Contributions to the Measurement of Public Opinion in Subpopulations

Dissertation zur Erlangung des akademischen Grades eines

Doktors der Sozialwissenschaften

vorgelegt von

Simon Munzert

an der

Universität Konstanz

Sektion Politik – Recht – Wirtschaft

Fachbereich Politik- und Verwaltungswissenschaft

Tag der mündlichen Prüfung: 09. Juli 2015

1. Referent: Prof. Dr. Susumu Shikano, Universität Konstanz

2. Referent: Prof. Dr. Peter Selb, Universität Konstanz

3. Referent: Prof. Michael D. Ward, PhD, Duke University

Konstanzer Online-Publikations-System (KOPS)
URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-0-296799

To Stefanie

Acknowledgments

I would not have been able to write and finish this dissertation without the help and support of many people. First of all, I want to thank my collaborators, Peter Selb and Paul Bauer, for their great work and support. Their contributions to the research presented in this thesis are of no small concern.

The single papers received manifold and invaluable feedback over the course of their creation. In particular, Peter Selb and I are grateful to Michael Herrmann, Thomas Hinz, Winfried Pohlmeier, Susumu Shikano as well as the editors and reviewers of *Political Analysis* for helpful comments and support. This work was also supported by the Center for Quantitative Methods and Survey Research at the University of Konstanz.

The second paper that I wrote together with Paul Bauer also got valuable support. We are grateful to Delia Baldassarri for providing materials of her and Andrew Gelman's analysis to us. We give the raters of the ALLBUS survey items our most sincere thanks for their contribution. Furthermore, we thank Klaus Armingeon, Matthias Fatke, Markus Freitag, Birte Gundelach, Daniel Stegmüller, Richard Traunmüller and Eva Zeglovits for helpful comments on previous versions of this paper. Additionally, we received extensive feedback from two anonymous reviewers of *Political Science Research & Methods*. Finally, I owe special thanks to Andreas Diekmann, who pointed me to Edward Ross and his early notes on cohesion and conflict in society.

With regards to the third paper, I would like to thank Peter Selb, Susumu Shikano and Michael Ward for many helpful comments and suggestions. Further, I thank the participants of the Konstanz-Essex PhD colloquium and the doctoral colloquia of the Graduate School for their feedback. I also thank Manfred Güllner for giving me early access to the *forsa* Bus.

I am heavily indebted to Peter Selb for his mentorship over the course of my studies in Konstanz. Finishing this project was a unexpectedly time-consuming endeavor, even more so as I took various detours along the way. I learned a lot while working on these side projects, but this would not have been possible without a permissive supervisor and enthusiastic collaborator. On that score, I am also grateful to Julian Bernauer, Michael Herrmann, Andreas Murr, Jordan Roberts, Thomas Schübel, and Susumu Shikano for great cooperation on various projects.

Further, I want to thank Michael Ward for hosting me in his lab at Duke University during summer and autumn of 2014. The time in Durham was very inspirational and gave me the opportunity to put the finishing touches on some of the papers in a superb environment.

Writing this thesis was no picnic, but I got tremendous support outside the ivory tower that sweetened it considerably. First of all, I want to thank my family for their everlasting support and love. I thank Johanna and Dirk for many trips to the Alps and a variety of kitchen experiments, and Peter, Dominic and Christian for the distraction of writing a book on web scraping. I am much obliged to Sebastian, Dag and Silvia, who were great companions during my time in Durham. Also, a big thanks goes out to Espen, Gwen, Jakob, Julian, Kristin, Matthias, Niklas, Paul, Peter, Sabine, Thomas, and Verena for pastime with good company.

And finally, I wish to thank Steffi, who motivated and supported me, for all the time and happy moments we spent together over the past five years. I hope that, at one time or another, I'll be able to make up for countless hours I spent at the desk and not on a meadow, together.

Abstract

Knowledge about public opinion in subpopulations is of increasing importance in political science research as well as in other disciplines. In this thesis, I present research that develops and extends methods which contribute to a better understanding of public opinion. In particular, the work targets political preferences at the local level, in socio-demographically defined subpopulations and across a wide range of applications.

In the first paper, we suggest a model that helps estimate preferences in constituencies based on survey data and geographic information. This is of particular importance when survey data at the constituency level are sparse and census information to stabilize survey-based estimates using poststratification is not available. The model can be used to inform studies on elite behavior and voters' preferences. It can also be used to improve survey-based estimates of any parameter for which the assumption of geographic clustering is justified.

In the second paper, we investigate polarization trends in the German public across time, issues, and various subpopulations. We replicate a study of mass polarization in the United States, presented by Baldassarri and Gelman (2008) with German data and investigate polarization trends on a large set of issues and in various subpopulations. We find that the German population has depolarized rather than polarized on many issues over the last 30 years.

In the third paper, I develop a framework for constituency-level forecasts of parliamentary elections that allows correcting individual constituency-level forecast models for systematic errors and combining them according to their past performance. I demonstrate the use of the method with an out-of-sample forecast of 299 district races from the 2013 German federal election.

Zusammenfassung

Wissen über öffentliche Meinung in Subpopulationen ist von zunehmender Bedeutung für politikwissenschaftliche Analysen und Forschung in anderen Disziplinen. In dieser Arbeit stelle ich Forschung vor, die Methoden entwickelt und erweitert, die zu einem besseren Verständnis öffentlicher Meinung beitragen. Die Arbeit beschäftigt sich insbesondere mit politischen Präferenzen auf lokaler Ebene und in auf Basis soziodemografischer Merkmale definierter Subgruppen für ein breites Spektrum von Anwendungen.

Im ersten Papier schlagen wir ein Modell vor, das dabei hilft durchschnittliche Präferenzen in Wahlkreisen auf Basis von Surveydaten und geografischer Information zu schätzen. Dies ist besonders dann relevant, wenn Umfragedaten auf Wahlkreisebene nur spärlich zur Verfügung stehen und Zensusdaten zur Stabilisierung umfragebasierter Schätzungen mittels Poststratifikation fehlen. Das Modell kann beispielsweise dazu verwendet werden, um Studien zu Eliteverhalten und Wählerpräferenzen zu informieren. Darüber hinaus kann sie dazu dienen, jegliche umfragebasierte Parameterschätzung zu verbessern, für welche die Annahme der geografischen Klumpung gerechtfertigt ist.

Im zweiten Papier untersuchen wir Polarisierungstrends öffentlicher Meinung über Zeit, Issues und Subpopulationen hinweg. Wir replizieren eine von Baldassarri und Gelman (2008) vorgelegte Studie zu Massenpolarisierung in den USA mit deutschen Daten und untersuchen Trends hinsichtlich einer großen Anzahl substantieller Fragen in verschiedenen Subpopulationen. Dabei zeigen wir, dass sich die deutsche Bevölkerung über die vergangenen 30 Jahre eher depolarisiert als polarisiert hat.

Im dritten Papier entwickle ich ein Framework zur Vorhersage von Wahlergebnissen bei Parlamentswahlen auf der Wahlkreisebene, das systematische Fehler einzelner Vorhersagemodelle korrigieren und mehrere Modelle gemäß ihres Abschneidens bei früheren Wahlen kombinieren kann. Anhand einer Out-of-sample-Vorhersage von 299 Rennen um das Direktmandat bei der Bundestagswahl 2013 demonstriere ich die Nützlichkeit der Methode.

Table of Contents

Chapter 1

Introduction

"All politics is local."—*Tip O'Neill*

When Thomas 'Tip' O'Neill was elected member of congress for the first time in 1952, he had already learned and internalized the mantra of his life, the phrase he should remain most famous for: "All politics is local."[1] As he describes in his autobiography *Man of the House* (O'Neill and Novak, 1988), he decided to run for his first political office in the Cambridge City Council at the age of 22 and came close to winning a seat although being considered a long shot candidate. However, this race (and by his own admission the only race he ever lost in his political career) held an important lesson: O'Neill writes that while he had come of well in many sections of the city, the vote share in his home section had been underwhelming: "I hadn't worked hard enough in my own backyard" (O'Neill and Novak, 1988, 26). O'Neill proved to have learned this lesson in later races. In the run-up to the 1982 Congressional elections, O'Neill brought a $1 billion jobs bill to the House that was publicly opposed by House Republican leader Robert H. Michel of Illinois.[2] When O'Neill went up to the rostrum, he started reading the names of ramshackle bridges located in Michel's home base Peoria and

[1]Note, however, that the quotation's origin is often falsely attributed to O'Neill, as it had appeared in The Frederick (Md.) News as early as 1932 (Shapiro, 2008).

[2]This anecdote is taken from Matthews' (1999) *Hardball*, pp.48-50.

which could be repaired under the proposed bill. The speech was broadcast live in Michel's district, exposing him during a national debate and, at the same time, at the local level.

For sure, O'Neill was not the first to recognize the importance of his maxim. Well before he coined the phrase, the notion that reelection-seeking candidates in single member plurality systems have strong incentives to stay connected with their district base by means of advertising, credit claiming, and position taking had already been formulated in political theories of dyadic representation (e.g., Eulau et al., 1959; Mayhew, 1974; Pitkin, 1967). Countless empirical assessments of dyadic representation, beginning with Miller and Stokes's (1963) seminal study, have since then provided evidence that local interest politics, both in terms of substantial distribution of pork and fuzzier ideological congruence are common practice in electoral systems with a local tier of representation.

The dissertation at hand does not provide another effort to assess the characteristics of dyadic representation in the United States or any other system. It targets the proverbial core of O'Neill's motto in another, methodological manner, and with implications not limited to the study of representation. While it is almost a truism that wherever the distribution of power has a local component, politics is local, learning about related implications for voter behavior, representation of local preferences and other phenomena is less trivial. An important building stone for locally rooted models is public opinion—including political preferences—, as it is an obvious indicator for demands and needs of the electorate. An insight which is at the heart of my dissertation papers is that the subnational analysis of public opinion is an increasingly powerful weapon to answer questions that are conceptually based on micro- or meso-level mechanisms, but have implications at a higher level, too. At the same time, more fine-grained analyses like these often require sophisticated modeling strategies and sufficiently large amounts of data. Consider the following example, which stems from a study I contributed to parallel to the work on my dissertation project.

Assessing the impact of strategic voting. Strategic (or tactical) voting in first-past-the-post elections has fascinated scholars of elections for many years. It is a textbook example of how institutions—in this case, electoral rules—shape human behavior. Voters who prefer a party which is lacking a realistic chance to succeed in a certain district face a conflict of goals: to support their mostly preferred party at the election on the one hand, and not to waste their vote on the other. Since a considerable amount of voters (estimates range between 5 and 12 percent) decide to cast their vote strategically in such settings, that is for a party they regard to be the lesser evil of two likely-to-win parties, this phenomenon may determine who comes to power in a close district race or even at the national level. While a tremendous amount of research has provided answers on important questions such as 'Who votes strategically?' or 'What is the overall amount of strategic voting?', it is a difficult endeavor to assess whether strategic voting actually matters for the allocation of parliamentary seats. This statistic cannot be derived from the overall amount of strategic voting in any straightforward manner, because strategic voting first and foremost matters on the district level. As Kim and Fording (2001) note:

> 'Ideally, to accurately gauge the political impact of tactical voting, we would like to compare observed election results (which reflect tactical voting) to the election results that would have been observed if all voters had voted sincerely. This would be easy to do if district-level public opinion polls measuring voters' sincere preferences were available prior to each election. Unfortunately, they are not.' (Kim and Fording, 2001, 296)

In Herrmann, Munzert and Selb (2015), we propose a method which helps determine the impact of strategic voting both at the district and the national level. We draw on massive amounts of survey data on party preferences that were collected prior to an election to estimate constituency-level party preferences. This is done using a small area estimation approach that we suggested previously in Selb and Munzert (2011), which is also the first part of my dissertation project. In a second step, we compare these estimates with election results, following a model of flows of strategic votes at the district level suggested by Herrmann (2012). Applying the model to the 1997 and 2001 Westminster elections, we find that the Liberal Democrats benefited most in terms of seats won due to strategic voting, although this party suffered most in terms of overall vote share because of the very same mechanism. Strategic voting varies tremendously across constituencies. Further, a considerable amount of strategic votes seems to be cast in the 'wrong direction', that is from parties that would have come second if everybody had voted sincerely to parties which would have come third. Patterns like these could not have been observed without the effort to identify political preferences at the constituency level, not to speak of constituency-level estimates of the impact of strategic voters on the race and estimated quantities for affected seats in the aggregate. In that sense, a more disaggregated analysis helped to give an educated answer on a more general question and opened avenues for further research.

The given example highlights that focusing on a more fine-grained picture of political preferences and electoral outcomes with advanced quantitative methods and sufficient data at hand is far from pedantry. Instead, this strategy offers deeper insight into prevailing questions—a constituency-level analysis helps combine insights drawn from many single pieces to the end of answering a larger puzzle.

However, it is a considerable methodological challenge to identify local needs and preferences. Disaggregated research designs use to be data hungry, and the overwhelming majority of both academic survey projects and commercial opinion polls are and will be designed as national-level studies, partly because the audience is first and foremost interested in inference on the general population, partly because any survey planned to allow design-based (not only model-based) inference on small geographic units would exceed any reasonable budget.[3] Therefore, the trend of disaggregated research perspectives that can be observed in many subfields of the discipline goes hand in hand with methodological progress in the modeling of local phenomena and the exploration of new data sources.[4] In fact, eyeballing leading polit-

[3] In a sense, the shift towards Big Data which has gained traction over the last few years can be seen as a desire for information which ordinal survey data cannot offer. Constant streams of reported opinion in social media have the potential to provide an even more fine-grained picture of sentiments in various segments of the population.

[4] See, for example, Gleditsch (2002) for the study of international relations, Weidmann, Kuse and Gleditsch (2010) for forecasting of local conflict, Murr (2011) for election forecasting at the constituency level, and Clinton (2006) for the study of subconstituency preferences.

ical science journals creates the impression that one of the major trends in the disciplines is '*all political science is local*.' This dissertation contributes to some of the named sub-fields by both developing and applying sophisticated modeling strategies that allow for geographically or demographically disaggregated analyses.

While all papers share the common goal of measuring public opinion in subpopulations, the more specific epistemological value of the work is threefold: The first paper is a methodological contribution, answering the question how geographic information can be used to improve survey-based estimates of constituency-level preferences. The second paper is largely descriptive in nature, analyzing trends of public opinion polarization in a previously unvetted society. The third paper serves the goal of forecasting, which is of scientific and general interest per se but is also increasingly popular to assess the empirical performance of theories and models in the social sciences. To achieve these goals, I draw on a variety of data sources—geographical, electoral, administrative and survey data. In the remainder of this introduction, I put each of the dissertation papers into context.

1.1 Estimating constituency-level public opinion

In electoral and representation research, recent data collection efforts have started to target the constituency level. Several comparative data projects are dedicated to systematically collect data about constituencies or are increasingly enriched with local information, like the Global Elections Database,[5] the Constituency-Level Elections Archive,[6] the Comparative Study of Electoral Systems[7] or the Comparative Candidate Survey.[8] It is inherent to many political systems around the world that electoral rules influence party or voter behavior at the local level, which in turn matters for political outcomes. For instance, they affect candidates' campaign strategies (Zittel and Gschwend, 2008), roll-call behavior (Stratmann, 2006; Sieberer, 2010),

[5] *http://globalelectionsdatabase.com/*
[6] *http://electiondataarchive.org/*
[7] *http://cses.org/*
[8] *http://comparativecandidates.org/*

pork-barrel politics (Lancaster and Patterson, 1990; Stratmann and Baur, 2002; Chen, 2010), or responsiveness to constituents' demands (Clinton, 2006; Ezrow et al., 2011; Bernauer and Munzert, 2014).

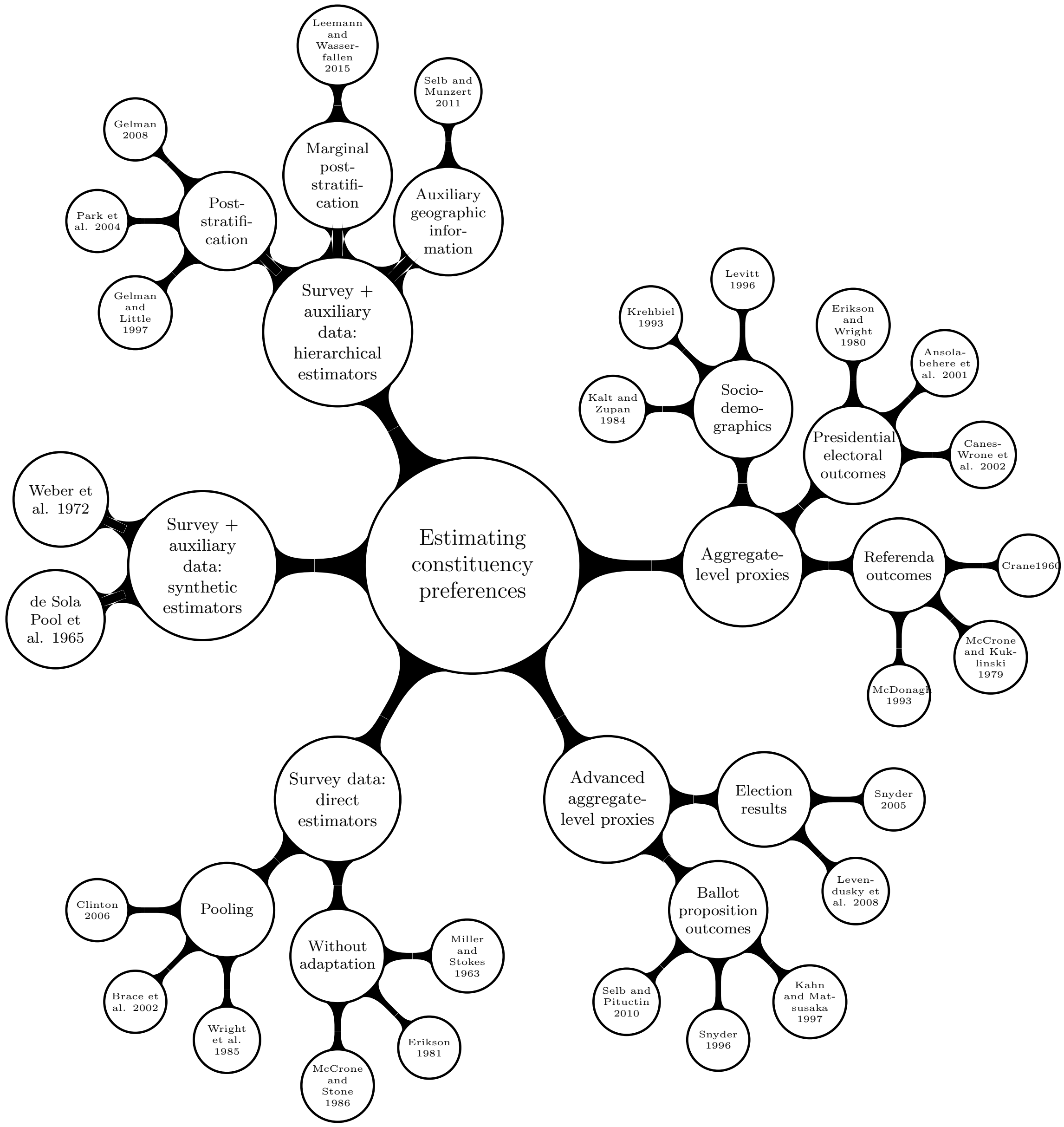
Technically, it has become more feasible to test models of voting, candidate behavior or public opinion at the local level. One reason for this is that due to the increased data collection efforts sketched above, empirical analyses stand on firmer ground. Additionally, methods to measure latent preferences have advanced over the last years, which prove especially useful when concepts of public opinion—usually captured with limited amounts of survey data—enter the equation. Model-based approaches that heavily draw on the multilevel toolbox (e.g., Gelman and Little, 1997; Park, Gelman and Bafumi, 2004) allow drawing inference from normal-sized general population surveys on district-level opinion. The first paper is a contribution to this streamline of research.

1.1.1 Previous approaches in constituency-level preference estimation

As existing work has been applied in manifold settings, the definition of constituencies varies over approaches, e.g., in focusing on relatively small units legislative districts or larger entities like federal states. Further, existing approaches differ in the data sources and the statistical techniques used for estimation. Figure 1.1 provides a schematic overview of those approaches, together with selected seminal or applied work.

The first branch comprises strategies that exploit various aggregate-level proxies of public opinion, like sociodemographic information (Kau and Rubin, 1979; Kalt and Zupan, 1984; Krehbiel, 1993; Levitt, 1996), presidential election outcomes (Schwarz and Fenmore, 1977; Erikson, 1980; Kalt and Zupan, 1984; Ansolabehere, Snyder and Stewart, 2001; Canes-Wrone, Brady and Cogan, 2002), or referendum voting returns (Crane, 1960; Hedlund and Friesema, 1972; Kuklinski, 1978; McCrone and Kuklinski, 1979; McDonagh, 1993; Gay, 2007). The benefits of these proxy variables are that they are—in contrast to large-scale state-level surveys—mostly readily available and, in the case of voting returns, reflect opinions of the voting constituency, which

Figure 1.1: Estimating constituency preferences - previous and current approaches



sometimes is the theoretically more relevant quantity of interest. On the other hand, while both sociodemographics and voting behavior are certainly correlated with political ideology or more specific preferences, using them as a surrogate raises questions of validity (Jackson, 1989; Cohen, 2006).

The second branch also relies on indirect measures—electoral outcomes—, but in this set of methods they are used to identify ideological distributions underlying multiple election (Snyder, 2005; Levendusky, Pope and Jackman, 2008; Kernell, 2009) or ballot proposition returns (Kahn and Matsusaka, 1997; Snyder, 1996; Selb and Pituctin, 2010). By identifying latent constituency partisanship via structural equation modeling or hierarchical modeling, these approaches are able to overcome problems of standard proxy-based methods like idiosyncrasies in specific election outcomes or the dubious validity of demographic characteristics. However, such estimates are limited to identifying rather general concepts of constituency preferences ('partisanship', 'ideology', 'electoral demand'), at least when they are based on elections or unrelated ballot propositions. Further, they partly assume that preferences do not change across elections (see Kernell, 2009) or require data that are only available in a very limited set of countries (ballot propositions).

This is where survey data come into play. Public opinion surveys offer direct measures of peoples' preferences on a vast set of issues. The main limitation is, however, that national public opinion surveys rarely provide sufficient information (i.e. respondents) to directly draw inferences about constituencies, let alone subgroups within legislative districts. Researchers trying to use such surveys therefore face a massive problem of data scarcity. This has become known as the 'Miller-Stokes problem' (Levendusky, Pope and Jackman, 2008, 737). In what is one of the first studies of political representation, Miller and Stokes (1963) studied the congruence between constituency opinion and representatives' roll call behavior in Congress. To do so, they used data from the University of Michigan's 1958 congressional election study which covered about 1,500 respondents in 116 of the 435 congressional districts (13 respondents on average) and asked for preferences on three policy dimensions (social welfare, foreign affairs and civil rights of African-Americans). This and similar approaches (Erikson, 1981; McCrone and Stone, 1986) suffer from extremely unstable estimates and a substantive lack of district coverage.

Several studies have tried to alleviate this problem by pooling survey data across time or pollsters in the first step and disaggregating them at the constituency level in the second. Wright, Erikson and McIver (1985) aggregate responses from 51 CBS News-New York Times phone polls over the period of 1976 to 1982 at the state level, leading to an overall sample size of about 75,000 respondents (see also Gibson, 1992; Erikson, Wright and McIver, 1993; Hill and Hinton-Anderson, 1995, for related approaches). In more recent work, Brace et al. (2002) draw on approximately 35,000 interviews from the General Social Survey conducted between 1974 and 1998 and derive a set of policy-specific measures (e.g., on political tolerance, abortion, homosexuality, and welfare). However, it is questionable whether preferences are stable over such long time periods. In another study, Clinton (2006) exploits the rare opportunity of combining two large-scale surveys conducted in 1999 and 2000 by Knowledge Networks and the National Annenberg Election Survey, leading to an overall sample size of about 100,000 respondents. The average sample size per congressional district is 232, with a minimum of 41 and a maximum of 2,099. However, it should be clear that such masses of polling data are usually not at hand, and if they are, just provide a snapshot in time because of budget constraints.[9] Further, massive pooling over pollsters usually limits the number of substantively interesting topics that are covered in all of them.

Synthetic or 'simulation-based' estimators were among the first approaches to promise a remedy to the problems of purely survey-based approaches. In conceptually groundbreaking work, de Sola Pool, Abelson and Popkin (1965) suggested to construct 480 synthetic voter types out of a variety of categorical socio-demographic variables (socio-economic status, city size, sex, religion, ethnicity, party, and region). As estimating preferences for each of these voter types would, again, have caused problems of data sparsity, they proceeded in two steps. In the first step, they estimated mean preferences for each of the categories separately using survey data. In a next step, voter types' preferences were computed as an additive function of the separate estimates. Finally, these voter type preferences were weighted using information from state-level census data to arrive at state-level, 'synthetic' preference estimates (*poststratification*). Using census data as auxiliary information not only helped stabilize the estimates, but also provided a mechanism to account for survey nonresponse or lack of coverage of specific states. Consequently, the selection of constituencies or states does not rest on the availability

[9]Note, however, that I can make use of such an exception in this thesis by drawing on massive polling data from the German *forsa* institute, which polls 500 respondents per business day since 1991 (see paper 3).

of survey data but on census information for each of the voter types, which is a relaxation in some settings. This approach was later expanded by Weber et al. (1972), who suggested the use of 960 synthetic voter types. It has to be noted, however, that synthetic approaches rest on arguably heroic assumptions. First, it is assumed that the factors used for poststratification essentially determine the distribution of preferences at the state level. As soon as other, ignored variables play a role, too, variability between states is likely to be underestimated. Secondly, the original specification does not allow for any nonlinearities in the effects of stratification variables on preferences. Thirdly, it is assumed that attitudes are unrelated to the environmental context—e.g., members of a synthetic voter type that represent a minority in one state are supposed to have the same mean attitude as the same type in another state where this groups is in the majority (see Seidman, 1975; Kuklinski, 1977, for a more extensive critique). In effect, the technique seems to be valid only if preferences are a direct and additive function of the census-based variables. Still, this approach laid the foundation for modern hierarchical approaches.

More recently, de Sola Pool, Abelson and Popkin's (1965) idea of synthetic voter types was revived by Gelman and Little (1997) and combined with a much more efficient estimation strategy (see also Park, Gelman and Bafumi, 2004; Gelman, 2008). In what is known as multilevel regression and poststratification (MRP), individual (binary) survey responses, i.e. preferences are modeled as a function of demographics, including the state of residence. Following the procedure of Park, Gelman and Bafumi (2004), this adds up to 3,264 cross-classifications. The multilevel setup allows partial pooling of voter types across states. In a second step (and in line with the well-known poststratification procedure), voter type preferences are weighted according to the proportion in each state, known from census data. Several validation efforts for this approach have shown that MRP outperforms prior methods of estimating subnational preferences (Lax and Phillips, 2009b; Warshaw and Rodden, 2012; Buttice and Highton, 2013; Hanretty, Lauderdale and Vivyan, 2014). Additionally, it corrects for coverage and nonresponse issues as long as census information for the districts not covered is available. At the same time, the latter—availability of census information—is a crucial condition for this approach to work, and one that might not be met in every scenario.

Therefore, variations of the hierarchical approach have been suggested. In very recent work, Leemann and Wasserfallen (2014) offer a method that relaxes the data requirement of MRP that census data provide information about the joint distributions of variables used for post-stratification (e.g., the fraction of 18-29 year old black men with a high school degree living in a certain subnational unit). Their method builds merely on known marginal distributions of variables which are used to estimate synthetic joint distributions, and idea somehow related to iterative proportional fitting ('raking'). This allows applying the hierarchical strategy in scenarios where only marginal distributions are known at the subnational level (therefore, they dub they apporach 'multilevel regression with *marginal* poststratification', MRmP). Further, while classic MRP is restricted to a very limited set of demographic variables, MRmP allows for more flexibility by adding strong individual-level predictors in the response model for which only the marginal distributions are known.

However, if neither joint nor marginal distributions of poststratification variables are known, these approaches are impractical. This is where our paper enters the debate. We suggest a method that gets along with minimal additional information from digitalized maps of the geographic unit of interest and geographic identifiers for these units in the survey data.

1.1.2 Contribution

In the first paper of this thesis (see also Selb and Munzert, 2011), we present a model which was originally developed for image restoration (Besag, York and Mollié, 1991) and illustrate its use in the context of the estimation of political preferences in local administrative units. Similar to Park, Gelman and Bafumi (2004), the approach we suggest mainly relies on survey data which are fed into a hierarchical model. However, it copes without additional census information on the marginals of certain strata in the constituency population. Instead, the model rests on the following logic: Based on survey data, preferences can be estimated at varying (geographic) levels with varying levels of precision. Preferences pooled over the entire survey sample contribute to the estimation of the overall mean, which can therefore be estimated quite precisely. Given dozens or even hundreds of constituencies, a 'direct' estimate of mean constituency preferences rests on very few respondents. In some cases, the survey may not even provide a single respondent for a district.

Therefore, we introduce a spatially structured random intercept that serves as an 'auxiliary layer.' This random intercept is assumed to follow an intrinsic conditional autoregressive (CAR) distribution, which was originally proposed for purposes of image restoration, archaeological work and epidemiological studies (Besag, 1975; Besag, York and Mollié, 1991). Technically, the mean of a CAR-structured random intercept for one district corresponds to the average value of intercepts for neighboring districts. The rationale is that this average can often be estimated with greater precision, as one district usually has several neighbors with likely more respondents covered. The degree of 'smoothing' induced by this structured (and another unstructured) district-level random effect is incorporated in the model and determined by the proportion of precision of these terms. Further, the model delivers estimates for constituencies that are not covered by the survey sample at all, as long as it can rely on neighboring districts with valid information. In sum, by providing minimal information—a discrete neighborhood structure of constituencies—estimates from survey data can be considerably improved.

1.1.3 Outlook

Our contribution opens avenues for research that builds on more sophisticated perceptions of mechanisms of political behavior and opinion forming, like the representation of subconstituency opinion in Congress (Fenno, 1978; Bishin, 2000; Clinton, 2006; Adams and Ezrow, 2009).

In the third paper of my dissertation project, I use this approach to generate poll-based sets of constituency-level election forecasts. While polling data play a major role in scholarly efforts to forecast election outcomes, they are hardly used to predict outcomes at the local level. In political systems where electoral outcomes in constituencies are crucial for the distribution of power at the national level (as is the case in parliamentary elections in the United Kingdom, for example), information about district races is valuable news for the national race, too. As I demonstrate, the method can be used to stabilize trend estimates in public opinion for election forecasting purposes. Apart from this application and the introductory example on measuring the impact of strategic voting, the method has also been used to inform various models of legislative behavior (Baumann, Debus and Müller, 2013; Bernauer and Munzert, 2014; Debus and Bäck, 2014; Hanretty, Lauderdale and Vivyan, 2014).

1.2 From elites to masses in the study of political polarization

One of the highly debated questions of our times is whether political elites (and the public) are drifting apart along various lines of political conflict, and what this implies for the process of policy-making, representation of peoples' opinion, and the society as a whole. The controversy about the degree of 'political polarization', as this phenomenon is frequently labeled, has gained traction especially among American academics over the last decade, when political elites have apparently lost their ability to agree on a compromise on various political issues, and radical movements like the Tea Party enjoy increased popularity (DiMaggio, Evans and Bryson, 1996; Layman and Carsey, 2002; Abramowitz and Saunders, 2005; Fiorina, Abrams and Pope, 2005; Layman, Carsey and Horowitz, 2006; Bafumi and Shapiro, 2009). While there is an on-going debate if the American public has polarized or just sorted along party lines, other societies have hardly been studied at all with respect to trends and levels of mass polarization. The second paper contributes to the study of polarization in contexts other than the United States. We apply an issue- and subgroup-based approach that was used by Baldassarri and Gelman (2008) to study polarization in the American public to a multiparty environment—Germany in the period between 1980 and 2010. We investigate polarization trends on a large set of economic, moral, immigration and gender issues and in various subpopulations and find that the German population has depolarized rather than polarized on many issues over the last 30 years.

1.2.1 Previous research on public opinion polarization

Among academics, the main attention has shifted from elite or party polarization towards polarization in the public, as highlighted in prominent reviews by Fiorina and Abrams (2008), Fischer and Mattson (2009), and Hetherington (2009). There is little doubt that elites in the American political system have become more polarized over the last 30 years; trends of Democrats and Republicans in Congress shifting apart reflect in virtually any common ideological measure (e.g., Poole and Rosenthal, 1984; Layman, Carsey and Horowitz, 2006; McCarty, Poole and Rosenthal, 2006). Evidence for consequences of this divergence is scarce, however. A narrative that has recently gained much attention is that partisans follow their parties and become more polarized as well. Hunter (1991) predicted that with the rise of moral issues that rarely change in peoples' views of life, e.g., abortion, gay marriage or the rights of ethnic minorities, the public would become inevitably more polarized. In other words, the mechanism underlying public opinion polarization which he suggests is that elites put contentious (moral) issues on the political agenda, which constrains people to rally around their party flag. While his conclusion of a more polarized society is not quite in line with recent findings on polarization dynamics, it has become the dominant view that a polarized society does not necessarily fragment along a single contested line like, for example, the classical liberal-conservative dimension, but is shaped by people who are aligned to a whole set of disputed issues. This 'multi-issue perspective' (as recently popularized by Baldassarri and Gelman, 2008; Levendusky, 2009, and others) is rather new wine in old wineskins than a novel insight. Conceptualizing polarization as attitude alignment over a variety of issues has, in essence, been identified almost 100 years ago. In '*Principles of Sociology*', Ross (1920) already notes:

> "A society, therefore, which is riven by a dozen oppositions along lines running in every direction, may actually be in less danger of being torn with violence or falling to pieces than one split along just one line. For each new cleavage contributes to narrow the cross clefts, so that one might say that society *is sewn together* by its inner conflicts." (Ross, 1920, 165)

By now, scholars can draw upon large and ever-expanding data to investigate claims like these and model dynamics of public opinion polarization over longer periods of time as well as within and between subgroups of the population. In a seminal study, DiMaggio, Evans and Bryson (1996) investigate opinion polarization in the U.S. using data both from the General Social Survey and the National Election Survey, spanning a period of 20 years. They find a trend towards consensus on racial, gender, and crime issues, stability on numerous others, and evidence of polarization only on attitudes toward abortion, the poor and, more recently, sexual morality. This study and others identify increasing trends of polarization only for the subgroups of partisans – either measured by party affiliation or self-identification between liberals and conservatives – and secular vs. religious voters (Abramowitz and Saunders 2005; Bafumi and Shapiro 2009; DiMaggio, Evans and Bryson 1996; Fiorina, Abrams and Pope 2005; Layman and Carsey 2002). In another study, which serves as a model for the empirical strategy of the second paper, Baldassarri and Gelman (2008) analyze polarization across a variety of attitudes and conclude that opinion changes in the U.S. are a result of people resorting to party labels rather than of greater polarization across attitudes. Moreover, they offer an innovative operationalization strategy, using pairwise correlations of issues as unit of analysis to track constraint over time. In the second paper, we draw upon the methodological and conceptual framework by Baldassarri and Gelman (2008) but shift the focus towards public opinion polarization in Germany.

Generally, little is known about trends in public opinion polarization in countries other than the United States. A study by Adams, Green and Milazzo (2012*a*) finds that the British public did not depolarize dramatically in terms of its policy preferences. Another study by Adams, Vries and Leiter (2011) provides similar findings for public opinion polarization dynamics for the Netherlands between 1986 and 1998. In the German case, the academic focus has been on party and party member polarization (e.g., Kirchheimer, 1965; Keil, 2004; Spier, 2011). Therefore, we target at a first descriptive, quantitative assessment of mass polarization in the German public.

1.2.2 Contribution

In the second paper of this thesis (see also Munzert and Bauer, 2013), we use a set of questions which is part of the German General Social Survey, a large survey project that was conducted 17 times between 1980 and 2010. The questions we use were administered several times and, analyzed together, they allow us to investigate trends in the relationships between many question pairs. Further, we classify all items in four issue subgroups—gender, moral, distribution and immigration issues. We can then identify several aspects of these trends, for example: Is the overall correlation between question pairs growing stronger (evidence for polarization) or getting weaker? What are the trends regarding specific policy fields? How do subgroups within the population differ in terms of polarization trends?

Our main finding is that polarization has generally decreased in the German public. What can be observed is that the correlation trend is negative for most of the item pairs over the last thirty years, but there is also interesting variation. Polarization has increased on the gender dimension, that is on items that concern the role of women in family and society, for example, whether mothers should care more about parenting than a career and whether a married woman should primarily support her husband's career or have one herself. This is a puzzling finding—on the one hand, recently debated topics like the introduction of gender quotas or child care issues have surely gained attention in the political arena and electoral campaigns as well, but were not the most salient issues in any campaign. On the other hand, gender equality might be one of the few remaining cleavages between the left and right camps in Germany, which have converged on many other highly debated topics such as unemployment assistance or the civil use of nuclear power. The rise of polarization on this domain is in contrast with the other domains we investigated.

By splitting the sample into subgroups, we are able to identify trends within (not between) parts of the population. It is generally argued that changes in public opinion (polarization) originate within intellectual elites and resonate first among the educated and the politically interested (Adams, Green and Milazzo, 2012*a*; Abramowitz and Saunders, 2005). Hence, these groups may be pioneers in what is regarded as trends in public opinion. Moreover, we deem it relevant to search for diverging trends among subpopulations of gender (Shapiro and Mahajan, 1986) and income (McCall and Manza, 2011). Further groups which merit attention are subpopulations of different religious denomination (Jelen, 1990). Finally, trends of public opinion may differ between East and West Germany making this another relevant subgroup division (Wegener and Liebig, 1995). Regarding the *overall trend*, there are no significant differences for subpopulations of gender, income and religious groups. However, the overall decreasing trend is much stronger among the highly educated, highly interested people, and somewhat stronger among respondents from West Germany.

1.2.3 Outlook

Our findings for the German case contrast evidence for the United States. On the other hand, recent studies on polarization of mass opinion in the UK (Adams, Green and Milazzo, 2012*a*,*b*) or the Netherlands (Adams, Vries and Leiter, 2011) have reported trends of depolarization rather than polarization as well. We hesitate to interpret existing evidence in this young research field as evidence for institutionally or elite induced movements in public opinion. Too little is known about the interplay between elite and mass polarization. However, our approach is useful to trace general trends of opinion polarization and to identify rise and fall of polarization on long-term issues. This could help to view current movements which seem to have caught fire on single issues like the conversion plans of the Taksim Gezi Park in Turkey, protest against massive surveillance of Internet traffic by intelligence agencies or recent protest against racist behavior by officials in Ferguson following the shooting of Michael Brown in a broader, temporal context.

Further, and in relation to the other contributions of this thesis, I see the key to understanding foundations of public opinion polarization in our focus on subpopulations—either defined by socio-economic, demographic, political or geographic strata—to see which groups are drifting apart and to which extent. These characteristics can have important implications from a democratic-theoretic view. For instance, if lines of polarization along a set of issues are aligned with partisanship of members of each group, one could argue that people are well represented by the supply side of the political system, that is, parties exist that represent their opinions of a vast array of issues.

1.3 The statistical turn in electoral forecasting

The last paper of my dissertation is located in the realm of electoral forecasting. It provides a method that helps improve forecasting at the constituency level. Election forecasting practice has been on a tear over the last few years, overcoming many of the drawbacks that were inherent to fundamentals-based regression approaches which used to dominate this subfield of electoral research (e.g., Abramowitz, 1988; Hibbs, 1982; Lewis-Beck and Rice, 1984*a*,*b*, 1992; Sigelman, 1979). While the limits of these methods where obvious very early,[10] more sophisticated approaches have emerged only recently,[11] but often with spectacular success.

Current forecasting models, specifically those located in the U.S. setting, have not much in common with their early predecessors. With regards to US presidential elections, Lock and Gelman (2010, 337) assert: 'The next level of sophistication (…) is to track the trends in state polls.' Indeed, methodological advances in combining information from the polls and previous elections have been used to generate astonishingly precise state-level forecasts of the US 2012 presidential election: Popular (and very successful) forecasts were given by Silver (2012), Linzer (2013), and Jackman (2012), just to name a few. These forecasts can rely on a massive amount of data from hundreds of trial-heat polls conducted at *state* level. Methodologically, they share several or all of the following features: (1) the combination of different data sources (overcoming the poll-based versus model-based dichotomy) and models, (2) dynamic fore-

[10] Although not necessarily to the authors themselves—for example, in reviewing Lewis-Beck and Rice's (1992) *Forecasting Elections*, Gelman (1993) already points to the fact that '[t]he book's weakness is its unquestioning faith in linear regression of aggregate electoral vote totals. We should always be suspicious of any grand claims made about a linear regression with five parameters and only 11 data points' (Gelman, 1993, 190). In a more recent critique, van der Eijk (2005) rightfully notices that with regards to the choice of their predictors, it is doubtful if standard regression-based forecasting models actually accumulate what the profession knows about voting.

[11] With some notable exceptions, e.g., Gelman and King (1993) and Rosenstone (1983).

casting with not one single prediction shot, but adaptations of forecasts with increasing levels of information (in line with horse race journalism based on the latest opinion polls), (3) probability-based forecasts (usually grounded on a Bayesian setting), and (4) sub-national levels of forecasts. My third paper aims at building upon recent progress in the field and developing a framework to forecast outcomes of elections at the constituency level; a task which is technically challenging but potentially rewarding in electoral systems where the distribution of power is an immediate consequence of district-level election outcomes.

Being able to rely on vast amounts of data to inform predictive models is a happenstance one is hardly befallen of in the case of constituency-level election forecasting. Genuine constituency-level polls are rare, so poll-averaging approaches are usually not an option for producing constituency-level estimates. Therefore the idea is to exploit information from past elections (and prediction efforts) to calibrate several model components and, in the spirit of model averaging techniques as propagated by Montgomery, Hollenbach and Ward (2012*a*,*b*), weight several single components according to their past performance to generate a combined forecast. A crucial component of the proposed model is an analysis of prediction residuals to correct single model components for unaccounted systematic bias.

1.3.1 State of the art in constituency-level election forecasting

The majority of established forecasting models that target the outcome of parliamentary elections can be broadly divided into five general types, depending on (a) the forecast level (national or constituency outcomes) and (b) the type of information used. Figure 1.2 provides a rough overview of these types, along with exemplary work.[12]

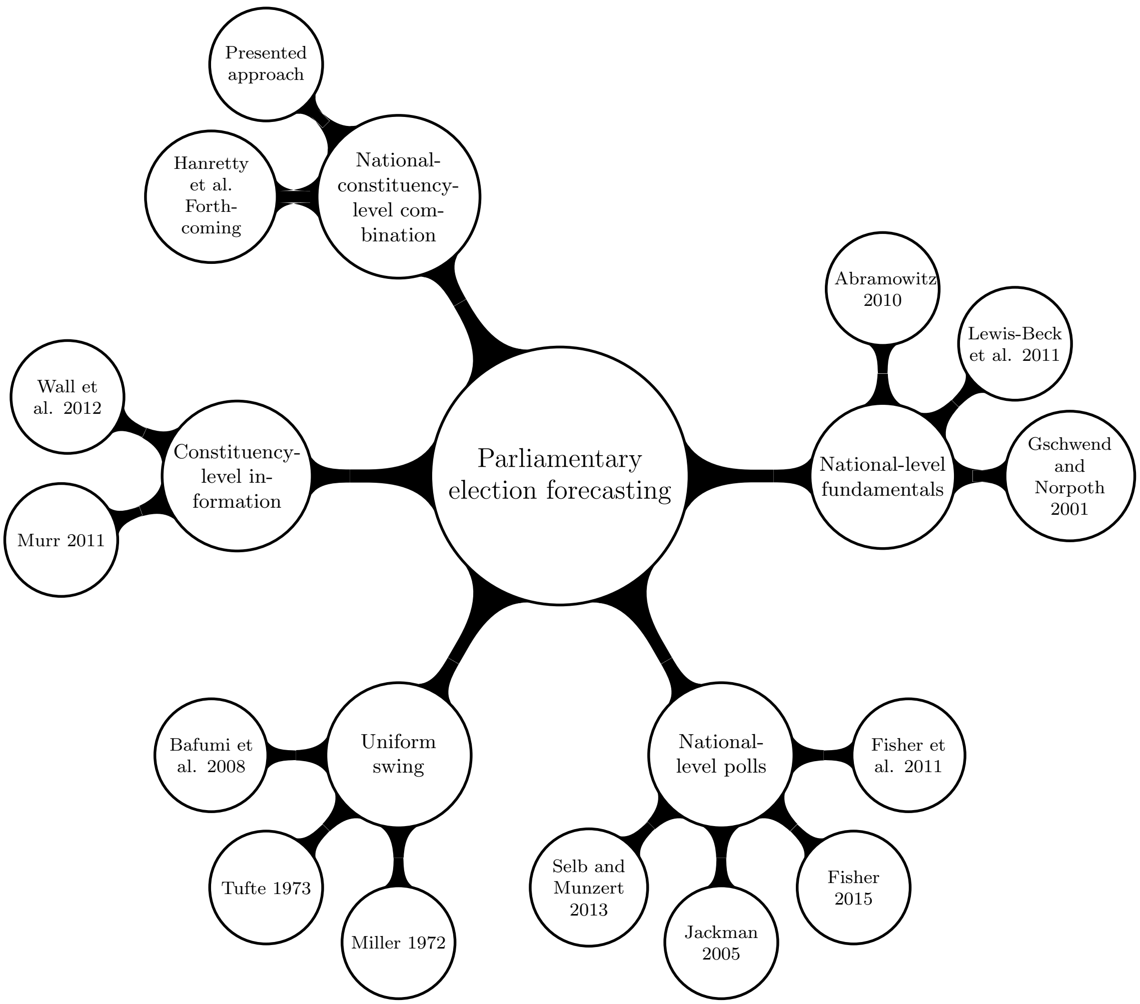
[12] In this review, I do not discuss poll-aggregating approaches which recently have been developed for U.S. Senate races (e.g., Linzer, 2013; Sides, Highton and McGhee, 2014; Silver, 2014), as they pose an exception in terms of data availability and electoral rules. They work on the state level and can draw on large amounts of polling data. Usually based on a firm Bayesian model framework and providing dynamic forecasts by design, such models are built to combine historical as well as polling data and give more weight to the latter as the election date approaches and more and more polls accumulate over the course of the campaign.

The first branch comprises national-level models that forecast aggregate vote or seat shares using structuralist information and regression analysis. They have been developed in the context of elections for the US House (for recent applications see, e.g., Abramowitz, 2010; Campbell, 2010; Lewis-Beck and Tien, 2010, 2012; Klarner, 2012), the British House of Commons (e.g., Lebo and Norpoth, 2011; Lewis-Beck, Nadeau and Bélanger, 2011; Whiteley et al., 2011), the German parliament (e.g., Gschwend and Norpoth, 2001, 2005; Jérôme, Jérôme-Speziari and Lewis-Beck, 2013; Kayser and Leininger, 2013) and other settings, and often merely build on a sparse set of economic and public opinion-based predictors which are shown to be strongly correlated with the party vote or seat shares of interest. While, given their sparseness, these models tend to produce fairly accurate forecasts, they come with a considerable amount of uncertainty and have been criticized for their weak data basis and tendency for overfitting (Gelman, 1993; van der Eijk, 2005). Moreover, lacking any local component, they do not inform about constituency-level campaign dynamics.

A second branch of models relies on information from national opinion polls. Instead of merely taking published vote intentions as a forecast, the idea is to exploit systematic trends in the historical relationship between polls and the vote which can be affected by institute- or party-specific biases or cyclical patterns that occur over the course of a campaign. Models of this flavor have been developed, among others, for Westminster elections (e.g., Fisher et al., 2011; Fisher, 2014), the Australian federal election (Jackman, 2005), and the German federal election (Selb and Munzert, 2013). Poll-based approaches exploit data which are an almost natural by-product of national-level election campaigns. In contrast to structural approaches, they often allow dynamic forecasts of public opinion.[13] On the other hand, such models have to operate on the level of the polls which is, more often than not, the national level. Furthermore, they are built to capitalize on the existence of systematic errors of trial-heat polls, which sometimes is an overly optimistic assumption (see Selb and Munzert, 2013). Depending on the context of elections, some of these models incorporate algorithms which take the seat-vote bias

[13] See, however, Gelman and King (1993) for a painstaking exploration of the volatility of pre-election polls and a discussion on why short-term variations in public opinion are rather endogenous to the campaign and do not provide crucial information for forecasting efforts.

Figure 1.2: Estimating constituency preferences - previous and current approaches



into account. This can be done by applying 'cube rule' (see Whiteley, 2005) or by predicting the bias with previous election data (see, e.g., Whiteley, 2005; Lebo and Norpoth, 2007; Nadeau, Lewis-Beck and Bélanger, 2009; Lebo and Norpoth, 2011; Whiteley et al., 2011). However, such approaches are hardly robust against 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. Essentially, these models 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 election results at the constituency level. As the assumption is 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; Tufte, 1973; Johnston and Hay, 1982; Butler and Beek, 1990; 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,[14] it rests on the critical assumption that swings are distributed equally over constituencies or are likely to 'cancel out' (Butler and Beek, 1990, 179). More sophisticated variants therefore introduce regional and tactical swing parameters or add information about incumbency status (Bafumi, Erikson and Wlezien, 2008, 2010). Ultimately, this approach generates constituency-level forecasts but does not incorporate any campaign information at this level, so the accuracy of this methodology essentially hinges on the forecast of the national trend.

More recent approaches turn to new data sources which 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 the crowds' effect. The idea is that aggregated group forecasts 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 at Westminster elections. For the U.S. house elections, Sides, Highton and McGhee (2014) incorporate fundraising data in the pre-primary model and substitute 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 which can easily be overseen in models that rely on overall trend measures. However, they still rely on exotic survey instruments or other data that are likely to be not available in many scenarios (as reported by Murr, 2011), or have been shown to add no

[14] 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.

predictive power compared to traditional approaches (see Wall, Sudulich and Cunningham, 2012).

Depending on the context, there are models which perform reasonably well in the aggregate but miss to generate reliable constituency-level forecasts. Models which 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.

1.3.2 Contribution

In the last paper of this thesis, I offer an approach that capitalizes on the power of constituency-level data and history on the one hand and model combination on the other. It integrates into existing models as it uses constituency-level forecasts regardless of the mechanism that produced them, e.g., a simple uniform swing approach or genuinely constituency-level based models, such as those presented by Wall, Sudulich and Cunningham (2012) and Murr (2011). However, it adds a correction and combination procedure which can help exploit past forecasts and several models more efficiently.

Depending upon the setting and the quality of available data, one could assume that the relative performance of different approaches varies by context. For instance, one could hypothesize that survey-based approaches are inferior when the survey data at hand are extraordinarily sparse or of bad quality. On the other hand, in settings where conditions have altered substantively between elections (e.g., because of massive redistricting, the appearance of a new party on the block or new candidates) one could argue that survey data should be expected to perform better, as they do not rely on previous (and potentially worthless) information. Identifying systematic relationships like these is a crucial component of my suggested procedure, as the different sources of information are weighted according to educated prior knowledge on model component performance. While the combination of pre-election polling data and historical data has recently been introduced for US presidential elections at the state level (see Linzer, 2013), comparative work for constituency-level forecasts is scarce. A notable exception is the (yet unpublished) work by Hanretty, Lauderdale and Vivyan (2015), who pursue an integrated approach of combining national and constituency polls, historical election results and census data to forecast the 2015 Westminster election.[15]

As an application I present a forecast for the German federal election held in September 2013—a multi-party setting with a considerable number of districts (299). The mixed member proportional system allows voters to cast a personal and a party vote. Candidates who run only on the district ticket or have a hopeless position on the list have relatively large incentives to focus on their own campaign. Lacking any credible polling data for their races, information from forecasting models can provide helpful advice. Furthermore, while the proportional component provides incentives for supporters of smaller parties to cast their second vote for their mostly preferred party, a first vote for a candidate running on a ticket of one of the smaller parties is likely to be wasted. Therefore, forecasts on this level can also inform voters who consider casting their vote strategically. To sum up, while a district-level forecasting approach in the German case might not add much of knowledge about the national outcome of an election, it can provide important information for local campaigns and voters.

1.3.3 Outlook

The benefits that can be derived from a constituency perspective on election forecasts are obvious. Such an approach would allow for more precise forecasts of the actual distribution of seats at the national level in electoral systems where the relationship between the share of votes and the share of seats is biased, e.g., by effects of plurality rule on district level. For example, at Westminster elections the relationship between the aggregated share of votes and the share of seats a party gains is blurred by the fact that seats are won in single member plurality districts. In the United Kingdom general election of 2010, the winning Conservatives gained 47% of the seats, but only 36% of the popular vote. However, the bias in the seats-votes curve is not constant over time, which makes it difficult to predict the distribution of seats from national-level

[15] See *http://www.electionforecast.co.uk/* for more details.

polls alone. In general, specific characteristics of electoral rules may prohibit forecasters to infer district-level outcomes from national-level trends which are identified in national polls. One way to overcome this problem is to generate estimates for constituency-level outcomes, which could be used to infer national level outcomes by means of aggregation. This is an advantage not only in purely majoritarian systems. The internal composition of parliaments in mixed electoral systems with a strong PR component such as the German electoral system (see Manow, 2011) could be more precisely predicted with constituency-level forecasts, too.

Chapter 2

Estimating Constituency Preferences from Sparse Survey Data Using Auxiliary Geographic Information

Peter Selb and Simon Munzert

Published in *Political Analysis* (2011) 19(4): 455-470.

Abstract

Measures of constituency preferences are of vital importance for the study of political representation and other research areas. Yet, such measures are often difficult to obtain. Previous survey-based estimates frequently lack precision and coverage due to small samples, rely on questionable assumptions, or require detailed auxiliary information about the constituencies' population characteristics. We propose an alternative Bayesian hierarchical approach that exploits minimal geographic information readily available from digitalized constituency maps. If at hand, social background data is easily integrated. To validate the method, we use national polls and district-level results from the 2009 German *Bundestag* election, an empirical case for which detailed structural information is missing.

'Everything is related to everything else, but near things are more related than distant things.'—*Waldo Tobler (1970)*

2.1 Introduction

Constituency preferences, commonly conceptualized in terms of the distribution of the voters' issue attitudes or value orientations within electorally relevant geographical units, are central for studying classic questions about democratic representation. For example, does the roll-call behavior of legislators reflect their constituents' policy views? Under what political and institutional conditions? What will be the electoral consequences if legislators fail to do so? Providing an empirical answer to such questions inevitably necessitates measures of constituency preferences. Yet, such measures are often difficult to obtain. Previous measures of constituency preferences based on aggregate proxies of public opinion such as constituency-level sociodemographics (e.g., Kalt and Zupan, 1984; Krehbiel, 1993; Levitt, 1996), electoral returns (e.g., Ansolabehere, Snyder and Stewart, 2001; Canes-Wrone, Brady and Cogan, 2002; Erikson, 1980), and referenda outcomes (e.g., Kuklinski, 1978; McCrone and Kuklinski, 1979; McDonagh, 1993) have been criticized both in validity and specificity terms (see, Jackson and King, 1989). Direct survey measures that average respondents' attitudes by district usually lack precision due to small samples and do not provide estimates for out-of-sample constituencies (e.g., Hurley and Hill, 2003; McCrone and Stone, 1986; Miller and Stokes, 1963), or pool surveys over long time periods and therefore have to assume temporally stable preferences (e.g., Brace et al., 2002; Erikson, Wright and McIver, 1993; Gibson, 1992; Wright, Erikson and McIver, 1985). Early model-based approaches that regress individual preferences on social background characteristics, and then use these parameter estimates to predict mean district preferences from the district-level distribution of covariates known from official sources (e.g., Erikson, 1978; de Sola Pool, Abelson and Popkin, 1965; Sullivan and Minns, 1976; Weber et al., 1972), are relatively insensitive to small samples and lack of district coverage. At the same time, however, such strategies involve cross-level inferences that rest on strong assumptions (see Seidman, 1975). Ardoin and Garand (2003) take a top-down approach in estimating mean preferences in congressional districts by first regressing (large-sample) state-level direct estimators of opinion on state-level characteristics. The first-stage regression weights are then used to predict mean district-level preferences from known population characteristics. Again, relationships between variables are not necessarily uniform across spatial scales, a phenomenon that is similar to the ecological inference problem and known as the 'modifiable areal unit problem' in the spatial statistics literature (see Openshaw and Taylor, 1979).

Current hierarchical estimators of constituency preferences have been demonstrated to outperform previous approaches regarding their accuracy and efficiency (e.g., Park, Gelman and Bafumi, 2004), but require detailed structural information that will often be missing, for example, if the boundaries of the geographic units of interest (e.g., electoral districts) do not coincide with the boundaries of the administrative units for which this information is collected or provided.

In this paper, we propose an alternative Bayesian hierarchical estimation strategy that exploits auxiliary geographic information about the constituencies' neighborhood structure and surface area, which is readily available from digitalized constituency maps alone. We essentially exploit the observation that political preferences tend to be spatially dependent—an observation closely related to Tobler's (1970) first law of geography (see the opening quote). If at hand, social background information is easily integrated in a manner compatible with other state-of-the-art strategies to estimating constituency preferences (in particular, Park, Gelman and Bafumi, 2004). In doing so, we draw on (and call attention to) recent developments in *small area estimation (SAE)* – a field that covers a variety of methods used to produce survey-based estimates for geographical areas in which the sample sizes are too small to provide reliable direct estimates (see Rao, 2005), but has been largely ignored by scholars of representation.[1]

[1] The only explicit references to SAE that we came across in the vast political science literature on measuring constituency preferences appear in Jackson (1989) and Park, Gelman and Bafumi (2004).

To validate the method, we estimate district-level party vote shares from two post-election surveys conducted after the 2009 German *Bundestag* election in order to compare them to official election results. Germany is a particularly relevant empirical case, since detailed census data as required by alternative approaches is not available – partly for structural reasons, i.e., non-nested administrative and electoral geographies, partly due to political sensibilities arguably dating back to the massive abuses of census data during the Nazi regime (see Luebke and Milton, 1994). We also demonstrate how our approach may be combined with Park, Gelman and Bafumi's (2004) methods provided that demographic information for poststratification is available.

2.2 The current gold standard

The major drawbacks of direct estimates of constituency preferences in normal circumstances, that is, their imprecision and potential lack of coverage with sparse survey data and large numbers of districts, arise from their exclusive reliance on survey information from within constituencies.[2] Alternative strategies use supplementary cross-district information in various ways (for a recent review, see Berry et al., 2010). Among those, an approach by Gelman and Little (1997) and Park, Gelman and Bafumi (2004) is currently emerging as a widely used gold standard for estimating constituency preferences from national surveys (e.g., Berkman and Plutzer, 2005; Christensen and Florence, 2008; Gelman, 2008; Hill, Herron and Lewis, 2010; Kastellec, Lax and Phillips, 2010; Lax and Phillips, 2009a,b).

Gelman and his colleagues pick up on de Sola Pool, Abelson and Popkin's (1965) and Weber et al.'s (1972) earlier idea to stratify survey samples into voter types, and to use census information for poststratification in order to obtain state-level constituency preferences. They develop a multilevel logistic model, with binary preferences y of individuals $i = 1, 2, \ldots, N$ being modeled as a function of $S = 64$ sociodemographic voter types s as defined by ethnicity

[2] For some notable exceptions using extraordinarily rich survey data that provide enough information to allow for accurate direct estimation, see Bafumi and Herron (2006), Clinton (2006) and Klingemann and Wessels (2001).

and sex (both binary), age and education (four categories k and l each) plus their respective interaction terms, nested in $J = 51$ states j (including Washington D.C.),[3]

$$\begin{aligned}
\Pr(y_i = 1) &= \operatorname{logit}^{-1}(\beta^0 + \beta^{\text{female}}\text{female}_i + \beta^{\text{black}}\text{black}_i \\
&\quad + \beta^{\text{female.black}}\,(\text{female} \times \text{black}) \\
&\quad + \alpha^{\text{age}}_{k[i]} + \alpha^{\text{edu}}_{l[i]} + \alpha^{\text{age.edu}}_{k[i],l[i]} + \alpha^{\text{state}}_{j[i]}),
\end{aligned} \tag{2.1}$$

making for a total of $64 \times 51 = 3{,}264$ synthetic subconstituencies. At first glance, stratification into a large number of subconstituencies thus seems to aggravate, not remedy, the problem of data scarcity. To compensate for small survey samples per (sub-)constituency, however, the hierarchical model 'borrows strength' by partially pooling voter types across states:

$$\begin{aligned}
\alpha^{\text{age}}_{k} &\sim \mathrm{N}\left(0, \sigma^2_{\text{age}}\right), \\
\alpha^{\text{edu}}_{l} &\sim \mathrm{N}\left(0, \sigma^2_{\text{edu}}\right), \\
\alpha^{\text{age.edu}}_{k,l} &\sim \mathrm{N}\left(0, \sigma^2_{\text{age.edu}}\right).
\end{aligned} \tag{2.2}$$

That is, all the respondents in a survey, no matter where they live, contribute information about differences in preferences between sociodemographic voter types to a degree warranted by the data. At the same time, the state of residency is used to estimate non-demographic state-level effects, which themselves are modeled using previous presidential vote shares as state-level predictors:

$$\alpha^{\text{state}}_{j} \sim \mathrm{N}\left(\alpha^{\text{region}}_{m[j]} + \beta^{\text{prev}}\text{prev}_j, \sigma^2_{\text{state}}\right). \tag{2.3}$$

Also, regional effects are included to account for unobserved similarities among $M = 5$ groups of states m (Northeast, Midwest, South, West, and the District of Columbia):

$$\alpha^{\text{region}}_{m} \sim \mathrm{N}\left(0, \sigma^2_{\text{region}}\right). \tag{2.4}$$

[3] We lay out the model in some detail here to set the stage for a replication analysis to follow in Section 2.6.

In a second step, Gelman and collaborators obtain measures of state-level preferences by summing over the predicted probabilities of holding the preference of interest π for each stratification cell in proportion to their population share within each state:

$$\pi_j = \frac{\sum_{s \in j}^{S} \pi_s N_s}{N_j}. \tag{2.5}$$

Park, Gelman and Bafumi (2004) use the model to estimate state-level vote shares for George Bush, Sr. (therefore the logistic specification) at the 1988 and 1992 Presidential elections based on CBS News-New York Times polls with regularly sized national samples of 2,193 and 4,650 respondents. Validation by comparison to the actual election returns yields a mean absolute error of about 4% in both cases – far better than predictions based on separate state analyses (a strategy that is akin to the direct estimator based on small samples), and on complete pooling of all respondents across states (an approach similar to earlier two-step strategies based on standard regression methods). Lax and Phillips (2009b) further compare this method to the direct estimator based on a large-scale sample from several merged polls, and conclude that the hierarchical estimator using small samples is as accurate as utilizing ten times as much survey data for direct estimation. Accordingly, the hierarchical estimator with poststratification appears to clearly outperform previous survey-based approaches in terms of its accuracy and efficiency. As opposed to the direct estimator, it offers *coverage* even for constituencies not included in the surveys used, provided constituency-level census information is available. Also, the hierarchical estimator incorporates the uncertainty inherent in direct estimation from survey data through *shrinkage* of directly observed differences between the constituencies contingent on the number of individual responses. Finally, by using census data for *poststratification*, this approach implicitly corrects for survey nonresponse, given that nonresponse is related to the observed sociodemographic characteristics.

On the other hand, though, the hierarchical estimator with poststratification requires detailed census information for the population of each synthetic voter type in each state (e.g., the number or population share of white females aged over 65 who are college graduates in Arizona) – information that might not be available for constituencies other than U.S. states, or at lower federal levels. Likewise, whenever the aim is to estimate constituency preferences at lower federal levels, the number of stratification cells will rapidly increase. For example, a model including the above 64 sociodemographic categories and the 435 congressional districts will yield 27,840 synthetic subconstituencies. Although multilevel models borrow strength from across all observations, and thus do not require that each and every stratification cell be populated with survey respondents, such complexities will inevitably necessitate larger numbers of survey responses.

2.3 An alternative approach using auxiliary geographic information

So what to do if the current gold standard in estimating constituency preferences is impractical due to data limitations? Our alternative approach starts from the observation that political predispositions, preferences and behaviors often come along in geographic clusters (see, i.a., Agnew, 2002; Gelman, 2008; Johnston and Pattie, 2006, and Tobler's opening quote). Explanations of these geographical clusterings emphasize regional differences in historical settlement and immigration patterns Elazar (1994), the geography of industrialization and urbanization Rodden (2010), and the resulting differences in current social composition Hero (1998), and local economic conditions Heppen (2003). Such broad historical processes are unlikely to abruptly halt at the geographical boundaries of constituencies, particularly if these boundaries are arbitrarily rather than historically drawn to satisfy numerical considerations. For example, single-member plurality (SMP) electoral districts are required to unite approximately equal numbers of voters to warrant evenly weighted votes, and are thus often subject to redistricting in order to reflect changing shares of electoral population. If the geographical distribution of political preferences cross-cuts constituency boundaries, this will lead to similar preference distributions among neighboring constituencies, that is, to *spatial dependence*. Potential spatial dependence is often conceived of as a nuisance for statistical estimation, as it may violate the assumption of independent errors that is typical of many regression models.

Far from being treated as a nuisance, spatial dependence can be exploited to help solve sparse data problems of the sort described: provided constituency preferences are spatially correlated, the preferences of neighboring constituencies are informative about preferences in any particular constituency. Moreover, since most constituencies have more than one neighbor, the mean preferences in the neighborhood of a given constituency can usually be estimated more precisely than in the target area itself (subject to the condition that the number of survey respondents per constituency is relatively constant). One of the main issues is then to formally represent that spatial dependence in order to utilize it in the estimation problem.

To this end, we set up a—preliminarily empty—hierarchical model of the (logit of the) probability that a respondent $i = 1, 2, \dots N$ in constituency $j = 1, 2, \dots J$ holds a certain preference y,[4]

$$\Pr(y_i = 1) \quad = \quad \text{logit}^{-1}\left(\alpha^0 + \phi_{j[i]} + \upsilon_{j[i]}\right), \tag{2.6}$$

where the constant term α^0 represents a global propensity to hold that preference, and ϕ_j and υ_j both represent constituency-specific deviations from that global propensity which are assumed to be independent of each other. In particular, ϕ_j depicts a spatially correlated random effect (RE) for which we adopt an intrinsic conditional autoregressive (CAR) distribution (see Besag, York and Mollié, 1991).[5] Under this specification, the conditional distribution of ϕ_j given the ϕ's in all the other constituencies $k \neq j$ depends only on the ϕ's in the constituencies adjacent to j,

$$\phi_j \mid \phi_k \quad \sim \quad \mathrm{N}\left(\frac{\sum_{k \neq j} w_{jk}\phi_k}{\sum_{k \neq j} w_{jk}}, \frac{\sigma^2_\phi}{\sum_{k \neq j} w_{jk}}\right), \tag{2.7}$$

[4]We use a logistic specification since, in the following empirical section, we will estimate party vote shares from post-election surveys in order to compare them to official election results (also see Park, Gelman and Bafumi, 2004). One could equally start from a linear model specification if indicated.

[5] Spatially correlated RE have relatively rarely been capitalized on in SAE until recently (see Pfefferman, 2002). As a simple alternative to this specification, one could assume a simultaneous autoregressive (SAR) process where the RE of any two constituencies are spatially correlated as a smooth function of their distance (see Pratesi and Salvati, 2008). However, we prefer the intrinsic CAR-specification, as SAR-processes involve assumptions (e.g., spatial stationarity) which are frequently violated in social science applications (see Cressie, 1993).

where the w_{jk} are elements of a $J \times J$ adjacency matrix assuming a value of 1 if units j and k are neighbors, i.e., have a common border or vertex, otherwise 0. Hence the expected conditional mean of ϕ in j corresponds to the average value of ϕ in the neighborhood of j, with its variance parameter, σ^2_ϕ, controlling how similar ϕ_j is to its neighbors. Deviation υ_j, on the other hand, is assumed to vary independent and identically across districts according to a normal distribution,

$$\upsilon_j \quad \sim \quad \mathrm{N}(0, \sigma^2_\upsilon). \tag{2.8}$$

Including both a spatially structured *and* an independent random component into the model will, in effect, pull the directly (but, due to small N_j, inaccurately) observed proportion of respondents holding the preference in constituency j toward both its neighborhood and the overall sample mean, with the amount of shrinkage increasing with decreasing N_j. That is, inferences for the district-level parameters, $\pi_j = \text{logit}^{-1}(\alpha^0 + \phi_j + \upsilon_j)$, reflect not just the direct survey information in district j, but also draw on relevant information in the neighboring districts (which will normally host more respondents than j), as well as in all the other districts (i.e., the whole survey sample). The relative amount of local versus global smoothing is then determined by the estimated variance of ϕ_j, σ^2_ϕ, in proportion to σ^2_υ. Further, by exploiting the conditional distribution of ϕ_j, the model equally informs estimates of constituency preferences for areas not covered by the survey, provided a constituency is not an island (i.e., it has neighbors to draw information from).[6]

If available, the model can easily be extended to include covariate information. In the following empirical application, we will use a single constituency-level covariate, namely the log inverse surface area, as a proxy of urbanity,[7] which has been repeatedly demonstrated to be a crucial contextual determinant of political preferences, and therefore has the potential to

[6] Note that, by exploiting second- and higher-order neighborhood relations, this is also true for constituencies that altogether lack in-sample neighbors, i.e., all its neighbors have missing values as well. However, inferences will be imprecise and unstable in these circumstances.

[7] If (and only if) the number of residents is relatively constant across electoral districts, as is usually the case in SMP systems, surface area is logically related to population density. In situations where some structural information is available one could, of course, directly use log population density as an indicator of urbanity. In fact, information about population density is available in our empirical case, and log inverse area and log population

disrupt otherwise smooth spatial preference distributions (see Rodden, 2010). To this end, we move the constant term α^0 from the individual-level model,

$$\Pr(y_i = 1) \quad = \quad \text{logit}^{-1}\left(\alpha_{j[i]} + \phi_{j[i]}\right), \tag{2.9}$$

to the district-level model, which now also includes the covariate:

$$\alpha_j \quad = \quad \alpha^0 + \beta^{\text{area}} \log\left(\text{area}_j^{-1}\right) + \upsilon_j. \tag{2.10}$$

To better be able to separate between eventual improvements in predictive power due to the inclusion of the covariate and due to the spatially structured random effect, we will pursue a sequential model strategy. First, we will fit an empty model containing only the unstructured RE, υ_j (Model 1), followed by a model including log inverse area as a district-level covariate plus the unstructured RE (Model 2). Finally, we will add the spatially structured RE, ϕ_j, and fit the full specification given in Equations 2.9 and 2.10 (Model 3). Validation results will also be reported for the direct estimator of the constituency proportions of respondents holding the preference of interest (Direct), that is, $\frac{\sum_{i\in j}^{N_j} y_i}{N_j}$.[8]

2.3.1 Estimation

Given the paucity of survey and auxiliary data one usually confronts, the random effects ϕ_j and υ_j are essential components of our model. However, the model is clearly overparameterized by including *two* REs per constituency, so the likelihood will only identify them jointly for each unit.[9] The nature of the prior distributions in Equations 2.7 and 2.8 allows posterior identifiability within a fully Bayesian setting, given the constraint that $\sum_{j=1}^{J} \phi_j = 0$ (see

density are correlated at 0.99. For the time being, we would like to see how far we get using minimal geographic information that is available from digitalized maps alone.

[8] The direct estimators may be weighted inversely proportional to the respondents' selection probability as determined by the sampling design of the survey. However, integrating such weights into a hierarchical model is not a trivial exercise (see Gelman, 2007).

[9] Additionally, due to its independent distribution, υ_j cannot be identified at all for constituencies which are not covered by the survey and thus has to be set to zero in these instances.

Carlin and Louis, 1996). Inference proceeds via Markov Chain Monte Carlo (MCMC) methods to compute a joint posterior density for all the model parameters. We use WinBUGS to run the required computations Lunn et al. (2000). Complex covariance structures as those in Equation 2.7 can be implemented using GeoBUGS, an add-on module to WinBUGS for manipulating spatial data Thomas et al. (2004).[10]

We use uninformative priors to let the data determine the parameters. In particular, we use flat priors for the fixed parameters, α^0 and the β's, and vague uniform priors within a range of (0, 2) for the standard deviations of the REs, σ_ϕ and σ_v.[11] To monitor convergence, we set up three chains with randomly chosen starting values for the parameters, each with 20,000 iterations, of which we discard the first 10,000 before summarizing the parameters' posterior probabilities.

2.3.2 Validation

To validate our models, we estimate district-level party vote shares from two post-election surveys conducted after the 2009 German *Bundestag* election within the framework of the German Longitudinal Election Study (GLES), and compare them to their true distributions which are known from official election results.[12] Germany's so-called 'mixed-member PR' electoral system divides the country into $J = 299$ primary electoral districts (*Wahlkreise*) which constitute our target areas. The districts are nested within a higher electoral tier, the 16 *Länder* (federal states). Each voter has two votes: a candidate vote (*Erststimme*) which governs the allocation of the district seats (*Direktmandate*) using SMP, and a second vote (*Zweitstimme*) for a *länder*-specific party list that translates into seats (*Listenmandate*) according to the Sainte-

[10] Replication data and code are available from the Political Analysis Dataverse at *http://hdl.handle.net/1902.1/16363*. Additional results are given in Appendix A.

[11] Although results should be relatively insensitive to prior choice with a large number of districts (see below), we have also used alternative inverse gamma priors for the variances, σ^2_ϕ and σ^2_v, yielding almost identical point estimates (though marginally lower coverage probabilities).

[12] Official election statistics, digitalized maps and some additional structural information can be downloaded from the website of the *Bundeswahlleiter* (federal elections officer) at *http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_09/*.

Laguë method, restricted by a national five per cent clause (for details, see Saalfeld, 2008). Since the *Erststimme* is particularly prone to strategic voting which may disrupt otherwise smooth spatial distributions of (partisan) preferences contingent on district-specific tactical incentives, we will focus our analysis on the *Zweitstimme*. All five parliamentary parties will be considered: the Christian Democrats (CDU/CSU), the Social Democrats (SPD), the Liberals (FDP), the Greens (B90/Die Grünen), and the Left (Die Linke). Altogether, these parties received 94% of the *Zweitstimmen* in 2009, which almost makes for the complete choice set with which the voters were confronted. Instead of adding a lumping category for marginal parties and setting up a more sensitive multinomial logistic regression that accommodates the fact that district-level party vote shares sum to unity, we model the party vote shares separately: first, in order to better be able to explore the party-specific conditions of how well or poorly the estimates perform; and second, to be consistent with the more general modeling strategy proposed in the previous section. While estimating (multi-)party vote shares from survey data allows us to validate our method using real data, we suspect that there are not too many substantively interesting applications that deal with such compositional data.[13]

In validation terms, one important criterion will be the distance between true values and their estimates expressed as the *mean absolute error (MAE)* of (the median of) the estimates' posterior probabilities. Clearly, smaller MAEs indicate better point estimates. As to these estimates' uncertainty, we calculate 90% Bayesian *credible intervals* from the highest posterior density regions that can be immediately interpreted in terms of the probability that the true value of the estimated parameter is inside a given interval.[14] A second validation criterion is the *coverage probability* of the credible intervals, that is, the proportion of the time that the intervals actually contain the true value of interest. The actual coverage probability should approximate the nominal level of 90% as closely as possible. A final validation criterion will

[13] See Katz and King (1999) for a discussion of the potential issues with the separate modeling of multinomial proportions.

[14] Assessing the uncertainty of the direct estimators via (frequentist) confidence intervals is far from straightforward with small samples (see Agresti and Coull, 1998). Moreover, frequentist and Bayesian intervals are hardly comparable, as the former treat the estimated parameters as fixed and the confidence intervals as random, so that the probability that the true parameter value is inside the given interval is either 0 or 1. In the subsequent empirical analysis, we will therefore only report MAEs for the direct estimators, but no uncertainty assessments.

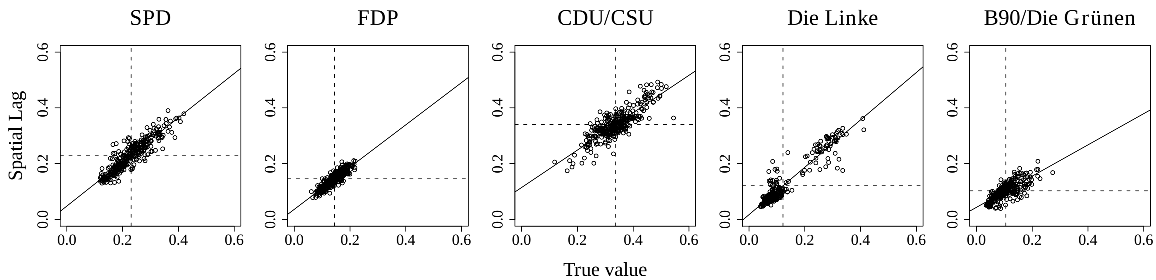
be the estimates' efficiency in terms of the *width* of the credible intervals. With the coverage probability being equal, a narrower interval is, of course, preferable.

In addition to these validation criteria, we will also use a diagnostic tool to assess our models' assumption that random effect v_j is, indeed, independently distributed across districts. To this end, we employ Moran's I, a widely used measure of spatial dependence. Moran's I indicates, in this case, the correlation between v_j with its *spatial lag*, $\frac{\sum_{k \neq j}^{J} w_{jk} v_k}{\sum_{k \neq j}^{J} w_{jk}}$, the latter of which gives the average value of v in the neighborhood of unit j (see the definition of w_{jk} following Equation 2.7).

2.3.3 An initial check of the model's central proposition

The Moran statistic can equally be utilized for an initial check of the basic premise of our modeling approach that political preferences often come along in geographical clusters. Figure 2.1 plots the official vote shares of the five parties (henceforth 'true values') against their spatial lags. The slopes of the solid lines is given by the Moran values. The dashed lines represent averages of the true values and their spatial lags. The average number of neighbors per district is 5.5, with a minimum value of 1 and a maximum of 10. Altogether, the magnitude of spatial autocorrelation is considerable, indicating that neighborhoods should be highly informative for estimating constituency preferences in any given district. In particular, the SPD vote shares exhibit the strongest spatial autocorrelation (I = 0.79), followed by the FDP (0.76), and the CDU/CSU (0.67). While, at first glance, there also seems to be tremendous autocorrelation in Die Linke vote shares (0.85), this turns out to be an artifact of its pronounced regional stronghold in eastern Germany (see the upper cloud in the graph). Separate calculations of Moran's I yield lower yet substantial values of 0.57 for eastern, and 0.49 for western districts. The same holds for B90/Die Grünen, whose district-level vote shares are spatially autocorrelated at 0.56. Such variability across parties will further provide us with the opportunity to see how our estimates' performance depends on the true spatial correlation of partisan preferences.

Figure 2.1: Moran's plots: (true) district-level party vote shares versus their spatial lags, i.e., the average party vote shares in the districts' neighborhoods. Official results from the German *Bundestag* elections 2009.



2.4 Survey data

Clearly, estimating constituency preferences for the 299 electoral districts from national polls puts some demands on the survey data even if model-based (let alone direct) estimation strategies are employed. We can benefit from the fortunate circumstance that the 2009 GLES conducted *two* surveys after the election on September 27:[15] a face-to-face (F2F) survey of a three-stage random sample of 2,117 eligible individuals, and the post-election wave of a rolling cross-section survey of 4,027 individuals conducted by telephone (CATI).[16] For the time being, we will pool these data, which makes for a total sample of 6,144 respondents, 5,067 of which are self-reported voters that indicated a party choice (recall that our aim is to estimate party vote shares, which are fractions of actual voters). Subsequently, we will also use the two surveys

[15] The survey data as well as technical reports are available from the GLES website (*www.gesis.org/gles*), study numbers ZA5301 and ZA5303.

[16] The respondents from the F2F survey were sampled using the standard design of the consortium of German market research institutes (*Arbeitsgemeinschaft deutscher Marktforschungsinstitute, ADM)*, with electoral wards (*Wahlbezirke*) as the PSUs being randomly sampled within regions (East and West), households being selected through random-route methods within electoral wards, and persons being randomly chosen within households. The sample of the CATI survey was drawn using the so-called Gabler-Häder design, where area codes within regions serve as the PSUs. None of these designs will normally warrant equal probability samples for the electoral districts. Absent the information necessary to construct weights that compensate for unequal selection probabilities within districts (and the quibbles with including such weights in hierarchical models; see Footnote 8), we will treat the district samples as if they were randomly drawn.

Table 2.1: Summary statistics of the survey data utilized: numbers of respondents, N, numbers of districts covered, J, average numbers of respondents per district, $\bar{N}_j$, their standard deviations, minimum and maximum values. F2F: face-to-face post-election survey. CATI: post-election wave of a computer-aided telephone rolling cross-sectional survey. Pooled: F2F and CATI surveys pooled.

	N	J	$\bar{N}_j$	$s.d.(N_j)$	$\min(N_j)$	$\max(N_j)$
F2F	1,527	201	7.60	5.21	1	29
CATI	3,540	291	12.16	5.60	2	45
Pooled	5,067	297	17.06	8.31	1	51

separately in order to see how our estimates behave with more common sample sizes. Table 2.1 gives an overview over the sample characteristics. The pooled sample covers 297 of the 299 districts, with an average of $\bar{N}_j = 17$ respondents per district, ranging from 1 to 51 – barely sufficient to directly estimate the vote shares of the five parties in any remotely reliable manner. Of course, this holds all the more true for the separate surveys. Otherwise, the global sample vote shares of the parties quite reasonably reflect their true national-level values, with B90/Die Grünen being the most notable exception (13.7% according to the pooled sample vs. an official return of 10.7%) – an issue we will return to in Section 2.6.

2.5 Results

Figure 2.2 plots the true party vote shares versus the direct estimates and the medians of the model-based estimates' posterior densities. Above all, it is easy to see that the direct estimates graphed in the leftmost panels are essentially useless in all cases. They are way too variable, and wide off the mark in most instances – in fact, so wide off the mark that some of them could not even be plotted within the displayed range of 0 to 60% of the votes. As the MAEs reported in Table 2.2 show, they mis-estimate the true vote shares by an unsatisfactory 7.4 to 10.1%. The estimates from the unstructured hierarchical model in the second column (Model 1) generally go to the other extreme and shrink the direct estimates toward the overall sample means (represented by the dashed lines in Figure 2.2) to an extent that there is barely any cross-district variability left. Adding log inverse area as a proxy of population density/urbanity to the unstructured RE model in the third column (Model 2) improves the estimates of the CDU/CSU, B90/Die Grünen, and very modestly, the SPD vote shares.[17] A more considerable improvement of the estimates results from adding the spatially structured RE, ϕ_j, in the final column (Model 3). The models' point estimates come out very close to their true values, with an average absolute error ranging from just 2.6 for the FDP to 4.5% for the CDU/CSU. The CDU/CSU estimates are too conservative in the tails, which also leads to coverage probabilities relatively far below their nominal level. Nevertheless, the coverage probabilities of the credible intervals of the spatially informed estimates are generally much closer to their nominal 90% level than those of the previous estimators, meanwhile the intervals are narrower and thus more efficient. Finally, as indicated by the Moran's values in Table 2.2, the inclusion of ϕ_j has helped to rid most of the spatial autocorrelation of v_j (which is assumed i.i.d.) in the previous models,[18] which supports the view that the CAR-structured RE provide a good representation of the prevalent spatial patterns. The only notable exception in this regard is the CDU/CSU model. Overall, our results suggest that the estimates from Model 3 perform remarkably well with the pooled data, particularly considering the models' exclusive reliance on auxiliary geographic information. The estimators' performance seemingly depends on the amount of spatial autocorrelation inherent in the underlying distributions of the true preferences. In particular, the model performed better in estimating SPD and FDP vote shares, and worse in estimating the vote shares of the other parties, most notably those of B90/Die Grünen, with an unacceptably low mean coverage probability of 0.76. A closer inspection of Figure 2.2 reveals, though, that the modest performance of the estimates of the Green vote shares is primarily due to the gen-

[17] CDU/CSU vote shares are negatively, and SPD and B90/Die Grünen vote shares positively related to urbanity (see the estimated coefficient plot in Figure 2.3). FDP and Die Linke vote shares are also related to urbanity, although in a non-monotonic way that is difficult to interpret, let alone predict, theoretically.

[18] Another violation obvious from Figure 2.2, the assumption that v_j is normally distributed, also seems difficult to maintain for the model of Left vote shares. Rather, there seem to be distinct modes for eastern and western districts. We therefore added a regional dummy to the vector of covariates which reduced the Left model's MAE to 0.02, although at the expense of over-confident estimates, as indicated by a decline in mean coverage probabilities to 0.83. Detailed results are available upon request.

eral upward bias in reported votes that has already been remarked in the data section above. We will further discuss this issue in the following section.

Finally, how well do the estimators perform with more regular sample sizes, that is, the F2F (N= 1,527) and the CATI (N= 3,540) surveys taken separately? Not surprisingly, the MAEs are generally larger than those for the pooled data, ranging at best (again with the spatially informed estimators) from 3.3 to 6.7% in the former, and from 3.3 to 4.9% in the latter case (see the two lower panels in Table 2.2). It is difficult to tell at first sight whether these differences are due to the CATI survey providing larger N (and thus, larger $\bar{N}_j$), or fuller cross-district coverage in terms of J (see Table 2.1), or due to the possibility that a telephone survey sample more evenly represents the district populations than the F2F sample based on single electoral wards within districts as the primary sampling units (see Footnote 16). On closer inspection (see Appendix A), the MAEs of the spatially informed estimates generally proved relatively insensitive to the number of respondents per district, N_j. Furthermore, even the estimates for the off-sample districts (i.e., $N_j = 0$) on average did not come out wider off target than those for the in-sample districts. Rather, the accuracy of the Model 3 estimates seem to moderately depend on the number of respondents in the neighborhood of a given district. Curiously, though, the estimators based on the CATI survey are over-confident, and thus exhibit lower coverage probabilities than those based on the F2F survey. In general, these results indicate that, for obtaining accurate model-based estimates for so many small areas, one still necessitates relatively large-scale national polls.

2.6 Combining both approaches

One of the drawbacks of our approach is that it does not naturally correct for survey non-response. This problem has, perhaps, become most evident in the general upward shift of our district estimates of Green party vote shares. In this final part of our empirical analysis, we will therefore demonstrate how our approach may be beneficially combined with Park, Gelman and Bafumi's (2004) method using poststratification to population controls in order to

Table 2.2: Validation results: mean absolute errors, mean widths of 90% Bayesian credible intervals, the intervals' coverage probabilities, and remaining spatial autocorrelation in the unstructured RE, v_j (Moran's I), for the estimates presented in Figure 2.2. F2F: face-to-face post-election survey. CATI: post-election wave of a computer-aided telephone rolling cross-sectional survey. Pooled: F2F and CATI surveys pooled.

	Mean absolute error				Mean width of 90%-CI			Coverage probabilities			Moran's *I*		
Pooled, $N = 5,067$	Direct	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
SPD	0.096	0.047	0.047	0.034	0.181	0.165	0.141	0.893	0.846	0.920	0.062*	0.063*	-0.037
FDP	0.079	0.029	0.030	0.026	0.157	0.158	0.145	0.946	0.943	0.963	0.011	0.014	-0.060
CDU/CSU	0.101	0.052	0.048	0.045	0.178	0.159	0.156	0.816	0.836	0.846	0.110*	0.106*	0.073*
Die Linke	0.074	0.044	0.044	0.032	0.168	0.168	0.126	0.886	0.890	0.906	0.273*	0.248*	0.032
B90/Die Grünen	0.075	0.037	0.033	0.033	0.146	0.132	0.123	0.783	0.786	0.796	-0.014	-0.007	-0.130
F2F, $N = 1,527$	Direct	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
SPD	0.101	0.052	0.051	0.043	0.160	0.158	0.212	0.702	0.722	0.953	0.085	0.095	0.016
FDP	0.100	0.032	0.033	0.037	0.140	0.146	0.204	0.796	0.816	0.970	0.058	0.051	-0.000
CDU/CSU	0.119	0.063	0.063	0.067	0.220	0.229	0.284	0.729	0.753	0.910	0.218*	0.212*	0.162*
Die Linke	0.093	0.064	0.064	0.054	0.175	0.174	0.211	0.605	0.619	0.876	0.177*	0.184*	0.115
B90/Die Grünen	0.076	0.032	0.029	0.033	0.101	0.107	0.146	0.742	0.793	0.930	0.324*	0.306*	0.297*
CATI, $N = 3,540$	Direct	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
SPD	0.107	0.051	0.051	0.043	0.136	0.123	0.135	0.726	0.635	0.793	0.043	0.035	-0.017
FDP	0.089	0.030	0.032	0.030	0.106	0.113	0.122	0.783	0.793	0.880	0.025	0.038	0.007
CDU/CSU	0.114	0.058	0.052	0.049	0.123	0.108	0.133	0.619	0.615	0.759	0.028	0.038	0.004
Die Linke	0.078	0.045	0.046	0.033	0.164	0.165	0.124	0.876	0.880	0.930	0.283*	0.281*	0.063
B90/Die Grünen	0.085	0.048	0.044	0.043	0.130	0.103	0.120	0.639	0.538	0.649	0.036	0.085*	-0.053

*$p < 0.10$ (for Moran values)

Figure 2.2: True district-level party vote shares versus estimates from alternative model specifications: direct estimates (Direct), estimates from an empty hierarchical model including an unstructured RE (Model 1), from a model including log inverse area as a covariate and an unstructured RE (Model 2), and from a model including the district-level covariate and both unstructured and spatially structured REs (Model 3).

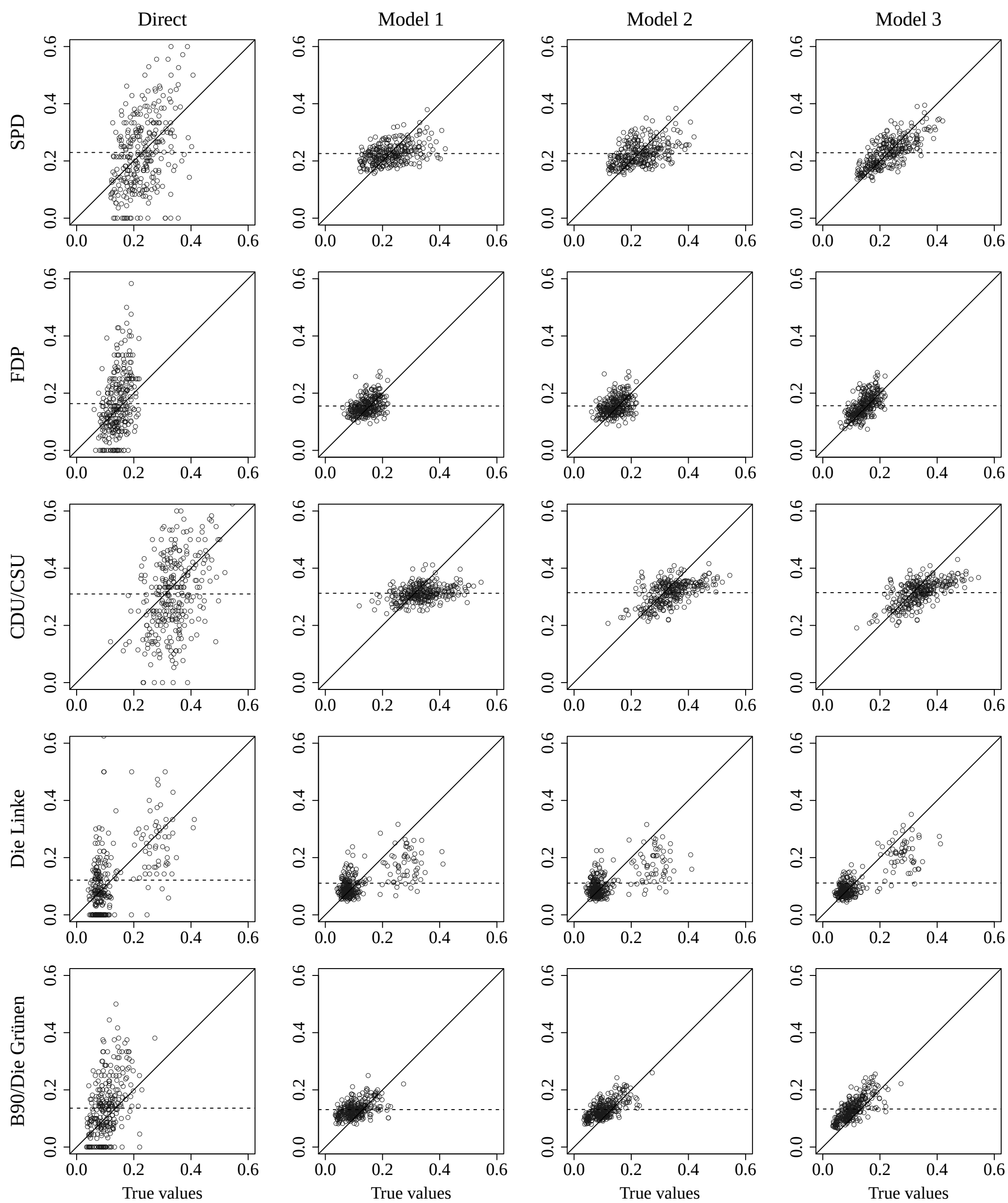
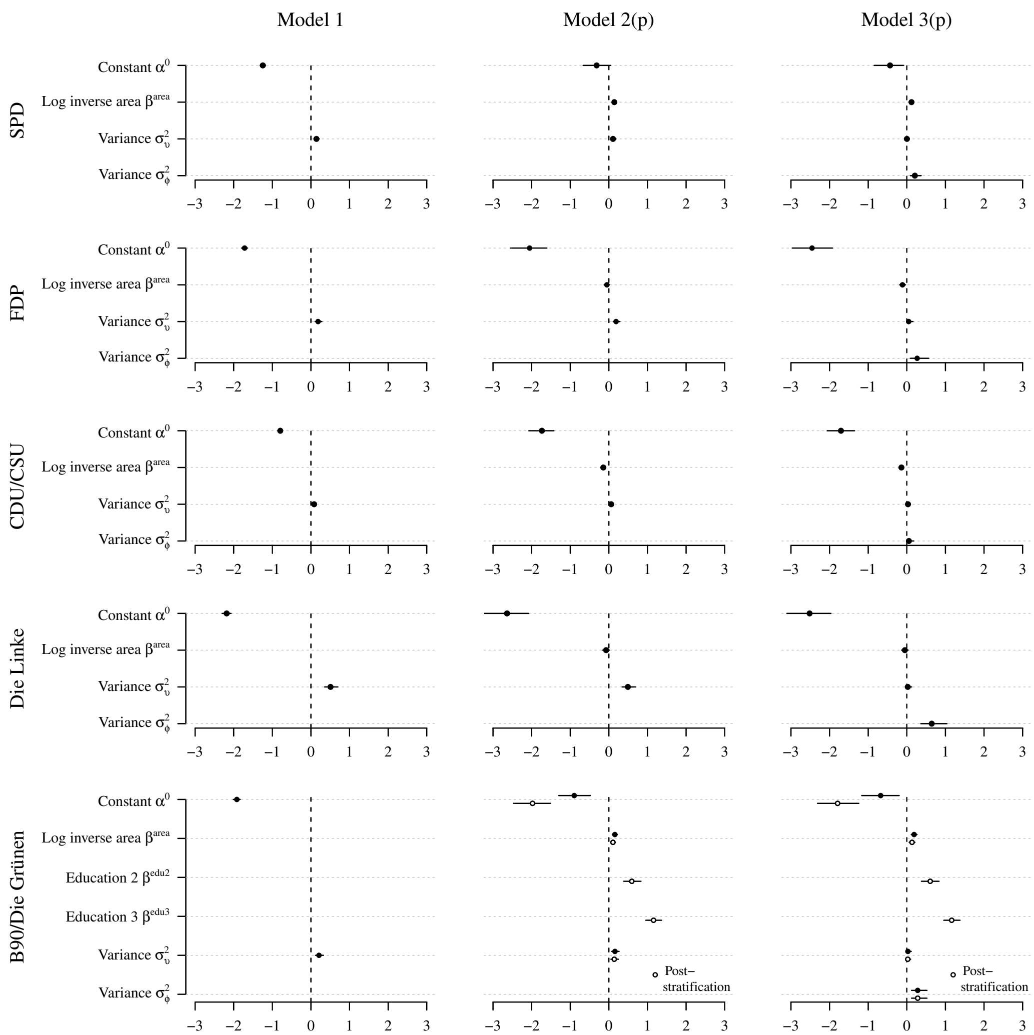


Figure 2.3: Medians and 90%-credible intervals of the coefficients' Bayesian posterior probabilities from different specifications of the hierarchical model of the respondents' partisan preferences: An empty model including an unstructured RE (Model 1), a model including log inverse area as a covariate and an unstructured RE (Model 2), and a model including the district-level covariate and both unstructured and geographically structured REs (Model 3). Green party vote shares have been additionally modeled including higher educational attainment levels for poststratification to district population controls (Models 2p and 3p).



correct for non-response bias in this case. Subsequently, we will turn the tables and replicate Park, Gelman and Bafumi's (2004) original analysis to see whether our neighborhood-based estimation strategy offers added value even in situations where structural information abounds.

As to the presumed non-response problem, it is well known that both the individual propensity to support postmaterialist parties and to participate in election surveys is positively related to educational attainment (e.g., Huefken, 2010). If so, we would expect the higher educated to be overrepresented in our sample, and the upward bias in our estimates of Green party vote shares to decrease after poststratification to the 'true' district-level distributions of attainment levels. While census data on education is not available, the *Bundeswahlleiter* at least provides district-level school graduation statistics as of 2007 which we, in the absence of better alternatives, use as proxies for the whole district populations (see Footnote 15). Indeed, individuals with tertiary education seem to be massively overrepresented in our sample (42% compared to 27% according to the official statistic). In contrast, those with secondary (33% compared with 41%) and primary education (24% compared with 32%) are underrepresented.[19] Therefore, we re-run our models of the Green party vote shares adding dummy variables for the two higher levels of educational attainment to the individual-level model in Equation 2.9 (see the bottom panel of figure 2.3 for the estimates),[20]

$$\Pr(y_i = 1) \quad = \quad \text{logit}^{-1}(\alpha_{j[i]} + \beta^{\text{edu2}}\text{edu2}_i + \beta^{\text{edu3}}\text{edu3}_i + \phi_{j[i]}), \tag{2.11}$$

and obtain predicted probabilities of having cast a Green vote for each of the $299 \times 3 = 897$ cross-classifications of electoral districts and attainment levels. The district estimate is then simply the population-weighted average of the attainment-specific predicted probabilities by district (see Equation 2.5). The results for the estimates from a model including only the unstructured RE in addition to the covariates (Model 2p), and both unstructured and CAR-structured REs (Model 3p) are reported in Table 2.3. Obviously, the upward bias in the estimates observed in the previous section is considerably reduced by using poststratification. Both MAEs are markedly lower (just 2.8 and 2.6%, respectively), and the coverage probabilities are more or less exactly at their nominal level of 90% for both estimators. Also, the inclusion of educational

[19] Considering that younger cohorts as those having graduated in 2007 are, on average, higher educated than older cohorts, we may even suspect the true educational bias in our sample to be stronger than indicated.

[20] As opposed to Park, Gelman and Bafumi (2004) in Equations 2.3 above, we model education with fixed, not random effects.

Table 2.3: Validation results: mean absolute errors, mean widths of 90% Bayesian credible intervals, the intervals' coverage probabilities, and remaining spatial autocorrelation in unstructured RE, v_j (Moran's I), for the hierarchical estimators of Green party vote shares with poststratification to district-level educational attainment. Specifications with an unstructured RE (Model 2p) and both unstructured and spatially structured REs (Model 3p).

Mean absolute error		Mean width of 90%-CI		Coverage probabilities		Moran's I	
Model 2p	Model 3p	Model 2p	Model 3p	Model 2p	Model 3p	Model 2p	Model 3p
0.028	0.026	0.143	0.139	0.886	0.916	0.001	-0.134

$^*p < 0.10$ (for Moran values)

attainment seems to have soaked up most of the spatial autocorrelation previously present in v_j, as indicated by the insignificant value of the Moran statistic. But still, the spatially informed estimator is slightly more precise in terms of MAE and more efficient in terms of a narrower credible interval.

Having demonstrated how socio-structural information can be fruitfully integrated into our model using poststratification, we now turn to Park, Gelman and Bafumi's (2004) original analysis of a 1988 election poll to see what our neighborhood-based estimation strategy has to offer in reverse.[21] As depicted in Equation 2.4, Park et al.'s model includes a regional RE and thus already provides some spatial structure. According to this specification, all the states within a given region are, in a way, considered neighbors that potentially share unobserved similarities. This seems perfectly reasonable given that the regional RE was explicitly built on prior knowledge of distinctive regional voting patterns (see Park, Gelman and Bafumi, 2004, 377). Conditional on the regional RE, however, the state-level RE is assumed independent within and across regions. To see this more clearly, we write out the state-level model in Equation 2.3 to explicitly include an error-term, v_j:

$$\begin{aligned} \alpha_j^{\text{state}} &= \alpha_{m[j]}^{\text{region}} + \beta^{\text{prev}}\text{prev}_j + v_j, \\ v_j &\sim \text{N}\left(0, \sigma_v^2\right). \end{aligned} \tag{2.12}$$

[21]Replication data and code for this analysis is available from Andrew Gelman's website (*http://www.stat.columbia.edu/£\sim£gelman/arm/examples/election88/*). We have used a digitalized states map from Kenneth Foote's website (*http://www.colorado.edu/geography/foote/maps/assign/hotspots/hotspots.html*).

Specifying an adjacency matrix for the U.S. states along the lines of our previous application allows us to directly check the assumption of spatially independent errors.[22] Results of our replication of Park et al.'s original model including the Moran test for remaining spatial autocorrelation in υ are given in Table 2.4.[23] Indeed, the Moran's statistic indicates moderate but significant spatial autocorrelation in υ (I = .11). Thus, Park et al.'s original model is, though to a modest extent, spatially misspecified in that the regional RE cannot fully account for unobserved similarities among contiguous states. Put differently, there seems some spatial information left to be utilized in the estimation problem. To do so, we replace the regional RE $\alpha^{\text{region}}_{m[j]}$ in Equation 2.12 with a constant term α^0, and add a CAR-structured RE ϕ_j instead:

$$\alpha^{\text{state}}_j = \alpha^0 + \beta^{\text{prev}}\text{prev}_j + \upsilon_j + \phi_j. \tag{2.13}$$

At the same time, we re-specify the adjacency matrix that determines ϕ_j in Equation 2.7 in order not to lose all the valuable prior information about regional differences in voting patterns and political preferences. In particular, we restrict our neighborhood definition to inner-regional contiguities so that states are considered as neighbors only to the extent that they share a common border or vertex *and* belong to the same region.[24] As indicated by a Moran value near zero in Table 2.4, the inclusion of ϕ_j has soaked up the remaining spatial autocorrelation in υ_j. This does not, however, lead to an appreciable improvement of the estimates' precision and efficiency.[25] Considering auxiliary geographic information, it seems, offers little added value in this particular case where ample information on the constituency populations, previous values of the preferences of interest, and affiliation to politico-cultural regions is available.

[22] Accordingly, each states has 4.3 neighbors on average, ranging from 0 for Alaska and Hawaii to 8 for Montana and Tennessee.

[23] Note that (Park, Gelman and Bafumi, 2004, 382) report an MAE of 4, not 4.5%. This discrepancy potentially relates to the fact that we did not achieve full convergence with the 5,000 iterations suggested by the authors (p. 379). Instead, we ran 10,000 iterations. Detailed results on the parameter estimates are given in Appendix A.

[24] This reduces the average number of neighbors to 3.3. Besides Alaska and Hawaii, Washington D.C. has no neighbors according to this definition, since Park et al. consider D.C. as a separate region. Therefore, ϕ_j has been set to 0 for these states.

[25] We have also fitted a model including both the regional and the CAR-structured RE, but we did not achieve convergence even after 30'000 iterations.

Table 2.4: Validation results: mean absolute errors, mean widths of 90% Bayesian credible intervals, the intervals' coverage probabilities, and remaining spatial autocorrelation in unstructured RE, v_j (Moran's I), for the Park et al.'s (2004) original model, and an alternative model with CAR-structured RE added.

Mean absolute error		Mean width of 90%-CI		Coverage probabilities		Moran's I	
Original	Alternative	Original	Alternative	Original	Alternative	Original	Alternative
0.045	0.044	0.170	0.174	0.922	0.882	0.107*	0.001

*$p < 0.10$ (for Moran values)

To generalize, though, more systematic assessments of the relative performance of alternative small area estimators under varying conditions are required (see Gómez-Rubio et al., 2010).

2.7 Discussion

In this paper, we proposed a Bayesian hierarchical approach to estimating constituency preferences from sparse survey samples that uses auxiliary geographic information in order to compensate for small numbers of respondents per constituency, incomplete constituency coverage, and lack of information about the constituencies' population characteristics. Our model exploited the fact that political preferences often come along in geographical clusters, and incorporates the spatial dependence emanating from these clusterings into the model's covariance structure. As a result, the model shrank directly observed constituency means in survey responses toward the mean responses in neighboring constituencies, and toward the whole sample mean. That way, the model borrowed strength both locally and globally to counterbalance small constituency-level samples. Additionally, we employed log inverse constituency area as a proxy of urbanization to account for eventual disruptions of otherwise smooth spatial preference distributions.

We used this model to obtain district estimates of party vote shares at the 2009 German *Bundestag* election from national post-election surveys that could be validated using official election results. On average, we had less than 20 survey responses per district available – very little information indeed to estimate the vote shares of five parties. Nevertheless, the mean absolute error of our predictions was only 3.4%, and even with smaller samples per district our estimators' properties were still reasonably satisfactory. Also, our model clearly outperformed direct estimates as well as spatially ignorant model-based alternatives.

Building upon these encouraging results, we locate the most fruitful areas of application in similar problems where the number of target areas is high relative to survey sample sizes, and where socio-structural data that could help inform alternative approaches such as Park, Gelman and Bafumi's (2004) method is missing. Single-member electoral districts (SMDs) constitute an obvious case (see Rodden, 2010), as their number is typically large (e.g., 299 in Germany, 435 in the U.S., 650 in the UK), and detailed census information is rarely available at this level. The same surely holds at even more disaggregated spatial scales, such as counties (e.g., Percival, Johnson and Neiman, 2009) or school districts (e.g., Berkman and Plutzer, 2005). In fact, spatial dependence often tends to be stronger, and thus more informative for our approach, at lower levels of spatial aggregation (see Openshaw and Taylor, 1979).

While our initial model exclusively relied on minimal geographic information readily available from digitalized constituency maps, we further demonstrated how our approach may be integrated with other state-of-the-art strategies that exploit additional information about the constituencies' population characteristics – specifically, Park, Gelman and Bafumi's (2004) hierarchical estimator with poststratification – to yield even more accurate estimates and to possibly correct for survey nonresponse.

In essence, the strategy we proposed boils down to the insight that spatial correlation is much more than a nuisance for statistical estimation. Rather, spatial dependence can be exploited to help solve problems typically arising when local survey samples are too small to provide reliable estimates of the quantities of interest. The main issue is then how to represent that spatial dependence in order to use it in the estimation problem. Our approach offers one solution to this issue.

Chapter 3

Political Depolarization in German Public Opinion, 1980-2010

Simon Munzert and Paul C. Bauer

Published in *Political Science Research & Methods* (2013) 1(1): 67-89.

Abstract

Little is known about political polarization in German public opinion. We offer an issue-based perspective and explore trends of opinion polarization in Germany. Public opinion polarization is conceptualized and measured as alignment of attitudes. Data from the German General Social Survey (1980 to 2010) comprise attitudes towards manifold issues which are classified into several dimensions. We estimate multilevel models which reveal general and issue- as well as dimension-specific levels and trends in attitude alignment for both the whole German population and subgroups. We find that public opinion polarization has decreased over the last three decades in Germany. In particular, highly educated and more politically interested people became less polarized over time. However, polarization seems to have increased in attitudes regarding gender issues. Our findings provide interesting contrasts to existing research on the American public.

> 'A society (...), which is riven by a dozen oppositions along lines running in every direction, may actually be in less danger of being torn with violence or falling to pieces than one split along just one line. For each new cleavage contributes to narrow the cross clefts, so that one might say that society *is sewn together* by its inner conflicts.'—*Edward Ross (1920)*

3.1 Introduction

Public opinion polarization (POP) affects a society's ability to reach consensus on a set of issues in the political arena. As POP grows, the likelihood rises that social groups with irreconcilable policy preferences will be formed (DiMaggio, Evans and Bryson, 1996, 693). As a consequence, public arenas are divided into adverse groups that are increasingly unable to cooperate because of their contradictory positions. This, in turn, may lead to an increase in conflicts within a society. Moreover, POP translates into the political sphere as citizens engage themselves in politics and politicians respond to the preferences of their respective constituencies, that is, electoral bases (Abramowitz, 2006, 73). This influence can impair a political system's capacity for reform and effective policy making (Galston and Nivola, 2006; Brady, Ferejohn and Harbridge, 2008).

Following discussions on elite polarization (e.g., Layman and Carsey, 2002; Layman, Carsey and Horowitz, 2006; Poole and Rosenthal, 1984; Stone, Rapoport and Abramowitz, 1990), several prominent review articles by Fiorina and Abrams (2008), Hetherington (2009), and Fischer and Mattson (2009) highlight the increasing scientific interest in POP. However, research on POP to date has been largely confined to the US context and centers around the question of whether American public opinion is more polarized today than in the past.[1] In a key study,

[1]Corresponding research mostly carries the label "political polarization." However, this term has suffered from ambiguity which produced different conclusions in what is regarded as polarization of the American public

DiMaggio, Evans and Bryson (1996) investigate opinion polarization in the United States, relying on different measures and using data both from the General Social Survey and the National Election Survey, spanning a period of 20 years. They find a trend towards consensus on racial, gender, and crime issues, stability on numerous others, and evidence of polarization only on attitudes toward abortion, the poor and (more recently) sexual morality. The conclusions drawn by DiMaggio, Evans and Bryson (1996) were confirmed by Evans (2003), who used additional data. These and other studies identify increasing trends of polarization only for subgroups of partisans—measured either by party affiliation or self-identification between liberals and conservatives—and secular versus religious voters (Abramowitz and Saunders, 2005; Bafumi and Shapiro, 2009; DiMaggio, Evans and Bryson, 1996; Fiorina, Abrams and Pope, 2005; Layman and Carsey, 2002). In another widely noticed study, Baldassarri and Gelman (2008) analyze polarization across a variety of attitudes and conclude that opinion changes in the United States reflect more a resorting of party labels among voters than greater polarization across issues.

Only recently have a few studies started to investigate POP in Europe. Adams, Green and Milazzo (2012*a*,*b*) find a depolarizing trend within the British public between 1987 and 2001 that mimics an analogous trend in the political system, that is, between Labour and the Tories. Adams, Vries and Leiter (2011) diagnose similar developments for the Netherlands between 1986 and 1998. Finally, Down and Wilson (2010) focus on a single opinion item (namely respondents' support for their country's EU membership) and investigate the causal link between opinion polarization and interparty competition on Europe.

Given the rising scholarly interest in the United States, and the relevance of the phenomenon for contemporary democratic states it is astonishing that, with the exceptions just mentioned, it has been largely neglected within the realms of European political science. In our study we pursue the following research question: *How has public opinion polarization in Germany de-*

(Hetherington, 2009, 447). Fiorina and Abrams (2008) convincingly argue that the many ways in which political polarization has been conceptualized do not properly reflect the concept of political polarization. According to Fiorina and Abrams, comparing political attitudes of respondents is straightforward, in contrast to several other ways of measurement (for example, vote choices as indicators of polarization). In our view it makes sense to clarify the respective focus on attitudes by using the term "public opinion polarization."

veloped over the last decades? To our knowledge, this is the first study which investigates POP in Germany. Following Baldassarri and Gelman (2008, 409) we conceive POP as *alignment along multiple attitude scales*.[2] As we will argue in detail, this conception and corresponding measurement has several advantages over other measures of POP. In general, we pursue a descriptive research agenda[3] and our study serves as a point of comparison for both existing[4] and future research. We find that public opinion polarization has decreased over the last three decades in Germany. In particular, the highly educated and more politically interested people have become less polarized over time. However, polarization seems to have increased regarding attitudes towards gender issues.

In the remainder of this paper, we first outline our conception of POP and demonstrate its theoretical advantages and the practical benefits of our measurement approach. Next, we present the data and offer an initial overview of the attitude scales taken into consideration. We then describe our modeling strategy and delineate the empirical results, first with regards to attitude alignment within the whole population, then focusing on several subgroups. Several robustness checks complete the analysis. The paper concludes with a discussion of the results and their implications for future research.

3.2 Conception and measurement of POP

A population may be polarized with regard to a single issue—for example, the question of legalizing abortion—or it may comprise people who are polarized on many issues. In the latter case, a population may be divided along multiple lines, for example pro-abortionists versus anti-abortionists, weapon haters versus weapon fanatics and proponents of state intervention versus advocates of a free market. However, we argue that POP only generates the previously mentioned negative effects if these groups overlap—for example, if anti-abortionists,

[2] We try to use a consistent terminology throughout this study. We use Converse's (1964) measure of constraint to study attitude alignment (issue alignment in the terms of Baldassarri and Gelman 2008). However, we refer only marginally to the debate on constraint in mass political belief systems.

[3] See Gerring (2012).

[4] For example, Baldassarri and Gelman (2008).

weapon haters, and state interventionists are in one camp that stands against the camp of anti-abortionists, pro-weapon fanatics and free market proponents. In other words: "[I]f people align along multiple, potentially divisive issues, even if they do not take extreme positions on single issues, the end result is a polarized society" (Baldassarri and Gelman, 2008, 409). This conceptualization of POP picks up an insight that had already been articulated by Ross (1920, see the opening quote), and later been dubbed 'criss cross' (Galtung, 1966). Following this logic, POP is conceived here as alignment along multiple scales that measure attitudes towards different issues. A conventional measure of alignment is the correlation coefficient which is higher when individuals have similar attitudes, that is, choose similar positions on two different attitude scales. Subsequently, one may take averages across pairwise correlations for groups of attitude scales that belong to certain issue dimensions or for all attitude scales.

The correlation measure features two useful characteristics: First, as the correlation between attitude scales increases, the variance of the average opinion score distribution grows as well (Baldassarri and Gelman, 2008, 419). An individual's average opinion score is simply the average of his or her positions on the combined scales. The average score distribution is then the distribution of all individual average opinion scores. In other words, the higher the correlation between these scales, the higher the share of individuals with extreme opinions on average across all scales. Hence, the correlation measure also accounts for extremities of attitudes. Secondly, populations can be divided into opposing clusters which captures the idea of camps opposing each other across a number of issues. Baldassarri and Gelman (2008, 419) show analytically for a four-dimensional case that the distance between individuals who belong to the same cluster decreases, while the distance between alternative clusters increases as the correlation between attitude scales increases. In other words, for high levels of polarization the average distance between individuals within a certain cluster is lower. As the correlation decreases, it is also harder to predict values on different attitude scales based on an individual's position on one specific attitude scale.[5]

[5] Note that one can use other measures of polarization, for example, kurtosis or variance, to identify other, specific aspects of polarization (see, however, Downey and Huffman (2001) for a detailed discussion of problems with such single scale measures). To take the latter as an example, variance as a measure of attitude dispersion is well-suited to identify opinion shifts on single items, with parts of the public taking more or less extreme

In the following description, which draws on DiMaggio, Evans and Bryson (1996, 696-697), it becomes apparent that the multi-issue conception of POP is better able to capture general trends of polarization than measures that focus on single issues: Imagine a society in which individuals are highly polarized—that is, they have extreme positions on different issues. At the same time, these individuals show no alignment with regards to their different attitude scale positions: two individuals may be close to each other on one scale, but take up completely opposed positions on the next scale. Instead of large stable camps that occupy the same position on several attitude scales such a society will be characterized by groups that huddle around positions on single attitude scales only to subsequently dissolve due to the heterogeneity of their positions on different scales. As a consequence, the probability for political conflict is lower than in the case of strongly correlated (aligned) attitude scale positions (Fiorina and Abrams, 2008, 577).

3.3 Methodology

3.3.1 Data, item selection and rating

We rely on time-series data generated from the *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften* (ALLBUS), the German equivalent of the General Social Survey. The ALLBUS was conducted 17 times (usually every two years) between 1980 and 2010.[6] In the first step, we identified 24 questions that query attitudes of political and societal relevance and have been asked at least three times in the period under consideration. Table 3.1 gives an overview of the respective items used in the analysis, the variable labels we chose and the length of the scale on which they are measured. Deciding whether an item is about political attitudes or not is naturally somewhat arbitrary. Some items directly target potential policy outcomes (for example,

positions. However, in order to capture the general trends of a society as a whole, a single-issue focus does not suffice and the multi-issue perspective is more revealing—as we will show in our analysis.

[6]Data are available from the GESIS archive free of charge after registration (*https://social-survey.gesis.org*, accessed August 2013).

immigration.asylum.seekers), while others encompass more general values that can, however, be perceived as fundamental for more concrete policies (for example, *gender.job.child*). The selection of items likely does not comprise all attitudes of relevance concerning political debates in Germany, due to the data source. One potentially interesting item was the question on the use of nuclear power, which could not be used as it was asked only once. Other subjects, like foreign policy evaluation, were never included in the ALLBUS survey. To check whether the results are heavily shaped by the specific item selection, we provide further analyses with subsets of the used variable set in Section 3.4.3.

Judging absolute levels of polarization is a rather difficult endeavor, since we do not have any absolute scales for polarization. In contrast, judging trends is easier and relevant when taking the current discussion into account (Fiorina and Abrams, 2008, 567). Our measure is therefore the correlation between attitudes towards different issues.[7] By using this measure, we avoid the problem of missing data on many variables for certain waves, as the units of observation are issue pairs rather than individuals.[8] The computed correlations are based on pairwise-complete observations. Table B.1.1 in Appendix B shows that some items were asked only two or three times, but by no means in every wave. This is not as problematic as it might seem for a trend analysis. Even though pairs of attitudinal items are not asked in every wave, their correlations—if present in at least two waves—serve as informative data points for the multilevel model which is described in more detail in the section below. We can make use of a total of 54,243 individual observations and 806 issue-year specific correlation pairs.

[7] The presented results are based on Pearson correlations. Rerunning the models with other correlation-based measures led to very similar results.

[8] When comparing correlation coefficients over independent samples (the ALLBUS is a trend study and has no panel component) we make the assumption that these samples are, plainly spoken, of similar quality. For example, if sample size varied a lot between studies, the coefficients would be estimated with varying uncertainty—uncertainty which we do not incorporate in our models. Varying item or unit nonresponse rates across samples could have similar effects. Besides, changes in the composition of the realized samples over time can be a problem if they do not mirror changes in the target population but are a consequence of different sample designs. For example, a study-induced overrepresentation of highly educated respondents could lead to a somewhat artificially increased correlation coefficient in comparison with the coefficients gained from other studies. However, the ALLBUS is a coherent study project with seventeen studies of similar size (each with around 3,000 respondents), same question wording on the compared items, and sampling schemes all based on random sampling. Hence, even if there is sample-induced noise it should not systematically affect our results. We also checked item nonresponse rates over items and years and did not find any suspect outliers.

Table 3.1: Question wording for items used in the analysis from the ALLBUS 1980-2010

Variable code	Variable label	Question/Statement wording (translated)[1]	Scale length
v318	*gender.job.child.1*	"A working mother can establish just as loving and secure a relationship with her children as a mother who doesn't work."	4
v320	*gender.job.child.2*	"A preschool child is likely to suffer if his or her mother works."	4
v322	*gender.job.child.3*	"A child actually benefits from his or her mother having a job rather than just concentrating on the home."	4
v323	*gender.job.marriage*	"A married woman should not work if there is a limited number of jobs and her husband is able to support the family."	4
v319	*gender.help.husband.1*	"It is more important for a wife to help her husband's career than to have one herself."	4
v321	*gender.help.husband.2*	"It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family."	4
v315	*moral.marriage*	Do you think one should get married if one is living with a partner on a permanent basis?	3
v357/v359/v361/v362	*moral.abortion*	"Should abortion be permitted ... if the woman is married and doesn't want any more children (... if the family has a very low income and can't afford more children, ... if the woman is unmarried and doesn't want to marry the child's father, ... if that is what the woman wants, regardless of her reasons)?"	5
v504	*moral.euthanasia*	"A doctor gives an incurably ill patient a lethal drug on his/her request."	4
v512	*moral.cannabis*	"Somebody smokes marijuana several times a week."	4
v513	*moral.homosexuality*	"A man has a homosexual relationship with another man."	4
v170	*distribution.state.provide.welfare*	"If social welfare benefits such as continued pay for sick workers, unemployment compensation and early retirement pensions are as high as they are now, it only makes people not want to work anymore."	4
v167	*distribution.profits.1*	"The economy can run only if the businessmen make good profits. That benefits everyone in the end."	4
v171	*distribution.profits.2*	"Generally speaking, business profits are distributed fairly in Germany."	4
v172	*distribution.social.inequality*	"Even if one wanted to, there is no way to reduce social inequality any further than here in Germany."	4
v192	*distribution.income.incentive*	" Only if differences in income and social standing are large enough is there an incentive for individual effort."	4
v193	*distribution.rank.difference.1*	"Differences in social standing between people are acceptable because they basically reflect what people made out of the opportunities they had."	4
v194	*distribution.rank.difference.2*	"All in all, I think the social differences in this country are just."	4
v236	*immigration.asylum.seekers*	People seeking asylum? (Should entry be restricted or not?)	3
v238	*immigration.non.eu.workers*	People from non-EU countries, e.g. Turkey coming to work here? (Should entry be restricted or not?)	3
v257	*immigration.lifestyle.adaption*	"The foreigners living in Germany should adapt their way of life a little more closely to the German way of life."	7
v260	*immigration.no.cross.marriage*	"Foreigners living in Germany should choose to marry people of their own nationality."	7
v258	*immigration.no.jobs.send.home*	"When jobs get scarce, the foreigners living in Germany should be sent home again."	7
v259	*immigration.political.rights*	"Foreigners living in Germany should be prohibited from taking part in any kind of political activity."	7

[1]Statements are generally introduced by an explanatory text and ask for the respondent's degree of agreement. For the full documentation, see the ALLBUS Data Handbook (*http://info1.gesis.org/dbksearch19/download.asp?db=D&id=36223*, accessed August 2013).

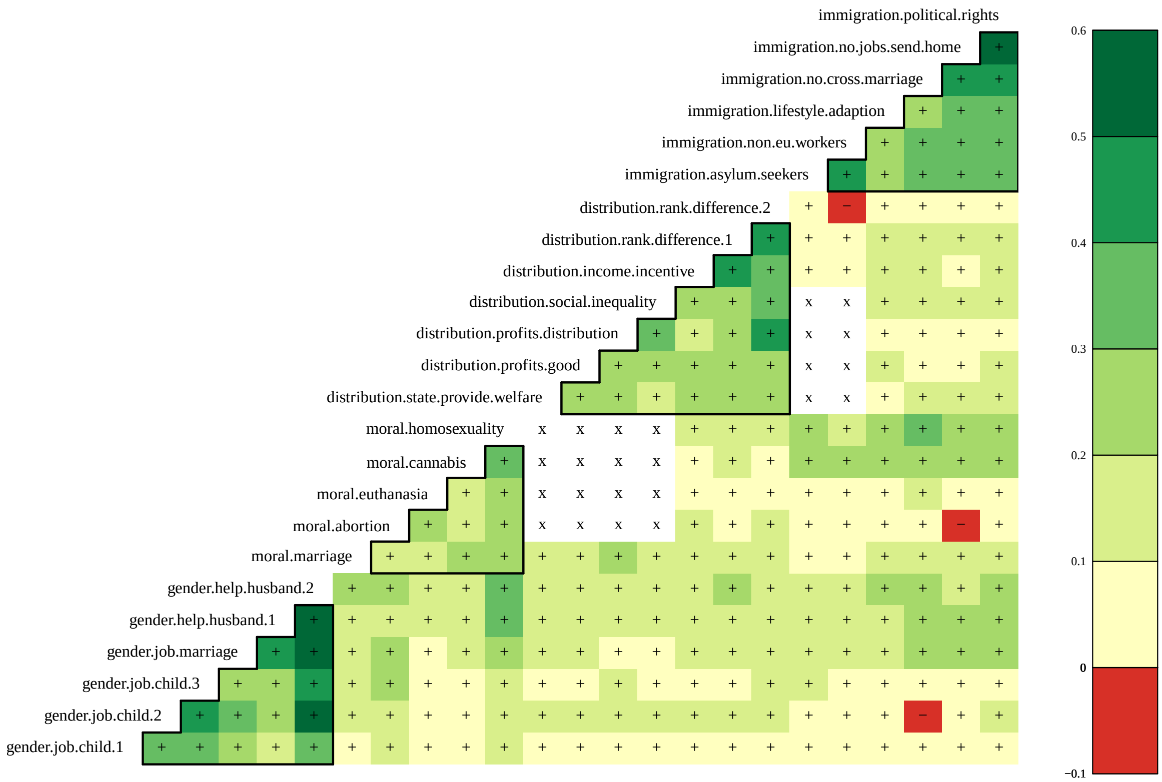
In the second step, students and researchers (seven in total) classified the items according to more general topics. This procedure could be plainly described as a "manual exploratory factor analysis". The resulting grouping enables us to trace different levels and trends of POP not only for the complete sample of attitude scales, but also for subsamples of attitude scales that belong to one of the finally identified issue dimensions. Four of the seven raters chose a four-dimensional solution, and most of the items were classified into similarly labeled dimensions. The only noteworthy disagreements concerned some of the (later labeled) immigration items, more specifically the ones stating that "foreigners living in Germany should choose to marry people of their own nationality" and "when jobs get scarce, the foreigners living in Germany should be sent home again" (see Table 3.1). The former was classified into the moral dimension, the latter into the economic. We ran the analysis again without these items, and the results remained largely unchanged. We therefore decided to use the four-dimensional solution with all items, and labeled the dimensions (according to suggestions from the raters) as 'gender', 'moral', 'distribution', and 'immigration.'[9] Finally, in cases where questions were formulated negatively with regard to the respective issue dimension they belonged to, we turned them around so that the answer categories follow the direction of the respective dimension (for example, liberal → conservative moral attitudes; see Baldassarri and Gelman 2008).[10] Hence, we expect most of the pairwise correlations to be positive, which simplifies the interpretation of the estimated trends.

Figure 3.1 provides an initial overview of the 252 attitude scale pairs under consideration. These correlations are also an initial, aggregated view of our outcome variable. As items were switched according to their respective dimension, most attitude scale pairs are positively cor-

[9] Baldassarri and Gelman (2008) report a similar procedure with four external raters. Certainly, one might object to this procedure and instead suggest to apply ordinary factor analyses or similar procedures to "let the data do the work" in identifying possible dimensions. We do not, for two reasons: First, the data structure—we do not have data for all items for all years—precludes estimating and comparing corresponding measurement models over time. Secondly and more importantly, the corresponding statistical methods suited to identify latent dimensions are often as subjective as the approach we follow (see, for example, Peffley and Hurwitz, 1985). In contrast, it is our strategy to be as transparent as possible so that readers can evaluate results both for single item pairs and for grouped dimensions of items.

[10] In their analysis, Baldassarri and Gelman (2008) orient attitude scales along the liberal-conservative continuum, which seems to make sense in the case of the United States. For the case of Germany, ordering according to the left-right scale would be too simplistic.

Figure 3.1: Mean correlations of the 252 attitude scale pairs (across all years for the specific pair). Positive correlations are colored light to dark green, negative correlations red. Correlation direction is further indicated with plus and minus signs. An 'X' indicates that no observations were available. Black triangles frame correlations of items within the four issue dimensions "gender," "moral," "distribution," and "immigration."



related, and the correlations among scales belonging to the same issue dimension (framed by black triangles) tend to be higher than correlations between scales belonging to different dimensions. Also note that for some attitude scale pairs, we did not observe any correlations (for example, the pair *moral.sexuality–distribution.profits.good*), as the corresponding items were never administered in the same waves. This is not a serious problem for the estimation method we apply. Although it provides only a static view of the set of attitude scale pairs, the plot gives a first impression of POP in Germany (if it is as a state rather than a process).[11] Higher correlations within the four issue dimensions (around 0.2 to 0.6) show that within policy do-

[11] See DiMaggio, Evans and Bryson (1996).

mains, public opinion is rather structured, that is, there is a tendency towards polarization on, for example, family or immigration politics. The correlation in the upper-right corner indicates that if a respondent thinks that "when jobs get scarce, foreigners living in Germany should be sent home" (item *immigration.no.jobs.send.home*), he also has a higher probability of approving the statement that "foreigners living in Germany should be prohibited from taking part in any kind of political activity" (*immigration.political.rights*). More important, however, is the observation that the correlation between dimensions is low to very low (almost always less than 0.2, and often there is no statistical relationship at all). This means that there is little alignment across different policy domains. If we know only that a respondent agrees with the statement that "a child actually benefits from his or her mother having a job rather than just concentrating on the home" (*gender.job.child.3*), we cannot infer much about how the respondent thinks about homosexuality (*moral.homosexuality*), social welfare benefits (*distribution.state.provide.welfare*) or asylum seekers (*immigration.asylum.seekers*)—the mean correlations with all of these items are virtually zero. This is the empirical manifestation of what we expressed theoretically above when we argued that POP is only a serious issue if people align along multiple issues and build overlapping groups. This does not seem to be the case when eyeballing the aggregated correlations. There are stronger relationships for some pairs of variables, for example, opinions on homosexuality and agreement with the statement that "foreigners living in Germany should choose to marry people of their own nationality" (*immigration.no.cross.marriage*). However, the average correlations in Figure 3.1 do not tell us anything about the evolution of alignment, that is, polarization. The corresponding trends are assessed in a slightly more complex modeling framework which is presented in the following section.

3.3.2 Estimating public opinion polarization trends

To estimate the time trends of polarization, we employ a multilevel approach, using the year-specific correlation of pairs of attitude scales as dependent variable and the pairs themselves as units on the second level. The explanatory variable is time t ranging from 1980 to 2010. For a better interpretation (and because we are interested in long-term effects), the time variable has been rescaled to decades and centered at 1994. The basic model (Model A) is specified as follows:

$$\rho_{pt} = \alpha_p + \beta_p t + \varepsilon_{pt}, \tag{3.1}$$

with ρ_{pt} representing the correlation of attitude scale pair p in the year t.[12] Both the intercept α and the trend coefficient β are allowed to vary over attitude scale pairs. In other words, we estimate attitude scale pair-specific levels of average correlation (intercept) and time trend (slope). ε_{pt} is the attitude scale pair year-specific error term which is assumed to follow a normal distribution. Implementing a multilevel approach serves at least two purposes: First, the estimation of some trends may be unstable, especially when a specific pair of issues is only observed for a few years. The varying-intercept, varying-slope approach allows us to explicitly capture this uncertainty in the estimates. Secondly, simply estimating an overall trend is only the first step of our analysis. In order to obtain a finer grained picture of POP trends, we are also interested in looking at issue-dimension-specific correlation trends. The estimated variation in the multilevel models gives us exactly this kind of information.

While Model A provides an estimate for the overall trend of POP (the average of β_p across all pairs, $\beta_p \rightarrow \beta$) and issue-pair-specific trends, we respecify the basic model in several ways: In Model B, we estimate trends of attitude scale pairs that are located within or between issue dimensions. More precisely, it allows us to evaluate separate trends for within-dimension attitude scale pairs and between-dimension attitude scale pairs. Here, the time trend variable is interacted with a dummy variable that indicates whether the two attitude scales belong to the same or different issue dimensions. Thus, we can determine whether attitude alignment between attitude scales belonging to a certain dimension has risen more than alignment be-

[12] Note that the modeled Pearson correlation coefficients have natural bounds at -1 and 1. We nevertheless chose a linear functional form, because empirically, the examined correlation coefficients are far away from these borders (see Figure 3.1), and the estimated trends do not seem to bring them anywhere near the technical limits. This approach is in line with Baldassarri and Gelman (2008) who handle correlation coefficients of a similar size.

tween attitude scales that belong to two different dimensions. Model C introduces four groups that correspond to the three issue dimensions and a group for cross-dimension pairs (that is, the group of attitude scale pairs that do not belong to the same dimension). As ever, ρ_{pt} is the correlation of a pair of attitude scales at point t in time and the group level is a specific pair. The trend term is now interacted with a variable that indicates the attitude scale pairs belonging to a certain issue dimension.

When moving to the subgroup analyses, we simply divide the sample of respondents according to certain characteristics (gender, education level, income, political interest, religious denomination, and East versus West). Statistically, the models follow the known specification.

3.4 Findings

3.4.1 Attitude alignment within the German population

Table 3.2 displays the results of our investigations of attitude alignment. As described in the previous section, Model A reports the trend across all attitude scale pairs. The average correlation between scales is 0.19 with a standard error of 0.01. The average correlation has decreased by 0.04 per decade, hence, there is a negative trend. This result is highly significant. However, when looking at the pair-specific trends, the picture becomes much more heterogeneous. Figure 3.2 displays the coefficients for all attitude scale pairs over time. They range from −0.15 to 0.04. One can observe that the correlation has decreased in most of the cases. Interestingly, the cluster of attitude scales belonging to the gender dimension (at the bottom left of the triangle plot) shows a systematically rising trend of correlations. These items have been asked seven to eight times since 1982, so opinions on these issues are well captured. We will elaborate on this finding later. Among the other item pairs, the size of the decrease varies to a certain extent. On average, it is strongest for pairs of items of the distribution and the immigration dimension. Three item pairs show a positive correlation trend, but there is no systematic alignment over different dimensions. In general, with the exception of the gender domain, Figure 3.2 reveals a

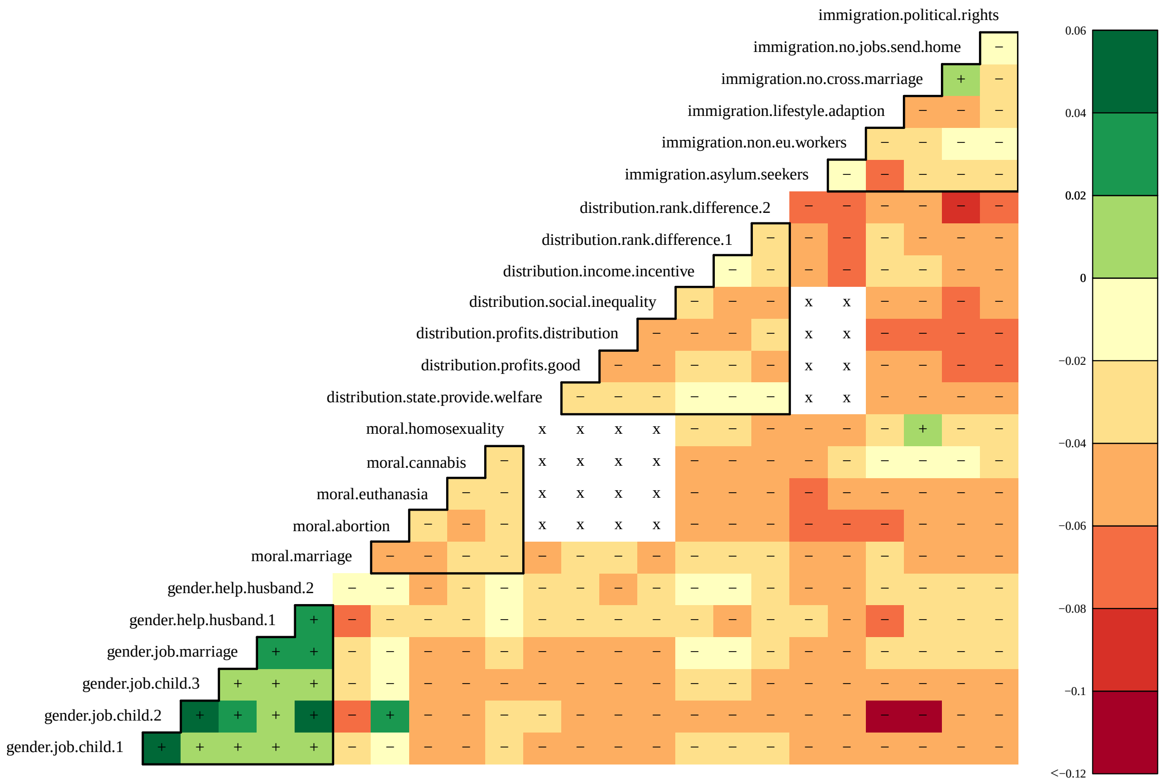
Table 3.2: Fitted multilevel models (varying-intercept and varying-slope) for attitude alignment

	Model[1]	$\rho \equiv$ Pairs of attitude scales
A	**No grouping of pairs:**	
	Intercept	0.19 (0.01)
	Time (decades)	-0.04 (0.00)
	Residual variance	
	Intercepts (σ^2_α)	0.01
	Trends (σ^2_β)	0.00
	Data (σ^2_ε)	0.00
B	**Within and between issue dimensions:**	
	Intercept	0.14 (0.01)
	Within dimension pairs	0.18 (0.01)
	Time (decades)	-0.04 (0.00)
	Time x Within dimension	0.02 (0.01)
	Residual variance	
	Intercepts (σ^2_α)	0.01
	Trends (σ^2_β)	0.00
	Data (σ^2_ε)	0.00
C	**By types of issue dimensions:**	
	Intercept	0.36 (0.02)
	Gender	baseline
	Moral	-0.12 (0.03)
	Distribution	-0.07 (0.02)
	Immigration	0.01 (0.03)
	Mixed pairs	-0.22 (0.02)
	Time (decades)	0.04 (0.01)
	Time x Gender	baseline
	Time x Moral	-0.09 (0.03)
	Time x Distribution	-0.08 (0.01)
	Time x Immigration	-0.08 (0.01)
	Time x Mixed pairs	-0.07 (0.01)
	Residual variance	
	Intercepts (σ^2_α)	0.00
	Trends (σ^2_β)	0.00
	Data (σ^2_ε)	0.00

[1] Number of observations: 806, number of item pairs: 252. Numbers in parentheses are standard errors. The intercepts correspond to the estimate of 1994 as the time variable is zeroed at this point of time (and rescaled to decades).

decrease in attitude alignment for all other issue pairs—both within and across domains. The overwhelming majority of negative trends (only 18 of the 252 estimated trends are positive), especially between dimensions, is a strong indicator of public opinion depolarization rather than polarization.

Figure 3.2: Correlation trends of the 252 item pairs over time: The plotted coefficients are based on Model A ('no grouping of pairs'). Decreasing trends are colored beige to red, increasing trends light to dark green. Trend direction is further indicated with positive and negative signs. An 'X' indicates that no observations were available. Black triangles frame correlations of items within the four issue dimensions "gender", "moral", "distribution", and "immigration".



Back to Table 3.2, Model B is slightly more sophisticated in that the attitude scale pairs are grouped into pairs that belong to the same issue dimension and pairs that belong to two different dimensions. Naturally, pairs that belong to the same dimension should display higher levels of correlation. Here, the estimate of intercept of 0.14 represents the average correlation for between-dimension pairs. The intercept for the group of within-dimension pairs is, as expected, much larger (0.14 + 0.18 = 0.32). Moreover, we can observe a decreasing trend of the average correlation of between-dimension pairs (−0.04), however, the trend among within-dimension pairs is less developed (−0.04 + 0.02 = −0.02). This finding strengthens our initial impressions of the data in Figure 3.1. Levels of alignment across different policy domains (which are not very high anyway) seem to decrease even more sharply than alignment within policy domains. There is no increasing trend of correlation over multiple issues, which is another strong indicator of public opinion depolarization.

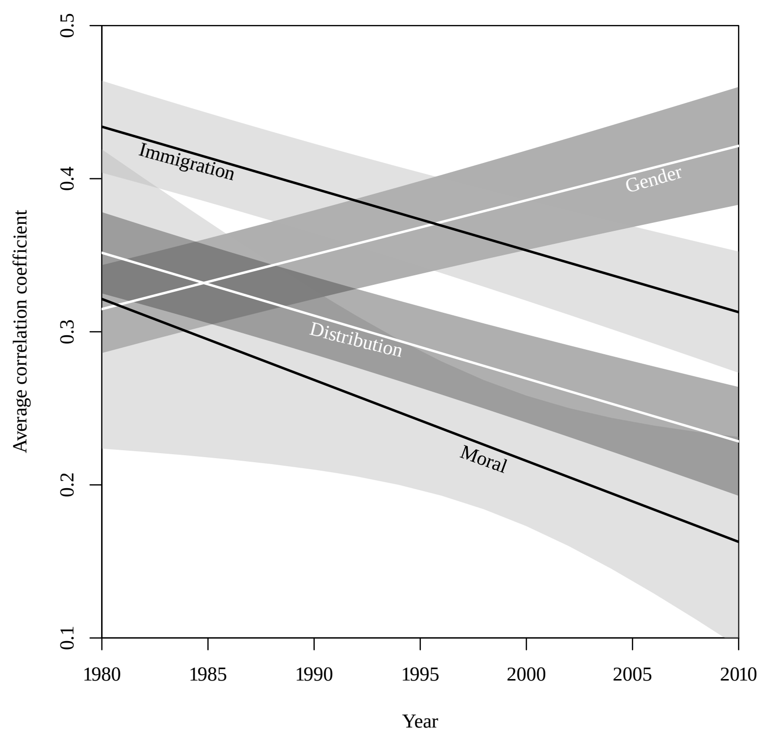
Since the results only differentiate within-dimension and between-dimension pairs on average, we set up another model that accounts for dimension-specific time trends (Model C) by interacting the trend variable with dimension indicators and an indicator for mixed pairs, that is, pairs from different dimensions. On average, the correlation within these dimensions is 0.36 for issue pairs from within the gender dimension, 0.21 for the moral dimension, 0.29 for the distribution dimension and 0.37 for the immigration dimension. What is already visible when looking at Figure 3.2 is now reflected in the numbers: While gender issues show a positive (that is increasing) correlation trend, the time effect of all other dimensions is negative. The dimension-specific trends are again visualized for the whole time period under consideration in Figure 3.3. Model C reveals the evidence for a trend of polarization on the gender dimension. The diagnosed trends are disentangled in a more disaggregated analysis in the following section.

3.4.2 Attitude alignment within subpopulations

So far, we have described several noticeable overall trends of POP. Nonetheless, it may be the case that these effects are driven by trends that only concern parts of the German population. One of the drawbacks of our conceptional and empirical approach to public opinion polarization is that we cannot directly integrate individual-level indicators of status, educational background and so forth, as our level of analysis is that of aggregated survey measures. In order to check whether the effects remain robust or vary over subgroups, we split the sample into several groups *a priori* according to socioeconomic and other strata, and then estimate group-specific levels and trends.

For instance, it is generally argued that changes in public opinion (polarization) originate within intellectual elites and resonate first among the highly educated and politically interested

Figure 3.3: Correlation trends by issue dimension, over time. The lines are based on the effects presented in Model C ('By types of issue dimensions'). Shaded areas represent 90% confidence intervals.



(Zaller 1992; Adams, Green and Milazzo 2012*a*; Baldassarri and Gelman 2008; Abramowitz and Saunders 2005). Hence, these groups may be pioneers in public opinion trends. Moreover, we deem it relevant to search for diverging trends among subpopulations of gender (Shapiro and Mahajan 1986) and income (McCall and Manza 2011).[13] A further subpopulation that merits attention is different religious denominations (Jelen 1990). Finally, public opinion trends may differ between East and West Germany, making this another relevant subgroup division (Wegener and Liebig 1995).

[13] The group of more educated people was defined as those who hold an academic high school diploma (*Abitur*), compared to the less educated who hold no certificate of school completion or at most a certificate of secondary education. People who are very strongly or strongly interested in politics make up the subgroup of high political interest and those with no or hardly any interest in politics the counterpart. People with a monthly income greater than 2,000 Euros are defined as having a high income, and those earning less than 500 Euros as having low income. Other specifications in which the thresholds of the categorization were varied led to very similar results.

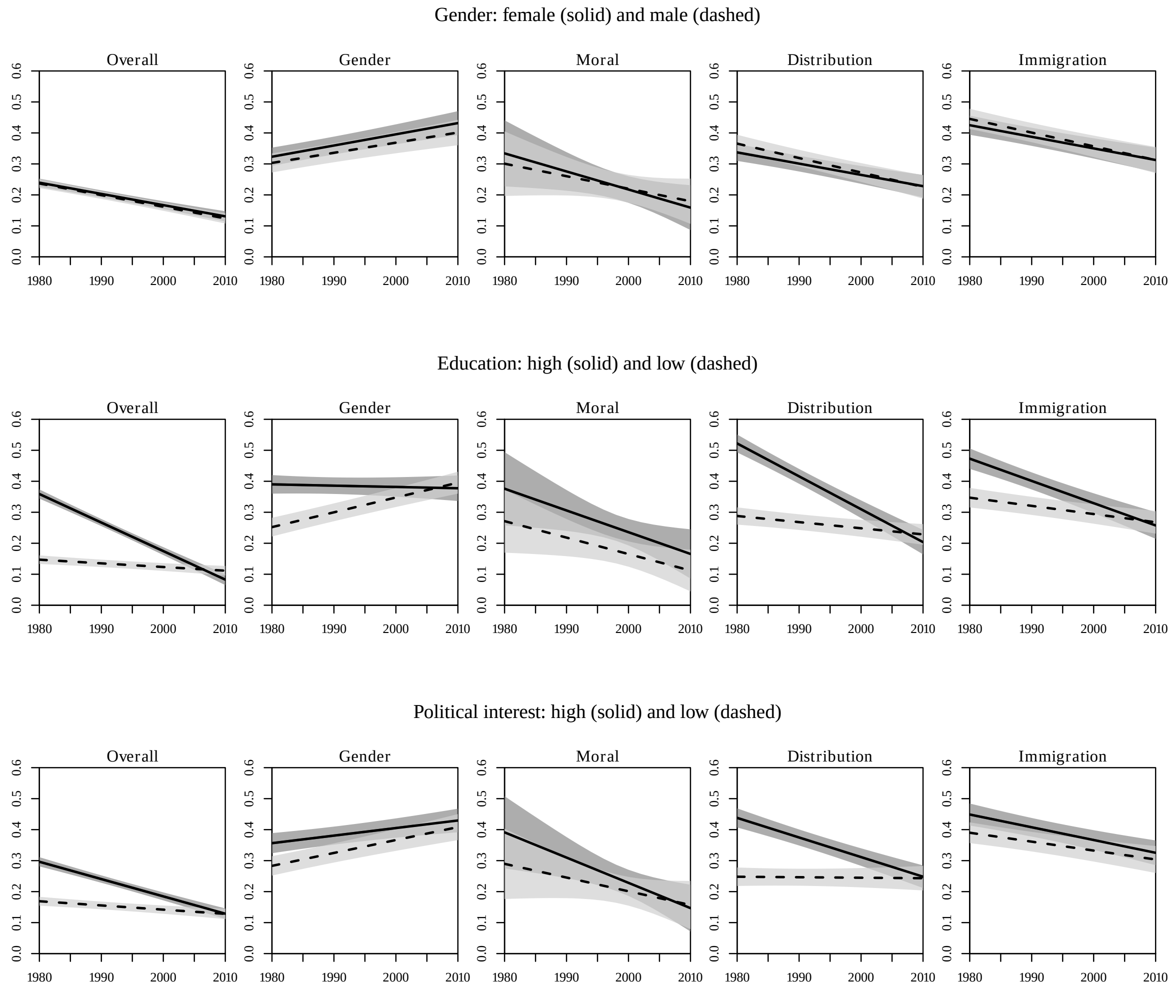
Figure 3.4 provides an overview of the estimated trends. All figures visualize results from multilevel models based on the same logic as the models presented above. Both intercepts and slopes are allowed to vary for each subgroup. Models visualized under the label 'Overall' display the average trend across all attitude scales and are specified as in Model A. The four other panels under each subgroup distinguish between issue-dimensions—that is, they report results from Model C specification. While each of the figures may tell its own complex story—and should be the subject of greater study in the future—our focus is on general patterns. Levels and trends differ considerably between certain subgroups, but not between all. We want to point out the findings we regard as the most illuminating.

Regarding the *overall trend* (Column 1) there are no significant differences for subpopulations of gender, income and religious groups. However, the overall decreasing trend is much stronger among highly educated, highly interested people, and somewhat stronger among respondents from West Germany.[14]

The trend of increasing polarization on the *gender dimension* (Column 2) is similar across most subpopulations. Only the difference between the more and less educated is striking. Whereas the more educated display no increase or decrease in attitude alignment, the increase in polarization seems to take place among the less educated on this dimension. We want to avoid in-depth speculation at this point, but there have been some developments on this topic that could put this finding into context. Germany has seen an improvement in women's rights over the last decades. The 1957 gender equality law (*Gleichberechtigungsgesetz*) abolished significant inequalities between women and men in marriage, for example, women could in principle choose to take a job without their husband's approval, and possession of goods was not solely assigned to the man (Bundesministerium für Justiz, 1957). Since then, several amendments have strengthened the position of women. Nonetheless, public debates on related topics are continuously boiling up such as the female quota, child care subsidies and child care in general. One could indeed argue that the gender question remains one of the few remaining cleavages (rooted in parts of the population) between the left and right camps in Germany,

[14]The models for the Eastern subgroup are naturally limited to the variable subsample from after 1990 and therefore trends are only displayed for this time span.

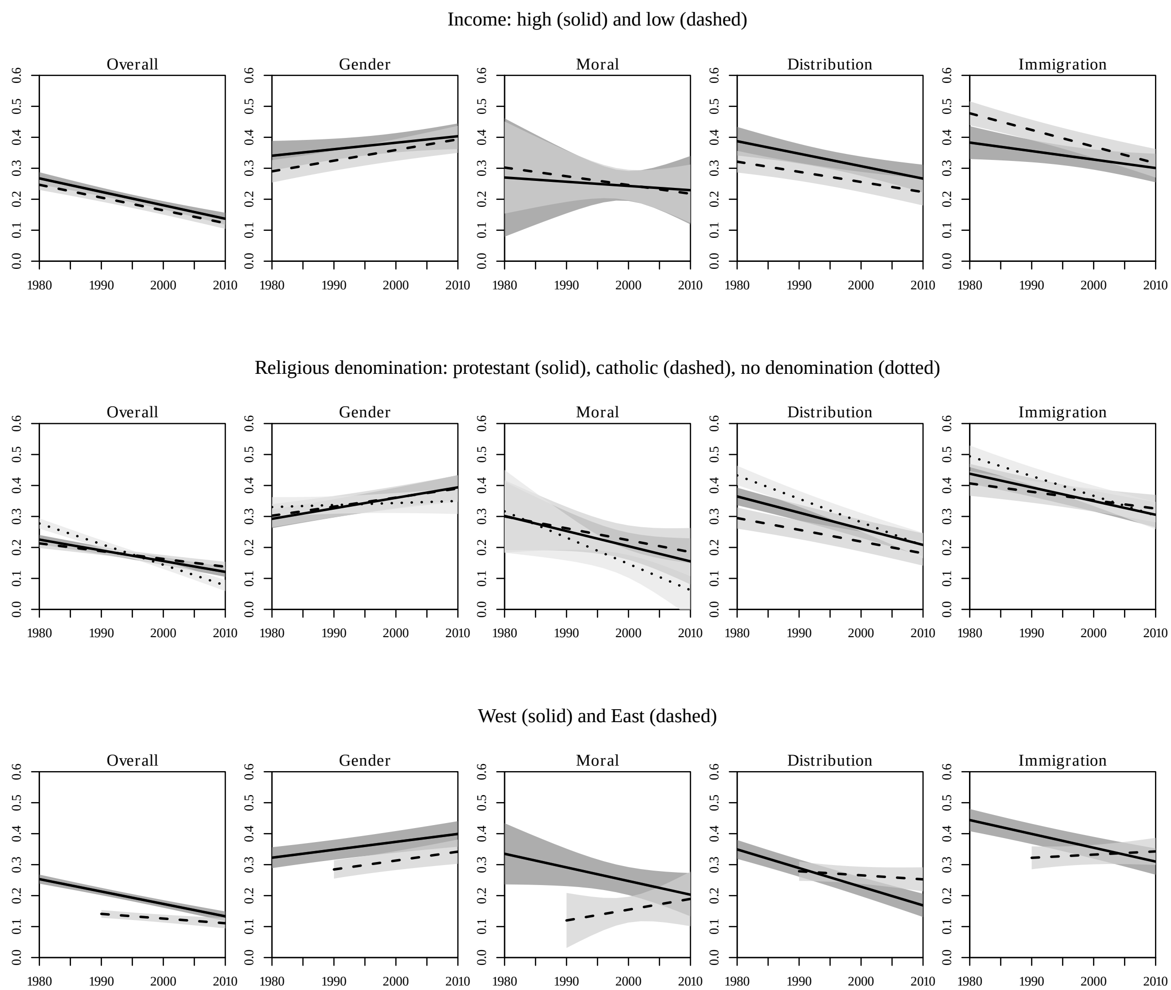
Figure 3.4: Polarization trends among several subgroups. Shaded areas around the effects (solid and dashed lines) represent 90% confidence intervals based on simulated responses from the model as a visualization of uncertainty.



which have converged on many other highly debated topics such as unemployment assistance or the civil use of nuclear power.

Concerning the *moral dimension* (Column 3) the plots reveal a general decreasing trend of polarization across several subpopulations. However, the levels of polarization differ among subpopulations of more and less educated. Moreover, there is a considerable difference between trends among respondents from West and East Germany. Public opinion polarization

Figure 3.4, *continued*: Polarization trends among several subgroups.



has decreased in the former and increased in the latter. These estimates, however, have to be taken with a grain of salt due to higher uncertainty (see uncertainty intervals in the plot, which are wider on the moral dimension due to the smaller amounts of data). On the *distribution dimension* (Column 4) we again find significant differences between subpopulations of education, political interest and respondents from West and East Germany. In contrast, trends are similar among the other subpopulations. On the *immigration dimension* (Column 5), trends are similar across most subpopulations. Nonetheless, the trend of depolarization is again strongest among the more educated, and there is a difference between West and East Germany.

To sum up, the subgroup analysis reveals some heterogeneity in the development of public opinion polarization among subpopulations. We will further comment on this discovery in the discussion and conclusion.

3.4.3 Further tests of robustness

We estimated various respecifications of our models to check the robustness of the presented results. To see whether some of the selected attitude scales (variables) have a significant impact on the estimated direction and size of the overall and dimension-specific effects, we reestimated Models A and C by dropping every possible combination of one (resulting in 24 different models), two (276 models), three (2,024), and four variables (10,626).[15] The results are reported in Figure B.1.1 in Appendix B. It can be seen that the point estimates of the trending effects go in the same direction for almost all of the 12,950 estimated models, and show medium to low variance (this is, of course, hardly surprising when recalling Figure 3.2, which revealed a remarkable homogeneity of trends within dimensions). There is an exception in the moral dimension, where some of the specified models show a positive effect.[16] However, the reassuring insight we draw from these models is that the reported trend estimates are remarkably stable for various subsets of the original variable sample.

Another set of respecifications targets possible question order and measurement issues. In general, the selected variables are scattered far and wide within the questionnaires. However, questions belonging to one dimension are sometimes asked within a question battery. For example, all questions related to the gender dimension are part of a set of questions focusing on

[15] The number of possible models results from $\frac{N!}{n!(N-n)!}$, with N being the number of variables and n the number of removed variables. Taking out more than four variables led to problems in model convergence because not all pairs of variables gave substantive information on correlation trends when they were collected together only a few times. Some of the estimates for the specifications with four variables dropped also suffered from lack of convergence and were excluded from the reported results.

[16] In every case, the positive effect was, however, not significantly different from zero.

the role of women in the family, and were asked in one row as statements. The respondent was then encouraged to agree or disagree. This is not the case for the other dimensions, in which only some items belonged to the same set of questions, and others were asked earlier or later in the questionnaire. Asking for attitudes in question batteries may potentially lead to higher correlations between the measured scores, as batteries are likely to provoke *response set* bias, that is, respondents answer similarly on each of the questions rather than take the trouble to evaluate every statement and think about the right answer.[17] While it seems plausible that this phenomenon may lead to biased estimates of polarization levels, we do not expect the trend estimates to be affected, as items of the same battery are almost always asked in the same waves, and because systematic volatility in acquiescence seems unlikely (see also Billiet and Davidov 2008). Our expectation is confirmed when we reestimate the models using only items within each dimension that are part of the same battery.[18] As can be seen in Figure B.1.2 in Appendix B, the overall levels of correlation rose in all other dimensions (about 0.02 points for the moral values, 0.06 points for the distribution and 0.04 points for the immigration dimension). The trend estimates, however, remained stable, indicating that our findings concerning the exceptional trend of POP regarding gender issues are not a measurement artifact, but are fairly robust against variable selection in the present sample of variables.

We also specified more flexible models by including trend polynomials of the second and third degree. Because the coefficients become harder to interpret, we only present the graphical output in Figure B.1.3 in Appendix B. Trend estimates for the moral dimension are not reported as they are highly unreliable in this nonlinear specification due to data sparsity in the first and last decade of the observation period. What can be observed is that the original linear specification might hide some nontrivial trend dynamics, especially for the last five years of the sample: the estimated slope of immigration issues rises again between 2005 and 2010. However, the increased widths of the confidence bands indicate that this trend has to be taken

[17] See, for example, Billiet and McClendon (2000), Green and Citrin (1994) and McClendon (1991).

[18] Specifically, from the gender dimension, we kept all variables. From the moral dimension, we kept the items on abortion, euthanasia, cannabis, and homosexuality. From the distribution dimension, we kept the items "income incentive," "rank difference 1," and "rank difference 2." From the immigration dimension, we used the items "lifestyle adaption," "no cross-marriage," "no jobs send home," and "political rights."

with a grain of salt, so we refrain from further post hoc speculations as to why there could be a recently rising trend on this dimension, and patiently wait for new data from the follow-up waves to shed more light on this trend with a richer data base.

Finally, we employed a different concept of public opinion polarization: attitude dispersion.[19] For this purpose, the standard deviation for each issue in every available year is computed and then modeled similarly to our standard approach. Single attitude items, not variable pairs, constitute the grouping variable, and standard deviations measured at time *t* are nested within these groups. We find a negative overall trend of dispersion, however, it is not significantly different from zero on the usual levels. Analogously to our main results dispersion is increasing for gender issues, and decreasing for other issues. Again, these effects are mostly insignificantly different from zero. These results are not surprising since both measures—correlation (see the Conception and Measurement of POP section) as a measure of attitude alignment and standard deviation as a measure of attitude dispersion—account for the extremity of positions on attitude scales. Thus, the fact that the trend coefficients point in the same direction for both measures is reassuring even if the coefficients for the dispersion measures are not significant. While this alternative specification is not a very strong test of the validity of our results,[20] we think it offers valuable information on the compatibility of different measures of polarization. In the presented scenario, insights from the attitude alignment perspective do not lead to contradictory findings when compared with results based on the attitude dispersion measure. We have already argued extensively why we prefer the former.

3.5 Discussion and conclusion

Studies on public opinion polarization have been largely confined to the U.S. context. We offer the first descriptive overview of POP trends in Germany between 1980 and 2010, and find

[19] See DiMaggio, Evans and Bryson (1996).

[20] Note, for example, that there is a scaling problem inherent to the measure: Not all of the items have the same scale range, which is between 3 and 7 for the selected variables; see Table 3.1. Variance, however, is a function of variable scaling: higher-scaled variables are expected to produce higher levels of variance.

that the overall level of POP, conceptualized and measured as attitude alignment, has declined. This result seems to mirror evidence from studies on party polarization, in which convergence on several policy dimensions can be observed (Spier 2011). Attitude alignment between domains is low and still decreasing (that is, we do not find a society in which POP extends over different issue dimensions), and our analyses of trends among subpopulations reveal that this overall negative trend mainly takes place among the highly educated and the politically interested. Stronger change among these subgroups seems to be in line with arguments that (de)polarization is an elite phenomenon.[21] Strikingly, the overall trend does not hold for gender issues. POP in Germany has grown for attitudes towards the role of women in the family. The trend for this issue dimension is similar and increasing among most subgroups. However, when we divide the sample into more and less educated respondents we clearly see that the development of polarization mainly took place among the less educated. This domain-specific finding, and the fact that the increasing trend only takes place among the less educated, is a puzzle that requires further research. Further, the differentiation between West and East German subpopulations reveals interesting subtrends. West Germans follow the general trend, while East Germans deviate somewhat with regard to moral, distribution and immigration issues. Although we observe no negative or positive trends regarding distribution and immigration issues, it remains positive for moral issues. While we refrain from speculations about why these differences may occur, we think they would provide interesting starting points for future studies.

Just like measurements of absolute values of polarization, trends are relative and more revealing when compared with other countries. Baldassarri and Gelman (2008, 431) present an overall increasing trend of attitude alignment (coefficient = 0.02 per decade) in the United States. The authors find this trend modest in comparison with trends of issue partisanship, that is alignment between attitude scales and the left-right scale. However, in light of the overall negative trend we find for the German case (coefficient = -0.04 per decade), Baldassarri and Gelman's results have to be reevaluated. In contrast to Germany, which is seeing an overall

[21] See, e.g., Abramowitz and Saunders (2005) and Adams, Green and Milazzo (2012*a*).

decreasing trend of polarization for attitudes towards most issues, POP seems to have slightly increased in the United States.[22] In addition, whereas the politically interested experience a stronger positive trend of attitude alignment than those who are less interested in the United States (see Baldassarri and Gelman 2008, 435: Figure 8, Panel A), the opposite is true for Germany (see Figure 3.4). Other European studies also find generally decreasing levels of public polarization, for example the British (Adams, Green and Milazzo, 2012*a*,*b*) and Dutch cases (Adams, Vries and Leiter, 2011). We can only speculate that the diverging trends in the United States and the examined European countries could be a consequence of elite polarization in the United States on the one hand, where conservative Democrats and liberal Republicans seem to have left Congress,[23] and depolarization in European party systems on the other (as described for the British case by Adams, Green and Milazzo 2012*a*). There is some agreement among German scholars that the party system has seen a trend of depolarization during the last decades (e.g., von Alemann, 2010; Spier, 2011). Thus, parties seem to follow the same broad decreasing trend we found for German public opinion. Nonetheless, we are not able to make any assertions concerning the causal connection between the two phenomena. We deliberately did not follow a partisan- or ideology-based perspective on polarization, as partisan camps are more difficult to identify in a multiparty setting than in a two-party case like the U.S., where such a perspective is more common (e.g., Baldassarri and Gelman, 2008; Hetherington, 2009). However, in order to get a deeper understanding of the mechanisms behind mass polarization, we suggest the connection between party polarization and POP from a comparative perspective as an urgent topic for further research.

Beyond that, our approach and findings may prove useful for scholars of public opinion in general as they open multiple avenues of further research. One avenue concerns the variety of substantive issues we have illuminated. Researchers of public opinion have compared attitudes towards the issues investigated here both between units and over time, for instance attitudes

[22] Surely, this comparison is somewhat impaired by different samples of variables. However, the estimates in both studies are very stable and rely on issues that are of true relevance as they shape the political debate in their own context.

[23] See, e.g., Fleisher and Bond (2004).

towards redistribution and the welfare state,[24] gender roles,[25] attitudes towards moral topics such as abortion, homosexuality and euthanasia,[26] and attitudes towards immigrants and immigration.[27] However, the perspective we pursued in this study changes the main focus from simple averages to relations between attitudes and distributions of attitudes among groups of people. We are convinced that this changing focus may lead to a better understanding of public opinion in general and reveal interesting insights, especially because the polarization aspect of attitudes— and its development—have largely been neglected outside the United States.

A second avenue concerns a drawback regarding our measurement of polarization. While it is possible to evaluate within-group polarization, between-group polarization received no scrutiny. Recently proposed measures[28] illustrate that such approaches may further enrich analyses. A possible way to preserve the advantages of the attitude alignment measure used here, but to account for between-group differences nonetheless, would be to calculate standardized Euclidean distances for a set of various attitude items, and subsequently evaluate them along the lines of different groups. Evaluation of such techniques should be the subject of future work.

A third avenue concerns examining explanations of trends and levels of POP in general. Studies that focus on single countries are insufficient in this regard. In our view, more comparative studies are needed that investigate causes of POP. Cross-country survey projects, like the Eurobarometer or the European Social Survey, might offer data that are better suited to look for macro-level factors that influence polarization trends. In this regard, studies such as Down and Wilson (2010) can be regarded as fruitful starting points.

[24] See, for example, Andress and Heien (2001), Bean and Papadakis (1998), Blekesaune and Quadagno (2003), and Larsen (2008).

[25] See Davis and Greenstein (2009) for an overview.

[26] See, for example, Carter, Carter and Dodge (2009), Cohen et al. (2006), Granberg and Granberg (1980), Loftus (2001), Scott (1998), and Yang (1997).

[27] See, for example, Mayda (2006), Meuleman, Davidov and Billiet (2009), and Wilkes and Corrigall-Brown (2011).

[28] For example, by Levendusky and Pope (2011).

Chapter 4

A Methodological Framework for Constituency-Level Election Forecasting

Simon Munzert

Abstract

Scholarly efforts to forecast parliamentary elections predominantly target the national level and disregard outcomes of constituency races. In doing so, they frequently fail to account for systematic bias in the seats-votes curve and are unable to provide candidates and campaign strategists with constituency-level information which could help in localized campaigning. On the other hand, existing accounts of constituency-level election forecasting suffer from data sparsity and, as a consequence, lack of precision. In this paper, I propose a methodological framework which allows correcting individual constituency-level forecast models for systematic errors and combining them according to their past performance. I demonstrate the use of the method with an out-of-sample forecast of 299 district races from the 2013 German federal election.

> 'I think you'll have a big increase in the number of poll-driven forecasts here. We need more polling here. If you don't know what's happening in individual constituencies that's a bit tricky. Especially if you have votes drifting off to Ukip and these third and fourth parties, that makes it more complicated than just assuming a uniform swing, potentially. And generally the volatility is higher the more parties are involved in the campaign, because voters can say "I want to vote for Ukip but they're not going to win in my constituency so I'll have to vote for the Tories instead."'—*Nate Silver (2013)*

4.1 Introduction

Academic election forecasting has flourished over the last years. Moreover, current practice has overcome many of the drawbacks inherent to fundamentals-based regression approaches which used to dominate this subfield of electoral research (e.g., Sigelman, 1979; Hibbs, 1982; Lewis-Beck and Rice, 1984*a*,*b*; Abramowitz, 1988; Lewis-Beck and Rice, 1992). New models which provide timely, dynamic and, most importantly, reliable forecasts are of great interest for campaign organizers and financial contributors who want to target their resources efficiently. Further, they offer valuable information for voters and are staple food for horse race journalism. From the scholarly perspective, election forecasts can inform research on the nexus between public opinion and campaign dynamics (Lodge, Steenbergen and Brau, 1995; Wlezien and Erikson, 2002; Zaller, 2004; Panagopoulos, 2009) and on methodological issues in the measurement of public opinion (Gelman and King, 1993; Graefe, 2014; Smidt, 2014).

Recent efforts to forecast state-level outcomes at U.S. Presidential elections have profited from both a substantial increase in pre-election polls at the state level and very efficient modeling strategies (Linzer, 2013; Lock and Gelman, 2010), and in the end obtained spectacular success.[1] However, these models have been developed for a very specific context—essentially, a set of two-candidate races with a long election record—and cannot be applied to other settings without further ado. The opening quote of Silver, whose attempt to forecast the 2010 UK General Election with a modified uniform swing model produced rather underwhelming results (Silver, 2010), illustrates a few of the challenges: If an election involves races of more than just two parties in a multitude of constituencies, one is hardly befallen of hundreds of constituency-level vote intention polls. This renders pure poll-averaging approaches useless if one is interested in constituency-level forecasts. Additionally, multiparty settings in combination with FPTP rules induces strategic voting incentives that blur the link between party preferences and actual voting behavior. Consequently, election forecasts in parliamentary settings usually target the national level.

In this paper, I propose a framework to forecast electoral outcomes at the constituency level. While local data to inform forecasting models about voter preferences are generally sparse at any single election, the electoral history of constituencies—and forecast models—can be used to unbias forecasts and ultimately improve model performance. Additionally, I suggest to pool information by combining several available models. The technical procedure breaks down to three stages. In the first, distinct constituency-level forecasts are produced (or collected) for past elections. In the application, these are a set of forecasts based on past election results and a set based on polling data. As the forecasts of these models might be both unreliable and biased because of data scarcity and disparity or fundamental flaws in the models, I unbias forecasts by regressing previous actual results on forecasts and a set of additional predictors (if at hand). Simultaneously, the out-of-sample forecasts are corrected. In the third step, corrected forecasts are combined and weighted according to their performance in past elections.

To demonstrate the use of the approach and to illustrate the gain in precision, I present a forecast of the German federal election held in September 2013—a multi-party setting with a considerable number of districts (299). The corrected and combined forecasts perform very well; the forecast distribution of seats is almost identical to the actual outcome of the election.

[1] For example, state-level forecasts offered by Silver (2012), Linzer (2013), and Jackman (2012) for the 2012 U.S. Presidential election remained flawless, each of them forecasting the winner in all of the 50 states correctly.

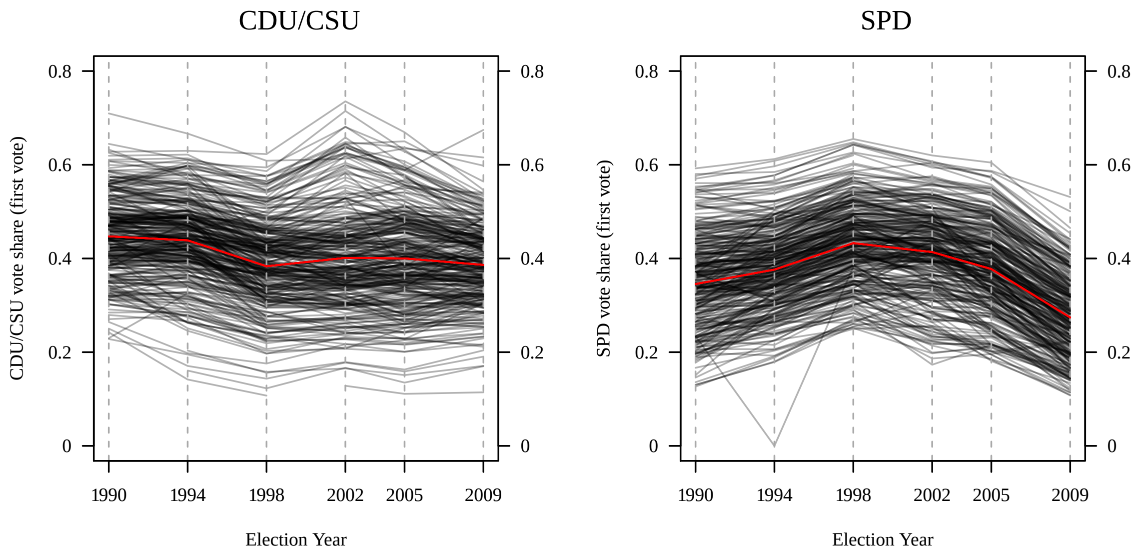
4.2 Potentials and challenges

Forecasting models that provide genuine information about constituency-level outcomes bear great potential. First, constituency-level outcomes are important for the actual distribution of seats at the national level in many electoral systems with a significantly biased votes-seats curve. 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 2010, the winning Conservatives gained 47% of the seats, but only 36% 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; Norris and Crewe, 1994; Tufte, 1973) which makes it difficult to forecast the distribution of seats from national-level polls alone. In general, specific characteristics of electoral rules may prohibit forecasters to infer 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 actually comes to power. Forecasting at the level where the race is decided is therefore best suited for answering questions about the prospective allocation of seats.

Constituency-level forecasts can also be useful in non-majoritarian settings. The internal composition of parliaments in mixed electoral systems with a strong PR component, as given in the German electoral system, could be forecast more precisely with constituency-level information (see Manow, 2011). Further, such forecasts would provide valuable for local candidates and party campaign strategists, as they may reduce uncertainty and give information where local effort should be concentrated.

Forecasting dozens or even hundreds of outcomes in one election is a more ambitious task than merely generating expected national-level vote shares for a few parties or candidates. To illustrate some of the involved challenges, Figure 4.1 provides an insight into the multitude of constituency-level election results for the two grand parties blocs (the Conservatives, CDU/CSU, and the Social Democrats, SPD) in Germany over the last six elections. There is

Figure 4.1: Election results on district level at German Bundestag elections for the parties CDU/CSU and SPD. The red line displays the mean share of the party's first vote over all districts; the black lines display constituency-level first vote shares.



considerable variance around the overall mean of results on the constituency level. Moreover, district deviations from the mean are not stable over time. Lines running 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 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, as, for example, candidates' campaigning skills, scandals, district history, specific constituency-level politics—in fact, the very factors which are also included in standard national-level and theory-driven forecasting models, but the effects of which might vary over contexts. As I will argue in the following section, existing forecasting models tend to ignore constituency-specific campaign data at the cost of precision. Those who try to implement current local information, on the other hand, often face problems of data scarcity.

The main challenge in disaggregated forecasting lies in collecting data that are informative at the constituency level, as this helps explain and anticipate the heterogeneity of results. In many of the current forecasting approaches, up-to-date public opinion data provide an important predictor. However, as political polls are usually conducted at the national level, information for single districts is sparse, and a considerable amount of districts might not even be covered by the survey sample at all. Further, in order to correct for population changes in districts' territories, electoral districts' boundaries are often subject to redistricting. This makes it difficult to inform models with district history predictors like the former winner, previous vote shares, or times a party has taken the district over a certain period of elections. These factors boil down to a fundamental dilemma of constituency-level election forecasting: the scarcity of information at this very level on the one hand, and a multitude of forecasts to be made on the other.

4.3 State of the art

The majority of established forecasting models that target the outcome of parliamentary elections can be broadly divided into five 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. They have been developed in the context of elections for the US House (for recent applications see, e.g., Abramowitz, 2010; Campbell, 2010; Lewis-Beck and Tien, 2010, 2012; Klarner, 2012), the British House of Commons (e.g., Lebo and Norpoth, 2011; Lewis-Beck, Nadeau and Bélanger, 2011; Whiteley et al., 2011), the German parliament (e.g., Gschwend and Norpoth, 2001, 2005; Jérôme, Jérôme-Speziari and Lewis-Beck, 2013; Kayser and Leininger, 2013) and other settings, and often merely build on a sparse set of economic and public opinion-based predictors which are shown to be strongly correlated with the party vote or seat shares of interest. While, given their sparseness, these

[2] In this review, I do not discuss poll-aggregating approaches which recently have been developed for U.S. Senate races (e.g., Linzer, 2013; Sides, Highton and McGhee, 2014; Silver, 2014), as they pose an exception in terms of data availability and electoral rules, and can hardly be translated into parliamentary settings where constituency-level outcomes are of relevant quantities. They work on the state level and can draw on large amounts of polling data. Usually based on a firm Bayesian model framework and providing dynamic forecasts by design, such models are built to combine historical as well as polling data and give more weight to the latter as the election date approaches and more and more polls accumulate over the course of the campaign.

models tend to produce fairly accurate forecasts, they come with a considerable amount of uncertainty and have been criticized for their weak data basis and tendency for overfitting (Gelman, 1993; van der Eijk, 2005). Moreover, lacking any local component, they do not inform about constituency-level campaign dynamics.

A second branch of models relies on information from national opinion polls. Instead of merely taking published vote intentions as a forecast, 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 flavor have been developed, among others, for Westminster elections (e.g., Fisher et al., 2011; Fisher, 2015), the Australian federal election (Jackman, 2005), and the German federal election (Selb and Munzert, 2015). Poll-based approaches exploit data that are an almost natural by-product of national-level election campaigns. In contrast to structural approaches, they often allow dynamic forecasts of public opinion.[3] On the other hand, such models have to operate on the level of the polls, which is, more often than not, the national level. Furthermore, they are built to capitalize on the existence of systematic errors of trial-heat polls. The existence of such errors sometimes is an overly optimistic assumption (see Selb and Munzert, 2015). Depending on the context of elections, some of these models incorporate algorithms which take the seat-vote bias into account. This can be done by applying 'cube rule' (see Whiteley, 2005) or by predicting the bias with previous election data (see, e.g., Whiteley, 2005; Lebo and Norpoth, 2007; Nadeau, Lewis-Beck and Bélanger, 2009; Lebo and Norpoth, 2011; Whiteley et al., 2011). However, such approaches are hardly robust against 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. Essentially, these models 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 election results at the constituency level. As the assumption is that the change in vote shares is constant over all

[3] See, however, Gelman and King (1993) for a painstaking exploration of the volatility of pre-election polls and a discussion on why short-term variations in public opinion are rather endogenous to the campaign and do not provide crucial information for forecasting efforts.

constituencies, the underlying mechanism is often referred to as 'Uniform Swing' or 'Uniform National Swing' (e.g., Miller, 1972; Tufte, 1973; Johnston and Hay, 1982; Butler and Beek, 1990; 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,[4] it rests on the critical assumption that swings are distributed equally over constituencies or are likely to 'cancel out' (Butler and Beek, 1990, 179). More sophisticated variants therefore introduce regional and tactical swing parameters or add information about incumbency status (Bafumi, Erikson and Wlezien, 2008, 2010). Ultimately, this approach generates constituency-level forecasts but does not incorporate any campaign information at this level, so the accuracy of this methodology essentially hinges on the forecast of the national trend.

More recent approaches turn to new data sources which 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 the crowds' effect. The idea is that aggregated group forecasts 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 at Westminster elections. For the U.S. house elections, Sides, Highton and McGhee (2014) incorporate fundraising data in the pre-primary model and substitute 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 overseen in models that rely on overall trend measures. However, they still rely on exotic survey instruments or other data that are likely to be not

[4] 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.

available in many scenarios (as reported by Murr, 2011), or have been shown to add no predictive power compared to traditional approaches (see Wall, Sudulich and Cunningham, 2012).

Depending on the context, there are models that perform reasonably well in the aggregate but miss to generate reliable 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.

A very recent constituency- and national-level forecasting effort has been presented by Hanretty, Lauderdale and Vivyan (Forthcoming) who suggest an integrated approach of combining national and constituency polls, historical election results and census data to forecast the 2015 Westminster election.[5] Their approach is customized for the British case, takes into account the compositional nature of multiparty election outcomes, and provides local as well as national, dynamic, and probability-based forecasts. As I will set forth in the following section, my suggested framework is similar in terms of goals–combining various sources of information at both the national and the constituency level–, but different regarding the technical procedure.

4.4 A correction-combination procedure

So, how to improve constituency-level forecasts? As I discussed above, the main challenge 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. I suggest two strategies to alleviate this problem: The first is model correction that is based on identifying bias in previous forecasts, the second is model combination. My approach does not start with the development of a singular forecast model, as data availability may vary by setting. Instead, it builds upon existing sets of constituency-level forecasts and offers a framework to improve their predictive performance. The basic intuition behind my approach is that while many existing forecasting models exploit past information, e.g., by incorporating previous constituency-level elec-

[5] See also *http://www.electionforecast.co.uk/* (accessed Monday 10$^{\text{th}}$ August, 2015).

tion 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 accumulate a vast number of forecasts from election to election, this holds a great source to further improve forecasting performance. Similar strategies to exploit forecast residuals to obtain better forecasts were suggested earlier (Wallace and Hussain, 1969; Issler and Lima, 2009) and have been used to improve forecasts of, e.g., weather conditions (Glahn and Lowry, 1972; Bao et al., 2010), daily bank transaction volumes (Mabert, 1978), and commodity prices (Issler, Rodrigues and Burjack, 2014). However, existing strategies use to capitalize on expansive time series data. 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 to borrow information both over time and constituencies.

Given 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.[6] As historical constituency-level outcomes are known, the performance of forecasts at 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 forecast itself weighted with β_p (both allowed to vary by party), a vector of further covariates $\mathbf{X}$ (with coefficient γ_p also allowed to vary by party) and party-district random effects (ξ_{pj}). Formally, I assume y_{pjk} to follow a normal distribution with mean μ_{pjk} and variance σ^2,

$$y_{pjk} \sim N(\mu_{pjk}, \sigma^2), \tag{4.1}$$

with

$$\mu_{pjk} = \alpha_{p[jk]} + \beta^{\text{forecast}}_{p[jk]} \mathrm{f}_{pjk} + \gamma_{p[jk]} \mathbf{X} + \xi_{pj}. \tag{4.2}$$

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

The core idea behind this model is that raw forecasts can carry election-invariant, systematic bias or serial correlation of constituency-specific forecasts. The error can be party- and/or district-specific and furthermore be correlated with omitted variables. By decomposing past outcomes into parts explained by the forecasting model and other systematic and random components, we can identify bias and correct for it, 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 $\alpha = 0$, $\beta^{\text{forecast}} = 1$, $\gamma = 0$ for all p, and $\xi = 0$ for all p and j. The proposed link function in Equation 4.2 could be further expanded, e.g., by introducing additional regressors which are motivated by the forecast scenario and method. For instance, a model that is ignorant towards local dynamics could be evaluated using information on the number of candidates which run in a district, their incumbency status or other campaign information, if available. If constituencies are frequently subject to politically motivated redistricting, a variable whether redistricting occurred in a district or not could added in this evaluation step. Further, one could include party-election-specific errors to absorb over- or underestimation of party vote shares at a specific election. However, election-specific errors cannot be identified for a true forecast and are therefore disregarded here. In its suggested specification, the model carries a linear interpolation of the original forecasts 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 superior or worse in specific districts relative to the national or regional trend.

Next, the estimated coefficients and error components can be used to correct forecasts made in a new setting at election k^*. One could do this separately by first estimating Equation 4.2 and correcting the set of raw forecasts f_{pjk^*} in a second step.[7] However, in a Bayesian estimation framework (which I follow in the application), a more natural way to correct an out-of-sample

[7] To do so, one would proceed as follows:

$$y^*_{pjk^*} = \hat{\alpha}_{p[jk*]} + \hat{\beta}_{p[jk*]}\mathrm{f}_{pjk*} + \hat{\beta}_{p[jk]}X + \hat{\xi}_{pj} \tag{4.3}$$

forecast 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 returns imputations from the posterior predictive density, conditional on the other parameter draws (Gelman and Hill, 2007, p.367). The advantage of this integrated imputation approach is that it generates a natural quantification of uncertainty through the 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.

This general way of evaluating forecasting performance and simultaneously correcting for historical bias enables forecasters to exploit information delivered by one 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 amount of benefit from this step to be related with 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 use. However, if one prefers to construct 'naïve' models in the first step or to borrow forecasts from others—probably even without full knowledge about the original model specification or if qualitative forecasts are used—, the evaluation overhead allows to improve their accuracy, 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 own constituency-level forecasting model.

While the first step allows incorporating further information into single models ex post to exploit forecast error patterns in the original model, the second is to pool multiple forecasts from different models. As has been discussed above, existing approaches of constituency-level election forecasting come with 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 exclusively focusing on past election results or other economic or political indicators naturally fail at predicting the success of new parties by construction. On the other hand, dynamics in public opinion make vote intention surveys a shaky prediction instrument (see Gelman and King, 1993). At the district level, it is not clear a priori if there are election-, or population- or data-specific conditions under which one of the approaches has a clear advantage. One could hypothesize that survey-based approaches are inferior when the survey data at hand are extraordinarily sparse or of bad quality. At the same time, in settings where 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, survey data should be expected to perform better, as they do not rely on previous (and less useful) information.

As has been demonstrated frequently, combining several forecasts helps reduce forecast error because more information is exploited and bias which is immanent in single models can be canceled out in the aggregate forecast (e.g., Bates and Granger, 1969; Armstrong, 2001; Clemen, 1989; Raftery et al., 2005; Graefe, 2014). The strategy to combine several forecasts has already be applied to forecast U.S. Presidential elections: Montgomery, Hollenbach and Ward (2012*a*,*b*) promote the use of Ensemble Bayesian Model Averaging which provides a weighting algorithm based on each components' past performance and uniqueness. Similarly, Graefe et al. (2014) pool forecasts within and between methods, but use equal weights.

The combination procedure I employ here is based on past performance of the corrected forecasts, measured at the constituency level. Suppose one has two different constituency-level forecasting models, the first relying mainly on historical information, producing a set of forecasts $\mathrm{f}^{\text{hist}}_{jk*}$, the other exploiting current polling information, leading to another set of forecasts $\mathrm{f}^{\text{polls}}_{jk*}$. As described above, these forecasts can be improved following the correction procedure, leading to $\mathrm{f}^{\text{hist}^*}_{jk*}$ and $\mathrm{f}^{\text{polls}^*}_{jk*}$. A naïve solution to combine both would be to employ equal weights and just take the mean of forecast values for each district. This, however, would disregard that models can perform differently in general or produce forecasts which, on average, fit well in some districts but not so well in others, even after correction. Therefore, I employ a marginally more sophisticated procedure: First, I estimate party-district-specific forecast variance, that is mean-square error $\sigma^2_{\cdot,pj}$, for every set of corrected forecasts.[8] This is a measure of how well

[8] The weights have to be party-district-specific because the models are run separately for each party.

the corrected forecast model performed in past elections. As it is plausible that one model performs better in a subset of districts and worse in another, forecast variance is estimated at the district level and for every party. In the two-forecasts case with $\mathrm{f}_{pjk^*}^{\mathrm{hist}^*}$ and $\mathrm{f}_{pjk^*}^{\mathrm{polls}^*}$, weighting proceeds as follows:

$$\mathrm{f}_{pjk^*}^{\mathrm{comb}} = \omega_{pj}\mathrm{f}_{pjk^*}^{\mathrm{hist}^*} + (1 - \omega_{pj})\mathrm{f}_{pjk^*}^{\mathrm{polls}^*}, \tag{4.4}$$

with

$$\omega_{pj} = \frac{\hat{\sigma}^2_{\mathrm{polls},pj}}{\hat{\sigma}^2_{\mathrm{polls},pj} + \hat{\sigma}^2_{\mathrm{hist},pj}} \tag{4.5}$$

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 for one specific district relative to another model, the more weight is attached to this more reliable model's forecast.

4.5 Application: forecast of the German Bundestag Election 2013

In the following, I present an out-of-sample forecast of outcomes of district races at the German Bundestag election 2013 that was produced two weeks before Election Day (September 22nd, 2013). It provides a setting with a considerable number of electoral districts (299) and regional heterogeneity in party support. Six parliamentary parties entered the race and, according to some published polls, two emerging parties—the German Pirate Party (*Piratenpartei*) and the *Alternative für Deutschland, AfD* (Alternative for Germany)— had realistic chances to enter parliament. In the German electoral system on federal level, people have two votes. The first is a personal vote for a candidate in the voter's district (candidates are elected via first-past-the-post on this level), the second is a party vote which essentially determines the distribution of seats in parliament (Saalfeld, 2008).

In this application, I focus on a forecast 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 the more challenging exercise. Further, forecasting first vote shares and district winners is the more relevant application for constituency-level forecasts because they directly affect the internal composition of parliamentary groups in parliament. In advance of the election, no published efforts of constituency-level election forecasting in Germany existed.[9] Therefore, I employ two different forecasting models: The first mainly relies on historical constituency-level data and a poll-driven national-level trend estimate, the second draws on a large individual-level polling database that is mapped on the constituency level.

4.5.1 Exploiting past election results

In established democracies, past political outcomes are usually a good predictor for future outcomes. In the German case, we can observe a strong relationship between first vote shares from one election to the other, especially for the two large parties CDU/CSU and SPD (see also Figure C.4.1 in the Supplementary Materials). Differences between current and past election outcomes can be explained with nation-wide trends as well as local factors like incumbent performance, the pool of local candidates, and other local or regional determinants.

I employ a very ascetic uniform swing approach to exploit past election results. Specifically, I draw on constituency-level first vote shares in the past election and add a uniform swing constant for every candidate (i.e. party) which mirrors the national trend of the respective party from one election to the other. As discussed above, the attractiveness of this approach lies in its simplicity. On the other hand, it rests on some heroic assumptions, one being that there is a stable set of parties over time, another presuming no regional variation in shifts. Besides, potential bottom or ceiling effects for constituency-level vote shares are not taken into account. Further, it is assumed that every district has a preceding district. However, districts are sometimes abandoned or newly created, and at most elections at least some of

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

the constituencies are subject to redistricting. In Appendix C.1, I present a way to construct a district panel data set which is employed in the forecast.

In order to identify the swing component in the model, I need 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 to avoid disturbances induced by polling data. For the current election, the trend is estimated by drawing on an existing forecast of the national-level outcome (Selb and Munzert, 2015).[10] Next, the constituency-level forecasts are computed as 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 4.1 shows that the uniform swing model performs generally well at past elections since 1994 (see also Figure C.4.2 in the Appendix for a graphical display of the model's performance). This reflects in rather small mean absolute errors (absolute difference between actual first vote share and predicted first vote share) which mostly do not exceed 3 percentage points. On average, 90% of the winners of the district are forecast correctly.

Next, I employ the suggested procedure of identifying and correcting forecast bias. Following Equations 4.2 and 4.3, 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 runs in the district for the first time.[11] I suspect that the *party dominance* dummy captures parts of the underestimates of the first vote shares which should occur when, e.g., party strongholds are less affected by national losses. 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 likely an underestimation of the actual outcome which is technically counterbalanced by a party-specific estimate of β^{pio}. I adopt

[10] 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%. 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%

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

Table 4.1: Predictive performance of the uniform swing model, uncorrected and corrected forecasts. The first five columns report mean absolute errors over all 299 districts at each election. The last column reports the percentage of correctly forecast districts (predicted winner equals actual winner). Cells where the corrected forecast outperforms the uncorrected forecast are highlighted in grey.

	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	0.020	90.6
1998	0.025	0.017	0.007	0.010	0.008	88.6
2002	0.031	0.028	0.009	0.011	0.013	87.6
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

MCMC methods for inference and use vague prior distributions for the model parameters. To run the simulations, I use WinBUGS (Lunn et al., 2000). The code of the sampler is reported in Appendix C.3.

The results are presented in Table 4.2. The near one-to-one relationship between true and forecast values mirrors in the estimates for 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 fully captured by the uniform swing model. The estimated effects for party dominance are negligible.[12] The pioneer predictor is very important to correct forecasts of SPD candidates in constituencies where the party did not run in the previous election (which is, however, a historically negligible scenario). The estimated party-constituency-level variance is zero, which indicates that the forecasts are not substantively distorted by other unaccounted district-level factors.

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

Table 4.2: Bayesian median estimates and 95% credible estimates for the model of party first vote shares, based on uniform swing model (see Equation 4.2).

Predictor		95% CI
Intercept α		
CDU/CSU	0.031	[0.025;0.038]
SPD	0.021	[0.016;0.025]
FDP	0.010	[0.007;0.013]
B'90/Die Grünen	−0.001	[-0.004;0.002]
Die Linke	0.001	[-0.001;0.004]
Uniform swing estimate β^{swing}		
CDU/CSU	0.918	[0.899;0.936]
SPD	0.948	[0.934;0.961]
FDP	0.806	[0.762;0.849]
B'90/Die Grünen	1.000	[0.959;1.042]
Die Linke	1.015	[0.996;1.033]
Party dominance β^{dom}		
CDU/CSU	0.005	[0.002;0.009]
SPD	−0.004	[-0.008;-0.001]
FDP	−0.176	[-0.228;-0.122]
B'90/Die Grünen	−0.042	[-0.563;0.468]
Die Linke	0.002	[-0.017;0.022]
Pioneer β^{pio}		
CDU/CSU	0.085	[-0.781;0.912]
SPD	0.304	[0.253;0.354]
FDP	0.014	[-0.011;0.040]
B'90/Die Grünen	0.026	[0.020;0.032]
Die Linke	−0.007	[-0.010;-0.004]
Party-constituency-level variance σ^2_ξ	0.000	[0.000;0.001]
Residual variance σ^2_η	0.026	[0.025;0.026]
N	8.970	

Using the corrected results for prediction in order to assess the performance of the unbiased model, it turns out that the corrections tend to improve the model's performance, although to a very modest extent. This mirrors in the lower part of Table 4.1. In a majority of cases, the mean absolute errors over all districts for each party at the previous five elections decrease, and in three of five elections, the overall share of correctly forecast districts increases marginally, with an average amount of about 90%. These results imply that the uncorrected uniform swing model is not severely affected by systematic bias that persists over time. This is probably not too surprising, as the model essentially builds on past information—while it obvious that a national trend projection leads to under- or overestimates of district-specific party vote shares, it is less likely that this procedure induces strong, time-persistent and party-specific bias.

The estimates are nevertheless 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 by Selb and Munzert (2015). According to this estimate, the uncorrected as well as the corrected forecasts suggest a distribution of the direct seats as shown in the first row of Table 4.5. The district-level forecasts are listed in Table C.4.2 in the Supplementary Materials.

4.5.2 Exploiting individual-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 towards local campaign dynamics as long as they are not explicitly incorporated in the correction procedure. Polls on voting intention can serve as a remedy to this problem. District-level polls 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 every working day and asks them, among other things, about their vote intention for the next general election. This allows me to construct a forecasting model which, in contrast to the uniform swing model, incorporates local information.[14]

In order to stabilize the constituency-level forecasts, I employ a Bayesian modeling strategy which has been 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 the 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)

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

[14] Note that alternative data sources like local betting markets or vote expectation surveys which target the local level were not available.

and two constituency-level random effects, one of which is assumed to vary independent and identically across districts following a normal distribution, and another which is imposed to follow an intrinsic conditional auto-regressive (CAR) distribution (see Besag, York and 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 and Bafumi, 2004). For the estimation procedure, I pool survey data in the period of five months to one month before the election date to be able to draw a reasonable number of respondents per district for the estimation procedure.[15] Table C.4.1 in the Appendix provides summary statistics for the utilized polling data.[16]

The prediction results at the previous three elections are reported in Table 4.3. Using the uncorrected estimates, no more than 85% of the district outcomes are forecast correctly.[17] Therefore, I again apply the correction strategy and regress actual first vote shares in every district in the three elections on the poll model predictions and party-constituency random effects (see again Equation 4.2). The results are presented in Table 4.4. While the relationship between actual first vote shares and the poll model forecast is, on average, nearly one-to-one, the party-specific slopes and intercepts reveal substantive bias in the original model. Specifically, SPD and Die Linke vote shares are underestimated in constituencies where the parties' candidates performed well, whereas the opposite is true for candidates the other parties, where the model corrects the original forecasts towards the mean (slopes < 1). Further, the estimated variance of the party-district-specific errors is substantive, indicating that there are other un-

[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 curtail the model of its desired feature to capture actual local preferences.

[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 primarily used for scientific purposes sometimes come with no geographical identifier at all or locate respondents in other than 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 Appendix C.2. In short, I identify all possible districts for each respondent and randomly assign the respondent to one of them. With regards to the forecasting method that uses respondents from neighboring districts to estimate vote intentions, an exact match should not substantively improve forecasting performance compared to this simplifying approach.

[17] See also Figure C.4.3 in the 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 significantly biased. For instance, SPD vote shares tend to be under-estimated.

Table 4.3: Predictive performance of the polling model, uncorrected and corrected forecasts. The first five columns report party specific mean absolute errors over all 299 districts in each election. The last column reports the percentage of correctly forecast districts (predicted winner equals actual winner). Cells where the corrected forecast outperforms the uncorrected forecast are highlighted in grey.

	CDU/CSU	SPD	FDP	B'90/Die Grünen	Die Linke	% Overall correct
Uncorrected						
2002	0.034	0.07	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	0.021	0.024	0.008	0.007	0.015	96.0
2005	0.031	0.035	0.005	0.008	0.015	80.9
2009	0.023	0.024	0.008	0.011	0.015	92.0

Table 4.4: Bayesian estimates of the model of party first vote shares, based on polls model

Predictor		95% CI
Intercept α		
CDU/CSU	0.126	[0.115;0.139]
SPD	0.038	[0.030;0.046]
FDP	0.004	[-0.003;0.010]
B'90/Die Grünen	0.011	[0.005;0.017]
Die Linke	0.003	[-0.002;0.007]
Polls estimate β^{polls}		
CDU/CSU	0.716	[0.685;0.744]
SPD	1.127	[1.099;1.153]
FDP	0.508	[0.532;0.634]
B'90/Die Grünen	0.894	[0.813;0.968]
Die Linke	1.291	[1.237;1.345]
Party-constituency-level variance σ^2_ξ	0.024	[0.023;0.026]
Residual variance σ^2_η	0.028	[0.027;0.028]
N	5.980	

accounted factors at the district level which play a role in second vote polling to first vote share transformation.

The gain of the correction procedure for the polling model is impressive (see Table 4.3 and Figure C.4.3 in the Appendix). The fit of the forecasts improves considerably over all parties and years, both in terms of mean absolute errors and correctly predicted district winners.

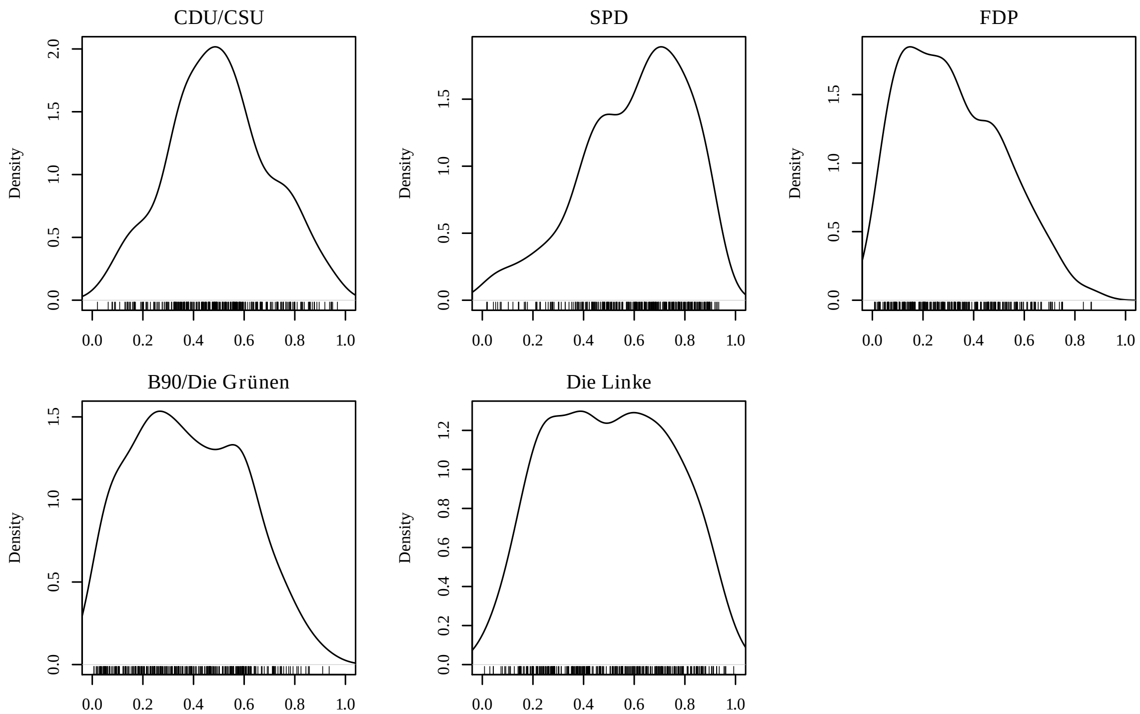
Next, the polling forecasts for the 2013 election are generated. According to the uncorrected forecast, 290 districts are attributed to the CDU/CSU and 9 to the SPD, mirroring the great advantage of the CDU/CSU in the raw polls. The corrected forecasts iron out this bias to a certain extent. Still, this forecast is significantly more favorable for the conservative parties than the corrected uniform swing model, with 261 vs. 224 seats for CDU/CSU and 34 vs. 70 seats for the SPD, respectively.

4.5.3 Forecast combination and retrospective evaluation

In the final step, the two forecasting components are combined following Equation 4.4. Therefore, I compute party-district-specific weights ω_{pj} based on the ratios of the estimated forecast variances of each model (see Equation 4.5). None of the forecast components seems to be superior for one party in general, but there is significant leverage at the constituency level. While more weight is attached to the historical model for SPD vote share forecasts, the polling model seems to have performed better in most of the districts for the FDP.

Finally, the weights are used to combine the single forecasting models and compute an aggregated forecast. This is done separately for every 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 aggregating the winners of each of the iterations and identifying the party with most wins. The aggregate results are reported in Table 4.5; the constituency-level forecasts are reported in Table C.4.2 in the Appendix. As outlined above, the simulations can also be used to derive a natural quantification of uncertainty of the estimates at both the constituency and the national level. Figure 4.3 provides probability distributions of the aggregate number of forecast seats per party, generated from all iterations. It shows that according to the model, one would expect the CDU/CSU to gain between 235 and 245 direct mandates in total, between 50 and 60 seats for the SPD, one seat of Bündnis 90/Die Grünen (which, according to the model, is a very safe seat), and three seats for Die Linke with a probability around 80%.

Figure 4.2: Distribution of forecast weight parameter (see Equation 4.5). 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.



The combined forecast provides a compromise between the uniform swing and the polling model. A comparison with the actual results reveals that the combined forecast was very close to the actual outcome. In total, 279 out of 299 districts or about 93% were forecast correctly. Note that the raw combined forecast performs equally well, too, but has a larger mean absolute error.

As an additional benchmark, I also inspected the share of correctly forecast districts in the subset of marginal seats. I defined those seats as marginal where the difference between first- and second-placed candidate was less than 10%.[18] Those cases are expected to be much more difficult to forecast. Accordingly, the share of correctly forecast districts in this subset is markedly lower for all of the models (see Table 4.5, next to last column), with the combined

[18] See also Wall, Sudulich and Cunningham (2012) for a similar procedure.

Table 4.5: Distribution of forecast winners over parties, by model

	CDU/CSU	SPD	FDP	B'90/Die Grünen	Die Linke	% correct	% correct, marg. dist.	MAE
Uniform swing, raw	224	70	0	1	4	92.3	78.8	0.031
Polling, raw	290	9	0	0	0	81.6	55.6	0.046
Combined, raw	229	65	0	1	4	93.3	80.8	0.028
Uniform swing	224	70	0	1	4	91.6	76.8	0.029
Polling	261	34	0	1	3	90.6	71.7	0.029
Combined	239	56	0	1	3	93.3	80.1	0.025
election.de (14.09.13)	224	69	0	1	5	?	?	?
spiegel.de (21.09.13)	181	89	0	3	13	?	?	?
Actual result	236	58	0	1	4			

uncorrected model performing best (81% of the cases correct) and the raw polling model worst (56% correct).

Disentangling actual versus predicted winners gives a more detailed picture of the forecast performance for the corrected and combined model (see Table 4.6). It shows that eleven seats won by the SPD were wrongly attributed to the CDU/CSU, whereas eight out of nine seats wrongly attributed to the SPD were actually won by a candidate running on a CDU/CSU ticket. Taken together, these wrong forecasts virtually cancel each other out. All other models, including the combined uncorrected model that performs equally well in terms of overall correctly forecast seats, are more biased towards one of the larger parties.

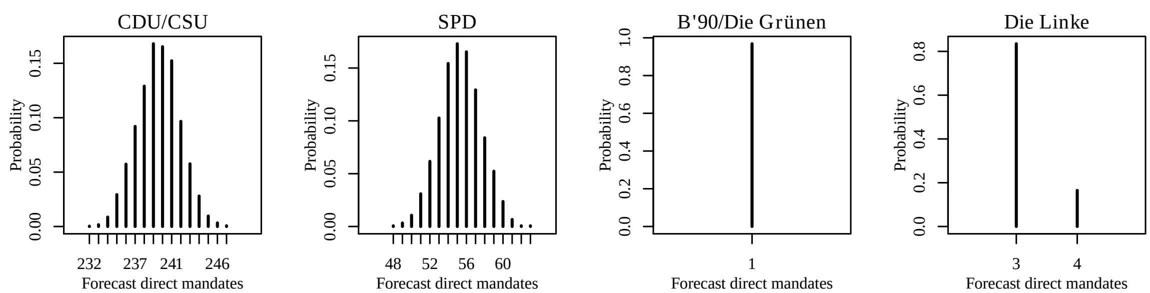
I found two other sources of constituency-level forecasts against which my forecasting models can be compared. The first is provided by the semi-commercial platform *election.de*.[19] The authors remain silent about details of their forecasting technique, but seem to implement a uniform swing-type model. Indeed, their forecasts are virtually identical with my uniform swing model (see Table 4.5) and underestimate the performance of the Conservatives. Another fore-

[19] See *http://www.election.de/cgi-bin/content.pl?url=/img/poll/btw_wp_130914.html* (accessed Monday 10th August, 2015).

Table 4.6: Actual (rows) vs. predicted (columns) winners, corrected and combined model

	CDU/ CSU	SPD	FDP	B'90/Die Grünen	Die Linke
CDU/CSU	228	8	0	0	0
SPD	11	47	0	0	0
FDP	0	0	0	0	0
B'90/Die Grünen	0	0	0	1	0
Die Linke	0	1	0	0	3

Figure 4.3: Distribution of forecast seats per party, corrected and combined model



cast approach was provided by the online newspaper site *spiegel.de*, which aggregated user expectations about the outcome in their home districts.[20] Apparently, the sample of users was strongly biased towards the left parties, as the aggregated forecast significantly underestimates the Conservatives' performance. This again indicates that individual models that mainly rest upon one source of information to inform 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 significantly biased. 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, following my suggested procedure.

What can be learned from the districts where the combined forecast failed? Table 4.7 lists the erroneous districts. By and large, the wrongly predicted cases were districts where the

[20] The original website *wahlwette.spiegel.de* has already been taken offline; a summary on the last forecasts which slightly depart from the reported figures can be accessed at *http://www.spiegel.de/politik/deutschland/bundestagswahl-wahlwette-von-spiegel-online-a-923650.html* (accessed Monday 10th August, 2015).

Table 4.7: Wrongly predicted districts (based on combination forecast of 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

winning margin, i.e. the difference between the winner's and the second runner's vote share, was very small – less than 4 percentage points in 75% of the cases. The more interesting cases are at the bottom of the table. I argued above that it is extraordinarily difficult to incorporate district-election specific information into the model. The polling model was built for this purpose, but is partly blurred by the smoothing mechanism of the model itself. In case of the Freiburg district, it was known in advance that the left parties SPD and Bündnis 90/Die Grünen both nominated popular candidates, thereby mutually taking off votes. Regarding the München-Nord district, the incumbent of the CSU managed to retain the seat against a newcomer SPD candidate. Indeed, none of the models considered incumbency status or integrated the popularity of candidates. One lesson to take away from these erroneously predicted cases is that it can be worth to collect and integrate more campaign information into the models. However, this often is a costly endeavor and, therefore, rarely done.[21]

4.6 Conclusion

Electoral research that takes the role of institutions seriously should—both in terms of theory and analysis—address the level where votes are transformed into seats. In many electoral systems, this is not (only) the national level but (also) the level of electoral districts. Consequently, efforts to forecast elections in such systems should be aware of local outcomes. In this paper, I proposed a correction-combination strategy to forecasting electoral outcomes at the constituency level. The approach accounts for bias in forecast time series that may occur because the original forecast model leaves important predictors unaccounted. Further, constituency-specific weighting of several forecasting components provides an easy way to exploit information from several sources. I demonstrated 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 the combination component contributed to an improvement of the forecast in terms of mean absolute error and share of correctly predicted outcomes.

To be sure, I do not claim that the correction-combination procedure involves any magic that leads to precise forecasts every time. The disaggregated evaluation of the steps has shown that not all models are subject to time-invariant bias that can be exploited (see the uniform swing model in the application), and that the success of combination procedure rests on the assumption that errors of individual models cancel each other out in the aggregate. This happened to be the case in the application, where the uniform swing model underestimated the

[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 expectation on the outcome in the home and neighboring constituencies. However, the response rate was underwhelming (less than 150 out of 1990 invited journalists took part), which kept me from integrating these data into the forecasting model.

conservatives' success while the polling model overestimated it. If several models err in the same direction or rely on similar information, the combination step can be of little use.[22]

The contribution of this paper to the current state of election forecasting is threefold. First, it offers an easy-to-implement approach to unbias any type of constituency-level forecasting model with a certain historical record. The procedure allows incorporating additional substantive predictors that are not part of the original model, but also works by merely identifying district-party specific bias in historical forecasts. Even if forecasters are not interested in modifying their model or combining it with other information or models, this can be a useful procedure to test whether a model produces locally or party-specific biased forecasts. Secondly, it demonstrates the benefit of model combination at the constituency level. Previous applications of combining several models for election forecasting purposes used to operate at the national level. Using party-district-specific weights, I offer a flexible way of exploiting past models' performance to generate a new, synthetic forecast. Finally, it offers an alternative strategy to approaches that rest upon massive amounts of data (as suggested by Silver (2013) in the opening quote), 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 one single forecasting model (e.g., Lock and Gelman, 2010; Jackman, 2012; Linzer, 2013; Hanretty, Lauderdale and Vivyan, Forthcoming).

The presented application provided a first test case for the method. There are other, potentially more relevant applications where 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 under such conditions, knowledge about the constituency-level distribution of voting preferences is of great value for estimating the distribution of seats in the parliament. Furthermore, constituency-level forecasts provide an opportunity for evaluating reasons of forecasting performance of different approaches (e.g., model-based vs. survey-based forecasts). In general, analyses of forecast bias are rare in election forecasting but could

[22] A fruitful extension of the approach could be to apply more sophisticated model averaging procedures like Ensemble Bayesian Model Averaging (Raftery et al., 2005; Montgomery, Hollenbach and Ward, 2012*a*) to take the uniqueness of individual model's forecasts into account.

move forward the discussion about strengths and weaknesses of the various approaches on the market.

Appendix A

A.1 WinBUGS code

Preliminary remark

We owe many of the computational tricks we use in this code to Bivand, Pebesma and Gómez-Rubio (2008), Gelman (2007), and Gómez-Rubio et al. (2008). The GeoBUGS module for WinBUGS Thomas et al. (2004) comes with the intrinsic Gaussian CAR prior distribution. We used the "R2WinBUGS" package to run WinBUGS from within R Sturtz, Ligges and Gelman (2005).

Model 1: empty model with unstructured random effects

```
model{
        for(i in 1:ns)
        {
                for(j in 1:n[i])
                {
y[cumn[i]+j] ~ dbern(p[cumn[i]+j])
logit(p[cumn[i]+j]) <- alpha + u[s[i]]
}
u[s[i]] ~ dnorm(0, tauu)
}

        for(i in 1:N)
        {
                mu[i] <- exp(alpha + u[i]) / (1 + exp(alpha + u[i]))
        }

tauu <- pow(sigmau, -2)
sigmau ~ dunif(0,2)
alpha ~ dflat()
}
```

Model 2: unstructured random effects plus district-level covariate log(distsize)

```
model
{
        for(i in 1:ns)
        {
                for(j in 1:n[i])
                {
y[cumn[i]+j] ~ dbern(p[cumn[i]+j])
logit(p[cumn[i]+j]) <- beta[s[i]]
}
u[s[i]] ~ dnorm(0, tauu)
}

        for(i in 1:N)
        {
beta[i] <- beta.0 + beta.distsize*log(distsize[i]) + u[i]
        }

        for(i in 1:N)
        {

                mu[i] <- exp(beta[i]) / (1 + exp(beta[i]))
        }

beta.0 ~ dnorm(0,.0001)
beta.distsize ~ dnorm(0,.0001)

tauu <- pow(sigmau, -2)
sigmau ~ dunif(0,2)

}
```

Model 3: CAR-structured random effects plus district-level covariate log(distsize^-1)

```
model
{
        for(i in 1:ns)
        {
                for(j in 1:n[i])
                {
                y[cumn[i]+j] ~ dbern(p[cumn[i]+j])
                logit(p[cumn[i]+j]) <- beta[s[i]] + v[s[i]]
                }
u[s[i]] ~ dnorm(0, tauu)
        }

        for(i in 1:N)
        {
beta[i] <- beta.0 + beta.distsize*log(distsize[i]) + u[i]
        }

        for(i in 1:N)
        {

                mu[i] <- exp(beta[i] + v[i]) / (1 + exp(beta[i] + v[i]))
        }

    v[1:N] ~ car.normal(nb[], weight[], num[], tauv)

    beta.0 ~ dnorm(0,.0001)
    beta.distsize ~ dnorm(0,.0001)

tauu <- pow(sigmau, -2)
sigmau ~ dunif(0,2)
tauv <- pow(sigmav, -2)
sigmav ~ dunif(0,2)
}
```

Model 3p: Combination with poststratification approach: CAR-structured REs plus district-level covariate log(distsize) plus individual level eduction dummies

```
model
{
        for(i in 1:ns)
        {
                for(j in 1:n[i])
                {
                y[cumn[i]+j] ~ dbern(p[cumn[i]+j])
                logit(p[cumn[i]+j]) <- beta[s[i]] + gamma2*edu2[cumn[i]+j]
                    + gamma3*edu3[cumn[i]+j]  + v[s[i]]
                }
u[s[i]] ~ dnorm(0, tauu)
        }

        for(i in 1:N)
        {
beta[i] <- beta.0 + beta.distsize*log(distsize[i]) + u[i]
        }

    v[1:N] ~ car.normal(nb[], weight[], num[], tauv)

    beta.0 ~ dflat()
    beta.distsize ~ dflat()
gamma2 ~ dflat()
gamma3 ~ dflat()

tauu <- pow(sigmau, -2)
sigmau ~ dunif(0,2)
tauv <- pow(sigmav, -2)
sigmav ~ dunif(0,2)
}
```

A.2 Tables and Figures

Figure A.2.1: Medians of the parameters' posterior probability with 90% credible interval, face-to-face data

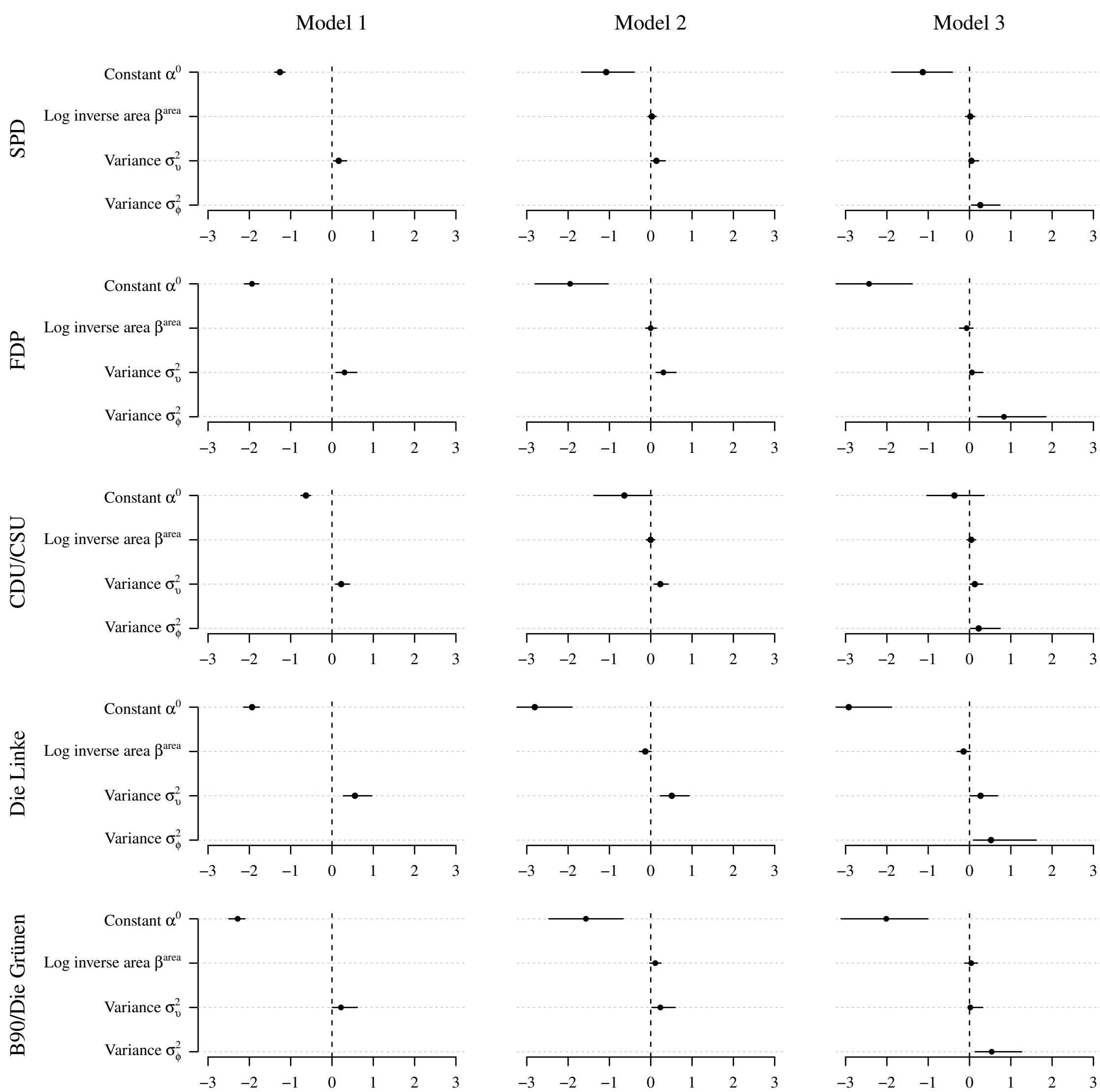


Figure A.2.2: Medians of the parameters' posterior probability with 90% credible interval, CATI data

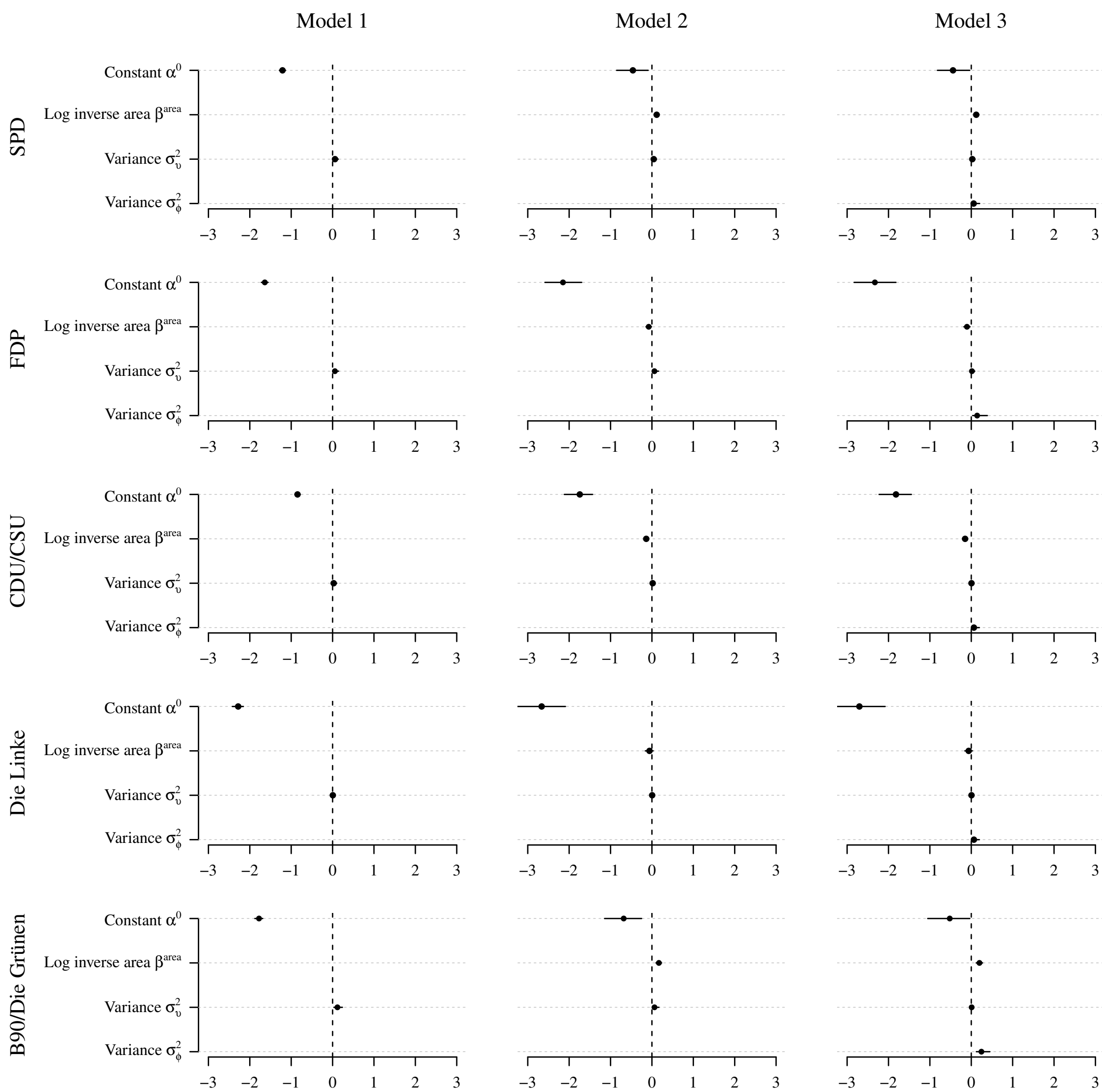


Figure A.2.3: MAE vs. number of observations in district, pooled data

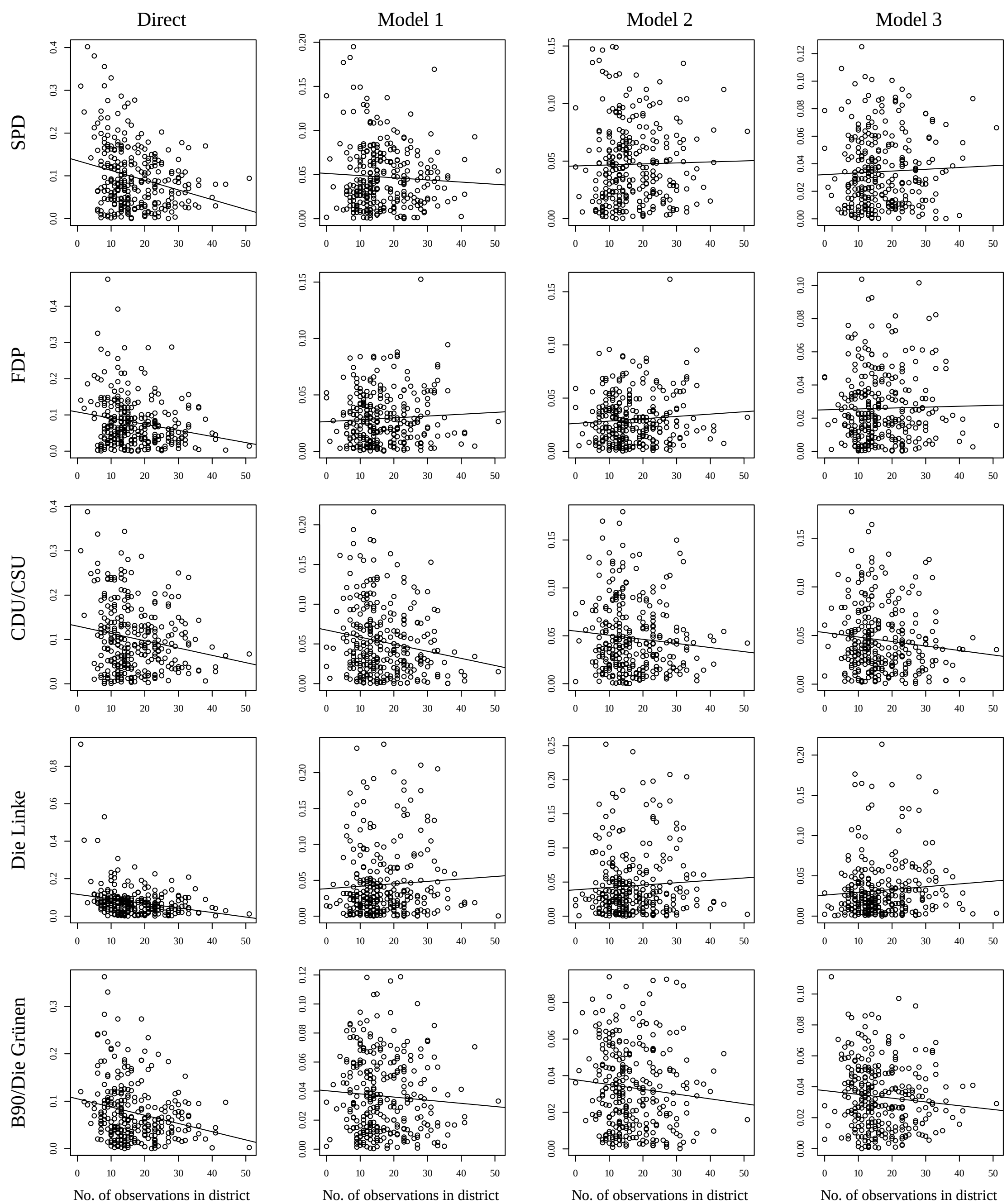


Figure A.2.4: MAE (CAR-specified model) vs. number of observations in neighbor districts

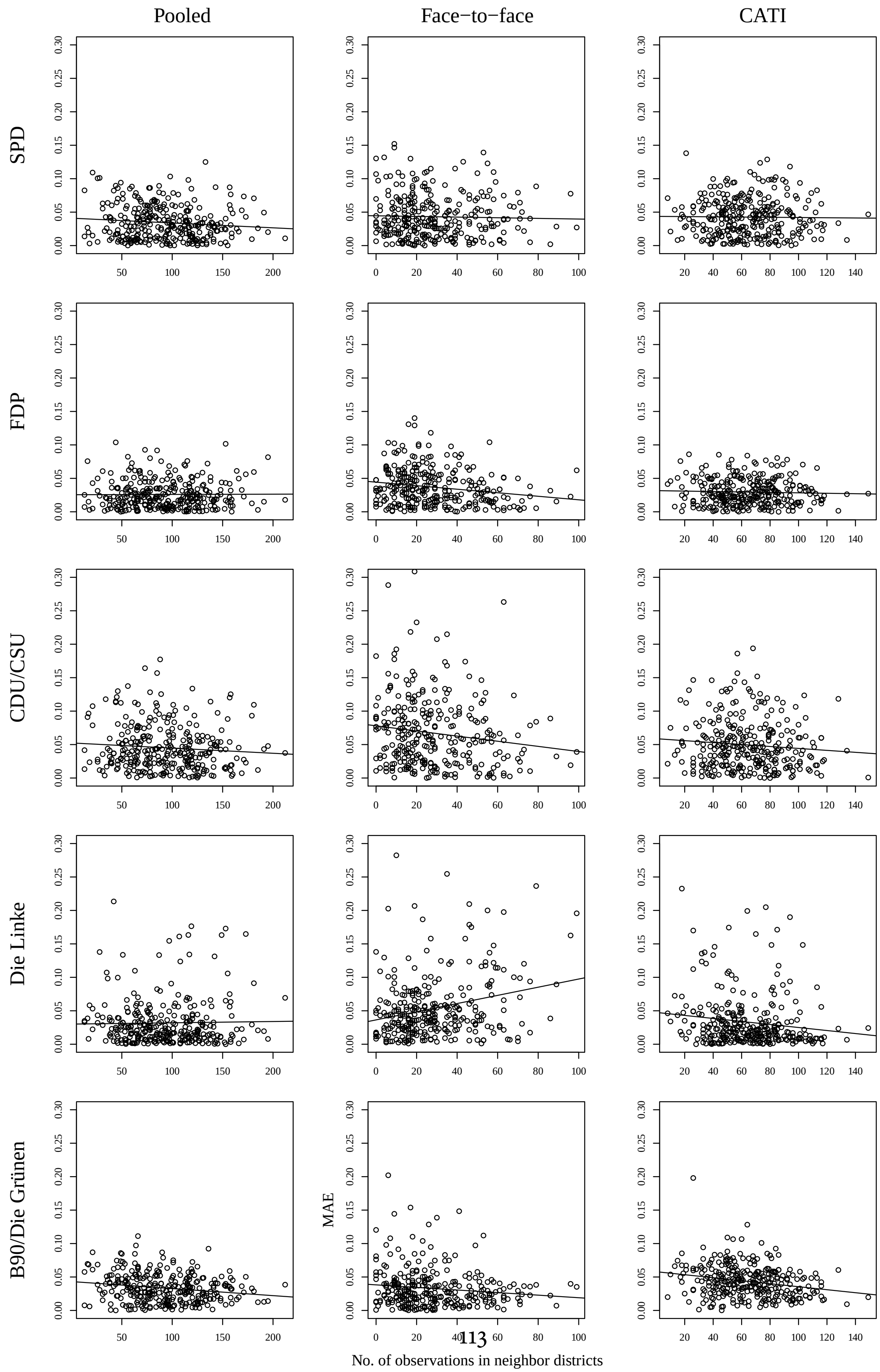
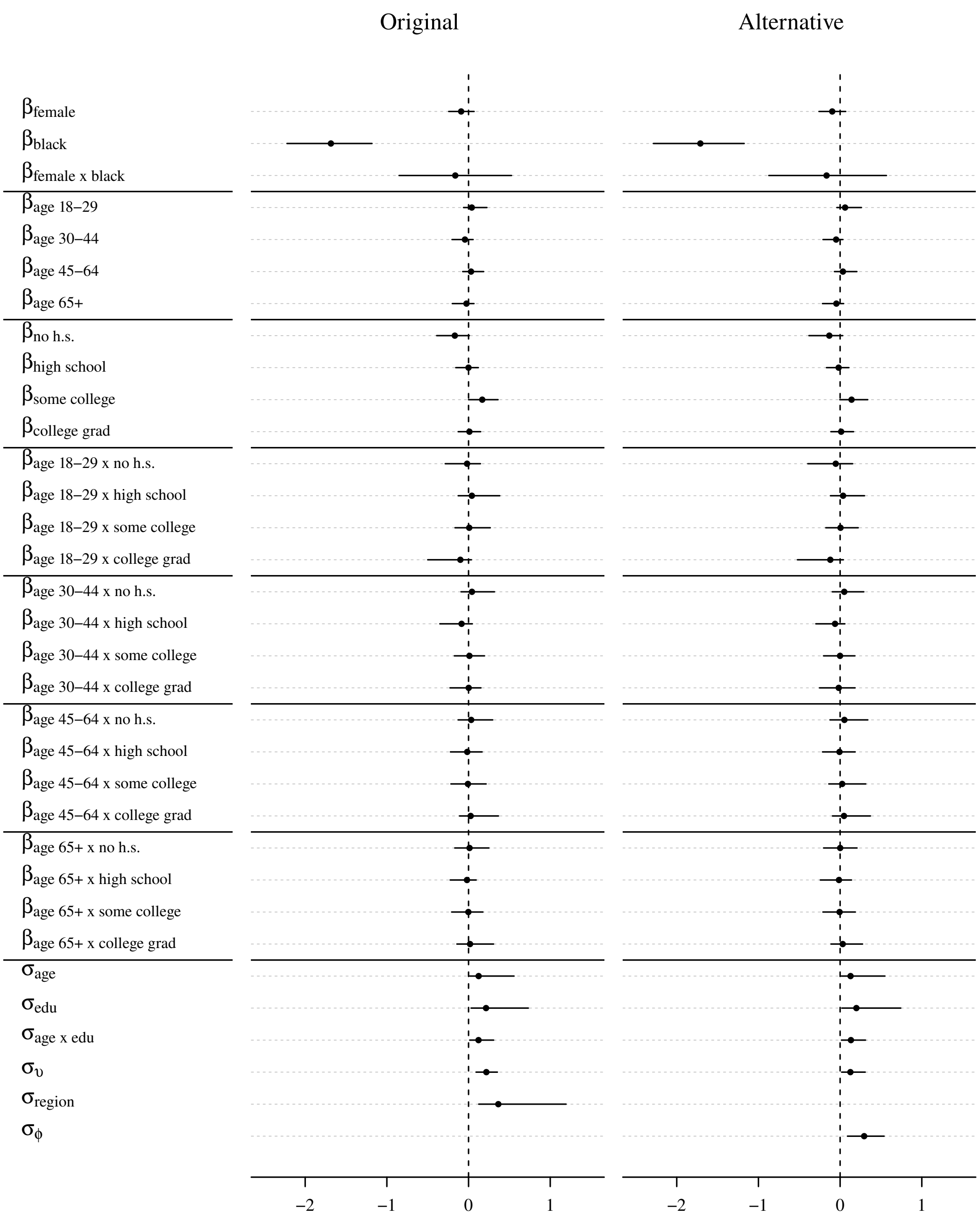


Figure A.2.5: Replication of Park, Gelman, and Bafumi (2004). Medians of the parameters' posterior probability with 90% credible interval.



Appendix B

B.1 Tables and Figures

Table B.1.1: Used variables from ALLBUS 1980-2010, separated in four issue blocks.

Variable	’80	’82	’84	’86	’88	’90	’91	’92	’94	’96	’98	’00	’02	’04	’06	’08	’10
gender.job.child.1		■					■	■		■		■		■		■	
gender.job.child.2		■					■	■		■		■		■		■	■
gender.job.child.3		■					■	■		■		■		■		■	
gender.job.marriage		■					■	■		■		■		■		■	
gender.help.husband.1		■					■	■		■		■		■		■	■
gender.help.husband.2		■					■	■		■		■		■		■	
moral.marriage	■		■		■		■	■		■		■	■		■		■
moral.abortion		■		■		■		■		■		■			■		
moral.euthanasia						■						■	■				
moral.cannabis						■						■	■				
moral.homosexuality						■						■	■				
distribution.state.provide.welfare			■						■					■			
distribution.profits.1			■						■					■			
distribution.profits.2			■						■					■			
distribution.social.inequality			■						■					■			
distribution.income.incentive			■		■		■		■		■	■		■		■	■
distribution.rank.difference.1			■		■		■		■		■	■		■		■	■
distribution.rank.difference.2			■		■		■		■		■	■		■		■	■
immigration.asylum.seekers						■	■	■		■		■			■		
immigration.non.eu.workers						■	■	■		■		■			■		
immigration.lifestyle.adaption	■		■		■	■			■	■		■	■		■		■
immigration.no.cross.marriage	■		■		■	■			■	■		■	■		■		■
immigration.no.jobs.send.home	■		■		■	■			■	■		■	■		■		
immigration.political.rights	■		■		■	■			■	■		■	■		■		

Figure B.1.1: Density curves of estimated overall and dimension-specific correlation trends for different sets of models. Shaded areas indicate effects opposed to the models reported in the paper based on the full set of variables (indicated with a vertical red line). The small vertical black lines indicate point trend estimates of a singular model.

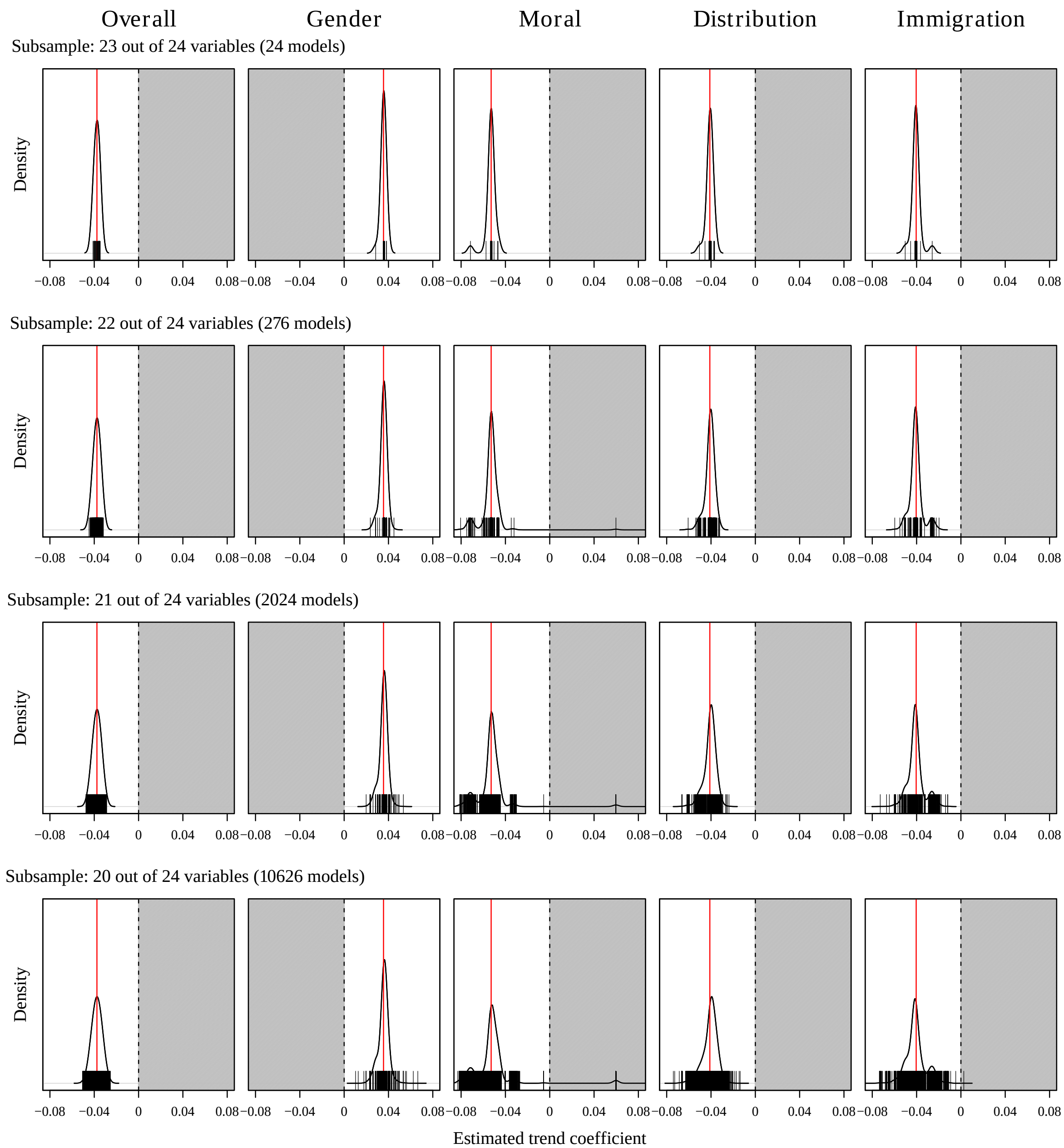


Figure B.1.2: Correlation trends by issue dimension, over time (reduced variable sample based on within-dimension sets of batteries). The lines are based on the effects as estimated in Model C ('By types of issue dimensions'). Shaded areas represent 90% confidence intervals.

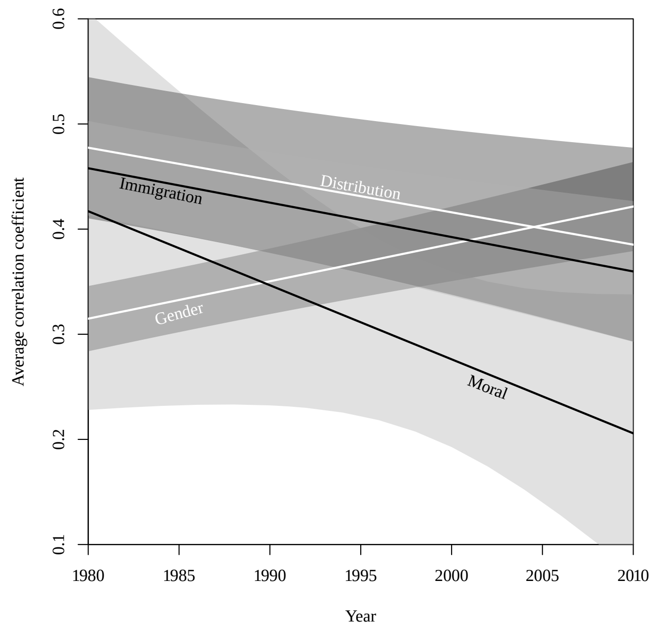
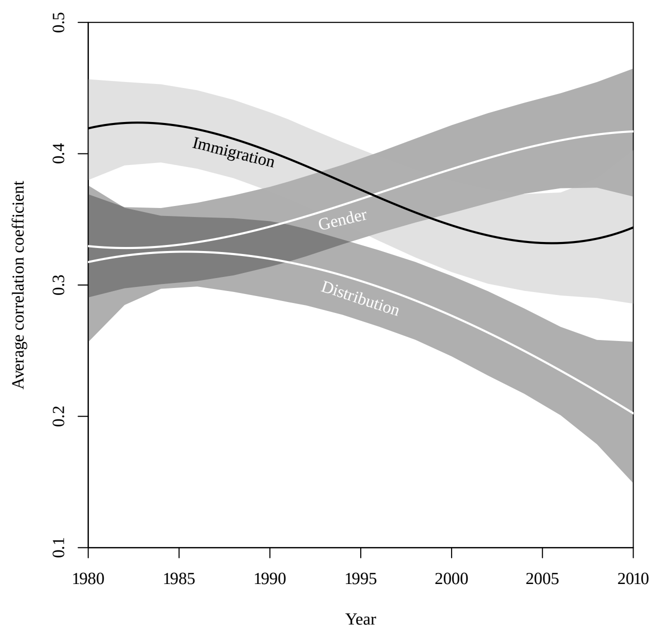


Figure B.1.3: Correlation trends by issue dimension, over time (based on cubic polynomial of trend specification). The lines are based on the effects as estimated in Model C ('By types of issue dimensions'). Shaded areas represent 90% confidence intervals.



Appendix C

C.1 Building a panel data set of electoral districts

Important determinants of constituency-level election results are historical information like, for instance, past election results or candidates' incumbency or seniority status. However, identifying such information is not a trivial task, because of redistricting (which affects the composition of the electorate), renaming and changes in the electoral system. For example, in advance of the 2002 German Bundestag election the number of districts was reduced from 328 to 299, disrupting the history of many districts.

In order to still be able to use past district information for the estimation and forecasting procedure, I constructed a panel data set of electoral districts. Therefore, I coded changes which are documented in amendments of the *Bundeswahlgesetz* (Federal Electoral Law) in which redistricting and renaming are published about a year before an election takes place. In the cases where districts where abandoned without replacement, a district's history comes to an end. Its area is usually distributed to neighboring districts which persist in the panel. Newly created districts are usually constructed from several existing districts. If all constituent districts persist as well, a new district is introduced in the panel. The constructed panel data set consists of 358 districts which, of course, have never been part of the electoral map simultaneously.

C.2 Matching respondents and electoral districts

In the best of all worlds, polls come with an identifier that matches every respondent to one and only one electoral district. However, this is not always the case. Many publicly available polls are rendered useless for constituency-level analysis as they are not delivered disaggregatedly or lack a geographical identifier. Still others are conducted for a more general purpose than predictive or even political interest, and therefore do not contain a constituency identifier but some other form of direct or indirect geographical variable. These might be dialing codes, zip codes, identifiers of other (mostly larger) political units like states, regions or provinces, or indicators for administrative units. As long as these units are not overly bigger than electoral units, that is, if the number of geographical identifiers is not much lesser than the number of constituencies, and if the geographical units do not typically contain many constituencies at once, the polls can still be considered useful for inference on constituency-level, because in this case respondents can be matched to one or only few constituencies.

The polling data available for the German showcase belong to the latter group. They are delivered with identifiers of administrative boundaries of *Kreise* and *Kreisfreie Städte*, which is the second-lowest administrative unit in Germany. Depending on the year, there are around 400 to 440 unique administrative units of that kind. A major problem for matching administrative with electoral units arises from the fact that they are not always congruent. To illustrate this disparity, Figure C.2.1 gives a view on a subset of electoral and administrative districts for the 2013 German Bundestag election. Qualitatively, there are two different cases of overlap between the geographical unit from the poll (PU) and from the election (EU) when it comes to an attribution of PU observations to EUs:[1]

1. The PU is completely contained in the EU. This is the case if the PU is congruent with the EU or if the EU contains one or more PUs entirely and some others partly. Consider the electoral district 63 in the map. The PUs "Oder-Spree" and "Frankfurt (Oder)" are entirely enclosed by the EU. In such cases poll respondents can be uniquely assigned to one EU.

2. The PU is dissected by one or several EU borders. This is the case for most PUs shown in Figure C.2.1. For instance, the PU "Teltow-Fläming" is part of the EUs 60, 61 and 62 (although the largest share of its area belongs to EU 62). To pick another example, the PU "Havelland" is part of the EUs 56 (not shown on the map), 58 and 60. In cases such as these uncertainty arises to which EU an observation from a PU should be assigned. The city of Berlin in the map is a special case of this category, and at the same time the worst case scenario in the data: This PU contains no less than 12 EUs, which makes the attribution of a respondent to one specific EU very uncertain.

[1] This is the only direction of interest in this case. If one wanted to transfer information in the other direction, that is from EUs to PUs, there were other, mirror-imaged cases to consider.

Figure C.2.1: Disparity between electoral district boundaries (grey) and administrative units (blue) for a subset of electoral districts of the 2013 German Bundestag election. The electoral districts are labeled with their respective numbers, the administrative units with their name.

In order to assign PU observations to EUs, I proceed in two steps. In the first step, possible EUs are assigned to PUs. I exploit the Federal Electoral Law which lists the administrative units of each electoral district. In the second step, one has to account for the attribution uncertainty in the estimate of EU-level aggregate statistics. There are two possible strategies here. The first is to account for this uncertainty in the assignment process and match observations from a PU to a EU with a probability of $p = \frac{area_{intersect}(EU,PU)}{area(PU)}$. The advantage of this solution is that such probabilities are easy to obtain when geographical data of both EU and PU units is available, and estimation of EU-level statistics is straightforward as there is a fixed set of respondents for every district once the assignment is done. On the downside, many different combinations of assignments are possible, and the areal share is only a proxy of population density. Another strategy would be to account for uncertainty in the estimation of EU-level quantities. Many different procedures are imaginable, such as cross-classifying observations into districts. The procedure I employed here is to assign respondents randomly to one of their assigned districts.

C.3 WinBUGS code

Uniform swing correction model (see Section 4.5.1)

```
model
{
    for (i in 1:N) {
        y[i] ~ dnorm(mu[i], tau.y)
        mu[i] <- alpha[party[i]] + b.project[party[i]] * project[i] +
            b.dominance[party[i]] * dominance[i] + b.newrun[party[i]] *
            newrun[i] + u[partyXwkr[i]]
    }
    for (j in 1:J) {
        alpha[j] ~ dnorm(alpha.hat[j], tau.alpha)
        b.project[j] ~ dnorm(b.project.hat[j], tau.b.project)
        b.dominance[j] ~ dnorm(b.dominance.hat[j], tau.b.dominance)
        b.newrun[j] ~ dnorm(b.newrun.hat[j], tau.b.newrun)
        alpha.hat[j] <- mu.alpha
        b.project.hat[j] <- mu.b.project
        b.dominance.hat[j] <- mu.b.dominance
        b.newrun.hat[j] <- mu.b.newrun
    }
    for (k in 1:K) {
        u[k] ~ dnorm(0.00000E+00, tau.u)
    }
    mu.alpha ~ dnorm(0.00000E+00, 1.00000E-04)
    mu.b.project ~ dnorm(0.00000E+00, 1.00000E-04)
    mu.b.dominance ~ dnorm(0.00000E+00, 1.00000E-04)
    mu.b.newrun ~ dnorm(0.00000E+00, 1.00000E-04)
    tau.y <- pow(sigma.y, -2)
    tau.alpha <- pow(sigma.alpha, -2)
    tau.b.project <- pow(sigma.b.project, -2)
    tau.b.dominance <- pow(sigma.b.dominance, -2)
    tau.b.newrun <- pow(sigma.b.newrun, -2)
    tau.u <- pow(sigma.u, -2)
    sigma.y ~ dunif(0.00000E+00, 100)
    sigma.alpha ~ dunif(0.00000E+00, 100)
    sigma.b.project ~ dunif(0.00000E+00, 100)
    sigma.b.dominance ~ dunif(0.00000E+00, 100)
    sigma.b.newrun ~ dunif(0.00000E+00, 100)
    sigma.u ~ dunif(0.00000E+00, 100)
}
```

Polls correction model (see Section 4.5.2)

```
model
{
    for (i in 1:N) {
        y[i] ~ dnorm(mu[i], tau.y)
        mu[i] <- alpha[party[i]] + b.polls[party[i]] * polls[i] +
            u[partyXwkr[i]]
    }
    for (j in 1:J) {
        alpha[j] ~ dnorm(alpha.hat[j], tau.alpha)
        b.polls[j] ~ dnorm(b.polls.hat[j], tau.b.polls)
        alpha.hat[j] <- mu.alpha
        b.polls.hat[j] <- mu.b.polls
    }
    for (k in 1:K) {
        u[k] ~ dnorm(0.00000E+00, tau.u)
    }
    mu.alpha ~ dnorm(0.00000E+00, 1.00000E-04)
    mu.b.polls ~ dnorm(0.00000E+00, 1.00000E-04)
    tau.y <- pow(sigma.y, -2)
    tau.alpha <- pow(sigma.alpha, -2)
    tau.b.polls <- pow(sigma.b.polls, -2)
    tau.u <- pow(sigma.u, -2)
    sigma.y ~ dunif(0.00000E+00, 100)
    sigma.alpha ~ dunif(0.00000E+00, 100)
    sigma.b.polls ~ dunif(0.00000E+00, 100)
    sigma.u ~ dunif(0.00000E+00, 100)
}
```

C.4 Tables and Figures

Table C.4.1: Summary statistics of the raw poll data utilized: numbers of respondents, N, numbers of districts covered, J, average numbers of respondents per district, $\bar{N}_j$, their standard deviations, minimum and maximum values.

	N	J	mean(N_j)	sd(N_j)	min(N_j)	max(N_j)
2002	30,627	298	103	49	5	321
2005	28,878	297	97	27	34	216
2009	28,216	297	95	29	35	233
2013	29,767	299	100	27	28	197

Table C.4.2: Overview of 2013 forecasts

NR	District name	Winner	Uniform Swing	Polls	Combined	Predicted vote shares, combined model				
						CDU/CSU	SPD	FDP	Greens	Left
1	Flensburg – Schleswi...	CDU	CDU	CDU	CDU	0.41	0.37	0.01	0.13	0.03
2	Nordfriesland – Dith...	CDU	CDU	CDU	CDU	0.45	0.32	0.02	0.11	0.03
3	Steinburg – Dithmars...	CDU	CDU	CDU	CDU	0.43	0.34	0.03	0.10	0.03
4	Rendsburg-Eckernförd...	CDU	CDU	CDU	CDU	0.42	0.37	0.01	0.12	0.02
5	Kiel	SPD	SPD	SPD	SPD	0.33	0.41	0.01	0.17	0.05
6	Plön – Neumünster	CDU	CDU	CDU	CDU	0.40	0.37	0.01	0.12	0.03
7	Pinneberg	CDU	CDU	CDU	CDU	0.43	0.36	0.01	0.12	0.03
8	Segeberg – Stormarn-...	CDU	CDU	CDU	CDU	0.42	0.34	0.02	0.12	0.03
9	Ostholstein – Storma...	CDU	CDU	CDU	CDU	0.41	0.36	0.02	0.11	0.03
10	Herzogtum Lauenburg ...	CDU	CDU	CDU	CDU	0.43	0.33	0.02	0.12	0.03
11	Lübeck	SPD	SPD	SPD	SPD	0.34	0.38	0.02	0.14	0.05
12	Schwerin – Ludwigslu...	CDU	CDU	CDU	CDU	0.37	0.23	0.02	0.07	0.18
13	Ludwigslust-Parchim ...	CDU	CDU	CDU	CDU	0.34	0.27	0.01	0.07	0.20
14	Rostock – Landkreis ...	CDU	CDU	CDU	CDU	0.31	0.23	0.01	0.11	0.22
15	Vorpommern-Rügen – V...	CDU	CDU	CDU	CDU	0.50	0.16	0.00	0.05	0.18
16	Mecklenburgische See...	CDU	CDU	CDU	CDU	0.41	0.17	0.02	0.06	0.17
17	Mecklenburgische See...	CDU	CDU	CDU	CDU	0.38	0.23	0.02	0.06	0.18
18	Hamburg-Mitte	SPD	SPD	SPD	SPD	0.31	0.37	0.01	0.18	0.08
19	Hamburg-Altona	SPD	SPD	CDU	SPD	0.33	0.37	0.01	0.17	0.06
20	Hamburg-Eimsbüttel	SPD	CDU	CDU	CDU	0.34	0.30	0.01	0.24	0.05
21	Hamburg-Nord	CDU	CDU	CDU	CDU	0.40	0.34	0.01	0.15	0.03
22	Hamburg-Wandsbek	SPD	CDU	CDU	CDU	0.39	0.38	0.00	0.12	0.05
23	Hamburg-Bergedorf – ...	SPD	SPD	SPD	SPD	0.37	0.42	0.00	0.11	0.06
24	Aurich – Emden	SPD	SPD	SPD	SPD	0.30	0.46	0.00	0.13	0.06
25	Unterems	CDU	CDU	CDU	CDU	0.48	0.33	0.02	0.08	0.04
26	Friesland – Wilhelms...	SPD	SPD	CDU	SPD	0.37	0.41	0.02	0.10	0.05
27	Oldenburg – Ammerlan...	SPD	CDU	CDU	CDU	0.39	0.35	0.02	0.16	0.05
28	Delmenhorst – Weserm...	CDU	SPD	CDU	CDU	0.39	0.37	0.02	0.11	0.05
29	Cuxhaven – Stade II	CDU	CDU	CDU	CDU	0.41	0.39	0.01	0.09	0.04
30	Stade I – Rotenburg ...	CDU	CDU	CDU	CDU	0.47	0.35	0.00	0.09	0.02
31	Mittelems	CDU	CDU	CDU	CDU	0.57	0.30	0.01	0.08	0.02
32	Cloppenburg – Vechta	CDU	CDU	CDU	CDU	0.64	0.23	0.02	0.06	0.02
33	Diepholz – Nienburg ...	CDU	CDU	CDU	CDU	0.42	0.37	0.02	0.11	0.03
34	Osterholz – Verden	CDU	SPD	CDU	CDU	0.40	0.40	0.01	0.11	0.05
35	Rotenburg I – Heidek...	CDU	CDU	CDU	CDU	0.43	0.39	0.01	0.10	0.03
36	Harburg	CDU	CDU	CDU	CDU	0.43	0.34	0.02	0.12	0.03
37	Lüchow-Dannenberg – ...	CDU	CDU	CDU	CDU	0.36	0.34	0.01	0.17	0.05
38	Osnabrück-Land	CDU	CDU	CDU	CDU	0.47	0.36	0.01	0.09	0.03
39	Stadt Osnabrück	CDU	CDU	CDU	CDU	0.41	0.37	0.02	0.13	0.03
40	Nienburg II – Schaum...	SPD	SPD	CDU	SPD	0.41	0.42	-0.01	0.10	0.02
41	Stadt Hannover I	SPD	SPD	SPD	SPD	0.36	0.38	0.01	0.12	0.04
42	Stadt Hannover II	SPD	SPD	SPD	SPD	0.34	0.39	-0.00	0.14	0.05
43	Hannover-Land I	CDU	SPD	CDU	CDU	0.41	0.39	0.00	0.10	0.02
44	Celle – Uelzen	CDU	CDU	CDU	CDU	0.45	0.36	0.00	0.10	0.03
45	Gifhorn – Peine	SPD	SPD	CDU	CDU	0.40	0.40	-0.01	0.09	0.03
46	Hameln-Pyrmont – Hol...	SPD	SPD	CDU	SPD	0.39	0.41	0.01	0.11	0.05
47	Hannover-Land II	SPD	SPD	CDU	SPD	0.39	0.40	-0.00	0.10	0.03
48	Hildesheim	CDU	SPD	CDU	SPD	0.39	0.42	-0.01	0.11	0.03
49	Salzgitter – Wolfenb...	SPD	SPD	SPD	SPD	0.36	0.42	0.00	0.09	0.04
50	Braunschweig	SPD	SPD	CDU	SPD	0.37	0.37	-0.01	0.11	0.04

Table C.4.2: Overview of 2013 forecasts

NR	District name	Winner	Uniform Swing	Polls	Combined	Predicted vote shares, combined model				
						CDU/CSU	SPD	FDP	Greens	Left
51	Helmstedt – Wolfsbur...	CDU	CDU	CDU	CDU	0.42	0.37	-0.00	0.10	0.03
52	Goslar – Northeim – ...	SPD	SPD	CDU	SPD	0.39	0.41	0.01	0.09	0.05
53	Göttingen	SPD	SPD	CDU	SPD	0.38	0.40	-0.00	0.15	0.04
54	Bremen I	SPD	SPD	SPD	SPD	0.31	0.36	0.01	0.18	0.07
55	Bremen II – Bremerha...	SPD	SPD	SPD	SPD	0.29	0.40	0.01	0.13	0.09
56	Prignitz – Ostprigni...	CDU	SPD	SPD	SPD	0.28	0.29	0.01	0.07	0.22
57	Uckermark – Barnim I	CDU	SPD	CDU	CDU	0.30	0.28	0.01	0.06	0.21
58	Oberhavel – Havellan...	CDU	SPD	CDU	CDU	0.32	0.29	0.02	0.09	0.16
59	Märkisch-Oderland – ...	CDU	Left	CDU	CDU	0.28	0.24	0.01	0.06	0.26
60	Brandenburg an der H...	SPD	SPD	SPD	SPD	0.29	0.30	0.01	0.06	0.19
61	Potsdam – Potsdam-Mi...	CDU	SPD	CDU	SPD	0.29	0.30	0.01	0.09	0.19
62	Dahme-Spreewald – Te...	CDU	SPD	CDU	CDU	0.30	0.29	0.02	0.07	0.18
63	Frankfurt (Oder) – O...	CDU	SPD	CDU	CDU	0.28	0.25	0.01	0.07	0.21
64	Cottbus – Spree-Neiß...	CDU	SPD	CDU	CDU	0.29	0.25	0.02	0.06	0.24
65	Elbe-Elster – Obersp...	CDU	CDU	CDU	CDU	0.33	0.26	0.02	0.05	0.20
66	Altmark	CDU	CDU	CDU	CDU	0.35	0.23	0.01	0.06	0.23
67	Börde – Jerichower L...	CDU	CDU	CDU	CDU	0.37	0.25	0.02	0.05	0.17
68	Harz	CDU	CDU	CDU	CDU	0.36	0.22	0.02	0.07	0.19
69	Magdeburg	CDU	CDU	CDU	CDU	0.35	0.25	0.01	0.07	0.23
70	Dessau – Wittenberg	CDU	CDU	CDU	CDU	0.39	0.20	0.02	0.06	0.18
71	Anhalt	CDU	CDU	CDU	CDU	0.36	0.23	0.02	0.05	0.19
72	Halle	CDU	CDU	CDU	CDU	0.34	0.19	0.01	0.09	0.28
73	Burgenland – Saalekr...	CDU	CDU	CDU	CDU	0.37	0.20	0.03	0.05	0.21
74	Mansfeld	CDU	CDU	CDU	CDU	0.36	0.20	0.02	0.05	0.23
75	Berlin-Mitte	SPD	SPD	SPD	SPD	0.25	0.28	0.01	0.21	0.14
76	Berlin-Pankow	Left	SPD	SPD	SPD	0.23	0.27	0.01	0.16	0.21
77	Berlin-Reinickendorf	CDU	CDU	CDU	CDU	0.41	0.30	0.02	0.12	0.04
78	Berlin-Spandau – Cha...	CDU	CDU	CDU	CDU	0.39	0.34	0.01	0.11	0.05
79	Berlin-Steglitz-Zehl...	CDU	CDU	CDU	CDU	0.41	0.31	0.02	0.14	0.02
80	Berlin-Charlottenbur...	CDU	SPD	CDU	SPD	0.34	0.34	0.03	0.18	0.03
81	Berlin-Tempelhof-Sch...	CDU	CDU	CDU	CDU	0.34	0.26	0.01	0.24	0.04
82	Berlin-Neukölln	SPD	CDU	CDU	CDU	0.34	0.29	0.01	0.15	0.08
83	Berlin-Friedrichshai...	Greens	Greens	Greens	Greens	0.17	0.19	0.01	0.39	0.14
84	Berlin-Treptow-Köpen...	Left	Left	Left	Left	0.25	0.23	0.00	0.09	0.35
85	Berlin-Marzahn-Helle...	Left	Left	Left	Left	0.23	0.19	0.01	0.07	0.31
86	Berlin-Lichtenberg	Left	Left	Left	Left	0.22	0.21	0.01	0.09	0.34
87	Aachen I	CDU	CDU	CDU	CDU	0.42	0.32	0.01	0.15	0.04
88	Aachen II	CDU	CDU	CDU	CDU	0.43	0.37	0.01	0.10	0.04
89	Heinsberg	CDU	CDU	CDU	CDU	0.53	0.28	0.01	0.09	0.03
90	Düren	CDU	CDU	CDU	CDU	0.48	0.35	-0.00	0.09	0.03
91	Rhein-Erft-Kreis I	CDU	CDU	CDU	CDU	0.42	0.38	0.02	0.09	0.02
92	Euskirchen – Rhein-E...	CDU	CDU	CDU	CDU	0.47	0.33	0.03	0.09	0.03
93	Köln I	SPD	SPD	CDU	CDU	0.37	0.37	0.01	0.13	0.04
94	Köln II	CDU	CDU	CDU	CDU	0.37	0.34	0.02	0.17	0.02
95	Köln III	SPD	SPD	SPD	SPD	0.33	0.38	0.00	0.18	0.05
96	Bonn	SPD	SPD	CDU	CDU	0.37	0.35	0.07	0.12	0.03
97	Rhein-Sieg-Kreis I	CDU	CDU	CDU	CDU	0.46	0.30	0.02	0.11	0.02
98	Rhein-Sieg-Kreis II	CDU	CDU	CDU	CDU	0.52	0.28	0.02	0.10	0.02
99	Oberbergischer Kreis	CDU	CDU	CDU	CDU	0.51	0.32	0.01	0.09	0.03
100	Rheinisch-Bergischer...	CDU	CDU	CDU	CDU	0.52	0.31	0.01	0.11	0.02
101	Leverkusen – Köln IV	SPD	SPD	CDU	CDU	0.38	0.38	0.00	0.12	0.04
102	Wuppertal I	SPD	SPD	CDU	SPD	0.37	0.38	0.01	0.11	0.06
103	Solingen – Remscheid...	CDU	CDU	CDU	CDU	0.42	0.38	0.01	0.10	0.04
104	Mettmann I	CDU	CDU	CDU	CDU	0.47	0.32	0.01	0.10	0.02
105	Mettmann II	CDU	CDU	CDU	CDU	0.43	0.38	0.01	0.09	0.04
106	Düsseldorf I	CDU	CDU	CDU	CDU	0.46	0.32	0.02	0.12	0.03
107	Düsseldorf II	CDU	CDU	CDU	CDU	0.40	0.37	0.01	0.12	0.05
108	Neuss I	CDU	CDU	CDU	CDU	0.49	0.34	0.01	0.08	0.02
109	Mönchengladbach	CDU	CDU	CDU	CDU	0.48	0.31	0.02	0.09	0.03
110	Krefeld I – Neuss II	CDU	CDU	CDU	CDU	0.46	0.33	0.03	0.09	0.02
111	Viersen	CDU	CDU	CDU	CDU	0.51	0.30	0.02	0.10	0.03
112	Kleve	CDU	CDU	CDU	CDU	0.51	0.31	0.01	0.09	0.02
113	Wesel I	CDU	SPD	CDU	CDU	0.42	0.41	0.00	0.09	0.03
114	Krefeld II – Wesel I...	SPD	SPD	CDU	SPD	0.39	0.42	0.01	0.10	0.04
115	Duisburg I	SPD	SPD	SPD	SPD	0.35	0.44	-0.01	0.10	0.05
116	Duisburg II	SPD	SPD	SPD	SPD	0.31	0.49	-0.01	0.08	0.06
117	Oberhausen – Wesel I...	SPD	SPD	SPD	SPD	0.32	0.47	-0.00	0.11	0.05
118	Mülheim – Essen I	SPD	SPD	SPD	SPD	0.36	0.42	0.01	0.10	0.04
119	Essen II	SPD	SPD	SPD	SPD	0.32	0.49	-0.00	0.09	0.06
120	Essen III	CDU	SPD	CDU	SPD	0.39	0.40	0.00	0.12	0.03
121	Recklinghausen I	SPD	SPD	SPD	SPD	0.35	0.45	0.00	0.09	0.06
122	Recklinghausen II	SPD	SPD	SPD	SPD	0.35	0.43	0.01	0.09	0.05
123	Gelsenkirchen	SPD	SPD	SPD	SPD	0.30	0.54	0.00	0.10	-0.02
124	Steinfurt I – Borken...	CDU	CDU	CDU	CDU	0.49	0.34	0.03	0.09	0.02
125	Bottrop – Recklingha...	SPD	SPD	SPD	SPD	0.37	0.45	-0.01	0.08	0.05
126	Borken II	CDU	CDU	CDU	CDU	0.55	0.29	0.01	0.09	0.02
127	Coesfeld – Steinfurt...	CDU	CDU	CDU	CDU	0.53	0.30	0.02	0.10	0.02
128	Steinfurt III	CDU	CDU	CDU	CDU	0.46	0.42	0.01	0.06	0.03
129	Münster	CDU	CDU	CDU	CDU	0.42	0.36	0.02	0.15	0.02
130	Warendorf	CDU	CDU	CDU	CDU	0.50	0.34	0.03	0.07	0.04
131	Gütersloh I	CDU	CDU	CDU	CDU	0.47	0.35	0.01	0.10	0.03
132	Bielefeld – Gütersl...	SPD	CDU	CDU	CDU	0.39	0.38	-0.00	0.15	0.05
133	Herford – Minden-Lüb...	SPD	SPD	CDU	SPD	0.40	0.41	0.01	0.09	0.04

Table C.4.2: Overview of 2013 forecasts

NR	District name	Winner	Uniform Swing	Polls	Combined	Predicted vote shares, combined model				
						CDU/CSU	SPD	FDP	Greens	Left
134	Minden-Lübbecke I	CDU	CDU	CDU	CDU	0.44	0.42	-0.00	0.08	0.03
135	Lippe I	SPD	SPD	CDU	SPD	0.40	0.42	0.02	0.09	0.03
136	Höxter – Lippe II	CDU	CDU	CDU	CDU	0.48	0.35	0.01	0.09	0.04
137	Paderborn – Güterslo...	CDU	CDU	CDU	CDU	0.54	0.28	0.02	0.09	0.03
138	Hagen – Ennepe-Ruhr-...	SPD	SPD	SPD	SPD	0.37	0.44	0.01	0.08	0.04
139	Ennepe-Ruhr-Kreis II	SPD	SPD	SPD	SPD	0.34	0.43	0.02	0.11	0.04
140	Bochum I	SPD	SPD	SPD	SPD	0.35	0.46	-0.01	0.11	0.05
141	Herne – Bochum II	SPD	SPD	SPD	SPD	0.31	0.52	-0.00	0.12	-0.02
142	Dortmund I	SPD	SPD	SPD	SPD	0.32	0.40	0.01	0.13	0.05
143	Dortmund II	SPD	SPD	SPD	SPD	0.33	0.45	-0.00	0.11	0.05
144	Unna I	SPD	SPD	SPD	SPD	0.36	0.45	-0.00	0.11	0.05
145	Hamm – Unna II	SPD	SPD	SPD	SPD	0.37	0.44	0.01	0.08	0.03
146	Soest	CDU	CDU	CDU	CDU	0.48	0.33	0.02	0.09	0.04
147	Hochsauerlandkreis	CDU	CDU	CDU	CDU	0.53	0.31	0.01	0.08	0.03
148	Siegen-Wittgenstein	CDU	CDU	CDU	CDU	0.44	0.40	0.02	0.06	0.04
149	Olpe – Märkischer Kr...	CDU	CDU	CDU	CDU	0.49	0.33	0.02	0.08	0.03
150	Märkischer Kreis II	SPD	SPD	CDU	SPD	0.41	0.41	0.00	0.07	0.04
151	Nordsachsen	CDU	CDU	CDU	CDU	0.44	0.19	0.02	0.05	0.16
152	Leipzig I	CDU	CDU	CDU	CDU	0.37	0.22	0.02	0.09	0.16
153	Leipzig II	CDU	CDU	CDU	CDU	0.33	0.25	0.02	0.13	0.17
154	Leipzig-Land	CDU	CDU	CDU	CDU	0.44	0.21	0.02	0.06	0.15
155	Meißen	CDU	CDU	CDU	CDU	0.47	0.17	0.02	0.07	0.14
156	Bautzen I	CDU	CDU	CDU	CDU	0.46	0.17	0.04	0.06	0.16
157	Görlitz	CDU	CDU	CDU	CDU	0.46	0.16	0.02	0.06	0.19
158	Sächsische Schweiz-O...	CDU	CDU	CDU	CDU	0.49	0.17	0.03	0.06	0.15
159	Dresden I	CDU	CDU	CDU	CDU	0.41	0.22	0.02	0.10	0.20
160	Dresden II – Bautzen...	CDU	CDU	CDU	CDU	0.41	0.19	0.04	0.13	0.14
161	Mittelsachsen	CDU	CDU	CDU	CDU	0.47	0.18	0.03	0.06	0.17
162	Chemnitz	CDU	CDU	CDU	CDU	0.37	0.24	0.03	0.07	0.19
163	Chemnitzer Umland – ...	CDU	CDU	CDU	CDU	0.45	0.19	0.03	0.06	0.15
164	Erzgebirgskreis I	CDU	CDU	CDU	CDU	0.45	0.15	0.05	0.05	0.16
165	Zwickau	CDU	CDU	CDU	CDU	0.43	0.19	0.02	0.06	0.19
166	Vogtlandkreis	CDU	CDU	CDU	CDU	0.44	0.19	0.02	0.06	0.17
167	Waldeck	CDU	SPD	CDU	SPD	0.39	0.40	0.02	0.11	0.04
168	Kassel	SPD	SPD	SPD	SPD	0.34	0.38	0.00	0.15	0.05
169	Werra-Meißner – Hers...	SPD	SPD	CDU	SPD	0.38	0.43	0.01	0.08	0.04
170	Schwalm-Eder	SPD	SPD	SPD	SPD	0.38	0.41	0.01	0.10	0.04
171	Marburg	SPD	SPD	CDU	SPD	0.39	0.41	0.00	0.11	0.04
172	Lahn-Dill	CDU	CDU	CDU	CDU	0.45	0.36	0.01	0.10	0.04
173	Gießen	CDU	CDU	CDU	CDU	0.40	0.37	0.03	0.11	0.03
174	Fulda	CDU	CDU	CDU	CDU	0.51	0.25	0.01	0.09	0.03
175	Main-Kinzig – Wetter...	CDU	CDU	CDU	CDU	0.40	0.37	0.03	0.12	0.03
176	Hochtaunus	CDU	CDU	CDU	CDU	0.48	0.28	0.03	0.11	0.03
177	Wetterau I	CDU	CDU	CDU	CDU	0.44	0.37	0.01	0.10	0.03
178	Rheingau-Taunus – Li...	CDU	CDU	CDU	CDU	0.49	0.31	0.02	0.10	0.02
179	Wiesbaden	CDU	CDU	CDU	CDU	0.43	0.34	0.01	0.11	0.03
180	Hanau	CDU	CDU	CDU	CDU	0.41	0.35	0.01	0.09	0.03
181	Main-Taunus	CDU	CDU	CDU	CDU	0.50	0.26	0.03	0.12	0.02
182	Frankfurt am Main I	CDU	CDU	CDU	CDU	0.37	0.31	0.02	0.15	0.05
183	Frankfurt am Main II	CDU	CDU	CDU	CDU	0.38	0.29	0.01	0.19	0.04
184	Groß-Gerau	CDU	CDU	CDU	CDU	0.39	0.38	0.00	0.12	0.04
185	Offenbach	CDU	CDU	CDU	CDU	0.43	0.32	0.02	0.12	0.04
186	Darmstadt	SPD	CDU	CDU	CDU	0.37	0.36	0.00	0.15	0.03
187	Odenwald	CDU	CDU	CDU	CDU	0.43	0.33	0.02	0.12	0.03
188	Bergstraße	CDU	CDU	CDU	CDU	0.47	0.35	0.01	0.10	0.02
189	Eichsfeld – Nordhaus...	CDU	CDU	CDU	CDU	0.46	0.21	0.02	0.06	0.18
190	Eisenach – Wartburgk...	CDU	CDU	CDU	CDU	0.38	0.24	0.01	0.06	0.19
191	Kyffhäuserkreis – Sö...	CDU	CDU	CDU	CDU	0.37	0.21	0.02	0.06	0.22
192	Gotha – Ilm-Kreis	CDU	CDU	CDU	CDU	0.34	0.26	0.02	0.07	0.19
193	Erfurt – Weimar – We...	CDU	CDU	CDU	CDU	0.34	0.22	0.01	0.09	0.22
194	Gera – Jena – Saale-...	CDU	CDU	CDU	CDU	0.33	0.22	0.03	0.08	0.24
195	Greiz – Altenburger ...	CDU	CDU	CDU	CDU	0.40	0.21	0.02	0.06	0.21
196	Sonneberg – Saalfeld...	CDU	CDU	CDU	CDU	0.35	0.22	0.02	0.05	0.22
197	Suhl – Schmalkalden-...	CDU	CDU	CDU	CDU	0.35	0.22	0.02	0.06	0.23
198	Neuwied	CDU	CDU	CDU	CDU	0.43	0.38	0.02	0.08	0.03
199	Ahrweiler	CDU	CDU	CDU	CDU	0.49	0.29	0.04	0.09	0.03
200	Koblenz	CDU	CDU	CDU	CDU	0.47	0.32	0.02	0.11	0.03
201	Mosel/Rhein-Hunsrück	CDU	CDU	CDU	CDU	0.50	0.28	0.04	0.09	0.03
202	Kreuznach	CDU	CDU	CDU	CDU	0.48	0.32	-0.00	0.09	0.04
203	Bitburg	CDU	CDU	CDU	CDU	0.50	0.28	0.04	0.09	0.03
204	Trier	CDU	CDU	CDU	CDU	0.48	0.30	0.01	0.10	0.04
205	Montabaur	CDU	CDU	CDU	CDU	0.46	0.31	0.02	0.09	0.04
206	Mainz	CDU	CDU	CDU	CDU	0.39	0.34	0.04	0.13	0.03
207	Worms	CDU	SPD	CDU	CDU	0.40	0.39	0.02	0.10	0.03
208	Ludwigshafen/Franken...	CDU	CDU	CDU	CDU	0.42	0.36	0.01	0.09	0.04
209	Neustadt – Speyer	CDU	CDU	CDU	CDU	0.47	0.29	0.02	0.11	0.03
210	Kaiserslautern	SPD	SPD	CDU	CDU	0.37	0.37	0.02	0.09	0.06
211	Pirmasens	CDU	CDU	CDU	CDU	0.43	0.30	0.03	0.09	0.06
212	Südpfalz	CDU	CDU	CDU	CDU	0.44	0.31	0.03	0.11	0.04
213	Altötting	CSU	CSU	CSU	CSU	0.60	0.17	0.01	0.09	0.02
214	Erding – Ebersberg	CSU	CSU	CSU	CSU	0.51	0.21	0.03	0.14	0.02
215	Freising	CSU	CSU	CSU	CSU	0.50	0.19	0.03	0.17	0.02
216	Fürstenfeldbruck	CSU	CSU	CSU	CSU	0.51	0.23	0.02	0.13	0.03

Table C.4.2: Overview of 2013 forecasts

NR	District name	Winner	Uniform Swing	Polls	Combined	Predicted vote shares, combined model				
						CDU/CSU	SPD	FDP	Greens	Left
217	Ingolstadt	CSU	CSU	CSU	CSU	0.58	0.18	0.01	0.10	0.03
218	München-Nord	CSU	CSU	CSU	SPD	0.39	0.39	0.02	0.12	0.03
219	München-Ost	CSU	CSU	CSU	CSU	0.39	0.31	0.04	0.16	0.03
220	München-Süd	CSU	CSU	CSU	CSU	0.39	0.33	0.02	0.16	0.03
221	München-West/Mitte	CSU	CSU	CSU	CSU	0.39	0.32	0.02	0.16	0.04
222	München-Land	CSU	CSU	CSU	CSU	0.48	0.24	0.03	0.13	0.02
223	Rosenheim	CSU	CSU	CSU	CSU	0.52	0.18	0.01	0.14	0.02
224	Starnberg	CSU	CSU	CSU	CSU	0.55	0.19	0.04	0.12	0.02
225	Traunstein	CSU	CSU	CSU	CSU	0.56	0.18	0.00	0.12	0.02
226	Weilheim	CSU	CSU	CSU	CSU	0.54	0.19	0.02	0.13	0.02
227	Deggendorf	CSU	CSU	CSU	CSU	0.54	0.19	0.03	0.09	0.03
228	Landshut	CSU	CSU	CSU	CSU	0.53	0.18	0.04	0.12	0.02
229	Passau	CSU	CSU	CSU	CSU	0.51	0.18	0.07	0.09	0.04
230	Rottal-Inn	CSU	CSU	CSU	CSU	0.56	0.19	0.02	0.08	0.02
231	Straubing	CSU	CSU	CSU	CSU	0.58	0.21	0.01	0.06	0.02
232	Amberg	CSU	CSU	CSU	CSU	0.50	0.21	0.03	0.11	0.03
233	Regensburg	CSU	CSU	CSU	CSU	0.48	0.23	0.03	0.10	0.02
234	Schwandorf	CSU	CSU	CSU	CSU	0.54	0.24	0.01	0.06	0.03
235	Weiden	CSU	CSU	CSU	CSU	0.48	0.22	0.00	0.07	0.02
236	Bamberg	CSU	CSU	CSU	CSU	0.52	0.22	0.02	0.12	0.02
237	Bayreuth	CSU	CSU	CSU	CSU	0.52	0.23	0.01	0.12	0.02
238	Coburg	CSU	CSU	CSU	CSU	0.49	0.27	0.01	0.08	0.03
239	Hof	CSU	CSU	CSU	CSU	0.49	0.30	0.00	0.08	0.04
240	Kulmbach	CSU	CSU	CSU	CSU	0.65	0.18	-0.01	0.07	0.02
241	Ansbach	CSU	CSU	CSU	CSU	0.49	0.25	0.02	0.11	0.02
242	Erlangen	CSU	CSU	CSU	CSU	0.48	0.28	0.02	0.13	0.02
243	Fürth	CSU	CSU	CSU	CSU	0.46	0.29	0.01	0.12	0.04
244	Nürnberg-Nord	CSU	CSU	CSU	CSU	0.39	0.33	0.01	0.14	0.04
245	Nürnberg-Süd	CSU	CSU	CSU	CSU	0.42	0.33	0.01	0.11	0.04
246	Roth	CSU	CSU	CSU	CSU	0.47	0.27	0.02	0.12	0.02
247	Aschaffenburg	CSU	CSU	CSU	CSU	0.46	0.23	0.02	0.18	0.03
248	Bad Kissingen	CSU	CSU	CSU	CSU	0.55	0.21	0.01	0.11	0.04
249	Main-Spessart	CSU	CSU	CSU	CSU	0.54	0.22	0.01	0.11	0.02
250	Schweinfurt	CSU	CSU	CSU	CSU	0.49	0.24	0.02	0.12	0.05
251	Würzburg	CSU	CSU	CSU	CSU	0.47	0.26	0.02	0.16	0.02
252	Augsburg-Stadt	CSU	CSU	CSU	CSU	0.44	0.22	0.02	0.16	0.03
253	Augsburg-Land	CSU	CSU	CSU	CSU	0.55	0.18	0.02	0.11	0.02
254	Donau-Ries	CSU	CSU	CSU	CSU	0.55	0.19	0.03	0.09	0.02
255	Neu-Ulm	CSU	CSU	CSU	CSU	0.53	0.20	0.03	0.12	0.03
256	Oberallgäu	CSU	CSU	CSU	CSU	0.55	0.16	0.02	0.16	0.02
257	Ostallgäu	CSU	CSU	CSU	CSU	0.53	0.15	0.03	0.13	0.02
258	Stuttgart I	CDU	CDU	CDU	CDU	0.38	0.24	0.02	0.25	0.04
259	Stuttgart II	CDU	CDU	CDU	CDU	0.38	0.30	0.02	0.18	0.04
260	Böblingen	CDU	CDU	CDU	CDU	0.48	0.24	0.02	0.15	0.03
261	Esslingen	CDU	CDU	CDU	CDU	0.46	0.31	0.01	0.16	0.04
262	Nürtingen	CDU	CDU	CDU	CDU	0.46	0.27	0.02	0.16	0.02
263	Göppingen	CDU	CDU	CDU	CDU	0.46	0.30	0.01	0.14	0.03
264	Waiblingen	CDU	CDU	CDU	CDU	0.46	0.28	0.02	0.14	0.03
265	Ludwigsburg	CDU	CDU	CDU	CDU	0.44	0.24	0.02	0.17	0.04
266	Neckar-Zaber	CDU	CDU	CDU	CDU	0.45	0.26	0.04	0.14	0.03
267	Heilbronn	CDU	CDU	CDU	CDU	0.48	0.26	0.02	0.12	0.03
268	Schwäbisch Hall – Ho...	CDU	CDU	CDU	CDU	0.46	0.25	0.04	0.14	0.03
269	Backnang – Schwäbisc...	CDU	CDU	CDU	CDU	0.48	0.29	0.02	0.13	0.03
270	Aalen – Heidenheim	CDU	CDU	CDU	CDU	0.48	0.30	0.02	0.12	0.04
271	Karlsruhe-Stadt	CDU	CDU	CDU	CDU	0.40	0.28	0.02	0.16	0.04
272	Karlsruhe-Land	CDU	CDU	CDU	CDU	0.48	0.30	0.02	0.13	0.02
273	Rastatt	CDU	CDU	CDU	CDU	0.51	0.27	0.01	0.14	0.03
274	Heidelberg	CDU	CDU	CDU	CDU	0.40	0.34	0.03	0.19	0.02
275	Mannheim	CDU	CDU	CDU	CDU	0.40	0.34	0.01	0.14	0.05
276	Odenwald – Tauber	CDU	CDU	CDU	CDU	0.53	0.25	0.02	0.11	0.03
277	Rhein-Neckar	CDU	CDU	CDU	CDU	0.45	0.30	0.03	0.13	0.03
278	Bruchsal – Schwetzin...	CDU	CDU	CDU	CDU	0.50	0.28	0.02	0.11	0.03
279	Pforzheim	CDU	CDU	CDU	CDU	0.45	0.26	0.05	0.12	0.04
280	Calw	CDU	CDU	CDU	CDU	0.50	0.25	0.03	0.13	0.03
281	Freiburg	CDU	SPD	CDU	SPD	0.33	0.33	0.01	0.20	0.04
282	Lörrach – Müllheim	CDU	CDU	CDU	CDU	0.42	0.35	0.02	0.13	0.03
283	Emmendingen – Lahr	CDU	CDU	CDU	CDU	0.45	0.26	0.02	0.17	0.03
284	Offenburg	CDU	CDU	CDU	CDU	0.50	0.23	0.02	0.14	0.03
285	Rottweil – Tuttlinge...	CDU	CDU	CDU	CDU	0.50	0.20	0.06	0.12	0.02
286	Schwarzwald-Baar	CDU	CDU	CDU	CDU	0.50	0.24	0.03	0.14	0.03
287	Konstanz	CDU	CDU	CDU	CDU	0.47	0.24	0.05	0.17	0.04
288	Waldshut	CDU	CDU	CDU	CDU	0.45	0.30	0.03	0.13	0.03
289	Reutlingen	CDU	CDU	CDU	CDU	0.44	0.26	0.03	0.16	0.04
290	Tübingen	CDU	CDU	CDU	CDU	0.42	0.24	0.01	0.20	0.04
291	Ulm	CDU	CDU	CDU	CDU	0.46	0.27	0.02	0.16	0.04
292	Biberach	CDU	CDU	CDU	CDU	0.48	0.18	0.04	0.15	0.03
293	Bodensee	CDU	CDU	CDU	CDU	0.48	0.19	0.04	0.19	0.03
294	Ravensburg	CDU	CDU	CDU	CDU	0.48	0.23	0.03	0.18	0.03
295	Zollernalb – Sigmari...	CDU	CDU	CDU	CDU	0.52	0.21	0.03	0.13	0.03
296	Saarbrücken	CDU	CDU	CDU	CDU	0.36	0.28	-0.00	0.09	0.14
297	Saarlouis	CDU	CDU	CDU	CDU	0.41	0.34	0.00	0.07	0.09
298	St. Wendel	CDU	CDU	CDU	CDU	0.43	0.32	-0.00	0.06	0.10
299	Homburg	CDU	CDU	CDU	CDU	0.37	0.32	0.01	0.07	0.13

Figure C.4.1: Constituency-level first vote shares from one election to the next, 1990 - 2009. The red lines are linear fits of bivariate linear models.

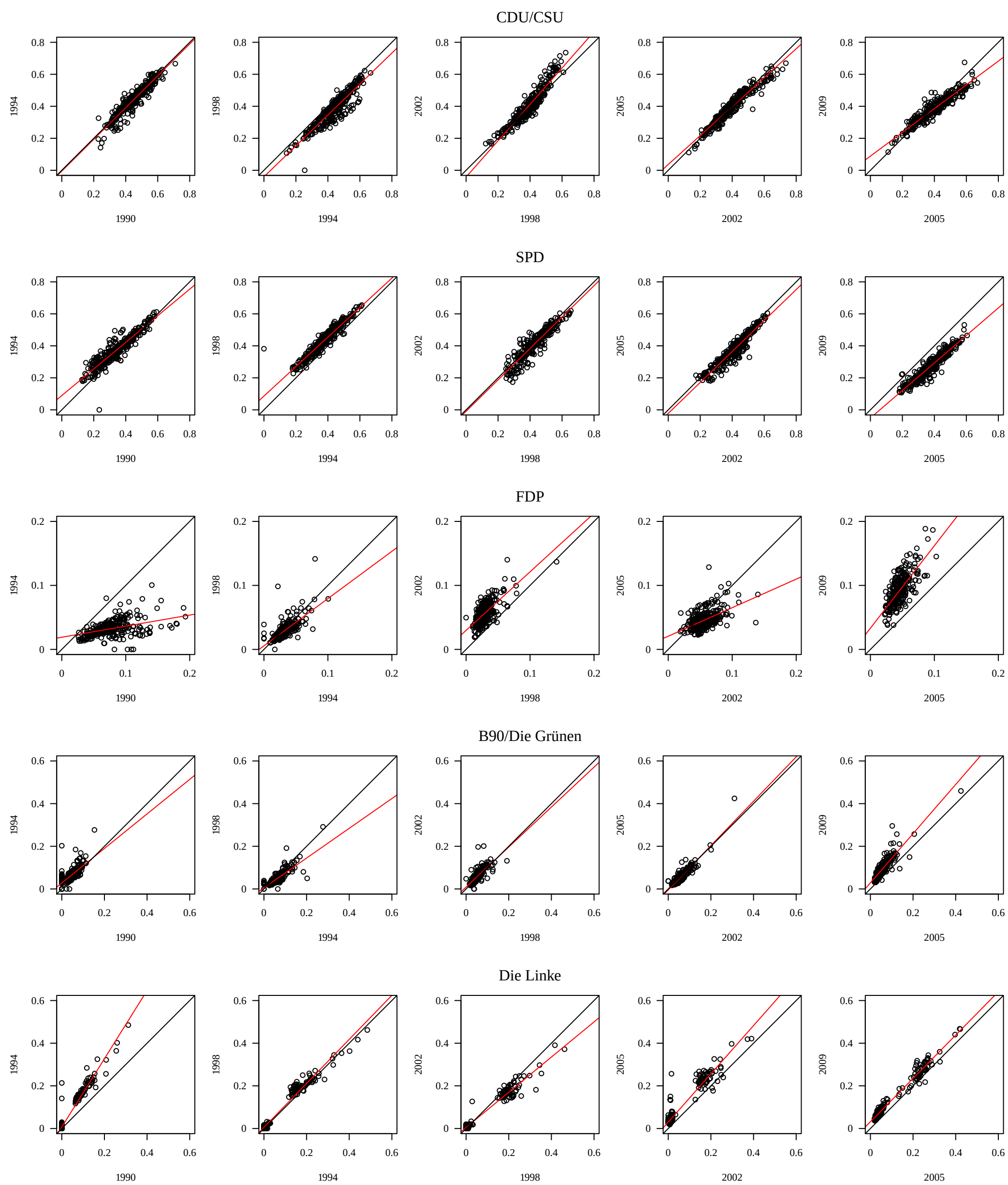


Figure C.4.2: Constituency-level estimation results the uniform swing forecasting model, before and after correction.

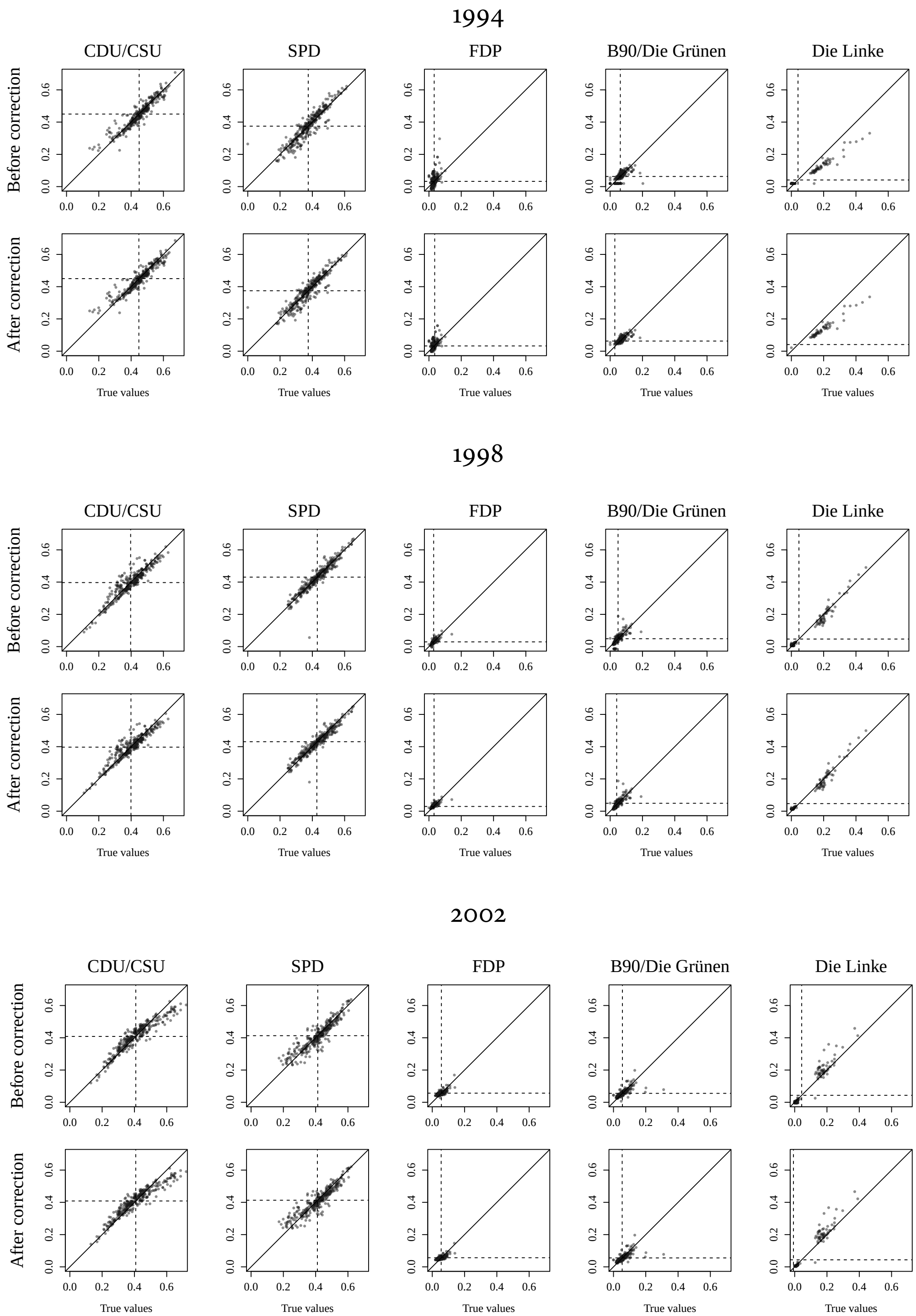
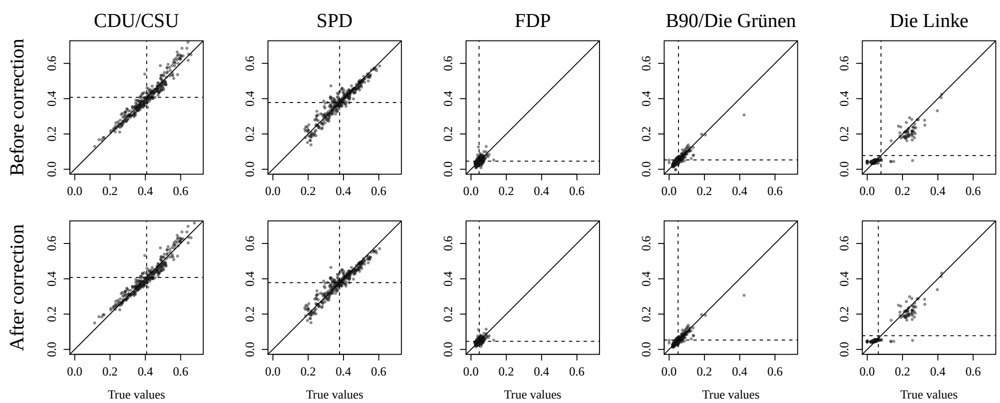


Figure C.4.2, *continued*: Constituency-level estimation results the uniform swing forecasting model, before and after correction.

2005



2009

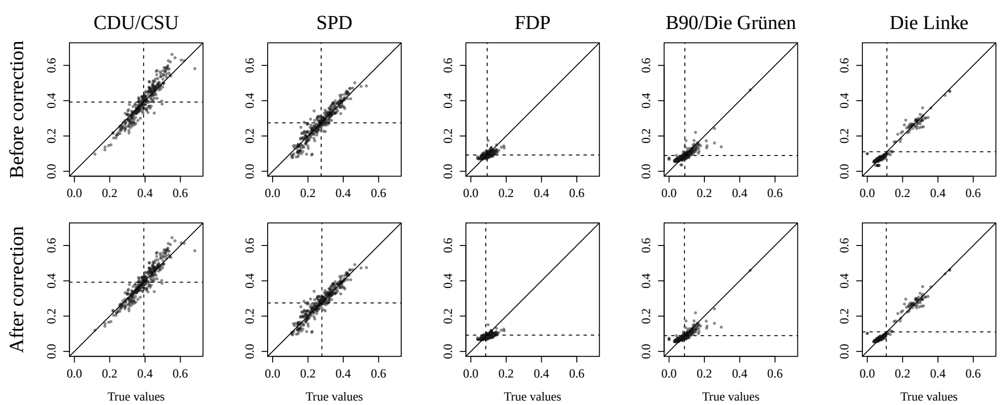
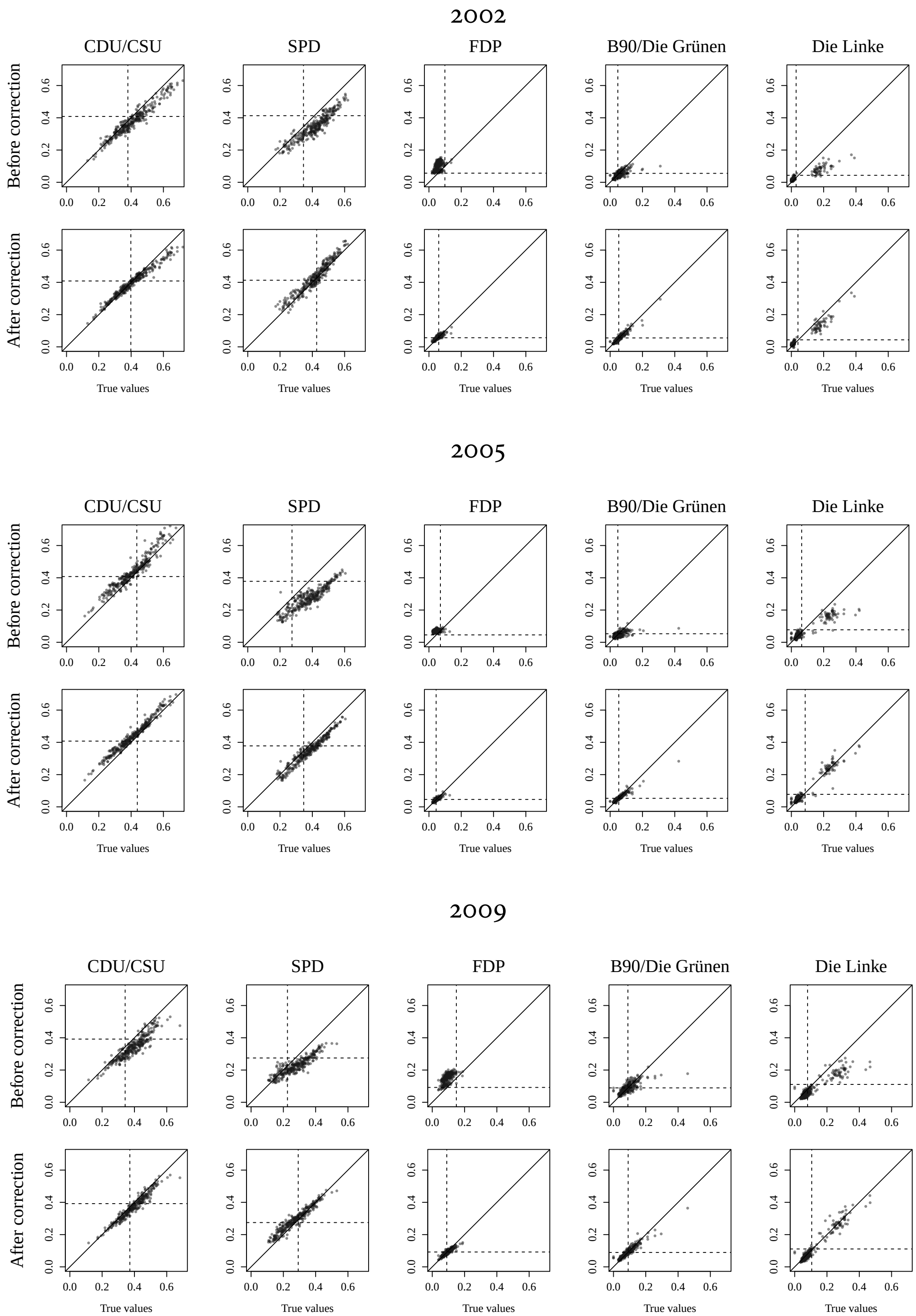


Figure C.4.3: Constituency-level estimation results the poll-based forecasting model, before and after correction.



Bibliography

Abramowitz, Alan I. 2010. "How Large a Wave? Using the Generic Ballot to Forecast the 2010 Midterm Elections."

Abramowitz, Alan I. 2006. Disconnected, or Joined at the Hip? In *Red and Blue Nation? Vol. 1: Characteristics and Causes of America's Polarized Politics*, ed. Pietro S. Nivola and David W. Brady. Washington, D.C.: Brookings Institution Press pp. 72—85.

Abramowitz, Alan I. 1988. "An Improved Model for Predicting Presidential Election Outcomes." *PS: Political Science and Politics* 21(4):843–847.

Abramowitz, Alan and Kyle Saunders. 2005. "Why Can't We All Just Get Along? The Reality of a Polarized America." *The Forum* 3(2):1–22.

Adams, James, Catherine E. De Vries and Debra Leiter. 2011. "Subconstituency Reactions to Elite Depolarization in the Netherlands: An Analysis of the Dutch Public's Policy Beliefs and Partisan Loyalties, 1986-98." *British Journal of Political Science* 42(1):81—105.

Adams, James and Lawrence Ezrow. 2009. "Who Do European Parties Represent? How Western European Parties Represent the Policy Preferences of Opinion Leaders." *Journal of Politics* 71(1):206–223.

Adams, James, Jane Green and Caitlin Milazzo. 2012*a*. "Has the British Public Depolarized Along With Political Elites? An American Perspective on British Public Opinion." *Comparative Political Studies* 45(4):507—530.

Adams, James, Jane Green and Caitlin Milazzo. 2012*b*. "Who moves? Elite and mass-level depolarization in Britain, 1987-2001." *Electoral Studies* 31:643—655.

Agnew, John. 2002. *The Making of Political Geography*. Oxford: Oxford University Press.

Agresti, Alan and Brent A. Coull. 1998. "Approximate Is Better than "Exact" for Interval Estimation of Binomial Proportions." *American Statistician* 52:119–126.

von Alemann, Ulrich. 2010. *Das Parteiensystem der Bundesrepublik Deutschland*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Andress, Hans-Jürgen and Thorsten Heien. 2001. "Four Worlds of Welfare State Attitudes? A Comparison of Germany, Norway, and the United States." *European Sociological Review* 17(4):337—356.

Ansolabehere, Stephen, Jr. Snyder, James M. and Stewart III, Charles. 2001. "Candidate Positioning in U.S. House Elections." *American Journal of Political Science* 45:136–159.

Ardoin, Phillip J. and James C. Garand. 2003. "Measuring Constituency Ideology in U.S. House Districts: A Top-down Simulation Approach." *Journal of Politics* 65:1165–1189.

Armstrong, J. Scott. 2001. Combining Forecasts. In *Principles of Forecasting: A Handbook for Researchers and Practitioners*, ed. J. Scott Armstrong. Kluwer p. Norwell.

Bafumi, Joseph, Robert S. Erikson and Christopher Wlezien. 2010. "Forecasting House Seats from Generic Congressional Polls: The 2010 Midterm Election." *PS: Political Science & Politics* 43(04):633–636.

Bafumi, Joseph, Robert S. Erikson and Christopher Wlezien. 2008. Forecasting House Seats from Generic Congressional Polls. In *Elections and Exit Polling*, ed. Wendy Alvey and Fritz Scheuren. New York: Wiley and Sons.

Bafumi, Joseph and Michael C. Herron. 2006. "Leapfrog Representation and Extremism: A Study of American Voters and Their Members in Congress." *American Political Science Review* 104:519–542.

Bafumi, Joseph and Robert Y. Shapiro. 2009. "A New Partisan Voter." *The Journal of Politics* 71(1):1–24.

Baldassarri, Delia and Andrew Gelman. 2008. "Partisans without constraint: Political polarization and trends in American public opinion." *American Journal of Sociology* 114(2):408–446.

Bao, Le, Tilmann Gneiting, Eric P. Grimit, Peter Guttorp and Adrian E. Raftery. 2010. "Bias Correction and Bayesian Model Averaging for Ensemble Forecasts of Surface Wind Direction." *Monthly Weather Review* 138:1811–1821.

Bates, J. M. and C. W. J. Granger. 1969. "The Combination of Forecasts." *Operational Research Society* 20(4):451–468.

Baumann, Markus, Marc Debus and Jochen Müller. 2013. "Das legislative Verhalten von Bundestagsabgeordneten zwischen persönlichen Charakteristika, Wahlkreisinteressen und Parteilinie." *Zeitschrift für Politikwissenschaft* 23(2):177–211.

Bean, Clive and Elim Papadakis. 1998. "A Comparison of Mass Attitudes towards the Welfare State in Different Institutional Regimes, 1985-1990." *International Journal of Public Opinion Research* 10(3):211—236.

Berkman, Michael B. and Eric Plutzer. 2005. *Ten Thousand Democracies: Politics and Public Opinion in America's School Districts*. Washington: Georgetown University Press.

Bernauer, Julian and Simon Munzert. 2014. "Loyal to the Game? Strategic Policy Representation in Mixed Electoral Systems." *Representation* 50(1):83–97.

Berry, William D., Richard C. Fording, Evan J. Ringquist, Russell L. Hanson and Carl Klarner. 2010. "Measuring Citizen and Government Ideology in the American States: A Re-Appraisal." *State Politics and Policy Quarterly* 10:117–135.

Besag, Julian. 1975. "Statistical Analysis of Non-Lattice Data." *Statistician* 24(3):179–195.

Besag, Julian, Jeremy York and Annie Mollié. 1991. "Bayesian image restoration, with two applications in spatial statistics." *Annals of the Institute of Statistical Mathematics* 43(1):1–20.

Billiet, Jaak B. and Eldad Davidov. 2008. "Testing the Stability of an Acquiescence Style Factor Behind Two Interrelated Substantive Variables in a Panel Design." *Sociological Methods and Research* 36(4):542—562.

Billiet, Jaak B. and McKee J. McClendon. 2000. "Modeling Acquiescence in Measurement Models for Two Balanced Sets of Items." *Structural Equation Modeling* 7(4):608—628.

Bishin, Benjamin G. 2000. "Constituency Influence in Congress: Does Subconstituency Matter?" *Legislative Studies Quarterly* 25(3):389–415.

Bivand, R. S., E. J. Pebesma and V. Gómez-Rubio. 2008. *Applied Spatial Data Analysis with R*. Springer.

Bivand, Roger, with contributions by Micah Altman, Luc Anselin, Renato Assunção, Olaf Berke, Andrew Bernat, Eric Blankmeyer, Marilia Carvalho, Yongwan Chun, Bjarke Christensen, Carsten Dormann, Stéphane Dray, Rein Halbersma, Elias Krainski, Nicholas Lewin-Koh, Hongfei Li, Jielai Ma, Giovanni Millo, Werner Mueller, Hisaji Ono, Pedro Peres-Neto, Gianfranco Piras, Markus Reder, Michael Tiefelsdorf, and Danlin Yu. 2010. *spdep: Spatial dependence: weighting schemes, statistics and models*. R package version 0.5-11.

Blau, Adrian. 2004. "A quadruple whammy for first-past-the-post." *Electoral Studies* 23:431–453.

Blekesaune, Morten and Jill Quadagno. 2003. "Public Attitudes toward Welfare State Policies: A Comparative Analysis of 24 Nations." *European Sociological Review* 19(5):415—427.

Brace, Paul, Kellie Sims-Butler, Kevin Arceneaux and Martin Johnson. 2002. "Public Opinion in the American States: New Perspectives Using National Survey Data." *American Journal of Political Science* 46(1):173–189.

Brady, David W., John Ferejohn and Laurel Harbridge. 2008. Polarization and Public Policy: A General Assessment. In *Red and Blue Nation? Vol. 2: Consequences and Correction of America's Polarized Politics*, ed. Pietro S. Nivola and David W. Brady. Washington, D.C.: Brookings Institution Press pp. 185—216.

Bundesministerium für Justiz [Federal Ministry of Justice]. 1957. "Gleichberechtigungsgesetz in der im Bundesgesetzblatt Teil III, Gliederungsnummer 400-3, veröffentlichten bereinigten Fassung, das durch Artikel 127 des Gesetzes vom 19. April 2006 (BGBl. I S. 866) geändert worden ist". [Law of equality in the form of April 19th, 2006.]

Butler, David and Stephen D. Van Beek. 1990. "Why Not Swing? Measuring Electoral Change." *PS: Political Science and Politics* 23(2):178–184.

Buttice, Matthew K. and Benjamin Highton. 2013. "How does multilevel regression and post-stratification perform with conventional national surveys?" *Political Analysis* 21(4):449–467.

Campbell, James E. 2010. "The Seats in Trouble Forecast of the 2010 Elections to the U.S. House." *PS: Political Science & Politics* 43(04):627–630.

Canes-Wrone, Brandice, David W. Brady and John F. Cogan. 2002. "Out of Step, Out of Office: Electoral Accountability and House Members' Voting." *American Political Science Review* 96(01):127–140.

Carlin, Bradley P. and Thomas A. Louis. 1996. *Bayes and Empirical Bayes Methods for Data Analysis*. London: Chapman & Hall.

Carter, J. Scott, Shannon Carter and Jamie Dodge. 2009. "Trends in Abortion Attitudes by Race and Gender: A Reassessment Over a Four-Decade Period." *Journal of Sociological Research* 1(1):1—17.

Chen, Jowei. 2010. "The Effect of Electoral Geography on Pork Barreling in Bicameral Legislatures." *American Journal of Political Science* 54(2):301–322.

Christensen, William F. and Lindsay Florence. 2008. "Predicting Presidential and Other Multistage Election Outcomes Using State-Level Pre-Election Polls" *The American Statistician* 62:1-10.

Clemen, R. T. 1989. "Combining forecasts: A review and annotated bibliography." *International Journal of Forecasting* 5:559–583.

Clinton, Joshua, D. 2006. "Representation in Congress: Constituents and Roll Calls in the 106th House." *Journal of Politics* 68(2):397–409.

Cohen, Jeffrey E. 2006. Conclusions: Where We Have Been, Where Should We Go. In *Public Opinion in State Politics*, ed. Jeffrey E. Cohen. Stanford University Press pp. 254–270.

Cohen, Joachim, Isabelle Marcoux, Johan Bilsen, Patrick Deboosere, Gerrit van der Wal and Luc Deliens. 2006. "Trends in acceptance of euthanasia among the general public in 12 European countries (1981-1999)." *European Journal of Public Health* 16(6):663—669.

Converse, Philip E. 1964. The Nature of Belief Systems in Mass Publics. In *Ideology and Discontent*, ed. David E. Apter. New York: The Free Press pp. 206—261.

Crane, Wilder W. 1960. "Do Representatives Represent?" *Journal of Politics* 22(2):295–299.

Cressie, Noel A.C. 1993. *Statistics for Spatial Data*. 2 ed. New York: Wiley.

Davis, Shannon N. and Theodore N. Greenstein. 2009. "Gender Ideology: Components, Predictors, and Consequences." *Annual Review of Sociology* 35:87—105.

Debus, Marc and Hanna Bäck. 2014. "Der Einfluss von Mandatstyp und Wahlkreiseigenschaften auf die inhaltlichen Positionen in Parlamentsreden: Eine Untersuchung anhand der parlamentarischen Debatten im Deutschen Bundestag von 1998 bis 2002." *Swiss Political Science Review* 20(2):330–353.

de Sola Pool, Ithiel, Robert P. Abelson and Samuel L. Popkin. 1965. *Candidates, Issues, and Strategies*. Cambridge: MIT Press.

DiMaggio, Paul, John Evans and Bethany Bryson. 1996. "Have American's Social Attitudes Become More Polarized?" *American Journal of Sociology* 102(3):690–755.

Down, Ian and Carole J. Wilson. 2010. "Opinion Polarization and Inter-Party Competion on Europe." *European Union Politics* 11(1):61—87.

Downey, Dennis J. and Matt L. Huffman. 2001. "Attitudinal Polarization and Trimodal Distributions: Measurement Problems and Theoretical Implications." *Social Science Quarterly* 82(3):494—505.

Elazar, Daniel. 1994. *The American Mosaic: The Impact of Space, Time, and Culture on American Politics*. Washington: Westview Press.

Erikson, Robert S. 1981. "Measuring Constituency Opinion: The 1978 U. S. Congressional Election Survey." *Legislative Studies Quarterly* 6(2):235–245.

Erikson, Robert S. 1980. "Policy Representation of Constituency Interests." *Political Behavior* 2(1):91–106.

Erikson, Robert S. 1978. "Constituency Opinion and Congressional Behavior: A Reexamination of the Miller-Stokes Representation Data." *American Journal of Political Science* 22:511–535.

Erikson, Robert S., Gerald D. Wright and John P. McIver. 1993. *Statehouse Democracy: Public Opinion and Policy in the American States*. Cambridge: Cambridge University Press.

Eulau, Heinz, John C. Wahlke, William Buchanan and Leroy C. Ferguson. 1959. "The Role of the Representative: Some Empirical Observations on the Theory of Edmund Burke." *American Political Science Review* 53(3):742–756.

Evans, John H. 2003. "Have Americans' Attitudes Become More Polarized? - An Update." *Social Science Quarterly* 84(1):71—90.

Ezrow, Lawrence, Catherine De Vries, Marco Steenbergen and Erica Edwards. 2011. "Mean voter representation and partisan constituency representation: Do parties respond to the mean voter position or to their supporters?" *Party Politics* 17(3):275–301.

Fenno, Richard F. 1978. *Home Style: House Members in Their Districts*. Boston: Little, Brown.

Fiorina, Morris P. and Samuel J. Abrams. 2008. "Political Polarization in the American Public." *The Annual Review of Political Science* 11:563–588.

Fiorina, Morris P., Samuel J. Abrams and Jeremy C. Pope. 2005. *Culture War? The Myth of Polarized America*. New York: Pearson Longman.

Fischer, C. and G. Mattson. 2009. "Is America Fragmenting?" *Annual Review of Sociology* 35:435.

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, Stephen D. 2014. "Predictable and unpredictable changes in party support: A method for long-range daily election forecasting from opinion polls." Unpublished Manuscript.

Fisher, S.D., R. Ford, W. Jennings, M. Pickup and C. Wlezien. 2011. "From polls to votes to seats: Forecasting the 2010 British general election." *Electoral Studies* 30(2):250–257.

Fleisher, Richard and John R. Bond. 2004. "The Shrinking Middle in the US Congress." *British Journal of Political Science* 34(3):429—451.

Galston, William A. and Pietro S. Nivola. 2006. Delineating the Problem. In *Red and Blue Nation? Vol. 1: Characteristics and Causes of America's Polarized Politics*, ed. Pietro S. Nivola and David W. Brady. Washington, D.C.: Brookings Institution Press pp. 1—47.

Galtung, Johan. 1966. Rank and Social Integration: A Multidimensional Approach. In *Sociological Theories in Progress*, ed. J. Berger, M. Zelditch and B. Anderson. Boston: Houghton-Mifflin pp. 145—198.

Gay, Claudine. 2007. "Legislating Without Constraints: The Effect of Minority Districting on Legislators' Responsiveness to Constituency Preferences." *Journal of Politics* 69(02):442–456.

Gelman, Andrew. 2008. *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*. Princeton: Princeton University Press.

Gelman, Andrew. 2007. "Struggles with Survey Weighting and Regression Modeling." *Statistical Science* 22:153–164.

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 and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Gelman, Andrew and Gary King. 1993. "Why Are American Presidential Election Campaign Polls So Variable When Votes Are So Predictable?" *British Journal of Political Science* 23(4):409–451.

Gelman, Andrew and Thomas C. Little. 1997. "Poststratification into Many Categories Using Hierarchical Logistic Regression." *Survey Methodology* 23:127–125.

Gerring, John. 2012. "Mere Description." *British Journal of Political Science* 4(4):721—746.

Gibson, James L. 1992. "The Political Consequences of Intolerance: Cultural Conformity and Political Freedom." *American Political Science Review* 86(2):338–356.

Glahn, Harry R. and Dale A. Lowry. 1972. "The Use of Model Output Statistics (MOS) in Objective Weather Forecasting." *Journal of Applied Meteorolgy* 11(8):1203–1211.

Gleditsch, Kristian Skrede. 2002. *All international politics is local: The diffusion of conflict, integration, and democratization*. University of Michigan Press.

Gómez-Rubio, Virgilio, Nicky Best, Sylvia Richardson, Guangquan Li and Philip Clarke. 2010. "Bayesian Statistics for Small Area Estimation." Unpublished Manuscript.

Gómez-Rubio, V., N. Best, S. Richardson, G. Li and P. Clarke. forthcoming. Bayesian Statistics Small Area Estimation. Unpublished Manuscript

Graefe, Andreas. 2014. "Accuracy of Vote Expectation Surveys in Forecasting Elections." *Public Opinion Quarterly* 78(S1):204–232.

Graefe, Andreas, J. Scott Armstrong Armstrong, Randall J. Jones Jr. and Alfred G. Cuzán. 2014. "Combining forecasts: An application to elections." *International Journal of Forecasting* 30(1):43–54.

Granberg, Donald and Beth Wellman Granberg. 1980. "Abortion Attitudes, 1965-1980: Trends and Determinants." *Family Planning Perspectives* 12(5):250—261.

Green, Donald P. and Jack Citrin. 1994. "Measurement Error and the Structure of Attitudes: Are Positive and Negative Judgments Opposites?" *American Journal of Political Science* 38(1):256—281.

Gschwend, T. and H. Norpoth. 2005. "Prognosemodell auf dem Prüfstand: Die Bundestagswahl 2005." *Politische Vierteljahresschrift* 46(4):682–688.

Gschwend, T. and H. Norpoth. 2001. 'Wenn am nächsten Sonntag…': Ein Prognosemodell für Bundestagswahlen. In *Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 1998*, ed. Hans-Dieter Kilngemann and Max Kaase. Wiesbaden: Westdeutscher Verlag pp. 473–499.

Hanretty, Chris, Benjamin Lauderdale and Nick Vivyan. Forthcoming. "2015 UK Parliamentary Election Forecast." *Electoral Studies*

Hanretty, Chris, Benjamin Lauderdale and Nick Vivyan. 2015. "2015 UK Parliamentary Election Forecast.". *http://www.electionforecast.co.uk/* (Last accessed Monday 10th August, 2015).

Hanretty, Chris, Benjamin E. Lauderdale and Nick Vivyan. 2014. "Dyadic representation in a Westminster system." Unpublished Manuscript.

Hedlund, Ronald D. and H. Paul Friesema. 1972. "Representatives' Perceptions of Constituency Opinion." *Journal of Politics* 34(3):730–752.

Heppen, John. 2003. "Racial and Social Diversity and U.S. Presidential Election Regions." *The Professional Geographer* 55:191–205.

Hero, Rodney E. 1998. *Faces of Inequality: Social Diversity in American Politics*. Oxford: Oxford University Press.

Herrmann, Michael. 2012. "Voter Uncertainty and Failure of Duverger's Law: An Empirical Analysis." *Public Choice* 151:63–90.

Herrmann, Michael, Simon Munzert and Peter Selb. 2015. "Determining the Impact of Strategic Voting on Election Results." Forthcoming at the *Journal of the Royal Statistical Society Series A*.

Hetherington, Marc J. 2009. "Putting Polarization in Perspective." *British Journal of Political Science* 39:413–448.

Hibbs, Douglas. 1982. "President Pragan's mandate from the 1980 elections: a shift to the right?" *American Politics Quarterly* 10:387–420.

Hill, Kim Quaile and Angela Hinton-Anderson. 1995. "Pathways of Representation: A Causal Analysis of Public Opinion-Policy Linkages." *American Journal of Political Science* 39:924–935.

Hill, Seth J., Michael C. Herron, Jeffrey B. Lewis. 2010. "Economic Crisis, Iraq, and Race: A Study of the 2008 Presidential Election." *Election Law Journal* 9:41–62.

Huefken, Volker. 2010. "Supplementary Questionnaire and Nonresponse-Results from the German ISSP Survey." *Quality & Quantity* 44:607–622.

Hunter, James Davison. 1991. *Culture Wars: The Struggle to Define America*. New York: Basic Books.

Hurley, Patricia A. and Kim Quaile Hill. 2003. "Beyond the Demand-Input Model: A Theory of Representational Linkages." *Journal of Politics* 65:304–326.

Issler, Joao Victor, Claudia Rodrigues and Rafael Burjack. 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. Understanding International Commodity Price Fluctuations.

Issler, Joao Victor and Luiz Renato Lima. 2009. "A panel data approach to economic forecasting: The bias-corrected average forecast." *Journal of Econometrics* 152(2):153–164.

Jackman, Simon. 2012. "Pollster Predictive Performance, 51 out of 51.". *http://www.huffingtonpost.com/simon-jackman/pollster-predictive-perfo_b_2087862.html* (Last accessed Monday 10th August, 2015).

Jackman, Simon. 2005. "Pooling the polls over an election campaign." *Australian Journal of Political Science* 40(4):499–517.

Jackman, Simon. 1994. "Measuring Electoral Bias: Australia, 1949-93." *British Journal of Political Science* 24(3):319–357.

Jackson, John E. 1989. "An Error-in-Variables Approach to Estimating Models with Small Area Data." *Political Analysis* 1:157–180.

Jackson, John E. and David C. King. 1989. "Public Goods, Private Interests, and Representation." *American Political Science Review* 83:1143–1164.

Jelen, Ted G. 1990. "Religious Belief and Attitude Constraint." *Journal for the Scientific Study of Religion* 29(1):118–125.

Jérôme, Bruno, Véronique Jérôme-Speziari and Michael S. Lewis-Beck. 2013. "A Political-Economy Forecast for the 20013 German Elections: Who to Rule with Angela Merkel?" *PS: Political Science & Politics* 46:479–480.

Johnston, R. J. and A. M. Hay. 1982. "On the parameters of uniform swing in single-member constituency electoral systems." *Environment and Planning* 14(1):61–74.

Johnston, Ron and Charles Pattie. 2006. *Putting Voters in their Place: Geography and Elections in Britain*. Oxford: Oxford University Press.

Jones, Jonathan. 2013. "Nate Silver on predicting the 2015 general election.". *http://blogs.spectator.co.uk/coffeehouse/2013/05/nate-silver-on-predicting-the-2015-general-election/* (Last accessed Monday 10th August, 2015).

Kahn, Matthew E. and John G. Matsusaka. 1997. "Demand for Environmental Goods: Evidence from Voting patterns on California Initiatives." *Journal of Law and Economics* 40:137–173.

Kalt, Joseph P. and Mark A. Zupan. 1984. "Capture and Ideology in the Economic Theory of Politics." *American Economic Review* 74(3):279–300.

Kastellec, Jonathan P., Jeffrey R. Lax and Justin H. Phillips. 2010. "Public Opinion and Senate Confirmation of Supreme Court Nominees." *Journal of Politics* 72:767–784.

Katz, Jonathan N. and Gary King. 1999. "A Statistical Model for Multiparty Electoral Data" *American Political Science Review* 93:15–32.

Kau, James B. and Paul H. Rubin. 1979. "Self-Interest, Ideology, and Logrolling in Congressional Voting." *Journal of Law and Economics* 22(2):365–384.

Kayser, Mark A. and Arndt Leininger. 2013. "A Benchmarking Forecast of the 2013 Bundestag Election." *http://mark-kayser.com/_researchblog/researchblog.html* (Last accessed Monday 10th August, 2015).

Keil, Silke I. 2004. Parteiprogrammatik in Wahlkampfanzeigen und Wahlprogrammen 1957-2002: Und es gibt ihn doch - den kleinen Unterschied. In *Die Bundestagswahl 2002. Analysen der Wahlergebnisse und des Wahlkampfes*, ed. Frank Brettschneider, Jan van Deth and Edeltraud Roller. Wiesbaden: VS Verlag für Sozialwissenschaften pp. 353–388.

Kernell, Georgia. 2009. "Giving Order to Districts: Estimating Voter Distributions with National Election Returns." *Political Analysis* 17(3):215–235.

Kim, Heemin and Richard Fording. 2001. "Does tactical voting matter? The political impact of tactical voting in recent British elections." *Comparative Political Studies* 34:294–311.

Kirchheimer, Otto. 1965. Der Wandel des Westeuropäischen Parteiensysterns. In *Vergleichende Regierungslehre*, ed. Theo Stammen. Darmstadt: Wissenschaftliche Buchgesellschaft pp. 253–287.

Klarner, Carl E. 2012. "State-Level Forecasts of the 2012 Federal and Gubernatorial Elections." *PS: Political Science & Politics* 45(04):655–662.

Klingemann, H.-D. and B. Wessels. 2001. "The Political Consequences of Germany's Mixed-Member System: Personalization at the Grass Roots?" In *Mixed-Member Electoral Systems: The Best of Both Worlds?*, ed. M.S. Shugart and M. Wattenberg. Oxford: Oxford University Press pp. 279–96.

Krehbiel, Keith. 1993. "Constituency characteristics and legislative preferences." *Public Choice* 76(1):21–37.

Kuklinski, James H. 1978. "Representativeness and Elections: A Policy Analysis." *American Political Science Review* 72(1):165–177.

Kuklinski, James H. 1977. "Constituency Opinion: A Test of the Surrogate Model." *Public Opinion Quarterly* 41:34–40.

Lancaster, Thomas D. and W. David Patterson. 1990. "Comparative Pork Barrel Politics: Perceptions from the West German Bundestag." *Comparative Political Studies* 22(4):458–477.

Larsen, Christian Albrekt. 2008. "The Institutional Logic of Welfare Attitudes: How Welfare Regimes Influence Public Support." *Comparative Political Studies* 41(2):145—168.

Lax, Jeffrey R. and Justin H. Phillips. 2009a. "Gay Rights in the States: Public Opinion and Policy Responsiveness." *American Political Science Review* 103:367–386.

Lax, Jeffrey R. and Justin H. Phillips. 2009b. "How Should We Estimate Public Opinion in The States?" *American Journal of Political Science* 53:107–121.

Layman, Geoffrey C. and Thomas M. Carsey. 2002. "Party Polarization and "Conflict Extension" in the American Electorate." *American Journal of Political Science* 46(4):786–802.

Layman, Geoffrey C., Thomas M. Carsey and Juliana Menasce Horowitz. 2006. "Party Polarization in American Politics: Characteristics, Causes, and Consequences." *Annual Review of Political Science* 9:83–110.

Lebo, Matthew and Helmut Norpoth. 2011. "Yes, Prime Minister: The key to forecasting British elections." *Electoral Studies* 30(2):258–263.

Lebo, M.J. and H. Norpoth. 2007. "The PM and the pendulum: dynamic forecasting of British elections." *British Journal of Political Science* 37(1):71.

Leemann, Lucas and Fabio Wasserfallen. 2014. "Extending the Use and Prediction Precision of Subnational Public Opinion Estimation." Unpublished Manuscript.

Levendusky, Matthew. 2009. *The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans*. Chicago: University of Chicago Press.

Levendusky, Matthew. S. and Jeremy C. Pope. 2011. "Red States vs. Blue States. Going Beyond the Mean." *Public Opinion Quarterly* 75(2):227—248.

Levendusky, Matthew S., Jeremy C. Pope and Simon D. Jackman. 2008. "Measuring District-Level Partisanship with Implications for the Analysis of U.S. Elections." *Journal of Politics* 70(03):736–753.

Levitt, Steven D. 1996. "How Do Senators Vote? Disentangling the Role of Voter Preferences, Party Affiliation, and Senator Ideology." *American Economic Review* 86(3):425–441.

Lewin-Koh, Nicholas J., Roger Bivand, contributions by Edzer J. Pebesma, Eric Archer, Adrian Baddeley, Hans-Joerg Bibiko, Stephane Dray, David Forrest, Michael Friendly, Patrick Giraudoux, Duncan Golicher, Virgilio Gómez-Rubio, Patrick Hausmann, Thomas Jagger, Sebastian P. Luque, Don MacQueen, Andrew Niccolai, Tom Short and Ben Stabler. 2010. *maptools: Tools for reading and handling spatial objects*. R package version 0.7-34.

Lewis-Beck, M.S., R. Nadeau and É. Bélanger. 2011. "Nowcasting V. polling: The 2010 UK election trials." *Electoral Studies* 30(2):284–287.

Lewis-Beck, Michael S. and Tom W. Rice. 1992. *Forecasting Elections*. Washington, DC: Concressional Quarterly.

Lewis-Beck, Michael S. and Tom W. Rice. 1984*a*. "Forecasting Presidential Elections: A Comparison of Naive Models." *Political Behavior* 6(1):9–21.

Lewis-Beck, Michael and Tom W. Rice. 1984*b*. "Forecasting U.S. House Elections." *Legislative Studies Quarterly* 9(3):475–486.

Lewis-Beck, Michael S. and Charles Tien. 2012. "Election Forecasting for Turbulent Times." *PS: Political Science & Politics* 45(04):625–629.

Lewis-Beck, Michael S. and Charles Tien. 2010. "The Referendum Model: A 2010 Congressional Forecast." *PS: Political Science & Politics* 43(04):637–638.

Linzer, Drew A. 2013. "Dynamic Bayesian Forecasting of Presidential Elections in the States." *Journal of the American Statistical Association* 108(501):124–134.

Lock, Kari and Andrew Gelman. 2010. "Bayesian Combination of State Polls and Election Forecasts." *Political Analysis* 18(3):337–348.

Lodge, Milton, Marco R. Steenbergen and Shawn Brau. 1995. "The Responsive Voter: Campaign Information and the Dynamics of Candidate Evaluation." *American Political Science Review* 89(2):309–326.

Loftus, Jeni. 2001. "America's Liberalization in Attitudes toward Homosexuality, 1973 to 1998." *American Sociological Review* 66(5):762—782.

Luebke, David Martin and Sybil Milton. 1994. "Locating the Victim: An Overview of Census-Taking, Tabulation Technology, and Persecution in Nazi Germany." *IEEE Annals of the History of Computing* 16:25–39.

Lunn, David J., Andrew Thomas, Nicky Best and David Spiegelhalter. 2000. "WinBUGS - a Bayesian Modelling Framework: Concepts, Structure, and Extensibility." *Statistics and Computing* 10:325–337.

Mabert, Vincent A. 1978. "Forecast Modification Based upon Residual Analysis: A Case Study of Check Volume Estimation." *Decision Sciences* 9:285–296.

Manow, Philip. 2011. "The Cube Rule in a Mixed Electoral System: Disproportionality in German Bundestag Elections." *West European Politics* 34(4):773–794.

Matthews, Christopher. 1999. *Hardball: How Politics Is Played, Told By One Who Knows the Game*. New York: Simon & Schuster.

Mayda, Anna Maria. 2006. "Who Is against Immigration? A Cross-Country Investigation of Individual Attitudes toward Immigrants." *The Review of Economics and Statistics* 88(3):510—530.

Mayhew, David. 1974. *Congress: The Electoral Connection*. New Haven, CT: Yale University Press.

McCall, Leslie and Jeff Manza. 2011. Class Differences in Social and Political Attitudes in the United States. In *The Oxford Handbook of American Public Opinion and the Media*, ed. Robert Shapiro and Lawrence Jacobs. Oxford University Press pp. 552–570.

McCarty, Nolan, Keith Poole and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches*. Cambridge, MA: MIT Press.

McClendon, McKee J. 1991. "Acquiescence and Recency Response-Order Effects in Interview Surveys." *Sociological Methods and Research* 20(1):60—103.

McCrone, Donald J. and James H. Kuklinski. 1979. "The Delegate Theory of Representation." *American Journal of Political Science* 23(2):278–300.

McCrone, Donald J. and Walter J. Stone. 1986. "The Structure of Constituency Representation: On Theory and Method." *Journal of Politics* 48:956–975.

McDonagh, Eileen Lorenzi. 1993. "Constituency Influence on House Roll-Call Votes in the Progressive Era, 1913-1915." *Legislative Studies Quarterly* 18(2):185–210.

Meuleman, Bart, Eldad Davidov and Jaak Billiet. 2009. "Changing attitudes toward immigration in Europe, 2002-2007: A dynamic group conflict theory approach." *Social Science Research* 38:352—365.

Miller, W. L. 1972. "Measures of Electoral Change Using Aggregate Data." *Journal of the Royal Statistical Society. Series A (General)* 135(1):122–142.

Miller, Warren E. and Donald E. Stokes. 1963. "Constituency Influence in Congress." *American Political Science Review* 57:45–56.

Montgomery, Jacob M., Florian Hollenbach and Michael D. Ward. 2012*a*. "Improving Predictions Using Ensemble Bayesian Model Averaging." *Political Analysis* 20(3):271–291.

Montgomery, Jacob M., Florian M. Hollenbach and Michael D. Ward. 2012*b*. "Ensemble Predictions of the 2012 US Presidential Election." *PS: Political Science & Politics* 45(04):651–654.

Munzert, Simon and Paul C. Bauer. 2013. "Political Depolarization in German Public Opinion, 1980–2010." *Political Science Research and Methods* 1(1):67–89.

Murr, Andreas Erwin. 2011. ""Wisdom of crowds"? A decentralised election forecasting model that uses citizens' local expectations." *Electoral Studies* 30:771–783.

Nadeau, R., M.S. Lewis-Beck and É. Bélanger. 2009. "Election forecasting in the United Kingdom: a two-step model." *Journal of Elections, Public Opinion and Parties* 19(3):333–358.

Norris, Pippa and Ivor Crewe. 1994. "Did the British Marginals Vanish? Proportionality and Exaggeration in the British Electoral System Revisited." *Electoral Studies* 13(3):201–221.

O'Neill, Tip and William Novak. 1988. *Man of the House: The Life and Political Memoirs of Speaker Tip O'Neill*. London: The Bodley Head Ltd.

Openshaw, S. and P. Taylor. 1979. "A Million or so Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem." In *Statistical Applications in the Spatial Sciences*, ed. N. Wrigey. London: Pion.

Panagopoulos, Costas. 2009. "Campaign Dynamics in Battleground and Nonbattleground States." *Public Opinion Quarterly* 73(1):119–129.

Park, David K., Andrew Gelman and Joseph Bafumi. 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.

Pebesma, E.J., R.S. Bivand. 2005. *sp: Classes and methods for spatial data in R.* R package version 0.9-69.

Peffley, Mark A. and Jon Hurwitz. 1985. "A Hierarchical Model of Attitude Constraint." *American Journal of Political Science* 29(4):871—890.

Percival, Garrick L., Martin Johnson and Max Neiman. 2009. "Representation and Local Policy: Relating County-Level Public Opinion to Policy Outputs." *Political Research Quarterly* 62:164–177.

Pfefferman, Danny. 2002. "Small Area Estimation. New Developments and Directions." *International Statistical Review* 70:125–43.

Pitkin, Hannah. 1967. *The Concept of Representation*. University of California Press.

Poole, Keith T. and Howard Rosenthal. 1984. "The Polarization of American Politics." *Journal of Politics* 46:1061–79.

Pratesi, Monica and Nicola Salvati. 2008. "Small area estimation: the EBLUP estimator based on spatially correlated random area effects." *Statistical Methods and Applications* 17:113–141.

Raftery, Adrian E., Tilmann Gneiting, Fadoua Balabdaoui and Michael Polakowski. 2005. "Using bayesian model averaging to calibrate forecast ensembles." *Monthly Weather Review* 133:1155–1174.

Rao, J.N.K. 2005. *Small Area Estimation*. New York: Wiley.

Rodden, Jonathan. 2010. "The Geographic Distribution of Political Preferences." *Annual Review of Political Science* 13:321–340.

Rosenstone, Steven J. 1983. *Forecasting Presidential Elections*. New Haven, Connecticut: Yale University Press.

Ross, Edward A. 1920. *The Principles of Sociology*. New York: The Century co.

Saalfeld, Thomas. 2008. "Germany: Stability and Strategy in a Mixed-Member Proportional System." In *The Politics of Electoral Systems*, ed. Michael Gallagher and Paul Mitchell. Oxford: Oxford University Press pp. 209–230.

Schwarz, John E. and Barton Fenmore. 1977. "Presidential election results and congressional roll call behavior: The cases of 1964, 1968, and 1972." *Legislative Studies Quarterly* 2:409–422.

Scott, Jacqueline. 1998. "Changing Attitudes to Sexual Morality: A Crossnational Comparison." *Sociology* 32(4):815—845.

Seidman, David. 1975. "Simulation of Public Opinion: A Caveat." *Public Opinion Quarterly* 39(3):331–342.

Selb, Peter and Simon Munzert. 2015. "Forecasting the 2013 German Bundestag Election Using Many Polls and Historical Election Results." Forthcoming at *German Politics*; pre-election version available at: http://dx.doi.org/10.2139/ssrn.2313845.

Selb, Peter and Simon Munzert. 2013. "Forecasting the 2013 Bundestag Election Using Data from Various Polls." Available at SSRN: http://dx.doi.org/10.2139/ssrn.2313845.

Selb, Peter and Simon Munzert. 2011. "Estimating Constituency Preferences from Sparse Survey Data Using Auxiliary Geographic Information." *Political Analysis* 19(4):455–470.

Selb, Peter and Sandrine Pituctin. 2010. "Methodological Issues in the Study of New Parties' Entry and Electoral Success." *Party Politics* 16(2):147–170.

Shapiro, Fred R. 2008. "Quote. . .Misquote." New York Times, July 21, 2008.

Shapiro, Robert Y. and Harpreet Mahajan. 1986. "Gender Differences in Policy Preferences: A Summary of Trends From the 1960s to the 1980s." *The Public Opinion Quarterly* 50(1):42–61.

Sides, John, Ben Highton and Eric McGhee. 2014. "How Election Lab Works." *http://www.washingtonpost.com/news/politics/wp/2014/05/05/how-election-lab-works/* (Last accessed Monday 10$^{\text{th}}$ August, 2015).

Sieberer, Ulrich. 2010. "Behavioral consequences of mixed electoral systems: Deviating voting behavior of district and list MPs in the German Bundestag." *Electoral Studies* 29:484–496.

Sigelman, Lee. 1979. "Presidential popularity and presidential elections." *Public Opinion Quarterly* 43:532–534.

Silver, Nate. 2014. "How The FiveThirtyEight Senate Forecast Model Works." *http://fivethirtyeight.com/features/how-the-fivethirtyeight-senate-forecast-model-works/* (Last accessed Monday 10th August, 2015).

Silver, Nate. 2012. "FiveThirtyEight Blog." *http://fivethirtyeight.blogs.nytimes.com/* (Last accessed Monday 10th August, 2015).

Silver, Nate. 2010. "UK Forecasting Retrospective.". *http://fivethirtyeight.com/features/uk-forecasting-retrospective/* (Last accessed Monday 10th August, 2015).

Smidt, Corwin D. 2014. "Dynamics in Parisanship during American Presidential Campaigns." *Public Opinion Quarterly* 78(S1):303–329.

Snyder, James M., Jr. 2005. "Estimating the Distribution of Voter Preferences Using Partially Aggregated Voting Data." *The Political Methodologist* 13(1):2–5.

Snyder, James M., Jr. 1996. "Constituency Preferences: California Ballot Propositions, 1974-90." *Legislative Studies Quarterly* 21(4):463–488.

Spier, Tim. 2011. Welche politischen Einstellungen haben die Mitglieder der Parteien? In *Parteimitglieder in Deutschland*. VS Verlag für Sozialwissenschaften pp. 121–137.

Stone, Walter J., Ronald B. Rapoport and Alan I. Abramowitz. 1990. The Reagan Revolution and Party Polarization in the 1980s. In *The Parties Respond*, ed. L. Sandy Maisel. Boulder, CO: Westview Press.

Stratmann, Thomas. 2006. Party-line voting and committee assignments in the German mixed-member system. In *Democratic Constitutional Design and Public Policy. Analysis and Evidence*, ed. R.D. Congleton and B. Swedenborg. Cambridge, MA: MIT Press pp. 111–130.

Stratmann, Thomas and Martin Baur. 2002. "Plurality Rule, Proportional Representation, and the German *Bundestag*: How Incentives to Pork-Barrel Differ Across Electoral Systems." *American Journal of Political Science* 46(3):506–514.

Sturtz, Sibylle, Uwe Ligges and Andrew Gelman. 2005. "R2WinBUGS: A Package for Running WinBUGS from R." *Journal of Statistical Software* 12:1–16.

Sullivan, John L. and Daniel Richard Minns. 1976. "Ideological Distance between Candidates: An Empirical Examination." *American Journal of Political Science* 20:439–468.

Thomas, Andrew, Nicky Best, Dave Lunn, Richard Arnold and David Spiegelhalter. 2004. *GeoBUGS User Manual*. 1.2 ed.

Tobler, Waldo. 1970. "A computer movie simulating urban growth in the Detroit region." *Economic Geography* 46(2):234–240.

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 & International Relations* 7:210–214.

Wall, Matthew, Maria Laura Sudulich and Kevin Cunningham. 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.

Wallace, T. D. and Ashiq Hussain. 1969. "The Use of Error Components Models in Combining Cross Section with Time Series Data." *Econometrica* 37(1):55–72.

Warshaw, Christopher and Jonathan Rodden. 2012. "How should we measure district-level public opinion on individual issues?" *Journal of Politics* 74(1):203–219.

Weber, Ronald E., Anne H. Hopkins, Michael L. Mezey and Frank J. Munger. 1972. "Computer Simulation of State Electorates." *Public Opinion Quaterly* 36(4):549–565.

Wegener, Bernd and Stefan Liebig. 1995. Dominant Ideologies and the Variation of Distributive Justice Norms: A Comparison of East and West Germany, and the United States. In *Social Justice and Political Change: Public Opinion in Capitalist and Post-Communist States*, ed. David S. Mason, James Kluegel and Bernd Wegener. New York: Aldine de Gruyter pp. 239–259.

Weidmann, Nils B, Doreen Kuse and Kristian Skrede Gleditsch. 2010. "The geography of the international system: The CShapes dataset." *International Interactions* 36(1):86–106.

Whiteley, P.F. 2005. "Forecasting seats from votes in British general elections." *The British Journal of Politics & International Relations* 7(2):165–173.

Whiteley, P., D. Sanders, M. Stewart and H. Clarke. 2011. "Aggregate level forecasting of the 2010 general election in Britain: The Seats-Votes model." *Electoral Studies* 30(2):278–283.

Wilkes, Rima and Catherine Corrigall-Brown. 2011. "Explaining time trends in public opinion: Attitudes towards immigration and immigrants." *International Journal of Comparative Sociology* 52(1-2):79—99.

Wlezien, Christopher and Robert S. Erikson. 2002. "The Timeline of Presidential Election Campaigns." *Journal of Politics* 64(4):969–993.

Wright, Gerald C., Robert S. Erikson and John P. McIver. 1985. "Measuring State Partisanship and Ideology with Survey Data." *Journal of Politics* 47:469–489.

Yang, Alan S. 1997. "Trends: Attitudes Toward Homosexuality." *Public Opinion Quarterly* 61(3):477—507.

Zaller, John. 2004. Floating voters in US presidential elections, 1948-2000. In *Studies in Public Opinion. Attitudes, Nonattitudes, Measurement Error, and Change*, ed. Willem E. Saris and Paul M. Sniderman. Princeton University Press pp. 166–212.

Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.

Zittel, Thomas and Thomas Gschwend. 2008. "Individualised Constituency Campaigns in Mixed-Member Electoral Systems: Candidates in the 2005 German Elections." *West European Politics* 31(5):879–1003.