strength of income-based voting. In such places, low-income voters—blacks, Native Americans, and, increasingly, Latinos—are consistently Democratic (Dawson 1995; Segura 2012). What distinguishes racially diverse areas from one another is the type of richer, disproportionately white, voters who live near geographic concentrations of poor minorities.

Data and Methods

We use two previously untapped data sources to analyze income-based voting at a low geographic level: a voter file containing a full census of party-registered voters in 29 states, and precinct-level results from 49 states. Past work on the political geography of income-based voting has relied primarily on two types of data: survey measures of self-reported income, party identification, and vote choice; or geographically coarse (usually, county-level) voting returns (Gelman et al. 2008). Because nationally representative surveys such as the American National Election Study, the National Annenberg Election Study, and the Cooperative Congressional Election Study (CCES) have insufficient samples for analysis within regions as low as state house districts, work based on these studies commonly aims to measure public opinion within states while sidestepping questions of *substate* geographic variation (Gelman and Little 1997; Lax and Phillips 2009; Park, Gelman and Bafumi 2004; Vigdor 2006). Scholars have developed novel approaches to leverage available survey data, including combination of data from multiple surveys and multilevel regression and poststratification (MRP) when data are sparse (Tausanovitch and Warshaw 2013). These approaches often accept strong modeling assumptions and rely on aggregate-level data to improve local estimates. They rarely entail multivariate analysis within congressional or state legislative districts (Tausanovitch and Warshaw 2013).[1]

While statistical models with strong distributional assumptions could be used to estimate the income-party relationship within small areas, they would rely strongly on sparse survey data that are rarely representative at the substate level. Although various surveys can be combined to yield a larger *n*, few of these surveys are designed to be representative at levels below the state or congressional district. Even in the presence of abundant data, contemporary surveys leave many substate areas unsampled or

---

[1] Other work has provided estimates of the voting behavior of racial and economic subgroups within states, also making geographic stationarity assumptions (Ghitza and Gelman 2013).

severely undersampled. Particularly when one wishes to analyze contextual effects, the consequences of low sample size and sampling bias become obvious. Even after the introduction of poststratification weights, national samples (especially online samples) tend to severely undersample the poor, rural, and minority areas such as the Black Belt and Rio Grande Valley. Counties in these areas typically have zero respondents, and survey-based inferences about the region could be based on only a few dozen respondents. Any inferences made about such areas would then depend on imputation of local values using data from other, quite dissimilar places, adopting a strong set of assumptions in the process.

Large national surveys are not designed to be representative at low levels of geography, and even the largest surveys are too small for statistical analysis at such fine granularity. As an illustration of this point, consider the geographic distribution of voters in the externally validated 2008 CCES compared with the national voter file. We focus on validated, registered voters residing in state house districts in which the majority of the population is nonwhite, in states that allow for party registration. The median number of respondents in such districts was four, and ten percent (31 of 313) of the districts had zero respondents. In addition, the block-group income of the average CCES respondent in these districts is considerably higher than for the average voter ($44,000 for CCES respondents versus $36,938 for the population of all registrants).

By using public voter registration records and precinct-level presidential election results, we aim to fill a gap in the existing literature on substate voting using data. We study voter behavior at the level of small-scale geographies, including block groups and precincts, permitting dramatically improved geographic resolution in analysis of voting behavior while accepting a modest trade-off by shifting the target of inference slightly away from the individual. We analyze data from two sources. The first is a voter file containing 73 million voter records from the 29 states that require party registration, assembled by Catalist, a leading vendor. We classify voters on the list according to their partisan registration, block-group racial composition, and block-group income.[2] The second is a record of the 2008 presidential election returns from 185,000 precincts compiled by the Harvard Election Data Archive. These data allow for estimation of the relationship between precinct-level income and the 2008 Republican presidential vote in the 49 states that report precinct-level

results (Ansolabehere and Rodden 2011). These data provide two different approaches to party affiliation. Registration numbers and election results both capture the strength of party affiliation. Party registration is a public expression of standing party affiliation, whereas presidential voting captures a less permanent party-affiliation decision associated with a particular election, even as in recent elections it has been highly correlated with the normal vote (Levendusky, Pope, and Jackman 2008).

The Catalist voter database permits us to directly enumerate (not estimate) the partisanship of all voters within block groups, according to block-group-level demographic characteristics. The academic version of the data links individual registrants to the Census block group of their residential addresses. The median household income of each registrant's block group in the 2000 Census is used as a proxy for individual-level income. In the Catalist data, block-group income is reported at $20,000 intervals, ranging from $20,000 or less to $200,000 (the top-coded value reported in the Census data). We also associate each voter with the racial composition of his or her block group.[3]

We analyze both the Catalist data and precinct-level data at the state house district (lower house) level for several reasons. State house districts are the lowest-level political geography with consistently partisan elections across all states, and they also happen to be the lowest-level office with influence over statewide economic policy. While the approximately 5,000 state house districts vary in size and population (with California's being much larger than other states' districts), across states they are functionally equivalent. They also must have approximately equal populations within states, making them a more desirable and comparable geographic unit than counties. To measure district-level income, we take the unweighted mean of the median household block-group income of registered voters in the district. We calculate each district's percent black and percent non-Hispanic white by averaging each group's percentage in two-party registrants' block groups, for each district. We are thus able to estimate the relationship between registered voters' partisanship (measured at the individual level), their block-group income, and district-level income and racial composition.

---

[2]These data are explained in more detail in Ansolabehere and Hersh (2012) and Hersh (2013, 2015).

[3]Individual-level race data are only available from voter files in some jurisdictions. Previous research has found that block-group demographics can yield accurate individual-level estimates of household income in homogeneous block groups (often in urban areas) but tend to be a less adequate proxy in rural and suburban areas, where block groups are larger in area and more heterogeneous. In the online supporting information (SI), we discuss the benefits of using block-group income and race as measures of economic class.

TABLE 1 Data Summary

	Catalist Data	HEDA Data
Individual Units	73,170,970 D. and R. registrants with block-group (BG) income and race	185,002 precincts with BG income and race
Aggregate Units	State house districts with aggregated BG income and race	Counties and state house districts with aggregated BG income and race
Coverage	29 party registration states	49 states with published data

While numerous scholars have opted to capture partisanship using voter registration, rather than presidential election returns (Abrams and Fiorina 2012; McGhee and Krimm 2009), we supplement our analysis with additional data from the Harvard Election Data Archive (HEDA; Ansolabehere and Rodden 2011), a geographic information system (GIS) database containing 2008 presidential election results from 49 states.[4] The precinct data are merged with block-group income and race data from the 2000 Census using a spatial join in ArcGIS.[5]

To test the relationship between racial and economic context and income-based voting, we adopt varying assumptions as we bore below state boundaries. We begin with the Catalist data. First, we nonparametrically summarize Republican registration by income group, examining only district partisanship by district income within states. These results suggest that racial, not income, composition is a likely explanation for observed differences in the income-party slope. Then we examine the relationship between block-group income and partisanship within state house districts within particular states using a direct enumeration approach, showing that district racial composition predicts much of the variation in income-based voting. We confirm these results using least squares regression on the same data. However, even

[4]The omitted state, Oregon, conducts only vote-by-mail elections and does not release precinct-level returns.

[5]Population data, including population by race, were obtained from the ESRI block-group layer (ESRI 2008). Block-group aggregate and median household income were obtained from the National Historical GIS database (Fitch and Ruggles 2003). Block-group polygon geographies were converted to points using the geometric centroid of each polygon, and data from all points within each precinct were averaged to generate the precinct-level data. Block groups from Oregon, the only state omitted from the HEDA data, were not used in this spatial matching procedure. After the Census data were aggregated by precinct, each precinct was spatially joined to the Census shapefile of 2006 state house districts and ESRI county shapefile (United States Bureau of the Census 2006). After this process, precincts' centroids were spatially joined to the district in which they fell. The resulting precinct-level data set contains 185,002 precincts distributed across districts and counties in the 49 precinct-data states. Summary statistics for these data appear in SI Table A.2. Table 1 summarizes the two data sources.

when using full-population data, the analysis is vulnerable to data sparseness due to rarely occurring types of districts (e.g., high-income, majority-minority districts). Adopting slightly stronger modeling assumptions, we estimate a series of hierarchical linear models in which the district-level income coefficients (both intercept and slope coefficients) are allowed to vary (Gelman and Hill 2007). We estimate district-specific income effects within state legislative districts in the 49 states with precinct data.

Racial Diversity and the Local Income-Party Relationship

To test the relative importance of income and racial context in the income-party relationship, we begin by analyzing aggregate data, plotting district-level partisanship against district-level income. Using the full data set of 29 party registration states, we plot, in Figure 1, the average household income and Republican percentage of each state house district. Following the income context hypothesis, we organize the party registration states into four approximately even groups based on state income, calculated by averaging median block-group household income for all voters in the state. District income is similarly calculated. We plot the Republican proportion of two-party registrants against the log of mean district income and then fit a lowess curve to each state's lower-house districts. We also plot a linear fit for each group of states, employing state fixed effects.

Figure 1 reveals four facets of the local income-party relationship that are missed in state-level analyses of individual-level survey data. First, the figure captures the large variance in district-level income within states. Average income in state house districts varies within states by as much as $130,000, whereas the range across the richest and poorest states in the party registration sample is only $35,000. Second, except in Oregon (highlighted in the upper-right plot) and Wyoming (not highlighted), richer districts are, on average, more Republican. Third, while

FIGURE 1   Regardless of State Income, Rich Districts Are Usually More Republican

District Partisanship by Income, Grouped by State Income
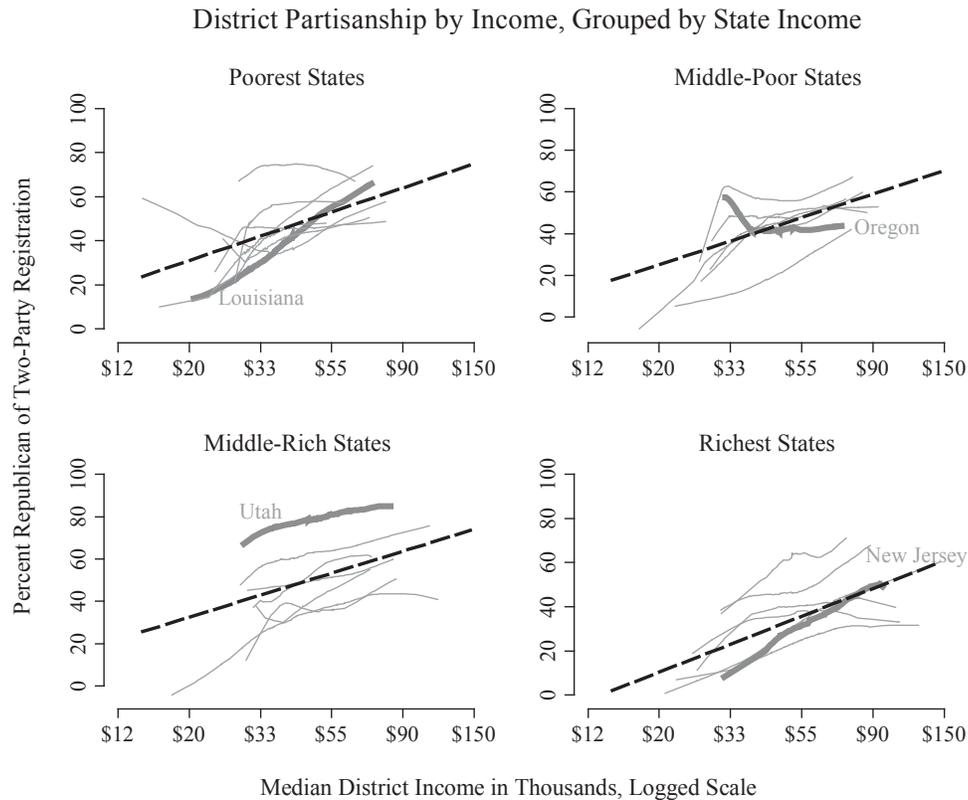
Note: Lowess fit lines are estimated based on the partisanship and income level of 2,575 state house districts in 29 party registration states, derived from Catalist data. District income is the average of the median household income of block groups in the district. The linear relationship between district income and partisanship for each grouping of states (using state fixed effects) is shown by a dashed line. Four illustrative states are highlighted. Upper left: KY, LA, ME, NM, OK, SD, WV, and WY. Upper right: IA, NC, FL, NE, OR, PA, and RI. Lower left: KS, AZ, DE, NV, NY, UT, and NH. Lower right: AK, CA, CO, CT, MA, MD, and NJ.

the relationship between district income and district partisanship varies across states, it appears unrelated to state-level income.[6] For instance, note that the dashed fit line shows a similar slope in each of the four plots.[7]

Fourth, while the income-party relationship appears unrelated to state-level income, race is an obvious lurking variable. Very Republican high-income districts and very Democratic low-income districts appear only in states with high-poverty minority areas. They include both rich states such as New Jersey and poor states such as Louisiana. Homogeneously white states evince only a modest income-party relationship in the aggregate

district-level data: States like Utah and Oregon (highlighted in the figure) as well as Wyoming, Iowa, and New Hampshire are less than 2% African American and have very minimal slopes. In other states, the relationship is nonlinear, and a few districts with concentrated poverty are the exception to an otherwise flat income-party relationship. A state's region and its history of racialized politics, not state-level income, appear to govern the strength of the income-party relationship at the aggregate district level.

To see this quite clearly, Figure 2 presents the income-party relationship for the same states, grouping them by percentage black rather than by income. Here, it becomes clear that the relationship between district income and partisanship is strong in states with larger proportions of African Americans, including both northern states like New Jersey and southern states like Louisiana, and weaker elsewhere. The difference in linear fit lines between

[6]This confirms a recent replication of Gelman et al. (2008) in Feller, Gelman, and Shor (2012) using 2012 election data.

[7]The slopes of the fit lines, with 95% confidence intervals, are 22 (CI: 18–26), 23 (CI: 19–26), 21 (CI: 17–25), and 25 (CI: 22–29), from poorest states to richest states.

FIGURE 2 The Relationship between District Income and Partisanship Is Strongest in States with Large Black Populations

District Partisanship by Income, Grouped by State Percent Black
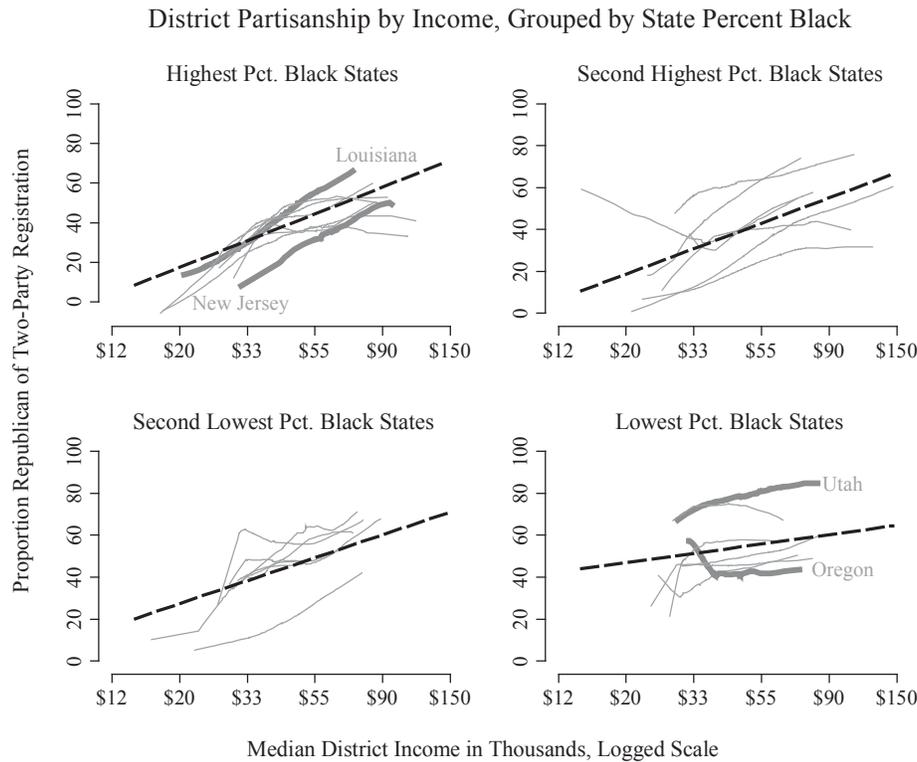
Note: Replication of Figure 1, with states organized by black population percentage. Four illustrative states are highlighted. Upper left: DE, FL, LA, MD, NC, NJ, NY, and PA. Upper right: CA, CT, KS, KY, MA, NV, and OK. Lower left: AK, AZ, CO, IA, NE, RI, and WV. Lower right: ME, NH, NM, OR, SD, UT, and WY.

Figures 1 and 2 provides a simple summary. In the first figure, the dashed fit lines are essentially indistinguishable across groups of states. In the second figure, the income-party slope is three times as large in the most black states versus the least black states.[8]

## Local Income, Racial Context, and Republican Support

We demonstrated that state income has little relationship to the district-level income-party relationship, and race seems to be a pertinent lurking variable. We now examine the income-party relationship at a geographic level much closer to the individual level, examining Republican registration rates among voters falling in each block-group income category, within state house districts and within states. Differences across *states* in the black pop-

[8]In Figure 2, the slopes of the fit lines, with 95% confidence intervals, are 27 (CI: 23–30), 24 (CI: 21–27), 22 (CI: 17–27), and 9 (CI: 4–13), from most black states to least black states.

ulation percentage conceal an even stronger contextual racial effect lurking below state lines. Unlike in Figure 2, where heavily black states like Pennsylvania and Louisiana looked alike in the aggregate income-party relationship, voters in rich and poor block groups in racially similar states behave very differently depending on the racial context of the local areas.

Consistent with the racial composition hypothesis, we expect that district racial composition, not state racial and economic composition, should explain differences in the block-group-level income-party relationship. Here, we employ the direct enumeration method described in the Data and Methods section. Within each state, district racial composition stratum, and block-group income level, we directly calculate the Republican proportion of two-party registrants. Figure 3 presents these nonparametric tests, demonstrating the importance of racial context in a lattice-style plot in which states are broken down by region and income. States are assigned to one of four regions: North, South, Midwest, and West. Districts within these states are placed into three groups based on

FIGURE 3 Across States, the Income-Party Relationship Is Linked to District-Level Racial Composition

Census Block-Group Income, in $10K
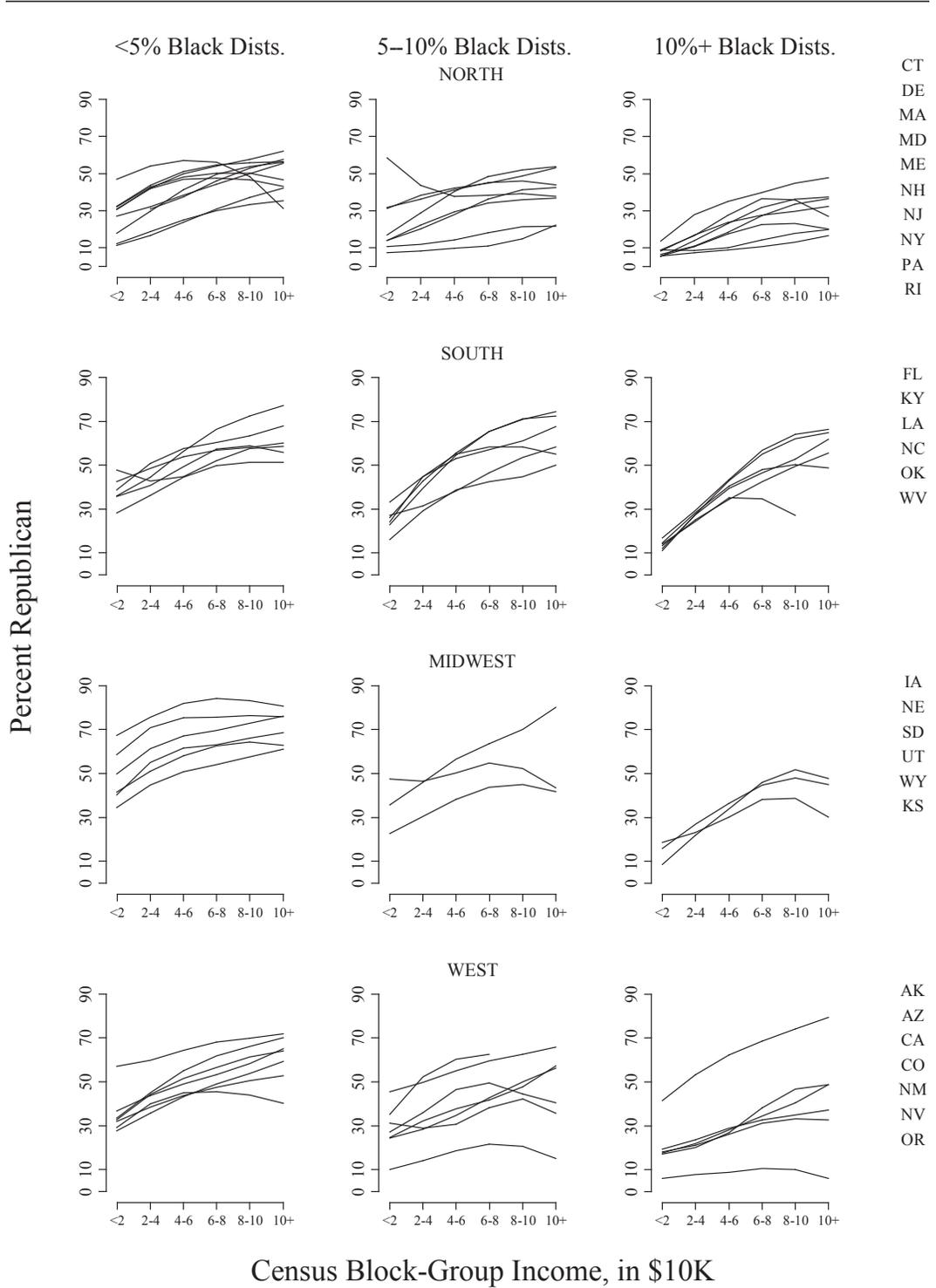
Note: Derived from the population of registered major party affiliates. N = 73,170,970. Each voter has an individual-level party affiliation and a block-group-level income group. The data include 108,532 block groups. Not all states have block groups in all racial-context categories.

FIGURE 4 The Effect of District-Level Income and Racial Context on the Bivariate Relationship between Block-Group Income and Voting, Inside and Outside the South

NONSOUTH   SOUTH

By District Income

By District Race
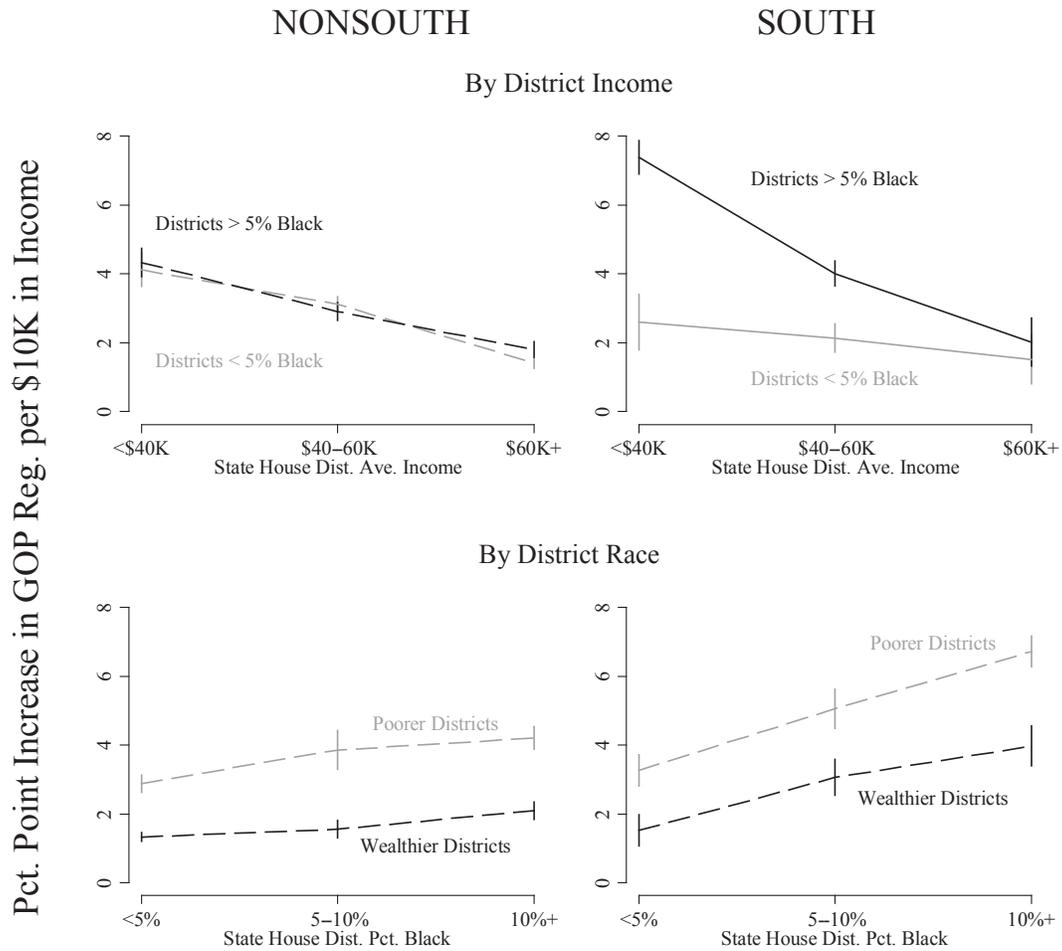
Note: Derived from a 1% simple random sample of registered major party affiliates from Catalist. N = 708,986. Each point is a slope coefficient estimate from a regression of Republican registration on block-group income with block-group clustered robust standard errors. On lower plots, poorer and wealthier districts are defined as those with average income less or greater than $50,000.

their racial minority composition. Since the most likely group explaining a strong income-party relationship is black voters, we classify districts according to black racial composition, organizing the data into districts that are up to 5%, 5 to 10%, and more than 10% black. In each district, the proportion of Republican registrants is estimated for two-party registered voters in each of six block-group-level income groups, ranging from $20,000 to $100,000, grouped in $20,000 intervals.[9] This exercise was done for each state, and a line graph connects the binned averages to represent the income-party relationship.

[9] These bins were designated by Catalist. We top-coded block groups with median household incomes above $100,000.

Income-based voting has generally been characterized in terms of region, with southern states identified as outliers (e.g., Feller, Gelman, and Shor 2012, 129). If southern voters were generally more likely to engage in income-based voting (Gelman et al. 2008), we would expect the lines across all three groups of districts in the South to have steeper slopes than districts elsewhere. Yet the differences between southern and nonsouthern states are most consistently pronounced only in districts of the South with high proportions of African Americans. In fact, looking down the first column of Figure 3, homogeneously nonblack districts have a weak income-party relationship across states. In such places, neither region nor state predicts the strength of income-based voting.

In districts with a larger proportion of African Americans, some regional differences emerge. In more African American districts in Louisiana, for example, fewer than 10% of voters in low-income block groups are Republicans, versus 70% of those in high-income block groups. By contrast, in the Midwest, the West, and especially the North, the relationship between block-group income and partisanship is much flatter in racially diverse districts. The poorest block groups in such districts are almost entirely Democratic, and affluent block groups in such districts are also much more likely to be Democratic. In all but a few states outside the South, Democrats are a majority of the party-registered voters living in the high-income block groups in such racially diverse districts.

This result persists even when replicating the analysis using the percentage non-Hispanic white, rather than the percentage black (SI Figure A.5). Furthermore, if we replicate Figure 3 only in block groups that are less than 10% black (as we do in SI Figure A.6), the same pattern emerges, though the magnitude of the difference between rich and poor block groups is smaller in the diverse districts in most regions. Much of the regional difference in party affiliation in diverse districts can be attributed to the differences in the behavior of richer voters in homogeneously white block groups.

To further illustrate interregional differences, as well as district-level income's moderating effect on the income-party relationship, we conduct a series of regression analyses. In these regressions, the dependent variable is a binary variable coded 1 for registered Republicans and 0 for registered Democrats. The independent variable is Census block-group income. We implement these regressions with a 1% simple random sample of registered voters taken from Catalist, and we employ clustered robust standard errors at the block-group level. Unlike the full-population data from Catalist that was used to compose Figure 3, which indicate block-group income but not individuals' block-group identifiers, the sample data from Catalist specify the block-group FIPS (Federal Information Processing Standard) code, allowing us to cluster standard errors appropriately. The 1% sample data contain 708,986 two-party registrants in the 29 party registration states.

Because our analysis has many moving parts (with race and income reported at different levels of aggregation), Figure 4 presents illustrative examples. In the upper plots, the x-axis reports results at three levels of district-level income, and districts are subsetted according to whether they are more or less than 5% black. In the lower plots, the x-axis represents three groups of district-level percentage black and divides the data into wealthier and poorer districts. Districts with an income of

$50,000 or less are considered poorer, whereas those above $50,000 are considered wealthier. The left panel displays nonsouthern states, the right panel southern states.

Figure 4, which displays income slope coefficients in different subgroups, shows that the income-party relationship varies minimally outside racially diverse areas of the South. As the upper plots show, the block-group income effect is strong in southern districts that are greater than 5% black, and weaker in homogeneously white southern areas and areas outside the South. Except in southern districts that are more than 5% black, the effect of block-group income on partisanship is only 1–2 percentage points higher in the richest districts than in the poorest districts. As shown in the lower-left plot for nonsouthern states, the income-party relationship is 1–2 percentage points stronger in poorer districts than in wealthier districts, but district-level racial context has a minimal effect on the income-party relationship. In southern districts, regardless of district-level income, the income-party effect grows with a growing district-level African American population.[10]

Testing Substate Variation Using Linear Mixed-Effect Models

While previous results were based almost entirely on direct enumeration of voter file data, large data sets do not solve the problem of data sparseness within state house districts. Districts with high proportions of racial minorities, for example, rarely have many high-income block groups. To estimate the income-party relationship for each state house district (rather than for groups of voters in block groups within *types* of districts), statistical modeling is necessary. To account for variation in the income-party relationship, we estimate multilevel models in which precinct-level presidential election returns are regressed on precinct-level income.

The quantities of interest are the district-specific estimates of the coefficient on precinct-level income. We use a varying-intercept, varying-slope model to estimate the income-party relationship using the 2008 precinct-level election returns clustered within state house districts. The 2008 McCain vote share is defined as the outcome of interest, and precinct-level income is defined using the mean of the block-group-level median household income values in each precinct. We then estimate a model in which the relationship between precinct-level income and 2008

---

[10]The figure does not include intercepts, which, as the previous figure indicates, meaningfully vary across states.

FIGURE 5 Income-Based Voting Is Stronger in Southern Black Districts
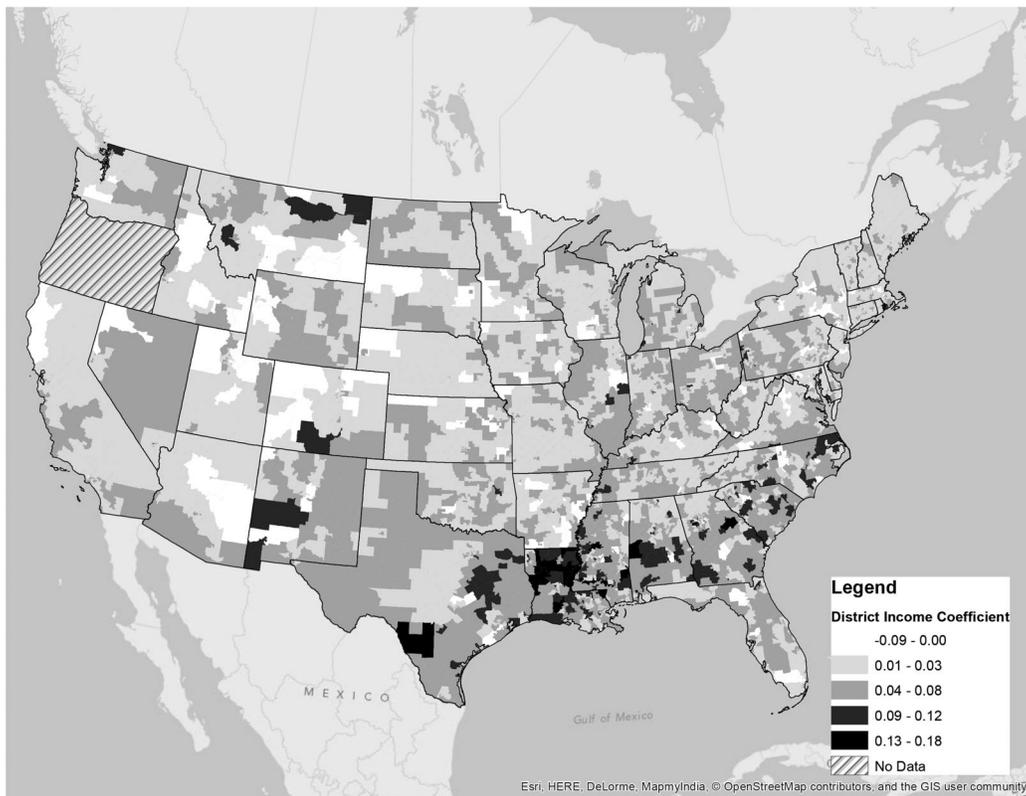
Note: Map of estimates of the district-level coefficient representing the relationship between precinct-level income (in thousands of dollars) and the Republican proportion of the two-party presidential vote in 2008. Darker shades indicate a more positive district-specific income effect.

two-party presidential vote is allowed to vary within state house districts. The hierarchical model for the income-party relationship for precinct i within district j is

$$y_{ij} = \alpha + \beta x_i + \delta_j + \gamma_j x_{i[j]} + \epsilon_{ij},$$

where α is a general intercept term; β is a general coefficient on x_i, the precinct-level median household income variable expressed in tens of thousands of dollars (after accounting for district-level random effects and slope estimates); and δ_j and γ_j are, respectively, the random intercept and income coefficients that shift the coefficient estimates for district j. From this model, we report the district-specific slope coefficient, represented by $\beta + \gamma_j$.[11]

We present the district-level income coefficients in choropleth maps in which the magnitude of the district-level slope is presented in grayscale. In Figure 5, the slope for each district is presented using five intervals

[11]These models were fitted using the lme4 package in R, which assumes that the district-level random effects are normally distributed.

selected to distinguish between positive and negative coefficients. Areas in which the marginal district-level difference is more positive appear at the dark end of the scale, whereas areas with a negative relationship appear at the light end of the scale. The fitted slope coefficients range from approximately –0.09 (a 9-point drop in the McCain vote for every $10,000 in precinct-level income) to 0.18 (an 18-point increase in the McCain vote for every $10,000). Especially prominent in this map is a band of districts in the Black Belt, the region extending from the Arkansas River delta across central Alabama and Mississippi and into eastern North Carolina. This area was once dominated by slavery and sharecropping. The racial segregation and polarized politics in this region, with affluent whites voting overwhelmingly Republican and poorer blacks voting overwhelmingly Democratic, ensure a large income coefficient.

To confirm that Figure 5 reflects differences between different types of racially heterogeneous areas, Figure 6 plots the district-level income coefficients, by district-level racial composition, in nonsouthern (left panel) and southern (right panel) states. On average, outside the

FIGURE 6 Income-Based Voting Is Stronger in African American Districts in the South

Effect of Income on Precinct Voting by Region and by District Race
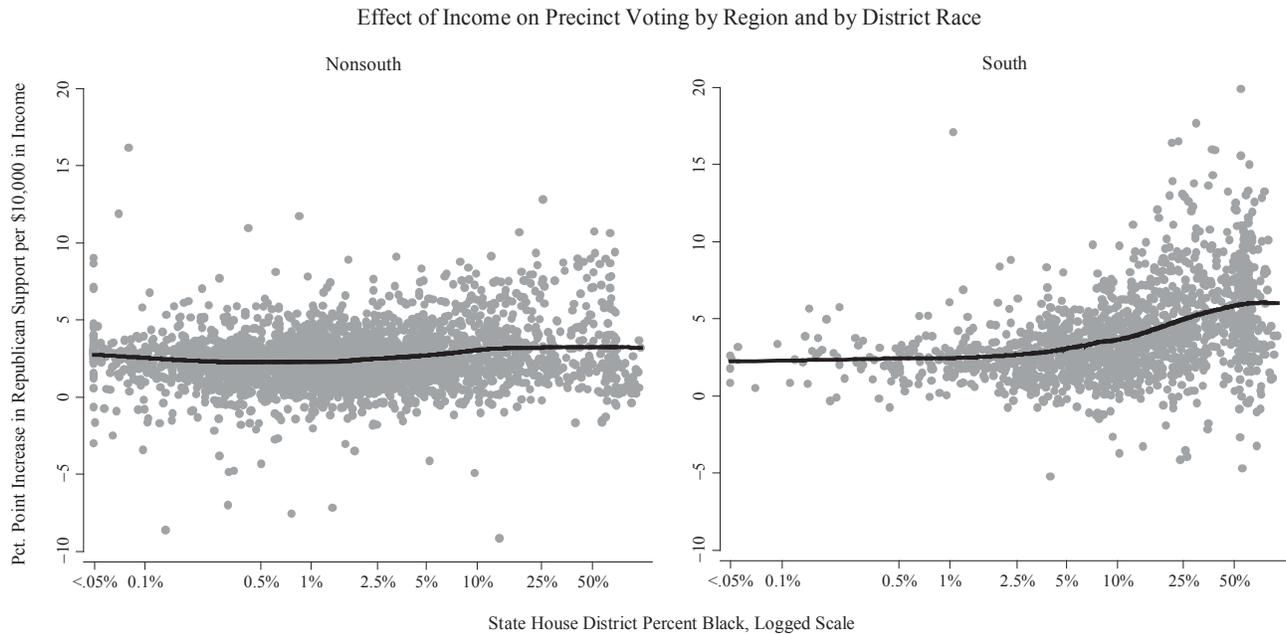
Note: Left panel displays the fitted slope coefficients from the random effects model for each state house district outside the South. Right panel displays the same coefficients from districts inside the South.

South, the income-party relationship is flat, something that holds with only a few exceptions, regardless of a district's racial composition. By contrast, in the South, income-based voting looks about the same in low African American districts as it does in similar districts elsewhere, but much stronger in the most African American districts. In districts in the South that are 25 to 50% black, the McCain vote increased by 5 points for every $10,000 in precinct-level income, whereas, on average, in the rest of the country his vote increased by only 2 to 2.5 points for each additional $10,000. In some areas of the South, this effect is 15 to 20 points.

Discussion

Our results have implications for the study of income-based voting, political geography, and contextual effects.

Our results suggest that the death of southern exceptionalism has been greatly exaggerated. When sufficient data permit small-area estimation with minimal assumptions, southern political behavior looks much as it has for many decades. Our results also demonstrate the exceptionalism of affluent urban voters in other parts of the country, especially in the Northeast Corridor and West Coast, where rich voters in racially diverse districts are, on average, almost as Democratic as poor voters (even when

the analysis is limited to include only white block groups). Substate variation is significant in explaining the red state, blue state "paradox" (Gelman et al. 2007). These localized exceptions hold, even as a moderate relationship between income and partisanship exists in most homogeneously white parts of both red and blue states. In short, while much evidence supports the "nationalization" of American politics, political behavior still cannot be captured in isolation from social and economic geography.

Some readers may be concerned about our use of block-group income in the place of individual-level income. This use of block-group income data is common in settings in which individual-level data cannot be easily collected (Geronimus and Bound 1998). While block-group data are an imperfect substitute for individual-level survey data, household-level income is known to be tightly correlated with block-group median household income. Moreover, the block-group measure may act as a suitable proxy for individual social class. By accepting this broader definition of income class and focusing on results within small geographies, we reap advantages of individual-level "big data" combined with small-scale aggregate geographic data. While resulting estimates are not directly comparable to individual-level survey estimates, they avoid other potential sources of error, including assumptions required of small-area estimation, particularly geographic

stationarity below state lines (Gelman and Little 1997; Rao 2003). Our analyses suggest that such assumptions are inappropriate.

Along with its substantive implications, our analysis suggests that political scientists should more frequently consider implications of the modifiable areal unit problem (MAUP). State-level data may yield findings that are of substantive interest for state politics, but they may yield incorrect inferences about more local geographic variation. For example, in the supporting information, we illustrate how the "red-blue paradox" is affected by the MAUP. To be sure, state house districts, our preferred geography, are, like any other geographic unit, vulnerable to the MAUP. In robustness checks, we addressed this concern. First, we replicated the multilevel models and associated maps using counties as the clustering geography, and we find no meaningful differences from the district-based analysis. Second, we estimate a local indicator of spatial autocorrelation, Local Moran's *I*, on the random effects drawn from the district-level models. We show that these random-effects estimates geographically cluster across multiple districts in the regions, and this clustering follows the same geographic patterns that we discuss in the text. We obtained similar results after conducting geographically weighted regressions on the precinct-level data, which disregard political boundaries (Brunsdon, Fotheringham, and Charlton 2010). These analyses show that our results are not a result of gerrymandering, but of voters' racial, economic, and political clustering below state lines (see, e.g., Chen and Rodden 2013).

Conclusion

What explains geographic variation in income-based voting? Our nonparametric and model-based findings based on microlevel partisanship data reinforce and extend findings based on survey data, while delving more deeply into the geographic origins of income-based voting. We find that the income-party relationship tends to be stronger than otherwise expected in places with high levels of racial diversity and a history of local racial exploitation, and weaker elsewhere. The relationship between income and partisanship is bound up with race.

Our evidence highlights the importance of local racial and ethnic heterogeneity in support for redistribution and the parties of the left. Economic voting models predict that rich and poor voters should have opposing political preferences oriented around redistribution (Meltzer and Richard 1981). We should then expect the rich and poor to be polarized consistently around the parties favoring redistribution. Explanations based on nonmaterialist or postmaterialist explanations may carry some weight, but they are less relevant than the persistence of local racial segregation and political cultures. While our results are consistent with Alesina and Glaeser (2004), we suspect that they reflect not so much contextual effects, but long-standing, geographically bounded racial poverty and associated local attitudes and institutions (King 1996). The places with unusual income-based voting are not "red states," but districts of the Old South and a few other areas where poor minorities share political geography with more affluent, conservative whites and that have a long history of racial, economic, and political inequality. The other unusual phenomenon is the rich voters outside the South who live among poorer voters but remain unusually Democratic in their affiliations. Whether these differences result from contextual effects or sorting of different types of white voters into different neighborhoods, a data-centered approach has allowed us to show that local racial and economic geography is an important explanation of state variation in income-based voting.

References

Abrams, Samuel, and Morris Fiorina. 2012. "'The Big Sort That Wasn't': A Skeptical Reexamination." *PS: Political Science and Politics* 45(2): 203–10.

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *Quarterly Journal of Economics* 114(4): 1243–84.

Alesina, Alberto, and Edward Glaeser. 2004. *Fighting Poverty in the United States and Europe: A World of Difference.* Oxford: Oxford University Press.

Ansolabehere, Stephen, and Eitan Hersh. 2012. "Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate." *Political Analysis* 20(4): 437–59.

Ansolabehere, Stephen, and Jonathan Rodden. 2011. "Harvard Election Data Archive." http://projects.iq.harvard.edu/eda.

Brunsdon, Chris, A. Stewart Fotheringham, and Martin E. Charlton. 2010. "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity." *Geographical Analysis* 28(4): 281–98.

Butler, David, and Donald Stokes. 1969. *Political Change in Britain: Forces Shaping Electoral Choice.* New York: St. Martin's Press.

Chen, Jowei, and Jonathan Rodden. 2013. "Unintentional Gerrymandering: Political Geography and Electoral bias in Legislatures." *Quarterly Journal of Political Science* 8(3): 239–69.

Cho, Wendy K. Tam, and Thomas J. Rudolph. 2008. "Emanating Political Participation: Untangling the Spatial Structure behind Participation." *British Journal of Political Science* 38(2): 273–89.

Dawson, Michael C. 1995. *Behind the Mule.* Princeton, NJ: Princeton University Press.

ESRI. 2008. U.S. Census Block Groups. In *StreetMap North America* http://hdl.handle.net/10573/42125.

Feller, Avi, Andrew Gelman, and Boris Shor. 2012. "Red State/Blue State Divisions in the 2012 Presidential Election." *The Forum* 10(4): 127–31.

Fitch, Catherine A., and Steven Ruggles. 2003. "Building the National Historical Geographic Information System." *Historical Methods* 36(1): 41–51.

Fotheringham, A. S., and D. W. S. Wong. 1991. "The Modifiable Areal Unit Problem in Multivariate Statistical Analysis." *Environment and Planning A* 23(7): 1025–44.

Gay, Claudine. 2006. "Seeing Difference: The Effect of Economic Disparity on Black Attitudes towards Latinos." *American Journal of Political Science* 50(4): 982–97.

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* New York: Cambridge University Press.

Gelman, Andrew, and Thomas Little. 1997. "Poststratification into Many Categories Using Hierarchical Logistic Regression." *Survey Methodology* 23(2): 127–35.

Gelman, Andrew, Boris Shor, Joseph Bafumi, and David Park. 2007. "Rich State, Poor State, Red State, Blue State: What's the Matter with Connecticut?" *Quarterly Journal of Political Science* 2(4): 345–67.

Gelman, Andrew, Boris Shor, Joseph Bafumi, David Park, and Jerome Cortina. 2008. *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do.* Princeton, NJ: Princeton University Press.

Geronimus, Arline T., and John Bound. 1998. "Use of Census-Based Aggregate Variables to Proxy for Socioeconomic Group: Evidence from National Samples." *American Journal of Epidemiology* 148(5): 475–86.

Ghitza, Yair, and Andrew Gelman. 2013. "Deep Interactions with MRP: Election Turnout and Voting Patterns among Small Electoral Subgroups." *American Journal of Political Science* 57(3): 762–76.

Gilens, Martin. 1999. *Why Americans Hate Welfare: Race, Media, and the Politics of Anti-Poverty Policy.* Chicago, IL: University of Chicago Press.

Gimpel, James, and Dante Chinni. 2011. *Our Patchwork Nation: The Surprising Truth about the "Real America."* New York: Gotham.

Gimpel, James G., and Jason E. Schuknecht. 2004. *Patchwork Nation: Sectionalism and Political Change in American Politics.* Ann Arbor, MI: University of Michigan Press.

Hersh, Eitan D. 2013. "Long-Term Effect of September 11 on the Political Behavior of Victims' Families and Neighbors." *Proceedings of the National Academy of Sciences* 110(52): 20959–63.

Hersh, Eitan D. 2015. *Hacking the Electorate: How Campaigns Perceive Voters.* Cambridge: Cambridge University Press. Forthcoming.

Hill, Kim Quaile, and Jan E. Leighley. 1992. "The Policy Consequences of Class Bias in State Electorates." *American Journal of Political Science* 36(2): 351–65.

Huber, John D., and Piero Stanig. 2007. "Why Do the Poor Support Right-Wing Parties? A Cross-National Analysis." Presented at the RSF Inequality Conference, UCLA.

Huckfeldt, Robert R., and John D. Sprague. 1995. *Citizens, Politics and Social Communication.* Cambridge: Cambridge University Press.

Inglehart, Ronald. 1990. *Culture Shift in Advanced Industrial Society.* Princeton, NJ: Princeton University Press.

Katz, Elihu, and Paul F. Lazarsfeld. 1955. *Personal Influence: The Part Played by People in the Flow of Mass Communications.* Glencoe, IL: Free Press.

King, Gary. 1996. "Why Context Should Not Count." *Political Geography* 15(2): 159–64.

Lax, Jeffrey, and Justin Phillips. 2009. "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science* 53(1): 107–21.

Lazarsfeld, Paul F. 1948. *The People's Choice: How the Voter Makes Up His Mind in a Presidential Election.* 2nd ed. New York: Columbia University Press.

Levendusky, Matthew, Jeremy Pope, and Simon Jackman. 2008. "Measuring District Level Partisanship with Implications for the Analysis of U.S. Elections." *Journal of Politics* 70(3): 736–53.

Luttmer, Erzo. 2001. "Group Loyalty and Taste for Redistribution." *Journal of Political Economy* 109(3): 500–28.

McGhee, Eric, and Daniel Krimm. 2009. "Party Registration and the Geography of Party Polarization." *Polity* 41(3): 345–67.

Meltzer, Allan H., and Scott F. Richard. 1981. "A Rational Theory of the Size of Government." *Journal of Political Economy* 89(5): 914–27.

Oliver, J. Eric. 1999. "The Effects of Metropolitan Economic Segregation on Local Civic Participation." *American Journal of Political Science* 43(1): 186–212.

Park, David, Andrew Gelman, and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12(4): 375–85.

Putnam, Robert D. 2007. "E Pluribus Unum: Diversity and Community in the Twenty-First Century." *Scandinavian Political Studies* 30(2): 137–74.

Rao, J. N. K. 2003. *Small-Area Estimation.* Hoboken, NJ: Wiley and Sons.

Romer, Thomas. 1975. "Individual Welfare, Majority Voting, and the Properties of a Linear Income Tax." *Journal of Public Economics* 4(2): 163–85.

Sampson, R. J., S. W. Raudenbush, and F. Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." *Science* 277(5328): 918–24.

Segura, Gary. 2012. "Latino Public Opinion and Realigning the American Electorate." *Daedalus* 141(4): 98–113.

Tausanovitch, Chris, and Christopher Warshaw. 2013. "Measuring Constituent Preferences in Congress, State Legislatures, and Cities." *Journal of Politics* 75(2): 330–42.

Tingsten, Herbert. 1937. *Political Behavior: Studies in Election Statistics.* London: P. S. King.

Tobler, Waldo R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46(supplement): 234–40.

United States Bureau of the Census. 2006. "State Legislative Districts." *Technical Report.* Washington, DC: United States

Bureau of the Census. http://www.census.gov/geo/www/cob/sld_metadata.html.

Vigdor, Jacob. 2006. "Fifty Million Voters Can't Be Wrong." NBER Working Papers (12371).

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure A.1: Party Registration Compared to Survey-Based Measures of Partisanship

Figure A.2: Party Registration Compared to Precinct Returns

Figure A.3: Party Registration Compared to 2010 District Representatives

Figure A.4: Relationship between Self-Reported Family Income and Census Block Group Median Household Income

Figure A.5: Replication of Figure 3 with District Race Constructed Based on Non-White Population

Figure A.6: Replication of Figure 3 with Block Groups Restricted to Only Those Less Than 10% Black

Figure A.7: Replication of Figure 4 Only Among Those Who Registered Before 2008

Figure A.8: State House District Aggregates Follow Individual-Level Income-Party Patterns, State Aggregates Do Not

Figure A.9: Replication of Figure A.8 using Self-Reported Partisanship and Income Fields

Figure A.10: Marginal Income-Based Voting Effect is Largest in Rural Areas with Racialized Poverty, Even After Accounting for District-Level Social Variables

Figure A.11: Income Effects are Weaker in Non-Southern Urban Areas: Inset Map of New York City

Figure A.12: Income Effects are Stronger in Southern, African American Areas: Inset Map of Louisiana

Figure A.13: Clustering of Income-Based Voting Is Strongest Across the Deep South and Inverted in the Mormon Corridor

Table A.1: Summary Statistics for State House Districts Used in Multilevel Modeling (n=2,570)

Table A.2: Summary Statistics of Harvard Election Data Archive Precinct Data Merged with 2000 Census Block-Group Data (n=165,631)