Online Appendix Forecasting Elections in Multi-Party Systems: A Bayesian Approach Combining Polls and Fundamentals*

Lukas F. Stoetzer†1, Marcel Neunhoeffer‡2, Thomas Gschwend§2, Simon Munzert¶3, and Sebastian Sternberg‖2

[1]University of Zurich
[2]University of Mannheim
[3]Hertie School of Governance

Contents

*NOTE: Replication materials are available online as a dataverse repository (Neunhoeffer et al., 2018) https://doi.org/10.7910/DVN/MLYNXO.

†Corresponding Author: PostDoc, Department of Political Science, University of Zurich, Affolternstrasse 56, 8050 Zürich, Switzerland (lukas.stoetzer@uzh.ch).

‡PhD Candidate, Department of Political Science, University of Mannheim, A 5, 6, 68131 Mannheim, Germany (marcel.neunhoeffer@gess.uni-mannheim.de).

§Professor, Department of Political Science, University of Mannheim, A 5, 6, 68131 Mannheim, Germany (gschwend@uni-mannheim.de).

¶Lecturer, Hertie School of Governance, Friedrichstr. 180, 10117 Berlin, Germany (munzert@hertie-school.org).

‖PhD Candidate, Department of Political Science, University of Mannheim, A 5, 6, 68131 Mannheim, Germany (sebastian.sternberg@gess.uni-mannheim.de).

A A Fundamentals-based Model to Forecast Multi-Party Elections

While we introduce in the main body of the text merely the idea that we also develop a fundamentals-based model, in this section we provide more details how we specify and implement it.[1] We start with describing the regression setup and later discuss the predictors used in our study.

A.1 Dirichlet Regression Model

We consider vote shares v_{pe} of party p $(= 1, \dots, P)$ at election e $(= 1, \dots, E)$, where election E is the upcoming election which we intend to forecast.[2] Our goal is to use the information from previous elections to predict v_{pE} for all parties that compete in the upcoming election. As vote shares for each election sum to one, we model them using a Dirichlet distribution with parameters $\boldsymbol{a}_e = (a_{1e}, ..., a_{Pe})$. Such a stochastic component assures that the vote shares for each election confirm to the simplex constraint, lie on the $0-1$ interval and sum up to 1. The vector of vote shares $\boldsymbol{v}_e = [v_{1e}, \dots, v_{Pe}]$ for all parties in election e is

$$\boldsymbol{v}_e \sim \mathcal{D}(\boldsymbol{a}_e), \tag{1}$$

while the systematic component models the log of the $\boldsymbol{\alpha}_e$ shape parameters as a linear-additive function of covariates x^k_{pe}

$$\alpha_{pe} = exp\left(\beta^0_e + \sum_{k=1}^{K} \beta^k_e x^k_{pe}\right). \tag{2}$$

[1]Fundamentals-based models have distinct advantages when predicting the outcome of elections (Lewis-Beck, 2005). First, while polls tend to exhibit relatively large forecasting variance when election day is still far away (Jennings and Wlezien, 2016), fundamentals-based models are more reliable early in the campaign cycle (Lewis-Beck and Dassonneville, 2015*a*,*b*). Second, they put current elections in a historical context, which helps build expectations about how special a particular election and its campaign really is. In contrast to many election observers who look merely how the current campaign plays out, fundamentals-based models allow us to learn from regularities across many elections and leverage them to forecast and explain the outcome of an upcoming election.

[2]In many applications, the number of parties will vary across elections. Out of notational convince we ignore this issue here. It would be straightforward to make the number of parties election specific P_e.

Often, fundamentals-based forecasting models are based on a time frame that includes elections from over 70 years. We propose that rather than assuming the same data-generating process across all these elections, the effect of different predictors should allow to vary over time. For example, it is well known that the German electorate is increasingly less partisan than it used to be (Arzheimer, 2006). If this dynamic process is supported by our data, we should expect the effect of long-term factors to decrease over time, and conversely, the effect of short-term factors to increase over time. In order to account for that, we allow the parameters of Equation 2 to vary across elections. Thus, we employ a hierarchical specification

$$\beta_e^k \sim N(\tilde{\beta_e^k}, \tau_k^2) \tag{3}$$

while we allow any parameter at election e to be a draw from a normal distribution with a mean that comprises an additive combination of the previous parameter and a drift parameter, i.e.:

$$\tilde{\beta_e^k} = \beta_{e-1}^k + \gamma_{\text{drift}}^k \tag{4}$$

Now we are in the position to derive both distributions introduced in the main text: The posterior distribution and the predictive posterior distribution for the upcoming election. The posterior distribution of our model requires us to specify priors about the parameters ($\boldsymbol{\theta}$). In our case those are the effect parameters β_e^k, the drift parameters γ_{drift}^k as well as the variance terms τ_k^2. We use independent prior distributions for for each covariates, thus $P(\boldsymbol{\theta}) = \prod_{k=0}^{K} P(\tau_k^2)P(\beta_1^k)P(\gamma_{\text{drift}}^k)$, where $\boldsymbol{\theta}$ is the set of all parameters in the model. The data in the likelihood of the election results are the past election results ($\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{E-1}]$) and an overall predictor matrix that collects the predictors for all elections ($\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{E-1}]$, where $\boldsymbol{x}_e = [x_{1e}, \ldots, x_{Pe}]$). Using this notation the posterior distribution can be written as introduced in the main text.

$$P(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{V}) \propto P(\boldsymbol{V}|\boldsymbol{\theta},\boldsymbol{X})P(\boldsymbol{\theta}) \tag{5}$$

The predictive posterior distribution then predicts the vote shares $\boldsymbol{v_E}$ for the upcoming election given the respective covariate values $\boldsymbol{x_E}$ while taking the uncertainty about the coefficients into account as follows:

$$P(\boldsymbol{v_E}|\boldsymbol{x_E}) = \int_{\theta} P(\boldsymbol{v_E}|\boldsymbol{\theta},\boldsymbol{x_E})P(\boldsymbol{\theta}|\boldsymbol{V},\boldsymbol{X})d\boldsymbol{\theta} \tag{6}$$

This forecast will not be viewed in isolation, though. Rather, it will serve as an anchor for our dynamic polls model by setting the log-ratio transformed values $\boldsymbol{v}_E$ equal to the latent state on Election Day.

A.2 Predictors in the fundamentals-based model

While any model that predicts each party's vote share could be used (e.g., Jérôme, 2013), most fundamentals-based models only forecast the vote share of governing parties (Magalhães, Aguiar-Conraria and Lewis-Beck, 2012; Kayser and Leininger, 2016; Norpoth and Gschwend, 2010). In addition, most models are tailored to the election at hand and not built to be applied in different contexts. For this project, we devised a *fundamentals-based model* that is general enough to be applied to multiple contexts. It builds on the idea that three core factors predict election outcomes: Long-term party attachment, short-term campaign dynamics, and institutional features.[3]

Long-term party attachment: Elections are not held in a political vacuum. It is well known that voters develop long-term stable attachments to political parties (Campbell et al., 1960). The distribution of such attachments in the aggregate allows us to form expectations about the outcome of a given election under normal circumstances (Converse, 1966). We operationalize such a normal-vote baseline as the party's vote share in the

[3]Economic performance, a factor that is central to forecast models in the U.S. (see Lewis-Beck and Stegmaier, 2000) and models of governing party vote share in multi-party elections (see e.g. Magalhães, Aguiar-Conraria and Lewis-Beck, 2012; Kayser and Leininger, 2016), is not included for two distinct reasons. First, economic accountability is arguably more ambiguous in multi-party governance, which particularly matters when modeling the vote share of multiple parties. Second, in our model most of the potential effect is already captured by the short-term campaign measurements.

previous election (which we set to '0' if the party competes for the first time).

Short-term campaign effects: Parties get support not only from their partisan base, but also from so-called undecided voters or even partisans of other parties. These voters might be motivated to support a different party by their preference for particular issues and/or candidates. The short-term campaign effects are captured using the average party support as published in polls available 230 to 200 days before the election.[4]

Institutional features: Our third predictor accounts for the fact that for every performance evaluation of the government, it is important which party leads the government. In parliamentary systems, this is the position of the chancellor or the prime minister. Credit and blame regarding the performance of the incumbent government most heavily registers with the support for the chancellor's party. The prime minister or chancellor is the most visible politician in government. We therefore construct an indicator variable scoring '1' for the party that holds the chancellorship.

[4]Selb and Munzert (2016) find that poll-based forecasting models generally perform better when using polls before the campaign.

B Application to the German Federal Elections

B.1 Fundamentals-based model

B.1.1 Data

To calibrate the fundamentals-based model, we leverage data on all 18 federal elections in Germany since 1949[5]. Until the 1976 election, we model vote shares of *CDU/CSU, SPD, FDP* and "*others*" (as combined share of all remaining competing parties). From 1980 on, we also model the vote shares of the *Greens* and since 1990 also the vote shares of the *Left Party* (originally *PDS*). Finally, the right-wing *AfD* is considered from 2013 onward. To build a comprehensive data-base of pre-election opinion polls, we rely on data initially collected by Groß (2010), later appended and made available by Schnell and Noack (2014).[6] For all polls published since 2009, we use data provided on the online platform `wahlrecht.de`.

[5] The result of the 1949 election is used as an indicator of long-term party identification, but is not part of the training set.

[6] Furthermore, we filled gaps in the time series with data made available by the polling company Allensbach.

B.1.2 Empirical relationship between predictors and motivation of predictors

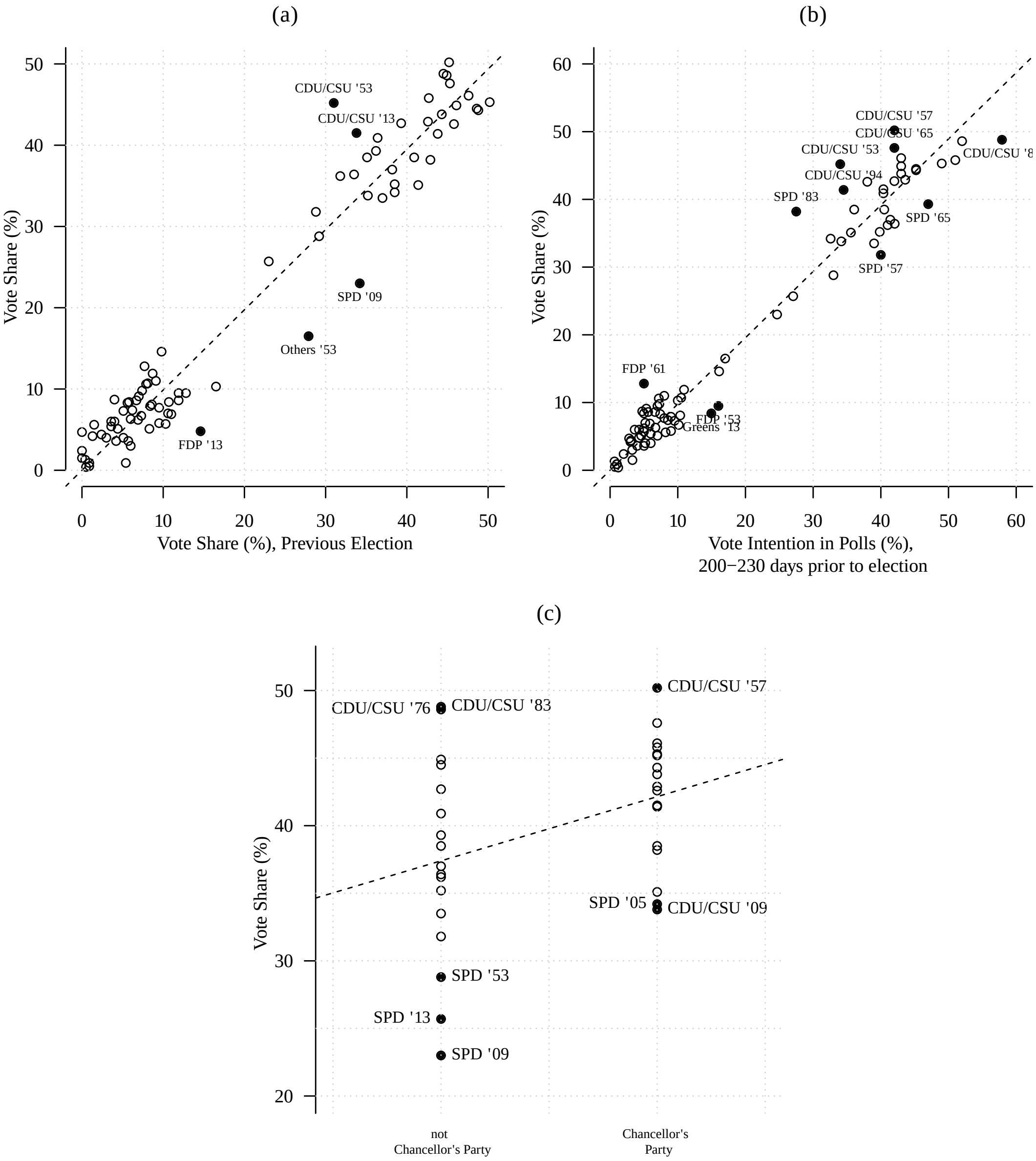

Figure 1: Relationship between predictors and vote share, 1953-2013.

Figure 1 shows the relationship between long-term effects, short-term and institutional features and the election outcomes. Panel (a) depicts the relationship between previous and current party vote shares across elections.[7] While this predictor clearly helps to separate small from large parties and also explains variation within these clusters, we can also see that our first predictor does not capture significant swings. Panel (b) in

[7]Kayser and Leininger (2017) use the same operationalization as predictor for their model while Gschwend and Norpoth's "chancellor model" operationalizes a party's normal-vote baseline as the average vote in the last three *Bundestag* elections (Norpoth and Gschwend, 2003, 2010, 2013).

Figure 1 shows that our second predictor, short term-effects, performs already quite well in predicting the actual vote shares on election day. Panel (c) in Figure 1 reveals that the party of the incumbent chancellor has, on average, a larger vote share than the respective other major party that does not hold the chancellorship (the small parties are not shown).[8]

B.1.3 Fundamentals-based model estimates and predictive performance

Figure 2 depicts the estimated β-coefficients for the last 17 elections. The pattern confirms our expectations: The effects of the predictors vary substantively across elections, and while the predictive importance of prior election results decreases over time, the polls become more predictive for the final outcome. The chancellor-party effect is not systematically different from 0. For the 2017 elections, we extrapolate the observed trends for all coefficients, given the estimates of the drift parameter and the random-walk component.

[8] The 1983 election is a special case because the party of the chancellor right before the election, the CDU/CSU, was not considered the incumbent that is to blame for the current situation. The SPD just lost the chancellorship a few months earlier through a reshuffling of the government. Similar to the coding strategy of the chancellor model (Gschwend and Norpoth, 2001), we therefore consider the SPD as incumbent party of the chancellor for the 1983 election.

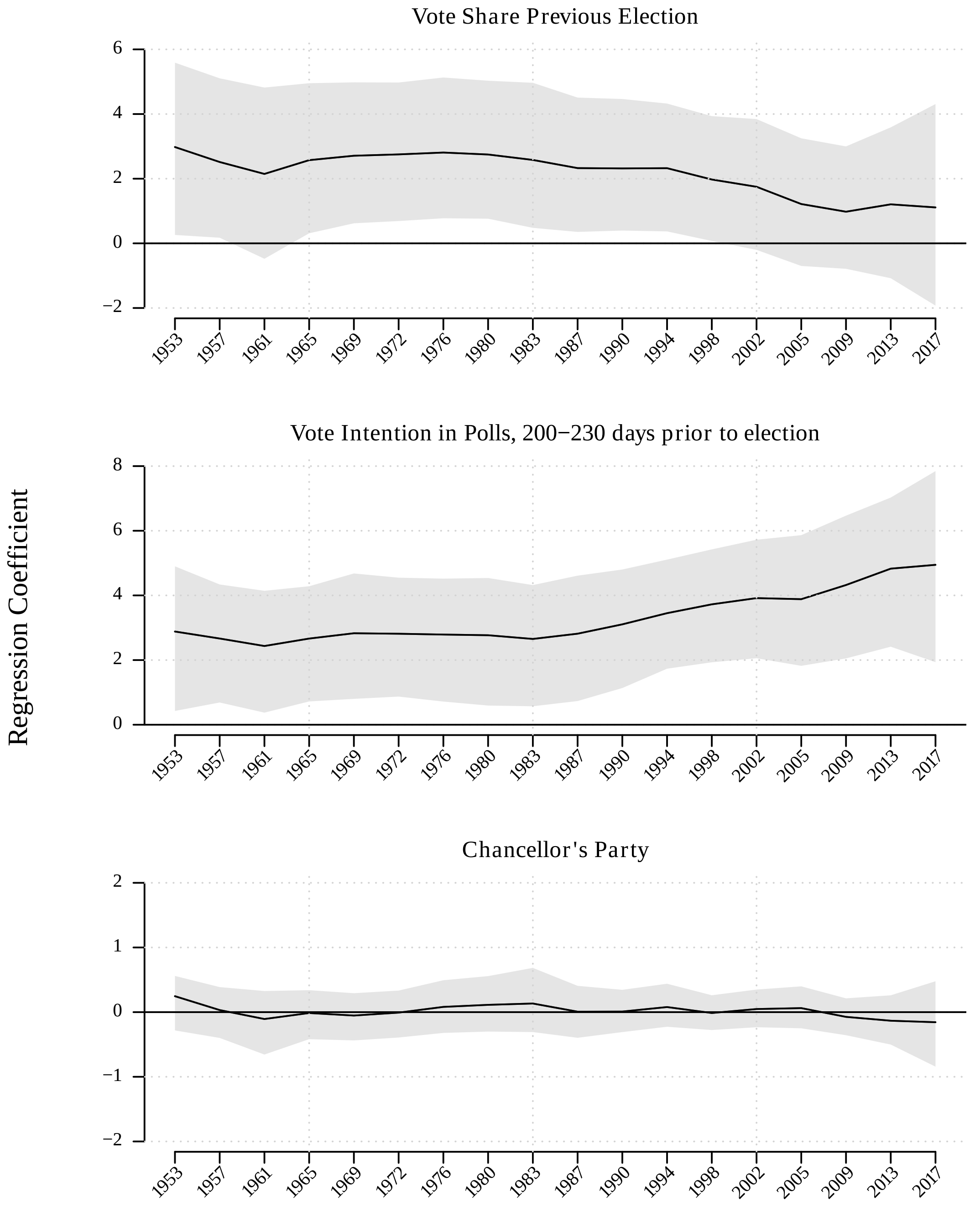

Figure 2: Coefficients for fundamentals-based forecasting model

B.2 Dynamic Bayesian measurement model for German federal elections

When using the dynamic Bayesian forecasting model, the forecast is sequentially updated. We use the polling results of the SPD during the 2017 Federal election campaign to illustrate how the anchoring process in the backwards random walk works. Figure 3 shows the results of the dynamic forecasting model for the SPD vote share over time, starting 148 days before the election. Using this example, we want to highlight two important features of the model. First, we see that the uncertainty about the SPD vote share considerably declines over time. 148 days before the election, the 95% credible interval reaches from 21.7% to 37.6%, whereas it is only between 20.5% and 24.5% eight days before the election.

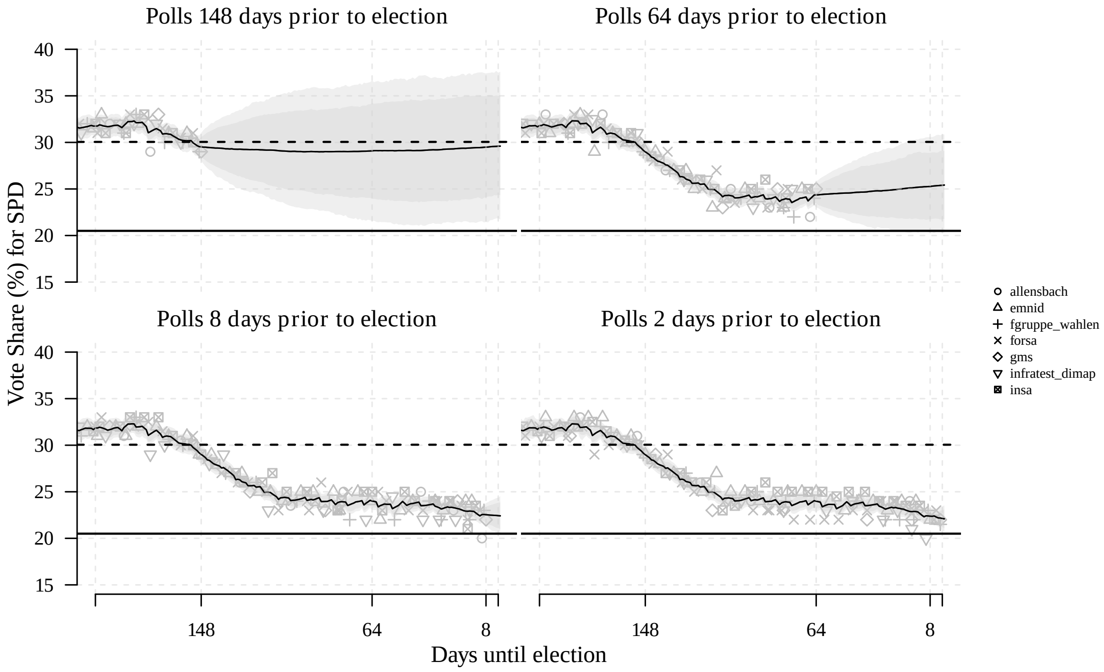

Figure 3: SPD vote share 2017 prediction based on the dynamic Bayesian forecasting model. The symbols represent the party supported reported in the respective polls. The solid line depicts the median latent SPD support of the posterior distribution; the shadowed area depicts the $\frac{5}{6}$ and 95% credible intervals. The observed 2017 SPD vote share is indicated by the solid horizontal line (20.5%), and the forecast of the fundamentals-based model is marked by the dashed horizontal line (30.1%).

Second, Figure 3 illustrates how the model's weight on the fundamentals-based forecast diminishes over time. With much time to go until the election at the beginning of the electoral campaign, the model puts more weight on forecast of the fundamentals-based model. As the election comes closer and more polls become available, the polling-based component becomes more influential. Consider the two panels in the upper part of the figure, 148 and 64 days before the election. Here, the predicted vote share for the SPD slowly approaches the horizontal dashed line that indicates the fundamentals-based forecast. In contrast, eight days and two days before the election (see the lower panels), the model diverges from the fundamental-based forecast and picks up the tendency of the reported support from the polls. This is a desired behavior of the model because the fundamentals-based model initially provides much information about the final election outcome, whereas polls become more accurate over time and are thus considered more and more important.[9]

B.3 Model Evaluation

In this section, we evaluate different model parts. We first compare the forecasts from our combined model to the forecasts from a pure fundamental-based model for the last five German federal elections. Next, we evaluate the 2017 model forecasts to two simpler specifications, without house effects and with a specification that only includes past vote share as a predictor. Finally, we take a look at the estimated evolution variances in the 2017 elections.

[9]The effect of this trade-off is especially strong for smaller parties, because the fundamentals-based model provides more accurate forecasts for smaller parties.

Table 1: The RMSE of the out-of-sample predictions by data basis and election year. Pure fundamentals-based model and dynamic Bayesian forecasting model arranged by time to election.

Model	RMSE					
	2002-17	2002	2005	2009	2013	2017
Pure Fundamentals-Based	3.88	2.29	4.66	4.28	3.50	4.22
Dynamic. 2 days prior to election	1.79	1.27	2.96	0.92	1.20	1.88
Dynamic. 8 days prior to election	1.95	0.92	3.12	1.12	1.67	2.13
Dynamic. 36 days prior to election	2.88	2.03	4.34	1.52	2.33	3.29
Dynamic. 64 days prior to election	3.07	1.84	5.02	1.37	2.38	3.36
Dynamic. 92 days prior to election	3.33	1.91	5.67	1.49	2.69	3.31
Dynamic. 116 days prior to election	3.31	2.58	5.29	1.74	2.46	3.42
Dynamic. 148 days prior to election	3.45	2.15	5.14	1.93	2.42	4.34

B.3.1 Comparison to fundamental-based model

The dynamic feature of our model considerably improves its predictive performance for the last four elections. Table 1 compares the RMSE (of the mean prediction) for the fundamentals-based model with our dynamic Bayesian forecasting model for different points in time of the campaign. The average error of the fundamentals-based model for the elections 2002–2017 is relatively small with 3.90. Our dynamic Bayesian forecasting model provides a slightly better accuracy 148 to 36 days before the election, and then strongly improves during the last eight days to, on average, 195. Two days before the election, the RMSE is 1.79. This pattern holds for the other elections as well: the predictive performance of both the fundamentals-based and the dynamic Bayesian forecasting model is about the same until 36 days to the election, but then substantially increases for the dynamic model. A good example for the strength of our dynamic approach is the 2013 Federal election. The forecast of the fundamentals-based model was quite off with an RMSE of 3.50, however its misleading predictions could be bolstered by the dynamic polls component of the dynamic Bayesian forecasting model (with a RMSE of 1.20 and 1.67 two and eight days before the election, respectively).

B.3.2 Evaluating different model parts for 2017

In this section, we evaluate two parts of the model for the 2017 election. We first use a simpler model specification in which we only include previous vote share. Table 2 reports that using this simpler specification results consistently in a higher RMSE. The pure

Table 2: The RMSE of the out-of-sample predictions for different model specifications and the 2017 election. Pure fundamentals-based model and dynamic Bayesian forecasting model arranged by time to election.

	Full Model	Vote Share t_{-1}	Model w/o house effects
Pure Fundamentals-Based	4.22	5.59	4.22
Dynamic. 2 days prior to election	1.88	2.02	2.01
Dynamic. 8 days prior to election	2.13	2.40	2.06
Dynamic. 36 days prior to election	3.29	4.32	3.12
Dynamic. 64 days prior to election	3.36	4.37	3.14
Dynamic. 92 days prior to election	3.31	4.59	3.39
Dynamic. 116 days prior to election	3.42	4.54	3.47
Dynamic. 148 days prior to election	4.34	5.19	3.92

fundamentals-based model RMSE is 5.59 instead of 4.29. In comparison to the dynamic forecast, we see that this matters in particular early during the campaign, where the fundamentals-based model still has a substantial influence on the forecast accuracy. 148 to 36 days, the RMSE is about 1 point larger for the model with only past vote share. The difference vanishes close to the election, where the two models give almost identical RMSE. This means that our modeling approach improves even inferior fundamental-based models over the campaign.

Next, we consider our dynamic Bayesian measurement model that does not include house effects. In this case, we see that the house effects do not particular improve our accuracy in forecasting the final outcome for the 2017 election. The evolution of the RMSE for the dynamic forecasts are very similar for the model with and without house effects. This potentially is a result of the identification restriction that the average house effects are zero.

B.3.3 Comparison to polls

Another way to benchmark our model is by comparing it to actual polls. It is informative to see how our forecasting model performs relative to monthly poll averages across parties (that have been shown already in figure 2 of the main text[10]). How much do we gain in terms of precision from our combined model that is anchored by a fundamental model compared to simple polling averages? The following table 3 compares the respective RMSEs of our model forecasts and the monthly poll averages across parties for each election separately. For the elections in 2002, 2005 and 2009 we provide forecasts for 6

[10]We calculate the monthly poll averages for each cutoff as the average across all polls that have been available up to 35 days to this cutoff date to make sure to have a reasonable number of polls available for a comparison. This is particularly an issue with the frequency of polls in New Zealand but for the interest of comparison between Germany and New Zealand we keep this arbitrary number.

(including others) parties while for the elections in 2013 and 2017 we provide forecasts for 7 (including others) parties. In order to provide an overall assessment of the model performance and monthly polling averages respectively, we calculate the respective RMSE across the vote shares of all 32 ($= 3 \cdot 6 + 2 \cdot 7$) parties that we predict between 2002 and 2017.[11] In the first column we document the overall precision. The respective smaller RMSE value for each cutoff is in bold.

Across almost all of the cutoff dates[12] our combined model performs better than the average of the poll averages across the respective 32 vote shares in the elections we look at (with the exception of 92 days before the election). While both, the performance of our combined model as well as the polls get better the closer election day is the respective RMSE based on our models is typically smaller indicating a superior performance than the poll averages. Essentially a similar picture holds if we compare the RMSEs for specific elections between 2002 and 2017 that are documented in the remaining columns. Our model outperforms the poll averages across parties most of the time.

Table 3: Benchmarking the combined model using monthly poll averages. A comparison of the RMSE for monthly poll averages and our forecast across all five elections and at several different points in time during the election campaign.

Cutoff	Model	RMSE					
		2002-17	2002	2005	2009	2013	2017
2 days	Combined Model	**1.79**	1.27	**2.96**	**0.92**	**1.20**	**1.88**
	Avg. of Polls	2.07	**1.23**	3.30	1.17	1.42	2.41
8 days	Combined Model	**1.95**	**0.92**	**3.12**	**1.12**	**1.67**	**2.13**
	Avg. of Polls	2.23	1.34	3.44	1.27	1.69	2.60
36 days	Combined Model	**2.88**	**2.03**	**4.34**	**1.52**	2.33	3.29
	Avg. of Polls	2.89	2.14	4.45	**1.52**	**2.27**	**3.20**
64 days	Combined Model	**3.07**	**1.84**	**5.02**	**1.37**	2.38	3.36
	Avg. of Polls	3.10	1.87	5.16	1.39	**2.36**	**3.31**
92 days	Combined Model	3.33	**1.91**	5.67	1.49	2.69	3.31
	Avg. of Polls	**3.27**	2.42	**5.48**	**1.46**	**2.33**	**3.29**
116 days	Combined Model	**3.31**	**2.58**	5.29	**1.74**	**2.46**	**3.42**
	Avg. of Polls	3.32	2.96	**4.95**	1.83	2.52	3.58
148 days	Combined Model	**3.45**	2.15	5.14	1.93	**2.42**	**4.34**
	Avg. of Polls	3.53	**2.06**	**4.91**	**1.65**	2.79	4.74

[11] Note, we do not take simply calculate the average of the RMSEs across all elections because the number of parties varies across elections.

[12] The same cutoffs that we have introduced in the main text.

Table 4: The estimated correlation matrix for the evolution variance in the dynamic Bayesian forecasting model for the 2017 election with polls up to 2 days prior to the election. Note that the SPD was the baseline party in the log-ratio transformation.

		AfD	CDU	FDP	Green	Left	Oth
AfD	Mean	1	0	0	0	0	0
	(95% CI)	(1; 1)	(0; 0)	(0; 0)	(0; 0)	(0; 0)	(0; 0)
CDU	Mean	-0.21	0.97	0	0	0	0
	(95% CI)	(-0.39; -0.02)	(0.92; 1)	(0; 0)	(0; 0)	(0; 0)	(0; 0)
FDP	Mean	0.14	0.06	0.98	0	0	0
	(95% CI)	(-0.04; 0.32)	(-0.1; 0.24)	(0.94; 1)	(0; 0)	(0; 0)	(0; 0)
Green	Mean	-0.04	0.28	0.06	0.94	0	0
	(95% CI)	(-0.23; 0.15)	(0.11; 0.44)	(-0.12; 0.23)	(0.88; 0.99)	(0; 0)	(0; 0)
Left	Mean	0.29	0.01	0.13	0.04	0.93	0
	(95% CI)	(0.11; 0.47)	(-0.18; 0.19)	(-0.05; 0.3)	(-0.14; 0.23)	(0.86; 0.98)	(0; 0)
Oth	Mean	-0.29	0.25	-0.01	-0.02	-0.04	0.9
	(95% CI)	(-0.45; -0.12)	(0.08; 0.42)	(-0.18; 0.16)	(-0.2; 0.15)	(-0.23; 0.14)	(0.82; 0.96)

B.3.4 Evolution correlations

The model estimates the correlation in how support for parties evolves via the evolution variance $\boldsymbol{W}$. While this is not the main focus of our analysis, we still want to take a look the estimates. The means and the 95% credible intervals of the correlation matrix for the evolution variance are reported in Table 4. The estimates are somewhat difficult to decipher as they are defined in terms of log-ratios, where the SPD is the baseline. While most log-ratios do not seem to covary (the respective credible intervals include 0), we would like to highlight two interesting patterns that we find in this example. A gain of the Left party relative to the SPD is positively correlated with a gain of the AfD relative to the SPD. This indicates that when the AfD gained support, the Left party is also likely to gain, showing a common populist trend similarly in favor of both, the left-wing and right-wing party in our data. Another interesting positive covariance is found between the evolution of Green and CDU/CSU support relative to the SPD potentially indicating that both parties benefited systematically from the decline of the SPD support during the campaign.

C Application to New Zealand General Elections 2017

C.1 Forecast for the New Zealand General Elections 2017

Figure 4 provides our final forecasts for the New Zealand election in September 2017, published one day before election day,[13] along with the respective $\frac{5}{6}$ credible intervals.

[13]The forecasts of an older version of our model were made available on as a blog entry on `http://http://zweitstimme.org/20170922_1_blog.html`.

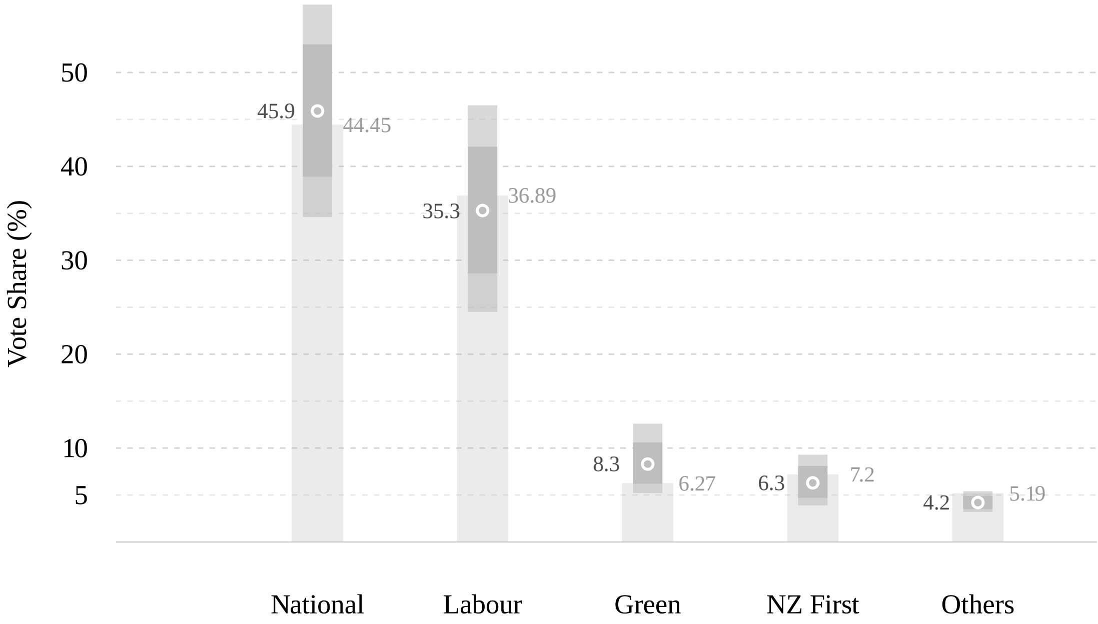

Figure 4: Forecast of the 2017 general election in New Zealand 2 days prior to the election. Point estimates along with $\frac{5}{6}(\approx 83\%)$ (dark grey) credible intervals and 95% (light grey) credible intervals, the light grey histogram bars represent the election results.

Accordingly, we predicted that the National Party will reach 45.9% [38.9%; 53%], the Labour Party 35.3% [28.6%; 42.1%], the Green Party 8.3% [6.2%; 10.6%], New Zealand First 6.3% [4.7%; 8.1%], and Others 4.2% [3.5%; 4.9%]. Similar to the German forecast, our predictions are reasonably close to the final results: for all parties except for the Other Parties, the actual election result is within the range of the $\frac{5}{6}$ credible intervals, and our final forecast has an RMSE of 1.46, which is a considerably small error in multi-party forecasting scenarios.

Finally, we calculated useful quantities of interest. We correctly predicted the National Party to become the strongest party (with a probability of 86.9%), and that both the Green Party (97.9%) and New Zealand First (86.1%) clear the 5% well ahead of time.

C.2 Fundamentals-based forecasting model

C.2.1 Data

To calibrate the fundamentals-based model, we ideally leverage data on all seven general elections since 1996[14]. However, we only had data on polls 200-230 days prior to the

[14] 1996 was the first election using the mixed-member proportional system in New Zealand. The result of the 1996 election is used as an indicator of long-term party identification, but is not part of the training

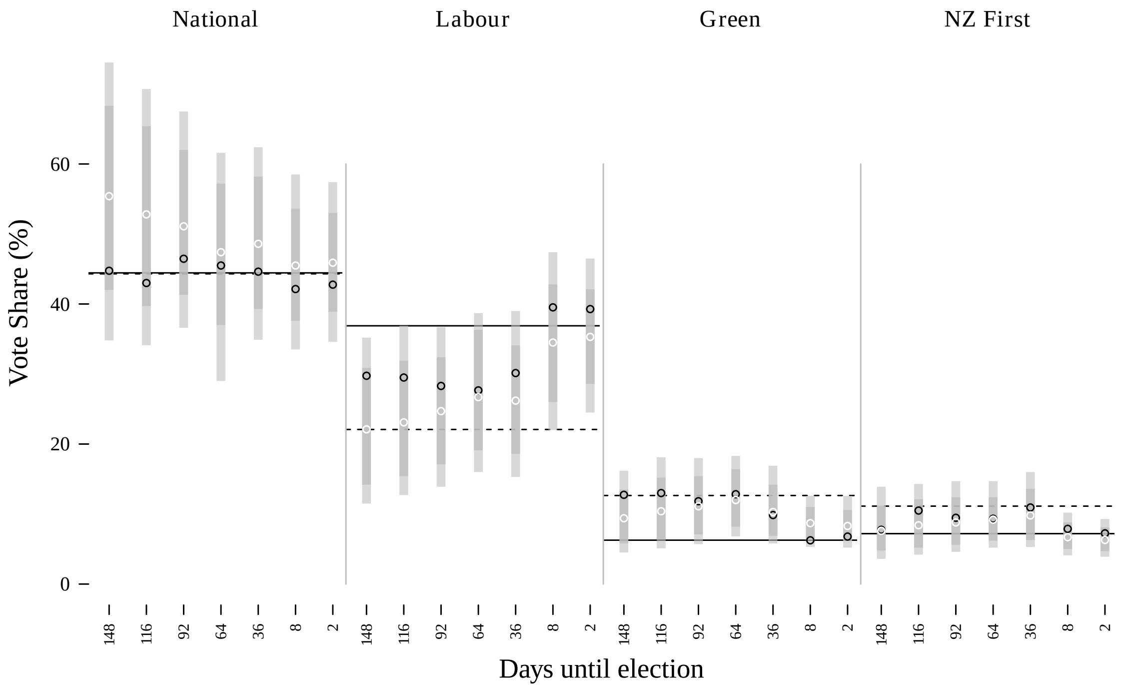

Figure 5: Development of the dynamic Bayesian forecasting model's vote share predictions over time for the general election in New Zealand 2017, starting 148 days until the final prediction one day before the election. The light points show the mean prediction; the dark grey bars depict the $\frac{5}{6}$ credible intervals and the light grey bars the 95% credible intervals. Each party's observed vote share is indicated by the solid horizontal line. The forecast of the fundamentals-based model is marked by the dashed horizontal line. The dark points indicate the monthly average of the polls included in the estimation of the dynamic Bayesian forecasting model.

respective election from 2005 on. We model the vote shares of *National*, *Labour*, *Green*, *NZ First* and "*others*", for all four elections since 2005. We rely on data collected by Peter Ellis made available through the R package `nzelect`.

C.2.2 Empirical relationship between predictors and motivation of predictors

Figure 6 shows the relationship between long-term effects, short-term and institutional features and the election outcomes of General Elections in New Zealand. Panel (a) depicts the relationship between previous and current party vote shares across elections from 1999 to 2014.

While this predictor clearly helps separate small from large parties and also explains variation to a certain degree, swings are, again, not fully captured. Panel (b) in Figure 6 shows that our second predictor, short term-effects as measured by vote intention in polls 230 to 200 days prior to the election, performs already quite well in predicting the actual

set.

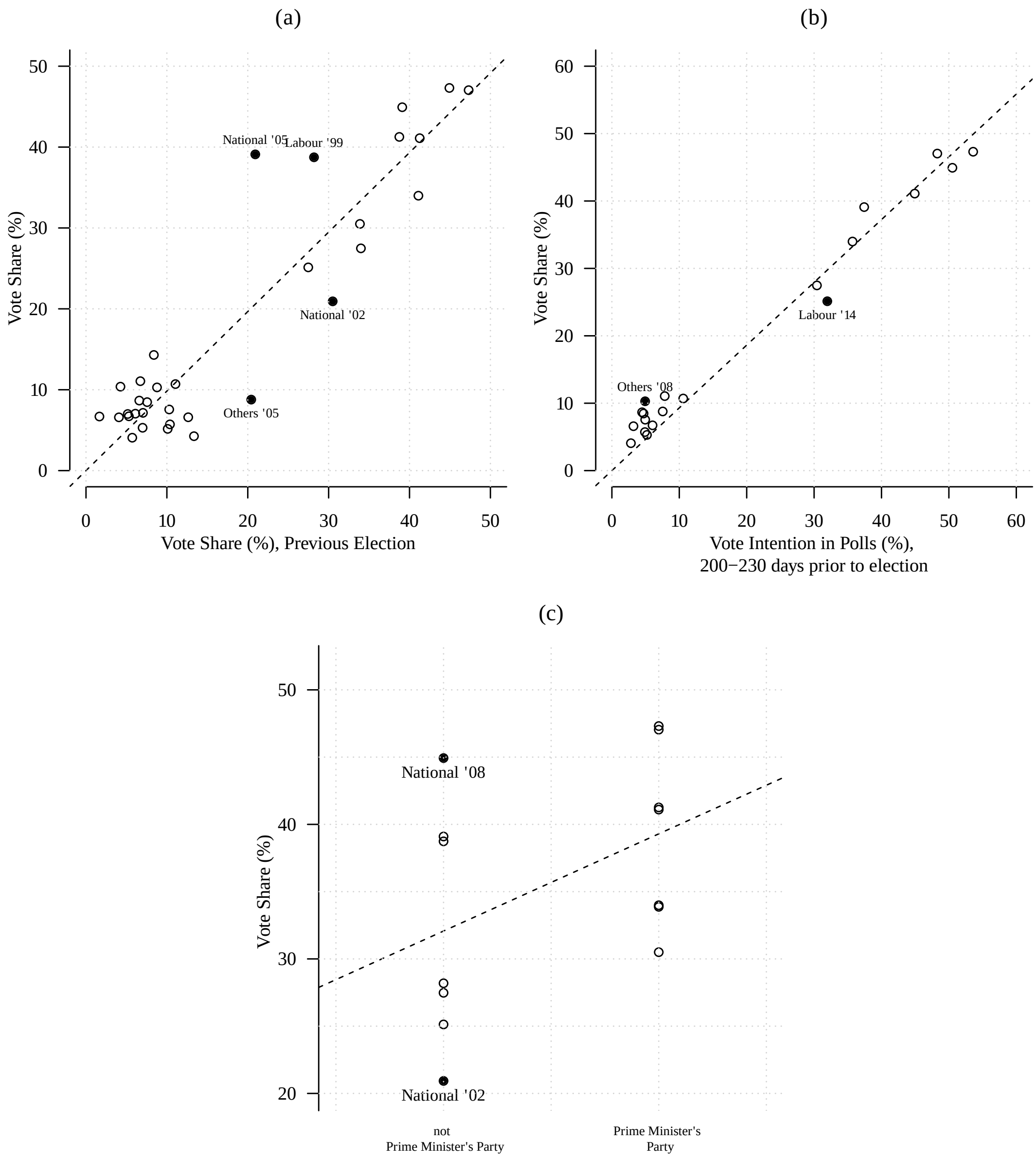

Figure 6: Relationship between predictors and vote share for New Zealand.

vote shares on election day. Panels (a) and (b) reflect what we also find for the German case in Figure 1. Panel (c) in Figure 6 reveals that the party of the incumbent prime minister has on average a larger vote share than the respective other large party that does not hold the prime minister office (the small parties are not shown). However, it is interesting to note that this incumbency effect seems to matter less than in the German case.

C.2.3 Fundamentals-based model estimates and predictive performance

Figure 7 reports the estimated β-coefficients for the last five general elections. As for Germany, the pattern confirms our expectation that the effect of the predictors vary over different elections. Yet, the story is a different one. Prior election results do not add much in the fundamentals-based model for New Zealand. Contrary to Germany, the polls get less predictive in foreseeing the final outcome. The prime minister's party effect is not distinguishable from 0 as for Germany. For the 2017 elections, we extrapolate the observed trends for all coefficients, given the estimates of the drift parameter and the random-walk component.

C.3 Dynamic Bayesian measurement model for New Zealand general elections

The dynamic Bayesian forecasting model sequentially updates the fundamentals-based forecast. We use the polling results of the Labour party during the 2017 campaign for the general election to illustrate how the anchoring process in the backwards random walk works. Figure 8 shows the results of the dynamic forecasting model for the Labour vote share over time, starting 148 days before the election. Using this example we want to highlight additional features of the model.

The Labour party replaced their candidate for the prime minister office just 53 days prior to the election. After Jacinda Ardern took over as new candidate, Labour began to rise in the following polls. This so called Jacindamania[15] is reflected in the forecasts of the dynamic model. The rise of the Labour party in the polls also considerably pulls our forecasts for the Labour party upwards. This is one of the strengths of our model. The

[15]https://www.theguardian.com/world/2017/sep/02/jacindamania-rocketing-rise-of-new-zealand-labours-fresh-political-hope

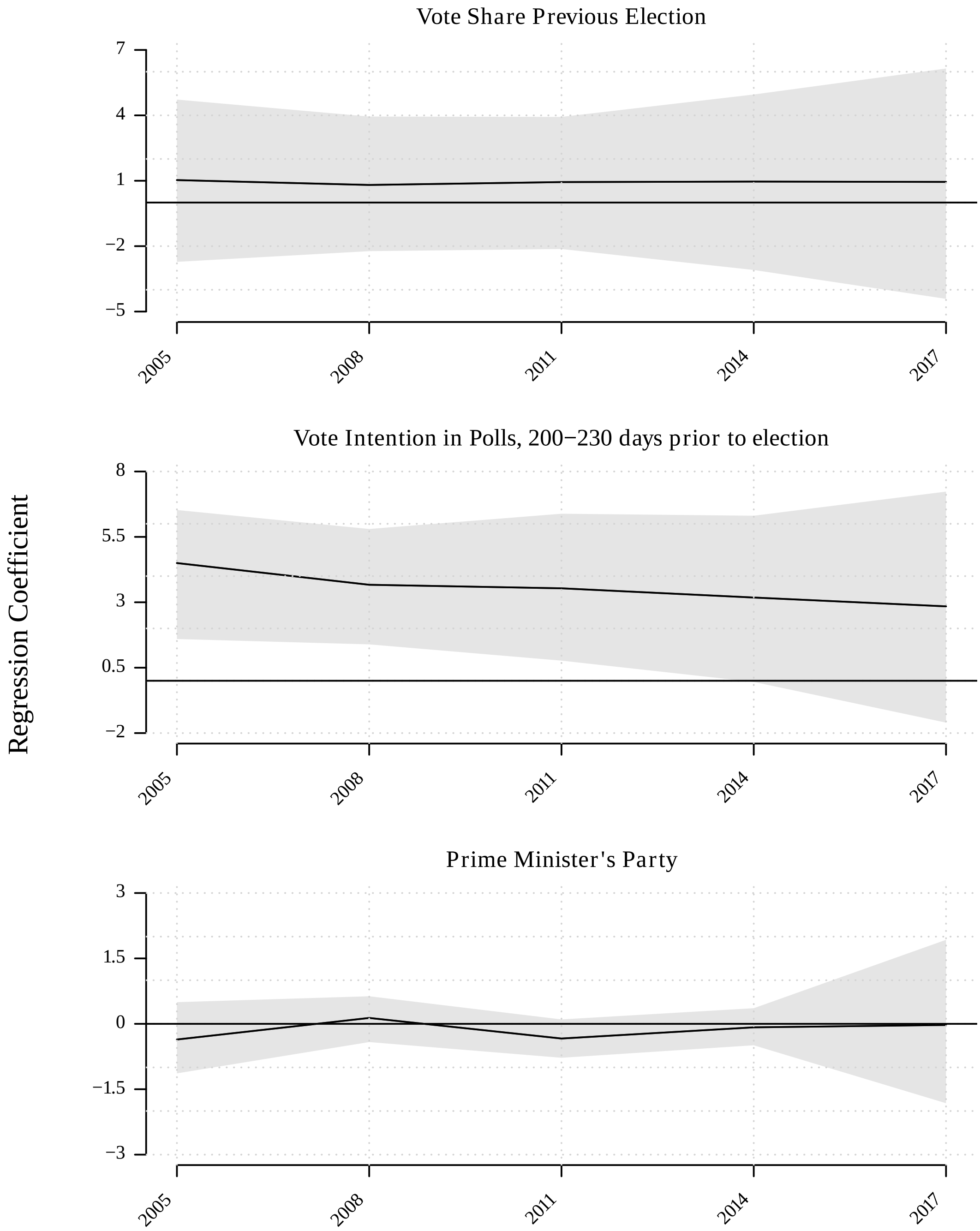

Figure 7: Coefficients for fundamentals-based forecasting model for New Zealand

dynamic Bayesian measurement model takes into account campaign dynamics that would have been missed by pure fundamentals-based models.

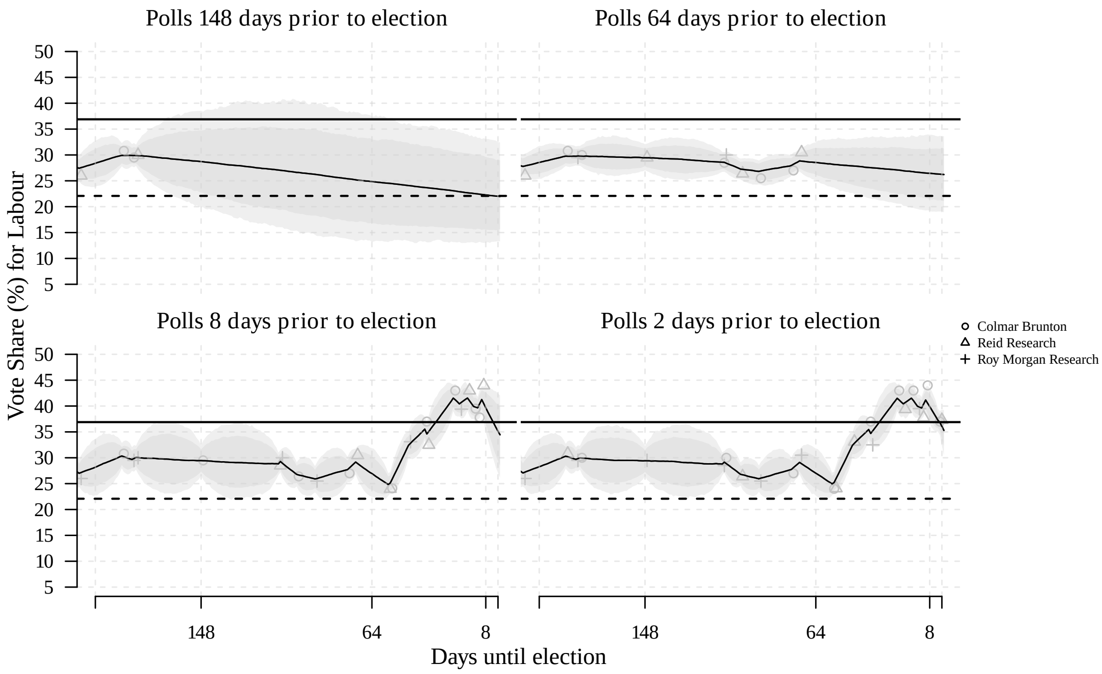

Figure 8: Labour vote share 2017 prediction based on the dynamic Bayesian forecasting model. The symbols represent the party support reported in the respective polls. The solid line depicts the mean latent Labour party support of the posterior distribution; the shadowed area depict the $\frac{5}{6}$ and 95% credible intervals. The observed 2017 Labour vote share is indicated by the solid horizontal line (36.89%), and the forecast of the fundamentals-based model is marked by the dashed horizontal line (22.08%).

Figure 8 also illustrates how the model's weight on the fundamentals-based forecast and the pre-electoral party support as measured in polls changes over time. At the beginning of the electoral campaign, the model puts more weight on the forecast of the fundamentals-based model, whereas it puts more and more "trust" into the polling trends when elections come temporally closer. Take the two panels in the upper part of the figure, 148 and 64 days before the election. Here, the predicted vote share for the Labour party slowly approaches towards the horizontal dashed line that indicates the fundamentals-based forecast. In contrast, eight days and one day before the election the model diverges from the fundamental-based forecast and approximates the tendency of the public support expressed in the polls.

Table 5: The RMSE of the out-of-sample predictions by data basis and election year for New Zealand. Pure fundamentals-based model and dynamic Bayesian forecasting model arranged by time to election.

Model	RMSE			
	2011-17	2011	2014	2017
Pure Fundamentals-Based	6.36	7.60	2.03	7.73
Dynamic. 2 days prior to election	1.84	2.42	1.48	1.46
Dynamic. 8 days prior to election	1.99	2.82	1.15	1.63
Dynamic. 36 days prior to election	4.66	5.00	3.00	5.58
Dynamic. 64 days prior to election	4.82	5.85	2.34	5.48
Dynamic. 92 days prior to election	5.37	5.83	2.99	6.60
Dynamic. 116 days prior to election	5.98	6.88	2.04	7.47
Dynamic. 148 days prior to election	6.73	7.88	1.96	8.36

C.4 Model Evaluation

C.4.1 Comparison to fundamental-based model

The case of New Zealand impressively shows that the dynamic feature of our model considerably improves its predictive performance for the last three elections. Table 5 compares the RMSE for the fundamentals-based model with our dynamic Bayesian forecasting model for different points in time of the campaign. The average error of the fundamentals-based model for the elections 2011–2017 is considerably large with about 6.36. Taking into account our dynamic Bayesian forecasting model, the accuracy especially improves during the last eight days to, on average, 1.99. Two days before the election the average RMSE is 1.84. This pattern holds for single elections: adding the dynamic model improves the predictive performance of the fundamentals-based model, it then slightly improves over time and then significantly improves one week prior to the election. The general elections in New Zealand, especially the 2017 election, are good examples for the strength of our dynamic approach. The forecast of the fundamentals-based model was really off with a RMSE of 7.71, however its misleading predictions could be bolstered by the dynamic polls component of the dynamic Bayesian forecasting model (with a RMSE of 1.46 two days before the election). This shows the merit of our approach even with a weak fundamentals-based model. With a more elaborate context-specific fundamentals-based model one could potentially improve the predictive performance months ahead of the election.

C.4.2 Comparison to polls

A way to benchmark our model for New Zealand is by comparing it to actual poll averages across parties as well. How does our combined model perform relative to monthly poll averages across parties? Similar to table 3 for the German case, table 6 compares the respective RMSEs of our model forecasts for New Zealand and the monthly poll averages across parties for each election separately. In order to provide an overall assessment of the model performance and monthly polling averages respectively, we calculate the respective RMSE across the vote shares of all parties that we predict between 2011 and 2017. In the first column we document the overall precision across all parties and all three elections. The respective smaller RMSE value for each cutoff is in bold. In the remaining columns we compare the model performances compared the the poll averages for each election individually.

Our model performs consistently worse than the respective poll averages early in the campaign. One reason for that might be that the fundamental model performs less well in New Zealand than in Germany. The closer we get to election day our model performance improves. Between 36 and 8 days before the election our model starts to improve considerably such that it beats the polls no matter which election we look at. Thus, our model consistently outperforms the poll averages across parties and all three elections 8 days before election day.

Table 6: Benchmarking the combined model using monthly poll averages. A comparison of the RMSE for monthly poll averages and our forecast across all three elections and at several different points in time during the election campaign.

Cutoff	Model	RMSE			
		2011-17	2011	2014	2017
2 days	Combined Model	**1.84**	**2.42**	**1.48**	1.46
	Avg. of Polls	2.12	2.92	1.82	**1.30**
8 days	Combined Model	**1.99**	**2.82**	**1.15**	**1.63**
	Avg. of Polls	2.41	3.35	1.88	1.65
36 days	Combined Model	4.66	5.00	3.00	5.58
	Avg. of Polls	**3.68**	**4.15**	**2.95**	**3.83**
64 days	Combined Model	4.82	5.85	**2.34**	5.48
	Avg. of Polls	**4.17**	**3.99**	3.09	**5.18**
92 days	Combined Model	5.37	5.83	**2.99**	6.60
	Avg. of Polls	**4.09**	**3.71**	3.65	**4.81**
116 days	Combined Model	5.98	6.88	**2.04**	7.47
	Avg. of Polls	**4.07**	**4.07**	3.22	**4.78**
148 days	Combined Model	6.73	7.88	**1.96**	8.36
	Avg. of Polls	**3.99**	**4.20**	3.38	**4.32**

References

Arzheimer, Kai. 2006. "'Dead men walking?' Party identification in Germany, 1977–2002." *Electoral Studies* 25(4):791–807.

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

Converse, Philip E. 1966. The Concept of Normal Vote. In *Elections and the Political Order*, ed. Angus Campbell, Philip E. Converse, Warren E. Miller and Donald E. Stokes. New York: Wiley pp. 9–39.

Groß, Jochen. 2010. *Die Prognose von Wahlergebnissen. Ansätze und empirische Leistungsfähigkeit*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Gschwend, Thomas and Helmut Norpoth. 2001. Wenn am nächsten Sonntag...Ein Prognosemodell für Bundestagswahlen. In *Wahlen und Wähler: Analysen aus Anlass der Bundestagswahl 1998 [Elections and Voters. Analysis of the Federal Election 1998]*, ed. Hans-Dieter Klingemann and Max Kaase. Wiesbaden: Westdeutscher Verlag pp. 473–499.

Jennings, Will and Christopher Wlezien. 2016. "The Timeline of Elections: A Comparative Perspective." *American Journal of Political Science* 60(1):219–233.

Jérôme, Bruno. 2013. "Editor's Introduction for Forecasting the 2013 German Elections." *PS: Political Science & Politics* 46(3):477–478.

Kayser, Mark A. and Arndt Leininger. 2017. "A Länder-based Forecast of the 2017 German Bundestag Election." *PS: Political Science & Politics* 50(3):689–692.

Kayser, Mark Andreas and Arndt Leininger. 2016. "A Predictive Test of Voters' Economic Benchmarking: The 2013 German Bundestag Election." *German Politics* 25(1):106–130.

Lewis-Beck, Michael S. 2005. "Election Forecasting: Principles and Practice." *The British Journal of Politics and International Relations* 7(2):145–164.

Lewis-Beck, Michael S and Mary Stegmaier. 2000. "Economic determinants of electoral outcomes." *Annual Review of Political Science* 3(1):183–219.

Lewis-Beck, Michael S. and Ruth Dassonneville. 2015*a*. "Comparative election forecasting: Further insights from synthetic models." *Electoral Studies* 39:275–283.

Lewis-Beck, Michael S and Ruth Dassonneville. 2015*b*. "Forecasting elections in Europe: Synthetic models." *Research & Politics* 2(1):1–11.

Magalhães, Pedro C., Luís Aguiar-Conraria and Michael S. Lewis-Beck. 2012. "Forecasting Spanish elections." *International Journal of Forecasting* 28(4):769–776.

Neunhoeffer, Marcel, Lukas F. Stoetzer, Thomas Gschwend, Simon Munzert and Sebastian Sternberg. 2018. "Replication Data for: Forecasting Elections in Multi-Party Systems: A Bayesian Approach Combining Polls and Fundamentals.".
URL: *https://doi.org/10.7910/DVN/MLYNX0*

Norpoth, Helmut and Thomas Gschwend. 2003. "Against All Odds? The Red-Green Victory." *German Politics and Society* 21(1):15–34.

Norpoth, Helmut and Thomas Gschwend. 2010. "The chancellor model: Forecasting German elections." *International Journal of Forecasting* 26(1):42–53.

Norpoth, Helmut and Thomas Gschwend. 2013. "Chancellor Model Picks Merkel in 2013 German Election." *PS: Political Science & Politics* 46(03):481–482.

Schnell, Rainer and Marcel Noack. 2014. "The Accuracy of Pre-Election Polling of German General Elections." *Methods, Data, Analyses* 8(1):5–24.

Selb, Peter and Simon Munzert. 2016. "Forecasting the 2013 German Bundestag Election Using Many Polls and Historical Election Results." *German Politics* 25(1):73–83.