Vote Overreporting

The Statistical and Policy Implications

Emily Jones

Introduction

This paper analyzes regional variations in vote overreporting. Vote overreporting occurs when an individual who did not vote claims to have voted on a survey. Survey data is the lifeblood of quantitative analysis and, not surprisingly, the validity of survey data holds great importance for researchers. Inaccurate data leads to inaccurate conclusions which lead, in turn, to flawed policies. An examination of vote overreporting is obviously a compelling topic in an election year. But in a broader sense, studies of voting behavior are important as an indication of the way in which the public exercises one of the most fundamental rights in a democracy, the right to vote. By accurately analyzing voting behavior, researchers can help contribute to the body of knowledge about voting behavior, with the ultimate goal of understanding and encouraging voting.

Self- or proxy-reports can lead to both random error and systematic bias. While random error might make associations between the dependent variable and any independent variables appear weaker by creating noise in the data, the systematic bias that results from demographically or regionally based patterns of vote overreporting can alter regression results quite dramatically, according to some studies. As a result, the study conclusions

may be inaccurate (Bernstein et al. 2001).

Previous research using individual-level National Election Survey (NES) data found demographic and regional patterns of overreporting. Since NES no longer validates voting as reported by respondents, this paper uses census data from the Current Population Survey (CPS) to examine regional variations in vote overreporting in the 2000 and 2004 U.S. Presidential elections. The main hypothesis, substantiated by a literature review, is that overreporting is highest in the South. This analysis uses two ways to measure overreporting, in reference to the whole population and in reference to nonvoters only. Both methods align with past trends, confirming the hypothesis that overreporting was slightly higher in the South for the 2000 election. The research finds no comparable statistically significant relationship for the 2004 presidential election.

Historical Background: Vote Overreporting

The National Election Studies

Researchers prize individual-level data that matches each reported vote with the individual's true election day action (or inaction). Individual-level data is much more useful than population-level data like the CPS because analysts can test hypotheses about voting behavior with regressions that control for individual-level characteristics and make corresponding generalizations.[1] With respect to research on voting, most recent studies on overreporting use the American National Election Studies (NES), an individual-level dataset on voting behavior.

To help researchers analyze the problem of vote overreporting, the NES conducted ten validated studies on presidential voting between 1964 and 1990. Following the survey, field researchers traveled to local election offices and confirmed whether specific respondents actually cast a ballot. Although it assisted policymakers attempting to understand voter overreporting and voter turnout, this survey methodology was very resource-intensive. Therefore, it seems unlikely that funding for this method of inquiry will return.

Toward the end of the 1980s, social researchers' intense interest in vote overreporting died out because many did not feel that bias from overreporting substantially affected analysis. As a result, the NES halted the validated studies in 1990. Some observers contend that "the discipline was premature in abandoning the collection of validated votes" (Bernstein et al. 2001, 24). Individual-level data may have been of a higher quality in this field of research than regional data, and no matter how robust the analytical methods, low-quality data can lead to erroneous conclusions.

Several researchers assert that some of the alleged overreporting in NES data might be false, arising instead from variations in the quality of voting records (Silver et al. 1986). This might be applicable to allegedly higher overreporting among blacks, southerners, and inner city residents. Such groups are more likely to live in areas where officials poorly maintain voting records, causing survey researchers who attempt vote validation to fail to accurately determine the number of individual votes. A review of the literature found only one study on the topic, conducted by Cassel (2004), which did not support the hypothesis that official record-keeping significantly impacted NES overreporting data. However, this study was seriously flawed as the researcher noted that the variables capturing the quality of voting records were poor indicators.[2] The field would benefit from additional studies on this topic. For the purposes of this study, which reviews data on the 2000 and 2004 Presidential elections, individual-level research does not exist.

Limitations of the Data

Since individual-level data does not exist, this study relies on Current Population Survey (CPS) data which may be doubly impacted by social desirability bias. First, people often overreport activities that society favorably views, such as voting. *Social desirability bias* refers to this issue and occurs when individuals change their survey answers to more closely conform to social norms. Second, the act of voting might be associated with survey participation if both actions correlate with political interest (Vooght and Saris 2003). Therefore, individuals who are likely to participate in the CPS may also be more likely to value political engagement. Such individuals

may be more likely to overreport voting than is the general population, though researchers cannot confirm this since respondents can refuse the CPS.[3] The CPS has a refusal rate of 13 percent, which may limit the validity of the data (Highton 2005, 113).

Data on the topic of vote overreporting may be generally biased. An additional challenge to credible data reflects human behavior. Overreporting does not necessarily mean that individuals intentionally lie about whether or not they voted. Instead, some genuinely cannot remember, especially if substantial time has elapsed between the day of the election and the day of the survey (Belli et al. 1999).

Social Desirability and Vote Overreporting

Researchers attempt to reduce overreporting within the constraints of state or regional studies by adjusting data collection methods and survey wording. Academic inquiry in other subject areas shows that the specific collection methods for survey data can affect the amount of overreporting. However, it appears to matter little how investigators collect survey data, whether in person (with one or several surveyors), by mail, or on the telephone (Silver et al. 1986).

Presser (1990) found no reduction in overreporting from changing the wording, content, and order of survey questions. In this study, surveyors attempted to mitigate overreporting by first asking subjects whether they had voted in the past, and prefacing the question with a kindly-worded litany of legitimate reasons why an otherwise respectable individual might fail to vote in an election. Although these methods would theoretically reduce social desirability bias, both failed to reduce overreporting. The same study also explored whether survey respondents overreport at the same rate when they had to identify their polling location. In this case, findings show that overreporting persisted despite the inherent necessity for nonvoters to know their polling place or to lie about it (Presser 1990).

Despite these conclusions, another study found that changes in question wording and context can reduce overreporting. Belli et al. (1999), in a random assignment experiment, specifically targeted both social desir-

ability and memory failure when designing a revised questionnaire. Survey subjects were presented with four options:

(1) I did not vote;

(2) I thought about voting but did not in this election;

(3) I usually vote but did not in this election; and

(4) I am sure I voted in this election.

This modified question lowered overreporting by 13 percentage points because the wording enables the subject to differentiate a current election from past elections (113). It also satisfies the need for social desirability by allowing the subject to report association with voting or past elections, while admitting to nonparticipation in the current election. Thus the literature is inconclusive about the importance of survey wording.

The CPS survey contains language that should address the social desirability bias using the wording, "In any election some people are not able to vote because they are sick or busy or have some other reason, and others do not want to vote. Did (you/name) vote in the election held on (date)?" (United States Census Bureau 2004, 9–1). However, the CPS does not go as far as the study documenting the effects of social desirability bias in attempting to normalize the perceived negative response.

For the years when NES data were available, most researchers examined vote overreporting as a percentage of all voting-age individuals. Conceptualized in this way, some studies indicate that individual characteristics correlate with an individual's propensity to falsely report voting. Race and region are the only demographic characteristics that researchers consistently correlate with the increased propensity to overreport. Specifically, blacks are more likely to overreport voting behavior than whites, even when controlling for education and region of residence. In addition, residing in the South correlates with statistically significantly overreporting (Kanazawa 2005; Abramson and Claggett 1991; Silver et al. 1986).

Another way to conceptualize overreporting is to calculate what proportion of nonvoters falsely claimed that they voted, rather than looking at the percentage of overreporting using the entire population as the denominator. Some researchers argue that this method more accurately measures

the behavior of the population "at risk" for overreporting. This measure shows that blacks are not significantly more likely to overreport voting. Instead, more educated nonvoters are most likely to falsely claim that they voted. Individuals who are more "politically efficacious," or involved with the political process, are more likely to overreport, as are individuals with a stronger sense of civic duty. Finally, people who care more about partisan differences and the election outcome overreport voting more often. This type of research shows that the people who are most strongly motivated to vote in the first place might be the most likely to falsely report voting (Silver et al. 1986).

Voting behavior researchers Bernstein et al. (2001) examined overreporting specifically within the nonvoter population, found similar associations, and proposed an underlying mechanism: the nonvoters who feel the most societal pressure to vote are most likely to experience guilt and shame when they fail to vote. They analyzed NES data through 1988 and found that among nonvoters, overreporting was most common among individuals who were more educated, partisan, and religious, as well as those who were contacted by a mobilization effort. The researchers also found that overreporting increases with a high concentration of minorities in a district, especially among whites in the Deep South, due to a significant interaction among race, region, and minority concentration. This social pressure theory is the most compelling underlying factor proposed in the research.

Data and Methods

The Census Bureau collects official information on actual voter turnout by state, as well as turnout as determined by self- or proxy-report in the CPS Voter Supplement. Proxy respondents, or individuals from the same household that answer for the research subject, reported voting behavior for 42 percent of the sample. Proxies are slightly less likely to overreport, with proxy-reported turnout averaging about four percentage points lower than self-reported turnout (Highton 2005, 115).

This study employs the average turnout and overreporting figures from the voting and registration portion of the CPS in 2000 and 2004. Averag-

ing two elections expands the dataset and gives the turnout estimates more power. In addition, it will mitigate any false trends resulting from important state- or local-level elections that occurred in 2000 or 2004. This study groups the states geographically into the official census regions of North, South, Midwest, and West.

This analysis is based on the Voting Age Population (VAP), since determining voting eligibility is not possible with the CPS data. The Census Bureau defines the VAP as all residents over 18 years of age; the number for this group is used as the denominator when calculating participation rates. Non-citizens, convicted felons, and prison inmates are not allowed to vote, yet they are included in the VAP. This rapidly increasing ineligible population might explain some of the reported decline in voter turnout (Adams 2005). Ineligible individuals, comprising ten percent of the population, are not spread evenly throughout the states. When adjusting the VAP to account for ineligible voters, one finds that the variation in turnout rates between states declines. This may particularly affect the South, since felons in this region are disenfranchised at a high rate, yet still included in the VAP. In addition, the South has a larger share than other regions of non-citizen immigrants who do not possess voting rights (McDonald 2002).

The statistical research begins with a presentation of Analysis of Variance (ANOVA) results that test whether average actual turnout variations by region are statistically significant. The researcher then presents post-hoc pairwise t-tests to show statistically significant differences in pairs of regions. After reviewing regional variations in overreporting, this paper reviews regional variation in actual turnout and examines the rise in turnout from 2000 to 2004 and relates this trend to changes in vote overreporting. This paper uses two different methods to measure overreporting, following trends in the literature: overreporting as a percentage of the voting age population and overreporting as a percentage of nonvoters only.

Overreporting Results: Regional Variation

Vote overreporting varies by geographic region. Specifically, more overreporting occurs in the South. Table 1 shows overreporting by region, with

Table 1:
Vote Overreporting by Region, Out of All Voters, 2000 and 2004

Region	2000 Gap	2004 Gap	Two-Year Average Gap
North	7.0	6.5	6.8
Midwest	6.4	5.3	5.9
West	6.6	7.9	7.2
South	10.5*	7.9	9.2
U.S. Average	*7.9*	*7.1*	*7.5*
ANOVA test statistic	2.71*	1.36	1.65

* Significant at the 90 percent confidence level

Source: Author's calculations based on data from the United States Census Bureau 2007.

overreporting measured as a percentage of all voting-age residents. The national average gap between actual and reported voting is 7.5 percentage points, and the regional values range from 5.9 percentage points in the Midwest to 9.2 percentage points in the South.

The South has the highest overreporting rate in 2000, a significant difference as calculated by the ANOVA test statistic. In 2004 the West shows the same rate as the South, 7.9 percent. Thus, the only statistically significant regional difference is in 2000, not in the 2004 or two-year average values. Ad-hoc t-tests show that overreporting in the South in 2000 is significantly higher than in the other regions.

When measuring overreporting as a percentage of nonvoters only, as shown in Table 2, the values are roughly twice the magnitude of the values constructed with the other measure since the nonvoting population is roughly half of the total voting-age population. The national two-year average gap is 16.5 percentage points, ranging from 14.9 in the Midwest to 18.7 in the South.

Using this measure, the South has the highest overreporting rate in

Table 2:
Vote Overreporting by Region, Out of Nonvoters Only, 2000 and 2004

Region	2000 Gap	2004 Gap	Two-Year Average Gap
North	15.4	15.3	15.4
Midwest	15.7	14.0	14.9
West	13.3	18.1	15.7
South	20.2*	17.2	18.7
U.S. Average	*16.5*	*16.4*	*16.5*
AN OVA est statistic	2.17	0.74	0.78

* Significant at the 90 percent confidence level

Source: Author's calculations based on data from the United States Census Bureau 2007.

2000, but the West has the highest rate in 2004. Although there are no significant regional differences at conventional levels, the test statistic of 2.17 for 2000 narrowly misses (by only 0.06) the 90 percent confidence level threshold for significance. Post-hoc pairwise t-tests again show that overreporting is only statistically significantly higher in the South in the 2000 election.

In sum, the two measures of overreporting show similar results: overreporting is statistically significantly higher in the South, but only in the year 2000. The magnitude of this difference is between five and seven percentage points in 2000. There were no other significant regional differences.

Actual Voter Turnout: Regional Variation

Overreporting, in every area but the West, fell from 2000 to 2004. The increase in voter turnout from 2000 to 2004, which did vary by region, is partially responsible for this change. Table 3 shows that the average turnout

Table 3:
Actual Voter Turnout by Region, 2000 and 2004

Region	2000 Turnout (%)	2004 Turnout (%)	Average Turnout (%)
North	56.2	60.0	58.1
Midwest	56.8*	62.4*	59.6*
West	52.5	57.8	55.2
South	48.3*	54.2	51.3*
U.S. Average	*52.7*	*57.9*	*55.3*
ANOVA test statistic	4.63***	3.60**	4.16**

* Significant at the 90 percent confidence level.

** Significant at the 95 percent confidence level.

*** Significant at the 99 percent confidence level.

Source: Author's calculations based on data from the United States Census Bureau 2007.

by region in 2000 and 2004 is 55.3 percent, ranging from a low of 51.3 percent in the South to a high of 59.6 percent in the Midwest. Regional turnouts are tightly clustered around the national average of 55.3 percent, with a range of eight percentage points.

Analysis of Variance (ANOVA) results indicate that the average turnout variation is statistically significant by region at the 95 percent confidence level. Ad-hoc pairwise *t*-tests on the difference in turnout between regions show that the only statistically significantly differences are in the South and the Midwest (using the 90 percent confidence level). In addition, average turnout is very close to being statistically significantly different in the South versus the North, with the test statistic falling short of the critical value by only 0.08.

In sum, turnout does vary by region, despite fairly tight clustering around the national average of 55.3 percent. The South had the lowest turnout and the Midwest had the highest, with a statistically significant

difference between them. Statistically significant differences at conventional levels do not exist between any other pairs of regions. These results mirror the overreporting trends, and a connection may exist between the two trends.

Increase in Turnout from 2000 to 2004

Researchers examined the increase in voter turnout from 2000 to 2004. One theory hypothesizes that the 2004 national turnout increase of 5.3 percentage points was the result of polarization. Polarization suggests that people believe there are important differences between the political parties—an idea which should fuel participation. Many authors have written about the way the nation, once mostly united behind President George W. Bush immediately after the September 11, 2001 terrorist attacks, has become bitterly divided in the years since. However, the trend of increasing polarization has impacted elections since at least the 1970s, while voter turnout has been declining over this period. While it may seem that polarization implies strength of emotion that should correlate with increased turnout, the two are not necessarily linked (Jacobson 2005).

Nonetheless, civic engagement was very strong in 2004. One in five Americans displayed yard signs, bumper stickers, or buttons versus one in ten Americans in 2000, and almost half of all Americans reported attempts to personally persuade someone else to vote for their favored candidate in 2004 (Abramowitz and Stone 2006, 142). The civic fervor of 2004 was not only fueled by opinions about Bush but also by an increasing rift between the parties. One study by researchers Abramowitz and Stone (2006) found that the effect of both party and ideological identification on evaluations of Bush increased from 2000 to 2004, even when controlling for demographic variables. The magnitude of the effect of political party identification increased by more than 75 percent, while the magnitude of the effect of ideology increased by only 30 percent. In addition, the explanatory power of the regression equation increased dramatically from 2000 to 2004, from accounting for 37.5 percent of the variation in opinions of Bush to explaining 53.7 percent of the variation (146). While the entire increase in explanatory power cannot be attributed to party and ideological variables, their

Table 4:
Actual Voter Turnout Increase by Region, 2000–2004

Region	Turnout Increase (percentage points)
North	3.8
Midwest	5.7
West	5.3
South	5.9
U.S. Average	5.3
ANOVA test statistic	2.22

Source: Author's calculations based on data from the United States Census Bureau 2007.

increasing importance likely accounts for much of this increased power. The hypothesized connection here is that a national focus on the election and on civic involvement should theoretically have encouraged more people to vote, lowering the need for civic-minded citizens to later falsely report voting.

Finally, the magnitude of the turnout increase is tied only loosely to the region, if at all (Table 4). The values cluster tightly around the national mean increase of 5.3 percentage points. Although glancing at the data does not reveal much variation, ANOVA results are very close to significant at the 90 percent confidence level, missing the critical value by only 0.01. Furthermore, post-hoc *t*-tests revealed an interesting result: the turnout increase in the South is more statistically significant than in other regions. This finding makes sense provided that turnout in the South was the lowest to start with, leaving significant room for an increase. In sum, turnout increased by about five percent nationally from 2000 to 2004. While the results are not significant at conventional levels, the turnout increase appears to be slightly larger in the South than in the other regions. However,

the magnitude of the difference is so small that, statistically speaking, the regional difference in turnout increase is all but nonexistent.

Discussion

This study adds to the literature by exploring whether regional differences in overreporting existed in the 2000 and 2004 Current Population Surveys. Using either method of measuring overreporting, the South shows higher overreporting in 2000, but not in 2004. No other significant regional differences exist. This may be linked to increased participation in 2004, but further research is needed to fully understand the forces behind regional variations in overreporting.

The key to understanding why significant regional overreporting variation disappeared by 2004 might be the increase in turnout from 2000 to 2004. Turnout increased nationally by 5.3 percentage points from 2000 to 2004, with the highest increase of 5.9 percentage points in the South. While the increase in actual turnout in the South is not statistically significant at conventional levels, it is close. One hypothesis that attempts to explain this relationship is that politically involved individuals who feel pressure to vote did actually vote in 2004, reducing their perceived need to overreport in this year. However, this hypothesis assumes a limited pool of politically participatory Americans, while the research indicates that more Americans became interested in politics from 2000 to 2004. Future research should explore this phenomenon.

The zenith of the study of individual-level overreporting ended with the final validated National Election Study (NES) survey in 1990, but research continues into the topic with an eye to mitigating overreporting. It is unfortunate that the demise of the validated NES led to a decline in research interest in overreporting. Vote overreporting is unique among survey data; individual-level data on voting behavior is publicly available, so it is possible to investigate individual-level variations in overreporting. The phenomenon of overreporting sheds light not only on the inherent dangers of self-reported survey data and the importance of careful statistical analy-

sis, but also on the motivations underlying voting behavior. Continued research on the topic will inform the discussion on overreporting, in addition to helping researchers design better surveys for social research.

Insight into the reasons that people vote is valuable to inform policy and encourage participation. As with all survey data, self-reported information on vote overreporting is prone to bias. There are several models of voting behavior and data problems, including overreporting, that can bias regression results and limit the interpretations of researchers. One needs to understand the demographic and regional patterns of overreporting in order to account for bias. This problem can be mitigated as much as possible with survey design and analysis that takes the possibility of overreporting into account.

The American electorate is still deeply divided, and turnout is likely to increase for the 2008 election. It will be interesting to see if overreporting remains as low as it was in 2004, or whether it increases to or rises above 2000 levels. The 2008 election will also reveal whether regional differences in overreporting will again manifest themselves, or whether they are disappearing as indicated in 2004. Future research using 2008 data will shed light on this phenomenon and inform researchers and policymakers in their efforts to increase the quality of data analysis and voter turnout itself.

References

Abramowitz, Alan and Walter Stone. 2006. The Bush effect: Polarization, turnout, and activism in the 2004 presidential election. *Presidential Studies Quarterly* 36(2): 141–154.

Abramson, Paul and William Claggett. 1991. Racial differences in self-reported and validated turnout in the 1988 presidential election. *The Journal of Politics* 53(1): 186–197.

Adams, William. 2005. *Election night news and voter turnout*. Boulder, CO: Lynne Rienner Publishers.

Belli, Robert, Michael Traugott, Margaret Young and Katherine McGonagle. 1999. Reducing vote overreporting in surveys: Social desirability, memory failure, and source monitoring. *Public Opinion Quarterly*

63(1): 90–108.
Bernstein, Robert, Anita Chadha and Robert Montjoy. 2001. Overreporting voting: Why it happens and why it matters. *Public Opinion Quarterly* 65(1): 22–44.
Cassel, Carol. 2004. Voting records and validated voting studies. *Public Opinion Quarterly* 68(1): 102–108.
Highton, Benjamin. 2005. Self-reported versus proxy-reported voter turnout in the current population survey. *Public Opinion Quarterly* 69(1): 113–123.
Jacobson, Gary. 2005. Polarized politics and the 2004 congressional and presidential elections. *Political Science Quarterly* 120(2): 199–218.
Kanazawa, Satoshi. 2005. Who lies on surveys, and what can we do about it? *The Journal of Social, Political, and Economic Studies* 30(3): 361–370.
McDonald, Michael. 2002. The turnout rate among eligible voters in the states, 1980–2000. *State Politics & Policy Quarterly* 2(2): 199–212.
Presser, Stanley. 1990. Can changes in context reduce vote overreporting in surveys? *Public Opinion Quarterly* 54(4): 586–593.
Silver, Brian, Barbara Anderson and Paul Abramson. 1986. Who overreports voting? *The American Political Science Review* 80(2): 613–624.
United States Census Bureau. 2007. 2007 Statistical abstract. Tables 406 and 408.
_____. 2004. Current population survey, November 2004: Voting and registration supplemental file. CPS-04. http://www.census.gov/apsd/techdoc/cps/cpsnov04.pdf
Vooght, Robert and Willem Saris. 2003. To participate or not to participate: The link between survey participation, electoral participation, and political interest. *Political Analysis* 11(1): 164–179.

Notes

1. Some researchers find CPS data to be more useful than NES data, regardless of the validation issue, because the CPS sample size is much larger (Abramson and Claggett 1991).

2. The three variables were "number of registration offices," "office workload," and "record quality and access index" (Cassel 2004, 104).
3. The effect of this possible systematic bias in whom refuses the CPS is likely very small, and it would be overcome by the fact that nonvoters might be more likely to refuse to participate in the CPS. This would counteract overreporting and actually create the impression of underreporting in the CPS (Vooght and Saris 2003).

Emily Jones is a Ph.D. candidate in the Trachtenberg School and her degree concentration is in Health Policy. She holds a Bachelor of Arts degree with honors from Brown University and a Masters of Public Policy degree from Georgetown University. Emily's interest in health policy was sparked during her four-year tenure at the Federal Trade Commission. She has also worked at the Urban Institute and the Georgetown Health Policy Institute. Her areas of interest include health disparities, community health centers, Medicaid policy, preventive medicine, and mental health policy.

The author would like to the student editors Jessica McKinney, Danielle Fleury, Christine Brown, and Eva DuGoff for their dedication to this article. The author also thanks Professor Bill Adams for his feedback.