International Journal of Forecasting 33 (2017) 467–481

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting elections at the constituency level: A correction–combination procedure

Simon Munzert

University of Mannheim, Mannheim Centre for European Social Research (MZES), 68131 Mannheim, Germany

ARTICLE INFO

Keywords:
Election forecasting
Parliamentary elections
Constituency
Bias
Correction
Combination
Germany

ABSTRACT

Scholarly efforts to forecast parliamentary elections have targeted the national level predominantly, disregarding the outcomes of constituency races. In doing so, they have frequently failed to account for systematic bias in the seats–votes curve, and been unable to provide candidates and campaign strategists with constituency-level information. On the other hand, existing accounts of constituency-level election forecasting suffer from data sparsity, leading to a lack of precision. This paper proposes a correction–combination procedure that allows for the correction of individual constituency-level forecast models for election-invariant bias, then combines these models based on their past performances. I demonstrate the use of this procedure through out-of-sample forecasts of 299 district races at the 2013 German federal election.

1. Introduction

Academic election forecasting has flourished over recent years. New models which provide timely, dynamic, and (most importantly) precise forecasts are of great interest for campaign organizers and financial contributors, who want their resources to be targeted efficiently. Furthermore, they provide valuable information for voters and are staples for horse race journalism. From a scholarly perspective, election forecasts can inform research on the nexus between public opinion and campaign dynamics (Lodge, Steenbergen, & Brau, 1995; Panagopoulos, 2009; Wlezien & Erikson, 2002), as well as on methodological issues in the measurement of public opinion (Gelman & King, 1993; Graefe, 2014).

Recent attempts to forecast US presidential elections have profited from both a substantial increase in pre-election polls at the state level and very efficient modeling strategies (Lauderdale & Linzer, 2015; Linzer, 2013), and have achieved spectacular success (Linzer, 2013; Silver, 2012). However, these models have been developed for a very specific context—essentially, two-candidate races with long historical records—and cannot be applied to other settings without further work. If an election involves races of more than just two parties and in a multitude of constituencies, forecasters can hardly operate on hundreds of constituency-level vote intention polls. This renders approaches that are driven largely by poll-averaging algorithms useless for anyone who is interested in constituency-level forecasts.

This paper proposes a framework for the forecasting of electoral outcomes at the constituency level. While local data on voter preferences are generally sparse at any individual election, the electoral history of constituencies and forecast models can be used to correct the forecasts for any election-invariant bias, thus ultimately improving the model performances. In addition, I also suggest the pooling of information by combining several available models. The technical procedure breaks down to three stages. In the first stage, distinct constituency-level forecasts are produced (or collected) for past elections. In this application,

E-mail address: simon.munzert@mzes.uni-mannheim.de.

http://dx.doi.org/10.1016/j.ijforecast.2016.12.001
0169-2070/

these are a set of forecasts based on past election results and a set based on constituency-level polling data. As the forecasts of these models might be both imprecise and biased because of issues of data scarcity and disparity and/or fundamental flaws in the models, I unbias the forecasts by regressing previous actual results on the forecasts and a set of additional predictors (if available). The out-of-sample forecasts are corrected simultaneously. In the third step, the corrected forecasts are combined and weighted according to their performances in past elections.

The use of the proposed procedure is demonstrated through the presentation of a forecast of the German federal election held in September 2013—a multi-party setting with a considerable number of districts (299). The forecast was published shortly before the election. The corrected and combined forecasts perform very well, with the forecast distribution of seats being almost identical to the actual outcome of the election. Thus, the forecast outperforms a set of existing and newly-implemented benchmark forecasts.

The paper makes several contributions to the literature on the forecasting of parliamentary election outcomes. First, it provides an alternative approach to integral models (as suggested by Hanretty, Lauderdale, & Vivyan, 2016, for example) that allows for flexible model inputs and helps in identifying and correcting for election-invariant local biases of individual forecasting components. This makes it easier for political theorists and practitioners alike to learn from past forecasting efforts. Secondly, it provides a straightforward way to synthesize information from several sources that may vary in terms of their availability over time. This adds to existing approaches that operate at the national or state level (e.g., Lewis-Beck & Dassonneville, 2015; Rothschild, 2015). In the suggested framework, different forecasting components can be combined at the constituency level, while accounting for their predictive performances in previous races. Thirdly, to the best of my knowledge, the paper presents the first comprehensive attempt to exploit the massive constituency-level polling data in order to forecast election outcomes in a Bayesian hierarchical modeling framework.[1]

2. Motivation

Forecasting models that provide genuine information about constituency-level outcomes have great potential. First, constituency-level outcomes are important for the actual distribution of seats at the national level in many electoral systems with biased votes–seats curves. For example, in majoritarian systems like the United Kingdom, the relationship between the aggregated share of votes and the share of seats a party gains is blurred by the plurality rule at the district level. In the general election of 2015, the winning Conservatives gained 51% of the seats, but only 37% of the popular vote. It has been shown that the bias in votes–seats curves is not constant over time (see e.g. Blau, 2004; Jackman, 1994; Tufte, 1973), which makes it difficult to forecast the distribution of seats from national-level polls alone. This has become manifest in a compendium of twelve academic forecasts of the very same 2015 British General Election, published in *Electoral Studies* (Fisher & Lewis-Beck, 2016): while all models substantively underestimated the tremendous success of the Conservatives, who won an absolute majority of seats, the model that came closest to the actual numbers of seats gained by the two major parties was the only one that did not rely on the national-level vote share but forecasted seat outcomes directly using data on citizens' expectations about constituency winners (Murr, 2016). In general, specific characteristics of electoral rules may prohibit forecasters from inferring both constituency- and national-level outcomes from trends that are identified in national-level polls. However, the most important statistic of interest in parliamentary elections is the distribution of seats, as it determines who effectively comes to power. Thus, forecasting at the level at which the race is decided is the most promising strategy for anticipating the prospective allocation of seats.

Constituency-level forecasts can also be useful in settings that are not purely majoritarian. The internal composition of parliaments in mixed electoral systems with a strong PR component, such as the German electoral system, could be forecast more precisely using constituency-level information (Manow, 2011). Moreover, such forecasts could prove valuable for local candidates and party campaign strategists, as they may reduce the uncertainty and provide information as to where local effort should be concentrated.

On the other hand, forecasting dozens or even hundreds of outcomes in one election is a more ambitious task than merely generating the expected national-level vote shares for a few parties or candidates. Fig. 1 illustrates some of the challenges involved by providing an overview of all constituency-level election results for the two major parties (the Conservatives, CDU/CSU, and the Social Democrats, SPD) in Germany for the last six elections. There is a considerable degree of variance around the overall mean of the constituency-level results. Moreover, the district deviations from the mean are not stable over time. Lines that run parallel to the aggregated trend indicate that time-invariant, district-specific components are highly predictive for constituency-level party shares. However, there is a considerable number of cases that run counter to the trend. Constituency-level outcomes are not generated as simple projections of a national-level swing on the local level. Regional and constituency-level factors are likely to play a role too, such as candidates' campaigning skills, scandals, district history, or constituency-level politics or economic performance—in fact, the very factors which are included in standard theory-driven national-level forecasting models, but the effects of which might vary depending on the context. As I will argue in the following section, the existing forecasting models tend to ignore constituency-specific information at the cost of precision. On the other hand, those who try to use current campaign data often face problems of data scarcity.

The main challenge of disaggregated forecasting lies in the collection of data that are informative at the constituency level, as this helps in explaining and anticipating the heterogeneity of the results. Up-to-date public opinion

[1] See Hanretty et al. (2016) for a recently published effort.

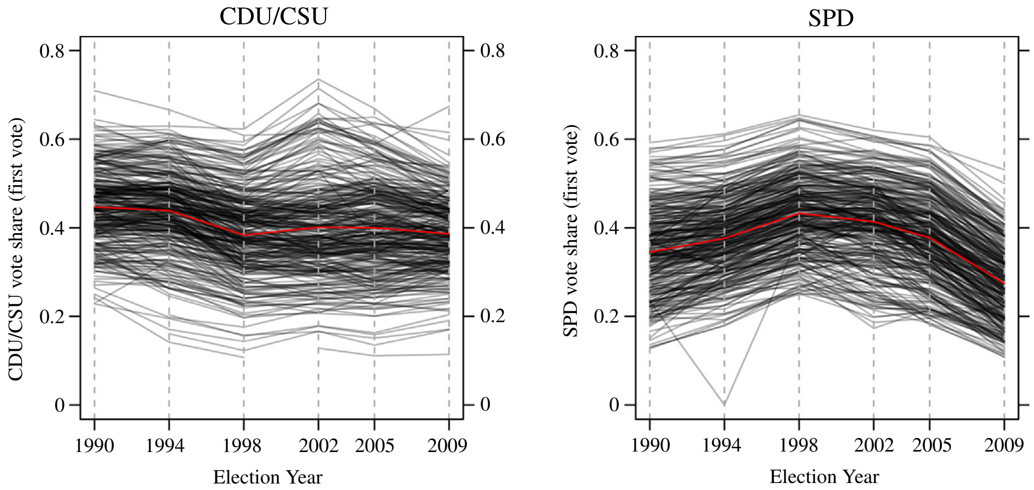

Fig. 1. District level election results from German Bundestag elections for the parties CDU/CSU and SPD. The red line is the mean share of the party's first vote over all districts; the black lines display constituency-level first vote shares.

data provide an important predictor in many of the current forecasting approaches. However, political polls are usually conducted at the national level, meaning that information on individual districts is sparse, and a considerable number of districts might not even be covered by the survey sample at all. Furthermore, the boundaries of electoral districts are often subject to redistricting so as to correct for population changes in districts' territories. This makes it difficult to inform models with district history predictors like the former winner, previous vote shares, or number of times a party has taken the district over a certain number of elections. These factors boil down to the fundamental dilemmas of constituency-level election forecasting: the scarcity of information at this very level on the one hand, and the multitude of forecasts to be made on the other.

3. State of the art

Most of the established forecasting models that target the outcomes of parliamentary elections can be divided into four general types, depending on (a) the forecast level (national or constituency outcomes) and (b) the type of information used.[2]

The first branch comprises national-level models that forecast aggregate vote or seat shares using structuralist information and regression analysis. These have been developed in the context of elections for the US House (for recent applications, see for example Abramowitz, 2010; Lewis-Beck & Tien, 2012), the British House of Commons (e.g., Lebo & Norpoth, 2011; Whiteley, Sanders, Stewart, & Clarke, 2011), the German Bundestag (e.g., Gschwend & Norpoth, 2005; Kayser & Leininger, 2016), the Spanish Cortes Generales (Magalhães, Conraria, & Lewis-Beck, 2012), and many other settings. These models commonly build on a sparse set of economic and public-opinion-based predictors, which are shown to be correlated strongly with the party vote or seat shares of interest. While these models tend to produce fairly accurate forecasts given their sparseness, they have been criticized for their weak data basis, misrepresentation of uncertainty, and tendency to overfit the data (Gelman, 1993; Lauderdale & Linzer, 2015; van der Eijk, 2005). Moreover, they are not informative about constituency-level campaign dynamics in the absence of any local component.

Another set of models relies on the information from national opinion polls. Instead of merely taking published vote intentions as forecasts, the idea is to exploit systematic trends in the historical relationship between polls and the vote that can be affected by institute- or party-specific biases or cyclical patterns that occur over the course of a campaign. Models of this type have been developed for Westminster elections (e.g., Fisher, 2015; Fisher, Ford, Jennings, Pickup, & Wlezien, 2011), the Australian federal election (Jackman, 2005), and the German federal election (Selb & Munzert, 2016), among others. Poll-based approaches exploit data that are almost a natural by-product of national-level election campaigns. In contrast to structural approaches, they often allow dynamic forecasts of public opinion. On the other hand, such models have to operate at the level of the polls, which is the national level more often than not. Furthermore, they are designed to capitalize on the existence of systematic errors (e.g., house effects) in trial-heat polls, which is not necessarily a given (see Selb & Munzert, 2016). Depending on the context of elections, some of these models incorporate algorithms which take the seat–vote bias into account. This can be

2 This review does not discuss the poll-aggregating approaches that have been developed recently for US Senate races (e.g., Linzer, 2013; Sides, Highton, & McGhee, 2014; Silver, 2014), as they pose an exception in terms of data availability and electoral rules, and hardly translate to parliamentary settings in which constituency-level outcomes are the quantities of interest. They work at the state level and can draw on large amounts of polling data. These types of models are usually based on a firm Bayesian model framework and provide dynamic forecasts by design; they are built to combine historical and polling data, placing more weight on the latter as the election date approaches and more and more polls accumulate over the course of the campaign.

done by either applying the 'cube rule' (see Whiteley, 2005) or predicting the bias using data from previous elections (see e.g. Nadeau, Lewis-Beck, & Bélanger, 2009; Whiteley et al., 2011). However, such approaches are hardly robust to regional shifts in party support or the emergence of new parties, which makes the volatility of the bias difficult to predict.

A third group of models combines national-level polling and historic constituency data. These models essentially project swings in public opinion, that is, shifts from one party to the other based on past election results and current polling trends, on previous constituency-level election results. Given the assumption that the change in vote shares is constant over all constituencies, the underlying mechanism is often referred to as 'Uniform Swing' or 'Uniform National Swing' (e.g., Miller, 1972; Payne, 1992). After projecting the anticipated swing on previous district results, the forecast local winners are aggregated again to arrive at a seat-share forecast. While the simplicity of this approach certainly has its own charm,[3] it relies on the critical assumption that swings are distributed equally over constituencies or are likely to 'cancel out' (Butler & Van Beek, 1990, p. 179). Thus, more sophisticated variants introduce regional and tactical swing parameters or add information about incumbency status (Bafumi, Erikson, & Wlezien, 2008, 2010). Ultimately, this approach generates constituency-level forecasts but does not incorporate any local campaign information, meaning that the accuracy of this methodology essentially hinges on the forecast of the national trend. A very recent forecasting effort to advance this branch of models was presented by Hanretty et al. (2016), who suggested an integrated model that combines current national constituency polls, historical election results and census data in order to forecast the 2015 Westminster election. Their procedure provides a set of desirable features, such as dynamic and consistent probabilistic forecasts for both district- and national-level outcomes, and takes into account the compositional nature of multiparty election outcomes. Thus far, this approach has only been customized for the British case.

Another group of models exploits new data sources that offer short-term information on candidates' winning odds at the constituency level. For example, Murr (2011) uses citizens' expectations about election outcomes in their own district, exploiting the wisdom-of-crowds effect. The idea is that aggregated group forecasts will outperform individual forecasts if the individual probability of a correct forecast averages more than 0.5. While it is difficult to generate precise constituency-level forecasts from national-level survey data on peoples' *vote intentions*, asking for local *expectations* is far more efficient. Following a similar logic, Wall, Sudulich, and Cunningham (2012) exploit constituency-level betting markets in Westminster elections. For US house elections, Sides et al. (2014) incorporate fundraising data in the pre-primary model and replace it with candidate information during the general campaign. These approaches provide important contributions to the existing set of forecasting tools, as they overcome the votes-to-seats problem by design, and are able to capture local campaign dynamics that can easily be overlooked in models that are based on overall trend measures. However, they still rely on exotic survey instruments or other data that are likely to be unavailable in many scenarios (as reported by Murr, 2011), or have been shown to have no additional predictive power relative to traditional approaches (see Wall et al., 2012).

Depending on the context, there are various models that perform reasonably well in the aggregate but are unable to generate precise constituency-level forecasts. Models that try to assess the local level often fail to incorporate actual local information on the race, lack the necessary data in most contexts, or produce underwhelming results. Hanretty et al. (2016) suggested one way of combining national-level with constituency polling data, but their strategy is not necessarily applicable to other contexts, though it works in the UK setting.

4. A correction–combination procedure

So, how can we improve constituency-level forecasts further? As was discussed above, the main challenge that needs to be addressed is data scarcity at the local level, especially when there is reason to assume that local campaign dynamics deviate from the national trend. This leads to the production of imprecise and potentially biased forecasts by simple models that operate on scarce local information. I suggest two strategies for alleviating this problem: first, model correction based on the identification of constituency-level bias in previous forecasts; second, model combination. This approach does not start with the development of a singular forecast model, as the data availability may vary by setting. Instead, it builds upon existing sets of constituency-level forecasts and offers a framework for the improvement of their predictive performances. The basic reasoning behind my approach is that, while many existing forecasting models exploit past information, e.g., by incorporating previous constituency-level election results into the regression equation or by evaluating the historical relationship between seats and votes, they neglect what can be learned from past forecasting efforts *per se*. Since constituency-level models produce vast numbers of forecasts from election to election, this provides a great source of information for further improvements in forecasting performance.

Similar bias-correcting strategies have been used previously to improve forecasts of weather conditions (Bao, Gneiting, Grimit, Guttorp, & Raftery, 2010; Glahn & Lowry, 1972), daily bank transaction volumes (Mabert, 1978), and commodity prices (Issler, Rodrigues, & Burjack, 2014), for example. However, the existing strategies capitalize on extensive time series datasets, whereas I present an approach that fits the given data structure of a panel of constituency-level election outcomes (N, usually large) for a set of elections (K, usually small), allowing information to be borrowed over both time and constituencies.

[3] In fact, the simple intuition of these models is the main reason for their popularity among pundits and the media; the approach has been featured on various online platforms, e.g., http://www.electoralcalculus.co.uk/userpoll.html for Westminster constituency forecasts.

4.1. Correcting for election-invariant bias

Consider a set of forecasts for previous elections, f_{pjk}, where $p = 1, \ldots, P$ indexes the pth party, $j = 1, \ldots, J$ the jth district, and k the corresponding election from a set of K past elections.[4] As historical constituency-level outcomes are known, forecast models performances on past elections can be evaluated. To this end, I model the actual vote share y of party p in district j at election k as a function of a constant α_p, the forecast itself weighted by β_p (both of which are allowed to vary by party), a vector of further covariates $\mathbf{X}$ (where the coefficient γ_p is also allowed to vary by party), and party-district random effects (ξ_{pj}). Formally, I assume that y_{pjk} follows a normal distribution with mean μ_{pjk} and party-specific variance σ_p^2,

$$y_{pjk} \sim N(\mu_{pjk}, \sigma_p^2), \tag{1}$$

with

$$\mu_{pjk} = \alpha_{p[jk]} + \beta_{p[jk]}^{\text{forecast}} \mathrm{f}_{pjk} + \gamma_{p[jk]}\mathbf{X} + \xi_{pj}. \tag{2}$$

The core idea behind this model is that existing forecasting models, such as the uniform swing model, can suffer from systematic election-invariant biases or a serial correlation of constituency- and/or party-specific forecasts. Decomposing past outcomes into the parts that are explained by the forecasting model and other systematic and random components allows us to identify and correct for biases, drawing on the weights of the link function, α_p, $\beta_p^{\text{forecast}}$ and γ_{pj}, as well as the party-district errors, ξ_{pj}. If the forecasting model provides unbiased forecasts, we would expect $\alpha = 0$, $\beta^{\text{forecast}} = 1$ and $\gamma = 0$ for all p, and $\xi = 0$ for all p and j. The proposed link function in Eq. (2) could be expanded further by, for example, introducing additional regressors which are motivated by the forecast scenario and method. For instance, a model that is ignorant of local dynamics could be evaluated using information on the number of candidates running in a district, the incumbency status of each, or other campaign information, if available. If constituencies are subject to frequent politically-motivated redistricting, a variable indicating whether or not redistricting occurred in a district could be included in this evaluation step. Furthermore, one could also include party election-specific errors in order to absorb over- or underestimations of party vote shares at a specific election. However, such election-specific errors cannot be identified for a true forecast, and are therefore disregarded here. The model's suggested specification includes a linear interpolation of the original forecasts that is specific to each party (the linear relationship is allowed to vary over parties) and captures party-district idiosyncratic effects that are missed by systematic components. This accounts for the fact that parties may perform systematically better or worse in specific districts, relative to the national or regional trend.

Next, the estimated coefficients and error components can be used to correct the forecasts made in a new setting for election k^*. This could be done separately by estimating Eq. (2) first and correcting the set of raw forecasts f_{pjk^*} in a second step.[5] However, a more natural way to correct an out-of-sample forecast in the Bayesian estimation framework that I follow in the application below is to integrate the to-be-forecast cases into the set of modeled data and keep the values of $y^*_{pjk^*}$ as missing data. The missing values will then be treated as stochastic nodes and the Gibbs sampler will return imputations from the posterior predictive density, conditional on the other parameter draws (Gelman & Hill, 2007, p. 367). The advantage of this integrated imputation approach is that it generates a natural quantification of the uncertainty through a simulation of the posterior predictive densities of all $y^*_{pjk^*}$ as a by-product of the correction procedure, which can be used to compute constituency-level winning probabilities, probability densities for the aggregated distribution of seats, or other quantities of interest.

One remaining caveat of this correction procedure concerns the danger of overfitting: the estimated regression weights are based on in-sample data from previous elections, with the only missing data being the to-be-forecast outcome of the current election. This induces overconfidence in the estimated weights, which will manifest itself in overconfident forecasts. To guard against this problem, I implement a simple leave-one-out procedure in the correction step: I run the model repeatedly, each time treating one of the previous elections as unobserved with regard to the electoral outcomes. At the end, I then pool the multiple runs into one simulation matrix and use this to identify the posterior quantities (see Tables 2 and 4 later). In addition, the out-of-sample simulations are used to calibrate the weights needed in the model combination step (see below).

This general way of simultaneously evaluating forecasting performances and correcting for historical bias enables forecasters to exploit the information from a single model more efficiently, as the set of constituency-level forecasts is large, which makes the identification of sources of bias computationally feasible. I expect the extent of the benefits of this step to be related to both model and election characteristics. If the original set of forecasts is generated using a model that already incorporates specific constituency-level predictors, introducing this information as a possible source of forecast bias in the evaluation step is of little assistance. However, if one prefers to construct 'naïve' models in the first step, or to borrow forecasts from others (probably even without a full knowledge of the original model specification), or when qualitative forecasts are used, the evaluation overhead allows their accuracy to be improved, as the model pulls itself up by its own bootstraps. Note that this procedure may also be of use in the process of developing and improving an new constituency-level forecasting model.

[4] There is an infinite number of possible ways to produce such forecasts, depending on the information that is available (and relevant) for specific elections or institutional settings. In the test case that I present below, I employ two different models. As the notation indicates, a necessary condition is that the single model provides constituency-specific forecasts for both the current election and a set of past elections ($K \geq 2$).

[5] To do so, one would proceed using an equation like: $y^*_{pjk^*} = \hat{\alpha}_{p[jk*]} + \hat{\beta}_{p[jk*]}\mathrm{f}_{pjk*} + \hat{\beta}_{p[jk]}X + \hat{\xi}_{pj}$.

4.2. Pooling models according to constituency-level prediction performance

While the first step allows more information to be incorporated into single models ex post in order to exploit forecast error patterns in the original model, the second involves the pooling of multiple forecasts from different models. As was discussed above, the existing approaches to constituency-level election forecasting come with their own unique strengths and weaknesses. For instance, uniform swing models are very easy to implement but do not use local-level campaign information. On the other hand, models which incorporate current district data (e.g., Murr, 2011) can capture recent local trends but may not cover the whole set of constituencies. More generally, models that focus exclusively on past election results or other economic or political indicators naturally fail to predict the success of new parties, by construction. On the other hand, the dynamics of public opinion make vote intention surveys a shaky prediction instrument (see Gelman & King, 1993). At the district level, it is not clear a priori whether there are election-, population-, or data-specific conditions under which one approach has a clear advantage. One could hypothesize that survey-based approaches are inferior when the survey data at hand are extraordinarily sparse or of poor quality. At the same time, survey data should be expected to perform relatively better when the electoral setting has altered substantively between elections, e.g., because of massive redistricting, local strongholds of a new party on the block or new candidates, because such data do not rely on previous (and less useful) information.

As has been demonstrated frequently, combining several forecasts helps to reduce the forecast variance because more information is exploited and bias which is immanent in single models can be cancelled out in the aggregate forecast (e.g., Armstrong, 2001; Bates & Granger, 1969; Clemen, 1989; Graefe, 2014; Raftery, Gneiting, Balabdaoui, & Polakowski, 2005). The strategy of combining several forecasts has already been applied to the forecasting of US presidential elections; for example, Montgomery, Hollenbach, and Ward (2012a,b) promote the use of ensemble Bayesian model averaging, which provides a weighting algorithm based on each component's past performance and uniqueness. Similarly, Graefe, Armstrong, Jones, and Cuzán (2014) pool forecasts within and between methods, but use equal weights.

The combination procedure that I employ here uses weights that are based on the past performances of the corrected forecasts, measured at the constituency level. Suppose that one has two different constituency-level forecasting models, where the first relies mainly on historical information, producing a set of forecasts f^{hist}_{jk*}, while the other exploits current polling information, leading to a different set of forecasts f^{polls}_{jk*}. As was described above, these forecasts can be improved by applying the correction procedure, leading to f^{hist*}_{jk*} and f^{polls*}_{jk*}. A simple solution for combining the two would be to employ equal weights and just take the mean of the forecast values for each district. However, this would disregard the fact that models can perform differently in general, or produce forecasts which, on average, fit well in some districts but less well in others, even after correction. I therefore employ a marginally more sophisticated procedure. First, I estimate the party district-specific forecast variance, that is, the mean square error $\sigma^2_{\cdot,pj}$, for every set of corrected forecasts.[6] This is a measure of how well the corrected forecast model performed in previous elections. As it is plausible that one model may perform better in one subset of districts and worse in another, the forecast variance is estimated for every party at the district level. In the two-forecasts case with f^{hist*}_{pjk*} and f^{polls*}_{pjk*}, the weighting proceeds as follows:

$$f^{comb}_{pjk*} = \omega_{pj} f^{hist*}_{pjk*} + (1 - \omega_{pj}) f^{polls*}_{pjk*}, \tag{3}$$

with

$$\omega_{pj} = \frac{\hat{\sigma}^2_{polls,pj}}{\hat{\sigma}^2_{polls,pj} + \hat{\sigma}^2_{hist,pj}}. \tag{4}$$

The weight ω_{pj} of the corrected district forecast is inversely proportional to the estimated district-specific forecast variance of previous corrected forecasts. In other words, the more reliable a forecast model proves to be for one specific district relative to another model, the more weight is attached to forecasts from this more reliable model.[7] Overall, the expected benefit from combining individual models is twofold. First, it helps to reduce any remaining model-specific bias that may be induced by factors that are idiosyncratic to the out-of-sample case and could not be eliminated in the first step. Secondly, as was shown by Bates and Granger (1969, pp. 463–464), it also tends to yield an improvement in the forecast variance, due simply to the fact that the combined forecast relies on more information. If individual models produce very similar forecasts, i.e., are highly correlated, the increase in precision from combination is expected to be smaller.

5. Application: forecast of the German Bundestag election 2013

This section presents out-of-sample forecasts of the outcomes of district races at the German Bundestag election 2013 that were produced two weeks before Election Day (September 22nd, 2013). It provides a setting that includes an appreciable number of electoral districts (299) and considerable regional heterogeneity in party support. Six parliamentary parties entered the race, and some published polls indicated that two emerging parties—the

[6] The weights have to be party district-specific because the models are run separately for each party. In addition, I also use out-of-sample models for previous elections to extract the predicted residuals.

[7] Dropping the election, constituency and party indices for ease of exposition, the combination formula easily generalizes to $f^{1*}, \ldots, f^{N*}$ forecast components (see Bates & Granger, 1969). The combined forecast is then calculated as $f^{comb} = \omega^1 f^{1*} + \omega^2 f^{2*} + \cdots + \omega^N f^{N*}$, with $\omega^n \in [0, 1]$ and $\sum_{x=1}^{N} \omega^x = 1$. The weights are normalized by the sum of all forecast variances across models, i.e., $\omega^n = \frac{\hat{\sigma}_n^{-2}}{\hat{\sigma}_1^{-2} + \hat{\sigma}_2^{-2} + \cdots + \hat{\sigma}_N^{-2}}$. Note that the forecast variances are inverted in order to ensure that the weight of the nth forecast component is inversely proportional to its variance.

Piratenpartei (Pirate Party) and the *Alternative für Deutschland, AfD* (Alternative for Germany)—had realistic chances of entering parliament. In the German electoral system at a federal level, people have two votes. The first is a personal vote for a candidate in the voter's district (at this level, candidates are elected via first-past-the-post voting), while the second is a party vote which essentially determines the distribution of the seats in parliament (Saalfeld, 2008).

This application focuses on forecasts of first vote shares, as they are more likely to be affected by local campaign dynamics than the second (party) vote shares, and thus pose a more challenging exercise. Furthermore, forecasting first vote shares and district winners is the more relevant application for constituency-level forecasts because they affect the internal composition of parliamentary groups in parliament directly. No published efforts at constituency-level election forecasting existed in Germany in the leadup to the election.[8] I therefore employ two different forecasting models. The first relies mainly on historical constituency-level data and poll-driven, national-level trend estimates, while the second draws on a large, individual-level polling database that is mapped to the constituency level.

5.1. *Exploiting past election results*

In established democracies, past political outcomes are usually a good predictor of future outcomes. In the German case, we can observe a strong relationship between first vote shares across elections, especially for the two large parties CDU/CSU and SPD (see also Fig. D.1 in the supplementary materials). Any differences between current and past election outcomes can be attributed to either nationwide trends or local factors such as incumbency, local campaign efforts, and other local or regional determinants.

I demonstrate the virtue of the suggested procedure by employing a very ascetic uniform swing approach for the exploitation of past election results. Specifically, I draw on constituency-level first vote shares in the past election and add a uniform swing constant for each candidate (i.e. party) which mirrors the national trend of the respective party from one election to the next. As was discussed above, the attractiveness of this approach lies in its simplicity. On the other hand, it rests on some heroic assumptions, such as that there is a stable set of parties over time, or that there is no regional variation in shifts. In addition, no potential bottom or ceiling effects for constituency-level vote shares are taken into account. Furthermore, it is assumed that every district has a preceding district, whereas in actual fact districts are sometimes abandoned or newly created, and most elections have at least some of the constituencies being subject to redistricting. Online Appendix A presents a way of constructing the district panel data set which is employed in the forecast.

The swing component of the model is identified by using an estimate of the current national trend to calibrate constituency-level election forecasts. For past elections, I use the final aggregated party vote shares of the first vote in order to avoid disturbances from polling data. For the current election, the trend is estimated by drawing on an existing forecast of the national-level outcome (Selb & Munzert, 2016).[9] Next, the constituency-level forecasts are computed as the difference between the current national-level vote share (or polling estimate) for each party and the relative position of a district for each party in the past election.

The upper panel of Table 1 shows that the uniform swing model generally performs well on past elections since 1994 (see also Fig. D.2 in the Online Appendix for a graphical display of the model's performance). This is reflected in the rather small mean absolute errors (absolute difference between the actual and predicted first vote shares), which mostly do not exceed three percentage points. On average, 90% of district winners are forecast correctly.

Next, I employ the suggested procedure for identifying and correcting forecast biases. Following Eq. (2), I model past constituency-level outcomes as a function of past forecasts, and include two further predictors: a dummy variable *party dominance*, which indicates whether a party has won the direct mandate in the district at least the last three times in a row, and a dummy variable *pioneer*, which indicates whether the party is running in the district for the first time.[10] I suspect that the *party dominance* dummy captures some of the underestimation of the first vote shares which should occur when party strongholds are affected less by national losses, for example.[11] The *pioneer variable* is used to correct for a natural problem in the uniform swing forecast: when a party runs in a district for the first time, the model predicts a vote share of zero (plus/minus the national swing). This is probably an underestimation of the actual outcome, which is technically counterbalanced by a party-specific estimate of β^{pio}. I adopt MCMC methods for inference and use vague prior distributions for the model parameters, and run the simulations using WinBUGS (Lunn, Thomas, Best,

8 The only exception that I am aware of is the semi-commercial platform election.de, which has sold constituency-level forecasts prior to the last three federal elections. These forecasts are not freely accessible and are therefore disregarded in the forecast here. However, I use the aggregated forecasts from this source to evaluate the overall performance of my model at a later stage.

9 The forecast party shares were: CDU/CSU 38.1%, SPD 28.2%, B'90/Die Grünen 13.5%, Die Linke 7.7%, and FDP 5.4%, which were originally published several weeks in advance of the election. This is equivalent to the following trends: CDU/CSU +4.3%, SPD +5.2%, B'90/Die Grünen +2.8%, Die Linke −4.2%, and FDP −9.2%.

10 This was the case in some districts in the early 1990s, where B'90/Die Grünen and Die Linke (formerly PDS) had not run for a direct mandate.

11 Furthermore, the indicator serves as a rough proxy for a party's incumbency status, i.e., whether the party's candidate is running for re-election in this constituency or not, information that was not available at the time of designing this model. Meanwhile, a research team led by Anna-Sophie Kurella, Franz Urban Pappi, and Thomas Bräuninger compiled these data in an effort to estimate the personal incumbency advantage (Kurella, Pappi, & Bräuninger, 2016), and thankfully made them available to me. This allowed me to see ex post how information on incumbency improves the forecasting performance. The accuracy gains were very modest in size (see Tables D.3 to D.5 in the online appendix).

Table 1
Predictive performance of the uniform swing model, uncorrected and corrected forecasts.

	CDU/CSU	SPD	FDP	B'90/Die Grünen	Die Linke	% overall correct
Uncorrected						
1994	0.021	0.025	0.019	0.014	0.024	90.0
1998	0.025	0.017	0.005	0.011	0.008	88.6
2002	0.030	0.027	0.009	0.011	0.012	87.0
2005	0.023	0.022	0.011	0.009	0.016	92.6
2009	0.031	0.024	0.013	0.014	0.011	89.6
Corrected						
1994	0.022	0.023	0.013	0.012	0.019	90.6
1998	0.025	0.017	0.007	0.010	0.008	88.3
2002	0.031	0.028	0.009	0.011	0.014	87.3
2005	0.022	0.023	0.010	0.009	0.017	92.3
2009	0.027	0.022	0.014	0.014	0.012	90.0

Notes: The first five columns report mean absolute errors over all 299 districts at each election. The last column reports the percentage of districts that were forecast correctly (predicted winner equals actual winner). Cells in which the corrected forecast outperforms the uncorrected forecast are highlighted in grey.

Table 2
Bayesian median estimates and 95% credible estimates for the model of party first vote shares, based on the uniform swing model (see Eq. (2)).

Predictor		95% CI
Intercept α		
—CDU/CSU	0.033	[0.001; 0.049]
—SPD	0.022	[0.011; 0.031]
—FDP	0.010	[0.002; 0.020]
—B90/Die Grünen	−0.002	[−0.006; 0.005]
—Die Linke	0.001	[−0.005; 0.006]
Uniform swing estimate β^{swing}		
—CDU/CSU	0.914	[0.873; 0.997]
—SPD	0.945	[0.919; 0.974]
—FDP	0.796	[0.528; 0.965]
—B90/Die Grünen	1.005	[0.894; 1.074]
—Die Linke	1.009	[0.972; 1.093]
Party dominance β^{dom}		
—CDU/CSU	0.004	[−0.004; 0.015]
—SPD	−0.004	[−0.010; 0.005]
—FDP	−0.161	[−0.210; 0.019]
—B90/Die Grünen	−0.013	[−0.492; 0.430]
—Die Linke	0.001	[−0.022; 0.030]
Pioneer β^{pio}		
—CDU/CSU	0.047	[−0.708; 0.848]
—SPD	0.293	[−0.066; 0.363]
—FDP	0.014	[−0.007; 0.034]
—B90/Die Grünen	0.026	[0.019; 0.048]
—Die Linke	−0.008	[−0.012; 0.006]
Party-constituency-level variance σ^2_ξ	0.000	[0.000; 0.001]
Residual variance σ^2_η		
—CDU/CSU	0.034	[0.031; 0.036]
—SPD	0.031	[0.029; 0.034]
—FDP	0.016	[0.013; 0.017]
—B90/Die Grünen	0.019	[0.017; 0.020]
—Die Linke	0.023	[0.020; 0.026]
N	8970	

& Spiegelhalter, 2000). The code of the sampler is reported in Online Appendix C.

The results are presented in Table 2. The near one-to-one relationship between the true and forecast values is mirrored in the estimates of α and β^{swing}, although there is some between-party variation in the effects: uniform swing forecasts for the two large parties (CDU/CSU and SPD) and the FDP are discounted to a certain extent, indicating some regression-to-the-mean dynamics that are not captured fully by the uniform swing model. The estimated effects for party dominance are negligible.[12] The pioneer predictor is useful for correcting the forecasts of SPD candidates in constituencies in which the party did not run in the previous election (though this is a historically

12 Note that the large negative coefficients for FDP and B'90/Die Grünen have no empirical relevance because those parties were virtually never dominant in any of the constituencies.

negligible scenario). The estimated party-constituency-level variance is zero, which indicates that the forecasts are not distorted substantively by other district-level factors that have not been accounted for.

Using the corrected results for prediction in order to assess the performance of the unbiased model, it turns out that the corrections do tend to improve the model's performance, although only to a very modest extent. This is mirrored in the lower part of Table 1. The mean absolute errors over all districts for each party at the previous five elections decrease in the majority of cases, and the overall share of districts that are forecast correctly increases marginally in three of the five elections, with an average amount of about 90%. These results imply that the uncorrected uniform swing model is not affected severely by systematic biases that persist over time. This is probably not too surprising, as the model essentially builds on past local information: while it is obvious that a national trend projection leads to under- or overestimates of district-specific party vote shares, the procedure tends to be robust to time-persistent, party-specific biases.

Nevertheless, the estimates are used to correct the forecast for the 2013 election, which is generated on the basis of a national swing forecast according to the party share estimates of Selb and Munzert (2016). According to this estimate, both the uncorrected and corrected forecasts suggest a distribution of the direct seats as shown in Table 5. The district-level forecasts are listed in Table D.2 in the online appendix.

5.2. Exploiting constituency-level polling data

The main limitation of uniform swing-based forecasts is that they fail to incorporate campaign-specific, constituency-level information. The uniform swing model is ignorant of local campaign dynamics unless they are incorporated in the correction procedure explicitly. Constituency-level polls on voting intentions can serve as a remedy to this problem, but such data are rarely available, and the 2013 German election campaign was no exception. However, I can capitalize on an exceptionally rich polling database, provided by the German *Forsa* institute.[13] *Forsa* surveys 500 respondents each working day and asks them about their vote intention for the next general election, among other things. This allows me to construct a forecasting model which incorporates local information, unlike the uniform swing model.[14]

I stabilize the constituency-level forecasts by employing the Bayesian modeling strategy suggested by Selb and Munzert (2011) to estimate constituency preferences using survey data and geographic information. The method is presented in depth by Selb and Munzert (2011), so I limit my description to the gist of the matter. The vote intention of a respondent in a constituency is modeled as a function of a global mean, the individual voting behavior at the last election, a constituency-level covariate (log inverse district size) and two constituency-level random effects, one of which is assumed to vary independently and identically across districts following a normal distribution, and another which is required to follow an intrinsic conditional autoregressive (CAR) distribution (see Besag, York, & Mollié, 1991). As true party vote shares for past elections are known at the district level, the predicted probabilities of voting for a party are then weighted according to the recalled voting coefficient (see Park, Gelman, & Bafumi, 2004). For the estimation procedure, I pool survey data over the period from five months to one month before the election date, so as to be able to draw reasonable numbers of respondents per district.[15] Table D.1 in the online appendix provides summary statistics for the polling data utilized.[16]

The prediction results from the previous three elections are reported in Table 3. No more than 85% of the district outcomes are forecast correctly when using the uncorrected estimates.[17] Therefore, I again apply the correction strategy and regress actual first vote shares in each district in the three elections on the poll model predictions and party–constituency random effects (see again Eq. (2)). The results are presented in Table 4. While the relationship between actual first vote shares and the poll model forecast is nearly one-to-one on average, the party-specific slopes and intercepts reveal a substantive bias in the original model. Specifically, the SPD and Die Linke vote shares are underestimated in constituencies in which the parties' candidates performed well, whereas the opposite is true for the other parties' candidates, for which the model corrects the original forecasts towards the mean (slopes < 1). Furthermore, the estimated variance of the party district-specific errors is substantive, indicating that there are other factors that are not accounted for at the district level which play a role in the transformation from second vote polling to first vote shares.

The gain in the polling model as a result of the correction procedure is impressive (see Table 3 and Fig. D.3 in the online appendix). The fit of the forecasts improves

13 Data for previous years are available from http://www.gesis.org/en/elections-home/other-surveys/forsa-bus/.

14 Note that no alternative data sources, such as local betting markets or vote expectation surveys which target the local level, were available.

15 If the chosen time window is too narrow, the constituency-level estimates would tend to rely more on the grand mean instead of the local (or neighboring) information, which would limit the ability of the model to capture actual local preferences as desired.

16 As has been described above, it is challenging to assign electoral districts to respondents. While most election studies provide such identification variables, polling data that are not used primarily for scientific purposes sometimes come with no geographical identifier at all, or locate respondents in some way other than by electoral units. The *Forsa* data come with identifier variables of German administrative units. There are several ways to attach district identifiers to respondents, which I discuss in more detail in Online Appendix B. In short, though, I identify all possible districts for each respondent and assign the respondent to one of them at random. With regard to the forecasting method that uses respondents from neighboring districts to estimate vote intentions, an exact match should not improve the forecasting performance substantively relative to this simplifying approach.

17 See also Fig. D.3 in the online appendix, which visualizes the relationship between the poll-based estimation results and the actual election results. Although the fit seems to be rather good, the polls are biased significantly. For instance, SPD vote shares tend to be underestimated.

Table 3
Predictive performance of the polling model, uncorrected and corrected forecasts.

	CDU/CSU	SPD	FDP	B'90/Die Grünen	Die Linke	% overall correct
Uncorrected						
2002	0.034	0.070	0.043	0.014	0.025	85.0
2005	0.033	0.105	0.026	0.016	0.024	62.5
2009	0.051	0.056	0.057	0.019	0.034	83.3
Corrected						
2002	0.026	0.027	0.007	0.007	0.016	93.6
2005	0.040	0.037	0.005	0.007	0.017	78.6
2009	0.032	0.025	0.007	0.010	0.017	88.6

Notes: The first five columns report party-specific mean absolute errors over all 299 districts in each election. The last column reports the percentage of districts forecast correctly (predicted winner equals actual winner). Cells in which the corrected forecast outperforms the uncorrected forecast are highlighted in grey.

Table 4
Bayesian estimates of the model of party first vote shares, based on the polls model.

Predictor		95% CI
Intercept α		
—CDU/CSU	0.024	[0.007; 0.111]
—SPD	0.007	[−0.051; 0.092]
—FDP	0.005	[−0.011; 0.021]
—B90/Die Grünen	0.010	[0.005; 0.015]
—Die Linke	−0.007	[−0.013; 0.002]
Polls estimate β^{polls}		
—CDU/CSU	0.948	[0.743; 1.084]
—SPD	1.182	[0.982; 1.510]
—FDP	0.582	[0.362; 0.690]
—B90/Die Grünen	0.904	[0.840; 1.020]
—Die Linke	1.442	[1.376; 1.537]
Party-constituency-level variance σ_ξ^2	0.024	[0.023; 0.026]
Residual variance σ_η^2		
—CDU/CSU	0.041	[0.028; 0.046]
—SPD	0.036	[0.031; 0.042]
—FDP	0.009	[0.008; 0.011]
—B90/Die Grünen	0.015	[0.011; 0.018]
—Die Linke	0.028	[0.025; 0.031]
N	5,980	

considerably over all parties and years, in terms of both mean absolute errors and correctly predicted district winners.

Next, I generate the polling forecasts for the 2013 election. According to the uncorrected forecast, 290 districts are attributed to the CDU/CSU and nine to the SPD, mirroring the great advantage of the CDU/CSU in the raw polls. The corrected forecasts iron out this bias to a certain extent, but this forecast is still significantly more favorable to the conservative parties than the corrected uniform swing model, with 270 vs. 224 seats for CDU/CSU and 27 vs. 70 seats for the SPD.

5.3. Forecast combination and retrospective evaluation

In the final step, the two forecasting components are combined, following Eq. (3). Thus, I compute party district-specific weights ω_{pj} based on the ratios of the estimated forecast variances of each out-of-sample model (see Eq. (4)). As Fig. 2 illustrates, neither of the forecast components seems to be superior in general. As the range of weight distributions shows, there is a substantial level of variation between constituencies. Furthermore, performances vary between parties: while more weight is attached to the historical swing model for SPD vote share forecasts, the polling model seems to perform better in most of the districts for the FDP.

Next, the weights are used to combine the individual forecasting models and compute an aggregated forecast. This is done separately for each party and district and over all iterations of the BUGS simulations. The party with the highest vote share after this computation is taken as the winner in each of the iterations, and the final forecast of the winner for a district is determined by aggregating the winner of each of the iterations and identifying the party with the most wins.

The aggregate results are reported in Table 5, while the constituency-level forecasts are reported in Table D.2 in the online appendix. The results are split by the underlying model (uniform swing, polling, or the combination of both) and whether the correction procedure was applied or not. In addition, I also present results for the combined model with equal weights (that is, where the uniform swing and the polling component contribute equally to the final forecast) in order to see whether taking performance-based weights mattered.

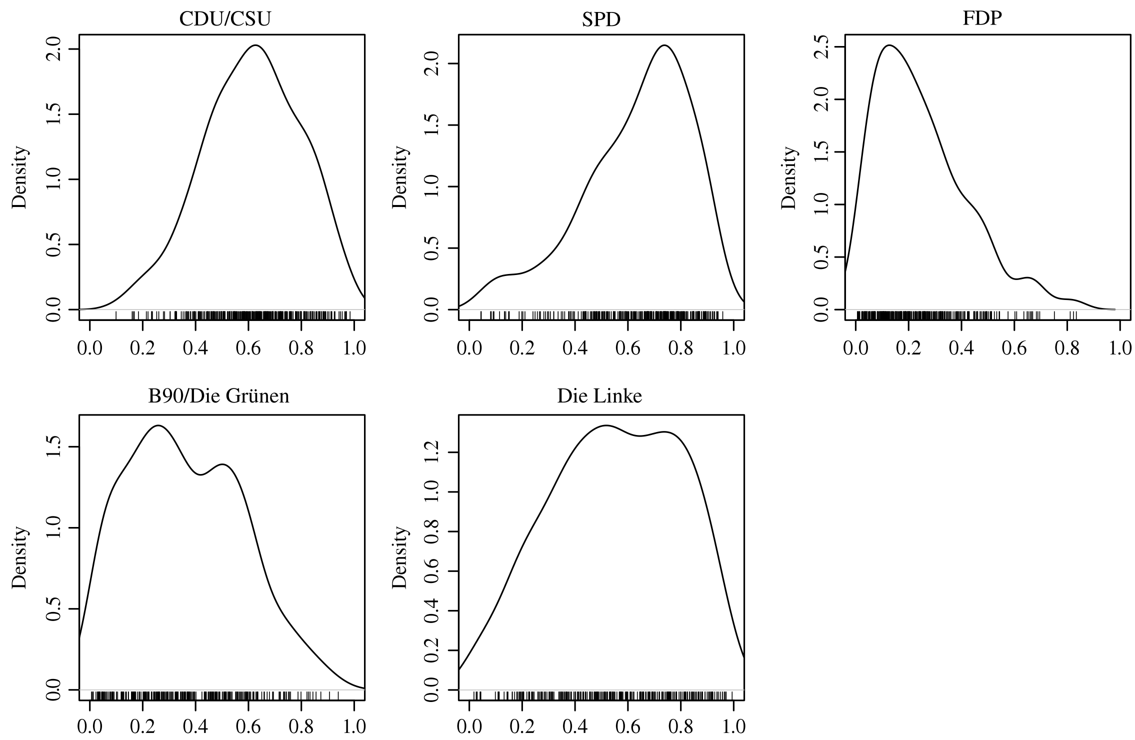

Fig. 2. Distribution of the forecast weight parameter (see Eq. (4)). Large values are weights in favor of the historical model, small values are weights in favor of the polling model. Ticks on the horizontal axis indicate single constituency weights.

Table 5
Distribution of forecast winners over parties, by model.

		Seats forecast (direct mandate), by party					Σ dev
		CD(S)U	SPD	FDP	B'90	Linke	
Uncorrected	Uniform swing	224	70	0	1	4	24
	Polling	290	9	0	0	0	108
	Combined (eq. wgt.)	264	31	0	1	4	55
	Combined	229	65	0	1	4	14
Corrected	Uniform swing	223	71	0	1	4	26
	Polling	270	27	0	1	1	68
	Combined (eq. wgt.)	245	50	0	1	3	18
	Combined	242	53	0	1	3	12
Other	*election.de* (14.09.13)	224	69	0	1	5	24
	spiegel.de (21.09.13)	181	89	0	3	13	97
Actual result		236	58	0	1	4	

Eyeballing the forecast distributions of seats per party by model, we see that the individual models arrive at substantively different forecasts: the uniform swing model favors the Social Democrats, while the polling model strongly favors the Conservatives. As expected from the results above, applying the correction procedure makes hardly any difference for forecasts based on the uniform swing model, but does make a difference for the polling model, in which the significant bias towards the conservative party is somewhat discounted. Furthermore, the combination step leads to a true compromise forecast, as a consequence of the evenly distributed weights. The decision as to which weights to employ does matter, but less so after the correction has been applied. As a first benchmark of forecasting performances, I compared the distribution of forecast seats with the true distribution (see last row in Table 5) by summing the absolute differences between the forecasted and actual numbers of seats per party (see the last column of the same table). According to this statistic, the combined and corrected model fares best, with a total deviation of only 12 out of 299 seats. However, the uncorrected combined model with performance-based weights performs similarly well. The uniform swing model performs slightly worse, with total deviations of 24 seats (uncorrected) and 26 seats (corrected), respectively. In contrast, the poll-based model performs worse but profits relatively more from the correction step. These results confirm the previously stated expectations. If the underlying model already incorporates historical data and is rather robust to election-invariant bias, the correction procedure may be of little use. If a model is driven largely by current information but is vulnerable to election-invariant bias, the correction step will contribute to better forecasts. Moreover, the combination procedure proved to be very powerful in this setting because the two individual components erred in different directions. However, combination might be less useful in other scenarios.

I found two other sources of constituency-level forecasts against which my forecasting models can be compared. The first is provided by the semi-commercial platform election.de.[18] The authors remain silent about the details of their forecasting technique, but seem to implement a uniform swing-type model. Indeed, their forecasts are virtually identical with those of my uniform swing model, and underestimate the performance of the Conservatives. Another forecast approach was provided by the online newspaper spiegel.de, which aggregated user expectations about the outcomes in their home districts.[19] Apparently, the sample of users was strongly biased towards the left parties, as the aggregated forecast underestimated the Conservatives' performance significantly. This again indicates that individual models that rely mainly on one source of information for informing constituency-level forecasts (a poll-derived national trend in the election.de model and expectations of the readership in the spiegel.de forecast) can be biased significantly. If components of this bias are stable over time (such as systematic, election-invariant trends in polls or readership preferences), they can be corrected in the future by following my suggested procedure.

Next, I computed additional benchmarks that compare actual and forecast vote shares, and derived winners at the constituency level. The statistics are presented in Table 6. First, I computed the mean absolute errors (MAE) of all constituency-level party vote share forecasts. One could regard this measure as the single most important benchmark because it is based on the raw quantity of interest, and other outcomes, such as the winning party or candidate, are derived from it. The MAE is consistently lower for the corrected set of forecasts, ranging from 2.5 percentage points for the corrected and combined model to 4.6 percentage points for the uncorrected polling model. Looking at the percentages of district outcomes that were forecast correctly, though, the pattern is less clear. According to this benchmark, the uncorrected combined forecast provides the best result, with 93.3% of the constituency winners being forecast correctly, but the differences from other well-performing models are minor. Again, the uniform swing models perform remarkably well given the limited information they use. We see a similar picture when focusing on the subset of marginal seats. I define seats as marginal when the difference between the two frontrunner candidates is less than 10%.[20] Those cases are expected to be much more difficult to forecast, and the share of districts that are forecast correctly is markedly lower in this subset for all of the models, with the combined uncorrected model performing best (80.8% of cases correct) and the uncorrected polling model performing worst (56% correct). I discuss these results in the concluding section.

As was outlined above, the simulations can also be used to derive a natural quantification of the uncertainty of the estimates at both the constituency and national levels. Fig. 3 provides probability distributions of the aggregate number of forecast seats per party, generated from all iterations. It shows that, according to the simulations, one would expect the CDU/CSU to gain roughly 225–260 direct mandates in total, the SPD between 35 and 70 seats, Bündnis 90/Die Grünen one seat (which, according to virtually all models, is a very safe one), and Die Linke three seats, with a probability around 80%. The multimodal distributions of the forecasted CDU/CSU and SPD seats might seem odd at first sight, but are in line with the correction procedure and the chosen leave-one-out approach in order to prevent overfitting. The forecasts are sensitive to the information available from past elections. Consequently, dropping parts of this estimation leads to different forecasts, and the forecasts mirror this uncertainty. If the estimates were based entirely on in-sample estimates, the distributions would be single-peaked, much narrower, and overconfident.

What can be learned from the districts for which the corrected and combined forecast failed? Table 7 lists the districts that correspond to erroneous forecasts. By and large, the districts that were predicted wrongly were those in which the winning margin, i.e., the difference between the winner's and the second runner's vote share, was very small—less than four percentage points in 75% of cases. The more interesting cases are at the bottom of the table. I argued above that it is extraordinarily difficult to incorporate district-level election-specific information into the model. The polling model was built with this aim, but this is blurred somewhat by the smoothing mechanism of the model itself. In the case of the Freiburg district, it was known in advance that the left parties SPD and Bündnis 90/Die Grünen had both nominated popular candidates, thereby mutually taking off votes. For the München-Nord district, the CSU incumbent managed to retain the seat against a newcomer SPD candidate. Indeed, none of the models considered either incumbency status or the popularity of candidates. One lesson that can be learned from these erroneously predicted cases is that it can be worthwhile to collect more campaign information and integrate it into the models. However, this is often a costly endeavor, and therefore is rarely done.[21]

18 See http://www.election.de/cgi-bin/content.pl?url=/img/poll/btw_wp_130914.html (accessed 7 October 2016).

19 The original website, wahlwette.spiegel.de, has already been taken offline; a summary of the last forecasts, which depart slightly from the reported figures, can be accessed at http://www.spiegel.de/politik/deutschland/bundestagswahl-wahlwette-von-spiegel-online-a-923650.html (accessed 7 October 2016).

20 See Wall et al. (2012) for a similar procedure.

21 In fact, one of my efforts to forecast the outcome of the 2013 election involved an online survey among local journalists, who were asked to report their expectations of the outcomes in their home and neighboring constituencies. However, the response rate was underwhelming (fewer than 150 of the 1990 journalists invited took part), which prevented me from integrating these data into the forecasting model.

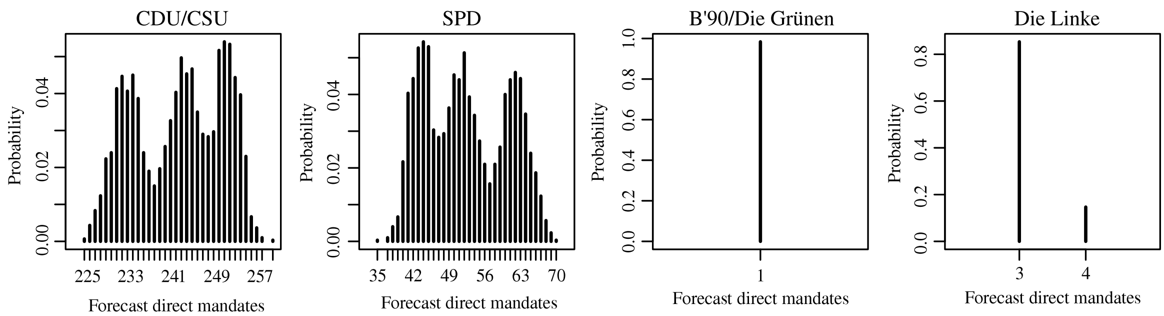

Fig. 3. Distribution of forecast seats per party, corrected and combined model.

Table 6
Forecast benchmark statistics, by model.

		MAE	% correct	% correct, marg. dist.
Uncorrected	Uniform swing	0.031	92.3	78.8
	Polling	0.046	81.6	55.6
	Combined (eq. wgt.)	0.031	89.6	68.7
	Combined	0.028	93.3	80.8
Corrected	Uniform swing	0.029	92.0	77.8
	Polling	0.030	88.3	66.7
	Combined (eq. wgt.)	0.026	92.0	75.8
	Combined	0.025	93.0	78.8

6. Discussion and conclusion

Electoral research that takes the role of institutions seriously should address the level at which votes are transformed into seats, in terms of both theory and analysis. In many electoral systems, this is not (only) the national level but (also) the electoral district level. Thus, any attempts to forecast elections in such systems should be aware of local outcomes. This paper proposes a correction–combination strategy for forecasting electoral outcomes at the constituency level. The approach accounts for biases in forecast time series that may occur because the original forecast model fails to account for important election-invariant predictors. Furthermore, the constituency-specific weighting of several forecasting components provides an easy way of exploiting information from several sources. I demonstrate the use of the procedure by forecasting first vote shares at the 2013 German Bundestag election, drawing on historical district and disaggregated polling data. Both the correction and combination components improve the forecasts relative to a variety of benchmarks.

While the out-of-sample performance of the suggested approach was remarkable—the model based on correcting the uniform swing and the polling component and combining the two based on past performances produced an MAE of only 2.5% and forecasted the winner of 93% of the 299 constituency races correctly—the concurring uniform swing model, which is very simple to implement, also produced rather satisfactory forecasts. So why should one consider employing a more complex procedure in the first place? First, staying with this example, the reduction in MAE from 3.1% to 2.5% may seem negligible, but can be decisive when the number of close races is higher. Secondly, it is surely possible for less sophisticated models to perform as well as or even better than more complex approaches in any single instance. However, the uniform swing model essentially hinges on getting the national-level trend correct; if one fails to do so, the consequences for the constituency-level forecasts are dramatic. Furthermore, even if the national trend is right, it can still lead to bad forecasts if the trend obscures regional deviations.[22] A combination of several approaches can safeguard against such a scenario. In addition, the lack of local information can be problematic when the election history is disrupted by factors such as the entry of new parties.

Obviously, I do not claim that the correction–combination procedure involves any magic that will lead to precise forecasts every time. The disaggregated evaluation of the steps has shown that not all models are subject to time-invariant biases that can be exploited, and that the success of the combination procedure relies on the assumption that individual model errors will cancel each other out in the aggregate. This happened to be the case in this application, where the uniform swing model underestimated the conservatives' success while the polling model overestimated it, but if several models err in the same direction or rely on similar information, the combination step will be of little use.[23]

The contribution of this paper to the current state of election forecasting is threefold. First, it offers an easy-to-implement approach to the removal of bias from

22 In fact, such a scenario seems likely for the next Bundestag election, where the right-wing populist party AfD is expected to become a major party in parliament and to perform particularly well in Eastern Germany.

23 A fruitful extension of the approach could be to apply more sophisticated model averaging procedures, such as ensemble Bayesian model averaging (Montgomery et al., 2012a; Raftery et al., 2005), in order to take the uniqueness of individual models' forecasts into account.

Table 7
Districts predicted wrongly (based on a combination forecast of the corrected models).

District name	Combined	Actual result	Margin
Essen III	SPD	CDU	0.000
Waldeck	SPD	CDU	0.002
Köln I	CDU	SPD	0.003
Potsdam	SPD	CDU	0.004
Oldenburg	CDU	SPD	0.006
Bonn	CDU	SPD	0.007
Bielefeld	CDU	SPD	0.008
Hildesheim	SPD	CDU	0.010
Gifhorn	CDU	SPD	0.012
Kaiserslautern	CDU	SPD	0.013
Darmstadt	CDU	SPD	0.014
Berlin-Neukölln	CDU	SPD	0.017
Leverkusen-Köln IV	CDU	SPD	0.020
Hamburg-Wandsbek	CDU	SPD	0.026
Prignitz	SPD	CDU	0.039
Hamburg-Eimsbüttel	CDU	SPD	0.042
Berlin-Pankow	SPD	Die Linke	0.044
Freiburg	SPD	CDU	0.049
Berlin-Charlottenburg	SPD	CDU	0.056
München-Nord	SPD	CSU	0.118

any type of constituency-level forecasting model with a certain historical record. The procedure allows for the incorporation of additional substantive predictors that are not part of the original model, but also works by merely identifying district-level party-specific biases in historical forecasts. This can be useful for testing whether a model produces locally or party-specific biased forecasts, even if forecasters are not interested in modifying their model or combining it with other information or models. Secondly, the paper demonstrates the benefit of model combination at the constituency level. Previous applications that combined several models for election forecasting purposes operated at the national level, but I use party- and district-specific weights to offer a flexible way of exploiting past models' performances in order to generate a new, synthetic forecast. Finally, the paper offers an alternative strategy to both approaches that rest upon massive amounts of data, which are unlikely to be realized due to time and money constraints, and highly sophisticated modeling efforts that try to integrate information from different sources into a single forecasting model (e.g., Hanretty et al., 2016; Linzer, 2013).

The application presented in the paper provided a first test case for the method. However, there are other, potentially more relevant, applications in which district-level outcomes are essential for the distribution of seats in parliament, as is the case in first-past-the-post systems such as the United Kingdom. It is easy to argue that a knowledge of the constituency-level distribution of voting preferences is of great value for estimating the distribution of seats in the parliament under such conditions. Furthermore, constituency-level forecasts provide an opportunity for evaluating the reasons for the forecasting performances of different approaches (e.g., model-based vs. survey-based forecasts). In general, analyses of forecast biases are rare in election forecasting; however, they could encourage discussions of the strengths and weaknesses of the various approaches on the market.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.ijforecast.2016.12.001.

References

Abramowitz, Alan I. (2010). How large a wave? Using the generic ballot to forecast the 2010 midterm elections. *PS: Political Science and Politics*, *43*(4), 631–632.
Armstrong, J. Scott (2001). Combining forecasts. In J. Scott Armstrong. (Ed.), *Principles of forecasting: A handbook for researchers and practitioners*. Kluwer p. Norwell.
Bafumi, Joseph, Erikson, Robert S., & Wlezien, Christopher (2008). Forecasting house seats from generic congressional polls. In Wendy Alvey, & Fritz Scheuren (Eds.), *Elections and exit polling*. New York: Wiley and Sons.
Bafumi, Joseph, Erikson, Robert S., & Wlezien, Christopher (2010). Forecasting house seats from generic congressional polls: The 2010 midterm election. *PS: Political Science and Politics*, *43*(04), 633–636.
Bao, Le, Gneiting, Tilmann, Grimit, Eric P., Guttorp, Peter, & Raftery, Adrian E. (2010). Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction. *Monthly Weather Review*, *138*, 1811–1821.
Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Society*, *20*(4), 451–468.
Besag, Julian, York, Jeremy, & Mollié, Annie (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, *43*(1), 1–20.
Blau, Adrian. (2004). A quadruple whammy for first-past-the-post. *Electoral Studies*, *23*, 431–453.
Butler, David, & Van Beek, Stephen D. (1990). Why not swing? Measuring electoral change. *PS: Political Science and Politics*, *23*(2), 178–184.
Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*, 559–583.
Fisher, Stephen D. (2015). Predictable and unpredictable changes in party support: A method for long-range daily election forecasting from opinion polls. *Journal of Elections, Public Opinion and Parties*, *25*(2), 137–158.
Fisher, S. D., Ford, R., Jennings, W., Pickup, M., & Wlezien, C. (2011). From polls to votes to seats: Forecasting the 2010 British general election. *Electoral Studies*, *30*(2), 250–257.
Fisher, Stephen D., & Lewis-Beck, Michael S. (2016). Forecasting the 2015 British general election: The 1992 debacle all over again? *Electoral Studies*, *41*, 225–229.
Gelman, Andrew. (1993). Review of *Forecasting elections* by Michael S. Lewis-Beck and Tom W. Rice, Washington, DC. *Public Opinion Quarterly*, *57*(1), 119–121.
Gelman, Andrew, & Hill, Jennifer (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
Gelman, Andrew, & King, Gary (1993). Why are american presidential election campaign polls so variable when votes are so predictable? *British Journal of Political Science*, *23*(4), 409–451.
Glahn, Harry R., & Lowry, Dale A. (1972). The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorolgy*, *11*(8), 1203–1211.
Graefe, Andreas. (2014). Accuracy of vote expectation surveys in forecasting elections. *Public Opinion Quarterly*, *78*(S1), 204–232.
Graefe, Andreas, Armstrong, J. Scott, Jones, Randall J., Jr., & Cuzán, Alfred G. (2014). Combining forecasts: An application to elections. *International Journal of Forecasting*, *30*(1), 43–54.
Gschwend, T., & Norpoth, H. (2005). Prognosemodell auf dem Prüfstand: Die bundestagswahl 2005. *Politische Vierteljahresschrift*, *46*(4), 682–688.
Hanretty, Chris, Lauderdale, Benjamin, & Vivyan, Nick (2016). Combining national and constituency polling for forecasting. *Electoral Studies*, *41*, 239–243.
Issler, Joao Victor, Rodrigues, Claudia, & Burjack, Rafael (2014). Using common features to understand the behavior of metal-commodity prices and forecast them at different horizons. *Journal of International Money and Finance*, *42*, 310–335.

Jackman, Simon. (1994). Measuring electoral bias: Australia, 1949–93. *British Journal of Political Science*, *24*(3), 319–357.
Jackman, Simon. (2005). Pooling the polls over an election campaign. *Australian Journal of Political Science*, *40*(4), 499–517.
Kayser, Mark A., & Leininger, Arndt (2016). A predictive test of voters' economic benchmarking: The 2013 German Bundestag election. *German Politics*, *25*(1), 106–130.
Kurella, Anna-Sophie, Pappi, Franz Urban, & Bräuninger, Thomas (2016). Using panel fixed effects analysis to estimate personal incumbency advantage without bias. Unpublished manuscript.
Lauderdale, Benjamin E., & Linzer, Drew (2015). Under-performing, over-performing, or just performing? The limitations of fundamentals-based presidential election forecasting. *International Journal of Forecasting*, *31*(3), 965–979.
Lebo, Matthew, & Norpoth, Helmut (2011). Yes, Prime Minister: The key to forecasting British elections. *Electoral Studies*, *30*(2), 258–263.
Lewis-Beck, Michael S., & Dassonneville, Ruth (2015). Forecasting elections in Europe: Synthetic models. *Research and Politics*, *2*(1).
Lewis-Beck, Michael S., & Tien, Charles (2012). Election forecasting for turbulent times. *PS: Political Science and Politics*, *45*(04), 625–629.
Linzer, Drew A. (2013). Dynamic Bayesian forecasting of presidential elections in the states. *Journal of the American Statistical Association*, *108*(501), 124–134.
Lodge, Milton, Steenbergen, Marco R., & Brau, Shawn (1995). The responsive voter: Campaign information and the dynamics of candidate evaluation. *American Political Science Review*, *89*(2), 309–326.
Lunn, David J., Thomas, Andrew, Best, Nicky, & Spiegelhalter, David (2000). WinBUGS–a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–337.
Mabert, Vicent A. (1978). Forecast modification based upon residual analysis: A case study of check volume estimation. *Decision Sciences*, *9*, 285–296.
Magalhães, P. C., Conraria, L. A., & Lewis-Beck, M. S. (2012). Forecasting Spanish elections. *International Journal of Forecasting*, *28*(4), 769–776.
Manow, Philip. (2011). The cube rule in a mixed electoral system: disproportionality in German Bundestag elections. *West European Politics*, *34*(4), 773–794.
Miller, W. L. (1972). Measures of electoral change using aggregate data. *Journal of the Royal Statistical Society, Series A (General)*, *135*(1), 122–142.
Montgomery, Jacob M., Hollenbach, Florian, & Ward, Michael D. (2012a). Improving predictions using ensemble bayesian model averaging. *Political Analysis*, *20*(3), 271–291.
Montgomery, Jacob M., Hollenbach, Florian M., & Ward, Michael D. (2012b). Ensemble predictions of the 2012 US presidential election. *PS: Political Science and Politics*, *45*(04), 651–654.
Murr, Andreas Erwin. (2011). ''Wisdom of crowds''? A decentralised election forecasting model that uses citizens' local expectations. *Electoral Studies*, *30*, 771–783.
Murr, Andreas E. (2016). The wisdom of crowds: What do citizens forecast for the 2015 British General Election? *Electoral Studies*, *41*, 283–288.
Nadeau, R., Lewis-Beck, M. S., & Bélanger, É (2009). Election forecasting in the United Kingdom: a two-step model. *Journal of Elections, Public Opinion and Parties*, *19*(3), 333–358.
Panagopoulos, Costas. (2009). Campaign dynamics in battleground and nonbattleground states. *Public Opinion Quarterly*, *73*(1), 119–129.
Park, David K., Gelman, Andrew, & Bafumi, Joseph (2004). Bayesian multilevel estimation with poststratification: State-level estimates from national polls. *Political Analysis*, *12*(4), 375–385.
Payne, Clive. (1992). Statistical methods for election forecasting in the United Kingdom 1970–90. *British Elections and Parties Yearbook*, *2*(1), 138–158.
Raftery, Adrian E., Gneiting, Tilmann, Balabdaoui, Fadoua, & Polakowski, Michael (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, *133*, 1155–1174.
Rothschild, David. (2015). Combining forecasts for elections: Accurate, relevant, and timely. *International Journal of Forecasting*, *31*(3), 952–964.
Saalfeld, Thomas. (2008). Germany: Stability and strategy in a mixed-member proportional system. In Michael Gallagher, & Paul Mitchell (Eds.), *The politics of electoral systems* (pp. 209–230). Oxford: Oxford University Press.
Selb, Peter, & Munzert, Simon (2011). Estimating constituency preferences from sparse survey data using auxiliary geographic information. *Political Analysis*, *19*(4), 455–470.
Selb, Peter, & Munzert, Simon (2016). Forecasting the 2013 German Bundestag election using many polls and historical election results. *German Politics*, *25*(1), 73–83. Pre-election version available at: http://dx.doi.org/10.2139/ssrn.2313845.
Sides, John, Highton, Ben, & McGhee, Eric (2014). How election lab works. http://www.washingtonpost.com/news/politics/wp/2014/05/05/how-election-lab-works/.
Silver, Nate (2012). http://fivethirtyeight.blogs.nytimes.com/fivethirtyeights-2012-forecast/.
Silver, Nate (2014). How the FiveThirtyEight senate forecast model works. http://fivethirtyeight.com/features/how-the-fivethirtyeight-senate-forecast-model-works/.
Tufte, Edward R. (1973). The relationship between seats and votes in two-party systems. *American Political Science Review*, *67*(2), 540–554.
van der Eijk, Cees. (2005). Election forecasting: A sceptical view. *The British Journal of Politics and International Relations*, *7*, 210–214.
Wall, Matthew, Sudulich, Maria Laura, & Cunningham, Kevin (2012). What are the odds? Using constituency-level betting markets to forecast seat shares in the 2010 UK General Elections. *Journal of Elections, Public Opinion and Parties*, *22*(1), 3–26.
Whiteley, P. F. (2005). Forecasting seats from votes in British general elections. *The British Journal of Politics and International Relations*, *7*(2), 165–173.
Whiteley, P., Sanders, D., Stewart, M., & Clarke, H. (2011). Aggregate level forecasting of the 2010 general election in Britain: The Seats–Votes model. *Electoral Studies*, *30*(2), 278–283.
Wlezien, Christopher, & Erikson, Robert S. (2002). The timeline of presidential election campaigns. *Journal of Politics*, *64*(4), 969–993.

Simon Munzert is postdoctoral researcher at the Data and Methods unit at the Mannheim Centre for European Social Research, University of Mannheim, Germany. His research interests include methods of election forecasting, public opinion measurement, and political behavior. He has published in several international journals, including *Political Analysis*, *Journal of the Royal Statistical Society* (Series A), *Political Science Research and Methods*, and *Social Science Computer Review*.