Technical Appendix

The Geography of Racial Stereotyping: Evidence and Implications for VRA "Preclearance" After *Shelby County*

Christopher S. Elmendorf
University of California, Davis
School of Law

Douglas M. Spencer
University of Connecticut
School of Law

Last updated: February 1, 2014

Contents

1 Findings Using Pooled vs. Unpooled Data

In this section we compare our results separately using data from the 2008 National Annenberg Election Survey (NAES) and the 2008 Cooperative Campaign Analysis Project (CCAP) survey. Results in the paper are based on a pooled sample of both surveys. As we show below, our analysis and findings are substantively similar whether we use NAES or CCAP data, reinforcing our decision to pool the samples for greater statistical power.

We note that racial stereotyping questions were presented differently to respondents in the two surveys. NAES respondents were asked to rate the work ethic, trustworthiness, and intelligence of their own ethnic group as well as Blacks using a 100-point slider for each question and ethnic group. The CCAP used a single question to elicit ratings of several different racial groups on two characteristics (work ethic and intelligence) using a 7-point scale.

Our measure of stereotyping is generated by aggregating respondents' ratings in each survey. We note that the NAES data is more reliable than the CCAP data for three reasons. First, the NAES utilized a random digit dialing selection process whereas the CCAP comprises an opt-in sample. Second, NAES respondents are asked to rate three characteristics compared to two for CCAP respondents. Researchers have shown that stability in public opinion increases as the number of component measures increases. Finally, the CCAP asked respondents to rate their own race and three different racial groups all at one time, using a grid. The NAES asked respondents to rate their own ethnic group on each measure separately, followed by ratings of blacks on each measure separately. By separating the ratings, the NAES made it more difficult for respondents to report identical scores for characteristics between races.

1.1 State-level estimates of stereotyping by nonblacks

On the next page we present plots that rank the states by the proportion of nonblack residents who stereotype blacks more negatively than (A) the national median, (B) 75% of all respondents, and (C) 90% of all respondents. Horizontal lines are 95% confidence intervals. Vertical lines represent the "average" state.

Looking at the states that were covered under Section 4 of the VRA before *Shelby County*, we see striking similarities in the rankings, regardless of the survey. Alaska and Arizona (states that are covered because of discrimination against Native Americans and Latinos, not blacks) consistently rank as the lowest two states on our measure of anti-black stereotyping. Mississippi, Louisiana, Alabama, Texas, and Georgia consistently rank in the top 15 states. Virginia and South Carolina fluctuate, though the fluctuations are similar whether using NAES or CCAP. The only exception is the top 10% plot, where South Carolina moves from the top ten using NAES data to the bottom ten using CCAP data. When we pool the data, South Carolina appears exactly in the middle (26 of 51).

1.1.1 CCAP only

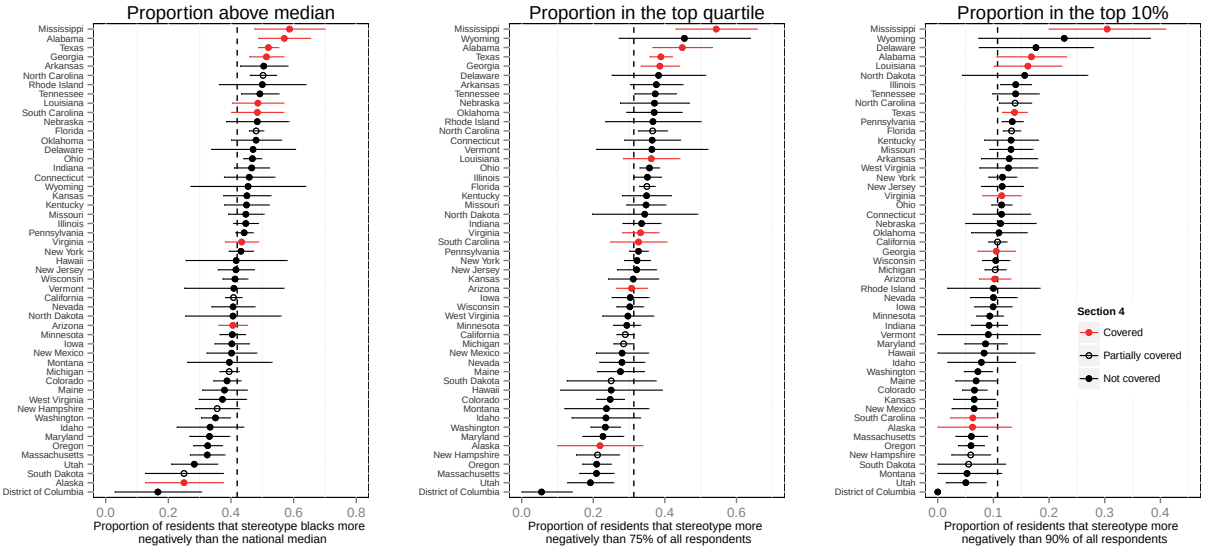


Figure A.3 Nonblack respondents to the 2008 CCAP survey (N=18,500).

1.1.2 NAES only

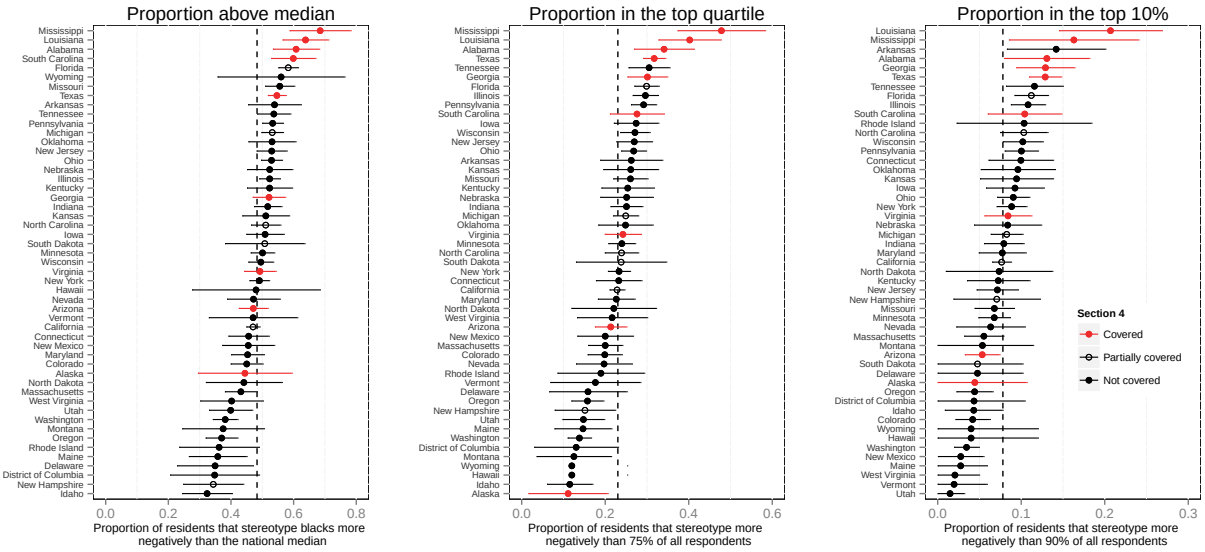


Figure A.2 Nonblack respondents to the 2008 NAES (N=19,325).

1.2 Predicting Political Behavior

1.2.1 Obama Vote Share

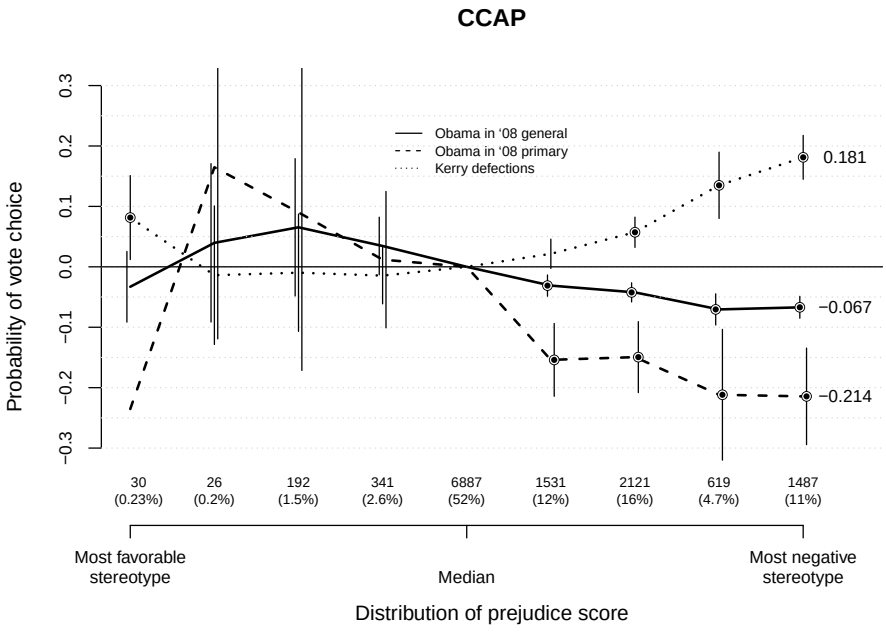
	(A) Obama '08 general election			(B) Obama '08 primary election			(C) '04 vote for Kerry/ '08 vote for McCain		
	CCAP	Pooled	NAES	CCAP	Pooled	NAES	CCAP	Pooled	NAES
Negative stereotype	-0.03^{***}	-0.03^{***}	-0.04^{***}	-0.09^{***}	-0.07^{***}	-0.06^{***}	0.06^{***}	0.06^{***}	0.06^{***}
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)
Age	-0.02^{***}	-0.01^{***}	-0.01^{*}	-0.06^{***}	-0.05^{***}	-0.04^{***}	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Conservativeness	-0.07^{***}	-0.07^{***}	-0.06^{***}	-0.01	-0.02^{***}	-0.03^{***}	0.03^{***}	0.03^{***}	0.04^{***}
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Party ID (7 point)	-0.12^{***}	-0.13^{***}	-0.13^{***}	-0.00	0.00	0.00	0.06^{***}	0.06^{***}	0.07^{***}
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female	0.02^{***}	0.00	-0.00	-0.07^{***}	-0.10^{***}	-0.11^{***}	0.02^{***}	0.02^{***}	0.02^{*}
	(0.01)	(0.00)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R is Hispanic	0.02	0.04^{***}	0.05^{***}	-0.04	-0.01	-0.00	-0.02	-0.03^{**}	-0.03^{*}
	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
R is "other" race	0.09^{**}	0.12^{***}	0.22^{**}	0.01	0.16^{*}	0.42^{***}	$-0.06^{\dagger}$	-0.09^{*}	-0.16^{*}
	(0.03)	(0.03)	(0.07)	(0.08)	(0.06)	(0.10)	(0.03)	(0.04)	(0.07)
Education	0.00	0.01^{***}	0.02^{***}	0.04^{***}	0.05^{***}	0.06^{***}	$-0.01^{\dagger}$	-0.01^{***}	-0.01^{**}
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Income	$-0.01^{\dagger}$	0.00	0.01^{*}	-0.00	$0.01^{\dagger}$	0.01^{*}	-0.00	-0.00	$-0.01^{\dagger}$
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Midwest	0.00	0.00	0.00	-0.02	0.04^{**}	0.07^{***}	0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
South	-0.02^{*}	-0.03^{***}	-0.04^{***}	-0.09^{**}	-0.05^{***}	-0.04^{*}	$0.02^{\dagger}$	0.02^{*}	$0.02^{\dagger}$
	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
West	-0.01	-0.01	-0.00	0.06^{*}	0.08^{***}	0.09^{***}	0.00	-0.01	-0.02
	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
(Intercept)	1.26^{***}	1.26^{***}	1.27^{***}	0.49^{***}	0.51^{***}	0.54^{***}	-0.11^{***}	-0.13^{***}	-0.14^{***}
	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
N	7161	21033	13872	2836	10163	7327	2521	7619	5098
R^2	0.78	0.66	0.60	0.07	0.07	0.07	0.22	0.21	0.20
adj. R^2	0.78	0.66	0.60	0.07	0.07	0.07	0.22	0.21	0.20
Resid. sd	0.23	0.29	0.31	0.48	0.48	0.48	0.18	0.24	0.26

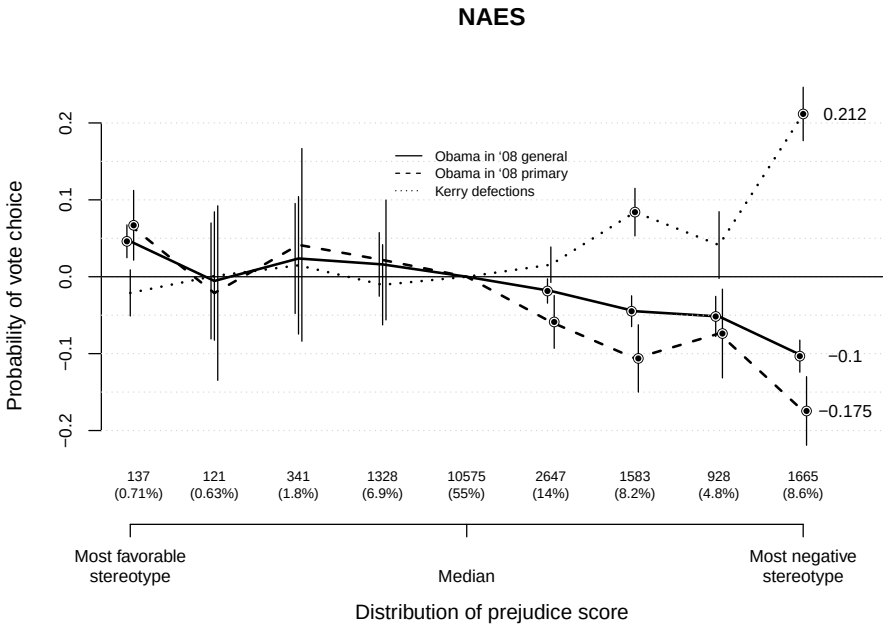
Standard errors in parentheses
† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table B.1 Linear probability models predicting votes for or against Obama. All non-dichotomous independent variables (prejudice, age, education, and income) have been normalized to facilitate interpretation. Model (A) predicts the probability of voting for Obama in the 2008 general election. Model (B) predicts the probability of voting for Obama in the 2008 primary election and only includes primary voters. Model (C) predicts the probability that a person who voted for John Kerry in 2004 defected and did not vote for Obama in 2008. The model only includes people who voted for Kerry in 2004.

Each of the three models fits the data almost identically whether we use CCAP data, NAES data, or a pooled sample. The sign and magnitude of the estimates are very similar in every model for every coefficient, including the intercept term. Additionally, the precision of each variable is nearly identical in all of the models even though the sample sizes vary greatly.

1.2.2 Vote choice by varying degrees of stereotyping





Although the CCAP sample is much smaller, meaning estimates are more noisy, we see a strikingly similar pattern in both datasets: (1) imprecise estimates near zero for those with favorable stereotypes of blacks and (2) more precise estimates of vote choice that increase in magnitude as stereotypes become more negative.

1.3 Racially Polarized Voting

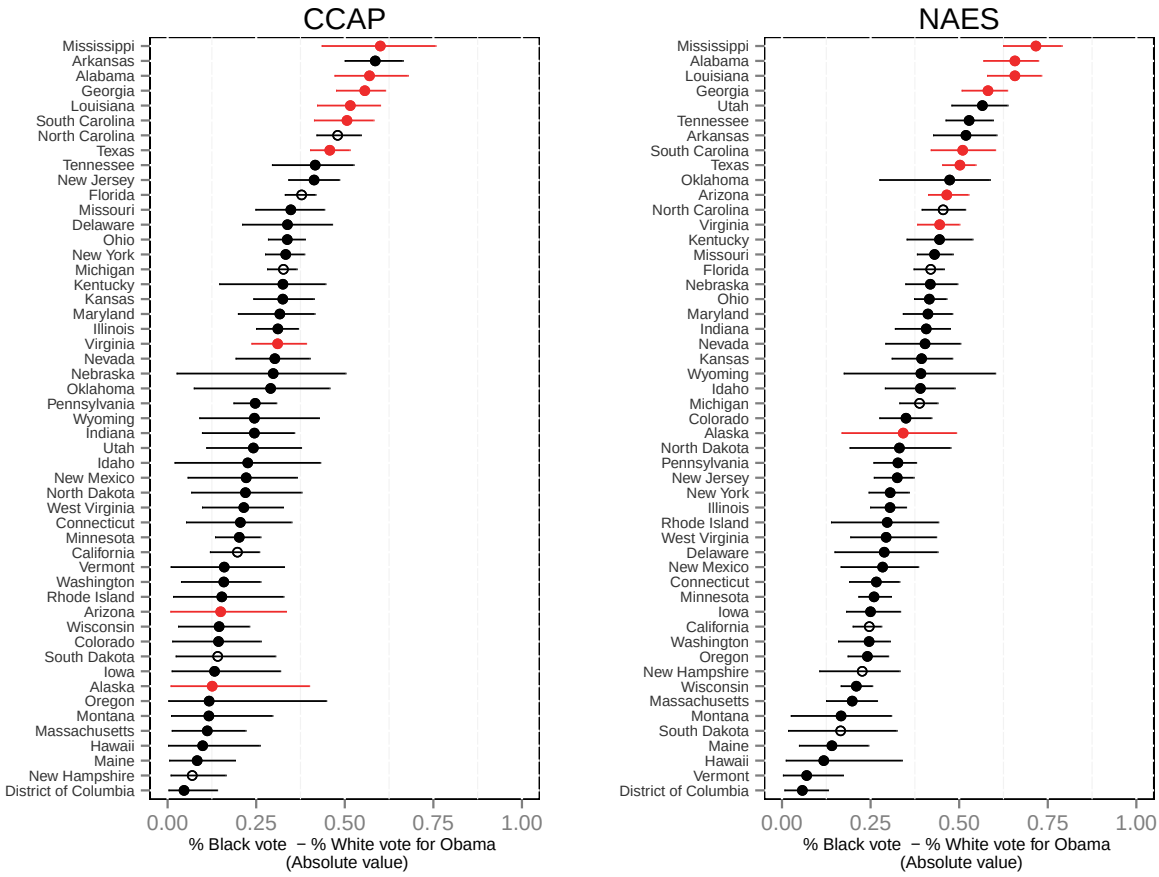


Figure C Racially polarized voting measured as the absolute difference between votes for Obama among black and white voters as reported by respondents to the 2008 CCAP (left panel) and NAES (right panel) surveys. We estimate black support for Obama using MRP because the sample of black respondents is very small in some states. Solid horizontal lines represent 95% confidence intervals.

2 MRP Meets the VRA

This section provides a technical description of multi-level regression with poststratification (MRP) and its application in the paper. Our intent is to provide political scientists and others interested in our methodological approach a transparent presentation of our assumptions and modeling choices.

2.1 County-level estimates of anti-black stereotyping

Because sub-state geographic identifiers are not available for the NAES sample, we estimate the prevalence of racial stereotyping at the county level using data from the CCAP survey. Our dependent variable of racial stereotyping (S) is an aggregated measure of ratings (R) such that

$$S_i = \sum_j R_{ij}^B - R_{ij}^O \tag{1}$$

where i indexes the respondent, j indexes the group attributes (work effort or intelligence; higher scores mean lazy or unintelligent), and O and B refer, respectively, to the respondent's own racial group and to blacks. S_i is positive if, on average, the respondent views her own racial group as better than blacks on these criteria; it is negative if the respondent deems blacks better than her group.
Using a varying intercept, constant slope multi-level regression, we model racial stereotypes as a function of nonblack respondents' demographic attributes (simplified into three race, two sex, and four education categories), and their county of residence:

$$S_i = \beta^0 + \alpha_{j[i]}^{sex,race} + \alpha_{k[i]}^{educ} + \alpha_{l[i]}^{county} + \varepsilon_i \tag{2}$$

where each demographic variable after the intercept (β^0) is modeled independently, drawn from a normal distribution with mean zero and some estimated variance.

$$\begin{aligned} \alpha_{j[i]}^{sex,race} &\sim N(0, \sigma_{sex,race}^2), \text{ for } j = 1, \ldots, 6 \\ \alpha_{k[i]}^{educ} &\sim N(0, \sigma_{educ}^2), \text{ for } k = 1, \ldots, 4 \end{aligned} \tag{3}$$

The county variable is also modeled independently, and in two iterations. The first iteration (presented in Figure 6a of the paper) models county as a function of the black share of each county's population:

$$\alpha_{c[i]}^{county} \sim N(\text{percent black}, \sigma_{county}^2), \text{ for } c = 1, \ldots, 3071 \tag{4}$$

The second iteration (presented in Figure 6b of the paper) adds an identifier for each state:

$$\alpha_{c[i]}^{county} \sim N(\alpha_{m[i]}^{state} + \beta^1(\text{percent black}), \sigma_{county}^2) \text{ for } c = 1, \ldots, 3071 \quad (5)$$

where the random effects state variable, including the District of Columbia, has mean zero and some estimated variance.

$$\alpha_{s[i]}^{state} \sim N(0, \sigma_{state}^2), \text{ for } s = 1, \ldots, 51 \quad (6)$$

Each combination of demographic and geographic variables defines a unique cell. In our current analysis we include 24 demographic variables in 3,071 counties, giving us a total of 73,704 unique cells (34 $\times$ 3,071). For example, the first cell includes all white females with no high school education in Autauga County, Alabama. The 73,704th cell includes all "Other" males with graduate degrees in Weston County, Wyoming. We use the results of the multi-level model to make predictions about the racial stereotyping of those in each cell. Suppose that θ_x represents the predicted value of each cell based on the relevant combination of coefficients from the multi-level model and that N_x represents the frequency of each cell in the population. To estimate the the stereotyping behavior in each county (c), we weight θ_x by the population frequency N_x of each cell in each county such that

$$\text{Stereotyping}_{\text{county } c}^{\text{MRP}} = \frac{\sum_{x \in c} N_x \theta_x}{\sum_{x \in c} N_x} \quad (7)$$

2.2 Racially Polarized Voting

We also use MRP to estimate the extent of racially polarized voting in each state. Our measure of racially polarized voting is the absolute difference between votes for Obama among black and white voters as reported by respondents to the 2008 NAES and CCAP surveys. We include responses to post-election waves only, but the state rankings are identical if we use the vote intention reported in the wave that immediately preceded the November election. Even though the sample size of the pooled surveys is greater than 40,000 the number of black respondents within states in quite small. See Table B.

State	# of black respondents	State	# of black respondents
Maine	0	Arizona	13
North Dakota	0	Wisconsin	15
Vermont	0	Massachusetts	18
Wyoming	0	Indiana	20
Hawaii	1	Colorado	21
Idaho	1	Mississippi	21
Montana	1	Missouri	21
New Hampshire	1	Washington	22
Rhode Island	1	Alabama	27
South Dakota	1	Tennessee	28
Utah	1	Louisiana	29
West Virginia	1	South Carolina	36
Alaska	2	New Jersey	40
Oregon	3	Maryland	45
Nebraska	4	Virginia	53
Iowa	5	New York	76
New Mexico	5	Illinois	78
Kansas	6	Pennsylvania	86
Minnesota	6	Georgia	92
Delaware	7	Michigan	95
Nevada	7	Ohio	98
District of Columbia	9	North Carolina	99
Arkansas	10	Texas	104
Oklahoma	10	California	124
Kentucky	11	Florida	134
Connecticut	12	**TOTAL**	**1,500**

Table B Number of black respondents to the 2008 NAES and CCAP surveys. The total number of respondents (of all races) is 41,384.

The lack of black respondents means that state-level estimates of black candidate preference are unreliable and in many states meaningless or impossible to generate. To overcome the limited sample size, we model votes for Obama in the 2008 general election as a function of sex, education, income, and state:

$$\text{Vote}_i = \beta^0 + \alpha^{sex}_{j[i]} + \alpha^{educ}_{k[i]} + \alpha^{income}_{l[i]} + \alpha^{state}_{m[i]} + \varepsilon_i \tag{8}$$

Each demographic variable after the intercept (β^0) is modeled independently, drawn from a normal distribution with mean zero and some estimated variance.

$$\begin{aligned} \alpha^{sex}_{j[i]} &\sim N(0, \sigma^2_{sex}), \text{ for } j = 1, 2 \\ \alpha^{educ}_{k[i]} &\sim N(0, \sigma^2_{educ}), \text{ for } k = 1, \ldots, 4 \\ \alpha^{income}_{l[i]} &\sim N(0, \sigma^2_{income}), \text{ for } l = 1, \ldots, 4 \end{aligned} \tag{9}$$

The state geographic variable is also modeled independently as a function of the geographic region and black population size.

$$\alpha_{m[i]}^{state} \sim N(\alpha_{n[i]}^{region} + \beta^1(\text{percent black}), \sigma_{state}^2) \text{ for } m = 1, \ldots, 51 \tag{10}$$

The random effects region variable is also modeled independently with mean zero and some estimated variance.

$$\alpha_{n[i]}^{region} \sim N(0, \sigma_{region}^2), \text{ for } n = 1, \ldots, 4 \tag{11}$$

The estimates generated from this model represent Obama's vote share among black voters in each state. We use disaggregation to estimate Obama's vote share among whites as all but four states have at least 100 white respondents. The difference between these estimates of vote share by race is presented as the measure of racially polarized voting in Figure 7c in the paper.