

The Timeline of Elections: A Comparative Perspective

American Journal of Political Science, forthcoming.

Will Jennings
University of Southampton

Christopher Wlezien
University of Texas at Austin

Abstract:

Scholars are only beginning to understand the evolution of electoral sentiment over time. How do preferences come into focus over the electoral cycle in different countries? Do they evolve in patterned ways? Does the evolution vary across countries? This paper addresses these issues. We consider differences in political institutions and how they might impact voter preferences over the course of the election cycle. We then outline an empirical analysis relating support for parties or candidates in pre-election polls to their final vote. The analysis relies on over 26,000 vote intention polls in 45 countries since 1942, covering 312 discrete electoral cycles. Our results indicate that early polls contain substantial information about the final result but that they become increasingly informative over the election cycle. Although the degree to which this is true varies across countries in important and understandable ways given differences in political institutions, the pattern is strikingly general.

Acknowledgements: Previously presented at the 2013 Meeting of the Elections, Public Opinion and Parties group of the UK Political Studies Association, Lancaster, the 2014 Meeting of the Southern Political Science Association, New Orleans, the 2014 Meeting of the Midwest Political Science Association, Chicago, the 2014 Meeting of the American Political Science Association, Washington, D.C., the 2015 Meeting of the American Association for Public Opinion Research, Hollywood, Florida, and also at the University of Amsterdam, Ludwig Maximilian University of Munich, the University of Manchester, the University of Mannheim.. Thanks to Shaun Bevan, Thomas Brauningner, Wouter van der Brug, Florian Foos, Andrew Gelman, Ben Goodrich, Armen Hahkverdian, Chris Hanretty, Yphtach Lelkes, Mike Lewis-Beck, Matt Loftis, Michael Marsh, Tom van der Meer, Julia Partheymuller, Mark Pickup, Sebastian Popa, Rudiger Schmitt-Beck, Mary Stegmaier, Patrick Sturgis and especially Stephen Jessee for comments on earlier versions of this paper. Thanks also to AJPS editor Bill Jacoby and the anonymous reviewers for various other suggestions. For poll data, special thanks to Chris Anderson, Joe Bafumi, Graziella Castro, Jan Drijvers (TNS-Global), Robert Erikson, Johanna Laurin Gulled (Ipsos Public Affairs), Chris Hanretty, Michael Marsh, Michael Thrasher, Jack Vowles, and Francisco José Veiga,. Thanks also to the Roper Center for Public Opinion Research, Australian Social Science Data Archive, Norwegian Social Science Data Services, Data Archiving and Networked Services in the Netherlands, Canadian Opinion Research Archive, and GESIS/Leibniz Institute for the Social Sciences. We are indebted to Yulia Marchenko from StataCorp for assistance with programming of multiple imputation, and David Baker from High Performance Computing at the University of Southampton.

The study of voters and elections has shed considerable light on people's vote choices and election outcomes (see, for example, Abramson et al 2007 Campbell et al 1960; Budge and Farlie 1983; van der Eijk and Franklin 1996; Clarke, et al 2004; for reviews, see Jacoby 2010; Heath 2010; Dalton and Klingemann 2007). Yet electoral scholars are only beginning to understand the evolution of electoral sentiment over time. How does the outcome come into focus over the electoral cycle? Do voters' preferences evolve in a patterned way?

Answers to these questions are interesting unto themselves, but also important. The structure and dynamics of electoral preferences matter in representative democracies. If preferences are highly structured and in place early on, voters are less subject to influence during the official election campaign. By contrast, if preferences are not highly structured and emerge only late, voters may be influenced by everything that happens between elections. These characterizations have roots in research on elections and voting, some of which reveals “minimal effects” of events (e.g., Berelson et al 1954; Lewis-Beck and Rice 1992; Finkel 1993) and others of which demonstrates substantial, lasting effects (e.g., Johnston et al 1992; Lodge et al 1995; Shaw 1999; Hillygus and Shields 2008).¹ Of course, it may be that both are at work, and that some effects last and others decay. Understanding the crystallization of preferences over time thus can reveal when (and how) cleavages come to matter on Election Day.

¹ An alternative view suggests that electoral competition moves voters towards the equilibrium set by the fundamentals (Gelman and King 1993; Campbell 2000; Stevenson and Vavreck 2000; Arceneaux 2005). Even from this perspective, campaigns still matter (especially see Vavreck 2009).

In this paper, we consider whether and how political institutions structure the evolution of electoral preferences in systematic ways. Institutions matter for electoral choice; they also should matter for the predictability of electoral choice and the evolution of voters' preferences. This is the subject of our investigation. First, we examine government institutions, focusing on differences between presidential and parliamentary systems. Second, we examine electoral institutions, particularly differences between proportional representation and single-member district plurality systems.

To begin with, we review the previous research and consider how differences in political institutions might impact voter preferences over the election cycle. We then outline an empirical analysis relating support for political parties or candidates in pre-election polls to their final Election Day vote. Finally, we undertake analysis of more than 26,000 polls from 312 electoral cycles in 45 countries.² The results reveal a general pattern: polls become increasingly informative about the vote over the election cycle but very early polls contain substantial information about the final result. They also reveal significant variation: the evolution of preferences differs across countries, reflecting differences in political institutions.

Polls and the Vote over the Election Timeline

Consider the timeline of elections (following Erikson and Wlezien 2012; also see Wlezien and Erikson 2002). We start the timeline immediately after the previous election. We end it on Election Day. Many events occur over the timeline, some very prominent and others routine.

² Full replication data and syntax are available from the American Journal of Political Science Dataverse: <https://thedata.harvard.edu/dvn/dv/ajps>.

We want to know whether these events have effects. We also want to know whether the effects last.

There is evidence of lasting influences from pre-election polls in US presidential elections (Erikson and Wlezien 2012). In these “trial heat” polls, survey organizations typically ask respondents which candidate they would vote for “if the election were held today.” The results of these polls tell us only a little about the persistence of specific events, as it is difficult to even identify their effects.³ The polls can tell us quite a lot about general patterns, however, as we can assess how poll results at different points of the election cycle match the final results. If polls are increasingly informative across the timeline, then we know that electoral preferences change and the some of the change lasts to impact the outcome. If polls are equally well informative across the timeline, then either (1) preferences do not change or else (2) preferences do change but these innovations do not persist, i.e., the fundamentals remain the same.⁴

Scholars have found that, at the beginning of the election year in the US, some 300 days before

³ It is difficult to characterize the effects of events for at least three reasons. First, the effects of most events are small, with exceptions such as party nominating conventions (e.g., Holbrook 1996; Shaw 1999; Erikson and Wlezien 2012) and possibly candidate debates (e.g., Johnston et al 1992; Holbrook 1996; Shaw 1999; Blais et al 2003). Second, survey error makes the effects of events hard to detect (see Wlezien and Erikson 2001; Zaller 2002). Third, the net effects of different campaign activities can cancel out. For additional details, see Erikson and Wlezien (2012).

⁴ Where the latter is true, we may see a late uptick in the correspondence between polls and the vote owing to short-term effects that arrive late and do not fully decay before Election Day.

the balloting, there is virtually no relationship between the results of presidential polls and the actual vote. At the end of the cycle, by contrast, poll results virtually match the final result. In between, polls become more and more accurate. This is revealing about voter preferences. It tells us that they change over the election year and in meaningful ways: although much of the change that we observe is short-lived, and dissipates before Election Day, a *substantial portion carries forward to impact the final outcome*.⁵

Polls and the vote in US Congressional elections exhibit a similar pattern (Bafumi et al 2010). In the “generic” ballot, survey organizations ask respondents which party’s candidate for Congress they would vote for in in their district. These measured preferences are more informative early in the election year. They also are less informative towards the end of the election cycle. Polls for parliamentary elections in the UK are informative earlier still (Wlezien et al 2013), starting to come into focus years before Election Day.

This research offers certain lessons. First, polls in each case become more revealing about the outcome the closer the election. This is as we might expect, but it indicates that electoral preferences evolve over time in both countries. Second, there is a hint of difference between presidential and legislative elections, where polls for the former are less predictive of the vote early in the election cycle and more predictive at the end of the timeline. This also as we might expect given that presidential polls capture preferences for the two candidates and the parliamentary and congressional polls tap party support in the various legislative districts. Third, it may be that early polls are more informative about UK parliamentary election outcomes than

⁵ Voters are, at least to some extent, “online processors,” updating their preferences based on new information about the parties and candidates (see Lodge et al 1995).

legislative elections in the US. This is suggestive about the effects of government institutions—that electoral preferences develop earlier in parliamentary systems. Whether the patterns are real and hold in other countries remains to be seen.

A Comparative Perspective on Polls and the Vote

Countries differ in many ways, and there is reason to expect that some of the differences—particularly in relation to political institutions—matter for electoral preferences (also see Shugart and Carey 1992). We consider both government and electoral institutions. They structure the set of choices faced by voters and this can influence the formation and stability of preferences across the timeline.

Government Institutions

The government structure can influence how and when electoral preferences form. Of special importance are the differences between presidential and parliamentary systems. There are two main differences: (1) between presidential and parliamentary elections and (2) between legislative elections in presidential and parliamentary systems.

First, in presidential elections voters select an *individual* to represent the country whereas in parliamentary elections they select a *legislature*, which in turn produces a government. In presidential elections there often is greater uncertainty over the identity of the candidates. Even to the extent the candidates are known, information about their attributes and policies often are not known until later still in the election cycle. By contrast, in parliamentary systems parties tend to dominate (Budge and Farlie 1983; Adams et al 2005). This is important because dispositions towards parties, while not fixed, are more durable than those toward candidates.

Even to the extent party leaders are important to voters in parliamentary systems (e.g. Clarke et al 2004), their identities typically are known well in advance, earlier than presidential candidates. As a result, we expect that voters' preferences crystallize earlier in parliamentary elections.⁶

Second, there also is reason to expect that the difference between *legislative elections* in parliamentary and presidential systems influences the formation of preferences. There actually are at least two reasons. To begin with, as discussed above, politics in parliamentary systems centers on parties even where single member districts are used (also see Cox 1997; Mair 1997). To the extent party dispositions matter more in parliamentary systems, preferences in legislative elections should harden earlier by comparison with preferences in presidential systems.⁷ In addition, legislative elections in presidential systems can be influenced by presidential elections themselves, in the form of coattails (Ferejohn and Calvert 1984; Golder 2006). As a result, electoral preferences in such systems may coalesce later in the election cycle.

⁶ That this would produce differences seems obvious when choosing among parties in proportional systems, but there is reason to think it will be true even where single member districts are used, as candidate-specific considerations are likely to cancel out when aggregating across districts.

⁷ Then again, elections are harder to anticipate in parliamentary systems, as most governments are able to select the time of the next election, so we might expect more—not less—change in these systems. This is implied by Stevenson and Vavreck's (2000) research. There is an alternative logic regarding the effects of government control of election timing as well. We consider these possibilities below.

Electoral Institutions

Electoral systems also may have consequences for how preferences evolve. There are two main expressions: (1) the general difference between proportional representation and single member districts, and (2) the specific differences in the number of political parties.

First, we might hypothesize that there is less change in preferences in systems where proportional representation is used as opposed to single member districts (SMDs). The intuition is fairly straightforward: voters choose among parties and not candidates, and so preferences may tend to be more stable over the election timeline. Of course this depends on the level of voter alignments with parties (van der Brug et al 2007).⁸ The basis for party alignments also is important (Clarke et al 2004). The stability of party coalitions is as well, since change in the coalition (or in the expected governing coalition) can cause voter preferences to change (Strøm 1997). Just as there are elements of instability in proportional systems, there are elements of stability in candidate-centered ones: parties matter there too and incumbency as well, and this can limit the effects of events (Abramowitz 1991; Cain et al 1984).⁹ It thus is not clear that proportionality should produce more stability. Further, some proportional systems have candidate-centric aspects to their electoral formula—the degree to which candidates have incentives to cultivate a personal vote (Carey and Shugart 1995). The degree of candidate- or

⁸ Where party alignments are weak we expect more “undecideds,” later decision-making, and greater susceptibility to campaign effects (see Fournier et al 2004).

⁹ Note that the “personal vote” also can matter in proportional systems that use open lists (see Shugart et al 2005), and the causal effect of incumbency is debatable (see Zaller 1998; Fenno 1978).

party-centricity actually might better account for variation across countries than the categorical distinction between proportional and SMD systems per se, which we consider in our analyses.

Second, we might even propose that proportionality actually leads to greater volatility in preferences. It is well known that electoral systems produce party systems, and the number of parties is one consequence (and indicator) of proportional election rules (Duverger 1954; also see Cox 1997). This may have implications for the evolution of electoral preferences. In systems where there are fewer parties, voters face less choice, and this may make preferences more stable and also more predictive of the final vote. That is, the propensity of individual voters to transfer their vote between parties is reduced as a function of the number of options available to them and the perceived average distance between the parties (Pedersen 1979). In multi-party systems, there is more choice and opportunity for change, and so the Election Day vote may be less clear in the polls until very late in the electoral cycle.

Data

Pollsters have sought to measure citizen's preferences for candidates or parties for almost three quarters of a century. While varying due to differences in context, most pre-election polls ask how citizens would vote "if the election were held today." We have compiled what we believe is the most extensive comparative dataset ever assembled of national polls of the vote intentions for presidential and legislative elections.¹⁰ The dataset consists of a total of 26,917 polls spanning

¹⁰ In every poll in our dataset respondents were asked for which candidate or party they would vote; we ignore cross-national and within-country differences in question wording. Lau (1994) shows that in the US such differences matter little for poll results, McDermott and Frankovic

the period from 1942 to 2013. The data cover a total of 312 elections (including 22 run-off elections) in 45 countries, 13 of which are pure presidential systems, 28 of which are parliamentary systems, and 4 of which are mixed, including a president and a parliament. All told, we have poll data for presidential elections in 23 countries and legislative elections in 31 countries, summarized in Table 1. Further details are provided in supplementary Appendix S1. On average we have 598 polls per country for approximately 7 elections per country, or about 86 polls per election cycle. Given the typical interval—1,195 days—between elections, we are missing polls on most dates and in many weeks and even months.

(2003) demonstrate that some are consequential. To the extent wording does matter, it serves to introduce error into our measure of electoral preferences.

Table 1. Poll Data in 45 Countries, 1942-2013

Country	System	Election	Rule	First poll	Last election
Australia	Parliamentary	Legislative (1 st Pref)	SMDP	1943	2013
		Legislative (2 nd Pref)*	SMDP	1993	2013
Belgium	Parliamentary	Legislative	PR	2004	2010
Bulgaria	Parliamentary	Legislative	PR	2009	2013
Canada	Parliamentary	Legislative	SMDP	1942	2011
Croatia	Parliamentary	Legislative	PR	2008	2011
		Presidential	Majority	2009	2010
Czech Republic	Parliamentary	Presidential	Majority	2012	2013
Denmark	Parliamentary	Legislative	PR	1960	2011
Finland	Parliamentary	Legislative	PR	2010	2011
Finland	Parliamentary	Presidential	Majority	2006	2012
Germany	Parliamentary	Legislative	PR	1961	2013
Greece	Parliamentary	Legislative	PR	2007	2012
Hungary	Parliamentary	Legislative	PR	2009	2010
Iceland	Parliamentary	Legislative	PR	2009	2012
		Presidential	Plurality	2012	2012
Ireland	Parliamentary	Legislative	PR	1974	2011
Italy	Parliamentary	Legislative	PR	2012	2013
Japan	Parliamentary	Legislative	PR	1998	2012
Malta	Parliamentary	Legislative	SMDP	2012	2013
Netherlands	Parliamentary	Legislative	PR	1964	2012
New Zealand	Parliamentary	Legislative	SMDP/PR	1975	2013
Norway	Parliamentary	Legislative	PR	1964	2013
Poland	Parliamentary	Legislative	PR	2010	2011
		Presidential	Majority	2011	2011
Serbia	Parliamentary	Legislative	PR	2008	2012
Slovakia	Parliamentary	Legislative	PR	2010	2012
Slovenia	Parliamentary	Presidential	Majority	2012	2012
Spain	Parliamentary	Legislative	PR	1980	2011

Sweden	Parliamentary	Legislative	PR	2000	2010
Switzerland	Parliamentary	Legislative	PR	2010	2011
Turkey	Parliamentary	Legislative	PR	2010	2011
U.K.	Parliamentary	Legislative	SMDP	1943	2010
Argentina	Presidential	Presidential	Majority	2006	2011
Brazil	Presidential	Presidential	Majority	2002	2010
Chile	Presidential	Presidential	Majority	2008	2010
Colombia	Presidential	Presidential	Majority	2010	2010
Cyprus	Presidential	Presidential	Majority	2007	2013
Ecuador	Presidential	Presidential	Majority	2010	2013
Mexico	Presidential	Presidential	Plurality	2005	2012
Paraguay	Presidential	Presidential	Plurality	2013	2013
Peru	Presidential	Presidential	Majority	2006	2011
Philippines	Presidential	Presidential	Plurality	2010	2010
South Korea	Presidential	Legislative	PR	2011	2012
		Presidential	Plurality	2012	2012
U.S.	Presidential	Legislative	SMDP	1942	2012
		Presidential	Electoral College	1952	2012
Venezuela	Presidential	Presidential	Plurality	2006	2013
Austria	Semi- Presidential	Legislative	PR	2006	2013
		Presidential	Majority	2010	2010
France	Semi- Presidential	Presidential	Majority	1965	2012
Portugal	Semi- Presidential	Legislative	PR	1985	2011
		Presidential	Majority	2010	2011
Romania	Semi- Presidential	Legislative	PR	2008	2012
		Presidential	Majority	2009	2009

* Polls of two-party preferences under Australia's transferable vote electoral system are excluded from analysis to avoid double-counting.

Methodology

Recall that we are interested in whether and how electoral preferences evolve over the timeline. For this, we need a method for assessing the correspondence between polls at different points in time and the eventual vote. Ideally we would use time series analysis. That is, we would examine the relationship between polls at different points in time within the various election years taken separately or pooled together. In theory, this would tell us exactly what we want—whether changes to preferences decay or last. In practice, it is not so straightforward because of data limitations. There are two main problems. First, as we have seen, data frequently are missing for weekly periods and even months. This has fairly obvious implications for what we can do with standard time series techniques. Second, the ratio of sampling variance to the true variance often is quite large. This has substantial, if less obvious, complications: the presence of sampling (and other survey) error makes it difficult to uncover the underlying process.

What can we do instead? Our approach is to follow Erikson and Wlezien (2012) and treat the data as a series of cross-sections—across elections—for each day of the election cycle. With the data organized as a series of cross sections, we can assess how polls and the vote across elections match up at different points in time. Sampling error is not a problem for such an exercise; whereas error may swamp the variance from true change when observing within-election polls, it is dwarfed by election-to-election (and party-to-party) differences in the cross-section.¹¹

¹¹ Consider that, when measured across presidential elections in the US between 1942 and 2008, the variance in the vote exceeds the error variance by a factor of 50 or more. For instance, when the vote is measured as 30-day cross-sections, the minimum of the estimated reliabilities is 0.98, i.e., virtually all of the difference *across elections* is real.

However, the problem that pre-election polls are sometimes sparse and conducted at irregular intervals remains.

When readings of electoral preferences are missing, we can interpolate daily voter preferences from available polls. For any date without a poll, an estimate is created as the weighted average from the most recent date of polling and the next date of polling. Weights are in proportion to the closeness of the surrounding earlier or later poll.¹² Where we interpolate, we also introduce a random component based on the poll variance -- controlling for country, party and election -- to reflect the uncertainty associated with the imputed values. We thus are able include in our analysis any election cycle from the moment the first poll is conducted in that cycle. This would not be acceptable in conventional time series analysis, as interpolating would compromise the independence of observations. Given that the methodology is explicitly cross-sectional, there is no such problem—interpolating actually permits a more fine-grained analysis. The main drawback of the approach is that we cannot assess whether dynamics differ across particular

¹² Specifically, given poll readings on days $t - x$ and $t + y$, the estimate for a particular day t is generated using the following formula:

$$\hat{V}_t = \{ [y * V_{t-x} + x * V_{t+y}] / (x + y) \} + \varepsilon,$$

where ε , is drawn from the following distribution: $\mu=0, \sigma=3.394$. Recall that for days in the timeline after the final poll before an election, we carry forward the numbers from the final poll. This has some consequence for the accuracy of poll predictions very close to Election Day, as we use polls from well before the end of the cycle in some cases.

elections. Importantly, the method allows us to assess patterns of correspondence across elections in different subsets of elections, e.g., across types of systems.¹³

For our analysis, we also need data on the actual vote shares parties and candidates receive on Election Day. We rely on a wide range of official sources and election data resources—details are reported in supplementary Appendix S2.

To analyze the relationship between polls and the vote over the election timeline, we estimate a series of daily equations predicting the vote share for different parties or candidates (j) in different elections (k) across countries (m) from vote intentions in the polls on each day of the timeline:

$$VOTE_{jkm} = a_{jmT} + b_T Poll_{jkmT} + e_{jkmT}, \quad (1)$$

where T designates the number of days before Election Day and a_{jmT} represents a separate intercept for each party or candidate j in country m . This is important because the level of electoral support varies systematically across parties. Now, suppose that our timeline covers the year before Election Day. We would estimate an equation using polls from 365 days before each election, and then do the same using polls from 364 days in advance, and so on up to Election

¹³ Some might think that we should analyze individual-level data. The problem is that we have such data for only a fraction of polls; we have self-reported vote choice against which to compare prior vote intention in fewer cases still, and these are clustered at the very end of the election cycle. As much as we would like to match individual-level preferences registered over the campaign timeline with their particular vote choices later in numerous election years in numerous countries, it just is not possible given the existing data.

Day itself. Multiple imputation (Rubin 1987) is used, which averages the coefficients across the imputed data series and adjusts the standard errors to reflect noise both due to imputation and residual variance.¹⁴ Using the resulting estimates, we can see whether and how preferences come into focus over time.

To begin with, we are interested in the variance explained by the polls (*R*-squared) over the timeline. This tells us how well the polls predict the vote at each point in time. If the *R*-squared increases over time, we know that polls become better predictors the closer the election. The improvement in predictability will reflect the variance of the shocks and the proportion that persists, bearing in mind that some changes may not last. An increase in *R*-squareds would not necessarily mean that the polls themselves—instead of poll-based predictions—are increasingly accurate. For this, we need to examine the regression coefficient (b_t) relating the polls and the vote. This tells us what proportion of the poll margin on each day carries forward to Election Day. If the coefficient increasingly approaches 1.0 as the *R*-squared increases over time, we know that the polls converge on the vote.

It is important to recognize that polls can be increasingly predictive of the vote—and produce larger *R*-squareds—while the coefficient relating the polls and the vote remains unchanged. This suggests that the *R*-squareds may be more relevant, as they reveal how predictable outcomes are.

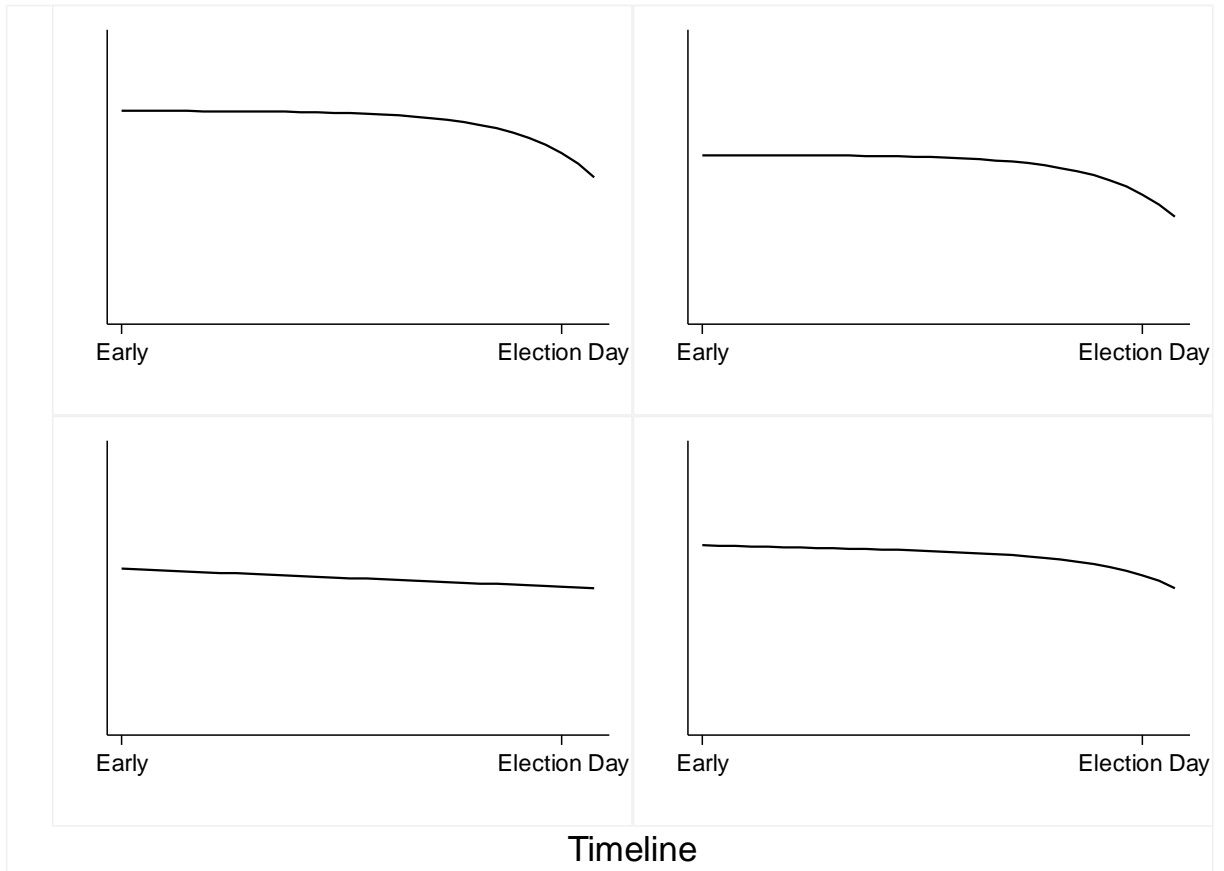
¹⁴ Rubin (1987) shows that where γ is the rate of missing data, estimates based on m imputations have an efficiency that approximates to a value of $(1 + \frac{\gamma}{m})^{-1}$. Since polls are missing on around 92% of days we use 50 imputed data series, which implies a relative efficiency of 0.98 compared to an infinite number of imputations.

The problem is that these are not very useful across different data, as they are standardized based on the total observed variance in the particular sample. Change the countries and we change the variance and so also change the R -squareds, and that change may tell us absolutely nothing about the effects of polls. We thus turn to unstandardized measures, specifically, the root mean squared error (RMSE). The RMSE indicates how much of the actual vote variance is unexplained, for instance, 2.5 percentage points for one set of countries by comparison with 1.5 points for another, and so is a particularly useful measure when forecasting, which in effect is what we are doing here.¹⁵

Let us consider some simple examples of what we might observe, focusing on the RMSE. Clearly, if preferences evolve at all, the RMSE would go down over time. That is, the polls would increasingly predict the vote. The exact functional form would depend on how much of the variance is due to long-term and short-term components, however. Figure 1 displays four possible models.

¹⁵ As the R -squared also is informative (Krueger and Lewis-Beck 2007), it is worth noting that those estimates and the RMSEs *always* are negatively correlated at 0.99 or higher for all of the analyses that follow (with the sole exception of the estimates for elections over 900 days, where the correlation is still a substantial 0.93). This indicates that when the RMSE is lower, the R -squared is, almost without exception, higher.

Figure 1. Different Functional Forms of the Evolution of Electoral Preferences



The panel in the upper-left corner presents the path of an RMSE from a baseline model where the vote is a function only of short-term factors, with all elections in different years sharing one common equilibrium value (V^*). In time series language, polls for all parties and elections evolve as a stationary series. Here, earlier polls tell us little about the Election Day vote. This is because any difference between early polls and the common mean would decay by Election Day, leaving V^* as the expectation in each year. Late polls would provide information, however, as they reflect effects that do not fully decay by Election Day. The result in terms of the RMSE is the nonlinear temporal pattern shown in the panel, where the RMSE starts at some high value and declines near the end of the cycle.

The assumption of a common equilibrium may not be very realistic, as it implies that elections are pretty much the same. We relax the assumption in the upper-right frame of Figure 1. Here the vote is a function of both short-term factors and long-term factors, where the equilibrium term (V^*_Y) varies by year Y but does not change over time. From a time series perspective, this represents a set of stationary series varying around different yearly means. (We also could introduce different means for parties and candidates (j), which matter much more, as we will see.) In this representation, earlier polls tell us more about the Election Day vote because of differences across elections (and parties). Later polls provide additional information because of late-arriving campaign effects. The result in terms of RMSE is much like what we saw in the upper-left panel, but consistently lower. This characterization is implicit in many election prediction models, e.g., Gelman and King (1993).

It may be that equilibria are not constant over each election cycle. The lower-left panel of Figure 1 reflects this possibility, where the equilibria evolve over the timeline. In this model, all shocks accumulate and so the time series evolves as a “random walk” and polls become consistently more informative over time. Of course, it may be that there are both short-term and long-term components, where the model combines a stationary series and a random walk. Here, polls would become more predictive but at an increasing rate as the election approaches, owing to late-arriving short-term forces. The expected pattern is depicted in the lower-right frame of Figure 1.¹⁶

¹⁶ Note that our expository analysis here assumes that shocks to preferences are fairly similar across the election timeline. If their impact is uneven at different points, the pattern of RMSEs would differ accordingly.

Which, if any, of these models should we expect as an accurate portrayal of the RMSE? That is an empirical question, of course. To begin with, we note that there are substantial differences in the average level of electoral support across parties and, to a lesser extent, elections. Consider analysis of variance (ANOVA) results, which indicate that 86% of the variance in the party vote in different countries and elections is due to systematic party differences. In actuality, since party variables are specific to each country, some of this party-explained variance reflects differences across countries; specifically, countries account for 37% and parties account for the other 49%. The remaining (14%) of the variance reflects differences across elections and time. ANOVA of daily polls indicates that an even smaller percentage is due to changes in preferences over the election timeline. That is, most of the variance in the polls is due to differences across parties, countries and elections. This leads us toward the models in the right-hand column of the Figure 1, where there are differences in equilibrium support. The question that remains then is whether the changes in voter preferences that we observe decay or last. If some substantial portion lasts, we would expect a pattern like what we see in the lower-right corner of the figure.

It may be that electoral preferences evolve over the election cycle in a similar way across political systems. It also may be that the pattern differs. Consider our discussion of political institutions. There we posited differences between presidential and parliamentary elections, where electoral preferences come into focus more quickly in the latter. In terms of Figure 1, we expect that the RMSEs would be consistently lower in parliamentary systems until the very end of the timeline, when presidential elections come into focus. We also posited differences between legislative elections in presidential and parliamentary systems, which might show a similar, if less pronounced pattern. Finally, we posited possible, contrasting differences between legislative elections in proportional and SMD systems. First, to the extent the former mostly

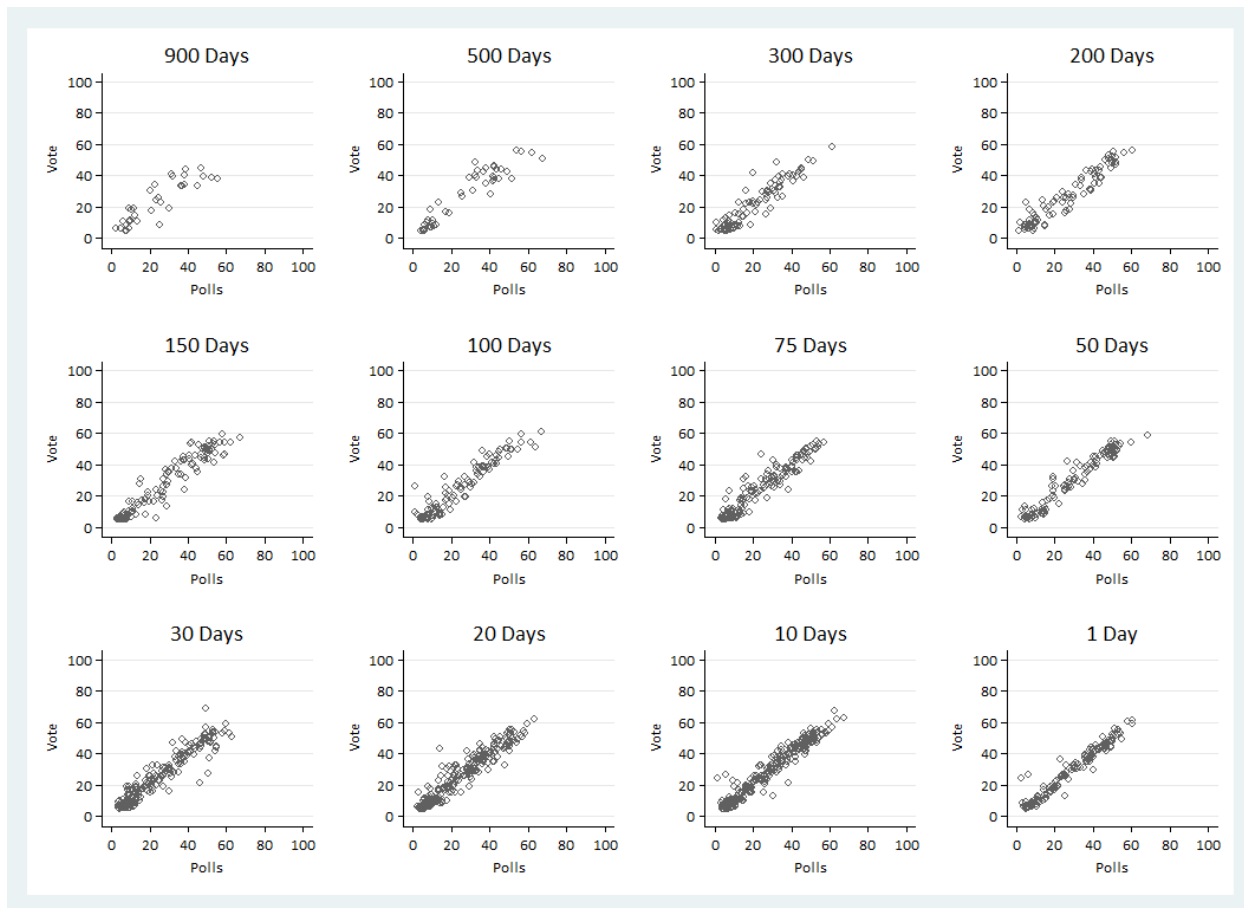
lead voters to place a greater emphasis on their party preferences we would expect electoral preferences in proportional systems to be more predictive of the vote throughout the election cycle, at least until the very end of the campaign. Second, to the extent these systems mostly provide voters with choice leading up to Election Day, we would expect preferences to be less predictive throughout, potentially right up to the very end.¹⁷ These are our hypotheses. Now let us see what the data reveal.

Results

To begin the analysis, let us consider the scatterplot between polls and the vote at various points of the election cycle. This is shown in Figure 2. The figure displays the poll share for all parties or candidates in all elections and countries for which we have actual polls, i.e., excluding imputed polls numbers. In the upper left-hand panel of the figure, using polls that are available 900 days before the election, fully two and a half years before an election, we see that there already is a discernible pattern. That is, the poll share and the vote share are positively related, though there also is a good amount of variation. At that point in time, we have polls in the field in approximately 40% of our cases, and this increases fairly steadily, reaching 75% one year before Election Day. As we turn to polls later in the election cycle, moving horizontally and then vertically through the figure, a clearer pattern emerges; the poll share and final vote share line up. Simply, as we get closer to the election, the polls tell us more about the outcome. It is as one would expect if preferences change and a nontrivial portion lasts. But how much do preferences evolve?

¹⁷ Of course, it might be the case that both effects are present but tend to cancel out, in which case the likelihood of observing effects owing to the electoral system would be reduced.

Figure 2. Scatterplot of Party Vote Share by Party Poll Share for Selected Days of the Election Cycle—Pooling all Elections



In Figure 3, we make a more fine-grained presentation. Here we display the cross-sectional RMSE from regressing the vote division on the poll division for each date starting 900 days before the election, now including the imputed data. (The number of cases on each day is shown in Appendix Figure A1.) To calculate the RMSE, one estimates the prediction errors from the regression, squares them, calculates the mean of those squared errors, and then takes the square root of the mean. The *R*-squareds and coefficients from the regressions are provided in supplementary Appendix S3, Figures S3.1 and S3.2.

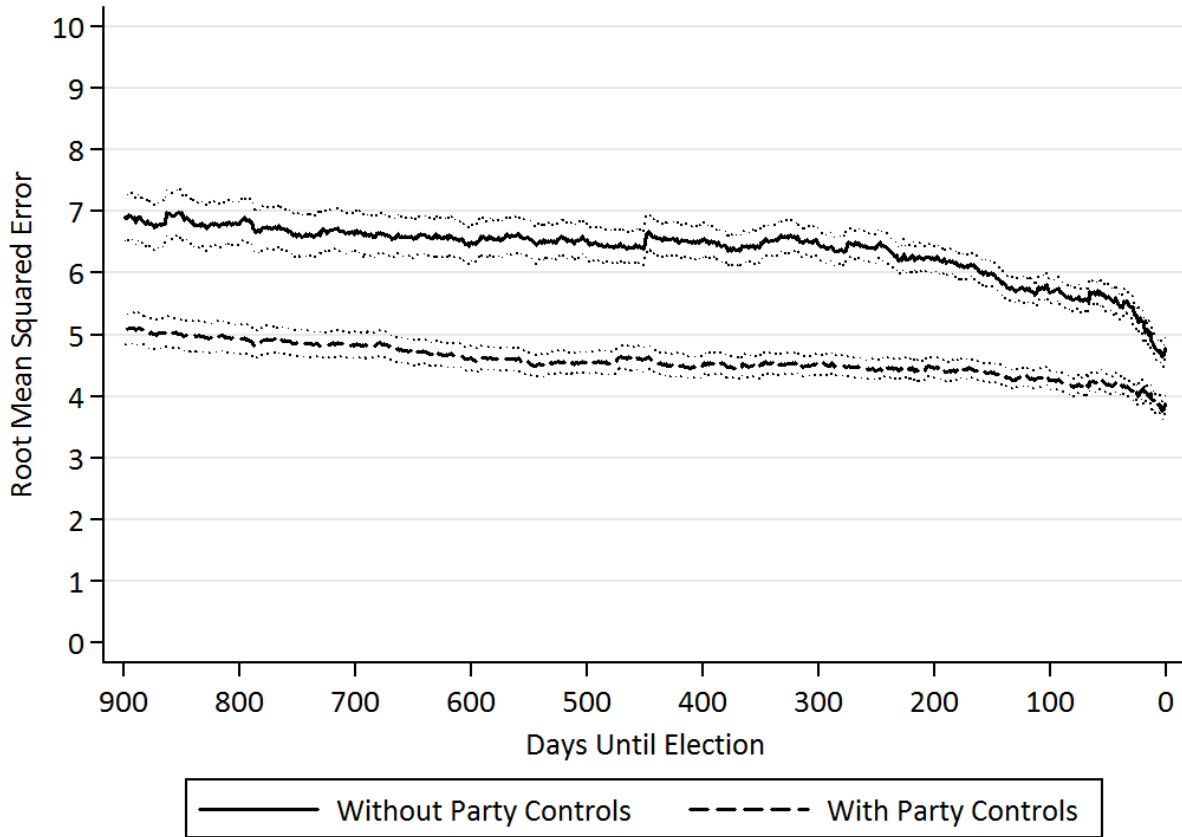
To begin with, we estimate equation 1 for all elections pooled together. This includes controls for different parties in different countries— which effectively accounts for differences across both countries and parties, recalling that party variables are country-specific. To show what effect the party controls have, we also estimate the equation without those controls. The regression equation is bootstrapped to estimate the standard errors of the RMSEs, enabling us to determine whether the relationship between polls and the vote differs significantly across institutional settings.¹⁸

The series of RMSEs in Figure 3 reveals that polls predict the final vote fairly well far in advance of an election. Consider first the results without party controls. Fully two and a half years out, the RMSE is about 7 percentage points. In terms of *R*-squareds, the poll share accounts for over 80% of the variance in the party/candidate vote share on Election Day. This early structure may surprise but is very much to be expected given the large cross-party variance in the poll and vote shares noted earlier. That is, shares of both the polls and the vote tend to be consistently higher for certain parties and consistently lower for others. Much as we could see in Figure 2, the RMSE decreases quite steadily over the election timeline, to about 5 points on

¹⁸ In bootstrapping the regression, we assume that our sample distribution (a total of 873,001 party/candidate*poll *days*) is representative of the general population of polls of vote intentions. This is not an unreasonable assumption, as our data set likely contains the majority of available polls. To bootstrap the estimates, the regression is estimated for randomly drawn resamples (with replacement) of the data repeated 1,000 times for each day of the election cycle. The model is estimated as a linear regression with one categorical factor that allows the effects of party controls to be absorbed.

Election Day. This tells us that preferences evolve, if only very slowly. The pattern almost perfectly matches the depiction in the lower-right hand panel of Figure 1.

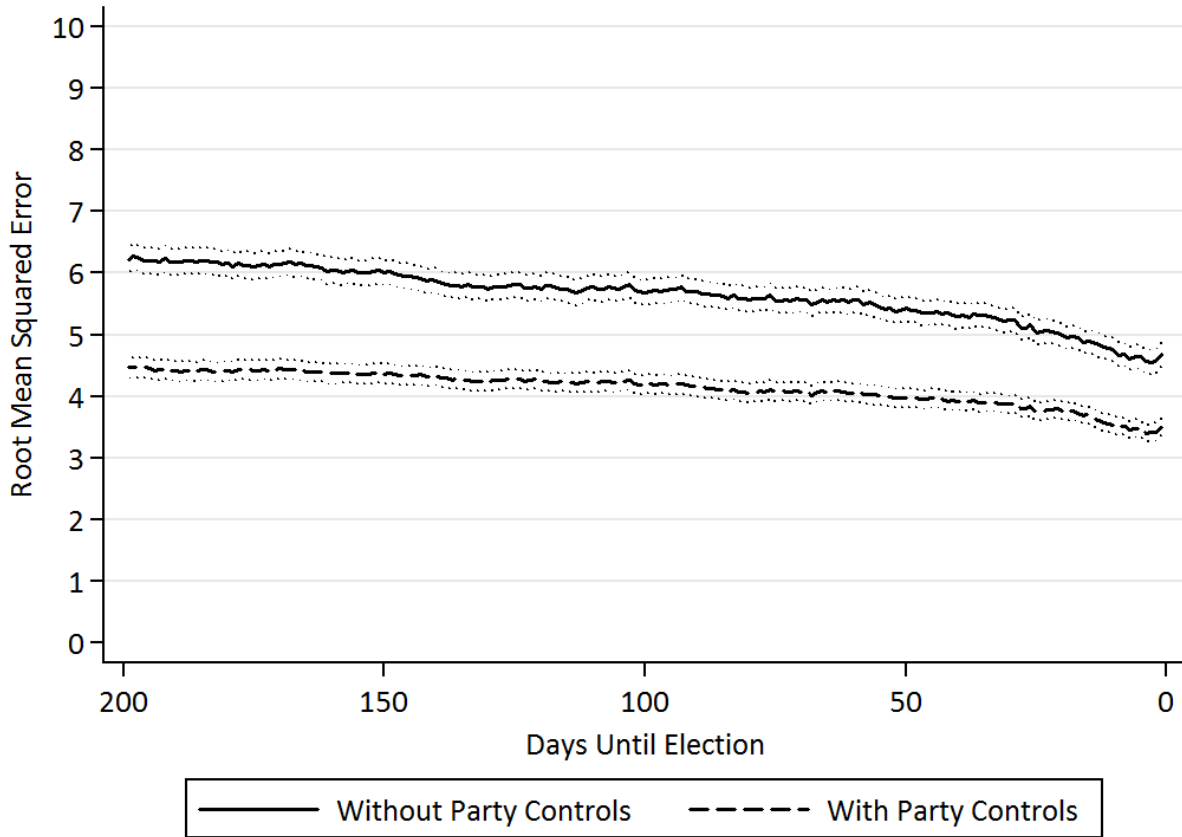
Figure 3. Root Mean Squared Errors (RMSEs) Predicting the Party Vote Share from the Poll Share, By Date in Election Cycle—Pooling all Elections



Results in Figure 3 from regressions including dummy variables for the different parties in different countries reveal a similar pattern. Not surprisingly, the RMSEs are smaller throughout the 900-day timeline, as the party controls help predict vote shares. The trend over the timeline is essentially the same, however. This indicates once again that preferences come into increasing

focus over time. They come into focus over the “long campaign” between elections but also during the short, official campaign in the weeks leading up to the election.

Figure 4. Root Mean Squared Errors for the Last 200 Days of the Election Cycle—Pooling all Elections with Continuous Poll Readings



The results in Figure 3 are based on a growing number of elections. At the beginning of the final year of the election cycle, for instance, the number of parties and candidates in our equation is about 750; by Election Day, the number is over 1,000. The fact that the cases (and the dependent variable) change can impact the trend in our results. We thus need to estimate the daily equation for the same set of cases. Figure 4 shows the resulting RMSEs from regressions for the 252

elections—and a total of 899 parties/candidates—for which we have poll readings for the final 200 days of the election cycle. These results are almost identical to what we observed in Figure 3; indeed, the correlation between the series is 0.98 ($p < 0.001$). The patterns we have observed thus are quite real. Electoral preferences demonstrate early structure but also evolve.¹⁹

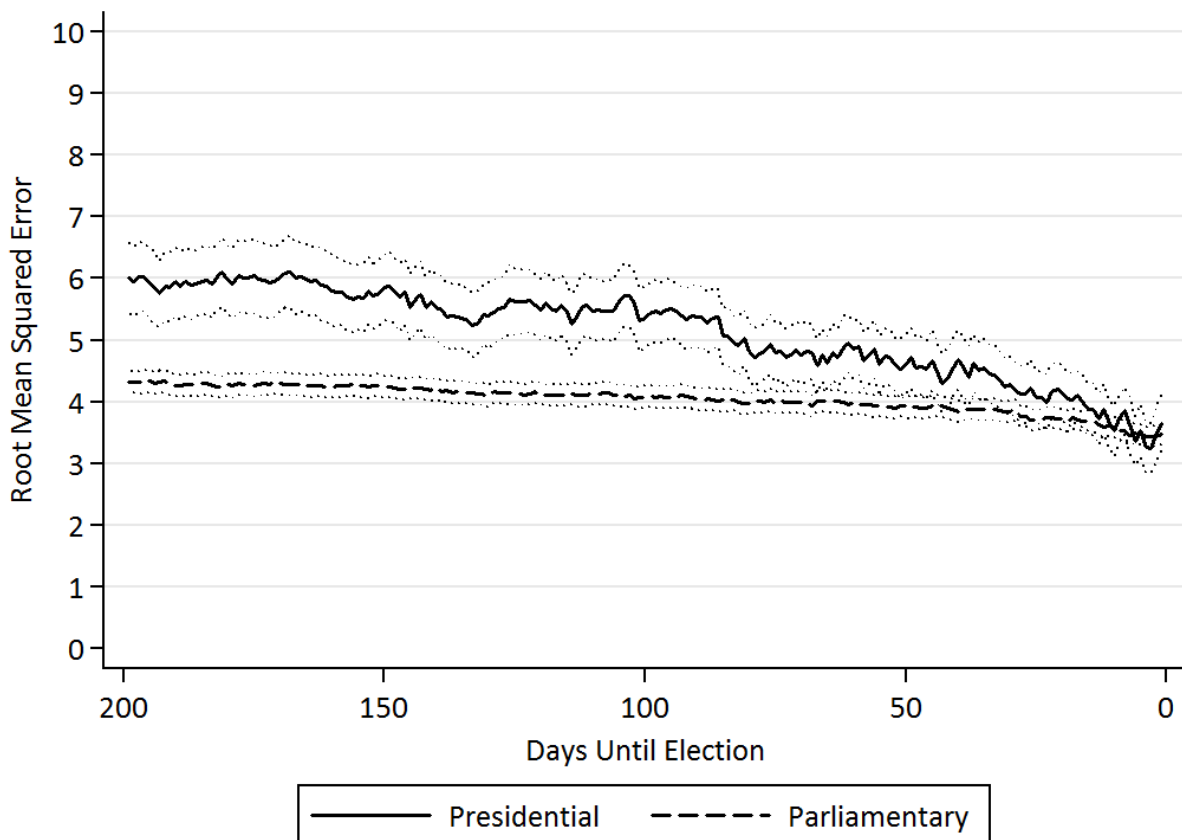
Government Institutions

Thus far our analysis has examined all elections pooled in a single analysis, regardless of institutional context. We have hypothesized that institutions matter, however. Specifically, we posited that political institutions structure the formation of preferences in a number of possible ways. Let us first consider the effects of government structure. Recall that we expect that voters' preferences crystallize earlier in the electoral cycle in parliamentary elections compared to presidential races. To test the hypothesis, we estimate separate equations relating poll and vote shares in the two types of elections. Figure 5 plots the resulting RMSEs over the final 200 days of the election cycle, based on models including party controls. The patterns in the figure are consistent with our expectations. At the beginning of the timeline, 200 days out, polls are much more informative in parliamentary elections, with an RMSE of 4 percentage points by comparison with 6 points for presidential elections. As can be seen in the figure, the difference

¹⁹ The regression coefficient (b) from the equation (1) relating polls and the vote offers additional, supporting information. The coefficient grows over the timeline as the RMSE shrinks. In regressions including party controls, the estimate is 0.6 using polls from 200 days in advance and increases to 0.8 using polls from the very end of the cycle. This tells us that an increasing portion of the polls lasts to impact the Election Day vote, i.e., the polls converge on the final result.

is statistically significant; that is, the confidence intervals do not overlap. The gap narrows over time, with preferences for presidential elections coming increasingly into focus, especially during the last 50 days. The two are virtually indistinguishable on Election Day. By that point in time, preferences in both types of elections are pretty much fully formed. There thus are important differences in the structure and evolution of preferences in presidential and parliamentary elections. This is as hypothesized.

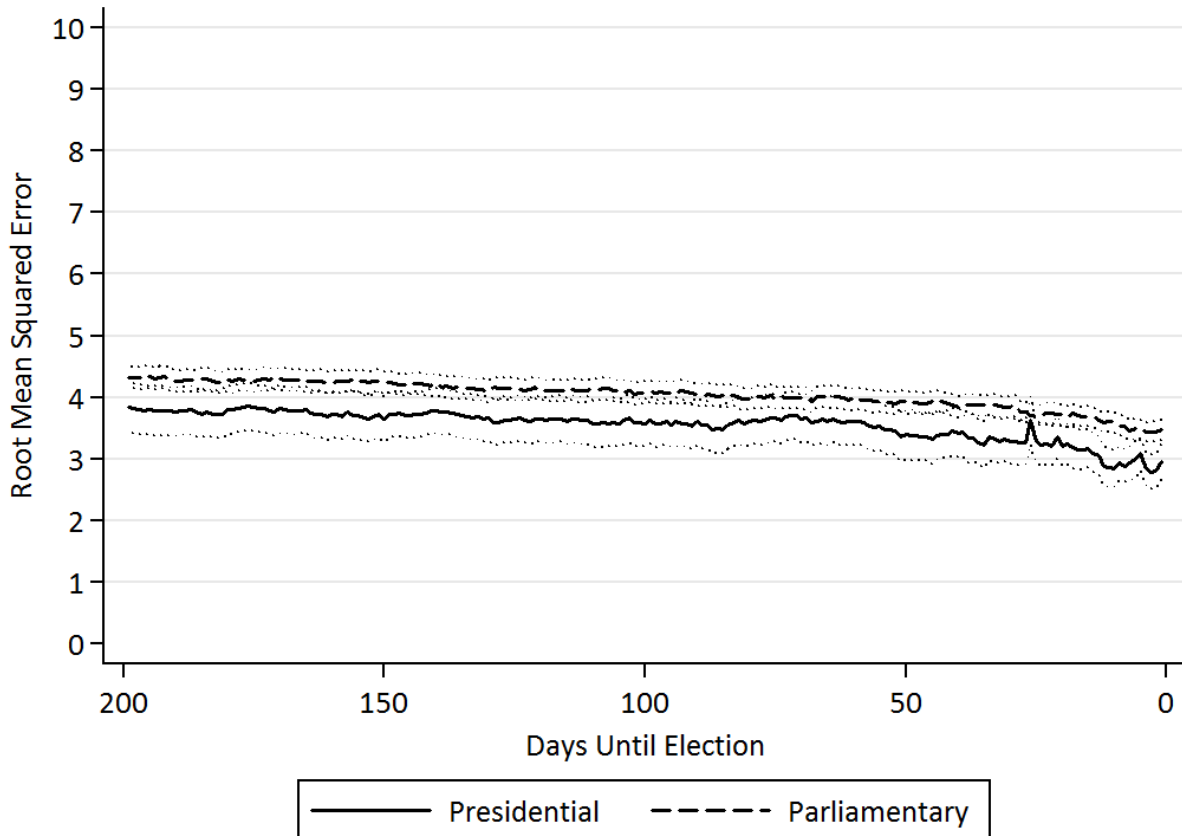
Figure 5. Root Mean Squared Errors for Presidential and Parliamentary Elections Taken Separately with Continuous Poll Readings



We next turn to differences between legislative elections in parliamentary and presidential systems. There is reason to suppose that party matters more in the former, so that electoral preferences crystallize earlier than legislative elections in the latter. Figure 6 shows the RMSEs for regressions relating the polls and the legislative vote in the two systems. The patterns in the figure indicate little difference in the evolution of preferences. While polls are slightly better predictors of the legislative vote in presidential systems, the differences are trivial. It seems that preferences for legislative elections evolve in much the same way regardless of the institutional context.²⁰ The only real difference relating to government institutions is between presidential elections on the one hand and legislative elections on the other.

²⁰ We have tested for differences between systems where the incumbent government is able to control the timing of legislative elections and those where it cannot (see Kayser 2005), and results suggest that preferences come into focus earlier in the former and remain so right up until the final days of the campaign. These differences hold – and do not vary significantly – across countries with different government and electoral institutions. Further details are in supplementary Appendix S3, Figure S3.3.

Figure 6. Root Mean Squared Errors for Legislative Elections in Presidential and Parliamentary Systems

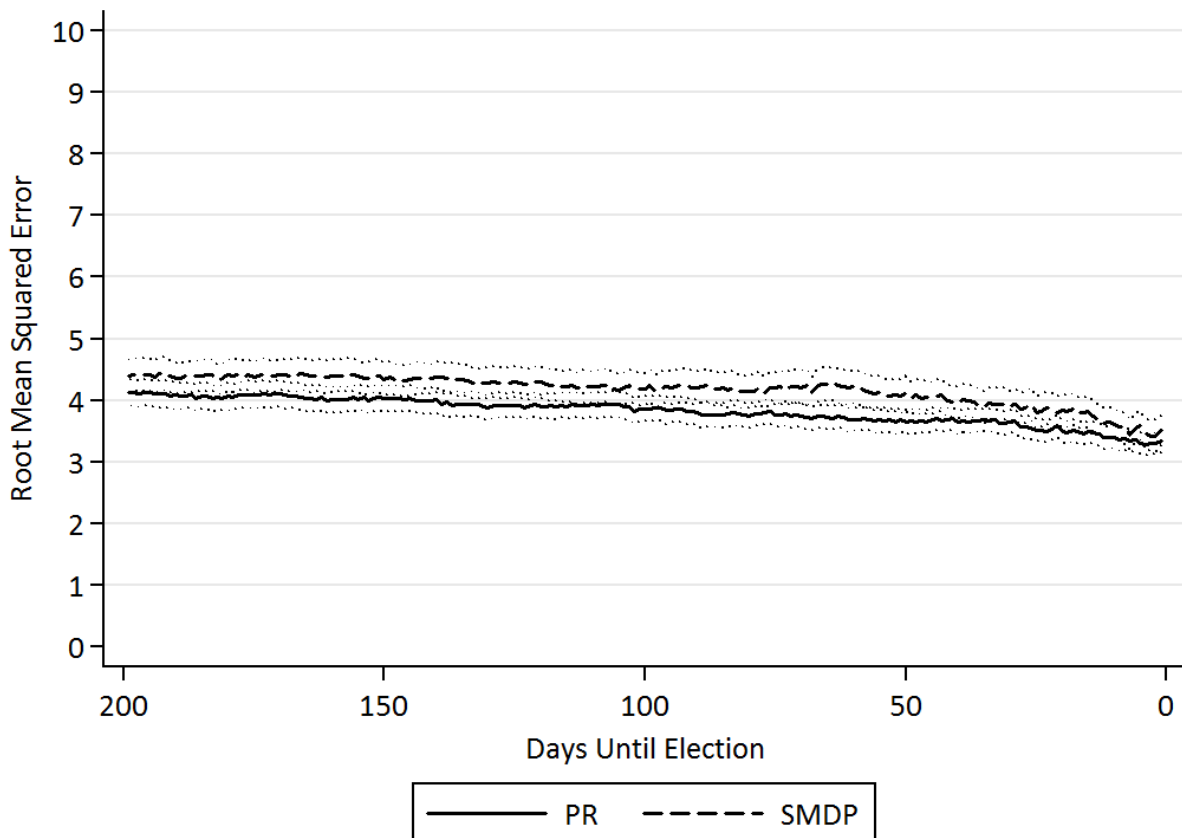


Electoral Institutions

Earlier, we posited that proportional and plurality systems might influence how electoral preferences evolve over the timeline. There are two, contrasting expectations: (1) that preferences crystallize earlier in proportional systems because of the greater importance of party support, and (2) that preferences crystallize late in proportional systems because of the greater choice, i.e., larger number of political parties.

To begin with, we examine the general differences between the two systems. Figure 7 plots the RMSEs for regressions relating the polls and the legislative vote in proportional and single member district plurality systems. In the figure we can see that polls are slightly more predictive in proportional systems. The differences are very small, however, and statistically significant only intermittently. If nothing else, the results indicate that preferences do not come into focus earlier and more completely in proportional systems.

Figure 7. Root Mean Squared Errors for Legislative Elections in Proportional and Plurality Systems



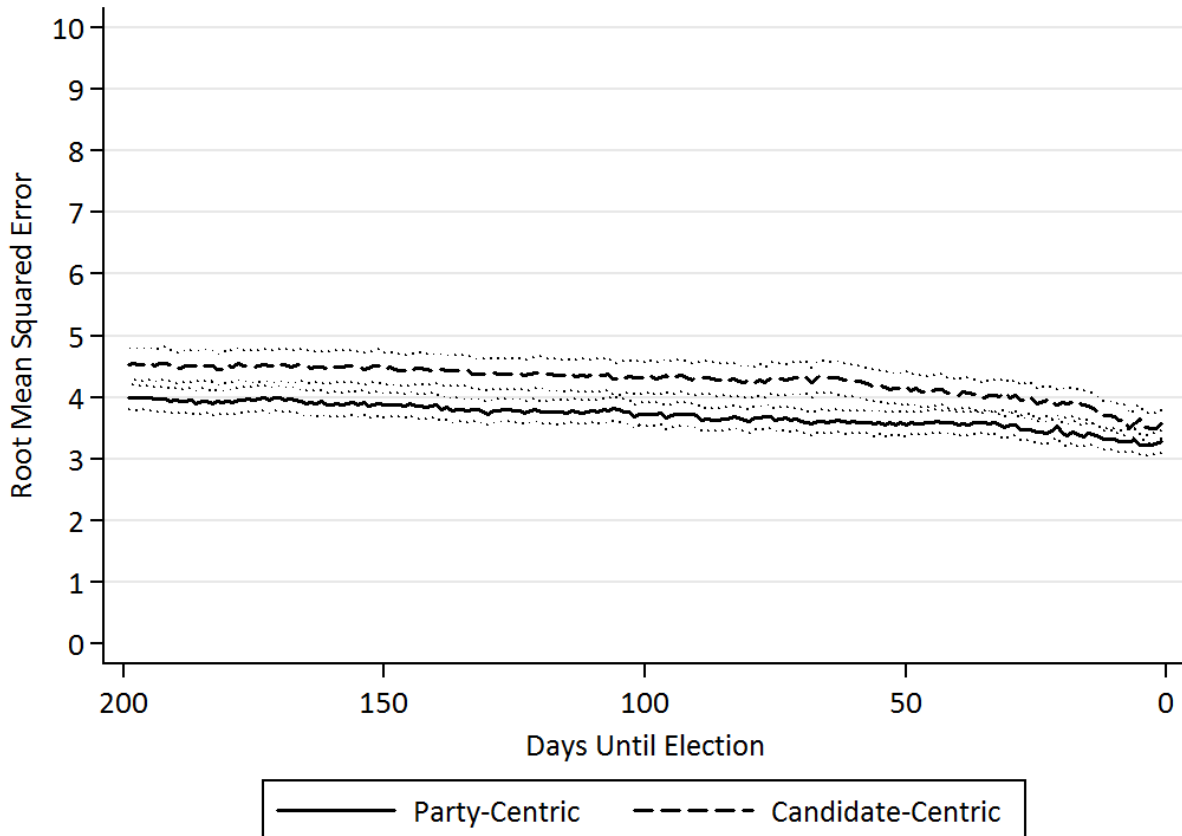
As discussed above, the potential impact of proportional and plurality systems may depend on electoral formula or rules, and the extent to which they structure vote choice by party preference or create incentives for candidates to cultivate a personal vote. We can test such differences. For this, we classify countries as either party- or candidate-centric based on the Johnson and Wallack (2013) index. Specifically, we code a country’s electoral rules as party-centric if it scores 1 to 6 on the index, and candidate-centric if it scores 7 to 12. Table 2 shows the overlap of the proportional-plurality and party-candidate categories – confirming that most but not all proportional systems are party-dominated and that most but not all plurality systems are candidate-centric.

Table 2. Proportional-Plurality and Party/Candidate Electoral Rules

	Party-centric	Candidate-centric	
PR	23	3	26
SMDP	1	5	6
	24	8	32

Figure 8 plots the RMSEs for party- and candidate-centric electoral systems. This reveals a larger and statistically significant difference. From 200 days out, preferences are clearer in party-centric systems, with the difference only narrowing slightly in the final 50 days, and even then only marginally. This suggests that, while there is weak evidence that proportional systems per se structure electoral preferences in meaningful ways, party-centric electoral rules that often are associated with PR systems do matter.

Figure 8. Root Mean Squared Errors for Legislative Elections in Party- and Candidate-Centric Systems



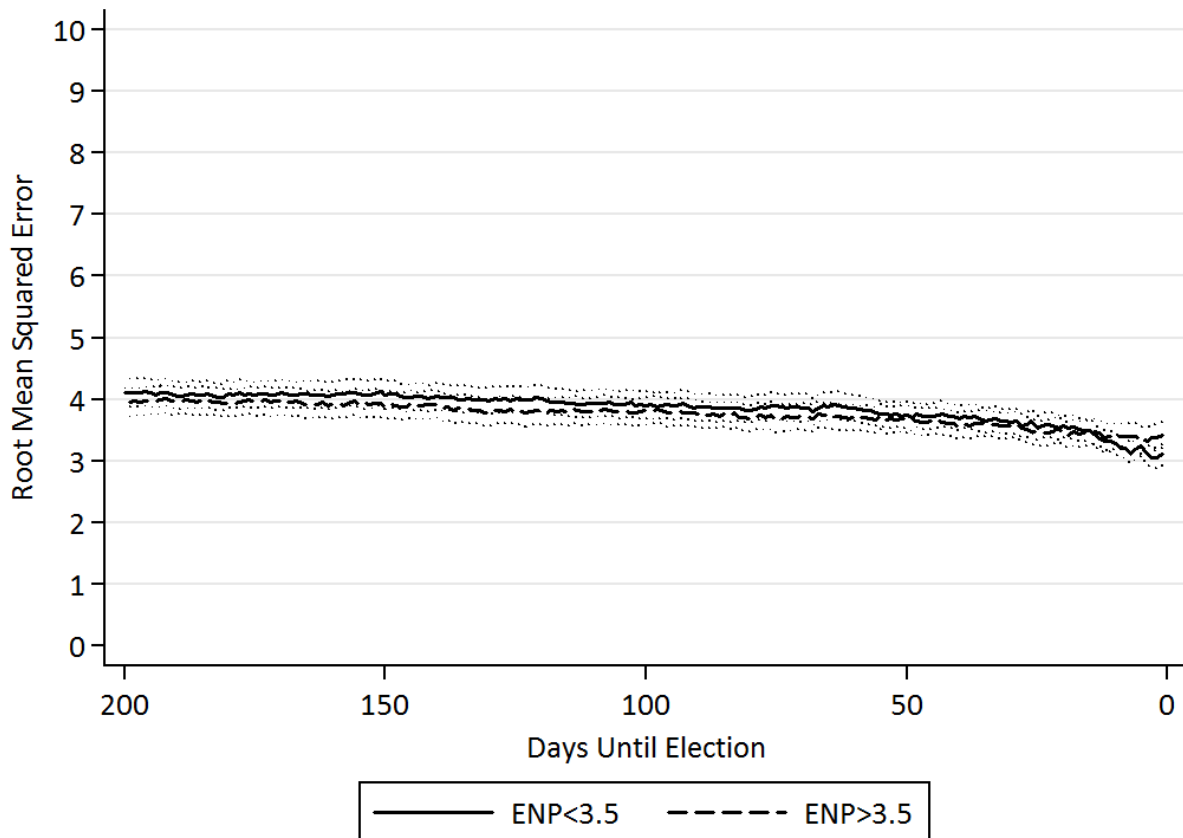
As discussed, there are other ways in which proportionality can influence the formation of preferences, including the number of choices that electoral systems present. We hypothesized that choice could reduce the predictability of elections from polls owing to volatility in preferences it induces. To test the possibility, Figure 9 plots the RMSEs from regressing vote shares on polls for those election cycles where the effective number of parties (ENP) is less than

or greater than 3.5.²¹ This threshold is the mean effective number of parties across all election cycles in legislative elections, enabling us to assess the difference between electoral systems that are highly fractionalized and those that are not. The differences here are inscrutable. While the RMSEs are lower over most of the time period for election cycles where there are fewer than three-and-a-half effective parties, the differences are trivial and statistically insignificant. There thus is little evidence to support the hypothesis of greater instability in preferences in multiparty systems.²² The main effect of electoral institutions thus seems to be in the degree to which they encourage party voting and so increase electoral predictability over the timeline of elections.

²¹ Following Laakso and Taagepera (1979), the ENP is calculated as the sum of the squared fraction of votes (V) for each party i , divided by one. That is, $ENP_e = \frac{1}{\sum_{i=1}^n V_i^2}$.

²² If an ENP of 2.5 is used as the threshold, the difference is consistent and significant, reflecting the differences between proportional and majoritarian systems suggested above.

Figure 9. Root Mean Squared Errors for Legislative Elections across the Effective Number of Political Parties



Discussion and Conclusion

Voter preferences evolve in a systematic way over the election timeline in a wide range of representative democracies. There is structure to preferences well in advance of elections, indeed, years before citizens actually vote. That is, very early polls predict the vote, at least to some extent. This largely reflects differences in the equilibrium support of parties and candidates. Polls do become increasingly informative over time, however, pointing to real evolution of preferences. That this pattern holds across countries is important and points towards

a general tendency in the formation of electoral preferences. But the pattern is not precisely the same in all countries. Political institutions structure the evolution of voters' preferences.²³

Government institutions are important. Preferences come into focus later in presidential elections than in parliamentary ones. A year out from Election Day, parliamentary elections are more predictable from the polls than are the outcomes of presidential races. This presumably reflects the greater uncertainties involved in the assessment of presidential candidates and also the time it takes for voters to directly factor in their dispositions toward the political parties (Erikson and Wlezien 2012). In parliamentary systems, by contrast, parties matter more early on. This is important because partisan dispositions, while not fixed, are more durable than those toward candidates. That preferences are in place much later in presidential systems thus comes as little surprise. That there is no real difference between legislative elections in presidential and parliamentary systems may surprise, however. It implies that parties do not matter consistently more to voters in the latter.

Electoral institutions also are important. Preferences in legislative elections come into focus more quickly and completely in proportional systems. We find limited evidence of general

²³ We also estimate more fully specified versions of our vote-polls equation using data on macro-partisanship and economic perceptions from the US, UK and Germany, countries where we have measures for numerous elections, to assess whether the findings are robust to their inclusion. Results confirm both the early structuration of electoral preferences and their evolution over time, as well as institutional differences. (See supplementary Appendix S3, Figures S3.5 to S3.12.) The results are robust to the inclusion of leader and party competence evaluations where available, indicating that the findings are not an artifact of under-specification.

differences across systems—that proportional representation per se is what matters. We find stronger evidence that the party-centricity of the systems matters most of all. Although closely related to proportionality, there is significant variation in party-centricity within both proportional and plurality systems, and this variation is of consequence for the formation of electoral preferences. The number of parties, meanwhile, appears to have little effect.

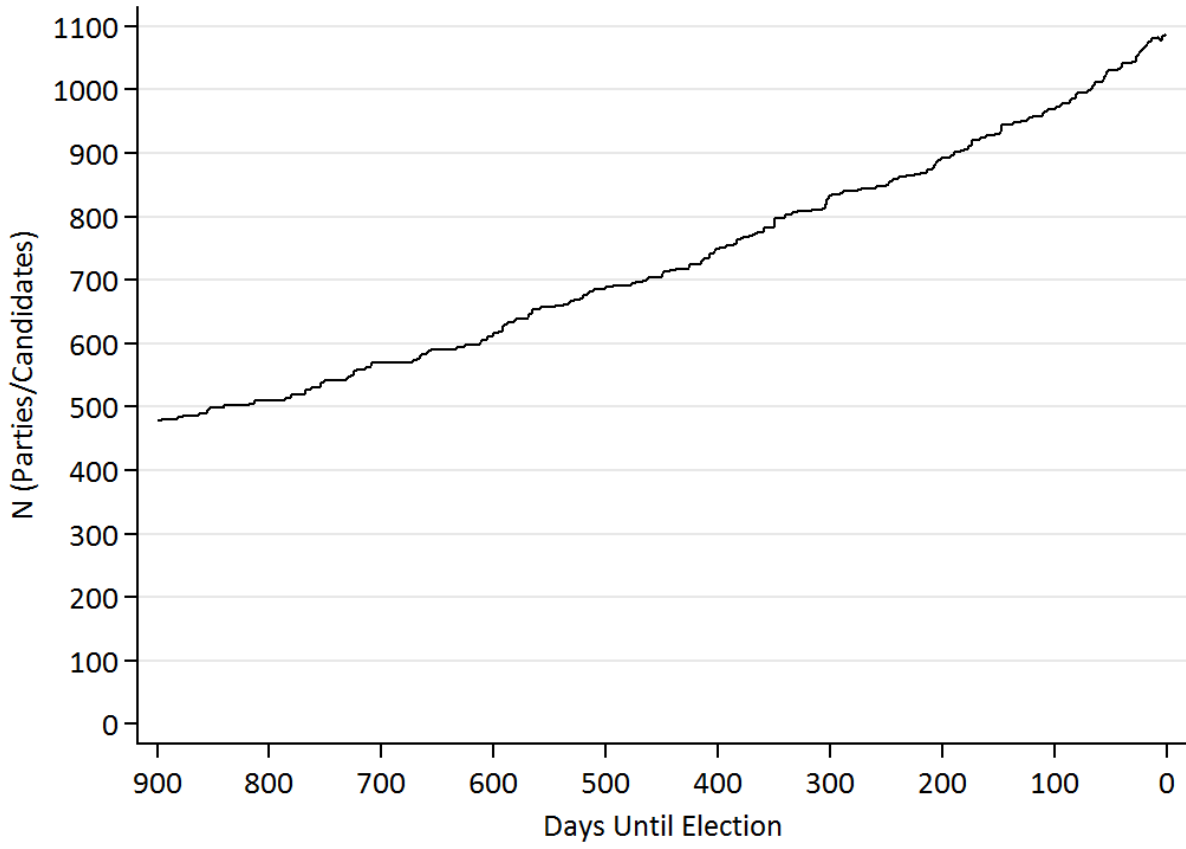
We have only scratched the surface of the variation in context. To begin with, political institutions differ in ways that we have not considered. Perhaps more importantly, there are other differences in context that we have not even begun to explore. Some of the differences relate to countries themselves. For instance, following Converse (1969), there is reason to think that the age of democracy is important to the formation and evolution of preferences. Other differences relate not to political institutions or the countries themselves, but to characteristics of political parties. There are numerous possibilities here, most notable of which may be whether parties are in government or opposition, as is suggested by the literature on economic voting (e.g. Fiorina 1981; Duch and Stevenson 2008). Another is whether parties are catch-all or niche. The age and size of parties also could matter. Clearly, much research remains to be done, and our methodology can guide the way.

That said, we have learned something about the general pattern relating preferences and the vote over the election timeline and the structuring influences of political institutions. We have shown that preferences are often in place far in advance of Election Day and that they evolve slowly over time. Indeed, the final outcome is fairly clear in the polls before the election campaign really begins. This is not to say that the campaign does not matter, as it does, particularly in

certain types of countries and elections where candidates are central. Even there, however, it is clear that the “long campaign” between elections matters most of all.

APPENDIX

Figure A1. Number of Parties and Candidates for which there are Poll Data



REFERENCES

- Abramowitz, Alan I. 1991. "Incumbency, Campaign Spending, and the Decline of Competition in U.S. House Elections." *Journal of Politics* 53(1): 34-56.
- Abramson, Paul R., John H. Aldrich, and David W. Rohde. 2007. *Change and Continuity in the 2004 and 2006 Elections*. Washington, D.C.: CQ Press.
- Adams, James, Samuel Merrill, III, and Bernard Grofman. 2005. *A Unified Theory of Party Competition: A Cross-National Analysis Integrating Spatial and Behavioral Factors*. Cambridge: Cambridge University Press.
- Arceneaux, Kevin. 2005. "Do Campaigns Help Voters Learn: A Cross-National Analysis." *British Journal of Political Science* 36(1): 159-173.
- Bafumi, Joseph, Robert S. Erikson, and Christopher Wlezien. 2010. "Balancing, Generic Polls and Midterm Congressional Elections." *Journal of Politics* 72(3):705-719.
- Berelson, Bernard R., Paul F. Lazarsfeld, and William N. McPhee. 1954. *Voting: a study of opinion formation in a presidential campaign*. Chicago: University of Chicago Press.
- Blais, Andre, Elisabeth Gidengil, Richard Nadeau, and Neil Nevitte. 2003. "Campaign Dynamics in the 2000 Canadian Election: How the Leader Debates Salvaged the Conservative Party." *PS: Political Science and Politics* 36(1): 45-50.
- van der Brug, Wouter, Cees van der Eijk and Mark Franklin. 2007. *The Economy and the Vote: Economic Condition and Elections in 15 Countries*. Cambridge: Cambridge University Press.
- Budge, Ian and Dennis Farlie. 1983. *Explaining and Predicting Elections*. London: Allen and Unwin.

- Cain, Bruce, John A. Ferejohn, and Morris P. Fiorina. 1984. "The Constituency Service Basis of the Personal Vote for U.S. Representatives and British Members of Parliament." *American Political Science Review* 78(1): 110-125.
- Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. New York: Wiley.
- Campbell, James E. 2000. *The American Campaign: U.S. Presidential Campaigns and the National Vote*. College Station, Texas: Texas A&M University Press.
- Clarke, Harold, David Sanders, Marianne Stewart and Paul Whiteley. 2004. *Political Choice in Britain*. Oxford: Oxford University Press.
- Converse, Philip E. 1969. "Of Time and Partisan Stability." *Comparative Political Studies* 2(2): 139-171.
- Cox, Gary. 1997. *Making Votes Count: Strategic Coordination in the World's Electoral Systems*. Cambridge: Cambridge University Press.
- Dalton, Russell J. and Hans-Dieter Klingemann (eds.). 2007. *Oxford Handbook of Political Behavior*. Oxford: Oxford University Press.
- Duch, Raymond and Randolph Stevenson. 2008. *The Economic Vote*. Cambridge: Cambridge University Press.
- Duverger, Maurice. 1954. *Political Parties: Their Organization and Activity in the Modern State*. New York: Wiley.
- van der Eijk, Cees and Mark Franklin. 1996. *Choosing Europe? The European Electorate and National Politics in the Face of Union*. Ann Arbor: University of Michigan Press.

- Erikson, Robert S. and Christopher Wlezien. 2012. *The Timeline of Presidential Elections: How Campaigns do (and do not) Matter*. Chicago: University of Chicago Press.
- Fenno, Richard F. 1978. *Home Style: House Members in their Districts*. Boston: Little, Brown.
- Ferejohn, John, and Randall Calvert. 1984. "Presidential Coat-Tails in Historical Perspective." *American Journal of Political Science* 28(1): 127-146.
- Finkel, Steven E. 1993. "Reexamining the "Minimal Effects" Model in Recent Presidential Campaigns." *Journal of Politics* 55(1): 1-21.
- Fiorina, Morris P. 1981. *Retrospective Voting in American Elections*. New Haven: Yale University Press.
- Fournier, Patrick, Richard Nadeau, Andre Blais, Elisabeth Gidengil, and Neil Nevitte. 2004. "Time of Voting Decision and Susceptibility to Campaign Effects." *Electoral Studies* 23(4): 661-681.
- Gelman, Andrew and Gary King. 1993. "Why are American Presidential Election Polls so Variable When Votes are so Predictable?" *British Journal of Political Science* 23(4): 409-451.
- Golder, Matt. 2006. "Presidential Coattails and Legislative Fragmentation." *American Journal of Political Science* 50(1): 34-48.
- Heath, Anthony. 2010. "Perspectives on Electoral Behavior." In Dalton, Russell J. and Hans-Dieter Klingemann (eds.), *Oxford Handbook of Political Behavior*. Oxford: Oxford University Press.
- Hillygus, D. Sunshine and Todd G. Shields. 2008. *The Persuadable Voter: Wedge Issues in Presidential Election Campaigns*. Princeton: Princeton University Press.

- Holbrook, Thomas. 1996. *Do Campaigns Matter?* Thousand Oaks, California: Sage.
- Jacoby, William. 2010. "The American Voter." In Jan E. Leighley (ed.), *The Oxford Handbook of American Elections and Political Behavior*. Oxford: Oxford University Press.
- Johnson, Joel W., and Jessica S. Wallack. 2013. "Electoral Systems and the Personal Vote." *Harvard Dataverse Network*, <http://hdl.handle.net/1902.1/17901>.
- Johnston, Richard, Andre Blais, Henry E. Brady, and Jean Crete. 1992. *Letting the People Decide: Dynamics of a Canadian Election*. Kingston, Canada: McGill-Queen's Press.
- Kayser, Mark. 2005. "Who Surfs, Who Manipulates?" *American Political Science Review* 99(1): 17-27.
- Krueger, James S., and Michael Lewis-Beck. 2007. 'Goodness-of-Fit: R-Squared, SEE and 'Best Practice'.' *The Political Methodologist* 15(1): pp.2-4.
- Laakso, Markku, and Rein Taagepera. 1979. "'Effective' Number of Parties: A Measure with Application to West Europe." *Comparative Political Studies* 12(1): 3-27.
- Lau, Richard. 1994. "An Analysis of the Accuracy of 'Trial-Heat' Polls during the 1992 Presidential Election." *Public Opinion Quarterly* 58(1): 2-20.
- Lewis-Beck, Michael, and Tom W. Rice. 1992. *Forecasting Elections*. Washington, DC: CQ Press.
- Lodge, Milton, Marco Steenbergen, and Shawn Brau. 1995. "The Responsive Voter: Campaign Information and the Dynamics of Candidate Evaluation." *American Political Science Review* 89(2): 309-326.

- Mair, Peter. 1997. *Party System Change: Approaches and Interpretations*. Oxford: Clarendon Press.
- McDermott, Monika and Kathleen Frankovic. 2003. "Horserace Polling and Survey Method Effects: Analysis of the 2000 Presidential Campaign." *Public Opinion Quarterly* 67(2): 244-264.
- Moore, David and Lydia Saad. 1997. "The Generic Ballot in Midterm Congressional Elections: Its Accuracy and Relationship to House Seats." *Public Opinion Quarterly* 61(4): 603-614.
- Pedersen, Morgen. 1979. "The Dynamics of European Party Systems: Changing Patterns of Electoral Volatility." *European Journal of Political Research* 7(1): 1-26.
- Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: J. Wiley & Sons.
- Shaw, Daron R. 1999. "A Study of Presidential Campaign Event Effects from 1952 to 1992." *Journal of Politics* 61(2): 387-422.
- Shugart, Matthew, and John Carey. 1992. *Presidents and Assemblies: Constitutional Design and Electoral Dynamics*. Cambridge: Cambridge University Press.
- Shugart, Matthew Sjøberg, Melody Ellis Valdini, and Kati Suominen. 2005. "Looking for Locals: Voter Information Demands and Personal Vote-Earning Attributes of Legislators under Proportional Representation." *American Journal of Political Science* 49(2): 437-449.
- Stevenson, Randolph and Lynn Vavreck. 2000. "Does Campaign Length Matter? Testing for Cross-National Effects." *British Journal of Political Science* 30(2): 217-235.
- Vavreck, Lynn. 2009. *The Message Matters*. Princeton: Princeton University Press.

- Wlezien, Christopher and Robert S. Erikson. 2001. "Campaign Effects in Theory and Practice." *American Politics Research* 29(5): 419-437.
- Wlezien, Christopher and Robert S. Erikson. 2002. "The Timeline of Presidential Election Campaigns." *Journal of Politics* 64(4): 969-993.
- Wlezien, Christopher, Will Jennings, Stephen Fisher, Robert Ford, and Mark Pickup. 2013. "Polls and the Vote in Britain." *Political Studies* 61(S1): 66-91.
- Zaller, John. 2002. "The Statistical Power of Election Studies to Detect Media Exposure Effects in Political Campaigns." *Electoral Studies* 21(2): 297-329.
- Zaller, John. 1998. "Politicians as Prize Fighters: Electoral Selection and Incumbency Advantage." In John Geer (ed.), *Party Politics and Politicians*. Baltimore: Johns Hopkins University Press.

SUPPLEMENTARY MATERIALS

APPENDIX S1. POLL DATA

In this supplementary appendix we summarize details of the poll data collected for this study. More than 26,000 polls were compiled from a large number of sources, and consistent procedures and coding rules were followed. Note also that exit polls are not included in our analysis.

There are several important points about these data. Firstly, we are compelled to work with vote intention figures that do not reflect consistent sampling or weighting strategies by different polling organizations or even by the same organization over time. Older polls are more likely to use face-to-face and quota samples, for example, whereas recent polls may include internet panels. While we ideally would like to work with data generated using a consistent methodology, assembling a time series that takes into account differences in weighting and sampling practices is impossible, as the required data are not available for most of the polls. We therefore use the headline figure vote intentions as the most consistent attainable time series of poll data—the numbers reflect the survey houses' best estimates of voter preferences at each point in time. Where a survey house changes their sampling or weighting strategies our poll data will reflect this change. Unfortunately, there is little alternative to using the headline figures, as these often are the only available data. It also is the norm in previous research.²⁴

²⁴ These decision rules might seem innocuous but the poll universe and, especially, weighting have been shown to affect the reported headline figures, particularly in recent elections (Moore and Saad 1997; Wlezien and Erikson 2001). This does not appear to influence the evolving accuracy of reported polls, at least based on analysis of presidential and congressional polls in the US (Erikson and Wlezien 2004; 2012; Bafumi et al. 2010).

Second, survey organizations typically conduct polls over multiple days, which requires a number of coding decisions. To begin with, for organizations reporting moving averages from a tracking poll, we use non-overlapping results. Since most polls are conducted over multiple days, where possible we “date” each poll by the middle day of the period that the survey is in the field. For days when more than one poll result is recorded, we pool the results together into a single poll of polls. During the later stages of the election cycle, we often have near day-to-day monitoring of vote intentions.²⁵

Third, the length of election cycles vary considerably. Some presidential elections involve a five- or six-year time interval, as do some legislative elections, while run-off elections can span just a couple of weeks, resulting in a very short election cycle. Because pollsters ask hypothetically about vote intentions for run-off elections we are able to extend our analysis beyond this period in some cases. In some countries there are legal restrictions on publication of poll results on or prior to Election Day (for a review see Spangenberg 2003). This means that in a few cases we have missing data over the final days of the campaign. In such circumstances, we carry forward the results from the final poll until the very end of the cycle. Thus, our analysis understates the strength of the relationship between polls and the election outcome at that point in time.

²⁵ It is important to note that polls on successive days are not truly independent. 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, which complicates a conventional time-serial analysis.

Non-Responses and Don't Knows

The norm in the polling industry is to adjust vote intention polls to exclude don't knows and non-responses. However in a small number of cases non-responses are included in the headline figures and we recalculate the poll numbers to ensure that the data are standardized. These sorts of adjustment are the exception but have been implemented consistently.

Dating the Polls

Since most polls are conducted over multiple days, where possible we “date” each poll by the middle day of the period that the survey is in the field. For surveys in the field for an even number of days, the fractional midpoint is rounded down to the earlier day. Information on fieldwork dates is not available for all polls and in those cases we follow careful procedures to calibrate the date assigned to each poll. The rules for poll dating are as follows, using the first possible option before moving onto the next when that possibility had been exhausted: (1) if both fieldwork dates available, the mid-point of the start and end dates is calculated, (2) if only one of the fieldwork dates is available, that date is used, (3) if only the date of publication of the poll in the media is available, that date is used, (4) if only information on the month or week of the poll is available, the mid-point of the corresponding month or week is used, (5) if only information on the month of the poll is available and is observed during the month of the election and is known to be prior to the election, the first of the month is used as the start date and the final day before the election day is used as the end date (and if the poll asks about voting on “... next Monday [or other day]”, then the start date is instead taken as seven days before the election).

Sources

Polls were compiled from a large number of sources, with additional cross-checks and triangulation conducted in the case of inconsistencies or missing data. Wherever possible, polls obtained from secondary poll aggregators were cross-checked and triangulated against other available sources, including the original cross-tabs or media reports. Some of our largest country datasets were collected from archival survey repositories. These included the Roper Center for Public Opinion Research's *iPoll* databank, the Norwegian Social Science Data Archive, the Australian Social Science Data Archive, the Netherlands' Data Archiving and Networked Services, Canadian Opinion Research Archive, and the GESIS/Leibniz Institute for the Social Sciences). A number of datasets were kindly shared with us by other scholars or pollsters. The sources of poll data for our largest poll collections are listed below.

- United States: presidential trial-heat polls are from Erikson and Wlezien (2012). Congressional poll data consist of 1,997 polls from Bafumi et al (2010), further supplemented with data from the Roper Center for Public Opinion Research's *iPoll* databank.
- United Kingdom: dataset of national surveys where respondents were asked about which party they would vote "if the election were held tomorrow" from Wlezien et al (2013), including data from Michael Thrasher, Mark Pack, Ipsos-MORI, YouGov, ICM Research Ltd, Gallup Political and Economic Index.
- Portugal: poll data kindly provided by Francisco José Veiga (see Veiga and Veiga 2004).

- Australia: historical data from the Australian Social Science Data Archive; additional data from Newspoll (www.newspoll.com.au) and Roy Morgan Research (<http://www.roymorgan.com/>).
- Ireland: poll data via Michael Marsh's Irish Opinion Poll Archive. (http://www.tcd.ie/Political_Science/IOPA/)
- Germany: Forschungsgruppe Wahlen "Politbarometer" data from GESIS/Leibniz Institute for the Social Sciences; additional poll data from the *Wahlrecht.de* website (<http://www.wahlrecht.de/>). Historical poll data from the Institut für Demoskopie, Allensbach were obtained from replication data for Christopher Anderson's (1995) *Blaming the Government*, via the Harvard Dataverse.
- Netherlands: the dataset "NIPO weeksurveys 1962-2000: NIWI/Steinmetz Archive study number P1654" from Data Archiving and Networked Services (DANS).
- Sweden: all companies' poll data from Johanna Laurin Gulled, Ipsos Public Affairs.
- Italy: all companies' poll data from Chris Hanretty and Graziella Castro.
- Norway: the following datasets from Norwegian Social Science Data Services -- "Respons Analyse AS, 2005-2012" (MMA0067), "ACNielsen, 1987-1994" (MMA0455), "Opinion, 1988-2003" (MMA0585), "Synovate (MMI), 1987-1998" (MMA0802), "TNS Gallup AS, 1964-2010" (MMA0952), "Opinion, 2007-2010" (MMA1119). More recent poll data was obtained from the TV2 Partibarometeret poll aggregator (<http://politisk.tv2.no/spesial/partibarometeret/>).

- Canada: poll data from monthly Gallup reports (1942-2000); data via the Canadian Opinion Research Archive.
- France: historical poll data from the publication Gallup Organization (1976) *The Gallup International Public Opinion Polls: France, 1939, 1944-1975*; contemporary poll data from TNS-Sofres (<http://www.tns-sofres.com>) and from other sources.
- Spain: data from El Centro de Investigaciones Sociológicas (CIS) (<http://www.cis.es/>) and other sources.

REFERENCES

- Anderson, Christopher. (1995). *Blaming the Government*. New York: M.E. Sharpe
- Bafumi, Joseph, Robert S. Erikson, and Christopher Wlezien. 2010. “Balancing, Generic Polls and Midterm Congressional Elections.” *Journal of Politics* 72:705-719.
- Erikson, Robert S. and Christopher Wlezien. 2012. *The Timeline of Presidential Elections: How Campaigns do (and do not) Matter*. Chicago: University of Chicago Press.
- Gallup Organization. (1976). *The Gallup International Public Opinion Polls: France, 1939, 1944-1975*. New York: Gallup Organization.
- Spangenberg, Frits. (2003). *The Freedom to Publish Opinion Poll Results: Report on a Worldwide Update*. Amsterdam: Foundation for Information.
- Veiga, Francisco José and Linda Gonçalves Veiga. 2004. “The Determinants of Vote Intentions in Portugal.” *Public Choice* 118(3/4): 341-364.

Wlezien, Christopher, Will Jennings, Stephen Fisher, Robert Ford, and Mark Pickup. 2013.

“Polls and the Vote in Britain.” *Political Studies* 61(S1): 66-91.

APPENDIX S2. ELECTION DATA

We rely on a wide range of official sources and election data resources. Official sources were preferred where possible. Where official sources were not readily available, resources such as the Election Guide database of the International Foundation for Electoral Systems (www.electionguide.org) were used as an alternative or were used to cross-check the reliability of data obtained from unofficial sources (such as the websites of opinion pollsters and academic or amateur poll spotters). Some of the older data is from Nohlen and Stöver (2010).

General Resources

The European Election Database of the Norwegian Social Science Data Services (NSD)

http://www.nsd.uib.no/european_election_database/

ElectionGuide, International Foundation for Electoral Systems

<http://www.electionguide.org/>

Political Database of the Americas: Electoral Systems and Data

<http://pdba.georgetown.edu/elecdata/arg/arg.html>

Election Resources

<http://electionresources.org/>

Nohlen, Dieter, and Philip Stöver. 2010. *Elections in Europe: A data handbook*. Baden-Baden: Nomos Verlagsgesellschaft.

Country-Specific Resources

Australian Politics and Elections Database at the University of Western Australia,

<http://elections.uwa.edu.au/>

Bundesministerium für Inneres, Austria,

http://www.bmi.gv.at/cms/BMI_wahlen/nationalrat/start.aspx

Federal Elections in Brazil, Brazil

http://electionresources.org/br/index_en.html

Bularian Parliament, Bulgaria

<http://www.parliament.bg/bg/electionassembly>

Elections Canada, Canada

<http://www.elections.ca/home.aspx>

Parliament of Canada, Canada

<http://www.parl.gc.ca/parlinfo/compilations/electionsandridings/ResultsParty.aspx>

Ministerio del Interior, Republica de Chile

<http://historico.servei.cl/>

Ministry of Interior, Cyprus

<http://www.ekloges.gov.cy/>

Consejo Nacional Electoral (National Electoral Council), Republic of Ecuador

<http://resultados.cne.gob.ec/>

Ministry of Justice, Finland

<http://www.vaalit.fi/>

Ministry of Interior, France

<http://www.interieur.gouv.fr/Elections/Les-resultats>

Der Bundeswahlleiter (the Federal Returning Officer), Germany

<http://www.bundeswahlleiter.de/en/index.html>

Ministry of the Interior, Greece

<http://ekloges.ypes.gr/>

Statistics Iceland

<http://www.statice.is/Statistics/Elections/>

Ministry of the Interior, Italy

<http://elezioni.interno.it/>

Ministry of Internal Affairs and Communications, Japan

http://www.soumu.go.jp/senkyo/senkyo_s/data/shugin44/index.html

Government of Malta

<http://www.gov.mt/en/Government/Government%20of%20Malta/Election%20Results/Pages/Elections-DOI-site.aspx>

Instituto Federal Electoral, Mexico

<http://www.ife.org.mx/portal/site/ifev2>

Statistics Norway, Norway

http://www.ssb.no/a/english/kortnavn/stortingsvalg_en/tab-2009-10-15-02-en.html

Electoral Commission, New Zealand

<http://www.electionresults.govt.nz/>

National Office of Electoral Processes, Peru

<http://www.onpe.gob.pe/inicio.php>

Commission on Elections, Republic of the Philippines

<http://www.comelec.gov.ph/>

Comissão Nacional de Eleições, Portugal

http://eleicoes.cne.pt/sel_eleicoes.cfm?m=raster

Ministry of the Interior, Spain

<http://www.infoelectoral.mir.es/min/>

Election Authority, Sweden

http://www.val.se/in_english/previous_elections/index.html

Federal Office of Statistics, Switzerland

<http://www.bfs.admin.ch/>

National Electoral Council, Venezuela

http://www.cne.gob.ve/web/estadisticas/index_resultados_elecciones.php

Rallings, C. and Thrasher, M. 2007. *British Electoral Facts, 1832–2006, seventh edition.*

Aldershot: Ashgate.

Rallings, C. and Thrasher, M. 2010. *Election 2010: The Official Results.* London: Biteback publishing.

APPENDIX S3. ADDITIONAL ANALYSES

Figure S3.1. Adjusted R-Squareds, 900 days

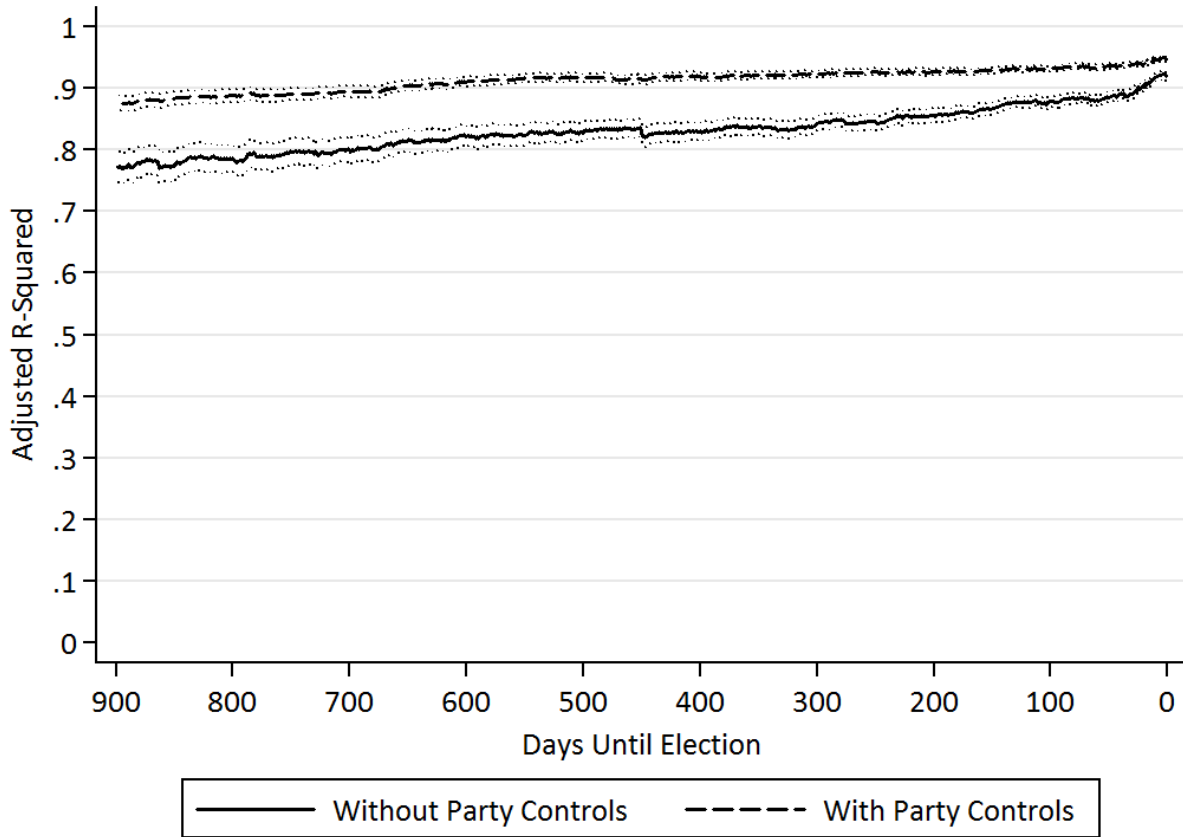


Figure S3.2. Regression Coefficients, 900 days

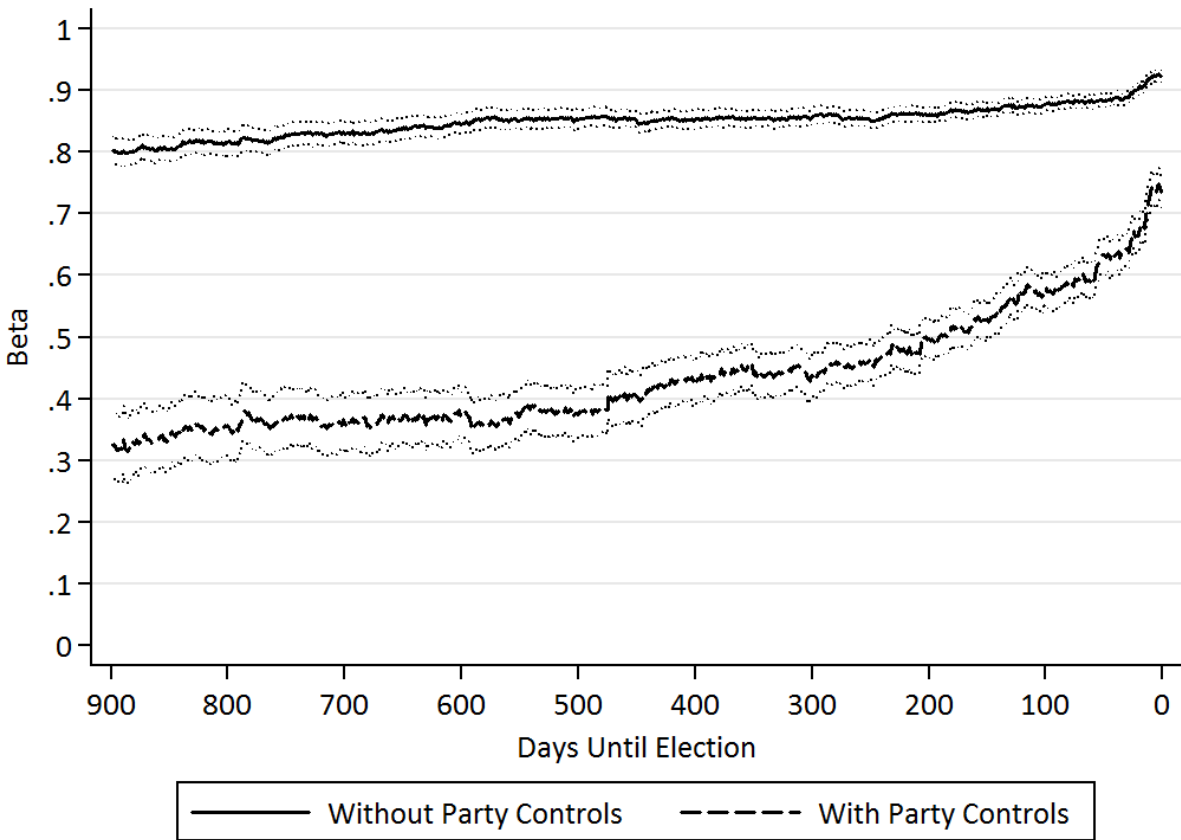


Figure S3.3. Government Control of Election Timing (Germany, Netherlands, Norway, Spain, US vs. Australia, Canada, Ireland, NZ, Portugal, UK)

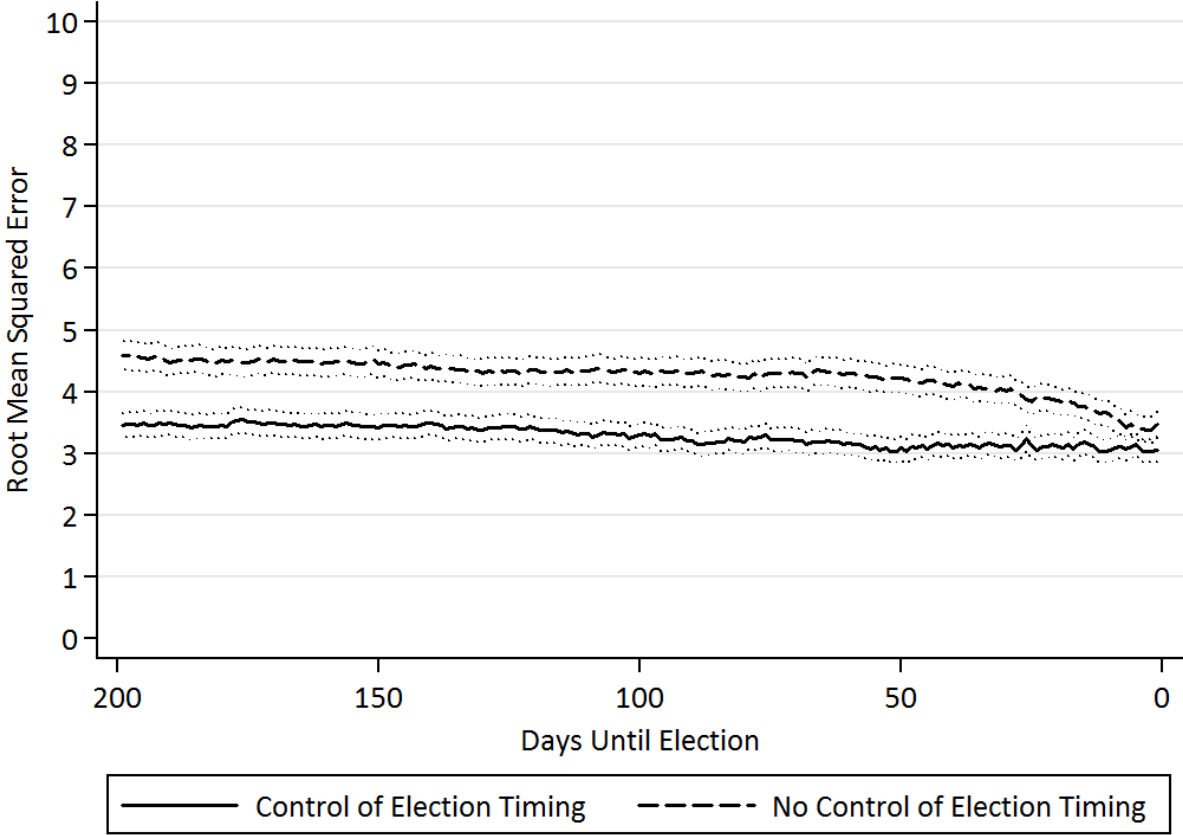


Figure S3.4. Old vs. New Democracies (post-1990 elections only)

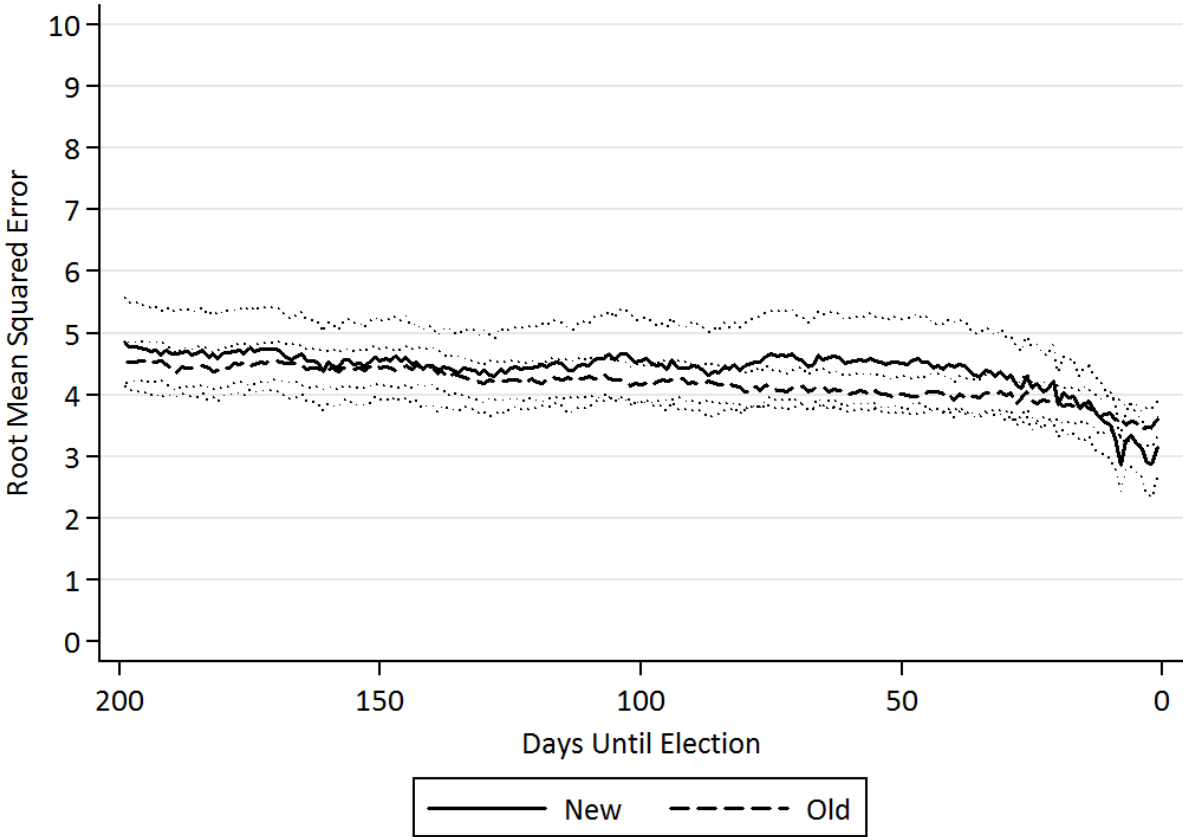


Figure S3.5. Regression Coefficients for the Polls (US, UK and Germany)

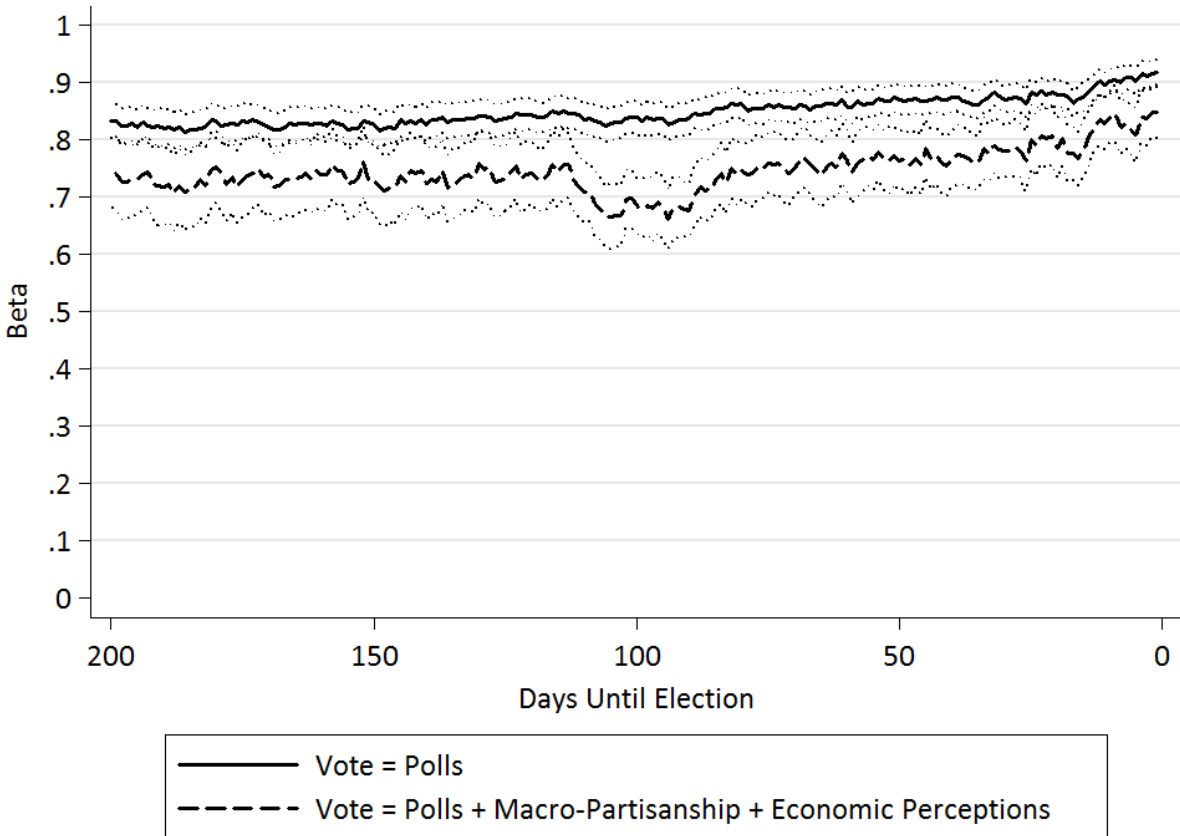


Figure S3.6. Root Mean Squared Errors (US, UK and Germany)

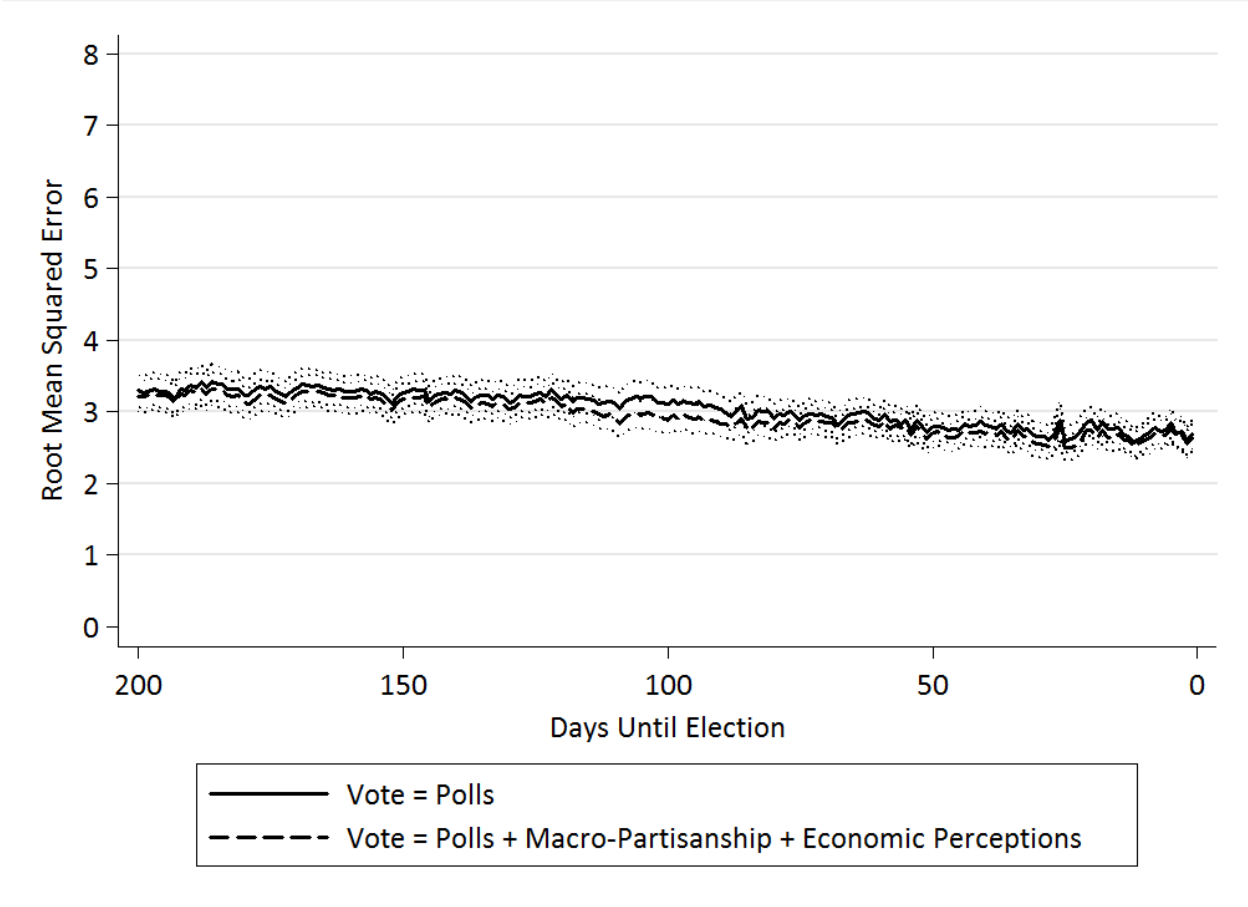


Figure S3.7. Root Mean Squared Errors (US, UK and Germany), Presidential vs. Legislative Elections, Vote = Polls

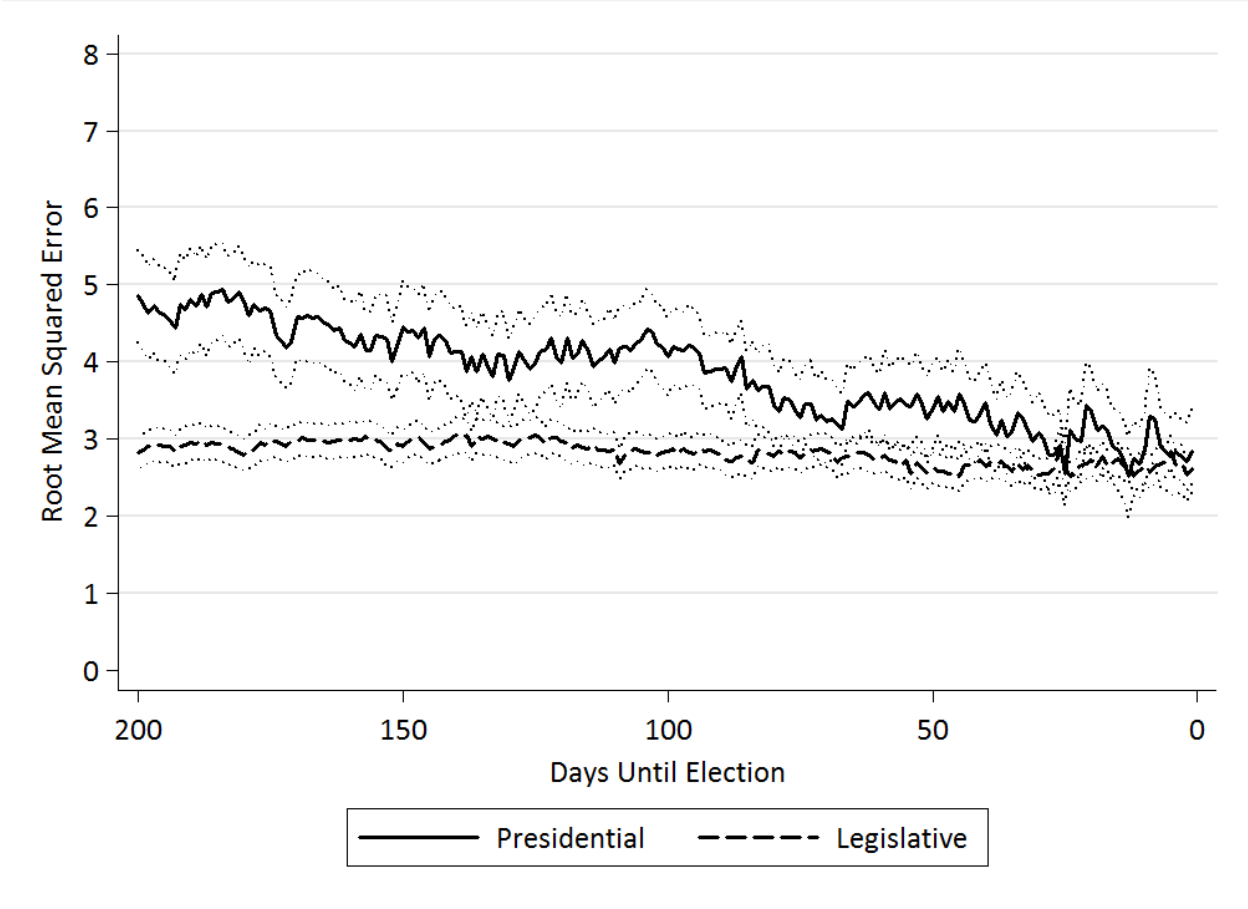


Figure S3.8. Root Mean Squared Errors (US, UK and Germany), Presidential vs. Legislative Elections, Vote = Polls + Macro-Partisanship + Economic Perceptions

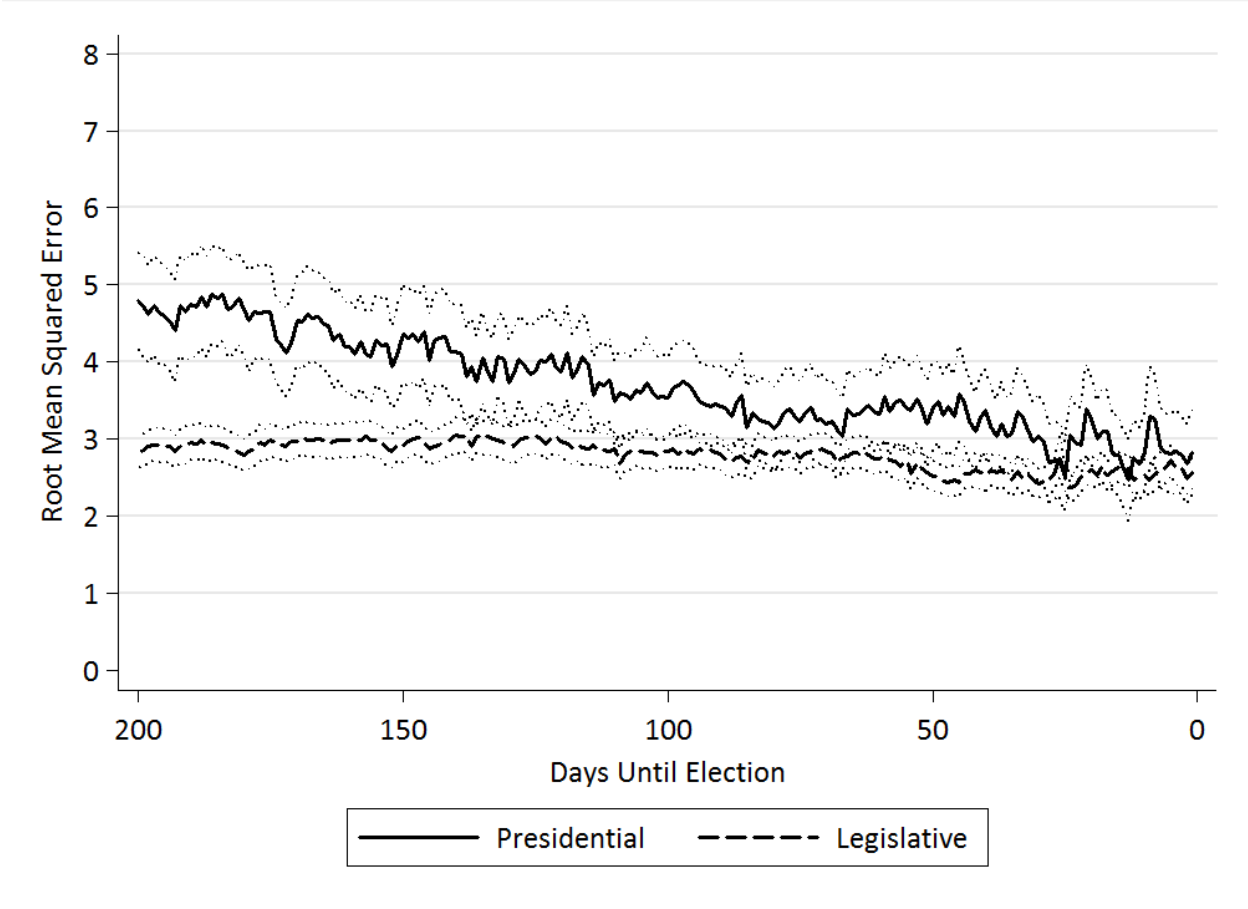


Figure S3.9. Root Mean Squared Errors (US, UK and Germany), Legislative Elections in Presidential vs. Legislative Systems, Vote = Polls

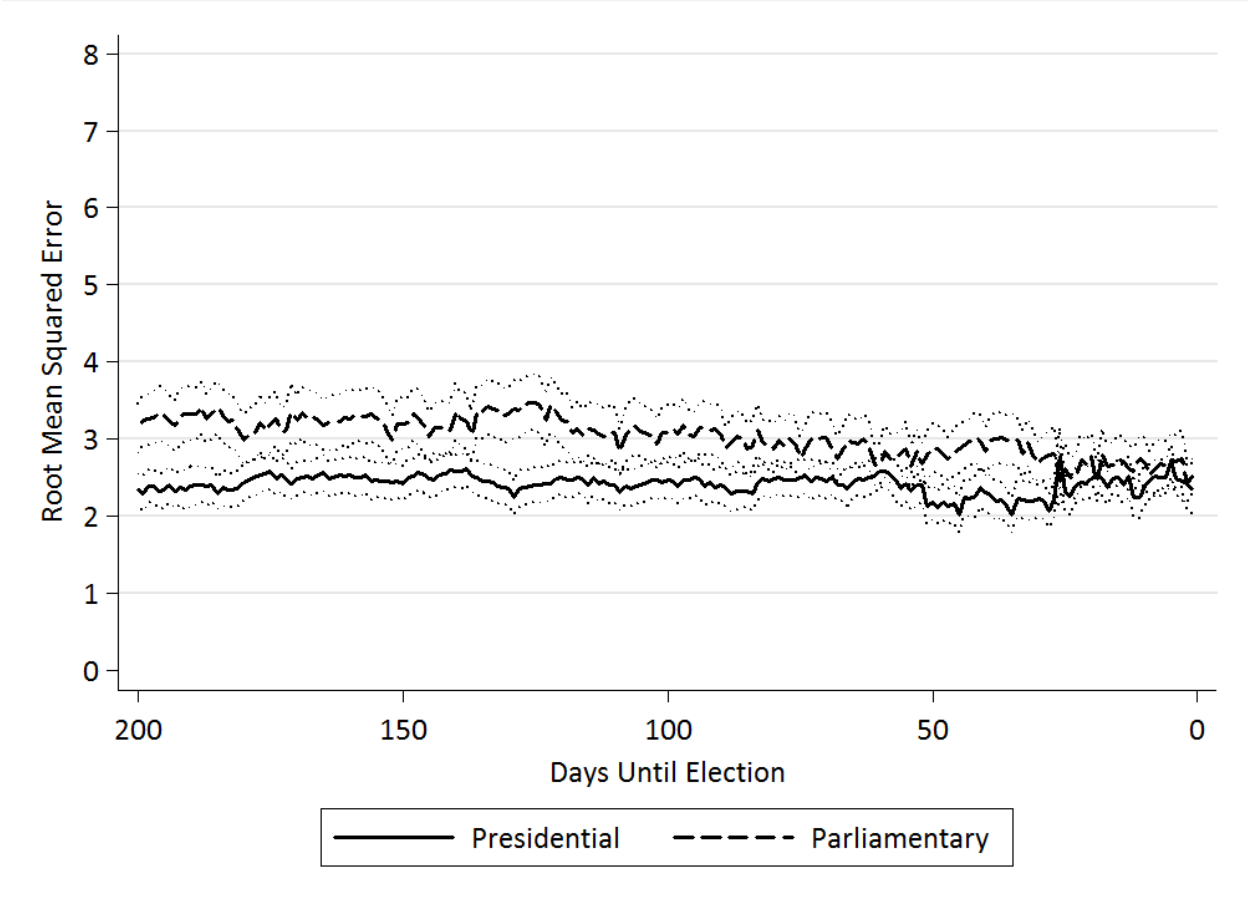


Figure S3.10. Root Mean Squared Errors (US, UK and Germany), Legislative Elections in Presidential vs. Legislative Systems, Vote = Polls + Macro-Partisanship + Economic Perceptions

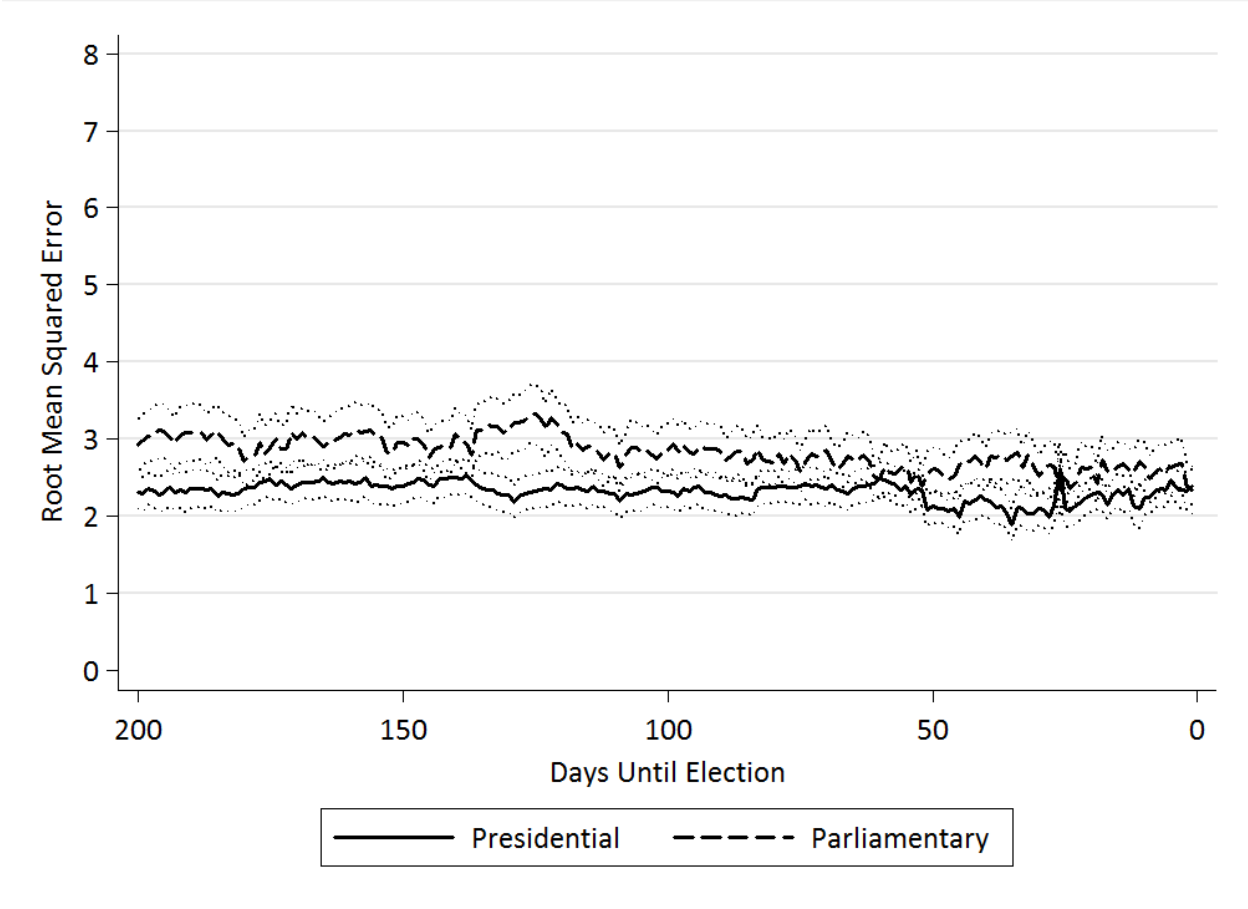


Figure S3.11. Root Mean Squared Errors (US, UK and Germany), Legislative Elections in Proportional vs. Plurality Systems, Vote = Polls

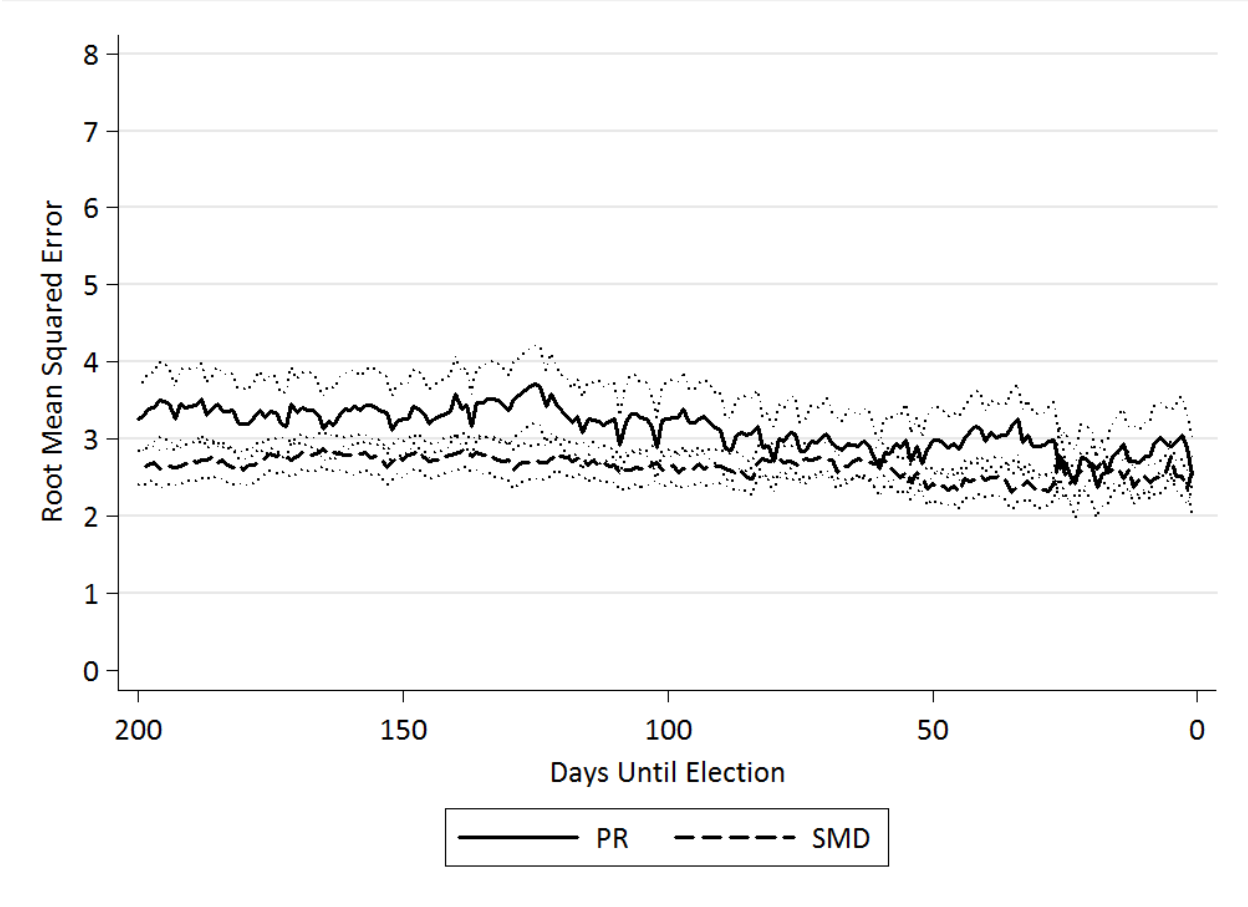
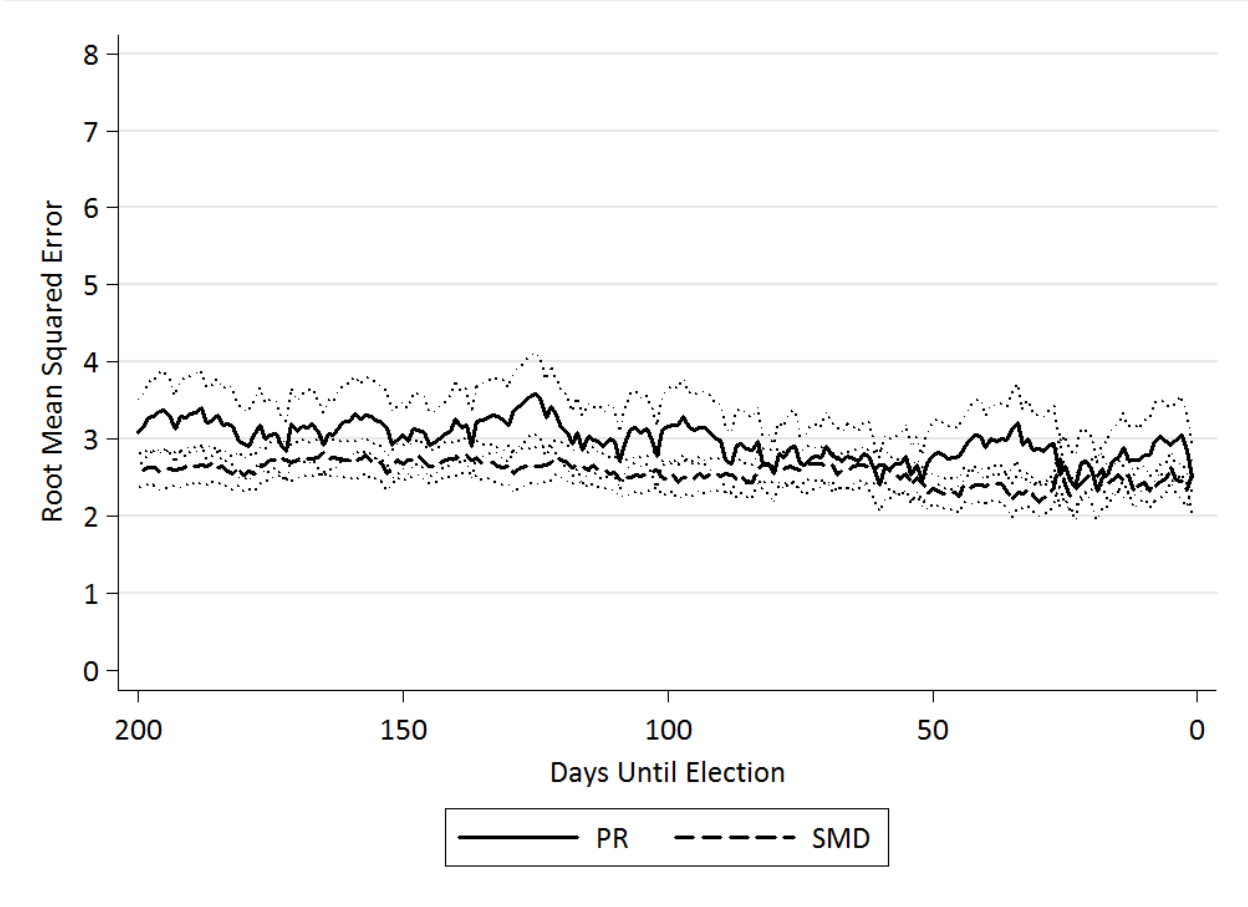


Figure S3.12. Root Mean Squared Errors (US, UK and Germany), Legislative Elections in Proportional vs. Plurality Systems, Vote = Polls + Macro-Partisanship + Economic Perceptions



Equations for Figures S3.5 to S3.12

$$VOTE_{jkm} = a_{jmT} + b_T Poll_{jkmT} + e_{jkmT} \quad (1)$$

$$VOTE_{jkm} = a_{jmT} + b_T Poll_{jkmT} + b_T Macro-Partisanship_{jkmT} + b_T Economic Perceptions_{jkmT} + e_{jkmT} \quad (2)$$

Measures

Macro-Partisanship: the percentage of party identifiers, based on aggregate survey data originally collected for MacKuen et al. (1989) and subsequently supplemented with more recent data. The UK data on macro-partisanship is from Bartle et al. (2011), and German data is aggregated from the Forschungsgruppe Wahlen “Politbarometer” stored at GESIS.

Consumer Sentiment: measures of consumer sentiment are obtained from the *OECD Business Tendency and Consumer Opinion Surveys (MEI) series*. The measure is inverted when parties are in opposition.

Macro-Competence: this measure is estimated using Stimson’s (1991) dyad ratios algorithm, extracting common variation in party issue handling evaluations across thousands of survey items in the UK, US and Germany, as introduced in Green & Jennings (2012).

Leader Approval: measures of party leader approval for the UK are from (Green & Jennings 2012), while data on presidential approval for the US is taken from the Roper Center.

References

Bartle, John, Sebastian, Dellepiane-Avellaneda, and James A. Stimson. 2011. “The Moving Centre: Preferences for Government Activity in Britain, 1950-2005.” *British Journal of Political Science* 41(2): 259-285.

Green, Jane, and Will Jennings. 2012. “Valence as Macro-Competence: An Analysis of Mood in Party Competence Evaluations in the U.K.” *British Journal of Political Science* 42(2):311-343.

MacKuen, Michael B., Robert S. Erikson, and James A. Stimson. 1989. “Macropartisanship.” *American Political Science Review* 83(4): 1125-1142.

Stimson, James A. 1991. *Public Opinion in America: Moods, Cycles, and Swings*. Boulder: Westview.