Bias and Variance in Multiparty Election Polls

Peter Selb, Sina Chen, John Körtner, Philipp Bosch

June 9, 2021

Abstract

Prominent polling failures highlight that election polls are prone to biases that the “margin of error” customarily reported with polls does not capture. However, such systematic errors are difficult to assess against the background noise of sampling variance. Shirani-Mehr et al. (2018) developed a hierarchical Bayesian model to disentangle random and systematic errors in poll estimates of two-party vote shares at the election level. We adapt the approach to multiparty elections and estimate variance and bias in more than 4,000 German national and regional election polls 1990–2021. Our analysis suggests that the average absolute election-level bias was about 1.8 p.p. in national and 2.8 p.p. in regional election polls. These biases tended to decrease as election day approached. We find little evidence of “house effects,” that is, tendencies of polling institutes to favor specific parties across elections. Common biases across institutes and little variance in excess of that implied by the standard margin of error may indicate “industry effects” due to similar polling methods and herding. Analyses like ours can inform realistic assessments of poll accuracy.
1 Introduction

Investigations into alleged polling misses such as the 2016 US presidential race (Kennedy et al., 2018) or the 2015 UK general election (Sturgis et al., 2018) document the many error sources that can bias polls: unrepresentative samples, failures to adjust for coverage and nonresponse problems, misspecified likely voter models, and misreporting of (or late swings in) vote intentions. The margin of error that is regularly reported to convey a poll’s uncertainty does not reflect any of these error sources. It solely captures the random fluctuation of an estimator across hypothetical replications of the sampling process. Statistical theory and information about the sampling design suffice to determine this type of error from a single probability sample. Assessing non-sampling errors can be complex as it requires validated records, population benchmarks, and possibly empirical replications of the sampling process. Election results provide the information necessary to measure the total survey error (TSE), which is the overall difference between a survey estimate \( p \) and its underlying population parameter \( P \) (say, a party vote share), \( TSE(p, P) = p - P \). While the ability to observe total errors after an election is a considerable methodological advantage of election polls over many other surveys (normally, the population parameter of interest remains unknown – learning about it is the reason why we are conducting surveys after all), it still does not allow us to distinguish between sampling variance and the aforementioned biases in a single poll. As Figure 1 illustrates, both variance and bias are defined in reference to the sampling distribution of an estimator. Bias corresponds to the difference between the expected value of the estimator and the underlying population parameter, \( E[p] - P \), and variance is defined as the average squared deviation between the estimates and their expected value, \( E[(p - E[p])^2] \).
Figure 1: Bias-variance decomposition of survey errors. $P$ is the population parameter of interest, $p_j$ represents estimates from four replications of the sampling process, TSE($p_1, P$) is the total error in the first estimate, $E[p]$ is the expected value, and Var($p$) is the variance of survey estimates across replications. Finally, Bias($p, P$) is the bias in the estimator with respect to $P$.

The proliferation of election polling over the past decades has created perhaps the only opportunity in the realm of survey research which brings us close to observing the sampling distribution of estimates from replications of approximately the same sampling process. It thus advances us towards identifying the bias-variance decomposition at the election level at which $P$ is observed. Most multi-survey studies of polling errors only look at the TSE or absolute error at the survey level, or the mean squared error (MSE) or mean absolute error (MAE) at the election level, but they do not decompose the error. See, among others, Crespi (1988) for US election polls, Sanders (2003) for Britain, Groß (2010) for Germany, Magalhães (2005) for Portugal, and Jennings and Wlezien (2018) for an epic study of more than 30,000 polls from 351 elections in 45 countries, 1942-2017. An early exception is Buchanan (1986), who analyzed 155 polls covering 68 elections from nine countries. Due to the small numbers of polls per election (2.3 on average), Buchanan limited his analysis to the top two competitors in each election and pooled all 155 two-party estimates as if they came from the same sampling distribution. Buchanan found poll biases to the detriment of conservative parties and variances more than twice as large as what sampling theory implies for simple random samples (SRS). Schnell and Noack (2014) in their study of 145 German Bundestag election polls 1957-2013 refined the margin of error for each poll to account for multiple parties and
design deviations from SRS. While the authors interpreted empirical coverage probabilities well below their nominal levels as indications of bias, they did not directly estimate election-level bias (nor variance, for that matter). Rather than completely pooling the polls or looking at each poll separately, Shirani-Mehr et al. (2018) take advantage of the multilevel data structure of polls nested in elections. They develop a Bayesian statistical model to estimate election-level variance and bias in polled two-party support, “borrowing strength” across all polls in all elections. In their empirical analysis of 4,221 polls from 608 US state-level elections between 1998 and 2014, the authors find an average election-level bias of about two percentage points and average election-level variance in excess of that implied by SRS.

Analyses like these are invaluable for making realistic judgments on the accuracy of polls and similar surveys. To extend the method’s scope beyond the two-party context of US elections, however, it needs to accommodate multiple parties. In the following section, we discuss the Shirani-Mehr et al. model in more detail before adapting it to multiparty elections. We then describe our data, which includes 1,102 polls from eight elections to the German national parliament (Bundestag) 1990-2017, as well as 2,096 polls on 68 regional parliamentary (Landtag) elections 1994-2021. Next, we present empirical results. The final section discusses the relevance of the approach for election polling and how it is reported to the public.

2 A model of polling errors in multiparty elections

Shirani-Mehr et al. (2018) model the two-party Republican vote share \( p_j = \frac{n_j^\text{Rep}}{n_j^\text{Rep} + n_j^\text{Dem}} \) as measured by poll \( j \) as a random draw from a normal distribution with mean \( \pi_j \) and variance \( \sigma_j^2 \) in the following manner:
\[ p_j \sim \text{Normal}(\pi_j, \sigma_j^2), \]  \hspace{1cm} (1)  
\[
\logit(\pi_j) = \logit(P_{r[j]} + \alpha_{r[j]} + \beta_{r[j]}t_j, \nabla^2_j = \frac{\pi_j(1 - \pi_j)}{n_j} + \phi_{r[j]}^2.
\]

That is, (the logit of) \( \pi_j \) is the sum of (the logit of) the actual two-party vote for the Republicans in election \( r \), \( P_{r[j]} \) and the bias of poll \( j \) with respect to population parameter \( P \), \( \alpha_{r[j]} \). The mean equation also contains a term that measures the temporal distance to the election, \( t_j \), to account for accuracy gains in polls as election day approaches. The logit scale ensures that estimated vote shares are bound between zero and one. The variance \( \sigma_j^2 \) is composed of first, the analytic sampling variance of a binomial proportion under SRS, \( (\pi_j(1 - \pi_j))/n_j \), where \( n_j \) is the valid sample size, and second, excess variance, \( \phi_{r[j]}^2 \), due to random measurement error, clustering, and other statistical inefficiencies in the design and analysis of surveys (see \cite{Frankel2010} for an overview of sampling design features).

For the case of multiparty elections, we introduce \( k \) to denote parties. Each poll \( j \) provides an estimate of each party \( k \)'s vote share, \( p_{jk} \). Analogous to Shirani-Mehr et al., we model the vote share in poll \( j \) as a random draw from a normal distribution, but for each party \( k \) separately, with mean \( \pi_{kj} \) and variance \( \sigma_{kj}^2 \):

\[ p_{kj} \sim \text{Normal}(\pi_{kj}, \sigma_{kj}^2), \]  \hspace{1cm} (2)  
\[
\logit(\pi_{kj}) = \logit(P_{r[kj]} + \alpha_{1kr[kj]} + \alpha_{2kl[kj]} + \beta_{1kr[kj]}t_j + \beta_{2kr[kj]}t_j^2, \nabla(\sigma_{kj}^2) = \log\left(\frac{\pi_{kj}(1 - \pi_{kj})}{n_j}\right) + \phi_{r[kj]}.
\]

The logit of \( \pi_{kj} \) equals the sum of the logit of the actual election result for party \( k \) in election \( r \), \( P_{kr[kj]} \), a party-election bias component \( \alpha_{1kr[kj]} \), and another bias component, \( \alpha_{2kl[kj]} \), which we introduce to identify party-specific “house effects” due to differences in
survey methods between polling institutes \(l\) (e.g., \citeauthor{Jackman2005}). Like \citeauthor{Shirani-Mehr et al.2018}, we include the temporal distance to election day, \(t_j\), as a predictor. Due to the marked curvilinear patterns observed in German Bundestag election polls (see Figure 2), however, we also add \(t_j^2\) for greater flexibility.

We model the party-specific total variance \(\sigma_{kj}^2\) as the sum of the analytic sampling variance of a multinomial proportion under SRS, \((\pi_{kj}(1 - \pi_{kj}))/n_j\), and excess variance \(\phi_{kr[kj]}\). In contrast to \citeauthor{Shirani-Mehr et al.2018}, we model the variance on the log scale. This adjustment allows the excess variance \(\phi_{kr[kj]}\) to be positive or negative (which can be achieved by efficient sampling schemes involving auxiliary information, see \citeauthor{Frankel2010}), while ensuring that the overall variance \(\sigma_{kj}^2\) is strictly positive. To account for the negative covariance of multinomial proportions, \(\alpha_{1kr[kj]}, \alpha_{2kr[kj]}, \beta_{1kr[kj]}, \beta_{2kr[kj]}\), and \(\phi_{kr[kj]}\) are given a multivariate normal distribution. The parameters are modeled as follows: \(\alpha_{1r} \sim \mathcal{N}_k(0, \Sigma^{\alpha_1}), \alpha_{2l} \sim \mathcal{N}_k(0, \Sigma^{\alpha_2}), \beta_{1r} \sim \mathcal{N}_k(0, \Sigma^{\beta_1}), \beta_{2r} \sim \mathcal{N}_k(0, \Sigma^{\beta_2}), \phi_r \sim \mathcal{N}_k(0, \Sigma^\phi)\).

For computational efficiency, priors on the covariance matrix \(\Sigma\) of each parameter are obtained by combining priors on the \(k \times k\) correlation matrix \(\Omega\) with a vector of priors on the standard deviations \(s_k\) of each component, which refers in the given context to the single parties, \(k\) \cite{Barnard, McCulloch and Meng 2000}: \(\Sigma = \text{diag}(s_k) \times \Omega \times \text{diag}(s_k)\), where \(\text{diag}(s_k)\) represents the diagonal matrix with diagonal elements \(s_k\). Instead of assigning a prior directly on the distribution of \(\Omega\), it is more efficient to decompose the correlation matrix into \(\Omega = L \times L'\) with \(L\) being the lower triangular matrix, also known as the Cholesky factor \cite{Stan Development Team 2020}. Now one can assign a prior on \(L\): \(L \sim \text{LKJCorr}(2)\).

For the standard deviations \(s_k\), weakly informative priors are chosen which allow for a five percentage point change in poll bias or variance (which corresponds to 0.2 on the logit scale and 0.05 on the log scale) \cite{Shirani-Mehr et al.2018}: \(s_k \sim \mathcal{N}_+(0, 0.2^2)\) for all parameters separately in the mean equation and \(s_k \sim \mathcal{N}_+(0, 0.05^2)\) for the excess variance \(\phi\). A half normal distribution is chosen for variance parameters to also allow values near 0 \cite{Shirani-Mehr et al.2018}. Details are given in Table 2 in the Supplementary Materials.
3 Data

Election polls have been conducted in the Federal Republic of Germany since the inaugural election of the Bundestag in 1949 (Gross 2010). However, the number of polls (and polling firms) was limited during the first decades and the party system was subject to some major changes (Zittel 2018). Our cross-election perspective requires some continuity, therefore we limit the main empirical analysis to the eight Bundestag elections after re-unification in 1990. We consider five major parties: the Christian Democrats (CDU/CSU), the Social Democrats (SPD), the Liberal Democrats (FDP), the Greens (B90/Grüne), and the Left (Die Linke). Other parties are lumped into a residual category. The Supplementary Materials contain separate analyses of the 2013 and 2017 Bundestag elections in which another major party, the populist Alternative für Deutschland (AfD), emerged. The Supplementary Materials report additional results for 2,096 polls on 68 regional (Landtag) elections from 1994 to 2021.

Figure 2: Results from 2,157 polls by party, Bundestag elections 1990-2017. Dashed horizontal lines indicate actual party vote shares. Dashed vertical lines indicate January 1st in each election year.
We scraped the polling data from **wahlrecht.de**, an independent website on elections, electoral rules, and voting rights in Germany, which provides a real-time collection of published vote intention surveys from eight prominent polling firms (IfD Allensbach, TNS Emnid/Kantar, Forsa, Forschungsgruppe Wahlen, GMS, Infratest dimap, INSA, YouGov), fully covering the five national election campaigns between 2002 and 2017. Earlier polling data were kindly provided by Jochen Groß (2010) through Simon Munzert and his colleagues (Stoetzer et al., 2019).

Bundestag elections are normally held every four years at the end of September. The main analysis is limited to polls published during election years, as campaigns gather pace and the density of polls increases (see the dashed vertical lines in Figure 2, which mark January 1 in each election year). Polls without publication date are excluded (11 polls). If information on the sample size was missing, the institute’s mean sample size was imputed (251 polls). A total of 1,102 polls is included in the main analysis, ranging from only 16 for the 1990 election to 192 in 2013, with an average of 138 per election. Figure 3 provides an exploratory view of the distribution of total errors. For comparison, we simulate a polling result for each actual poll \( j \) by drawing a SRS of size \( n_j \) and party vote share \( p_{jk} \) in the corresponding election \( r \). The dashed curves represent these theoretical distributions of polling errors assuming SRS. The plots highlight two points. First, not all of the empirical distributions seem to be exactly centered at zero. On average, the polls tend to overestimate the larger parties (CDU/CSU: 1.8 p.p.; SPD: 0.4 p.p., Greens: 0.6 p.p.) while the vote shares of smaller parties are underestimated (FDP: -0.9, Left: -0.4, Other: -1.4). Given the moderate number of elections included (8), however, one should not overinterpret this result. Second, the observed election-level variance of the polls about their mean is substantially larger than expected under SRS, as indicated by the wider tails of the empirical distributions in comparison to the theoretical distributions. A detailed analysis of election-level biases and variances based on the statistical model developed above follows in the next section.
Figure 3: Distribution of total survey errors by party, Bundestag elections 1990-2017. For comparison, the dashed curves show the theoretical distributions of survey errors assuming simple random sampling.

4 Empirical results

Figure 4 shows the estimated average election-day biases by party and Bundestag election 1990-2017. *Election-day* bias means that the temporal distance to election day, \( t_j \), in Equation 2 is set to zero, thereby accounting for possible pre-election trends in polling. With the exception of the 1990 and 2013 Bundestag elections, the CDU/CSU has been regularly overestimated in the polls, with peaks occurring in the 2005 and 2017 Bundestag elections. By contrast, the minor parties grouped in the “others” category have been consistently underestimated. A detailed analysis of the 2013 and 2017 Bundestag elections suggests that at least in 2013 polling, this bias was largely driven by the then new Alternative für Deutschland (see Figure 8 in the Supplementary Materials). For the remaining parties, the patterns have been more nuanced.

Figure 5 presents possible party-institute effects relative to the election-day poll average.
(i.e., *house effects*). While every poll firm except IfD Allensbach and INSA has overestimated the CDU/CSU to some extent, the magnitude of the average bias has varied across institutes. The seemingly stark house effects in favor of the SPD pertain to relatively young institutes – INSA (active since the 2013 election campaign) and YouGov (2017 only; thus the wider credible intervals) – and thus need to be treated with caution. Other than that, party-specific biases tend to occur consistently across institutes and thus indicate *industry bias* (e.g., due to common methods or herding) rather than house effects.

![Figure 4: Estimated average party election-level bias on election day. Negative values indicate that polls, on average, underestimated election results and vice versa. Horizontal lines represent 95% and 50% credible intervals.](image-url)
Figure 5: Estimated average party-institute biases across elections. Negative values indicate that an institute underestimated a party’s vote share relative to the election-day poll average. Horizontal lines represent 95% and 50% credible intervals.

Figure 6 plots estimates of average total standard errors relative to analytic standard errors assuming SRS. With few exceptions, the estimated average total standard errors are relatively inconspicuous and suggest that, most of the time, polls have not been much more variable than expected under SRS.

Finally, Table 1 provides summary statistics of estimated average absolute biases and total and excess standard errors by party (see Section A in the Supplementary Materials for formal descriptions of the errors.). Estimates of average election-level absolute bias over the election year vary from about 1 p.p. for the LINKE to 3 p.p. for the SPD in Bundestag election polls and from 1.8 p.p. for the GRUENE to 4 p.p. for the CDU/CSU in Landtag election polls. These biases tend to decrease as election day approaches (see the second row of the tables). The estimated average total standard error at the election level tends to be somewhat higher than what would be expected under SRS by 15 to 40%.
Figure 6: Estimates of average election-level total standard error relative to SRS standard error.

### Table 1: Mean posterior estimates of average absolute bias, average absolute election-day bias, average total standard error, and average excess standard error in Bundestag election polls 1990-2017, and Landtag election polls 1994-2021 (in p.p.). The standard deviation of the posterior is given in parentheses. See section A for formal descriptions of errors.
5 Conclusion

In this paper, we extended the scope of Shirani-Mehr et al.’s (2018) method of disentangling variance and bias in election polls to accommodate multiple parties. Our empirical analysis of German election polls 1990-2021 largely resonates with what Shirani-Mehr et al. have found for US elections 1998-2014: the average absolute election-level bias was about 1.8 p.p. in national and 2.8 p.p. in regional election polls. We also looked for house effects but found mostly consistent party-specific biases across polling institutes. Contrary to Shirani-Mehr et al., we found little variance beyond what would be expected based on a simple random sample for most parties in most elections. Common biases across institutes and little variance in excess of that implied by the standard margin of error may indicate “industry effects” due to similar polling methods and herding.

After the 2020 US Presidential election poll miss, academic pollsters have again questioned whether the margin of error is a useful metric, but a suitable alternative is not clear (Schulson, 2020). As Shirani-Mehr et al. (2018, 608) point out, there is little prospect of a general statistical theory of non-sampling errors to inform an alternative metric. In the absence of a convenient analytic measure, however, the convergence of empirical evidence from diverse electoral contexts is reassuring and suggests that there are regularities that can and should be taken into account by pollsters and journalists in realistic assessments of the accuracy of polls and similar surveys.

Assessing bias and variance in past polls alone does not help to predict the magnitude and direction of errors in current polls. The modelling approach advocated here is easily extended to incorporate predictor variables at various levels to account for contextual correlates of polling misses. Such an extension could thus prospectively identify elections in which substantial polling errors are likely to occur (e.g. Crespi, 1988; Jennings and Wlezien, 2018).

Besides that application, the ability to distinguish between variance and bias in election polling is crucial to improve election forecasts. If survey errors were in large part random
we could try to reduce the variance of estimates, for instance, by increasing sample sizes. This principle is the basic idea underlying poll aggregators which average over many polls to forecast election outcomes, effectively increasing sample size and reducing sampling variance compared to estimates based on any individual poll covered (for an overview, see Jackson, 2018). If survey errors are systematic, however, little can be gained from poll aggregation other than over-confidence in biased estimates.
References


**URL:** https://undark.org/2020/11/23/polls-margin-of-error-gets-new-scrutiny/


**URL:** http://mc-stan.org/


Zittel, Thomas. 2018. Electoral systems in context: Germany. In *The Oxford Handbook of*
Supplementary Materials

A Formal descriptions of errors

In this section, we provide formal descriptions of the quantities based on the model in Equation 2 which are reported in Table 1 and 3.

**Average absolute bias** at the election level for party $k$ is defined as

$$\mu_{b_k} = \frac{1}{R} \sum_{r=1}^{R} |b_{kr}|,$$

with $R$ representing the total number of elections. $b_{kr}$ is the bias for party $k$ in election $r$,

$$b_{kr} = \frac{1}{S_{kr}} \sum_{j \in S_{kr}} (p_{kj} - P_{kr}),$$

with $S_{kr}$ being the set of polls for party $k$ in election $r$. Accordingly, the **average absolute election-day bias** for party $k$ is obtained by:

$$\mu_{b0_k} = \frac{1}{R} \sum_{r=1}^{R} |b_{0kr}|,$$

where the election-day bias for party $k$ in election $r$ is defined as

$$b_{0kr} = \frac{1}{S_{kr}} \sum_{j \in S_{kr}} (p_{0kj} - P_{kr}).$$

$p_{0kj}$ is estimated by setting the temporal distance between the poll and election day $t_j$ to zero:

$$\logit(p_{0kr}) = \logit(P_{kr}) + \alpha_1_{kr} + \alpha_2_{kl}.$$
The **average total standard deviation** at the election level for party $k$ is defined as

$$\mu_{\sigma_k} = \frac{1}{R} \sum_{r=1}^{R} |\sigma_{kr}|,$$

where $\sigma_{kr}$ is the average party-election standard deviation obtained by:

$$\sigma_{kr} = \frac{1}{S_{kr}} \sum_{j \in S_{kr}} \sigma_{kj}.$$  

Finally, we define the **average excess standard deviation** as

$$\mu_{ex_k} = \frac{1}{R} \sum_{r=1}^{R} |ex_{kr}|.$$  

$ex_{kr}$ represents the party-election level excess standard deviation, which is the average over all polls $j$ for party $k$ in election $r$ of the square root of the difference between the total estimated variance $\sigma^2_{kj}$ and the analytic sampling variance of a multinomial proportion under SRS, $(\pi_{jk}(1-\pi_{jk}))/n_j$. In our case this can be defined as:

$$\log(ex_{kr}) = \sqrt{[\log(\pi_{jk}(1-\pi_{jk}))/n_j + \phi_{kr}] - \log(\pi_{jk}(1-\pi_{jk}))/n_j} = \sqrt{\phi_{kr}}.$$  

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Hyperprior</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>$L^{\alpha_1} \sim \text{LKJCorr}(2)$</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>$L^{\alpha_2} \sim \text{LKJCorr}(2)$</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>$L^{\beta_1} \sim \text{LKJCorr}(2)$</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>$L^{\beta_2} \sim \text{LKJCorr}(2)$</td>
</tr>
<tr>
<td>$\phi$</td>
<td>$L^{\phi} \sim \text{LKJCorr}(2)$</td>
</tr>
</tbody>
</table>

Table 2: Hyperpriors for covariance matrices.
B 2013 and 2017 Bundestag elections: Detailed results

In this section, we provide detailed results for the 2013 and 2017 Bundestag elections, reporting bias and variance estimates for the Alternative für Deutschland (AfD) separately from the “others” category which we formed in the main text to facilitate cross-election analyses. Founded in February 2013, the AfD narrowly failed to clear the 5 percent threshold with 4.7 percent in the Bundestag elections in September 2013. After several successful Landtag (state) elections, the AfD entered the Bundestag in the 2017 election with 12.6 percent of the vote and has since gained representation in all regional parliaments, with vote shares ranging from 5.9% (Schleswig-Holstein 2017) to 27.5% (Saxony 2019). Figure 7 tracks 1,691 polls over the election cycles 2009-2013 and 2013-2017, with the “others” category from the main text broken down to distinguish the AfD from minor parties. Figure 10 gives empirical versus theoretical distributions of total polling errors. Clearly, the polls systematically underestimated the AfD both in 2013 and 2017. This underestimation is a frequently observed phenomenon with right-wing parties and candidates and is often ascribed to socially desirable survey responses and selective participation in polls (e.g., Kennedy et al. 2018).

According to our statistical model, the average election-day bias for the AfD is -1.87% in 2013 and -0.14% in 2017, while the election-level variances are estimated to be somewhat greater than what SRS theory would have one expect. Due to otherwise sparse data, the model is fitted to all the polls, not just those conducted during election years, and it includes a linear time trend instead of the quadratic term to improve convergence.

<table>
<thead>
<tr>
<th></th>
<th>CDU/CSU</th>
<th>SPD</th>
<th>GRUENE</th>
<th>FDP</th>
<th>LINKE</th>
<th>AfD</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average abs. bias ($\hat{\mu}_{bk}$)</td>
<td>5.50 (0.08)</td>
<td>2.32 (0.08)</td>
<td>4.08 (0.09)</td>
<td>3.03 (0.05)</td>
<td>0.11 (0.03)</td>
<td>3.26 (0.07)</td>
<td>0.95 (0.06)</td>
</tr>
<tr>
<td>Av. abs. election-day bias ($\hat{\mu}_{b0}$)</td>
<td>0.93 (1.18)</td>
<td>2.64 (0.46)</td>
<td>3.14 (0.60)</td>
<td>1.57 (0.36)</td>
<td>1.31 (0.20)</td>
<td>1.01 (0.26)</td>
<td>1.82 (0.25)</td>
</tr>
<tr>
<td>Av. total standard error ($\hat{\mu}_{\sigma}$)</td>
<td>2.48 (0.06)</td>
<td>2.36 (0.06)</td>
<td>2.56 (0.06)</td>
<td>1.39 (0.03)</td>
<td>1.00 (0.02)</td>
<td>1.48 (0.05)</td>
<td>1.69 (0.04)</td>
</tr>
<tr>
<td>Av. excess standard error ($\hat{\mu}_{exx}$)</td>
<td>1.35 (0.06)</td>
<td>1.34 (0.06)</td>
<td>1.78 (0.06)</td>
<td>0.88 (0.03)</td>
<td>0.34 (0.02)</td>
<td>0.97 (0.05)</td>
<td>1.11 (0.04)</td>
</tr>
</tbody>
</table>

Table 3: Mean posterior estimates of average absolute bias, average absolute election-day bias, average total standard error, and average excess standard error for Bundestag election polls, 2013 and 2017 (in p.p.). The standard deviation of the posterior is given in parentheses. See section A for formal descriptions of errors.
Figure 7: Election-year poll results by party, Bundestag elections 2013-2017. Dashed horizontal lines indicate actual party vote shares.

Figure 8: Average party-election level bias on election day. Negative values indicate that polls, on average, underestimated election results and vice versa. Horizontal lines represent 95% and 50% credible intervals.

C Regional (Landtag) election polling, 1994-2021.

Germany consists of 16 federal states (Bundesländer), each of which conducts parliamentary (Landtag) elections every four or five years. In this section we provide additional results for 2,096 polls in 68 regional (Landtag) elections from 1994 to 2021. The data were retrieved from wahlrecht.de (cutoff date: 2021-03-19). For the sake of comparability across elections, we restrict our analysis to polls which cover the five major parties already considered in the main text. This method led to the exclusion of 15 elections (273 polls). As mentioned
Figure 9: House effects across elections. Negative values indicate that an institute underestimated a party’s vote share relative to the election-day poll average. Horizontal lines represent 95% and 50% credible intervals.

Figure 10: Distribution of total survey errors by party, Bundestag elections 2013 and 2017. For comparison, the dashed curves show the theoretical distribution of survey errors assuming simple random sampling.
above, if information on the sample size was missing, the institute’s mean sample size was imputed (1,017 polls). If there was no information on the sample size for all polls conducted by an institute, 1,000 was imputed (60 polls). Furthermore, we do not model house effects, since there were over 70 institutes active in polling Landtag elections, with few overlaps between states. As in the previous section, we include all the polls conducted during the observation period to increase statistical efficiency in the face of increased numbers of groups / parameters. Time to election is specified quadratically in the model. Figure 11 gives average election-day bias estimates for each party in each election included in the analysis. Overall, absolute bias amounts to an average of 2.8% across parties and elections, ranging from 1.8% for the GRUENE to 4% for the CDU/CSU.
Figure 11: Average party-election level bias on election day for German Landtag election polls. Negative values indicate that the polls on average underestimated a party’s vote share and vice versa. Horizontal lines represent 95% and 50% credible intervals.