Jumping on the Bandwagon or Jumping Ship?
Testing Alternative Theories of Vote Share Overestimation

David Crow[1]
Division of International Studies
Centro de Investigación y Docencia Económicas
david.crow@cide.edu

Martin Johnson
Department of Political Science
University of California, Riverside
martin.johnson@ucr.edu

Shaun Bowler
Department of Political Science
University of California, Riverside
shaun.bowler@ucr.edu

Abstract

Post-election poll results typically overstate the proportion of people who voted for winning candidates at all levels of government. We test several alternative explanations of this apparent post-election "bandwagon" effect using data from the American National Election Study: conventional ones include expectations that respondents misrepresent how they voted to save face, genuinely forget how they voted, or experience shifts in opinion just before an election. We develop and test an unexplored alternative hypothesis, that post-election surveys inflate the winner's vote because a greater proportion of people who voted for the winning side want to participate in a post-election survey than people who voted for the loser. We devise empirical tests to distinguish and test each of these hypotheses. We find evidence that, rather than misrepresenting their votes to poll-takers, people who voted for the losing side are less likely to participate in post-election surveys.

Introduction

Exit polls and other public opinion surveys conducted after an election often indicate winning candidates have a higher level of support than they did in the election itself. The long-standing interpretation of this exaggerated support for the winner is that survey respondents incorrectly indicate support for the winning side (Atkeson 1999, Wright 1990). On the face of it, this may be a welcome, if puzzling, insight. It is a welcome finding because it helps to underscore the value of elections in providing legitimacy to political leaders (Harrop and Miller; 1987: 244-269). The very fact of winning an election seems to confer on winners several additional points of support among voters.

As comforting as this pattern may be for assessing the value of elections, a puzzle remains of why we see this overestimation. Beginning with Thomsen's 1938 paper, a series of studies seek to explain this post-election bump in support (e.g., Atkeson 1999, Beasley and Joslyn 2001, Belli et al. 1999, Ginsberg and Weissberg 1978, Wright 1990). Grounded in the psychology of voters, this research infers that losers do not simply become reconciled to winners but, in effect, misrepresent or misremember having supported the winner all along. This is the case not simply for primary elections (Atkeson 1999) but for general elections, too. The most prominent explanation for vote share, as well as turnout, overestimation[2] is that survey respondents deliberately dissemble their voting behavior to conform to perceived norms of social desirability. That is, they "overreport," claiming to have voted for the winner when they really voted for someone else (and to have voted when they really did not). In Atkeson's turn of phrase, respondents jump on the "bandwagon" after the election (1999).

We offer a heretofore unexplored alternative explanation for the pattern: non-response bias. We anticipate that people who voted for the losing side in an electoral contest may actually decline to answer post-election surveys proportionally more often than those who voted for the winning side. Rather than jumping on the winner's bandwagon in post-election polls, they "jump ship" and refuse to take part in the political discourse represented by the polling enterprise. In this case, vote share overestimation owes to overrepresentation of electoral winners and under-representation of losers in the sample, rather than to willful overreporting or memory failure. Despite an intense preoccupation with non-response among survey methodologists, scholars of elections have paid surprisingly little attention to non-response bias as a cause of vote share overestimation.[3] We show this non-response bias to hold in both candidate and, in a new area for this literature, ballot proposition elections.

We exploit the panel design of the American National Election Study (ANES), comparing re-interviewed respondents' pre-election vote preferences with those who dropped out of the panel. If non-response bias obtains, re-interview rates should be higher for t_1 respondents who intended to vote for the eventual winner than those who intended to vote for the losing candidate.

Our research has important methodological and substantive implications. Substantively, the frequent overestimation of winning candidates' vote shares may create perceptions that the winners received more votes than they actually did and, thus, inflate victorious candidates' putative mandate following an election (Wright 1990, Atkeson 1999). Methodologically, understanding the mechanism underlying vote overestimation should inform choices for *post hoc* statistical adjustments to survey data, such as appropriate weighting schemes and selection models.[4] We also hope to shed light on potential survey respondents' psychological motivations in choosing whether to answer survey questions truthfully and whether to participate in the first place. We begin by reviewing the puzzle itself, overreport of winner's vote share in post-election surveys.

Exaggerated Support for Winners

Studies that seek to explain why people vote have long noted that post-election polls routinely overestimate the percentage of people who report having voted (e.g., Belli, Traugott, Young and McGonagle 1999). Wolfinger and Rosenstone found that in the American National

Election Study (ANES), taken after every presidential and mid-term election since 1948, the percentage of respondents who report having voted is always between 5% and 20% higher than official turnout figures provided by the Federal Electoral Commission (1980: 115, see also Deufel and Kedar 2000: 24). Validated vote studies comparing self-reported voting behavior on post-electoral surveys to voting records maintained by county registrars, also find large differences between self-reported and actual turnout (Silver, Anderson and Abramson 1986).

Research investigating why people vote as they do also finds that post-election polls overestimate the percentage of people who report having voted for the winning candidate (Wright 1990). Averaging over ANES studies since 1952, Wright calculated that the "pro-winner" bias was 4.0% in U.S. Senate races, 4.7% in gubernatorial contests, and (between 1978 and 1988) 7.0% in races for the U.S. House of Representatives (1993: 295). Eubank and Gao (1984) find even bigger effects, a disparity of 14.0% between the average survey-reported vote share for incumbents in House races and their average share on ballot returns. Atkeson (1999) shows systematic post-election survey vote overestimation for presidential primary winners 1972-1992.

Prevalent explanations of both turnout and winning vote share overestimation largely point to respondent misreports—and more specifically, to overreports—of their own behavior as the culprit. Survey respondents overreport when they inaccurately claim to have voted when they did not and to have voted for the winning candidate when they voted for someone else. Conventional thinking on this problem anticipates that overreporting occurs because respondents misrepresent or misremember whether and for whom they voted.[5]

Respondents may also "forward telescope a remote voting experience," transforming prior votes into a vote cast in the most recent election (Belli et al. 1999: 91)—especially if, after an election, media coverage of the winning candidate emphasizes positive attributes of the winner and his/her opponent's campaign falls silent. The consequent revision of (remembered) history may well be especially prevalent among less sophisticated voters. When such voters cannot reconstruct how they voted accurately, they substitute judgments they make at the time of the survey for those they made at the time of the election, and may claim to have voted for someone other than the candidate for whom they really voted (Wright 1993).[6]

We also explore another alternative, the possibility that some voters have an 11^{th}-hour change in voting intentions. If large numbers of people who had supported a disadvantaged candidate change their minds after the last pre-election interviews and vote for the winning candidate, this could account for at least some of the sharp discrepancies observed between the pre- and post-election polls, and between the pre-election poll and actual voting results. Of course, the longer the pre-election poll is taken before the election, the less accurately it will predict election results (Crespi 1988).

Here, however, we focus on a (surprisingly) over-looked explanation: *non-response bias*. People who abstained or voted for the losing side may disproportionately refuse to take the survey, compared to actual supporters of the winner. Higher non-response among electoral losers and abstainers would result in overrepresentation of voters and winners and, consequently, overestimation of the percentages of citizens who voted, or voted for winning candidates. Citizens who cast ballots for losing candidates, momentarily dispirited, may be disinclined to take a survey about an election whose outcome they find disagreeable. It is possible, in theory, for overestimation to occur even when all respondents report their voting behavior truthfully, although it is likelier that misreporting and non-response bias contribute to overestimation in tandem.

While this is a straightforward argument, in practice it involves quite complex issues. Research on sampling and survey methodology has long grappled with the potential biases produced by non-response (e.g., Berinsky 2004, Cochran 1977, Groves and Couper 1998, Groves et al. 2002, Kish 1965). Practitioners distinguish between item non-response, in which respondents neglect to answer some (but not all) questions on a survey, and unit non-response, in which some people selected into the sample fail to take the survey altogether. We are concerned here with unit non-response, which may bias estimation when the probability of taking a survey is different for different segments of a population and there are significant differences between segments: "For the bias to be important, a large nonresponse must coincide with large differences between the means of ... two segments" (Kish 1965: 535).

Some of the existing literature does mention non-response bias in passing but avoids a more thorough exploration of the possibility. For example, Atkeson shows considerable pro-winner bias in the ANES 1988 Super Tuesday post-election poll. Since African American voters were underrepresented in the sample, the survey results understated support for Jesse Jackson and overstated support for the eventual winner, Michael Dukakis. *Post hoc* weighting adjustments brought vote share estimates in line with actual results, which raises the possibility that non-response bias drove overestimation of Dukakis's vote share—and, conceivably, other results (Atkeson 1999: 207). Similarly, in his study of presidential and congressional races, Wright speculates that "hostile" respondents are not likely to misreport vote choice intentionally, but rather "would generally refuse to be interviewed in the first place" (1993: 293). Neither Atkeson nor Wright, however, develops these asides into a full-blown consideration of non-response bias as a cause of vote preference overestimation.

Distinguishing Explanations of Winner Vote Overestimation

We identify four potential explanations for overestimating the vote share of winners in post-election surveys:

1. **Social desirability**: Respondents recall how they voted but deliberately misreport their electoral preference, embarrassed to admit voting for the losing side.

2. **Non-response bias**: The survey sample overrepresents citizens who voted for the winning side because those who voted for the losing side or abstained are less likely to

participate in a post-election poll.

3. **Memory lapse**: Survey respondents, unable to recall how they voted, misreport their electoral preference.

4. **Late opinion shift**: Large numbers of voters change their minds too late for pre-election polls do not capture the shift, which is registered in the post-election survey.

The two main hypotheses under consideration—socially desirable overreporting and non-response bias—both favor the winner. Neither is directly testable because it is impossible to know who survey respondents really voted for. The ANES "vote validation" studies compared individual, survey-reported voting to country registrar records of actual voting; thus, researchers know which respondents reported having voted accurately and which mis-reported. While we can know who voted, the secret ballot means we cannot know for whom. Detecting evidence of our hypotheses thus necessarily implies drawing inferences indirectly from patterns we observe in the data measured against patterns we would expect to observe under both hypotheses.

We compare re-interviewed respondents' reported pre-election vote preferences with those of pre-election respondents who dropped out of the panel. Non-response bias would imply higher re-interview rates among t_1 respondents who intended to vote for the (ultimately) winning candidate than among those who intended to vote for the (ultimately) losing candidate. On the other hand, more or less equal re-interview rates among the two groups are consistent with social desirability-induced overreporting for the winning side. The difference between survey-reported support for the winner and that actually obtained at the polls is not attributable to non-response bias, but to the "post-election bandwagon effect."

Of course, panel attrition occurs for reasons other than losing-side voters' turning down the follow-up interview, including low interest in politics, belonging to disadvantaged social and ethnic groups, and other factors (Groves and Couper 1998). So, we also model the decision to participate in the follow-up interview as a function of these factors as well as intended voice choice. This controls for potentially confounding variables, yielding cleaner estimates of pre-election vote preferences' effect on t_2 survey response. Both prongs of our research strategy afford evidence of non-response bias rather than social desirability-induced overreporting.

In light of our analytical strategy, these are formal statements of our hypotheses:

Non-Response Bias Hypothesis I: In a cross-classification of predicted by reported vote for both the pre- and post-election cross-sections, classification error produced by predicting a vote for the losing side-given a reported vote for the winning side—will be the same in the pre- and post-election samples:

$$\Pr(PV_{win} = 0|RV_{win} = 1, T = 2) = \Pr(PV_{win} = 0|RV_{win} = 1, T = 1) \ ,$$

where PV_{win} is a predicted vote for the winner, RV_{win} is a reported vote for the winner, and T is the survey period (1 = pre-election, 2 = post-election).

Social Desirability (Overreporting) Hypothesis I: In a cross-classification of predicted by reported vote for both the pre- and post-election cross-sections, classification error produced by predicting a vote for the winning side—given a reported vote for the losing side—will be higher in the post-election than in the pre-election sample:

$$\Pr(PV_{win} = 0|RV_{win} = 1, T = 2) > \Pr(PV_{win} = 0|RV_{win} = 1, T = 1) \ ,$$

where the notation is as before.

Non-Response Bias Hypothesis II: Re-interview rates will be higher for t_1 survey respondents who intended to vote for the (ultimately) winning candidate than for t_1 respondents who intended to vote for the (ultimately) losing candidate, *ceteris paribus*:

$$\Pr(R_{post} = 1|RV_{win,t_1} = 1, x) \ ¿Pr(\mathrm{R}_{post} = 1|RV_{win,t_1} = 0, x) \ ,$$

where R_{post} is survey response in the post-election wave, RV_{win,t_1} is intent to vote for the winner declared in the pre-election wave, and x is a vector of covariates related to panel attrition.

Social Desirability (Overreporting) Hypothesis II: Re-interview rates will be the same, within sample error, for pre-election respondents who intended to vote for the (ultimately) winning candidate as for pre-election respondents who intended to vote for the (ultimately) losing candidate, *ceteris paribus*:

$$\Pr(R_{post} = 1|RV_{win,t_1} = 1, x) = \Pr(R_{post} = 1|RV_{win,t_1} = 0, x),$$

where the notation is as before.

Detecting evidence of the two remaining hypotheses, memory lapse and genuine late shifts in voter preferences, involves investigating the course of electoral preferences over the duration of the survey. If people forget who they voted for and then systematically misremember voting for the winner when they did not, we should observe a trend toward greater self-reported voting for the winning side after Election Day. Similarly, if late shifts of opinion affect survey overestimation of winner support, we should see a trend toward greater support for the winner across days leading up to the election, controlling for other attributes of survey respondents.

Memory Lapse Hypothesis: Support for the winning side of an election will increase over time *after* Election Day, *ceteris paribus*.

$$\Pr(RV_{win} = 1|t > T) > \Pr(RV_{win} = 1|t \leq T)\forall t > 0$$

where RV_{win} is a reported vote for the winner, t is the day on which the survey was taken, T is an arbitrarily fixed reference day, and 0 is Election Day.

Late Opinion Shift Hypothesis: Support for the winning side of an election will increase over time *before* Election Day, *ceteris paribus*. So,

$$\Pr(RV_{win} = 1 | t > T) > \Pr(RV_{win} = 1 | t \leq T) \forall t < 0,$$

where all notation is as before.

U.S. Presidential Elections, 1952-2008

The American National Election Studies have been carried out every U.S. presidential election since 1948. As Table 4 (seventh column) shows, in nine of the 16 presidential contests since 1952[7] the ANES overestimated the vote share of the winning candidate. Overestimation averaged 1.44% points over all 16 elections, and 3.06% taking into account just the elections that overestimated the winners' vote tally-outside the 1.4% margin of error for the smallest ANES sample of 1,212 in 2004. In contrast, underestimation, which averages -1.00%, can likely be chalked up to sampling error.

[Table 4 about here]

Test and results

We test our *Social Desirability I* and *Non-Response Bias I* hypotheses on the nine elections in which winner bias obtained—1952, 1956, 1964, 1968, 1972, 1980, 1992, 1996, and 2008 (see Table 4)—excluding years that underestimate winning vote share as irrelevant to a study on *overestimation*, and because underestimation appears to reflect random fluctuation of samples around the true vote share. The pre-post panel design of the ANES presents a challenge of potential "consistency bias" (respondents' tendency to remember and give the same answers they gave in previous waves) could artificially deflate post-election vote overreporting for the winner, stacking the deck against the overreporting hypothesis. So, we emulated pre- and post-election cross-sections by randomly dividing ANES respondents into two halves, modeling t_1 vote intention on one half and using t_1 model coefficients to predict

t_2 vote choice on the other half. We repeated this process 1,000 times to ensure that our results do not depend on which observations were selected into each sample half.

Developing a model of winning candidate support for all presidential elections since 1952 presents a challenge for at least two reasons. First, virtually no attitudinal or behavioral variables that might explain winning candidate support (such as performance ratings, economic evaluations, etc.) are available for the entire time series.[8] Second, for nearly all variables measured at both t_1 and t_2, the ANES cumulative data file reports only the t_1 measure. Given these limitations, our independent variables are a winning party identification dummy variable equal to 1 if the respondent identifies with the same party as the winning candidate and 0 otherwise; a Republican Party identification dummy variable (1 = Republican, 0 = Other); African American ethnicity; a residual, "Other" ethnicity category; age; sex (1 = Male, 0 = Female); education (four categories treated linearly); and family income (five categories representing percentile ranges). We anticipate these characteristics will be associated with support for Republicans (age, education, and income) or Democrats (minority ethnicity).

We interacted each demographic variable with the Republican Party ID dummy. Thus, the explanatory variables' main effects correspond to non-Republican respondents, and interactive effects—the two component variables' main effects plus the interaction coefficient—to Republicans. Since Republicans won five of the nine contests considered here, we expect Republican identification will increase the likelihood of support for winning candidates, both alone and in combination with the socio-demographic variables. That is, demographic variables' effects on winning candidate support should be positive for Republican respondents and greater than for non-Republicans. Finally, to control for election-specific circumstances we include dummy variables for each election year (with 1952 as the reference category).

Winning party ID accounts for the lion's share of the model's explanatory power, and Republican Party ID (that is, when all other variables are equal to 0) is also strong and highly significant. Other significant predictors are African American ethnicity (positive for Republican African Americans, negative for non-Republicans), age (which had a negative effect for Republican respondents), education (positive effect for non-Republican respondents, negative for Republicans), and all the year dummies except for 1956. Model fit was reasonable: Pseudo-R^2 averaged .54 over the 1,000 half samples (about 8,800 respondents each).[9]

[Figure 3 about here]

Figure 3 depicts their substantive results, superimposing two histograms of, respectively, classification error at t_1 (dark gray bars) and at t_2 (white bars), with the overlapping area in light gray. The t_1 median is 20.2%, with a 95% confidence interval of (17.5%, 22.9%) and the t_2 median, 22.9%, with a 95% confidence interval of (20.5%, 25.7%). It may be that the time-invariant predictors measured at t_1—the only ones available in the ANES cumulative data file—do a worse job predicting votes at t_2 than t_1. Still, overlap between t_1 and t_2 classification errors is large (2.5 percentage points between the t_2 confidence interval lower bound and the t_1 upper bound), and the t_2 median is below the t_1 upper bound, taking into account rounding error.

We also subject the *Social Desirability* and *Non-Response Bias* hypotheses to the second prong of our analytic strategy: comparing re-interview rates. The last three columns of Table 4 report, respectively, the percentage of pre-election supporters of winning party candidates who took the post-election poll, the percentage of losing party candidates who did so, and the difference between the two. Since 1952, post-election survey response rates were, on average, 0.9 points higher for pre-election supporters of winning candidates than for supporters of losing candidates; electoral losers' response probability was actually higher than winners' on five occasions (four of which were Democratic victories). We compare re-interview rates for the same subset of election years used in the Classification Error Comparison Method, minus the 1956 contest. The 100% panel retention rate that year can shed no light on differences in re-interview rates between supporters of winning and losing candidates.

[Table 5 about here]

Table 5 presents the results of three logistic models of post-election survey response on pre-election support for winning candidates. In Model 1, the coefficient for winning candidate support, the only explanatory variable in the model, is ($\beta = .25, p = .000.$). The model-predicted probability that a respondent who intended to vote, in the pre-election survey, for the ultimate winner is .913. It is .891 for respondents who intended to vote for a losing candidate, for a difference of 2.2 percentage points. To ensure that this difference is not spurious or attributable to omitted variables, we progressively control for potentially confounding variables by adding election year dummy variables to Model 2 and the year dummies plus a full complement of explanatory variables (African American ethnicity, other, non-white ethnicity, age, education, income, sex, interest in politics, and party ID) to Model 3. The effect of winning candidate support on post-election survey response (Model 2 $\beta = .25, p = .000$, Model 3 $\beta = .23, p = .000$) is robust to the addition of control variables. Predicted probabilities of survey response are .897 for "losers" and .919 for "winners" (difference of 2.2 points) under Model 2, and .903 and .922, respectively (difference of 1.9 points) under Model 3, with all differences significant at $p < .001$. The results of these tests, then, suggest that non-response bias accounts for 60% to 70% of the 3.06% winning vote share overestimation.

Finally, we examine the *Late Opinion Shift* and *Memory Lapse* hypotheses, regressing pre- and post-election reports of winning candidate support on a day-of-interview counter and the explanatory variables used to estimate classification error. We find no evidence for either. The pre-election counter's coefficient is indistinguishable from 0 (β_{pre} = -.002, p = .170) and, on average, support shifts away from winning candidates after the election (β_{post} = -.005, p = .007).

Discussion

The problem of post-election polls' overestimating winning vote share pervades survey research. Existing explanations for overestimation center on psychological factors that cause survey respondents to misreport their votes. Respondents are seen either remembering their votes inaccurately or, because they wish to present themselves as having engaged in socially sanctioned behavior, deliberately misrepresent their vote. We propose—and find evidence for—an alternative hypothesis: voters for the losing side may not lie about how they voted, but rather choose not to participate in a post-election survey in the first place. Despite a plethora of research on survey non-response and the reasons for it, scholars have not taken it into account in explaining overestimation of winning vote share. We find no evidence (in the elections we consider, at any rate) of two other possible explanations, late opinion shift and memory lapse.

Our findings suggest that survey researchers need to revise our understanding of survey psychology and respondents' motivations for participating in surveys. We know from vote validation turnout studies that survey participants will prevaricate when responding truthfully is embarrassing. Scholars assumed that this explanation also accounted for survey overestimation of winning candidates' vote shares: respondents would say they voted for the winner because they are ashamed to say they voted for the loser. This study, however, raises the possibility that overestimation occurs not because voters find it awkward to admit they voted for the losing side, but because they are simply less interested in taking a survey in the first place.

That survey respondents would lie about having voted more than they would about having voted for the winner stands to reason. Voting is a civic duty, but voting for a winning candidate is not. Greater shame probably attaches not having voted than to having cast a ballot for a candidate who ultimately lost. Thus, incentives to overreport having voted may be stronger than the incentives to overreport voting for winning candidates. This may be part of the reason turnout overestimation rates are generally much higher than winning vote share overestimation rates. Simply put, there is greater reason to lie about having voted than about whom one voted for.

Although we have called non-response bias and social desirability-induced overreporting alternative explanations of winning side vote share overestimation, they are perhaps better seen as complementary. The tests we carry out favor non-response bias over social desirability, suggesting that as much as half of the overestimation detected in our re-interview comparison analysis of the ANES cumulative dataset may be attributable to non-response bias. Devising a means for assessing the relative contributions of social desirability and non-response to overestimation remains an area for further research. We intimate possibilities—including classification error as a percentage of total overestimation—but developing them is outside the scope of this paper.

Our study's main implication for political research is that we need to do a better job of representing losers (and subpopulations that supported losing candidates) in post-election survey samples—especially given the importance of survey research in shaping voters' and political elites' perceptions of the public will. The fault for winner vote overestimation may not lie with deceitful respondents, but with survey recruitment techniques. Rather than unjustly excoriating survey respondents for giving dishonest but socially desirable responses, then, survey researchers turn a critical eye to the problem of making potential respondents who have suffered recent political loss feel more welcome in political discourse.

Notes

[1]Corresponding author. A version of this paper was presented at the 2010 annual meetings of the Midwest Political Science Association, Chicago, IL, and American Political Science Association, Washington, DC. The authors thank Matt Barretto, Peter Esaisson, Jon Krosnick, and Joanne Miller for helpful comments.

[2]We prefer the term vote "overestimation" to the more widely used "overreporting." "Overreporting" implies that respondents supply inaccurate information when taking a survey. But overestimation can also result from overrepresentation of citizens who voted for the winner in the sample. We therefore use "overestimation" in the general case where survey estimates are higher than actual vote share, and "overreporting" in the specific case where overestimation results from respondents' inaccurately claiming to have voted for the winner.

[3]Though this study focuses on vote share overestimation, we refer frequently to studies on turnout overestimation (in which survey-reported turnout rates exceed actual turnout). The psychological underpinnings of overreporting and survey participation are largely the same in both cases.

[4]See Brehm (1999) and Heckman (1979) for sample selection models and Raghunathan (2004) for sample reweighting techniques.

[5]Some studies argue overreporting is partly an artifact of the survey instrument. Eubank and Gow (1983) point to question order effects. The ANES's placement of the question on vote choice after questions that identified incumbents slanted responses in favor of more recognized, incumbent candidates. Jacobson and Rivers (1993) argue that the phrasing of questions on vote choice induces respondents to report voting for winning candidates and incumbents. The switch in 1978 from an open-ended question about whom respondents voted for to one that provided name and party cues boosted reported vote shares for incumbents, likely due to the higher name recognition enjoyed by winners.

[6]Studies have also established that misreporting increases the later the post-election survey is taken after the election. Using validated voting data in an Oregon study, Belli and his coauthors found that an experimental question wording designed to prod respondents' memories increased reliability of self-reported voting data for surveys carried out later in the data collection period. The authors inferred that misreporting increased the more time had elapsed between the election and the survey (1999: 99). Atkeson (1999) noted that memory failure was an especially important explanation for vote misreporting given the large number of days between the primary election and the administration of the ANES, taken after the general election.

[7]We omit the 1948 study, in which the pre-election poll contained only seven questions about the coming election.

[8]In the on-line appendix, we explore the influence of retrospective pocketbook voting using the 1996 as a case study. The results substantively echo the results from the ANES cumulative file.

[9]An example of these statistical models are presented in the on-line appendix.

References

Atkeson, Lonna Rae. 1999. "'Sure I Voted for the Winner!' Overreport of the Primary Vote for the Party Nominee in the National Elections," *Political Behavior*, 21(3):197-215.

Beasley, Ryan K., and Mark R. Joslyn. 2001. "Cognitive Dissonance and Post-Decision Attitude Change in Six Presidential Elections," *Political Psychology* 22(3):521-540.

Belli, Robert F., Michael W. Traugott, Margaret Young, and Katherine A. McGonagle. 1999. "Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring," *Public Opinion Quarterly*, 63(1):90-108.

Berinsky, Adam. 2004. *Silent Voices: Public Opinion and Political Participation in America.* Princeton NJ: Princeton University Press.

Brehm, John. 1993. *The Phantom Respondents. Opinion Surveys and Political Representation.* Ann Arbor: University of Michigan Press.

Brehm, John. 1999. "Alternative Corrections for Sample Truncation: Applications to the 1988, 1990, and 1992 Senate Election Studies," *Political Analysis*, 8(2):183-199.

Cochran, William G. 1977. *Sampling Techniques.* New York: John Wiley Sons.

Crespi, Irving J. 1988. *Pre-election Polling. Sources of Accuracy and Error.* New York: Russell Sage Foundation.

Deufel, Benjamin J., and Orit Kedar. 2000. "But Did They Really Vote? A Method to Correct for Overreporting of Turnout." Presented at the 2000 Annual Meeting of the American Political Science Association, Washington, D.C.

Ginsberg, Benjamin, and Robert Weissberg. 1978. "Elections and the Mobilization of Popular Support.," *American Journal of Political Science*, 22(1):31-55.

Groves, Robert M., and Mick P. Couper. 1998. *Nonresponse in Household Interview Surveys.* New York: John Wiley Sons.

Groves, Robert M., Don A. Dillman, John L. Eltinge, and Roderick J.A. Little, eds. 2002. *Survey Nonresponse*. New York: John Wiley Sons.

Harrop, Martin, and William Miller. 1987. *Elections and Voters: A Comparative Introduction*. New Amsterdam Books.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1):153-161.

Herbst, Susan. 1998. *Reading Public Opinion. How Political Actors View the Democratic Process*. Chicago: University of Chicago Press.

Johnson, Marcia K., Shahin Hashtroudi, and D. Stephen Lindsay. 1993. "Source Monitoring," *Psychological Bulletin*, 114:3-28.

Kish, Leslie. 1965. *Survey Sampling*. New York: John Wiley Sons.

Little, Roderick J.A., and Donald B. Rubin, 2002. *Statistical Analysis with Missing Data*, 2^{nd} ed. New York: John Wiley Sons.

Noelle-Neumann, Elisabeth. 1993. *The Spiral of Silence: Public Opinion-Our Social Skin*. Chicago, IL: University of Chicago Press.

Powers, Daniel, and Yu Xie. 2000. *Statistical Methods for Categorical Data Analysis*. San Diego, CA: Academic Press.

Presser, Stanley. 1990. "Can Context Changes Reduce Vote Overreporting?" *Public Opinion Quarterly*, 54:586-593.

Price, Vincent. 1999. *Public Opinion*. Thousand Oaks: Sage Publications.

Raghunathan, Trivillore E. 2004. "What Do We Do with Missing Data? Some Options for Analysis of Incomplete Data," *Annual Review of Public Health*, 25:99-117.

Silver, Brian D., Barbara A. Anderson, and Paul R. Abramson. 1986. "Who Overreports Voting?" *American Political Science Review*, 80(2):613-624.

Southwell, Priscilla L. 1986. "The Politics of Disgruntlement: Nonvoting and Defection among Supporters of Nomination Losers, 1964-1984," *Political Behavior*, 8:81-95.

Thomsen, Arnold. 1938. "What Voters Think of Candidates Before and After Election," *Public Opinion Quarterly*, 2(2):269-274.

Tomz, Michael, Jason Wittenberg, and Gary King. 2001 CLARIFY: Software for Interpreting and Presenting Statistical Results (Version 2.0). Cambridge, MA: Harvard University.

Wolfinger, Raymond, and Steven J. Rosenstone. 1980. *Who Votes?* New Haven, CT: Yale University Press.

Wright, Gerald C. 1990. "Misreports of Vote Choice in the 1988 NES Senate Election Study," *Legislative Studies Quarterly*, 15(4):543-563.

Table 1: Winning Vote Share, Overestimation, and Participation in Post- Election Polls by Pre-Election Vote Preference, ANES 1952-2008.

Year	*Winner*	*Party*	*Vote%*	*Victory Margin*	*ANES%*	*Overest.*	*%Winners Post-Elec.*	*%Losers Post-Elec.*	*Diff. Win-Lose*
1952	Eisenhower	R	55.2%	10.9%	57.9%	2.74%	90.9	91.0%	-0.11%
1956	Eisenhower	R	57.4%	15.4%	59.5%	2.08%	100.0	100.0%	0.00%
1960	Kennedy	D	49.7%	0.2%	49.3%	-0.38%	92.9	96.1%	-3.25%
1964	Johnson	D	61.1%	22.6%	67.4%	6.34%	94.4	92.6%	1.74%
1968	Nixon	R	43.4%	0.7%	47.6%	4.20%	88.7	86.7%	2.02%
1972	Nixon	R	60.7%	23.2%	63.6%	2.94%	87.0	84.6%	2.46%
1976	Carter	D	50.1%	2.1%	49.7%	-0.38%	84.5	88.2%	-3.76%
1980	Reagan	R	50.8%	9.7%	50.8%	0.07%	90.1	87.0%	3.11%
1984	Reagan	R	58.8%	18.2%	57.7%	-1.10%	90.9	89.4%	1.54%
1988	Bush	R	53.4%	7.7%	52.3%	-1.10%	90.1	88.2%	1.84%
1992	Clinton	D	43.0%	5.6%	47.6%	4.62%	90.8	90.9%	-0.03%
1996	Clinton	D	49.2%	8.5%	52.9%	3.68%	90.7	90.1%	0.55%
2000	Bush	R	47.9%	-0.5%	45.5%	-2.38%	88.7	85.8%	2.99%
2004	Bush	R	50.7%	2.5%	50.1%	-0.67%	91.2	86.7%	4.50%
2008	Obama	D	52.9%	7.3%	53.7%	0.87%	91.8	91.9%	-0.13%

Table 2: Logistic Regression of ANES Cumulative File Post-Election Survey Response Probability on Pre-Election Vote Preference and Other Covariates

	Model 1	Model 2 w/year effects	Model 3 w/year effects
	β	β	β
	(s.e.)	**(s.e.)**	**(s.e.)**
Pre-election vote preference	0.25***	0.26**	0.23***
(1=Winner, 0=Other)	(0.05)	(0.05)	(0.06)
African American			-0.15
(1=Yes, 0=No)			(0.09)
Other non-white			-0.43***
(1=Yes, 0=No)			(0.29)
Age			0.00
			(0.00)
High school or less			0.10
(1=Yes, 0=No)			(0.09)
Some college			0.25*
(1=Yes, 0=No)			(0.11)
College or advanced degree			0.43**
(1=Yes, 0=No)			(0.12)
Income			-0.05
			(0.02)
Gender			-0.09
(1=Male, 0=Female)			(0.05)
Interest in politics			0.16
(1=Not much, 3=Very interested)			(0.04)***
Party identification			0.03*
(1=Strong Dem., 7=Strong Repub.)			(0.01)
Intercept	2.06***	2.83***	2.43***
	(0.03)	(0.89)	(0.17)
N	16,726	16,726	14,358
Model χ^2	23.94***	267.92***	310.22***
$Pseudo\,R^2$	0.002	0.025	0.034

Significance levels: *** $p < .001$ ** $p < .01$ * $p < 05$ † $p < .10$

Coefficients for year fixed effects in models 2 and 3 available from authors.

Figure 1: Histogram of Classification Error Rates for ANES 1952-2008 Simulated Pre- and Post-Election Cross-sections.