Dynamics of (National) Electoral Preferences during the 2016 US Presidential Race*

Christopher Wlezien
University of Texas at Austin
Wlezien@austin.texas.edu

Stuart Soroka
University of Michigan
ssoroka@umich.edu

George Elliott Morris
University of Texas at Austin
g.e.morris@utexas.edu

*Prepared for presentation at the Conference on The US Elections of 2016: Domestic and International Aspects, the Interdisciplinary Center (IDC), Herzliya, Israel, January 8-9, 2017. This paper builds on Wlezien's previous research with Robert S. Erikson and relies on Lexicoder software development by Mark Daku and Lori Young. We owe special thanks to Amnon Cavari for getting us to do the data work and analysis for this paper and also for comments along the way.

The study of voters and elections has taught us a lot about individuals' vote choices and election outcomes themselves. We know that voters behave in fairly understandable ways on Election Day (see, e.g., Alvarez, 1997; Campbell, 2000; Gelman and King, 1993, Johnston, et al., 1992; Lewis-Beck, 1988; Vavreck, 2009; Erikson and Wlezien, 2012). We also know that the actual outcomes are fairly predictable, including the 2016 popular vote (see, e.g., Campbell, 2017). Of course, what we do know is imperfect, both about the popular vote and the decisive electoral college result. The 2016 election outcome has made this especially clear.

Research also has taught us quite a lot about how preferences come into focus over the course of campaigns (Campbell, 2000; Wlezien and Erikson, 2002; Erikson and Wlezien, 2012; Wlezien et al, 2013; Jennings and Wlezien, 2016). Much of this work focuses on the relationship between election results for a set of years and trial-heat poll readings at varying points in time during the election cycle.[1] What it shows is that the predictability of outcomes increases in proportion to the closeness of the polling date to Election Day: The closer we are to the end of the race, the more the polls capture the Election Day "fundamentals," and tell us about the ultimate outcome.[2] Although this may not surprise, it is important.

That said, this work tells us little about dynamics in particular years. This is not surprising given that we lack anything approaching a daily time series of voter preferences until the most recent elections. Erikson and Wlezien (1999) provided an initial foray into poll dynamics during the 1996 US presidential election. Follow-up research examined the 2000 election and incorporated time-series analysis techniques (Wlezien, 2003). This paper takes another step, focusing on the 2016 election. It attempts to answer two primary questions. First, to what extent does the observable variation in poll results reflect real change in electoral preferences as opposed to survey

[1] There is some research on Congressional elections as well (Erikson and Sigelman, 1995; 1996).
[2] The fundamentals come in two varieties, one "internal" to voters and the other "external" (Erikson and Wlezien, 2012).

error? Second, to the extent that poll results reflect real change in preferences, does this change in preferences actually last, or decay? It also begins to answer one additional question: What role did mass media coverage play? Answers to these questions tell us a lot about the evolution of electoral sentiment during the 2016 presidential race, and also about the effects of the election campaign itself.

The Polls

For the 2016 election year itself, the Huffington Post website contains national polls of the Clinton-Trump (-Johnson/-Stein) division reported by different survey organizations. In each of the polls, respondents were asked about how they would vote "if the election were held today" (albeit with slight differences in question wording). Where multiple results for different universes were reported for the same polling organizations and dates, data for the universe that best approximates the actual voting electorate is used, e.g., a sample of likely voters over a sample of registered voters. Most importantly, all overlap in the polls--typically tracking polls--conducted by the same survey houses for the same reporting organizations is removed. 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. This leaves 307 separate national polls during the election year, 100 after the unofficial Labor Day kickoff of the general election campaign. Wherever possible, respondents who were undecided but leaned toward one of the candidates were included in the tallies.

Figure 1 displays results for the complete set of 307 polls. Specifically, it shows Clinton's percentage share of the two-party vote (ignoring Johnson and Stein and others) for each poll. (To see the candidates' shares of the total vote in the polls, including other candidates, see Figure S1.) Since most polls are conducted over multiple days, it is necessary to date the polls. To begin with, we date by the middle day of the period the survey is in the field.[3] Doing this, the 307 polls allow readings for 188 separate days during 2016, 50 of which are after Labor Day, permitting a virtual day-to-day monitoring of preferences during the general election campaign, particularly for the final six weeks. It is important to note that polls on successive days are not truly independent, however. Although they do not share respondents, they do share overlapping *polling periods*. Thus, polls on neighboring days will capture a lot of the same things by definition, which is of consequence for our analysis of dynamics.

-- Figures 1 and 2 about here --

The data in Figure 1 indicate some patterned movement in the polls over time. For any given date, the poll results nevertheless differ quite considerably. Some of the noise is mere sampling error; but there are other sources of survey error, of course. The daily poll-of-polls in Figure 2 reveals more distinct pattern. The observations in the figure represent Clinton's share for all respondents aggregated by the mid-date of the reported polling period. (To see Clinton and Trump's shares of all respondents, including those supporting Johnson, Stein and others, see Figure S2.) We see clearly that Clinton retained a lead in the raw national polls on virtually every day during the election year – all but ten days, four at the very beginning of the year, four in the middle of summer prior to and during the Republican convention, and then one day in the Spring and one other during the fall. The average margin was 53.0-47.0. We also see that her lead varied over time, increasing through the Spring before declining into the summer, after which poll shares shifted fairly predictably with political events during the summer months, out of which Clinton emerged with a slight lead heading into the autumn. Thereafter polls shifted in a narrow range and ended up (at 51.8%) just slightly above the two-party vote share of 51.1%.

[3] For surveys in the field for an even number of days, the fractional midpoint is rounded up to the following day. There is a good amount of variation in the number of days surveys are in the field: The mean number of days is 4.49; the standard deviation is 1.80 days.

Survey Error and the Polls

Trial-heat poll results represent a combination of true preferences and survey error. Survey error comes in many forms, the most basic of which is sampling error. All polls contain some degree of sampling error. Thus, even when the division of candidate preferences does not change, we will observe changes from poll to poll. This is well known. All survey results also contain design effects, the consequences of the departure in practice from simple random sampling that results from clustering, stratifying, and the like (Groves, 1989). When studying election polls, the main source of design effects relates to the polling universe. It is not easy to determine who will vote on Election Day: When we draw our samples, all we can do is estimate the voting population. Survey organizations typically rely on likely voter screens. In addition, most organizations use some sort of weighting procedure, e.g., weighting by a selected distribution of party identification or some other variable that tend to predict the Election Day vote. How organizations screen and weight has important consequences both for the cross-sectional poll margins at each point in time and for the variance in the polls over time (Wlezien and Erikson, 2001).

Given that we are combining polls from various survey organizations, house effects are another source of error. Different houses employ different methodologies and these can affect poll results. Much of the observed difference in results across survey houses may reflect differences in screening and weighting practices noted above. 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; also see Crespi, 1988). As with design effects, poll results will vary from day to day because the polls reported on different days are conducted by different houses.

Now, we cannot perfectly correct for survey error. We cannot eliminate sampling error. We also cannot perfectly correct for design and house effects, as we have only limited information about what survey organizations actually do. We can to some extent account for these effects, however. That is, we can control for the polling universe—at least broadly defined—as well as the different survey houses. The data include poll results from 39 different survey organizations. These organizations sampled three different polling universes during the year, specifically, adults, registered voters, and likely voters. (As noted above, these universes do not necessarily mean the same things to different organizations, particularly as relates to "likely" voters.) In order to adjust for possible design and house effects, the results for all 307 polls are regressed on a set survey house dummy variables and two polling universe dummy variables. Specifically, we include dummy variables for each house with five or more polls in the field during the year.[4] Dummy variables also were included for each of the 188 dates with at least one survey observation. With no intercept, the coefficients for the survey dates constitute estimates of public preferences over the course of the campaign.

-- Tables 1 and 2 about here --

The results of the analysis are shown in Table 1. Here, we can see that the general polling universe did not meaningfully affect poll results during 2016, controlling for survey house and date. This is not entirely surprising, as the same was true of polls in previous election years (see Erikson and Wlezien, 1999; Wlezien, 2003). Table 1 also shows that survey house did matter in 2016. Table 2 summarizes selected effects relative to the median house, which was Fox. For the full election year, the range of the house effects is just less than 4.0 percentage points.[5] These are big

[4] Restricting to houses with fewer polls makes little statistical difference to the resulting portrait of preferences or the analyses that follow.
[5] This may appear to be less than we might expect given the variation in results from different firms and previous elections (see, e.g., Wlezien, 2003), which partly reflects the focus on firms with five or more polls in the field during the year.

differences and ones that are difficult to fully explain given the available information about the practices of different survey houses. There is reason to suppose that the observed differences across houses largely reflect underlying differences in design (Wlezien and Erikson, 2001), though we cannot be sure. Whatever the source, the differences have consequences for our portrait of preferences during 2016, as poll results will differ from day-to-day merely because different houses report on different days.

-- Figures 3-4 about here --

Also notice in Table 1 that survey date effects easily meet conventional levels of statistical significance (p <.001) during the full election year but are less apparent (p = .09) during the post-Labor Day period. The results imply that underlying electoral preferences changed mostly over the course of the *long campaign*. Figure 3 displays the survey date estimates from the first regression in Table 1. Because of substantial house effects, it was necessary to re-center the estimates, and the median house (Fox) from the analysis of variance was used.[6] These readings exhibit less variance (3.9 versus 4.4 percentage points) than the unadjusted polls in Figure 2 and slightly more pattern. There still is some evidence of noise, partly the result of sampling error. There also is reason to think that the series contains design and house effects that are not easily captured statistically, that is, because they vary over time. The problem is that we do not know. We thus must accept that our series of polls, even adjusted for systematic house and design effects, is imperfect. These data still can tell us quite a lot, as we will see.

When including all polling organizations regardless of the number of polls they conducted, house effects are quite a bit larger, approaching 6.8 points.

[6] This seems reasonable, although it may not be quite right; the problem is that we cannot tell for sure. It may be tempting to use the house that best predicted the final outcome, though this is even more tenuous. Given that all polls, getting it right at the end is as much the result of good luck as it is good survey design.

An Analysis of Poll Variance

As noted, our house-adjusted poll estimates also contain random sampling error. We cannot simply separate this sampling error from reported preferences, however, as we cannot determine how much error variance there is in every poll. We nevertheless can ask: What portion of the total variance is real? Assuming random sampling or its equivalent by pollsters, the answer is relatively easy to compute using the frequencies and sample sizes of the actual polls (Heise, 1969). That is, we can determine the observed variance of poll results where underlying preferences do not change.[7]

For each daily poll-of-polls, the estimated error variance is $\frac{p(1-p)}{N}$, where p = the proportion who prefer for the Democratic candidate rather than the Republican and N = the number of respondents offering preferences for either candidate. For the value of p, we simply insert the observed Clinton proportion of major-party preferences in the daily poll reading. For N, we take the number of respondents offering Clinton or Trump preferences on that day. Simple calculations give the estimated error variance for each daily poll-of-polls. For example, 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 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

[7] The assumption of simple random sampling seems a fairly safe one historically, though less so in recent election years given the increasing use of weighting procedures. To the extent the weighting variables are exogenous to the campaign and predict the vote on Election Day, weighting will reduce observed sampling error. By implication, an analysis of variance based on the assumption of simple random sampling will tend to overstate the actual error variance due to sampling. That said, it will understate the total survey, which has various sources, as noted in the text. Indeed, weighting itself introduces error.

error variance. The ratio between the estimated true variance and the observed variance is the statistical *reliability*.

The results of conducting such an analysis using our series of adjusted readings for the 2016 presidential election are shown in Table 3a. The first row of the table contains results for the full election year. Specifically, it shows the average daily variance of our adjusted series of polls along with the estimated error variance, the resulting “true” variance, and the corresponding reliability statistic. Of greatest importance is that most (71 percent) of the variance we observe over the election year is real, not the mere result of sampling error. The estimated real variance in preferences over the election year is less than 3 percentage points. The standard deviation is roughly 1.7 points, i.e., the square root of 2.77, which implies an average daily confidence interval of plus-or-minus 3.3 points around the observed percent for Clinton. The estimated range of real movement thus was less than 7 points. The number is lower still over the last 200 days of the campaign, by which point in time Clinton and Trump had emerged as the leading party nominees. During this period, the range of real preferences was less than 5.5 percentage points.

-- Table 3 about here –

In the final row of the Table 3a, we can see that the estimates for the post-Labor Day period are smaller still, as the observed poll variance during the period is less than two percentage points. Adjusting for sampling error leaves less the estimated true variance at less than 1.0 percentage point, which implies a range of real movement of about 4 percentage points during the autumn. The number is small both in absolute terms and also by historical standards. As is clear in Table 3b, over the 16 presidential elections between 1952 and 2012, the mean estimated true variance during the last 60 days of the cycle is only 2.30 percentage points and the median is 1.65 points.[8] (For

[8] These historical averages exaggerate the difference with 2016 since they are not based on house-adjusted polls. Still, it makes little difference, as the house adjustment only trivially reduces the observed variance in the polls, from 1.99 to 1.88 percentage points.

specifics, see Erikson and Wlezien, 2012; 2014.) This implies an average range of movement of about double what we saw in 2016. Based on our analysis, then, it appears that campaign events had fairly modest net effects on preferences in 2016, at least much smaller than in most previous presidential elections.

This is all the more striking considering that our presentation may overstate the true change in preferences during the 2016 campaign. Recall that our presentation thus far has aggregated polls by the middle date of the polling period. We alternatively could pool all polls on each day they are in the field, weighting each by the inverse of the number of days in the field, i.e., weighting results from 3-day polls by one-third, results from 4-day polls by one-fourth, and so on. The results of pooling house-adjusted results in this way are shown in Figure 4. Figure 5 zooms in on the post-Labor Day period. Here the observed variance is less than 1.0 percentage point (0.97). We can see that Clinton began the fall with a slight lead, about what she polled at the end of the campaign and received on Election Day.

-- Figures 4-5 about here --

The polls still do indicate real, if modest, change in voter preferences during the 2016 campaign. What exactly happened to produce the evident ebb and flow in electoral support? It is hard to tell for sure, as we know that campaigns represent a series of many events, many of which are difficult to identify conceptually or empirically. Indeed, when studying campaign effects, political scientists typically focus on the effects of very prominent events, such as nominating conventions and general election debates in the US (see, e.g., Holbrook, 1996; Shaw, 1999).[9] The

[9] This focus is understandable for a number of reasons (see Wlezien and Erikson, 2001). First, we know that conventions and debates very visible, where large numbers of people watch on televisions and/or acquire information about them in other ways. Second, we can anticipate these events, so our interpretation of their effects is not subject to the post hoc, ergo propter hoc reduction that characterizes interpretations of the seeming effects of many other campaign events. Third, there already is evidence that they matter a lot more than other events, or at least that they can.

previous research indicates that conventions have big effects that mostly cancel out but can alter the state of the race. Debates have smaller, less consequential effects.

One way to assess the effects of conventions and debates, is to compare the poll shares before and after the two seasons. One can compare results of polls before the first convention with results of polls taken after the second convention. Likewise, one can compare results from before the first debate with those from after the last debate. This is what Erikson and Wlezien (2012; 2014) do in their analysis of election campaigns between 1952 and 2012. We update their analysis to include 2016 and plot the results in Figures 6-7.

In Figure 6, we can see that poll results one week before and then two weeks after the conventions are correlated but that the two usually differ. Most importantly, we can see that the leader in the polls often changes as a result of the conventions, specifically, in five of the 17 years: 1968, 1988, 1992, 2000, and 2004.[10] The leaders in these years – as in all other years in the figure – went on to win the popular vote. The conventions did not change the leader in 2016 but did have a sizable net effect, as Clinton's share increased from 51.1% percent before the Republican convention to 53.7% after the Democratic convention. The effect may have been exaggerated some by her sharp drown in the polls a week before the convention season began, in the wake of FBI Director Comey's July 5 announcement regarding Clinton's e-mails. Also note that the lead that she emerged with did not last very long, as Clinton entered the fall campaign with 51.6% in the polls, just about where she ended the campaign.

As can be seen in Figure 7, with the sole of exception of 1976, debates have smaller net effects. Perhaps most importantly, they are less consequential, as the lead has changed hands after the debates only one time out of the twelve presidential election campaigns during which debates

[10] In 1980, Carter moved into a tie after the conventions, only to watch Reagan recapture the lead in the fall and increase his advantage by Election Day.

have been held, namely, in 2000, when George W. Bush took the lead from Gore only to lose it and the actual popular vote. Of course, we cannot be absolutely sure that the debates were the reason(s) that Bush gained a temporary advantage in 2000, but we do know that it was the only year in which the lead did change during the debate season. Debates still may have had effects in other years, including 2016, when Clinton's share increased from 51.9% to 52.8%, seemingly enough to win 270 electoral college votes and then some. As we have seen, the lead did not last, costing her the White House.

-- Figures 6-7 about here --

While conventions and debates are – or can be – important, they account for only a modest portion of the variance in poll results. Even attributing all poll movement during the *entire* convention and debate seasons leaves 70-75 percent of the variance unaccounted for over the full election year (Wlezien and Erikson, 2001; Wlezien, 2003). These results tell us that the numerous other events, mostly small, when taken together, have a greater impact than the handful of very visible events that occupy most media and scholarly attention. The problem, of course, is that it is difficult to identify these "other" events and even harder to detect their effects. Some may be easier than others, however, such as Comey's two intrusions into the 2016 campaign.

An Analysis of Dynamics

Thus far, we have seen that electoral preferences changed, if modestly, during the course of the 2016 presidential election campaign, though we are not sure about what exactly caused the observable change.[11] We also want to know whether the evident effects really mattered, that is, the

[11] These analyses—and any others that rely on poll aggregates—tend to understate campaign effects. First, they register the net effect of many different campaign activities, which can cancel out on a daily basis. For instance, one candidate's campaign may shift preferences by 0.5 percent on a particular day but the other campaign may shift preferences by 0.3 in the opposite direction. Thus, while the total effect of the two campaigns is 0.8 percent, the net effect is only 0.2 percent. Second, the net effects of campaigns on different days can cancel out. One candidate may gain 0.2 percent on a particular

extent to which they lasted to impact the outcome on Election Day. Even small changes can make a difference on Election Day, after all, particularly if the race is close and if campaign effects accumulate in one direction or the other. The question then is whether the shocks from events take the form of temporary "bounces" or permanent "bumps" (Erikson and Wlezien 2012). Simply put, do the effects decay or else last? If campaign effects are bounces, they dissipate over time. In this scenario, preferences revert to an "equilibrium" that is set in each particular election year (also see Gelman and King 1993). The final outcome is the equilibrium plus the effects of any very late events that do not fully dissipate before Election Day. If campaign effects are bumps, conversely, they last to affect the outcome. The election outcome is the sum of all the bumps—mainly small in size individually—that happen during the campaign, keeping in mind that they can go in both directions and cancel each other out. Of course, it may be that campaign events produce both bounces and bumps: some effects may dissipate and others last, or a portion of effects may dissipate and the rest last. The bumps and not the bounces are what matter in the long run. They cumulate over time. Figure 8 illustrates the different types of campaign effects—the bump, the bounce and the hybrid effect.

-- Figure 8 about here --

In theory, we could tell whether bumps or bounces predominate from regression analysis of poll results on consecutive days. That is, we could model the time series of aggregate voter preferences (V_t) during the 2016 presidential election as follows:

(1) $$V_t = \alpha + \beta V_{t-1} + e_t,$$

day but lose 0.1 percent the following day. Here the total effect for the two days is 0.3 percent but the net effect is only 0.1 percent. Unfortunately, the separate effects of the different campaigns are almost impossible to disentangle.

where V_t is one candidate's percentage share in the polls and e is a series of independent campaign shocks drawn from a normal distribution. That is, preferences on one day are modeled as a function of preferences on the preceding day and the new effect of campaign events, broadly defined. In theory, dynamics are directly evident from the coefficient β in equation 1.

If $0 \leq \beta < 1$, effects on preferences decay. As an "autoregressive" (AR) process, preferences tend toward the equilibrium of the series, which is $\alpha / (1 - \beta)$. This equilibrium does not represent the final outcome, as what happens on Election Day also will reflect late campaign effects that have not fully dissipated by the time voters go to the polls. That said, the degree to which late campaign effects do matter is evident from β, which captures the rate of carryover from one point in time to the next. The smaller the β the more quickly effects decay. If this is the correct model, daily campaign effects would be of no electoral relevance except for those that occur at the *very end* of the race, on Election Day itself. In one sense, this characterization of campaign effects is implicit in most forecasting models of election outcomes, where the "fundamentals" are assumed to be constant for much of the campaign (Gelman and King, 1993).

Now, if β equals 1.00, campaign effects cumulate. Each shock makes a permanent contribution to voter preferences. As an "integrated" process, preferences wander over the election cycle and become more and more telling about the final result. The actual outcome is simply the sum of all shocks that have occurred during the campaign up to and including Election Day. This clearly is a very strong model of campaign effects. It is the one implied by on-line processing models of voter preferences (see, e.g., Lodge, Steenbergen, and Brau, 1995).

It may be that neither one of these models applies strictly to all campaign effects. That is, preferences may not evolve as either a pure autoregressive or integrated process but as a "combined" process, where some effects decay and others persist (Wlezien, 2000). Certain events may have temporary effects and others permanent ones. Some events may produce both effects, where preferences move and then bounce back though to a different level.[12] As a combined series, preferences would evolve much like an integrated process, though with less drift.[13]

In theory, then, we can tell quite a lot about the persistence of campaign effects in 2016 by simply estimating equation 1 using our series of adjusted polls. We want to know whether the AR(1) parameter is equal to or less than 1.0: if the parameter equals 1.0, we can conclude that *at least* some effects persist; if the parameter is less than 1.0, campaign effects would appear to decay. Analysis of the 2016 data produces an AR parameter of 0.79 (s.e. = 0.03) for the full election year and 0.91 (s.e. = 0.05) for the post-Labor Day period. Both of these estimates are below 1.0, which implies that the effects of preferences to decay. Dickey-Fuller (DF) tests indicate that only the first parameter is significantly less than 1.00, however, and so it may be that much of what happened during the fall actually lasted to impact the final voted.[14]

---

[12] We can represent this process as follows:

$V_t = V^*_{t-1} + \beta\,(V_{t-1} - V^*_{t-1}) + g_t, + u_t,$

where $0 \leq \beta < 1$ and u_t represents the series of shocks to the fundamentals. Notice that this is the equivalent of an error correction model. In this model, some effects (u_t) persist–and form part of the moving equilibrium V^*_t–and the rest (g_t) decay. The ultimate outcome is the Election Day equilibrium V^*_{ED}, which is the final tally of u_t, plus the effects of late-arriving campaign effects (g_t) that have not fully decayed.

[13] Indeed, statistical theory (Granger, 1980) tells us that any series that contains an integrated component is itself integrated. The intuition is that, because of its expanding variance over time, the integrated component will dominate in the long run. Over finite time, however, it may not always work out so neatly (see Wlezien, 2000).

[14] The DF test is commonly used to assess whether a time series is integrated (the null hypothesis) or else stationary. To conduct a DF test, one simply regresses the first difference ($V_t - V_{t-1}$) of a variable on its lagged level and then compares the *t*-statistic of the coefficient. The intuition is straightforward. If the coefficient for the lagged level variable is indistinguishable from 0, we know that future change in the variable is unrelated to its current level. This tells us that the variable is integrated. If the coefficient is negative and significantly different from 0, future change is related to the current level in predictable ways, that is, the variable regresses to the equilibrium or long-term mean of the series. This implies that the series is stationary. Note that appropriate critical values for the DF *t*-tests are nonstandard (see MacKinnon, 1991).

The pattern of results, particularly the seeming persistence during the fall campaign, may be deceiving, as the pooled poll readings on consecutive days are not completely independent. (They probably are best viewed as a moving average.) The apparent day-to-day relationship thus may be to some extent artifactual. The most obvious and basic way to address the issue is to examine the autocorrelation function (ACF) for our series of poll readings across different lags. We are particularly interested in whether correlations remain fairly stable or else decline as the gap between readings widens. If the process is stationary and campaign effects decay, the correlation between poll readings will decline geometrically. Specifically, if the AR(1) parameter is ρ, the correlation between poll readings x days apart is ρ^x. If campaign effects persist, conversely, the correlation would remain fairly constant as the number of days between poll readings increases.

-- Table 4 about here –

Table 4 presents the correlations between adjusted poll results and their lagged values over 1 to 10 days, separately for the entire election year and the post-Labor Day period. Consider first the pattern of correlations for the full year in the first column. For expository purposes, the correlations are plotted in Figure 9. The correlations with the poll reading at days *t*-1 and t-2 are understandably high -- 0.79 and 0.62 -- given the overlapping results on successive days. For lags of 3-10 days, however, the correlations change only slightly. This tells us that preferences are not strictly stationary--the correlations across lags do not decay geometrically. The pattern of correlations is actually what we would expect of a combined process, where some effects last indefinitely and others decay (Wlezien, 2000). It implies that some meaningful portion of effects in 2016 did not decay and instead persisted over time.[15]

[15] The pattern also is what we might expect of a fractionally integrated (FI) series, where effects decay but much more slowly than in a stationary series (Box-Steffensmeier and Smith, 1998; Lebo, Walker and Clarke, 2000). Unfortunately, identifying such a process is difficult; that is, the statistical power of the tests is low (DeBoef, 2000; Wlezien, 2000).

-- Figures 9-10 about here –

The correlations for the post-Labor Day period in the second column of Table 4 reveal a similar pattern. Also see Figure 10. Although these correlations start off slightly higher and decline more sharply, they nevertheless do not decay geometrically. Of course, we have not yet taken into account sampling error, which dampens the observed autocorrelations. To obtain true over-time correlations, we simply divide the observed correlations for long lags, i.e., where there is no overlap in the polls, by the statistical reliability of the readings (Heise, 1969). Assuming the upper-bound estimate of 0.49 from Table 3a, the resulting correlations for lags 7-10 hovers between 0.60 and 0.90. That the correlations are below 1.0 implies that some campaign effects did not last. That the correlations are so high at long lags, however, suggests that a significant portion of campaign effects actually did persist over time to impact the outcome. What happened during the 2016 presidential election campaign, it appears, really mattered.

Mass Media Coverage during the Election Cycle

As noted above, it is difficult to capture the impact of a campaign. We have seen that specific events – namely, conventions and, to a lesser extent, the debates – produced shifts in voter preferences in 2016. But how might we capture that campaign more broadly? One possibility is to focus on media coverage. The modern campaign is experienced by most voters only through mass media, after all, though alongside discussion with friends and family, who also are experiencing the campaign mainly through mass media. And while news coverage does not capture parties' direct appeals to voters (through ads or canvassing, for instance), it should reflect much if not most of the parties' national campaigns, alongside related news, horserace coverage, etc. It follows that one relatively straightforward approach to capturing shifts in the presidential campaign is to focus on the "tone" of media coverage of both candidates, across a range of media outlets, over the course of the election timeline.

The tone of media coverage of each candidate is captured here using content from January 1st to November 9th 2016, from nine major US newspapers: the Chicago Sun-Times, Denver Post, Houston Chronicle, LA Times, New York Times, Philadelphia Inquirer, St. Louis Post-Dispatch, USA Today, and Washington Post. We note that this list includes three of what we would regard as the two "papers of record" (New York Times and Washington Post), alongside USA Today and six broadsheets from different regions of the country. We use any article that mentions either Clinton or Trump, based on a full-text search in Lexis-Nexis. Articles are downloaded using the Lexis-Nexis Web Services Kit (WSK), which allows for large-scale downloads of xml-formatted data. In the end, our databased includes 29,288 articles over this time period.

The tone of coverage is identified using the Lexicoder Sentiment Dictionary (LSD) in Lexicoder (Daku et al. 2015), a multi- platform Java-based automated content software available at lexicoder.com. The software applies a simple dictionary-based approach to content analysis – in this case, it counts the number of positive and negative words in the LSD.[16] We do not want the tone of the entire article, however – we want to focus here on the tone of coverage connected to one candidate or the other. We thus use a hierarchical dictionary count (or proximity count), in which we count positive and negative words only in sentences that mention one or the other candidate. The LSD is discussed and tested in some detail in Young and Soroka (2012). Using the use of the tone of media coverage – either through human or automated coding through a hierarchical dictionary count – to capture the tone of coverage for candidates and parties has been tested with some success in

[16] The dictionary includes roughly 4,567 entries in total, so full details are available via lexicoder.com. Even so, by way of example, negative words include, e.g., abandon, cheat, harm, reckless, tainted, and positive words include, e.g., assure, dependable, honest, resourceful.

Soroka et al. (2009) and Belanger and Soroka (2012). Given that we are averaging across different papers, we weight the counts by the average daily circulations of the newspapers.

-- Figures 11-13 about here –

The only way in which our measure differs from this past work (aside from its application to 2016 US data, of course) is that we rely on a slightly different approach to measuring "net tone," i.e., the overall tone of candidate coverage based on the difference in positive versus negative words co-occurring with each candidate name. Here, we use a measure suggested in Lowe et al. (2011), and tested in Proksch et al. (2016), as follows: log [(pos counts + .05) / (neg counts + .05)], which is an empirical logit, slightly smoothed towards zero.[17]

The resulting daily data are summarized in Figure 11, with three-day moving averages depicted in Figures 12 and 13. Specifically, the figures show the net Clinton minus Trump tone over the election year. What is most evident – and perhaps surprising to some – is that the tone of coverage tended to be more positive toward Clinton than Trump. The mean 3-day moving average is 0.10 with a range of between -0.19 and 0.34 (s.d. = 0.11). Net tone was negative on only ¼ of the days during the election year, particularly during the Republican convention and in the weeks preceding it, perhaps owing to the Comey's July announcement. Importantly, the patterns of generally positive Clinton coverage are virtually identical during the fall campaign, during which time the trend in tone actually was positive on balance (see Figure 13)

-- Table 5 about here –

Media coverage exhibits very different dynamics to what we saw for voter preferences. Whereas the latter demonstrate persistence, the tone of coverage does not. This is clear from the ACFs in Table 5. There is a very weak correlation (0.32) between net media tone on successive days

[17] Previous tests suggest that this measure is very highly correlated with other approaches, at >.95.

and the correlations drop off fairly sharply using longer lags, especially after Labor Day. Even over the long campaign, there is little momentum in media tone. It's as if this aspect of media coverage is created anew each day, or every other day, which has potentially meaningful consequences. Indeed, it fits with the metaphor of the campaign as a series of (perhaps biased) daily coin flips.

That said, net tone is only modestly correlated with polls in 2016. Over the long campaign, the correlation using daily net tone peaks 0.21 using media readings at day t-2, and is only slightly larger (0.23) using 3-day moving averages. After Labor Day, the correlations actually are negative.[18] Net tone is meaningful, however. This can be seen in the regression analyses in Table 6, where daily net tone is regressed on lagged tone and measures capturing different political events. The first column simply includes variables for the two conventions, each of which takes the value "1" during each of the four days of the convention and "0" otherwise. The results indicate that both conventions moved media tone in the expected directions; indeed, the coefficients – -0.21 and 0.18 – are almost equal and opposite, implying almost identical effects. Given slight momentum in coverage indicated by the coefficient (0.29) for the lagged tone variable, the effects of the conventions actually increase some over time, and peak at -.30 and 0.27, respectively. Each is approximately one-third of the range of the daily tone variable and two-thirds that of the 3-day moving average.

-- Tables 6 and 7 about here –

The second column of Table 6 shows that the effects of other events – the first Comey announcement in July, the first debate, the Trump "Access Hollywood" tape, and the Comey letter in October – on our measure of media tone are uneven. The estimated equation actually includes three lagged dummy variables for each of the events, and the table simply displays the average of the three coefficients and also the test of the significance of the sum of the coefficients. The first Comey

[18] Not surprisingly, the correlations for the pre-Labor Day period actually are larger – 0.27 and 0.31 – than they are for the full year.

announcement had the expected negative effect, and one that greater than the effects of the conventions. The same is not true for first debate, the sex tape, and even Comey's October letter, which appeared to matter little for media tone, at least based on the equation. The latter is a special surprise, as an effect seems clear from Figure 13, but keep in mind that tone even then is only slightly below average and fairly neutral in absolute terms.

The measure of media tone does matter for electoral preferences. This can be seen in Table 7, which summarizes regression analysis of the pooled poll results including lagged polls and media tone. The first column shows that Clinton-Trump tone positively predicts Clinton's support over the full election year, though the coefficient (0.92) implies a modest effect. That is, a one standard deviation (.15) shift in measured tone predicts a 0.13 increase in the poll share. Based on the results, Comey's summer announcement and the two conventions had real, if small, mostly short-lived effects on preferences, and the primary political events of the fall did not matter for voter sentiment. This is borne out by analysis (not reported here) that includes the variables from Table 7 into the equations in Table 8 The second column reveals that the effect of tone does not hold during the Fall campaign, as the coefficient is negative, though insignificant. Mass media coverage just does not matter equally throughout the election year. Cross-correlation analysis in Table 8 suggests that the effects might even be reversed during the Fall, whereby public support for candidates drives future media coverage, not the other way around. While surprising, it is not entirely counterintuitive, as there is reason to suppose that media coverage reflects public preferences, and there is evidence for such an effect in media coverage of the economy, which seems to follow (as well as lead) public perceptions (Soroka, et al, 2015). Whether the same holds true for electoral preferences still remains to be seen.

-- Table 8 about here –

Discussion

Political scientists debate the role of political campaigns and campaign events in electoral decision-making (see, for example, Alvarez 1997; Campbell, 2000; Gelman and King 1993; Holbrook 1996; Erikson and Wlezien, 2012). An examination of polls during the 2016 presidential election cycle suggests that for this particular campaign, events mattered a little, at least by comparison with previous elections. Although much of the variance of poll results over time is the result of survey error, a portion appears to reflect change in underlying preferences, both over the course of the full election year and the autumn campaign itself. Perhaps more importantly, there is evidence that at least some of the effects of the campaign did not decay but actually persisted over time to affect the outcome. The 2016 presidential election campaign, it appears, mattered.

Of course, the analysis has focused on a single election in a single year. What about presidential elections in other years? What does the research tell us about the effects of election campaigns more generally? Clearly, one cannot generalize based on the foregoing analysis. After all, a fairly similar analysis of polls during the 1996 general election campaign revealed almost no effects whatsoever (Erikson and Wlezien, 1999) and another of the 2000 general election found much larger effects (Wlezien, 2003). Even in 2016, it is not clear how the campaign mattered. What events had effects? Which ones lasted and which decayed? The foregoing analyses, even those including media coverage offer only limited insight. What they offer are very general conclusions about the dynamics of electoral preferences in a particular year. Given the data, this may be all that we can hope to provide.

References

Alvarez, R. Michael. (1997). *Information and Elections*. Ann Arbor: University of Michigan Press.

Bartels, Larry M., and John Zaller. 2001. "Presidential Vote Models: A Recount." *PS* 33:9-20.

Belanger, Eric and Stuart Soroka. 2012. "Campaigns and the Prediction of Election Outcomes: Can Historical and Campaign-Period Prediction Models Be Combined?" *Electoral Studies* 31: 702-714.

Box-Steffensmeier, Janet, & Smith, Renee. (1998). "Investigating Political Dynamics Using Fractional Integration Models." *American Journal of Political Science* 42:661-689.

Campbell, Angus, Converse, Philip E., Miller, Warren E., & Stokes, Donald E. (1960). *The American Voter*. New York: Wiley.

Campbell, James E. (2017). "Political Scientists' Forecast of the 2016 Election." *PS: Political Science and Politics*, forthcoming.

Campbell, James E., (2000). *The American Campaign: U.S. Presidential Campaigns and the National Vote*. College Station, Texas: Texas A&M University Press.

-----, & Garand, James C. (2000). *Before the Vote: Forecasting American National Elections*. Thousand Oaks, Calif.: Sage Publications.

Converse, Philip E., & Traugott, Michael W. (1986). "Assessing the Accuracy of Polls and Surveys." *Science* 234:1094-1098.

Crespi, Irving. (1988). *Pre-Election Polling: Sources of Accuracy and Error*. New York: Russell Sage.

Daku, Mark, Stuart Soroka, and Lori Young. 2015. Lexicoder, version 3.0. Available at: www.lexicoder.com.

DeBoef, Suzanna. (2000). “Persistence and Aggregations of Survey Data Over Time: From Microfoundations to Macropersistence.” *Electoral Studies* 19: 9-29.

Erikson, Robert S., Joseph Bafumi, and Bret Wilson. (2001). “Was the 2000 Presidential Election Predictable?” *PS* 33:815-819.

Erikson, Robert S., & Wlezien, Christopher. (2014). *The 2012 Campaign and the Timeline of Presidential Elections*. Chicago: University of Chicago Press.

------. (2012). *The Timeline of Presidential Elections. Chicago: University of Chicago Press*.

-----. (1999). “Presidential Polls as a Time Series: The Case of 1996.” *Public Opinion Quarterly* 63:163-177.

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.

Erikson, Robert S., and Christopher Wlezien. 2012. *The Timeline of Presidential Elections*. Chicago: University of Chicago Press.

Gelman, Andrew, & King, Gary. (1993). “Why are American Presidential Election Polls so Variable When Votes are so Predictable?” *British Journal of Political Science* 23: 409-519.

Groves, Robert M. (1989). *Survey Errors and Survey Costs*. New York: Wiley.

Heise, D.R. (1969). “Separating Reliability and Stability in Test-Retest Correlations.” *American Sociological Review* 34:93-101.

Holbrook, Thomas. (1996). *Do Campaigns Matter?* Thousand Oaks, California: Sage.

Johnston, Richard, Blais, Andre, Brady, Henry E., & Crete, Jean. (1992). *Letting the People Decide: Dynamics of a Canadian Election*. Kingston, Canada: McGill-Queen's Press.

Lau, Richard. (1994). “An Analysis of the Accuracy of ‘Trial Heat’ Polls During the 1992

Presidential Election." *Public Opinion Quarterly* 58:2-20.

Lazarsfeld, Paul F., Berelson, Bernard R., & Gaudet, Hazel. (1944). *The People's Choice*. New York: Columbia University Press.

Lebo, Matthew J., Walker, Robert W., & Clarke, Harold D. (2000). "You Must Remember This: Dealing With Long Memory in Political Analyses." *Electoral Studies* 19:31-48.

Lewis-Beck, Michael S. (1988). *Economics and Elections*. Ann Arbor: University of Michigan Press.

Lodge, Milton, Steenbergen, Marco, & Brau, Sean. (1995). "The Responsive Voter: Campaign Information and the Dynamics of Candidate Evaluation." *American Political Science Review* 89:309-326.

Lowe, Will, Kenneth Benoit, Slava Mikhaylov and Michael Laver. 2011. "Scaling Policy Preferences from Coded Political Texts." *Legislative Studies Quarterly* 36(1):123–155

MacKinnon, James G. 1991. "Critical Values for Cointegration Tests." In Robert F. Engle and Clive W. Granger, eds., *Long Run Economic Relationships*. New York: Oxford University Press.

Proksch, Sven-Oliver, Will Lowe and Stuart Soroka. 2016. "Multilingual Sentiment Analysis: A New Approach to Measuring Conflict in Parliamentary Speeches." Paper presented at the Annual Meeting of the American Political Science Association, Philadelphia PA.

Shaw, Daron R. (1999). "A Study of Presidential Campaign Event Effects from 1952 to 1992." *Journal of Politics* 61:387-422.

Soroka, Stuart, Marc Andre Bodet, Lori Young and Blake Andrew. 2009. "Campaign News and Vote Intentions." *Journal of Elections, Public Opinion and Parties* 19(4): 359-376.

Soroka, Stuart N., Dominik A. Stecula, and Christopher Wlezien. (2015). "It's (Change in) the (Future) Economy, Stupid: Economic Indicators, the Media, and Public Opinion." *American Journal of Political Science* 59:457-474, Traugott, Michael W. 2001. "Assessing Poll Performance in the 2000 Campaign." *Public Opinion Quarterly* 63:389-419.

Vavreck, Lynn. (2009). The Message Matters. Princeton: Princeton University Press.

Wlezien, Christopher. (2003). "Presidential Election Polls in 2000: A Study in Dynamics." *Presidential Studies Quarterly* 33: 1:172-186.

-----. (2001). "On Forecasting the Presidential Vote." *PS: Political Science and Politics* 34:25-31.

-----. (2000). "An Essay on 'Combined' Time Series Processes." *Electoral Studies* 19:77-93.

Wlezien, Christopher, and Erikson, Robert S. (2002). "The Timeline of Presidential Election Campaigns." *Journal of Politics* 64:969-993.

-----. (2001). "Campaign Effects in Theory and Practice." *American Politics Research* 29:419-437.

Young, Lori and Stuart Soroka. 2012. "Affective News: The Automated Coding of Sentiment in Political Texts." *Political Communication* 29: 205-231.

Zaller, John. (2002). "Assessing the Statistical Power of Election Studies to Detect Communication Effects in Political Campaigns." *Electoral Studies* 21:297-329.

Figure 1: All Trial-Heat Presidential Polls by Date, 2016

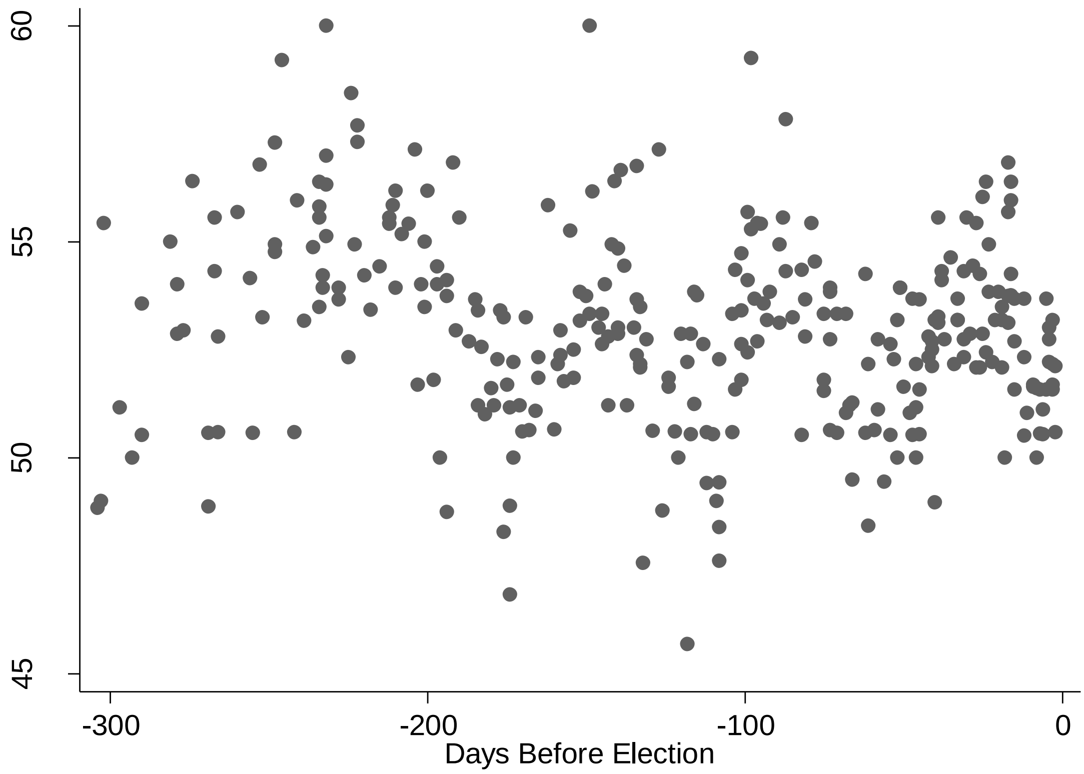


Figure 2: Trial-Heat Presidential Polls Aggregated by Date, 2016

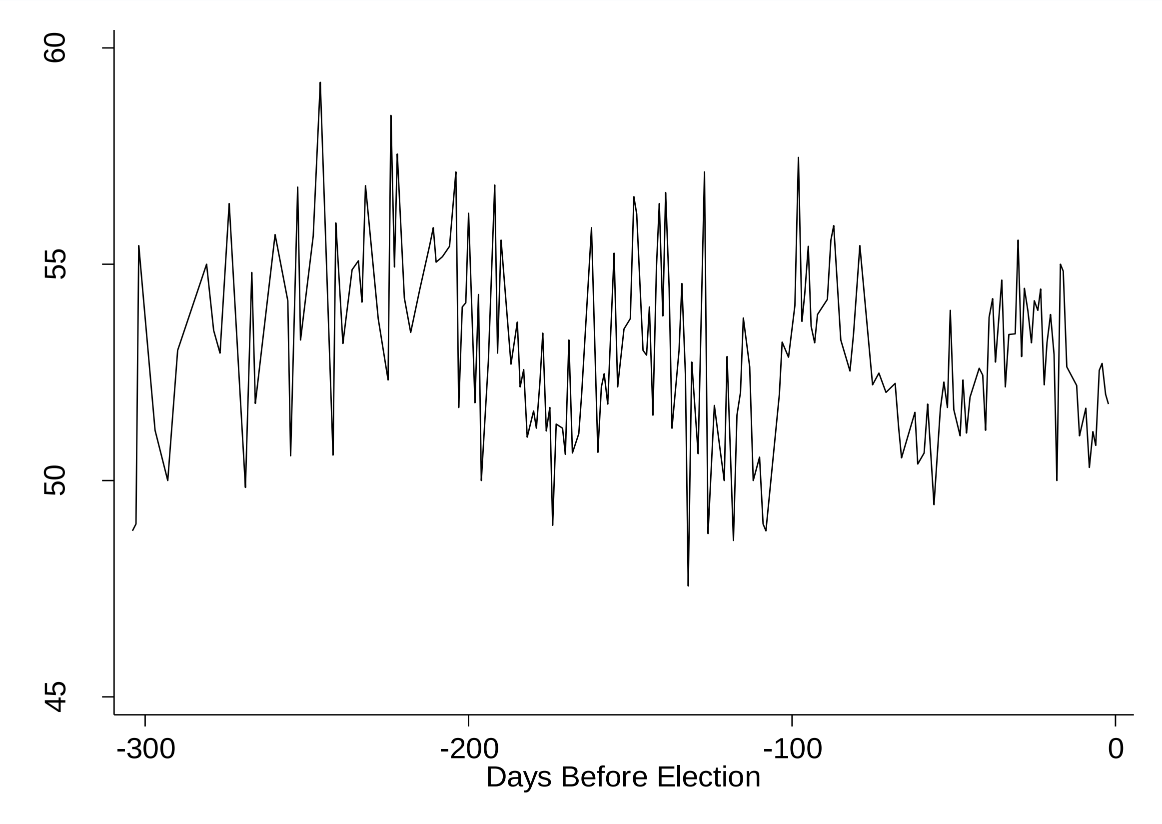


Table 1: An Analysis of House and Date Effects on Presidential Election Polls, 2016

Variable	Election Year	After Labor Day
Poll Universe	0.01	0.26
	(0.91)	(0.61)
Survey House	3.38	2.39
	(0.00)	(0.02)
Survey Date	2.97	1.61
	(0.00)	(0.09)
R-squared	0.90	0.87
Adjusted *R*-squared	0.68	0.57
Mean Squared Error	1.23	1.09
Number of polls	307	100
Number of respondents	420,260	109,135

Note: The numbers corresponding to the variables are F-statistics. The numbers in parentheses are two-tailed *p*-values.

Table 2: Selected House Effects on Presidential Election Polls, 2016

House	Effect on Polls (Relative to Median House)
UPI/CVOTER	-2.04
ARG	-1.46
Politico/Morning Consult	-1.04
IBD/TIPP	-0.94
Morning Consult	-0.53
YouGov/Economist	-0.47
CBS/Times	-0.39
Quinnipiac	-0.37
Gravis Marketing/OANN	-0.16
FOX	0.00
CNN	0.06
Ipsos/Reuters	0.17
NBC/SurveyMonkey	0.41
ABC/Post	0.55
CBS	0.75
NBC/WSJ	0.88
Suffolk/USA Today	1.09
Monmouth University	1.59
McClatchy/Marist	1.62
Bloomberg/Selzer	1.89
PPP (D)	2.35

Figure 3: House-Adjusted Trial-Heat Presidential Polls by Date, 2016

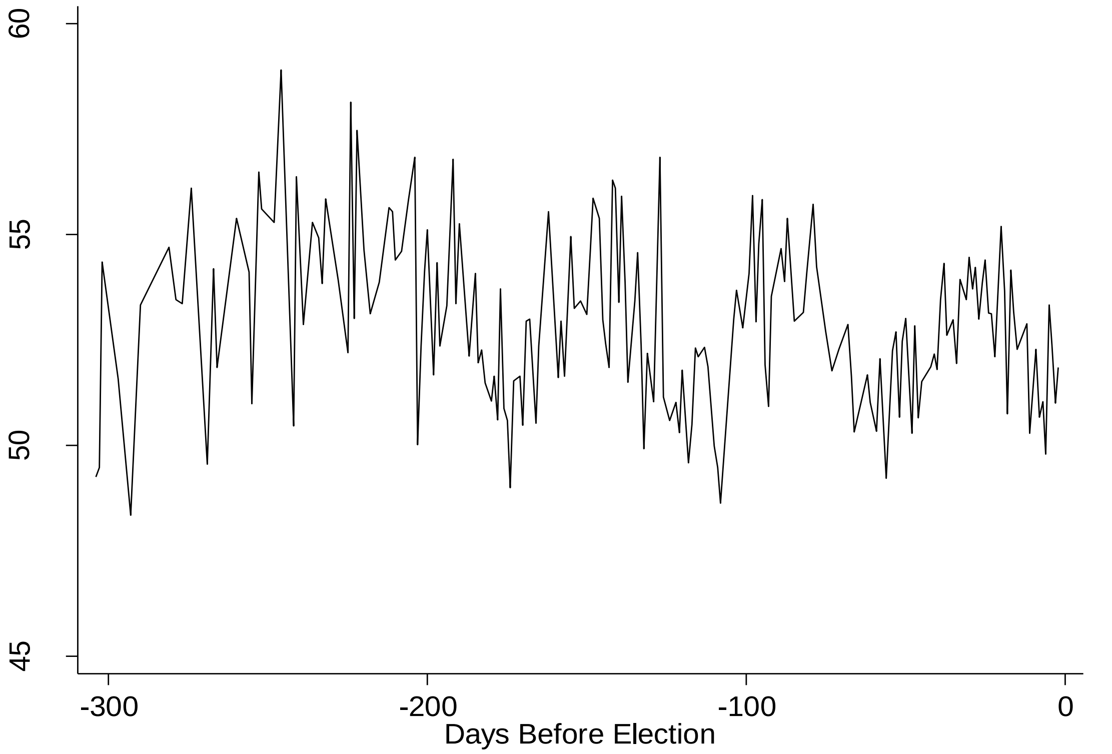


Table 3a. Variance of Daily Trial Heat Poll Readings in 2016

Time Frame	*Total Variance*	*Error Variance*	*True Variance*	*Reliability*
Full year	3.88	1.11	2.77	.71
Last 200 days	2.92	1.09	1.83	.63
Last 64 days	1.88	0.96	0.92	.49

Table 3b. Daily Variance of Trial Heat Polls During the Final 60 Days, 1952-2012

	Total Variance	*Error Variance*	*True Variance*	*Reliability*
Mean	3.69	1.86	2.30	.45
Median	3.71	1.77	1.65	.45

Note: Based on polls that have not been adjusted for house effects.

Figure 4: House-Adjusted Trial-Heat Presidential Polls Pooled by Date, 2016

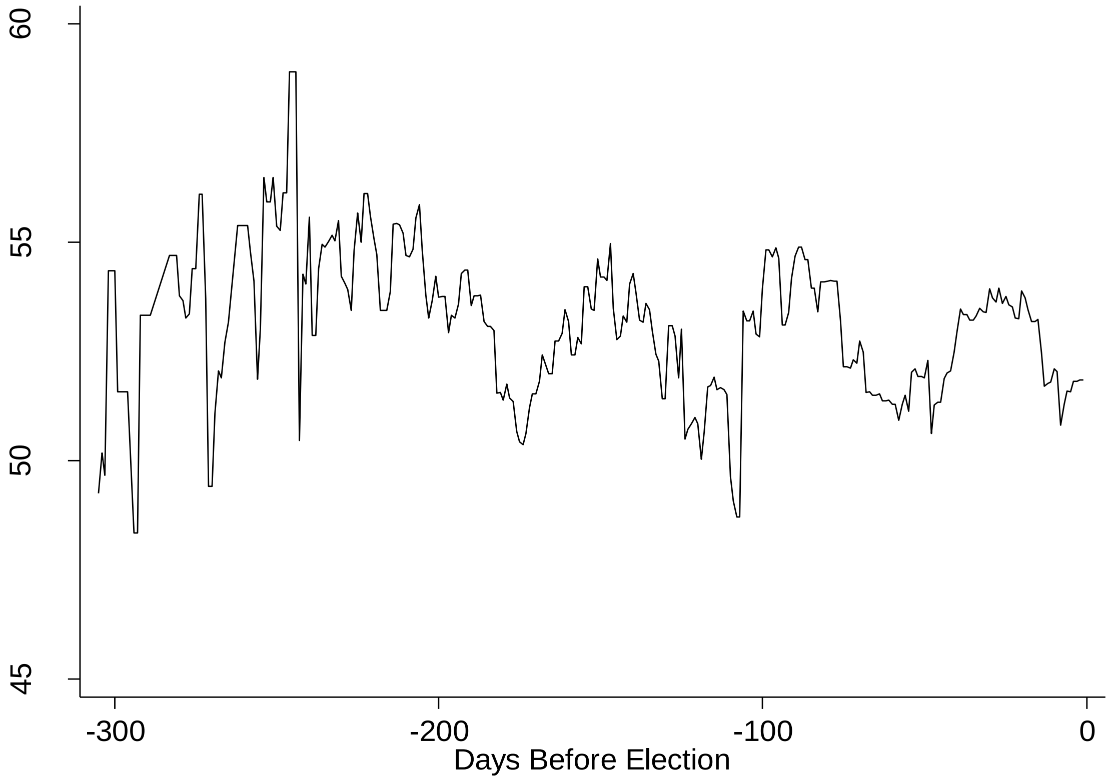


Figure 5: Pooled House-Adjusted Trial-Heat Presidential Polls by Date, Labor Day to Election Day, 2016

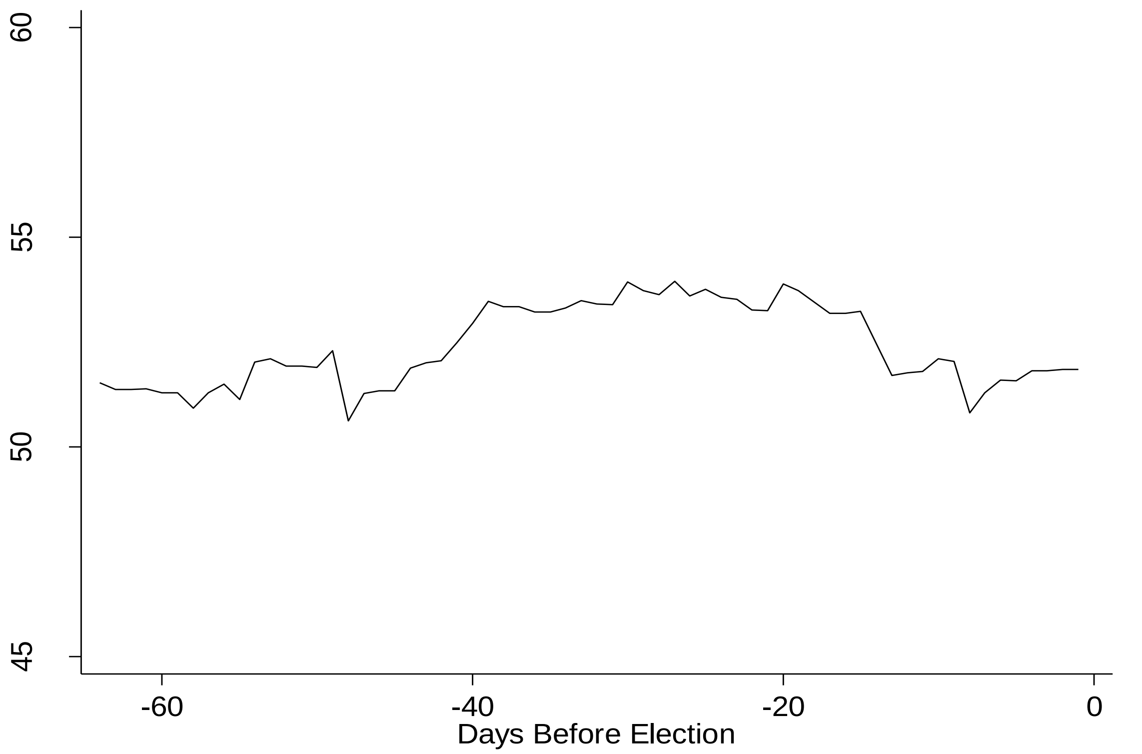


Figure 6: Trial-Heat Polls Before and After the Conventions, 2016

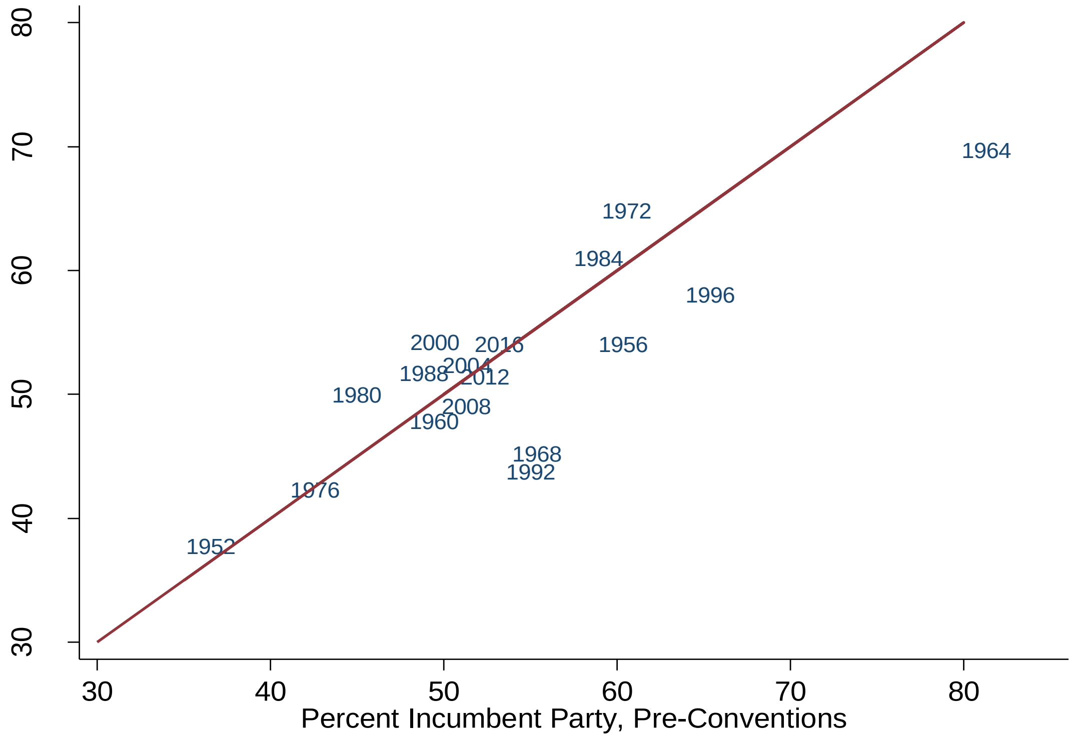


Figure 7: Trial-Heat Polls Before and After the Debates, 2016

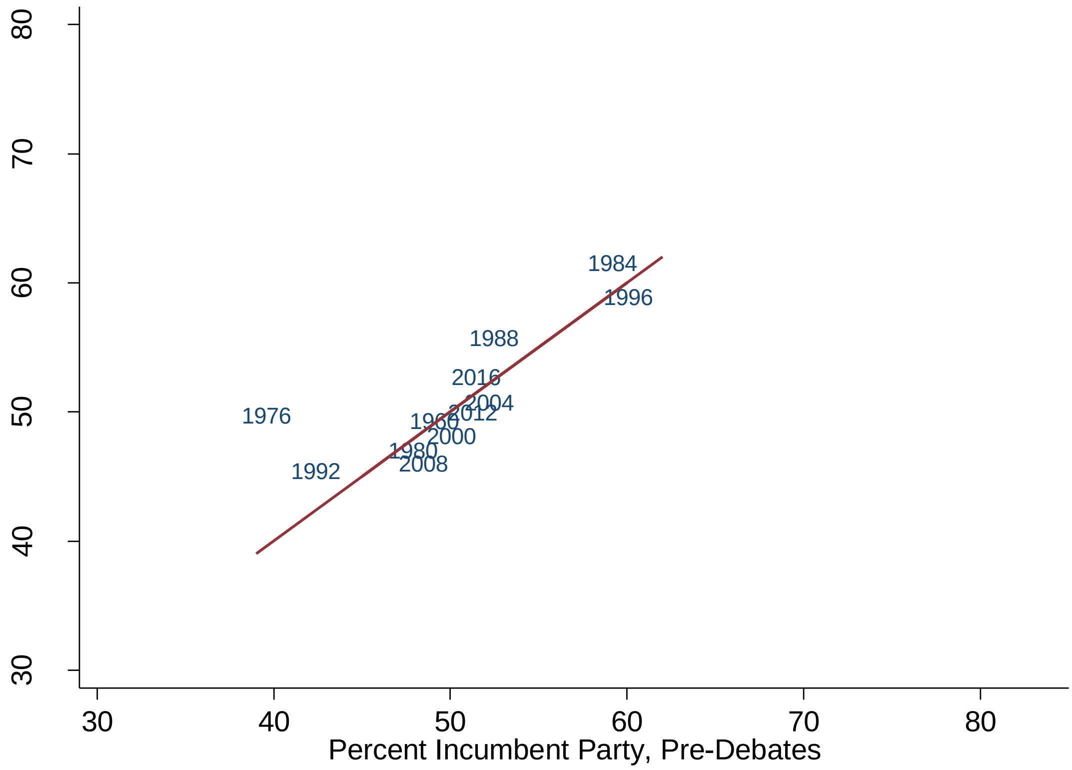
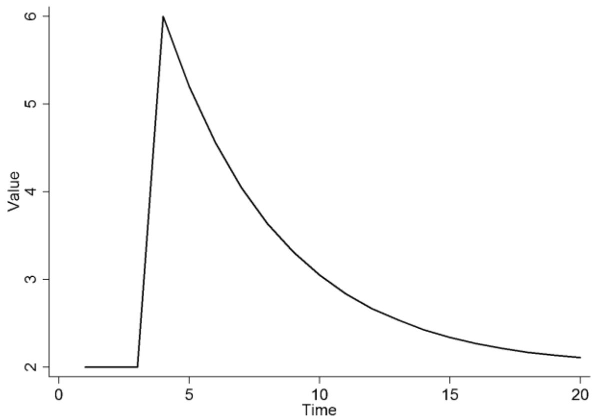
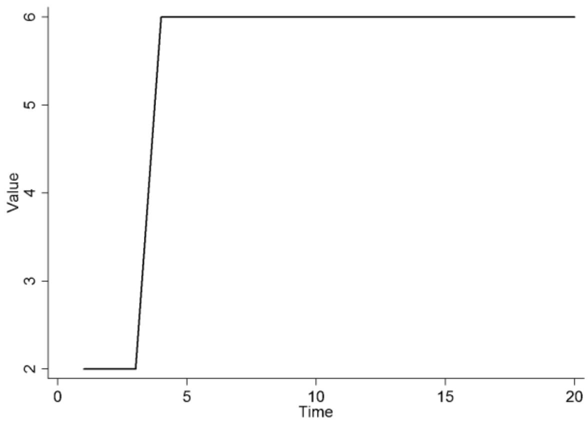


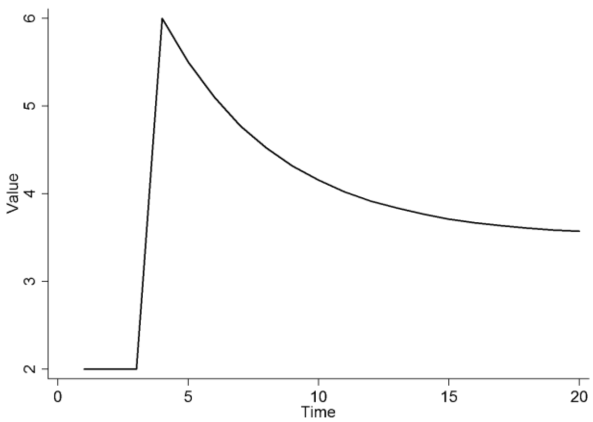
Figure 8: Bounces and Bumps



a, A Bounce



c. A Bump



c. A Hybrid Effect

Table 4. Autocorrelation Functions for Allocated House-Adjusted Polls

	Correlation	
Lag	*Election Year*	*Post-Labor Day*
t-1	.79	.91
t-2	.62	.84
t-3	.49	.78
t-4	.47	.71
t-5	.45	.65
t-6	.44	.55
t-7	.42	.46
t-8	.39	.38
t-9	.37	.33
t-10	.33	.29

Figure 9: Autocorrelations for Allocated House-Adjusted Polls, Full Year, 2016

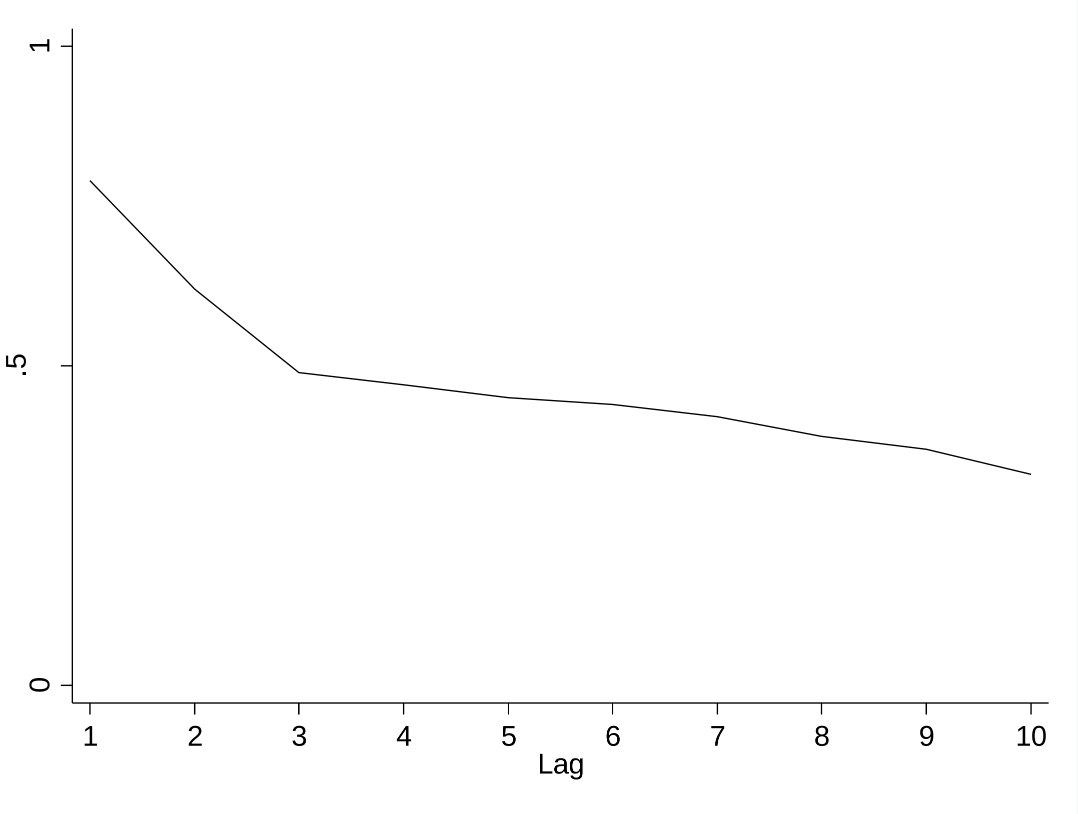


Figure 10: Autocorrelations for Allocated House-Adjusted Polls,
Labor Day to Election Day, 2016

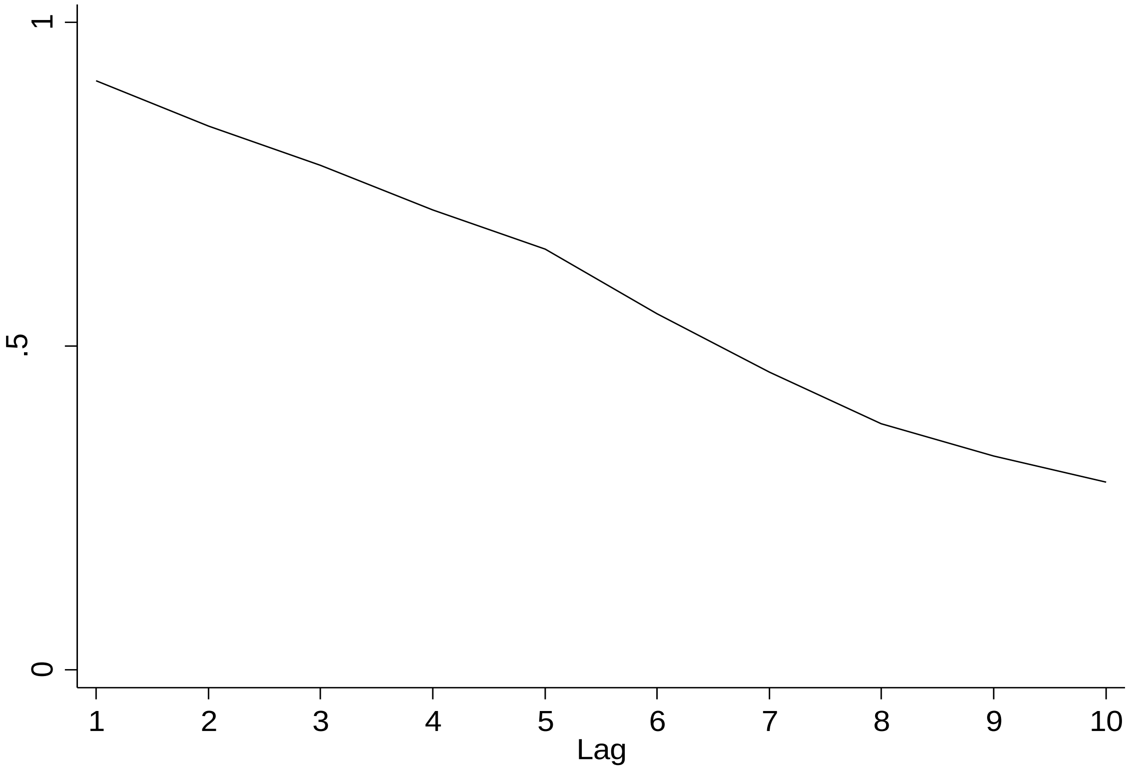


Figure 11: Net Clinton minus Trump Media Tone by Date, 2016

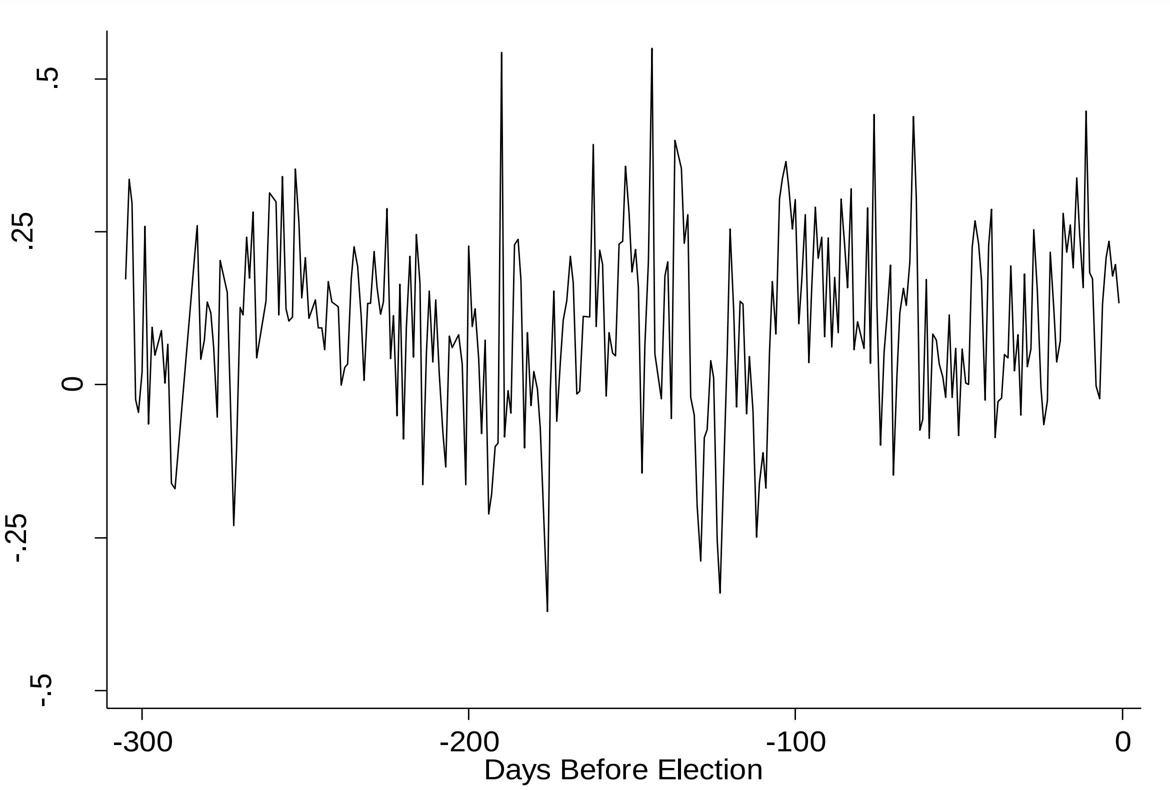


Figure 12: Net Clinton minus Trump Media Tone, 3-day Moving Averages, 2016

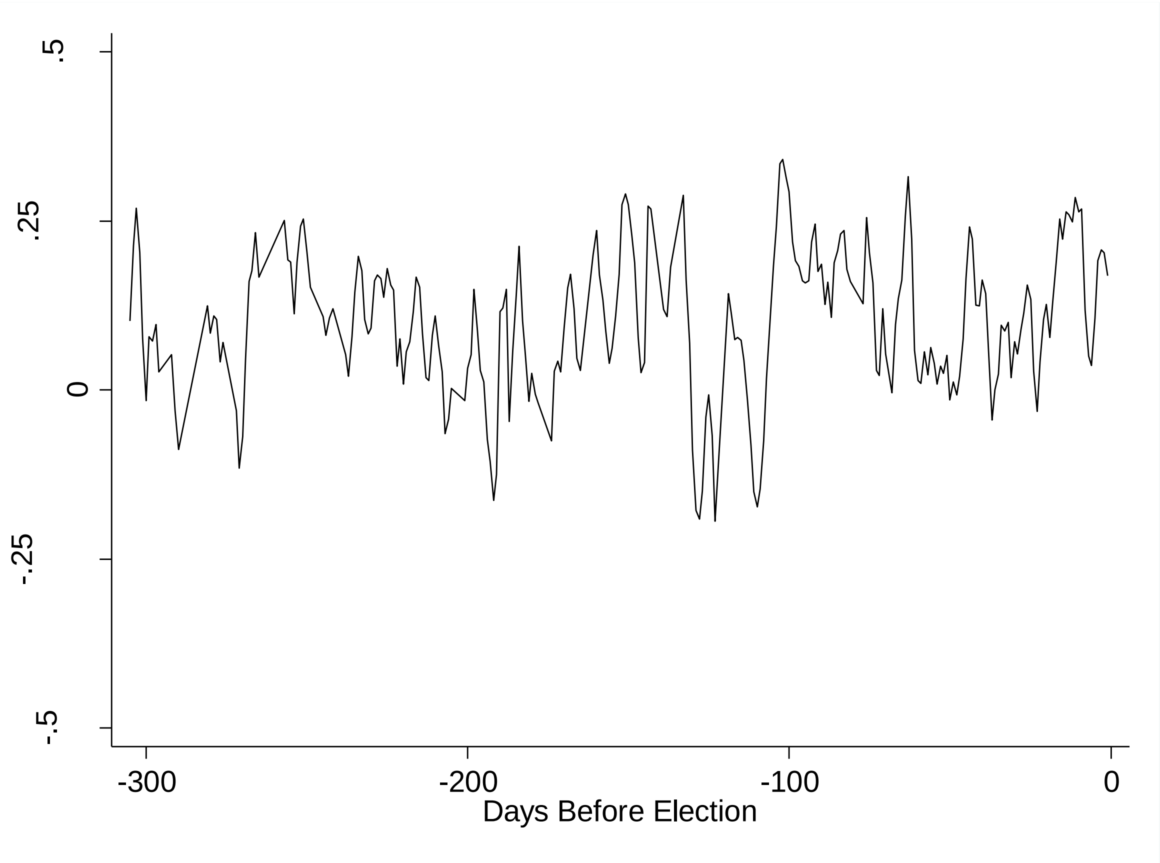


Figure 13: Net Clinton minus Trump Media Tone, 3-day Moving Averages, Labor Day to Election Day, 2016

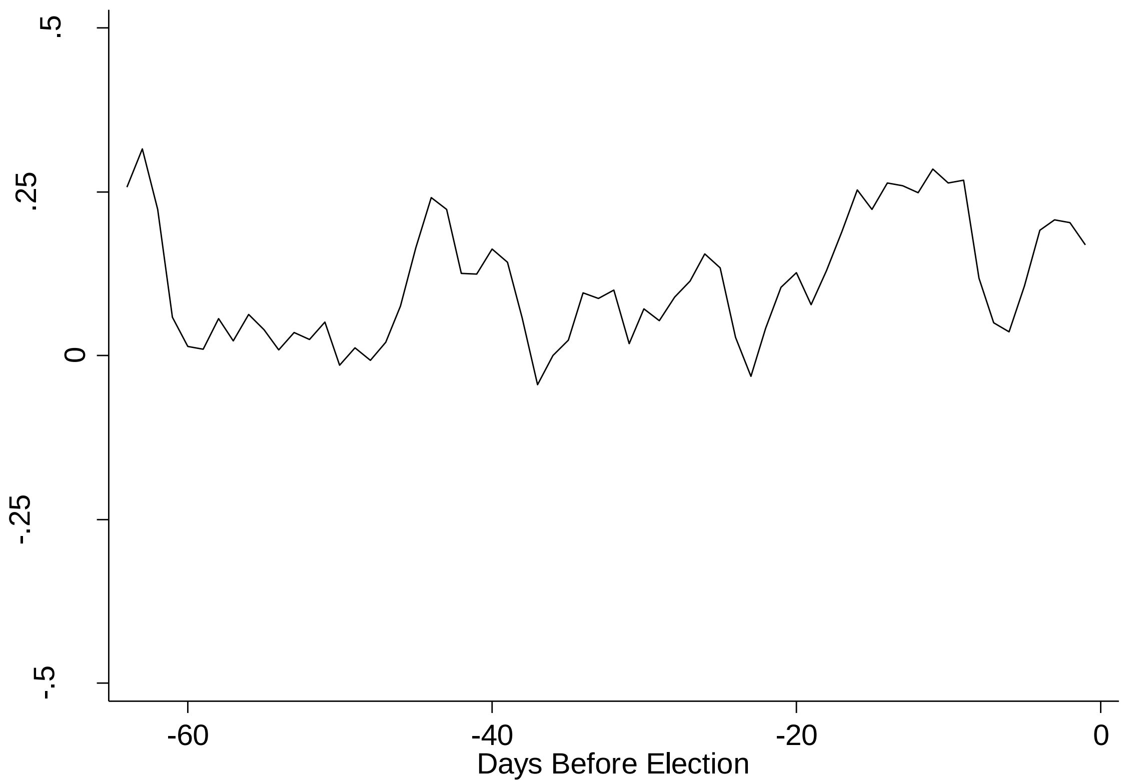


Table 5. Autocorrelation Function for Daily (and 3-day Moving Averages of) Media Tone, 2016

	Correlation	
Lag	*Election Year*	*Post-Labor Day*
t-1	.32 (.77)	.32 (.79)
t-2	.22 (.48)	.08 (.47)
t-3	.10 (.24)	.09 (.21)
t-4	.08 (.16)	.12 (.15)
t-5	.12 (.12)	.08 (.12)
t-6	.02 (.09)	.02 (.08)
t-7	.06 (.07)	.02 (.03)
t-8	.04 (.04)	.01 (-.03)
t-9	.04 (.02)	-.06 (-.07)
t-10	.01 (.02)	-.07 (-.05)

Table 6: Lagged Dependent Variable (LDV) Models of Daily Media Tone, 2016

	With Conventions	Including Other Events
Media Tone$_{t-1}$	0.29	0.27
	(0.06)	(0.06)
Republican Convention$_{t-1}$	-0.21	-0.21
	(0.07)	(0.07)
Democratic Convention$_{t-1}$	0.18	0.18
	(0.07)	(0.07)
Comey in July[a] $_{t-1,\, t-2,\, t-3}$	---	-0.25
		(0.00)
The First Debate[a] $_{t-1,\, t-2,\, t-3}$	---	0.05
		(0.49)
The Sex Tape[a] $_{t-1,\, t-2,\, t-3}$	---	-0.04
		(0.59)
Comey in October[a] $_{t-1,\, t-2,\, t-3}$	---	0.02
		(0.05)
Constant	0.07	0.08
	(0.01)	(0.01)
R-squared	0.17	0.23
Adjusted *R*-squared	0.16	0.18
Root Mean Squared Error	0.13	0.13

Note: Number of cases = 271; the numbers in parentheses are standard errors.

[a] The numbers are the average of the coefficients for the three lagged variables and the test of the significance of their sum.

Table 7: Lagged Dependent Variable (LDV) Models of Pooled House-Adjusted Polls, 2016

	Election Year	Post-Labor Day
Poll Reading$_{t-1}$	0.77	0.91
	(0.03)	(0.05)
Net Media Tone$_{t-1}$	0.92	-0.27
	(0.39)	(0.38)
Constant	12.12	4.83
	(1.77)	(2.70)
R-squared	0.66	0.84
Adjusted *R*-squared	0.66	0.83
Root Mean Squared Error	0.96	0.40
Number of Cases	283	64

Note: The numbers in parentheses are standard errors.

Table 8. Cross-correlation Function between Daily Media Tone and Pooled House-Adjusted Polls, 2016

Lead/Lag of Tone	*Correlation*	
	Election Year	*Post-Labor Day*
t+10	-.04	.24
t+9	.01	.29
t+8	.01	.23
t+7	.03	.18
t+6	.07	.21
t+5	.04	.18
t+4	.05	.11
t+3	.12	.06
t+2	.10	-.01
t+1	.10	-.00
t	.11	-.04
t-1	.13	-.08
t-2	.21	-.05
t-3	.20	-.14
t-4	.21	-.12
t-5	.20	-.18
t-6	.17	-.16
t-7	.18	-.10
t-8	.15	-.19
t-9	.08	-.22
t-10	.04	-.16

Figure S1: Clinton and Trump Shares of Total Respondents by Date, 2016

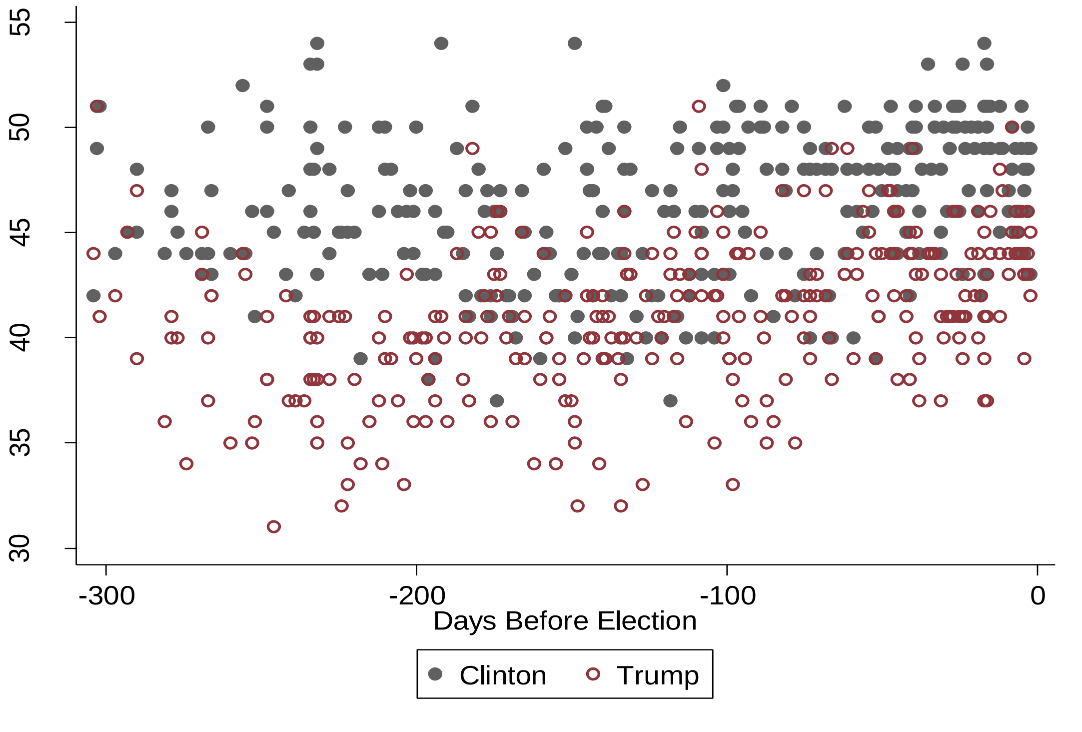


Figure S2: Candidate Shares of Total Respondents Aggregated by Date, 2016