# How Well Does the Current Population Survey Measure the Composition of the U.S. Voting Population?

Kurt Bauman
Education and Social Stratification Branch
Social, Housing and Economic Statistics Division
U.S. Census Bureau

July 6, 2018

Social, Economic and Housing Statistics Working Paper Series
SEHSD Working Paper Number 2018-25

# How Well Does the Current Population Survey Measure the Composition of the U.S. Voting Population?

## INTRODUCTION

Since the Census Bureau first collected data on voting in the Current Population Survey (CPS) in 1964, the estimates of voter turnout have almost always been higher than those reported by election officials. In that year, the Census Bureau estimate exceeded reported turnout by 6 million out of 70 million votes cast (Powers and Dodge 1965). The Census Bureau estimates have continued to be higher than official tallies since that time, though they vary from year to year and state to state (Bauman and Julian 2010). CPS is not alone in overestimating voter participation. Researchers have found the problem to universally affect survey-based estimates of voting in the United States and around the world (Clausen 1968, Selb and Munzert 2013).

There have been many studies examining the problems with survey estimates of voting. However, there has been little progress in improving actual surveys in the field. In light of this condition, data users have been encouraged to use survey results to look at demographic and socio-economic characteristics of the voting and non-voting population rather than using them to look at the level of turnout, per se. This recommendation was put forth with the original release of the CPS data in 1964, where readers were encouraged to understand the statistics as "generally reflecting real differences in voting behavior among classes of the population," despite differing from official records in terms of absolute level of turnout (Powers and Dodge 1965, p. 5). A paper by Hur and Achen (2013) specifically recommended adjusting CPS totals to match official turnout levels, while still making use of the CPS estimates of the characteristics of the voting population.

While it would be nice to believe that survey estimates of the composition of the voting population (i.e., people who cast a ballot in a given election) are unaffected by the overestimate of turnout, this is almost certainly more than can be hoped. Research comparing survey responses to

actual turnout has found that people's characteristics do matter with respect to accuracy of reporting, at least to some extent (Bernstein, Chadha and Montjoy 2001, Cassel 2004, Ansolabehere and Hersh 2012). The question that faces us is not whether there is bias in CPS estimates of the characteristics of voters, but, rather, whether the bias is small enough that we should be comfortable using these estimates of the composition of the voting population as reasonable approximations to the truth. Sigelman (1982) concluded that estimates from the National Election Survey were close enough to "validated" results as to not affect most substantive conclusions, but cautioned that the effect of race on misreport is strong enough to be of concern (see also Abramson and Clagett 1992, Deufel and Kedar 2010). McDonald (2007) compared the overall demographic composition of the voting population estimated by CPS and other surveys to turnout recorded by election officials in the few states where officials report on the characteristics of voters. While survey results did not always match the composition revealed from voting records, in most cases CPS estimates were judged to be reasonably close. However, these comparisons were made for a single point in time (the election of 2004), while the amount by which CPS estimates exceed administrative reports has varied over time (Bauman and Julian 2010). Further exploration of this topic is greatly needed.

This paper approaches the issue in two ways. First, building on earlier work by Jennings (1990), a set of regression models measure the degree to which state-level population characteristics are associated with higher overall CPS voting estimates across states relative to official state turnout reports. The question becomes, "is the overall CPS estimate higher (or lower) relative to official counts in states with higher percentage of a given demographic group?" For reasons explained below, this does not allow us to say that higher CPS estimates resulted from misreports by members of a given group, but it does tell us whether population characteristics seem to matter, and can be used to provide rough estimates of the size of the bias under certain assumptions.

The second approach is to replicate the analysis conducted by McDonald in 2007, focusing on the difference between estimates of the racial makeup of the voting population in CPS and in state voter turnout reports across nine elections in the four states for which data were available.

## OVERREPORTING OF VOTER TURNOUT IN THE CURRENT POPULATION SURVEY

In this paper, I will use the term "overreport" to refer to the total difference in turnout estimates between surveys and official voter records. Many explanations for survey overreport of voting have been offered and examined. For example, the possibility that voter records, rather than surveys are at fault has been looked at in various ways without finding serious systematic deficiencies. Importantly, voter turnout by race does not seem to be systematically misreported by election authorities (Abramson and Claggett 1992).

Most research is now focused on two explanations. The first is the tendency of respondents to provide "socially desirable" answers to questions (Clausen 1968, Silver, Anderson and Abramson 1986, DeBell et al. 2015). The second explanation is non-response bias (Burden 2000, Sciarini and Goldberg 2016).

### *Social desirability bias*

The presence of social desirability bias was detected by matching survey respondents to voter rolls, which are often available to the public (Clausen 1968). This is referred to as "validation." It's fairly rare for survey respondents to say they didn't vote when they are listed on the rolls as having voted. However, those who are recorded as not voting have a fairly high rate of reporting that they did (Tourangeau, Groves and Redline 2010). The tendency of survey respondents to report voting when the record shows they did not will be referred to here as "misreporting."

Analysis of respondents who misreported their voting status shows that they are not a simple cross-section of the voting population. Most of the factors that predict misreport are the same as those that predict voting itself. Relative to people who correctly report their voting status, misreporters tend to be politically involved and partisan. As regards basic socioeconomic and demographic characteristics, findings can be summarized as follows:

- Education has consistently been identified as being positively associated with misreport of voting in validated samples of the National Elections Studies (NES) (Silver, Anderson and Abramson 1986, Anderson and Silver 1986, Bernstein, Chadha and Montjoy 2001, Belli, Traugott and Beckmann 2001, Ansolabehere & Hersh 2012), as well as the 2008 Cooperative Congressional Election Study (CCES) (Ansolabehere & Hersh 2012). European surveys examined by Granberg and Holmberg (1991) and Karp and Brockington (2005) do not always show a significant effect, but are also generally positive.
- Income was found by Ansolabehere and Hersh (2012) to be positively associated with misreport in NES and CCES.[1] Unfortunately, income has not been available as a measure in surveys in other countries.
- Age was found to be positively related to misreport in NES and CCES (Belli, Traugott and Beckmann 2001, Cassel 2004, Ansolabehere and Hersh 2012). Granberg and Homberg (1991) also found a positive effect in Swedish data. However, in their study of voting in several countries, Karp and Brockington (2005) found this relation to only hold in the United States.
- Males have been more likely to misreport voting status in the NES and CCES (Belli, Traugott and Beckmann 2001, Cassel 2004, Ansolabehere and Hersh 2012), although the findings

[1] Tragott and Katosh (1979) find a negative relation between income and misreport. However, their research used the total population, rather than the nonvoting population as a basis. The negative selection into the nonvoting group may be larger than the positive effect of income on misreporting in this case.

have been less consistent than with other variables. Karp and Brockington (2005) did not find a relation even in U.S. data.

- Blacks have been found to be more likely to misreport in numerous NES studies (Bernstein, Chadha and Montjoy 2001, Belli, Traugott and Beckmann 2001, Ansolabehere and Hersh 2012). Not surprisingly, European studies did not examine this variable. Probably due to small sample size, few studies have examined misreport by Hispanic origin, but Cassel (2004) found a small negative effect.[2]

Based on these results, it should be expected that the voting population identified through surveys (the "survey-based voting population") would be older, better-off, more educated, and more often African-American than the actual voting population.

Thinking about the effects of misreport, this survey-based voting population can be conceptually divided into two groups. Those who said they voted and actually did vote (the "actual voting population") would be more-or-less accurately represented in the survey-based voting population. Those who said they voted but didn't ("actual nonvoters") would be the ones who bring in the bias. People who misreport their voting introduce bias to the degree they don't have the same socio-demographic makeup as the actual voters.

The degree of bias introduced by the misreporters is limited by two factors. First, as stated above, the characteristics of misreporters make them seem more like actual voters. If it were poor people who were more likely to misreport, the average income of the survey-based voting population would be biased downward. However, it being the well-to-do who are most likely to misreport, the bias

[2] Some of the variables found to be associated with misreport were political involvement, religious attendance, sense of civic duty, partisanship, and voting expectations (Silver, Anderson and Abramson 1986, Cassell 2004, Ansolabehere and Hersh 2012). Measures of these attributes are not readily available in the CPS and were not considered here. The CPS has included supplemental surveys on Civic Engagement since 2008 and Volunteering since 2005. However, these were not collected at the same time as the voting supplement except for the Civic Engagement supplement in 2008 and 2010. In other instances, analysis using cross-month or cross-year matching could be used for some types of analyses.

is softened or possibly even eliminated, as the well-to-do are already more likely to vote. The second factor in reducing the bias is that the population subject to this bias is small. For example, take the results of the 1964 election when there were 70 million actual voters and 76 million in the survey-based voting population. Since a little over 90 percent of the survey-based voting population is presumed to be accurately classified as to social and demographic characteristics, any bias in the remaining part would be dampened when considering the whole.

Another implication of thinking of misreport as the basis for overreport is that the amount of bias depends on the size of the non-voting population. In low-turnout elections, there are more nonvoters, and unless the propensity to misreport falls drastically, the greater number of nonvoters would result in a greater number of misreporters. (Although there is evidence that the propensity to misreport increases with higher turnout (Deufel and Kedar 2010), the difference between the size of the survey-based voting population and that of the actual voting population is generally higher in low-turnout elections, as this paper shows below.) At the same time, the number of actual voters in low-turnout elections is smaller, making for a smaller number in the unbiased share of the survey-based voting population. The net result would be greater bias in low-turnout elections.

### *Non-response bias*

The second explanation for overreport is non-response bias. The standard assumption underlying survey research is that those who don't respond would provide the same answers as respondents. When non-respondents are different, it is no longer legitimate to project a characteristic such as voting from the sample to the full population. This has been generally recognized as a factor affecting survey estimates of voting, but has been subject to less research. Recently, Sciarini and Goldberg (2016) used a population register in Switzerland to provide a broader match to survey respondents and non-respondents and showed that non-respondents are less likely to vote than those who respond.

Since survey response rates typically vary by age, race, and sex, along with other factors, non-response bias can be slightly different for each of these groups. Unfortunately, there has been no good way to detect how differential nonresponse and differential weighting might figure into overall patterns of overreport, because data on voting records by demographic characteristics are not broadly available for comparison.

The effect of weighting is to take the characteristics of interviewed populations and project them to those that weren't reached. If unreached respondents are less likely to vote than respondents, then groups with higher proportions who were not reached would be most affected by weighting. Kojetin (1994) looked at the characteristics of nonrespondents to the CPS by matching them to census records and found that non-responding households were more likely to be single individuals, to be renters, to live in multi-unit buildings, and to be made up of people under 65. Coverage ratios, which indicate how well the total sample (including nonrespondents) match census-based populations, show nonrespondents to be younger, non-White, and male (U.S. Census Bureau 2016a). These considerations would lead to the supposition that non-response and undercoverage would raise the estimated voting rates of younger, single, non-White, male, and non-homeowning populations relative to others. However, this supposition relies on the assumption that the effect of non-response on voting doesn't vary among these groups, which has no basis in theory or evidence.

### *Regression models of overreport*

Although it may not be possible to look directly at non-response bias, it may be beneficial to take a step back and look at overreport generally. One way to do this is to look at voting patterns at the state level in the U.S. Jennings (1990) examined the effect of state-level factors on the degree to which CPS overreported voting, by state, in 1980 to 1988. He found a

negative effect of turnout levels, per capita income, and being in the South, along with a positive relation to the percent of state population that was Black.

In looking at state-level voting reports, it is important to recognize that there is no necessary link between the tendency of a demographic group to misreport their voting status and the demographic makeup of the total overreporting population. Bernstein, Chadha and Montjoy (2001), for example, found that racial characteristics of a congressional district had a distinct effect from individual race on a person's tendency to overreport. Similarly, they found overreport to be more common in the "deep South," particularly among Whites. The demographic makeup of an area, along with other factors, can have an effect on all voters.

Social and demographic characteristics of a state can be expected to affect voting in a number of ways. First, the presence of people with a characteristic associated with vote misreporting can increase it directly. Second, the presence of this group may also affect attitudes of others in a way that might make them feel they should say they voted even if they didn't. Third, the presence of the group might be associated with economic, institutional or political features of a state that might influence the likelihood of overreport. Finally, the demographic composition of a state might also influence non-response bias. Non-response bias would be higher for groups that have low survey response rates, for example.

Regardless of mechanism, however, if there is an association between a state-level characteristic and the degree to which voting is overreported in a state, we can say that some degree of bias is present. Even if this is not so within the state, it is so at the national level. For example, let's say there were a state where the presence of a large Black population made voting contentious and led all groups to claim they voted when they did not. Survey estimates of the composition of the voting population would not be biased for this state because, in this example, all groups are claiming to have

voted more than they actually did. Estimates would be biased at the national level when this high-turnout state with its large Black population was added to national totals.

A stronger statement can be made about bias in those cases where administrative voting data for a state includes demographic characteristics. Simply comparing the estimates of turnout from CPS to those released by the state can show whether bias is present. On the other hand, there are limitations to this method, as well. As mentioned above, researchers have generally found that problems with vote collection and recording are not large enough to create systematic problems with official voting estimates. Less is known, however, about the quality of demographic information associated with state voting records. The forms used to collect data on age, sex, and race do not always use similar approaches and categories as national data sets like CPS. The data come from registration systems that are often controlled at the county or local level, and the reporting of statistics to state authorities might suffer from differential norms and practices. One influence that might have improved this situation is the Help America Vote Act of 2002 (U.S. Electoral Assistance Administration 2016). The legislation provided funding to states to introduce more uniformity in their systems for registration and tallying of votes, and included a requirement that registration records be "computerized" at the state level.

This paper takes a look at voting levels by characteristics measured at the state level, using a regression analysis to examine how these factors affect overreport, followed by an analysis of four states where data on race are available from voter registration systems. The next section discusses data and methods. This is followed by an examination of patterns and trends in overreport and its relation to overall election turnout. This is done first by looking over the years, then across states. Subsequently, the paper examines the results of a set of regressions across states from the period 1980 to 2014 in order to determine whether, and to what extent, the makeup of a state in terms of age, race, education, and income influence the degree of voting overreport. The last analytical section compares CPS

estimates of turnout by race with those of four states which keep records on this topic. This is followed by a discussion of results and conclusion.

## DATA AND METHODS

This paper makes use of CPS data from the voting and registration supplement, administered in November of each even-numbered year. Voting and demographic characteristics are summarized for 50 states and the District of Columbia, covering the period 1980 to 2014. One observation for one year was suppressed as an outlier (Louisiana, 1982). There was no voting on federal offices that November because Louisiana's unique primary system resulted in no congressional or senate seats at stake. The total number of observations for 18 elections across the 51 states was 917. The administrative voting totals were taken from the United States Election Project website, maintained by Michael McDonald (McDonald 2016).

The measure of overreport used here was the gap between administrative and CPS turnout estimates as a percentage of the non-voting population. Expressing overreport in relation to the non-voting population has been advocated by researchers as a better representation of the phenomenon because it relates overreporting to the population that is "at risk" of overreport (Anderson and Silver 1986). The non-voting population was measured as the difference between CPS estimates of the citizen voting-age population and administrative estimates of voters. Data from the November CPS files were used to calculate most control variables, all as a percent of the citizen population 18 and over. These were percent Black, percent Hispanic, percent under 30 years, percent 65 and older, and percent with a college (bachelor's or higher) degree.[3] Median household income by state was taken from tables published by the Census Bureau for years 1984 forward (U.S. Census Bureau 2016b). Analyses

[3] Before 1992, education was measured in years completed, rather than degrees. For those years, people who reported having completed four years of college were treated as having a college degree.

using median household income were restricted to the available years of 1984 to 2014. Voter turnout was calculated as a ratio of votes cast from administrative sources to the CPS estimate of citizen population 18 and over.

A major part of the analysis makes use of multiple regression of voting overreport by state on turnout and other state-level characteristics. Weighted least squares regressions were used, with state citizen voting-age population as the basis for weights. Initial attempts to include all control variables in a single regression were abandoned due to problems of multicollinearity. Instead, a larger number of regressions were used to examine the impact of state-level characteristics one or two at a time.

One influence that operates at the state level, at least to a certain extent, is the level of turnout in an election, which strongly affects overreport, as we have already seen. By looking at state-level overreport in a multiple regression framework we are able to control for the influence of turnout. To see the importance of this, take the influence of education. Research on voter turnout shows that education is positively correlated with turnout (File 2015). Because turnout is negatively associated with overreport, we would expect high education to lead to less overreport. On the other hand, the studies cited above showed that more highly educated people are more likely to report they have voted when they have not. In this case the effects are offsetting – we might see no effect of education on overreport unless we can control for the influence of turnout first. Similar arguments might be made about any of the other socio-demographic influences on voting.

The final analytical section of the report makes use of a collection of data from the states of Florida, Georgia, Louisiana, and North Carolina. Voting records by race and Hispanic origin were available for elections from 1998 to 2014. In Georgia and Louisiana, records were available for all nine

elections, for North Carolina, eight elections, and Florida, two, making for a total of 28 election results for comparison.[4] Results for Hispanic and Asian voters were only available for 16 of these.

## VOTING OVERREPORT IN CPS FROM 1980 TO 2014

Table 1 shows national voting totals from administrative sources and from the CPS. In only one year, the presidential election of 2008, did CPS turnout estimates fall below the number from administrative sources. In other years the gap between CPS and administrative data (shown in column 5) ranged from 2 million to 15 million in favor of the CPS estimate. The last three columns show different ways the gap or "overestimate" can be expressed as a percentage. For reasons explained below, this paper mainly focuses on the overestimate as a percent of people who did not vote.

Figure 1 shows voting as a percentage of the citizen population 18 and over from CPS and the administrative sources. Presidential elections took place every four years from 1980 to 2012 and had high turnout. The other years, termed "congressional" election years by the Census Bureau, had lower turnout. Patterns of turnout over time always show this sawtooth pattern, with peaks in the presidential years and troughs in the congressional years.

The second aspect of the trend shown in Figure 1 is the seeming convergence of the time series. This is visible in Figure 2, which shows the gap or overestimate of turnout in three ways – as a percent of the total citizen population and as a percent of the population who did and did not vote.[5] Expressing

[4] Unlike the other states, which reported turnout by race in a set of tables compiled from election results, the Florida data were available as an extract of registration records as of July 2015, when a request for the extract was submitted. Voters who were removed from the roles due to migration or death during the period from the election to the creation of the extract were not available for analysis. The extract contained 6,000,563 voters, compared with 6,026,802 recorded as having voted in November 2014. The extract contained 8,211,553 voters, compared with 8,538,264 recorded as having voted in November 2012. Especially in the latter case, the possibility of shifts in the composition of the voting population should be taken into account when interpreting results.

[5] The difference between the overreport measured in 1980 to 1996 versus that measured in 1998 to 2014 was significant at the 5% level when overreport was measured as a percentage of non-voters. When measured as a percentage of voters or as a percentage of citizens there was no significant difference between the two periods.

overreport in relation to the non-voting population has attraction especially for those who focus on misreport by non-voters who provide "socially desirable" answers. For those concerned about the possibility of non-response bias, the population of interest would be non-respondents rather than non-voters. Since levels of non-response are not readily available with the data at hand, that will be taken up in later work. For this paper, the focus will be on non-response as a percentage of those not voting.

Moving from the 1980s and early 1990s to the later years in the chart, a decline in voting overreport is seen in relation to the non-voting population. In the earlier years, overreport measured in relation to the non-voting population, seemed to be relatively stable, as if it were a relatively stable characteristic of those who report voting for reasons of social desirability. However, the difference in variance from one period to the next is not statistically significant.

Without further evidence, it is hard to say what exactly is going on here. However, at least part of the trend seems to be related to increased levels of non-response in CPS: McDonald (2014) showed that correcting for non-response largely eliminated the apparent convergence between CPS and administrative estimates of turnout over the years. What is clear is that patterns of overreport have shifted over time, and that they are not fixed in relation to any of the populations measured here.

### *Overreport by state*

The data collected for this project allow us to look at turnout levels by state and the District of Columbia for each election from 1980 to 2014. Figure 3 shows the level of over-report or under-report for all states in a typical year, 1998. States that are a medium hue (orange) had rates of overreport that were within five percentage points of the national average. The darker, red states were five or more points above the average, the light-shaded states five or more points below. Although it cannot be seen on the map, the District of Columbia falls into the red, high overreport, category, along with South Carolina, Mississippi, Tennessee, North Dakota, New Mexico and Colorado. At the other extreme, 11

states were at least five points below the national average in the degree to which CPS overestimated turnout.

The distribution across states does not appear to be random. Some of the states where CPS overestimates voting by the greatest margin are in the South. Some of the states where CPS has lower overestimates of voting (relative to the national average) are sparsely settled states such as Alaska, Vermont, and Wyoming. What is more, the direction of overreport is stable across the period 1980 to 2014, at least for certain states. Mississippi and the District of Columbia were above the national average in all 18 elections in the degree to which CPS overestimated voting. At the other extreme Ohio, Alaska, and Pennsylvania were never significantly above the national average in any of the 18 elections. However, most states had some elections in which they were above the national average and some when they were below.

Overall, the correlations across years in the degree to which CPS overestimated voting among states were moderate (Table 2). Of the 153 pairwise comparisons in Table 2, 101 were significant and positive, ranging from 0.24 to 0.70. Moreover, the pairwise correlations across years very distant from one another were sometimes as large as those in adjacent or nearby years. In four of six cases, the correlations were significant even across a span of 30 years or more.

In summary, we have overreport across states that shows variation from year to year, but that seems to be at least partially stable over time. We can't say exactly what makes for that stability, but it is interesting to note that some of the states with the highest overestimates are ones with large African-American populations (Mississippi, District of Columbia, Alabama, South Carolina). This leads to the hypothesis that the two may be associated.

Thus far we have looked at overreport across time, nationally, and across states, holding time constant. In so doing, we have identified two factors that might be associated with how CPS

overestimates voting – the level of turnout and the size of the African-American population. In the next section we will combine approaches, pooling states and election years into a single analysis.

### *Overreport by social and demographic characteristics*

Table 3 shows the results of a series of regression analyses of the relationship of voting overreport to turnout and state characteristics. Each box in Table 3 represents a separate regression result. In the first column, we see the results of regressing overreport on turnout. The first set of regressions were zero-order regressions (no other control variables). The first regression was on the entire data set, followed by separate regressions for the early part and later part of the timespan under consideration. Confirming what was evident from Figure 2, turnout had a significantly negative effect on overreport overall and in each of the two periods. As was also evident in Figure 2, the relationship between turnout and overreport was smaller in the early period (1980 to 1996) than in the later period (1998 to 2014).

The second set of regressions included dummy variables for states, which means the coefficient of turnout reflects its relationship with overreport over time, both within states and overall. Once again, turnout was significant and negative, with a smaller coefficient in the 1980 to 1996 period than in 1998 to 2014. The third set of regressions includes dummy variables for years, rather than states, so the turnout coefficient is more strongly influenced by cross-sectional differences. Once again, the effect of turnout is negative and the effect varies by time period (1980 to 1996 vs 1998 to 2014).

In a separate regression, not shown here, it was tested whether the effect of turnout was due mainly to the difference between presidential and congressional elections. Separate coefficients for turnout in these two types were both negative, and the effect of turnout was slightly larger in presidential election years (-0.28, compared with -0.24 in congressional election years).

For each of the other variables – percent Black, percent Hispanic, percent with college education, percent under 30, percent 65 and over, median household income -- the same set of regressions appear.[6] In many but not all cases, these variables had a significant effect on overreport. Note that the effect of black population percentage was negative in the regression with state dummies, implying that on a year-to-year basis overreport decreased when Black population increased within a state and vice-versa. Other regressions showed a positive effect of percent Black, indicating that states with higher Black populations were more likely to exhibit overreporting. A similar relationship seems to be evident for percent Hispanic (particularly for the period 1998-2014), but if it exists, it is much more muted, and the overall effect of Hispanic population is to lower overreport. Another result worth mentioning is the reversal in sign between 1982 to 1996 and 1998 to 2014 in the impact of percent of population under 30 in the regressions with year dummy variables. In the first period, states with more young people had lower overreport, in the second period states with more young people had higher overreport.

The bottom panel of the table shows the effect of these variables on overreport after controlling for turnout. Even with this control, several variables still have a significant effect on overreport, including percent Black, percent under 30, and median household income. In the 1998 to 2014 period, five of the six variables under consideration remained significant even with control for turnout. This confirms the evidence from Jennings (1990) that some demographic characteristics are biased in the CPS data.

The demographic variables measured at the state level here didn't always behave in the same way as they do in relation to individual voting misreport, for reasons discussed earlier. The effect of age was found to be positive at the individual level in prior research, but appears to be negative here when aggregated to the state level, even with control for turnout (positive effect of population under 30, zero

[6] Regressions on median household income include data from 1984 to 2014 due to data availability.

or negative for over 65). Education follows the scenario discussed in the "hypothetical" discussion above. It was found to be positive in past research at the individual level, negative here at the state level, but close to zero or even positive when controlling for turnout. The effect of income has generally been found to be positive or zero at the individual level, but measured at the state level here has a negative effect, even with control for turnout. Race and Hispanic origin, on the other hand, generally behave at the state level in the same way they do at the individual, with the effect of Black population on overreport positive, and the effect of Hispanic population weakly negative.

### *Effect on CPS estimates of the composition of the voting population*

Although we have found that there is potential bias in the CPS estimates of voter turnout due to differential association with overreport, this alone does not allow us to estimate the degree to which this bias is manifested in estimates of the composition of the voting population. A high effect of the Black population on overreport could be due to misreports by Blacks themselves, to misreports by Whites in reaction to the presence of Blacks, to general misreporting due to social and institutional factors in states with large Black populations, to aspects of nonresponse bias associated with Black populations, or to other factors.

We can, however, calculate hypothetical estimates if we want to assume that all the overreport found in a state with a certain proportion of people in a demographic group is directly due to misreported voting by individuals in that group. Take, for example, a state with an overall 50 percent turnout rate and a 40 percent turnout rate reported by Blacks, who make up 20 percent of the population. Using the parameters of the bottom regression in the second column of Table 3, we can estimate that the actual turnout of Blacks would be 25 percent rather than the reported 40 percent. The share of the voting population that is Black would appear to be 16 percent in conventional estimates, but 11 percent based on the model. That is to say, the scale of effects that emerge from the

model show the potential (but not necessarily the reality) of moderately substantial differences between CPS estimates of characteristics of the voting population and the underlying reality.

## COMPARISON WITH INDIVIDUAL STATES

Another way to approach the problem is to examine turnout by race for individual states. Most states do not collect data on the race of people who vote, but several do. McDonald (2007) was able to obtain registration files from several states in order to make reasonably accurate comparisons of voter turnout in the election of 2004 between state records and survey estimates, including those from CPS. He was able to compare the proportion of voters belonging to various age groups in 10 states. He found that CPS estimates of the share of the voting population aged 65 and over was generally higher than the administrative reports by two points or less in most states, and that the differences between administrative estimates and CPS estimates were statistically significant in 4 of the 10. In the five states in which he was able to obtain estimates of the share of the voting population by sex, he found no differences between administrative and CPS estimates. In three states, he was able to compare estimates by race and found a significant difference only in North Carolina, where CPS overestimated the black share of the voting population by 3 percentage points relative to administrative reports.

This paper makes use of voting records by race and Hispanic origin from four states for 1998 to 2014. The number of voters recorded by the election authorities in each state is compared to the number of voters estimated from CPS in Table 4. As we have seen already, turnout estimates are higher in CPS overall, and for all but one comparison within race groups, the CPS estimates were the same or higher here. Of the 28 elections, the differences were significant for Whites in 13 cases and for Blacks in 20 cases. Of the 16 elections where comparisons were possible, Asians were higher in CPS in four elections and Hispanics higher in five.

In Table 5, estimates of turnout were obtained by dividing the number of reported voters by the citizen population 18 and over (from CPS). Once again, turnout rates were mostly higher when estimated using CPS data rather than official voting records. This is especially evident for Black voters whose CPS turnout estimate was often higher than the official estimate by 10 percentage points or more (11 of the 28 comparisons). There were similar problems for estimates of Asian and Hispanic turnout rates. The CPS estimates were rather volatile for these groups, varying from 7 percent to 60 percent turnout for Asians, and from 13 to 65 percent for Hispanics, in North Carolina. Note that the administrative turnout estimates for Asians and Hispanics from election authorities varied from year-to-year also. In general, estimates of state-level populations of small populations in the CPS are subject to a high degree of error.

Table 6 focuses on only the Black and White populations, and compares the share each group made of the voting population as estimated from CPS and administrative sources. The results are also summarized in Figure 4. In most cases (17 of 28), the White share of the voting population is understated in CPS relative to the estimate from official counts. In 15 of the 28 cases the Black share is overstated. On the other hand, most of the Black overstatement is not large – in only four cases was it 5 percentage points or more. Those few cases came in the elections of 1998 to 2002.

These results confirm that there is potential bias in CPS estimates of the demographic composition of the voting population. However, in most cases, the bias does not reach the level of the "hypothetical" predictions from the regression models above.

## DISCUSSION

This paper has taken a broad look at overreport of voting in the CPS at the state level. Overreport has been recognized as a problem with CPS voting estimates, as with other survey-based voting estimates, from the inception of data collection in 1964. While the impact of overreport on

estimates of overall turnout has long been recognized, both the Census Bureau and academic users of these data have maintained that the data provide good estimates of the composition of the voting population. This paper, along with others that have looked at different aspects of the problem, do not so much refute this claim but provide a warning to users. It has been shown that CPS estimates of demographic composition of the voting population are biased.

The biases are large enough to be of concern to anyone using the data. Regression models of overall output and comparisons with state reports show that estimates of Black share of the voting population, for example, can be off by 5 percentage points or more. Current estimates of the margin of error for CPS voting estimates understate the problem. Users of CPS data should take the following steps.

1. ***Always take account of margins of error.*** Margins of error understate the total error in estimates of voting but are still useful. Differences that don't exceed the margins of error should not be treated as if they were real.

2. ***Treat margins of error as a minimum.*** Statisticians are familiar with the idea that total survey error is the sum of sampling error and bias. We can't know the size or direction of bias in advance, so caution dictates that we look for differences that are truly large before (tentatively) talking about them as if they were real. How large? We have seen here that bias commonly produces differences of 4 or 5 percentage points with respect to race. McDonald's work (2007) shows slightly lower differences by age and sex.

   It probably does not help to get larger sample sizes or repeated observations. While sampling error diminishes with sample size, bias does not. We have also seen that the

direction of overreport is somewhat stable over time. Thus, high turnout for a specific subgroup election after election might simply reflect a stable aspect of bias.

3. ***Look for corroborating evidence.*** If a difference in voting behavior is observed in CPS data, there could be traces of evidence elsewhere. Was there an organized campaign to stimulate or suppress voting in the group in question? Were there issues or candidates that might have affected turnout for that group? Is there evidence from rallies and public events? Do exit polls show the same difference? Do institutional or historical factors play a role?

In the end, CPS data remain an important source of data on voting behavior in the United States. If used cautiously and judiciously, they can provide important insights into our voting system and the characteristics of the voting population. As with all survey data, there is potential to over-interpret small differences.

Looking at the election of 2004, McDonald (2007) examined both the CPS and the National Election Poll, a large-scale election exit poll conducted jointly by a consortium of national news providers (see Edison Project 2017). Looking at election results by age, sex, and race for those states where records were available, he found them to be "generally consistent" with CPS results. The NEP, on the other hand, was found to show an voting population that is "younger and perhaps containing more minorities" (McDonald 2007 p. 599). This is of interest because the methodology of election polls should, in theory, lack the problems identified with CPS. Since people are being questioned at polling places, there is little question of voting misreport due to "social desirability bias." Nonresponse bias with regard to voting is likewise not a problem. There is no guarantee that CPS is a consistently better source of data than exit polls – 2004 was a good year for CPS, as can be seen from Table 1 and Figure 4.

Nonetheless it is clear that CPS remains an important element in the endeavor to understand voting in the United States.

## REFERENCES


Abramson, Paul R. and William Claggett. 1992. "The Quality of Record Keeping and Racial Differences in Validated Turnout." *Journal of Politics* 54(3):871-80.

Anderson, Barbara A. and Brian D. Silver. 1986. "Measurement and Mismeasurement of the Validity of the Self-reported Vote." *American Journal of Political Science* 80:771-785.

Ansolabehere, Stephen, and Eitan Hersh. 2012. "Validation: What big data reveal about survey misreporting and the real electorate." *Political Analysis*: mps023.

Bauman, Kurt J. and Tiffany Julian. 2010. "A Summary of Data Collection Procedures and Reports of Voter Turnout from the Current Population Survey." *Social, Economic and Housing Statistics Working Paper* 2011-11. U.S. Census Bureau, Washington, D.C.

Belli, Robert F., Michael W. Traugott, and Matthew N. Beckmann. 2001. "What leads to voting overreports? Contrasts of overreporters to validated voters and admitted nonvoters in the American National Election Studies." *Journal of Official Statistics* 17(4): 479.

Bernstein, Robert, Anita Chadha, and Robert Montjoy. 2001. "Overreporting Voting: Why it Happens and Why it Matters." *Public Opinion Quarterly* 65:22-44.

Burden, Barry C. 2000. "Voter turnout and the national election studies." *Political Analysis* 8(4): 389-398.

Cassel, Carol A. 2004. "Voting Records and Validated Voting Studies." *Public Opinion Quarterly* 68(1):102-08.

Clausen, Aage R. 1968. "Response Validity: Vote Report." *Public Opinion Quarterly* 32:588-606.

DeBell, Matthew, Katie Gera, Jon A. Krosnick, David Yeager, and Michael McDonald. 2015. "Facing the Turnout Gap in Surveys: Why So Wrong for So Long?" Unpublished draft.

Deufel, Benjamin J., and Orit Kedar. 2010. "Race and turnout in US elections exposing hidden effects." *Public Opinion Quarterly* 74(2): 286-318.

Edison Project. 2017. "The National Election Pool's exclusive provider of election exit poll data." Web page (accessed 2/21/2017) http://www.edisonresearch.com/election-polling/

File, Thomas. 2015. "Voting and Registration in the Election of November 2014." *Current Population Reports,* Series P20-577. U.S. Census Bureau. Washington, D.C.

Granberg, Donald, and Soren Holmberg. 1991. "Self-reported turnout and voter validation." *American Journal of Political Science* (1991): 448-459.

Hur, Aram and Christopher H. Achen. 2013. "Coding Voter Turnout Responses in the Current Population Survey." *Public Opinion Quarterly* 77(4): 985-993

Jennings, Jerry T. 1990. "Estimating Voter Turnout in the Current Population Survey." Studies in the Measurement of Voter Turnout. *Current Population Reports,* Special Studies. P-23 No. 168. Washington DC: U.S. Census Bureau. Pp. 21-29.

Karp, Jeffrey A., and David Brockington. 2005. "Social desirability and response validity: A comparative analysis of overreporting voter turnout in five countries." *Journal of Politics* 67(3): 825-840.

Kojetin, Brian A. 1994. "Characteristics of Non-Respondents to the Current Population Survey (CPS) and Consumer Expenditure Interview Survey (CEIS)." *Proceedings of the American Statistical Association* 966-971.

McDonald, Michael P. "United States Election Project." < http://www.electproject.org/>. Accessed on July 21, 2016.

McDonald, Michael P. 2007. "The true electorate a cross-validation of voter registration files and election survey demographics." *Public Opinion Quarterly* 71(4): 588-602.

McDonald, Michael P. 2014. "What's Wrong with the CPS?" Presented at the 2014 American Political Science Association Conference, Washington, D.C. (August)

Powers, Mary G. and Richard W. Dodge. 1965. "Voter Participation in the National Election: November 1964." *Current Population Reports*, Series P-20 No. 143. U.S. Census Bureau, Washington, D.C.

Sciarini, Pascal, and Andreas C. Goldberg. 2016. "Turnout Bias in Postelection Surveys: Political Involvement, Survey Participation, and Vote Overreporting." *Journal of Survey Statistics and Methodology* 4(1): 110-137.

Selb, Peter, and Simon Munzert. 2013. "Voter Overrepresentation, Vote Misreporting, and Turnout Bias in Postelection Surveys." *Electoral Studies* 32(1): 186-196.

Sigelman, Lee. "The nonvoting voter in voting research." *American Journal of Political Science* (1982): 47-56.

Silver, Brian D., Barbara A. Anderson, and Paul R. Abramson. 1986. "Who overreports voting?." *American Political Science Review* 80(02): 613-624.

Tourangeau, Roger, Robert M. Groves, and Cleo D. Redline. 2010. "Sensitive topics and reluctant respondents demonstrating a link between nonresponse bias and measurement error." *Public Opinion Quarterly* (2010): nfq004.

Traugott, Michael W., and John P. Katosh. 1979. "Response validity in surveys of voting behavior." *Public Opinion Quarterly* 43(3): 359-377.

U.S. Census Bureau. 2016a. "Current Population Survey November 2016 Voting and Registration File Technical Documentation." CPS—16. https://www2.census.gov/programs-surveys/cps/techdocs/cpsnov16.pdf (Accessed March 2018)

U.S. Census Bureau. 2016b. "Table H-8. Median Household Income by State." From the page, Historical Income Tables: Households. http://www.census.gov/data/tables/time-series/demo/income-poverty/historical-income-households.html (Accessed December 2016)

U.S. Electoral Assistance Administration. 2016. "Help America Vote Act." https://www.eac.gov/about_the_eac/help_america_vote_act.aspx (Accessed December 2016)

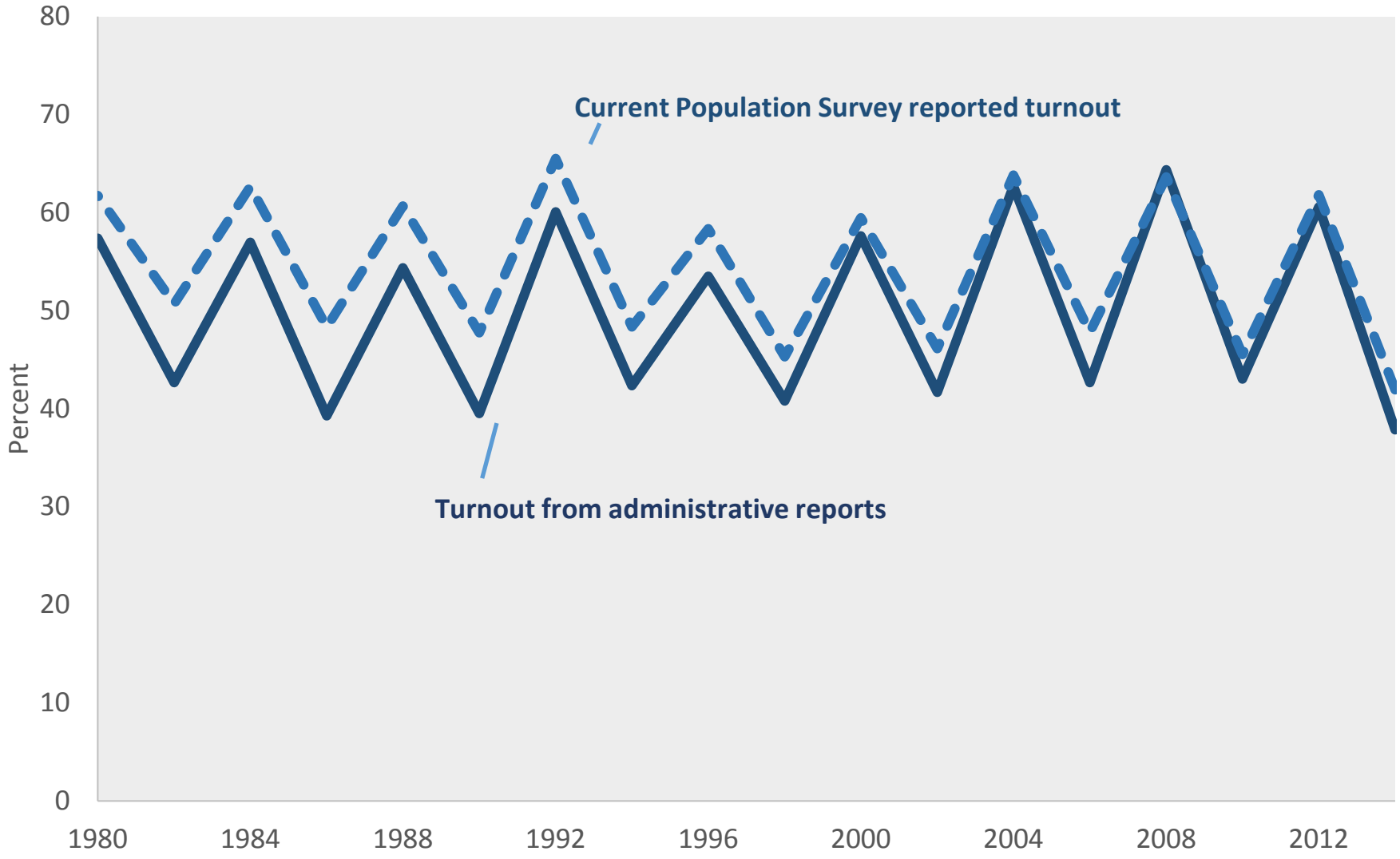
Table 1. United States Administrative Vote Estimates (Ballots Counted) and Current Population Survey (CPS) Vote Estimates, 1980-2014
Numbers in thousands

| Year | Number of voters | | Adult citizens (CPS estimate) | Non-Voters (adult citizens minus administrative count of voters) | CPS overestimate | | | |
|---|---|---|---|---|---|---|---|---|
| | Administrative count | Current Population Survey | | | Number | Overestimate as percent of voters (Administrative count) | Overestimate as percent of non-voters | Overestimate as percent of citizens |
| 2014 | 83,262 | 92,251 ** | 219,941 | 136,679 | 8,989 | 10.8 | 6.6 | 4.1 |
| 2012 | 130,292 | 132,948 ** | 215,081 | 84,789 | 2,656 | 2.0 | 3.1 | 1.2 |
| 2010 | 90,912 | 95,987 ** | 210,800 | 119,888 | 5,075 | 5.6 | 4.2 | 2.4 |
| 2008 | 132,609 | 131,144 ** | 206,072 | 73,463 | -1,465 | -1.1 | -2.0 | -0.7 |
| 2006 | 85,769 | 96,119 ** | 201,073 | 115,304 | 10,350 | 12.1 | 9.0 | 5.1 |
| 2004 | 123,536 | 125,736 ** | 197,005 | 73,469 | 2,200 | 1.8 | 3.0 | 1.1 |
| 2002 | 80,295 | 88,903 ** | 192,656 | 112,361 | 8,608 | 10.7 | 7.7 | 4.5 |
| 2000 | 107,390 | 110,826 ** | 186,359 | 78,969 | 3,436 | 3.2 | 4.4 | 1.8 |
| 1998 | 74,826 | 83,098 ** | 183,451 | 108,625 | 8,272 | 11.1 | 7.6 | 4.5 |
| 1996 | 96,263 | 105,017 ** | 179,935 | 83,672 | 8,754 | 9.1 | 10.5 | 4.9 |
| 1994 | 75,106 | 85,702 ** | 177,260 | 102,154 | 10,596 | 14.1 | 10.4 | 6.0 |
| 1992 | 104,405 | 113,866 ** | 173,784 | 69,379 | 9,461 | 9.1 | 13.6 | 5.4 |
| 1990 | 67,859 | 81,991 ** | 171,659 | 103,800 | 14,132 | 20.8 | 13.6 | 8.2 |
| 1988 | 91,595 | 102,224 ** | 168,495 | 76,900 | 10,629 | 11.6 | 13.8 | 6.3 |
| 1986 | 64,991 | 79,954 ** | 165,438 | 100,447 | 14,963 | 23.0 | 14.9 | 9.0 |
| 1984 | 92,653 | 101,878 ** | 162,627 | 69,974 | 9,225 | 10.0 | 13.2 | 5.7 |
| 1982 | 67,616 | 80,310 ** | 158,424 | 90,808 | 12,694 | 18.8 | 14.0 | 8.0 |
| 1980 | 86,515 | 93,066 ** | 150,742 | 64,227 | 6,551 | 7.6 | 10.2 | 4.3 |

Source: McDonald, Michael, United States Elections Project; U.S. Census Bureau, Current Population Survey

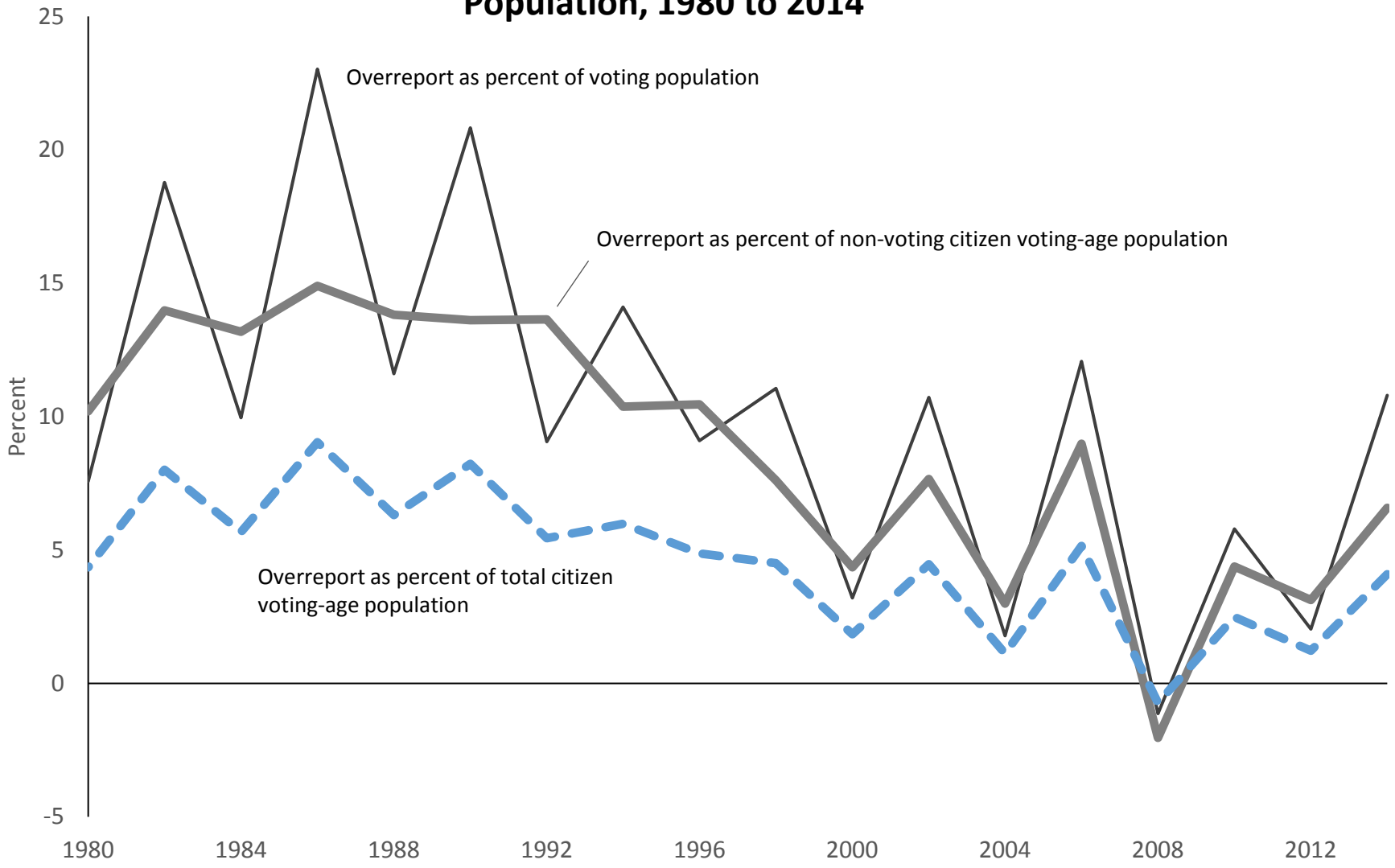
Difference between CPS estimates and administrative counts are significant at the .05 level **, .10 level *, not significant -.

## Figure 1: Reported Turnout in National Elections as Percent of Citizen Population 18 and Older, 1980 to 2014




Source: U.S. Census Bureau, Current Population Survey, November 1980 to November 2014, U.S. Election Project: www.electproject.org

## Figure 2: Voting Overreport as Percent of Citizen Population 18 and Over, as Percent of Voting Population, and as Percent of Non-Voting Population, 1980 to 2014




Source: U.S. Census Bureau, Current Population Survey, November 1980 to November 2014, U.S. Election Project: www.electproject.org

Table 2.

Correlations Between Overreport Rates Across States, by Year, 1980 to 2014

| | 1980 | 1982 | 1984 | 1986 | 1988 | 1990 | 1992 | 1994 | 1996 | 1998 | 2000 | 2002 | 2004 | 2006 | 2008 | 2010 | 2012 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1982 | **0.52** ** | | | | | | | | | | | | | | | | |
| 1984 | **0.60** ** | **0.61** ** | | | | | | | | | | | | | | | |
| 1986 | 0.15 - | 0.32 ** | 0.38 ** | | | | | | | | | | | | | | |
| 1988 | **0.43** ** | **0.47** ** | 0.40 ** | 0.19 - | | | | | | | | | | | | | |
| 1990 | 0.22 - | 0.24 * | 0.31 ** | 0.35 ** | 0.32 ** | | | | | | | | | | | | |
| 1992 | **0.45** ** | **0.55** ** | **0.49** ** | 0.11 - | 0.37 ** | 0.23 - | | | | | | | | | | | |
| 1994 | 0.12 - | -0.03 - | 0.08 - | 0.20 - | -0.22 - | 0.15 - | 0.00 - | | | | | | | | | | |
| 1996 | 0.11 - | 0.16 - | 0.35 ** | 0.07 - | 0.06 - | 0.29 ** | 0.39 ** | **0.41** ** | | | | | | | | | |
| 1998 | 0.35 ** | 0.22 - | 0.34 ** | 0.21 - | 0.14 - | 0.34 ** | 0.26 * | **0.41** ** | **0.43** ** | | | | | | | | |
| 2000 | 0.29 ** | 0.33 ** | **0.46** ** | 0.09 - | 0.25 * | 0.21 - | **0.43** ** | 0.20 - | 0.57 ** | 0.39 ** | | | | | | | |
| 2002 | 0.36 ** | 0.32 ** | **0.44** ** | 0.34 ** | 0.18 - | 0.23 - | 0.20 - | 0.17 - | 0.38 ** | 0.19 - | **0.55** ** | | | | | | |
| 2004 | 0.18 - | 0.26 * | **0.46** ** | 0.38 ** | -0.04 - | 0.21 - | 0.39 ** | 0.27 * | 0.34 ** | 0.28 ** | **0.55** ** | **0.41** ** | | | | | |
| 2006 | 0.09 - | 0.18 - | 0.37 ** | 0.39 ** | -0.03 - | 0.12 - | 0.34 ** | 0.33 ** | 0.35 ** | 0.30 ** | **0.40** ** | 0.36 ** | **0.52** ** | | | | |
| 2008 | 0.10 - | 0.18 - | 0.39 ** | 0.15 - | 0.05 - | 0.11 - | 0.36 ** | 0.23 - | **0.42** ** | 0.16 - | **0.57** ** | 0.35 ** | **0.59** ** | **0.57** ** | | | |
| 2010 | 0.24 * | 0.27 * | 0.26 * | **0.51** ** | 0.22 - | 0.24 * | 0.33 ** | 0.13 - | 0.10 - | 0.05 - | 0.26 * | **0.44** ** | **0.51** ** | 0.19 - | **0.54** ** | | |
| 2012 | 0.34 ** | 0.28 ** | **0.45** ** | 0.35 ** | 0.06 - | 0.27 * | 0.32 ** | 0.28 ** | 0.30 ** | 0.26 * | **0.52** ** | **0.50** ** | **0.66** ** | **0.45** ** | **0.70** ** | **0.63** ** | |
| 2014 | -0.02 - | 0.28 ** | 0.22 - | 0.14 - | 0.03 - | 0.09 - | 0.32 ** | 0.34 ** | **0.48** ** | **0.43** ** | **0.50** ** | 0.31 ** | 0.30 ** | **0.46** ** | **0.55** ** | 0.26 * | 0.39 ** |

Source: U.S. Census Bureau, Current Population Survey, November supplement files 1980 to 2014. Turnout data from United States Elections Project (McDonald).

Correlation coefficient significant at .10 level *, at .05 level **. Not significant -.

Correlations greater than or equal to 0.40 are shown in bold.

| Table 3: Regression results: Voting Overreport as Percent of Non-Voting Population by State, 1980 to 2014 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | State turnout rate (administrative count) | Percent of state population Black | Percent of state population Hispanic | Percent of state population with college education | Percent of state population under 30 | Percent of state population 65 or older | State median household income[1] |
| Zero order regressions | | | | | | | | |
| All years | Coefficient | -0.223 ** | 0.196 ** | -0.092 ** | -0.400 ** | 0.670 ** | -0.444 ** | -0.083 ** |
| | R-squared | 0.134 | 0.053 | 0.011 | 0.133 | 0.129 | 0.029 | 0.028 |
| 1980 to 1996 | Coefficient | -0.115 ** | 0.213 ** | -0.034 - | -0.048 - | 0.045 - | 0.070 - | -0.058 ** |
| | R-squared | 0.049 | 0.101 | 0.001 | 0.002 | 0.001 | 0.001 | 0.025 |
| 1998 to 2014 | Coefficient | -0.310 ** | 0.227 ** | 0.032 - | -0.197 ** | 0.644 ** | -0.410 ** | -0.099 ** |
| | R-squared | 0.312 ‡ | 0.073 | 0.002 | 0.024 † | 0.041 ‡ | 0.024 ‡ | 0.040 |
| Regressions with state dummies | | | | | | | | |
| All years | Coefficient | -0.203 ** | -1.271 ** | -0.697 ** | -0.583 ** | 0.639 ** | -0.579 ** | -0.033 - |
| | R-squared | 0.304 | 0.282 | 0.309 | 0.385 | 0.316 | 0.236 | 0.221 |
| 1980 to 1996 | Coefficient | -0.075 ** | -0.681 ** | -0.156 - | 0.037 - | -0.020 - | 0.406 ** | -0.028 - |
| | R-squared | 0.347 | 0.361 | 0.334 | 0.333 | 0.333 | 0.343 | 0.324 |
| 1998 to 2014 | Coefficient | -0.284 ** | -0.220 - | -0.282 ** | -0.226 ** | -0.233 - | 0.117 - | 0.057 - |
| | R-squared | 0.557 ‡ | 0.350 | 0.355 | 0.357 ‡ | 0.351 | 0.350 | 0.350 |
| Regressions with year dummies | | | | | | | | |
| All years | Coefficient | -0.278 ** | 0.223 ** | 0.032 - | -0.069 - | 0.229 ** | -0.138 * | -0.082 ** |
| | R-squared | 0.438 | 0.422 | 0.355 | 0.356 | 0.359 | 0.356 | 0.407 |
| 1980 to 1996 | Coefficient | -0.200 ** | 0.219 ** | -0.014 - | 0.037 - | -0.302 ** | 0.109 - | -0.057 ** |
| | R-squared | 0.136 | 0.158 | 0.052 | 0.052 | 0.064 | 0.054 | 0.108 |
| 1998 to 2014 | Coefficient | -0.392 ** | 0.228 ** | 0.051 * | -0.139 ** | 0.703 ** | -0.428 ** | -0.102 ** |
| | R-squared | 0.404 ‡ | 0.338 | 0.269 | 0.274 ‡ | 0.313 ‡ | 0.286 ‡ | 0.307 † |
| Regressions with year dummies and control for state turnout | | | | | | | | |
| All years | Coefficient | | 0.147 ** | -0.028 - | 0.024 - | 0.155 * | -0.099 - | -0.045 ** |
| | R-squared | | 0.462 | 0.439 | 0.438 | 0.440 | 0.439 | 0.462 |
| 1980 to 1996 | Coefficient | | 0.163 ** | -0.032 - | 0.119 * | -0.200 * | 0.095 - | -0.035 * |
| | R-squared | | 0.181 | 0.137 | 0.143 | 0.141 | 0.138 | 0.150 |
| 1998 to 2014 | Coefficient | | 0.143 ** | -0.058 ** | -0.023 - | 0.356 ** | -0.282 ** | -0.048 ** |
| | R-squared | | 0.430 | 0.409 | 0.404 ‡ | 0.415 ‡ | 0.413 ‡ | 0.412 |

Source: U.S. Census Bureau, Current Population Survey, November supplement files 1980 to 2014. Turnout data from United States Elections Project (McDonald).
Regression coefficient significant at .10 level *, at .05 level **. Not significant -.
Difference between coefficient for 1980 to 1996 and coefficient for 1998 to 2014 significant at the .10 level †, at .05 level ‡.

1. Regressions on median household income cover years 1984 to 2014, due to data availability.

Table 4: Estimates of the Number of Voters by Race and Hispanic Origin from Current Population Survey and Administrative Sources, Selected States, 1998-2014

Numbers in thousands

| | White | | Black | | Asian | | Hispanic | |
|---|---|---|---|---|---|---|---|---|
| | CPS | State | CPS | State | CPS | State | CPS | State |
| Florida | | | | | | | | |
| 1998 | | | | | | | | |
| 2000 | | | | | | | | |
| 2002 | | | | | | | | |
| 2004 | | | | | | | | |
| 2006 | | | | | | | | |
| 2008 | | | | | | | | |
| 2010 | | | | | | | | |
| 2012 | 5,430 - | 5,336 | 1,104 - | 1,119 | 155 - | 119 | 1,399 ** | 1,007 |
| 2014 | 4,237 - | 4,376 | 929 ** | 734 | 128 ** | 71 | 892 ** | 592 |
| Georgia | | | | | | | | |
| 1998 | 1,377 - | 1,383 | 700 ** | 416 | | | | |
| 2000 | 1,788 ** | 1,993 | 1,008 ** | 616 | | | | |
| 2002 | 1,695 * | 1,537 | 684 ** | 459 | | | | |
| 2004 | 2,327 - | 2,345 | 919 - | 834 | 31 - | 21 | 26 - | 18 |
| 2006 | 1,871 ** | 1,556 | 722 ** | 514 | 26 - | 11 | 37 - | 12 |
| 2008 | 2,683 * | 2,522 | 1,309 - | 1,183 | 53 - | 36 | 128 ** | 44 |
| 2010 | 1,845 - | 1,739 | 929 ** | 741 | 23 - | 17 | 46 - | 19 |
| 2012 | 2,590 ** | 2,399 | 1,340 ** | 1,168 | 83 ** | 40 | 114 ** | 52 |
| 2014 | 1,835 ** | 1,648 | 925 ** | 743 | 62 * | 21 | 62 - | 26 |
| Louisiana | | | | | | | | |
| 1998 | 724 - | 680 | 450 ** | 297 | | | | |
| 2000 | 1,441 ** | 1,262 | 565 * | 472 | | | | |
| 2002 | 1,045 ** | 913 | 434 ** | 328 | | | | |
| 2004 | 1,430 - | 1,363 | 606 - | 532 | | | | |
| 2006 | 849 ** | 705 | 319 ** | 227 | | | | |
| 2008 | 1,458 * | 1,329 | 625 - | 584 | | | | |
| 2010 | 1,087 ** | 911 | 480 ** | 348 | | | | |
| 2012 | 1,389 - | 1,325 | 692 - | 617 | | | | |
| 2014 | 1,065 - | 1,029 | 533 ** | 436 | | | | |
| North Carolina | | | | | | | | |
| 1998 | | | | | | | | |
| 2000 | 2,293 - | 2,396 | 629 ** | 511 | 2 - | 3 | 22 - | 4 |
| 2002 | 1,967 - | 1,916 | 497 ** | 400 | 13 - | 3 | 22 - | 3 |
| 2004 | 2,713 - | 2,789 | 784 ** | 661 | 43 ** | 13 | 27 - | 18 |
| 2006 | 1,986 ** | 1,653 | 408 ** | 316 | 9 - | 5 | 9 - | 7 |
| 2008 | 3,270 - | 3,169 | 919 - | 967 | 43 - | 27 | 77 - | 44 |
| 2010 | 2,207 * | 2,065 | 582 - | 541 | 31 - | 10 | 80 ** | 16 |
| 2012 | 3,194 - | 3,229 | 1,203 ** | 1,048 | 34 - | 35 | 95 - | 62 |
| 2014 | 2,290 - | 2,179 | 735 * | 628 | 26 - | 17 | 53 - | 26 |

Source: U.S. Census Bureau, Current Population Survey; Florida Division of Elections, Voter extract file; Georgia Secretary of State, Voter registration system tabulations; Louisiana Secretary of State, Post-election statistical report; North Carolina Board of Elections, Election data files.
Difference between CPS estimates and administrative counts are significant at the .05 level **, .10 level *, not significant -.

Table 5: Estimates of the Turnout Rate (Voters as a Percent of Citizens 18 and Older) by Race and Hispanic Origin from Current Population Survey and Administrative Sources, Selected States, 1998-2014

| | White | | Black | | Asian | | Hispanic | |
|---|---|---|---|---|---|---|---|---|
| | CPS | State | CPS | State | CPS | State | CPS | State |
| Florida | | | | | | | | |
| 1998 | | | | | | | | |
| 2000 | | | | | | | | |
| 2002 | | | | | | | | |
| 2004 | | | | | | | | |
| 2006 | | | | | | | | |
| 2008 | | | | | | | | |
| 2010 | | | | | | | | |
| 2012 | 61.9 - | 60.9 | 57.6 - | 58.3 | 43.0 * | 33.0 | 62.3 ** | 44.8 |
| 2014 | 47.5 * | 49.0 | 44.0 ** | 34.8 | 43.0 ** | 23.8 | 36.0 ** | 23.9 |
| Georgia | | | | | | | | |
| 1998 | 38.3 - | 38.4 | 40.9 ** | 24.3 | | | | |
| 2000 | 50.3 ** | 56.1 | 53.2 ** | 32.5 | | | | |
| 2002 | 43.6 ** | 39.5 | 40.2 ** | 27.0 | | | | |
| 2004 | 58.3 - | 58.7 | 57.5 - | 52.2 | 23.4 - | 15.6 | 30.6 - | 21.3 |
| 2006 | 46.9 ** | 39.0 | 40.7 ** | 28.9 | 25.7 - | 11.1 | 21.0 * | 6.6 |
| 2008 | 64.1 ** | 60.3 | 67.9 ** | 61.3 | 34.1 - | 23.5 | 54.5 ** | 18.6 |
| 2010 | 43.7 * | 41.1 | 46.8 ** | 37.4 | 14.6 - | 10.6 | 23.9 - | 9.9 |
| 2012 | 62.0 ** | 57.4 | 65.0 ** | 56.6 | 41.9 ** | 20.1 | 48.7 ** | 22.2 |
| 2014 | 44.8 ** | 40.3 | 42.9 ** | 34.5 | 25.7 ** | 8.6 | 29.8 - | 12.7 |
| Louisiana | | | | | | | | |
| 1998 | 36.3 - | 34.1 | 46.3 ** | 30.5 | | | | |
| 2000 | 67.5 ** | 59.2 | 63.3 ** | 52.9 | | | | |
| 2002 | 52.5 ** | 45.9 | 47.5 ** | 36.0 | | | | |
| 2004 | 65.6 * | 62.5 | 62.7 ** | 55.0 | | | | |
| 2006 | 41.9 ** | 34.8 | 36.3 ** | 25.8 | | | | |
| 2008 | 72.4 ** | 66.0 | 66.2 - | 61.9 | | | | |
| 2010 | 51.4 ** | 43.1 | 48.9 ** | 35.5 | | | | |
| 2012 | 65.1 * | 62.1 | 69.5 ** | 62.0 | | | | |
| 2014 | 51.3 - | 49.5 | 51.7 ** | 42.3 | | | | |
| North Carolina | | | | | | | | |
| 1998 | | | | | | | | |
| 2000 | 60.0 * | 62.7 | 47.6 ** | 38.7 | 7.3 - | 12.1 | 25.9 - | 4.3 |
| 2002 | 46.4 - | 45.2 | 42.9 ** | 34.6 | 30.6 - | 6.6 | 26.1 - | 4.1 |
| 2004 | 62.3 - | 64.0 | 64.6 ** | 54.5 | 50.8 ** | 15.5 | 24.4 - | 16.2 |
| 2006 | 44.1 ** | 36.7 | 32.2 ** | 25.0 | 21.4 - | 12.0 | 13.1 - | 10.7 |
| 2008 | 68.3 * | 66.2 | 68.3 - | 71.9 | 53.3 - | 33.1 | 65.4 ** | 37.2 |
| 2010 | 47.0 ** | 44.0 | 41.0 - | 38.1 | 32.4 * | 11.1 | 43.8 ** | 9.0 |
| 2012 | 66.3 - | 67.0 | 80.2 ** | 69.8 | 59.1 - | 62.2 | 56.0 * | 36.8 |
| 2014 | 47.2 * | 44.9 | 48.1 ** | 41.1 | 25.9 - | 17.5 | 26.0 - | 12.8 |

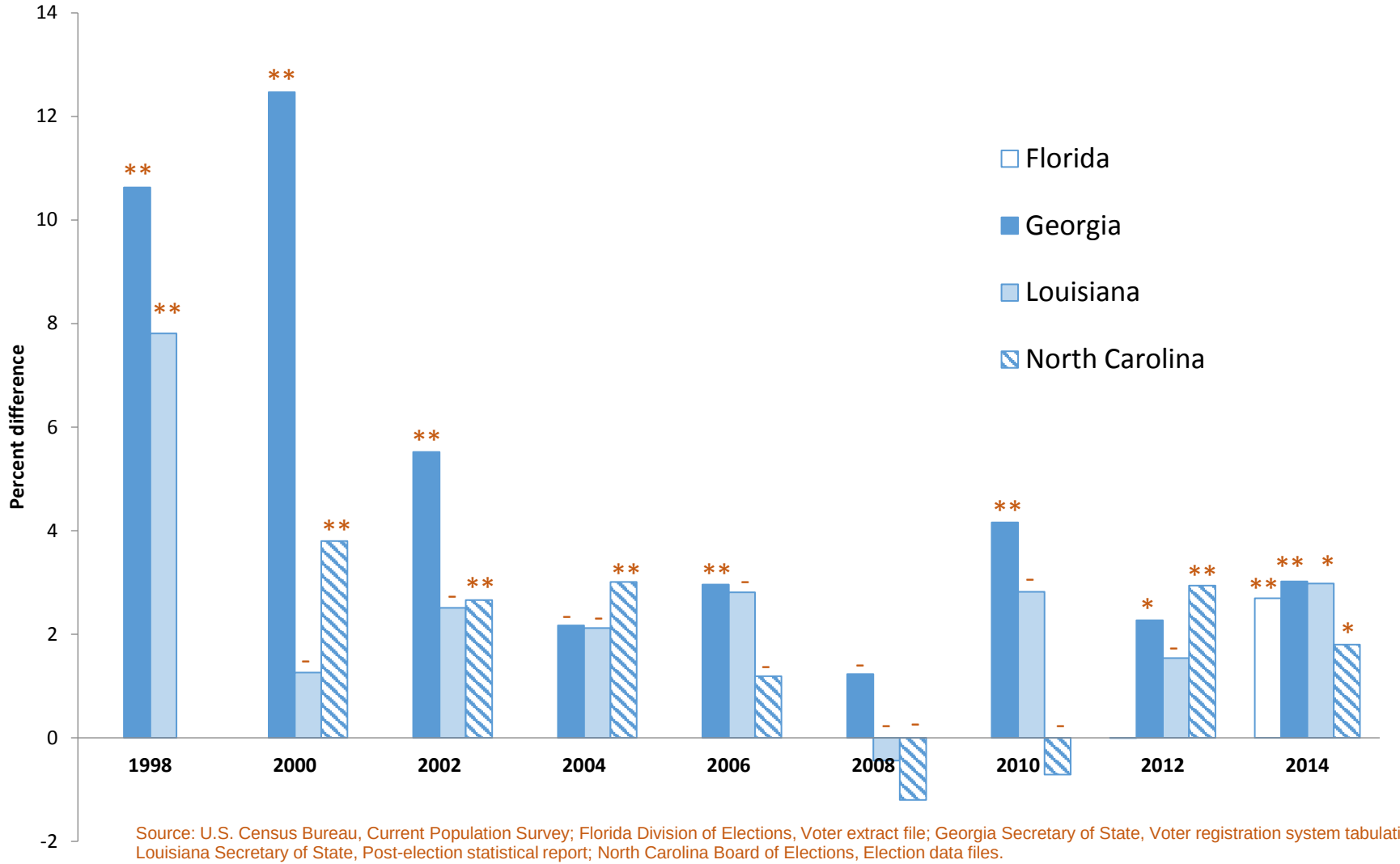
Source: U.S. Census Bureau, Current Population Survey; Florida Division of Elections, Voter extract file; Georgia Secretary of State, Voter registration system tabulations; Louisiana Secretary of State, Post-election statistical report; North Carolina Board of Elections, Election data files.
Difference between CPS estimates and administrative counts are significant at the .05 level **, .10 level *, not significant -.

Table 6: Estimates of the Vote Share (Voters in a Group as a Percent of All Voters) by Race and Hispanic Origin from Current Population Survey and Administrative Sources, Selected States, 1998-2014

| | White | | Black | |
|---|---|---|---|---|
| | CPS | State | CPS | State |
| Florida | | | | |
| 1998 | | | | |
| 2000 | | | | |
| 2002 | | | | |
| 2004 | | | | |
| 2006 | | | | |
| 2008 | | | | |
| 2010 | | | | |
| 2012 | 67.0 ** | 65.0 | 13.6 - | 13.6 |
| 2014 | 68.1 ** | 72.9 | 14.9 ** | 12.2 |
| Georgia | | | | |
| 1998 | 65.9 ** | 76.1 | 33.5 ** | 22.9 |
| 2000 | 63.3 ** | 75.1 | 35.7 ** | 23.2 |
| 2002 | 69.7 ** | 75.7 | 28.1 ** | 22.6 |
| 2004 | 69.8 - | 71.4 | 27.6 - | 25.4 |
| 2006 | 70.0 ** | 72.9 | 27.0 ** | 24.1 |
| 2008 | 64.1 - | 64.1 | 31.3 - | 30.1 |
| 2010 | 64.4 * | 66.3 | 32.4 ** | 28.3 |
| 2012 | 62.1 - | 61.4 | 32.2 * | 29.9 |
| 2014 | 62.9 - | 63.5 | 31.7 ** | 28.7 |
| Louisiana | | | | |
| 1998 | 60.7 ** | 68.7 | 37.8 ** | 29.9 |
| 2000 | 70.9 - | 71.0 | 27.8 - | 26.6 |
| 2002 | 68.4 ** | 72.1 | 28.4 - | 25.9 |
| 2004 | 69.2 - | 69.7 | 29.3 - | 27.2 |
| 2006 | 70.7 ** | 73.8 | 26.6 - | 23.8 |
| 2008 | 67.9 - | 67.1 | 29.1 - | 29.5 |
| 2010 | 67.2 ** | 70.3 | 29.7 - | 26.8 |
| 2012 | 64.6 - | 65.8 | 32.2 - | 30.6 |
| 2014 | 63.5 ** | 68.0 | 31.8 * | 28.8 |
| North Carolina | | | | |
| 1998 | | | | |
| 2000 | 76.5 ** | 80.6 | 21.0 ** | 17.2 |
| 2002 | 77.5 ** | 81.1 | 19.6 ** | 16.9 |
| 2004 | 74.5 ** | 78.2 | 21.5 ** | 18.5 |
| 2006 | 82.0 - | 81.8 | 16.8 - | 15.6 |
| 2008 | 74.8 ** | 72.9 | 21.0 - | 22.2 |
| 2010 | 73.3 ** | 76.5 | 19.3 - | 20.0 |
| 2012 | 69.1 * | 71.1 | 26.0 ** | 23.1 |
| 2014 | 72.3 ** | 74.1 | 23.2 * | 21.4 |

Source: U.S. Census Bureau, Current Population Survey; Florida Division of Elections, Voter extract file; Georgia Secretary of State, Voter registration system tabulations; Louisiana Secretary of State, Post-election statistical report; North Carolina Board of Elections, Election data files.
Difference between CPS estimates and administrative counts are significant at the .05 level **, .10 level *, not significant -.

## Figure 4: Percentage Point Difference in Estimates of Black Share of Voting Population from Current Population Survey and from State Voting Records, Florida, Georgia, Louisiana, and North Carolina, 1998-2014




Source: U.S. Census Bureau, Current Population Survey; Florida Division of Elections, Voter extract file; Georgia Secretary of State, Voter registration system tabulations; Louisiana Secretary of State, Post-election statistical report; North Carolina Board of Elections, Election data files.
Difference between CPS share estimate and that reported by state is significant at the .05 level **, .10 level *, not significant -.