German Election Forecasting:
Comparing and Combining Methods for 2013

Forthcoming (subject to changes) in *German Politics*

Andreas Graefe
Department of Communication Science and Media Research
LMU Munich, Germany
a.graefe@lmu.de

Abstract. The present study reviews the accuracy of four methods for forecasting the 2013 German election: polls, prediction markets, expert judgment, and quantitative models. On average, across the two months prior to the election, polls were most accurate, with a mean absolute error of 1.4 percentage points, followed by quantitative models (1.6), expert judgment (2.1), and prediction markets (2.3). In addition, the study provides new evidence for the benefits of combining forecasts. Averaging all available forecasts within and across the four methods provided more accurate predictions than the typical component forecast. The error reductions achieved through combining forecasts ranged from 5% (compared to polls) to 41% (compared to prediction markets). The results conform to prior research on US presidential elections, which showed that combining is one of the most effective methods to generating accurate election forecasts.

Keywords. Election forecasting, combining forecasts, PollyVote, accuracy

Acknowledgments. Mario Haim helped developing the PollyVote.de website and collecting the forecast data. Bettina Zerwes collected the names for the expert survey. This work was supported by an LMUexcellent research fellowship from the *Center for Advanced Studies* at LMU Munich.

Introduction

Forecasting elections has become increasingly popular in recent years and researchers have developed a variety of methods that can help to predict election outcomes. While traditional polls remain most common in media campaign coverage, other important methods include quantitative models[1], prediction markets[2], and expectation surveys of citizens and experts.[3]

Prior evidence on US presidential elections suggests that there is no clear order in terms of the methods' relative accuracy. The reason is that a method's accuracy is heavily influenced by the idiosyncrasies of a particular election and thus varies both across elections as well as within a single campaign. One study provides evidence on the relative accuracy of polls, prediction markets, expert surveys, and quantitative models across the six elections from 1992 to 2012: methods that provided the most accurate forecasts in one election were often among the least accurate in another election.[4]

At the time of making a prediction, it is usually difficult – if not impossible – to know which of several available forecasts will be most accurate. In such situations, a valuable strategy to produce accurate forecasts is to combine all available forecasts, instead of relying on a single forecast.[5] Often, the combined forecast is more accurate than even the most accurate individual forecast.[6] It is, however, important that the procedure for how to combine the forecasts is specified prior to seeing the forecasts. This avoids that forecasters weight forecasts in a way that the result suits their biases.[7]

Combining forecasts works for two reasons. First, a combined forecast includes more information than forecasts from any single component method. Second, individual component forecasts are often associated with systematic (and random) errors, which are likely to cancel out in the aggregate. Therefore, combining is particularly useful if one can draw on many forecasts that were generated with different methods and that rely on different data.[8]

The benefits of combining are known for almost half a century[9] and have recently also entered election forecasting. While the *Economist* published the first polling average in 1992, online polling aggregators such as *RealClearPolitics'* Poll Average or the *Huffington Post* Pollster are becoming increasingly popular for reporting polling results in the US.[10] Since 2004, the PollyVote.com project demonstrates the benefits of combining for forecasting the national popular vote in US presidential elections by averaging forecasts within and across four methods: polls, prediction markets, expert judgment, and quantitative models. On Election Eve prior to the three elections from 2004 to 2012, the combined PollyVote forecast missed the final vote share on average by 0.6 percentage points. In comparison, the corresponding error of the final Gallup poll was more than three times higher.[11] Furthermore, an ex post analysis across the last 100 days prior to the six elections from 1992 to 2012 found that the PollyVote provided more accurate forecasts than each component method. Compared to prediction markets, the most accurate component, the PollyVote reduced forecast error by 16%. Compared to a typical poll, the method that is most prominent in media campaign coverage, error was reduced by 59% on average.[12] One study used a similar approach to demonstrate the power of

combining for forecasting the 2012 US Electoral College and senatorial elections, a situation in which fewer methods and limited data are available. The combined forecast of prediction markets, polls, and a quantitative model provided robust predictions and was often more accurate than the best component method.[13]

The available evidence on combining forecasts is limited to US elections, however. Little is known about whether the approach is equally valuable for predicting election results in more complex electoral systems, such as multi-party systems with proportional representation. The present study provides evidence from applying the PollyVote approach to such a situation, namely the 2013 German election.

Applying the PollyVote to the 2013 German election

The present study follows the approach that has previously been used in the US version of the PollyVote. That is, combined forecasts of the 2013 German election were calculated by averaging forecasts within and across four component methods: polls, prediction markets, expert judgment, and quantitative models. This section provides a brief overview of each component method (and the available forecasts for the 2013 election) and describes the calculation of the combined PollyVote forecast.

Polls

Polls ask people for whom they intend to vote if the election was held today. Thus, polls measure public opinion at a certain point in time; they do not provide predictions of what will happen on Election Day. Yet, polling results are commonly projected to Election Day and interpreted as forecasts.[14]

As in the US, online polling aggregators have become increasingly popular in Germany. Prior to the 2013 election, two websites reported polling averages. Wahlumfrage.de calculated simple unweighted averages of the most recent polls conducted by six established German pollsters, namely, Allensbach, Emnid, Forsa, Forschungsgruppe Wahlen (FGW), GMS, and Infratest dimap. Pollytix.de calculated weighted averages and also included polls from two other survey institutes (TNS and INSA/YouGov). The weighting approach assigned higher weights to surveys with larger samples and to surveys that were conducted more recently.

Prediction markets

Betting on election outcomes has a long history and can be traced back to 16th century Italy, where such markets were common for civic and papal elections.[15] Long before the emergence of scientific polling, such markets were also popular for US presidential elections and betting odds were published as forecasts in leading newspapers such as the *New York Times*.[16] The University of Iowa established the first online prediction market, the *Iowa Electronic Markets* (IEM), which has provided highly accurate predictions of US presidential election outcomes since 1988.[17]

The IEM is a so-called real-money market. That is, participants can open an account with up to $500. The money can then be used to buy and sell shares of political parties. Participants win (or lose) money depending on the accuracy of their predictions, and thus have an incentive to make accurate predictions. The market price of the shares provides the forecast of the election result.

Prior to the 2013 German election, five websites ran a total of six prediction markets. These websites were eix-market.de, politikprognosen.de, prognosys.de, spiegel.de, and wahlfieber.de (which ran two markets). Politikprognosen.de used a similar design as the real-money IEM but the maximum investment was limited to 20 Euros. Spiegel.de did not use a real market mechanism to aggregate the individual predictions. Instead, participants could sign up and submit predictions, which were averaged into a combined forecast. The remaining five markets were play-money markets. That is, participants received a certain amount of play money that they could use for trading. Performance on play-money markets is measured by rankings and, in some markets, the best performing participants can win prizes.

Expert judgment

Expert surveys have a long history as a method to forecast election outcomes.[18] Experts are assumed to provide accurate forecasts due to their domain knowledge. Experts may, for example, be able to correctly interpret polls and project their results to Election Day, by taking into account potential impacts of recent and future campaign events.

Prior to the 2013 election, two groups of experts (i.e., political journalists and German election scholars) were asked to participate in an online survey. The names of political journalists were obtained from the public relations and media database zimpel.de. Names of German election scholars with at least a doctoral degree were collected from websites of universities and think tanks.

Respondents were asked to predict the vote-shares received by the seven largest parties (i.e., CDU/CSU, SPD, Grüne, Linke, FDP, AfD, and Piraten) as well as the remaining vote share for all other parties combined (i.e., Sonstige). Respondents were told that the vote shares should sum up to 100 but the online questionnaire did not enforce this.

Five waves were conducted prior to the election. The five waves started 67, 40, 19, 12, and 5 days prior to Election Day. On average across the five waves, 53 journalists and 69 scholars participated.

Quantitative models

For more than three decades, political scientists and economists have developed quantitative models for predicting election outcomes. Most of these models rely on the idea of retrospective voting. That is, voters are assumed to reward (or punish) the incumbent government based on past performance. Given that most models use economic and political (and/or public opinion) variables to measure performance, they are often referred to as political economy models.[19]

Three political economy models were available prior to the 2013 election. The model by

Jérôme, Jérôme-Speziari, and Lewis-Beck has been used in modified form since 1998.[20] The model predicts the vote shares of all parties that are represented in the outgoing parliament based on the unemployment rate and several poll-based measures (e.g., the popularity of the Chancellor candidates of CDU/CSU and SPD, the popularity of the FDP as a coalition partner, and vote intention for the smaller parties). The *Chancellor model* by Norpoth and Gschwend, which was first published prior to the 2002 election, is based on three variables: (1) the outgoing coalition's average vote share across the three preceding elections, (2) the support for the chancellor in public opinion polls, and (3) attrition, measured as the number of terms in office.[21] The *Benchmark model* by Kayser and Leininger builds on this idea and uses slightly different measures for the three variables: (1) the outgoing coalition's average vote share in the last election, (2) the share of voters that identify with the parties that form the outgoing coalition, and (3) the number of years the outgoing coalition was in power. In addition, the model includes a fourth variable, the benchmark variable, which measures the performance of the German economy relative to the economies of France, Italy, and the UK.[22] Both the Chancellor and the Benchmark model predict the aggregate vote share of the outgoing coalition (i.e., in 2013, the sum of the vote shares gained by CDU/CSU and FDP).

In addition, two models were available that predict the election outcome based on polling data. The model by Selb and Munzert used historical polling data to first estimate the relationship between polls and election outcomes. Then, this relationship was used to predict the election outcome from polls published prior to the 2013 election.[23] Finally, the model published at the website election.de used published polls and adjusted the figures based on information derived from historical data such as the parties' vote shares and voters' vote splitting (i.e., strategic voting) in past elections.[24]

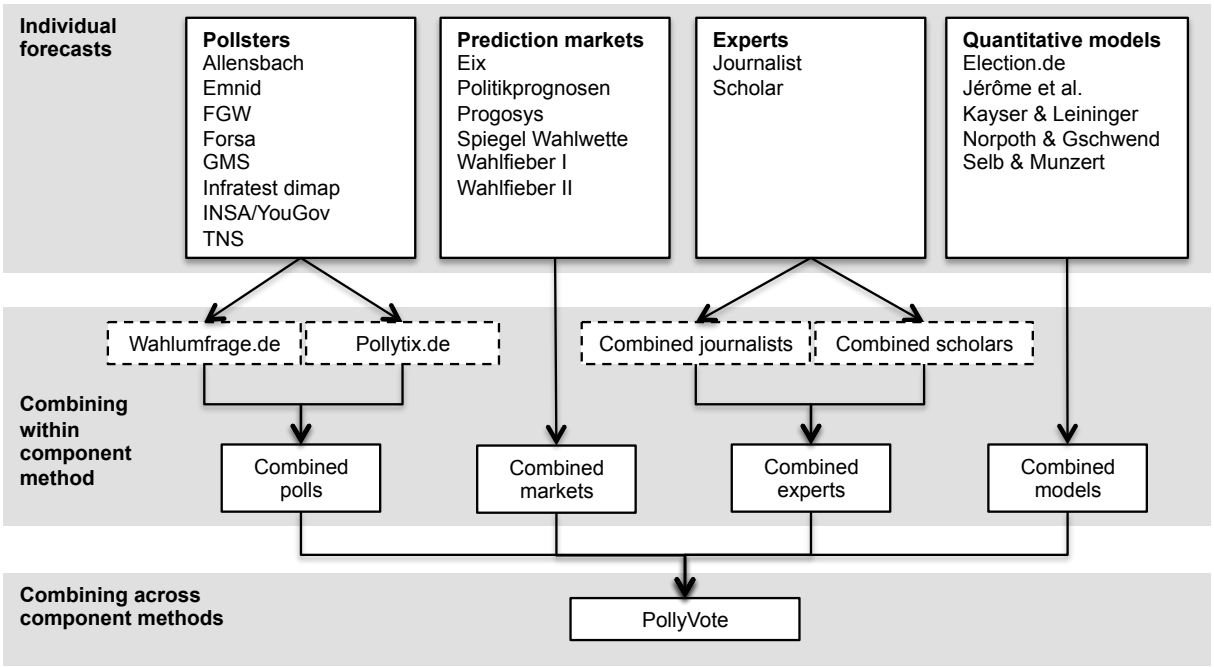
Combined PollyVote forecast

As shown in Figure 1, daily forecasts of the combined PollyVote were calculated using a two-step procedure. In the first step, the individual forecasts were averaged *within* each component method. The combined polls forecast was calculated as the simple average of the latest figures of the two polling aggregators Wahlumfrage.de and Pollytix.de, which already combined polls from different pollsters. The combined markets forecast was calculated as the average of the six prediction markets. The combined expert forecast was determined by, first, averaging the individual forecasts within both groups and, then, averaging the resulting combined forecasts across both groups. Individual forecasts were added (or updated) over time as they became available. Finally, the combined models forecast was the simple average of the five available models. In the second step, the combined PollyVote forecast was calculated by averaging the combined forecasts *across* the four component methods.

Simple averages may appear as a naïve approach to combining forecasts. However, prior research that aimed to develop sophisticated methods for combining forecasts did not improve upon the accuracy of the simple average. An early review of more than 200 papers found no evidence that complex combining procedures provide more accurate forecasts than simple averages.[25] A review of published studies since then found that the results still hold today. In addition, that study provided new

evidence for combining forecasts from six quantitative models for predicting US presidential election outcomes. Across the elections from 1976 to 2012, the error of simple averages of forecasts from six election-forecasting models was 25% lower than the corresponding error of the forecasts from a sophisticated Bayesian approach to combining forecasts.[26]

Figure 1: Procedure for calculating the combined PollyVote for forecasting the 2013 German election



Results

Forecast accuracy was analyzed across the 58-day period from July 26 to September 21 (Election Eve), which is when forecasts from all four components were available. A forecast's absolute error on a particular day was calculated by averaging the absolute differences between the predicted and actual vote shares of the seven largest parties and the remaining share for all other parties combined (i.e., CDU/CSU, SPD, Grüne, Linke, FDP, AfD, Piraten, and Sonstige).

This section first reports the mean absolute error (MAE) of the typical individual component forecast, which is the error that one would achieve by randomly picking a forecast within a certain component. Then, gains in accuracy are reported from combining *within*, combining *across*, and combining *within and across* component methods. All data and calculations are publicly available.[27]

Errors of the typical component forecast

Table 1 shows the MAE of the typical forecast in each component method, calculated across the full 58-day period. On average, a randomly picked poll achieved a MAE of 1.44 percentage points and was thus more accurate than the typical model (1.57 percentage points), the typical expert (2.13 percentage points), and the typical prediction market (2.33 percentage points).

Gains from combining *within* component methods

Table 1 also shows the MAE of the combined component forecasts. For example, the combined polls forecast achieved an MAE of 1.27 percentage points. Dividing this figure by the MAE of the typical component forecast yields the error ratio for combining *within* polls, which is 0.88. This means that combining polls reduced the error of the typical poll by 12% (= 1 – 0.88). A similar error reduction (11%) was achieved by combining the forecasts from the six prediction markets. Error reductions were larger for combining forecasts from experts (21%) and models (28%).

Table 1: Accuracy gains from combining *within*, *across* and *within and across* component methods

	Polls	Prediction markets	Experts	Models
Mean absolute error				
Typical	1.44	2.33	2.13	1.57
Combined	1.27	2.08	1.67	1.14
Error ratios				
Combining within	0.88	0.89	0.79	0.72
Combining across	1.08	0.66	0.82	1.21
Combining within and across	0.95	0.59	0.64	0.87

Notes:
- Mean absolute errors were calculated across the last 58 days prior to the election.
- The mean absolute error of the combined PollyVote forecast was 1.37.
- Underlined error ratios mean that the PollyVote was less accurate than the benchmark.

Gains from combining *across* component methods

Across the 58-day period, the PollyVote forecast, which simply averaged forecasts across the four components, achieved an MAE of 1.37. If one divides this figure by the respective errors of the combined component forecasts, one achieves the error ratios for combining *across* components. For example, as shown in Table 1, the error ratio of the PollyVote relative to the combined experts was 0.79. That is, the PollyVote error was 21% lower than the error of combined expert forecasts. Compared to combined prediction markets, the PollyVote reduced error by 34%. The error of the PollyVote was, however, 8% higher than forecasts from combined polls and 21% higher than forecasts from combined models.

Gains from combining *within and across* component methods

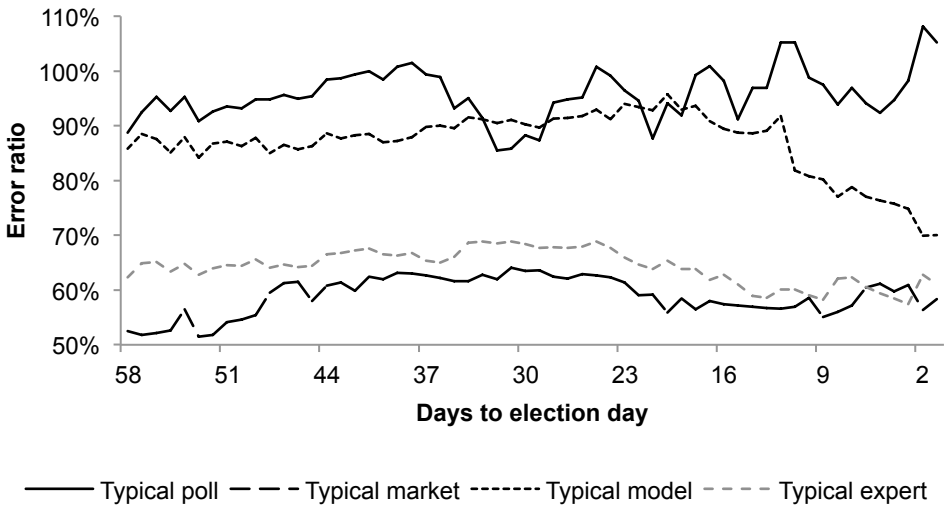
Dividing the PollyVote's MAE (i.e., 1.37 percentage points) by the respective errors of the typical component forecasts yields the error reductions achieved by combining *within and across* component methods. This is the critical figure, as it reveals the error reduction that could be obtained by relying on the PollyVote rather than randomly picking any of the individual forecasts.

As shown in Table 1, on average across the 58-day period, the PollyVote provided more accurate forecasts than the typical forecast in each component. Error reductions ranged from 5% (compared to the typical poll) to 41% (compared to the typical prediction market). Compared to the typical model and expert, the PollyVote reduced error by 13% and 36%, respectively.

Figure 2 shows the daily error ratios of the combined PollyVote relative to the typical

component forecast. Compared to the typical prediction market and the typical expert, the error reductions achieved by the PollyVote were relatively stable and did not change much over to course of the campaign. The relative accuracy of the typical poll, however, increased closer to the election. At times, the typical poll was even more accurate than the PollyVote forecast. The results thus conform to the well-known finding that polls become more accurate closer to Election Day.[28] In contrast, the relative accuracy of the typical model decreased over time, which is not surprising given that the political economy models published their forecasts weeks before the election.

Figure 2: Daily error ratios of the combined PollyVote relative to the typical component forecast



Discussion

The present study applied the principle of combining forecasts to predicting the 2013 German election, following the PollyVote approach that has been successfully used for forecasting US presidential elections since 2004. The results provide further evidence for the benefits of combining election forecasts. As in the US case, averaging forecasts *within and across* four component methods (i.e., polls, prediction markets, expert judgment, and quantitative models) produced accurate predictions of the election outcome. Across the 58 days for which forecasts from all four components were available, the combined PollyVote forecast was more accurate than each component's typical forecast. Error reductions ranged from 5% (compared to a typical poll) to 41% (compared to a typical prediction market).

Given the PollyVote's small error reduction compared to a typical poll, one might wonder whether the efforts required to combine forecasts are justified. However, the results are based on only a single election. As noted in the introduction, the methods' relative accuracy often varies substantially across elections. While 2013 was a very good year for German pollsters, there is no guarantee that polls will perform equally well in future elections. This becomes clear when looking at the relative performance of polls and prediction markets in previous German elections. While

prediction markets performed poorly in the 2013 election, they provided more accurate forecasts than polls in both the 2005 and the 2009 election.[29]

Given this uncertainty, the best course of action is to combine forecasts from different methods that use different data, an approach that is well established in the general forecasting literature. The benefits of combining forecasts for German (and other) elections will become stronger as additional data on future elections will be gained.

NOTES

[1] See, for example, the special symposiums in *Political Methodologist* 5(2), *American Politics Research* 24(4)
[2] J. E. Berg and T. A. Rietz, 'Market design, manipulation, and accuracy in political prediction markets: Lessons from the Iowa Electronic Markets', *PS: Political Science & Politics* 47/2 (2014), pp. 293-296.
[3] M. S. Lewis-Beck and C. Tien, 'Voters as forecasters: a micromodel of election prediction', *International Journal of Forecasting* 15/2 (1999), pp. 175-184; M. S. Lewis-Beck and M. Stegmaier, 'Citizen forecasting: Can UK voters see the future?', *Electoral Studies* 30/2 (2011), pp. 264-268; A. Graefe, 'Accuracy of vote expectation surveys in forecasting elections', *Public Opinion Quarterly* 78/S1 (2014), pp. 204-232.
[4] A. Graefe, J. S. Armstrong, R. J. Jones Jr. and A. G. Cuzán, 'Combining forecasts: An application to elections', *International Journal of Forecasting* 30/1 (2014), pp. 43-54.
[5] J. M. Bates and C. W. J. Granger, 'The combination of forecasts', *OR* 20/4 (1969), pp. 451-468; R. T. Clemen, 'Combining forecasts: A review and annotated bibliography', *International Journal of Forecasting* 5/4 (1989), 559-583; J. S. Armstrong, K. C. Green and A. Graefe, 'Golden Rule of Forecasting: Be conservative', *Journal of Business Research* (forthcoming), Available at: www.goldenruleofforecasting.com.
[6] J. S. Armstrong, 'Combining forecasts', in J. S. Armstrong (ed), *Principles of Forecasting: A Handbook for Researchers and Practitioners* (New York: Springer, 2010), pp. 417-439.
[7] A. Graefe, J. S. Armstrong, R. J. Jones Jr. and A. G. Cuzán, 'Combining forecasts: An application to elections', *International Journal of Forecasting* 30/1 (2014), pp. 43-54.
[8] J. S. Armstrong, 'Combining forecasts', in J. S. Armstrong (ed), *Principles of Forecasting: A Handbook for Researchers and Practitioners* (New York: Springer, 2010), pp. 417-439.
[9] J. M. Bates and C. W. J. Granger, 'The combination of forecasts', *OR* 20/4 (1969), pp. 451-468
[10] M. Blumenthal, 'Polls, forecasts, and aggregators', *PS: Political Science & Politics* 47/2 (2014), pp. 297-300.
[11] A. Graefe, J. S. Armstrong, R. J. Jones Jr. and A. G. Cuzán, 'Accuracy of Combined Forecasts for the 2012 Presidential Election: The PollyVote', PS: Political Science & Politics 47/2 (2014), pp. 427-431.
[12] A. Graefe, J. S. Armstrong, R. J. Jones Jr. and A. G. Cuzán, 'Combining forecasts: An application to elections', *International Journal of Forecasting* 30/1 (2014), pp. 43-54.
[13] D. Rothschild, 'Combining forecasts for elections: Accurate, relevant, and timely', *International Journal of Forecasting* (2014), http://dx.doi.org/10.1016/j.ijforecast.2014.1008.1006.
[14] D. S. Hillygus, 'The evolution of election polling in the United States', *Public Opinion Quarterly* 75/5 (2011), pp. 962-981.
[15] P.W. Rhode and K. S. Strumpf, 'The long history of political betting', in L. Vaughan Williams and D. S. Siegel (eds), *The Oxford Handbook of the Economics of Gambling*, (Oxford: Oxford University Press, 2014), pp. 560-586.
[16] P. W. Rhode and K. S. Strumpf, 'Historical presidential betting markets', *Journal of Economic Perspectives* 18/2 (2004), pp. 127-141.
[17] J. E. Berg and T. A. Rietz, 'Market design, manipulation, and accuracy in political prediction markets: Lessons from the Iowa Electronic Markets', *PS: Political Science & Politics* 47/2 (2014), pp. 293-296.
[18] S. Kernell, 'Life before polls: Ohio politicians predict the 1828 presidential vote', *PS: Political Science and Politics* 33/3 (2000), pp. 569-574.
[19] See, for example, the models published for US presidential elections since 1992 in special symposiums in *Political Methodologist* 5(2), *American Politics Research* 24(4) and *PS: Political Science and Politics* 34(1), 37(4), 41(4), and 45(4).
[20] B. Jérôme, V. Jérôme-Speziari and M. S. Lewis-Beck, 'A political-economy forecast for the 2013 German elections: Who to rule with Angela Merkel?', *PS: Political Science & Politics* 46/3 (2013), pp. 479-480.
[21] H. Norpoth and T. Gschwend, 'Chancellor model picks Merkel in 2013 German election', *PS: Political Science & Politics* 46/3 (2013), pp. 481-482.
[22] M. Kayser and A. Leininger, 'A benchmarking forecast and post-mortem of the 2013 Bundestag election', *Under review with German Politics* (2014).

[23] P. Selb and S. Munzert, 'Forecasting the 2013 Bundestag election using data from various polls' *Under review with German Politic*s (2014).
[24] Only election.de provided forecasts for each party. For the remaining models, the forecasts for the individual parties were calculated by using the distribution of the PollyVote forecast from the preceding day. All data and calculations are publicly available: Graefe, A. (2015). Replication data for: German election forecasting: Comparing and combining methods for 2013, Harvard Dataverse Network, http://dx.doi.org/10.7910/DVN/GERMANPOLLYVOTE2013..
[25] R. T. Clemen, 'Combining forecasts: A review and annotated bibliography', *International Journal of Forecasting* 5/4 (1989), 559-583.
[26] A. Graefe, H. Küchenhoff, V. Stierle and B. Riedl, 'Limitations of Ensemble Bayesian Model Averaging for forecasting social science problems', *International Journal of Forecasting* (forthcoming, 2014): DOI: 10.2139/ssrn.2266307.
[27] Graefe, A. (2015). Replication data for: German election forecasting: Comparing and combining methods for 2013, Harvard Dataverse Network. http://dx.doi.org/10.7910/DVN/GERMANPOLLYVOTE2013.
[28] R. S. Erikson and C. Wlezien, *The Timeline of Presidential Elections: How Campaigns Do (And Do Not) Matter* (Chicago: University of Chicago Press).
[29] L-M. Schaffer and G. Schneider, 'Die Prognosegüte von Wahlbörsen und Meinungsumfragen zur Bundestagswahl 2005', *Politische Vierteljahresschrift* 46/4 (2005), pp. 674-681; J. Groß, 'Märkte und Prognosen', in N. Braun, M. Keuschnigg and T. Wolbring (eds), *Wirtschaftssoziologie II: Anwendungen* (München: Oldenbourg, 2012), pp. 111-126.