

Patterns of Poll Movement*

Public Perspective, forthcoming

Christopher Wlezien is Reader in Comparative Government and Fellow of Nuffield College,
University of Oxford

Robert S. Erikson is a Professor of political science at Columbia University

* We thank Joseph Bafumi, Christopher Carman, Bruce Carroll, Jenny Hill, Joe Howard, Jeff May, Adrian Shepherd, and Amy Ware for assistance with data collection over the years. We also thank Steve Fisher for comments. The larger project has been supported by grants from the National Science Foundation (SBR-9731308 and SBR-0112856) and the Institute for Social and Economic Research at Columbia University.

The growth of pre-election polling is well known. In 1968, there were some 22 published “trial-heat” polls pitting the two major-party candidates against each other. By 1980, the number of presidential polls exceeded 100. In 2000, results for more than 500 polls of the Bush-Gore “vote” are listed on the well-known website, PollingReport.com. What can we learn from this large (and growing) body of polls? What do the polls tell us about voters’ preferences as the campaign evolves? In this essay, we consider these questions. We first describe the sources of poll movement in a very general way. We then offer certain observations based on an historical analysis of presidential polls.

Sources of Poll Movement

Trial-heat preferences from sample surveys represent a combination of true preferences and survey error. Poll results can change due to (1) survey error and (2) real preference change.

There are many manifestations of survey error, the most basic of which is sampling error. All polls contain some degree of sampling error. Thus, even when the division of candidate preferences is constant and unchanging, we will observe changes from poll to poll. This is well known, but not easy to address: We cannot separate sampling error from reported poll preferences.

Survey results also reflect design effects. These effects represent the consequences of the departure in practice from simple random sampling that results from clustering, stratifying, weighting, and the like (Groves, 1989). For election polls, the major source of design effects relates to the polling universe. Determining who will vote on Election Day is not easy: All we

can do is estimate the voting population. Survey organizations typically use “screens” to identify likely voters. Many organizations also weight their samples by selected distributions of party identification or other variables so as to approximate the likely voting electorate. How one screens and weights, of course, has important consequences both for poll margins at each point in time *and* for the variance in poll results over time (Wlezien and Erikson, 2001).

When dealing with polls from different survey organizations, house effects also are a problem. These are the result of survey houses employing different methodologies, including survey design itself. Indeed, much of the difference across houses may reflect underlying differences in design, especially screening and weighting procedures. Results also can differ across houses due to data collection mode, interviewer training, procedures for coping with refusals, and the like (see, e.g., Converse and Traugott, 1986; Lau, 1994). As with design effects, poll results will vary from day to day because the polls reported on different days are conducted by different houses. The effects can be quite pronounced (see Erikson and Wlezien, 1999).

Putting aside survey error, the remaining changes in poll results reflects actual movement in voters’ preferences. This is straightforward. Now, let us see what the polls reveal.

An Examination of Presidential Polls

As part of our ongoing project of estimating campaign effects in modern presidential elections, we attempted to locate all national polls that included the actual Democratic and Republican nominees for the fifteen elections between 1944 and 2000. In the polls, respondents

were asked about how they would vote “if the election were held today” or, less frequently, who they “would like to see win.” The bulk of the data were drawn from Roper Center’s POLL database but other sources also were used, including *The Gallup Report*, *Public Opinion*, and *The Public Perspective*. For 1996, the data were drawn primarily from the now-defunct *PoliticsNow* website, supplemented by data from *The Public Perspective* and the Roper Center. For 2000, the data were taken entirely from PollingReport.com.

Where multiple results, reflecting different sampling universes, were reported for the same polling organizations and dates, we use data for the universe that in theory best approximates the actual voting electorate. For example, where a survey house reports poll results for samples of registered and likely voters, we use the data for the latter. We also removed all overlap in polls, typically tracking polls, conducted by the same survey houses for the same reporting organizations. For example, where a survey house operates a tracking poll and reports 3-day moving averages, we only use poll results for every third day. Following this procedure leaves 1,429 separate national presidential preference polls between 1944 and 2000.

Since most polls are conducted over multiple days, we date each poll by the middle day of the period the survey is in the field.¹ Using this method, the 1,429 polls allow readings (often multiple) for 984 different days from 1944 to 2000. We then generate a daily poll-of-polls for the 15 election years. The numbers represent the Democratic share of the two-party vote intention--ignoring all other candidates--for all respondents aggregated by the mid-date of the

¹ For polls that were in the field for an even number of days, we round up the fractional midpoint, e.g., for a poll in the field four days, we center the poll on the third day.

reported polling period. Wherever possible, respondents who were undecided but leaned toward one of the candidates were included in the tallies. Now, what do these polls reveal?

On Poll Movement

To begin with, we observe that the polls exhibit a lot more volatility in some years than in others. In 1992, for example, the polls shifted dramatically during the last 200 days of the campaign. This can be seen in Figure 1. In 1996, conversely, the polls did not move much at all, as is clear in Figure 2. The 2000 election was somewhere in between (see Figure 3). Put simply, campaign dynamics differ meaningfully across elections. Of course, as we already have discussed, observed poll results combine real movement in preferences and survey error. Although we cannot fully disentangle survey error from reported preferences, we nevertheless can ask: How much of the observed variance is real?

-- Figures 1-3 about here --

We can provide a basic answer to this question by comparing the variance we observe with what we would expect to observe, given the survey N s (or sample sizes) and simple random sampling, if electoral preferences were in fact constant over the campaign. The surplus indicates the true variance of preferences, and it is relatively easy to compute.² Carrying out the calculations indicates that the true variance itself differs significantly from year to year. In 1992,

² For each daily poll-of-polls, the estimated error variance is $p(1-p)/N$, where p = the proportion voting for the Democratic candidate rather than the Republican. Simple calculations give us the estimated error variance for each daily poll-of-polls, e.g., where preferences are divided 50-50 in a sample of 1000, the estimated error variance is 0.00025 or 2.5 when expressed in terms of percentage points ($50*50/1000$). The error variance for all polls-of-polls in a particular election cycle is simply the mean error variance over the campaign. The estimated true variance is the arithmetic difference between the variance we observe from the poll readings and the estimated error variance.

sampling error accounts for only 5 percent of the variation in the polls during the last 200 days. In 1996 and 2000, the same amount of sampling error accounted for more than 30 percent of the observed variation. In some years, up to half of the variance is sampling error. This nevertheless tells us that most of the variance in poll results over the long campaign is “real.” On average, over the 15 presidential elections 1944-2000, the ratio of true variance to error variance is about four to one. We know this because the variance we observe is about four times what we would expect from sampling error alone, that is, if survey respondents made choices by flipping coins.

Most of the variation in preferences is concentrated in the period leading up to the fall general election campaign, however. After Labor Day, the unofficial kick-off of the general election campaign, more than half of the poll variance is due to sampling error alone. Less than half is real. And keep in mind that this estimate is based only on allowing for sampling error without also taking into account house and design effects. Although relatively small, the late movement in preferences matters quite a lot, as we will see.

Patterns of Poll Movement

Now, let us shift our focus, and consider the temporal pattern to poll dynamics across the 15 elections. Figure 4 shows selected reading of the poll of polls over 25-day intervals for each of the 15 elections. To provide comparability across years, we subtracted out the actual Democrat share of the two-party vote; thus, the numbers reflect the degree to which poll results differ from the final vote. For each of the many days without polls centered on those dates, we

interpolate from the most recent date with polls and the next date with polls. For days after the last poll in particular election years, we assume the numbers from the last pre-election poll.

The data in Figure 4 reveal that election outcomes come into focus as the election cycle evolves. At the beginning of our timeline, 200 days before the election, reported poll results and the final vote differ quite a lot, by about 6.5 points on average. The differences remain remarkably stable over the next 100 days, and then diminish dramatically thereafter. At the 100-day mark, about the time of the national party conventions, the average difference between the polls and the ultimate vote is 5.6 points. By the very end of the campaign, the average difference is a mere 2.2 points.³ The polls clearly tell us more and more about the outcome as the campaign unfolds. This is not especially surprising. What may be surprising is that much of the improvement in predictability occurs during the general election campaign after Labor Day. Indeed, the relatively small real movement in preferences during the fall appears to matter most.

– Figures 4-5 about here –

Figure 5 displays another interesting pattern in the polls over time. It shows the poll lead for the ultimate winner in each election over the last 200 days of the campaign. In the figure, we can see that the early polls tell us something about the Election Day result. At 200 days out, the winner has the poll lead in 10 of the 15 elections. Important sorting continues through the summer. By 100 days out, the winner has the lead in 12 of the elections; by Labor Day, this is

³ Even this estimate is inflated because the final pre-election polls often end well in advance of the election, and we simply carry forward the results.

true in all but one year, 1948.⁴

Once in the lead, however, the winner's margin tends to shrink. Leads from the Labor Day period, for instance, eventually are halved by Election Day. Even the final pre-election polls distort the lead, though this partly reflects the lack of late polls in a number of years, as noted above. These results imply a persistent underdog effect, where the projected loser gains support as the campaign persists. With such an effect, the drift of the polls is as if one candidate emerges as the favorite after the conventions and then watches the lead shrink.⁵ But, why does this happen? Why does the underdog typically gain even though the leader—and ultimate winner in virtually every case--evidently is the stronger candidate?

The Puzzle of Poll Movement

The basic patterns of poll movement constitute a puzzle of sorts. Why do results converge both toward the final vote and toward 50-50 as the campaign unfolds? That the polls increase in accuracy leading up to Election Day indicates that something happens to change voter preferences and in meaningful ways. Indeed, it appears that election campaigns really do matter.⁶ We start with the knowledge that because the electorate's preferences do change, campaign events (broadly defined) must be exerting some sort of impact. The question then is whether these shocks from campaign events take the form of temporary "bounces" or permanent "bumps." Simply put, do the effects decay or else last? If campaign effects are

⁴ As Gore won the popular vote in the 2000 election, he is considered the winner for the purposes of our analysis.

⁵ A similar pattern also is evident in Congressional polls (Erikson and Sigelman, 1995; 1996).

⁶ For a more detailed consideration of this issue, see our article on "The Timeline of Election Campaigns" in the

bounces, they dissipate over time. Preferences tend to revert to an “equilibrium” that is set early in each particular election year, say, about the time of the conventions. The final outcome is the simple sum of this equilibrium plus the effects of very late events that do not fully dissipate before Election Day. If campaign effects are bumps, conversely, they last to affect the outcome. In effect, the equilibrium drifts over time. The election outcome is the sum of all the bumps—often small in size—that happen during the campaign.

The answer may be that campaign events produce both bounces and bumps. It may be that some effects dissipate and others last. It may be that effects combine both bounces and bumps. Statistically, it is the bumps and not the bounces that matter in the long run. They cumulate over time.

We see the evidence of permanent bumps in the fact that the polls are increasingly accurate over the fall, general election campaign. If this were not so—if the effects of events dissipated—the accuracy of polls would vary little except at the very end of the campaign, reflecting the effects of late events. Something clearly happens during the fall to change voters’ preferences. Beforehand, at least prior to the conventions, we see a very different pattern. The polls during this period do not vary much in their accuracy: Indeed, it is as if they bounce around an equilibrium that is constant for the particular election. The conventions have important, often realigning effects. The fall campaign then generates change as the accumulation of seemingly small bumps for one candidate or the other. We do not know what exactly causes preferences to change during this period. We also cannot predict it in advance.

current issue of the *Journal of Politics* (Wlezien and Erikson, 2002).

Now, what explains the shrinking margins? Recall that trial heat surveys show considerable movement early in the campaign, often with one candidate surging to a large lead. As the campaign progresses, the electorate's net vote intention hardens, typically moving in the direction of a tightening outcome. This result is as if people make tentative choices early, based on the political news of the moment. In other words, early in the campaign, survey respondents act in a relatively nonpartisan or independent manner. Responding to the prevailing news about the candidates, the early-campaign electorate drifts toward the early favorite, much in the manner of "independent" voters generally.

As the campaign evolves, much of these effects dissipate and preferences polarize, with a widening attitudinal gulf between the supporters of the two major-party candidates. It may be that the campaign activates voters' predispositions, causing them to gravitate toward their partisan "equilibrium" or some broader underlying preference, or it just may be that individuals react differently to the events of the campaign. Regardless of its particular underpinnings, the polarization of underlying preferences over the campaign will produce a predictable decline in poll margins (see Wlezien and Erikson, 2001). This, of course, is exactly what we observe.

On the Importance of Election Campaigns

Shifting poll results often represent chance variation due to survey error. Beneath the surface we nevertheless can see that the electorate's preferences change over the course of a campaign. Early on, the likely winner holds a large initial lead. As the campaign unfolds, the race typically tightens and becomes more stable, as preferences harden. The polls also increase

in accuracy leading up to Election Day. These patterns tell us that election campaigns do matter and that the general election campaign matters most of all. They do not tell us *how* campaigns actually matter. What events had effects? Which ones lasted and which ones decayed? We simply do not know. This mystery remains unsolved.

References

- Campbell, James E. 2000. *The American Campaign: U.S. Presidential Campaigns and the National Vote*. College Station, Texas: Texas A&M University Press.
- Campbell, James E. and James C. Garand. 2000. *Before the Vote: Forecasting American National Elections*. Thousand Oaks, Calif.: Sage Publications.
- Converse, Phillip E. and Michael W. Traugott. 1986. "Assessing the Accuracy of Polls and Surveys." *Science* 234(Nov. 28):1094-1098.
- Erikson, Robert S. and Christopher Wlezien. 1999. "Presidential Polls as a Time Series: The Case of 1996." *Public Opinion Quarterly* 63(Summer):163-177.
- Erikson, Robert S., and Lee Sigelman. 1996. 'Poll-Based Forecasts of the House Vote in Presidential Election Years.' *American Politics Quarterly* 24:52-531.
- , 1995. 'Poll-Based Forecasts of Midterm Congressional Elections: Do the Pollsters Get it Right?' *Public Opinion Quarterly* 59:589-605.
- Groves, Robert M. 1989. *Survey Errors and Survey Costs*. New York: Wiley.
- Lau, Richard R. 1994. "An Analysis of the Accuracy of 'Trial Heat' Polls During the 1992 Presidential Election." *Public Opinion Quarterly* 58(Winter):2-20.
- Wlezien, Christopher and Robert S. Erikson. 2002. "The Timeline of Presidential Election Campaigns." *Journal of Politics* 64:969-993.
- , 2001. "Campaign Effects in Theory and Practice." *American Politics Research* 29(September):419-437.

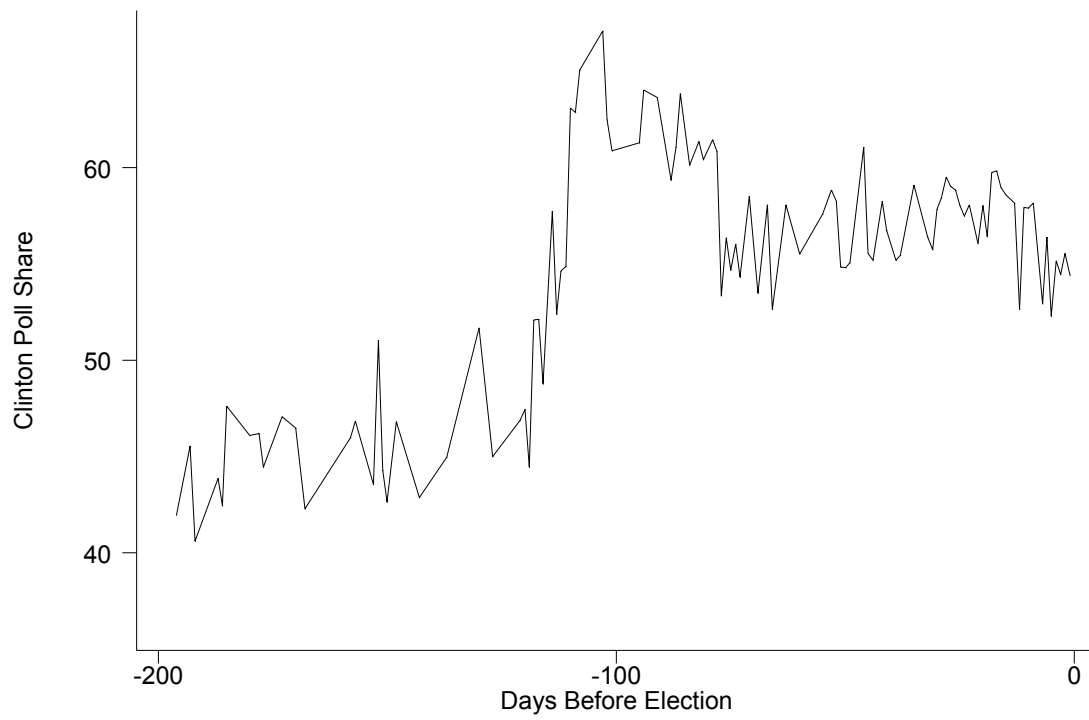


Fig. 1. Clinton Poll Share over the Last 200 Days of the 1992 Campaign

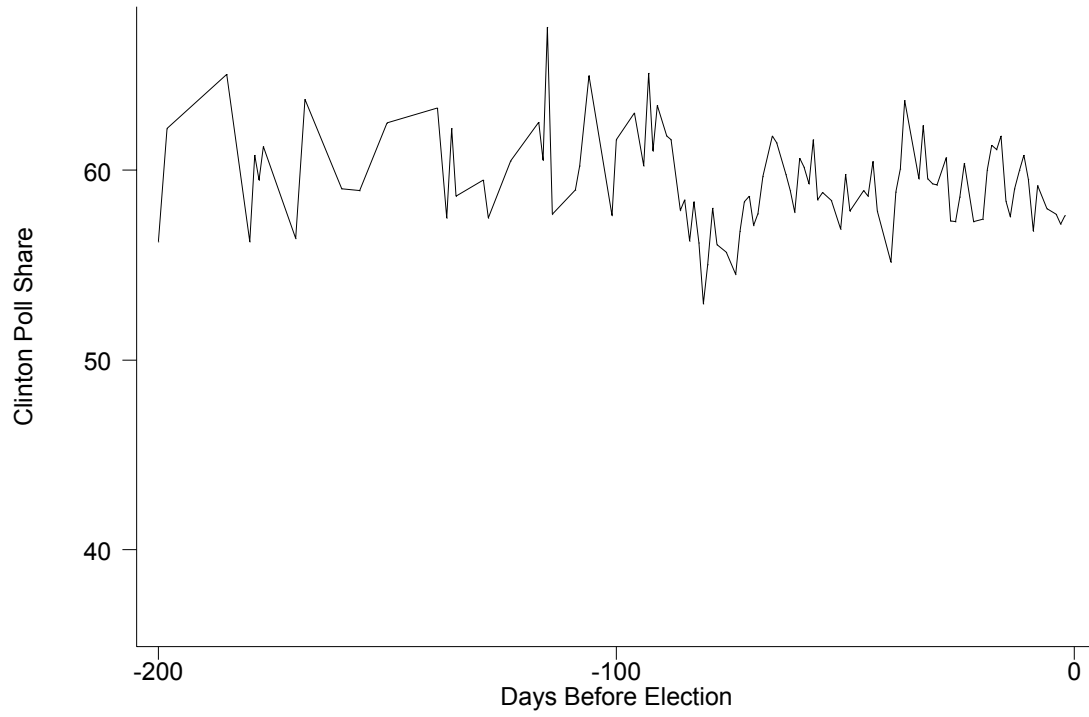


Fig. 2. Clinton Poll Share over the Last 200 Days of the 1996 Campaign

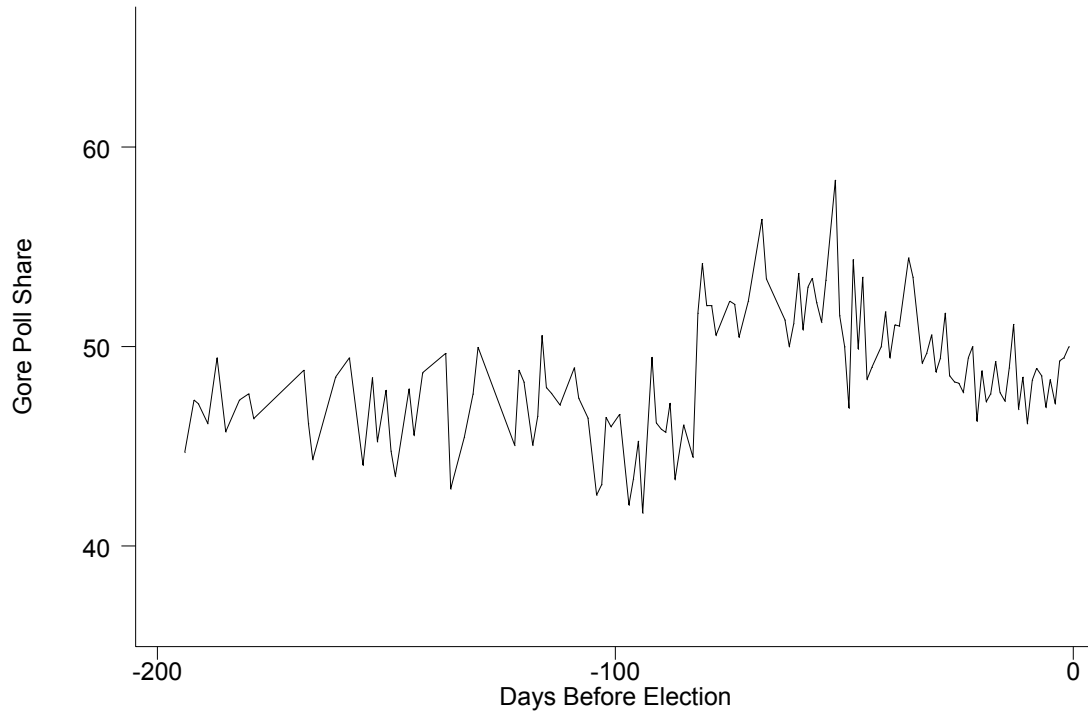


Fig. 3. Gore Poll Share over the Last 200 Days of the 2000 Campaign

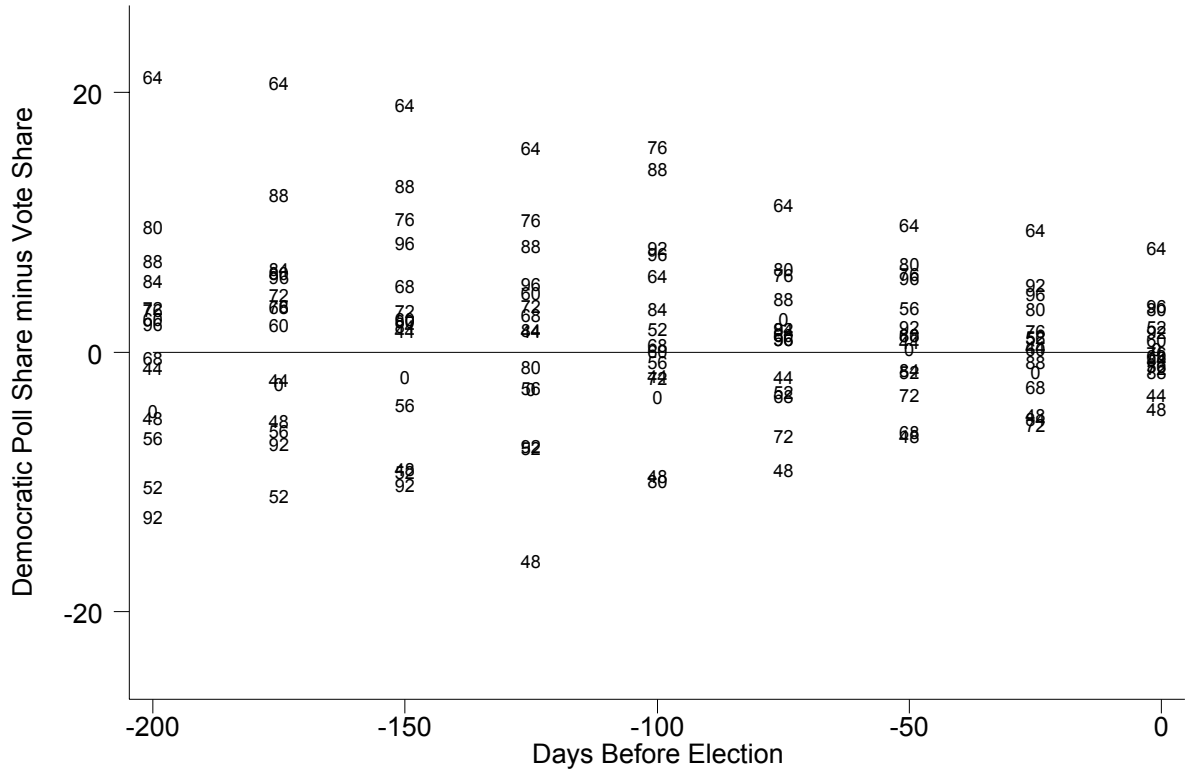


Fig. 4. Democratic Poll Share minus the Actual Vote Share for Selected Days of the Election Cycle, 1944-2000

