The role of polls for election forecasting in German state elections

DISSERTATION

zur Erlangung des akademischen Grades eines

Doktors der Wirtschaftswissenschaften an der Universität Passau

vorgelegt von

M.Sc. Sandra Huber

Passau, September 2017

Disputation: 07.06.2018

Erstgutachter: Prof. Dr. Harry Haupt, Universität Passau

Zweitgutachter: Prof. Dr. Heinrich Oberreuter, Universität Passau

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Voting is a fundamental right of people in a democracy. Article 20 of the German Constitutional Law displays that "all state power emanates from the people." Elections are an instrument to concentrate the people's power to parties' delegates, who represent their opinion in the government.

In this thesis, the focus lies on state elections, which fill the gap of the four years between the national elections, taking place in the years 2013 and 2017. State elections are so called "second-order elections" which were of lower interest for the voters. Hence, over the last years state elections grew in importance, visible by the growing research on state elections and growing interest of the voters (increasing voter turnout rates). They are seen as barometer elections or direction elections for the first order (national) elections. Research of Lock and Gelman (2010) indicates, that elections are decided in swing states in case of the United States (US). During the campaign, the media, public and politicians see elections as competition, who will win the election (see Oberreuter, 2012). To forecast the winner of an election or in the multi-party case the vote share of the parties, election forecasting models are applied. Fundamental models, depending on socio-economic variables like unemployment rate were prevalent over the last years, whereas this thesis investigates forecasts for state elections based on polls. Especially for state election forecasts, fundamental variables on state level basis are not available and variables on national level are not suitable for state elections. The reasons for polls as data base are the German multi-party system, the changing party landscape, blurring borders between the parties affiliations and resulting changes in voting behavior. The German multi-party system is more challenging to forecast because it is more open in its outcome than two-party systems. So called "catch all parties" like the SPD and CDU, who were part of the government since the Second World War became less important in elections (see Münch and Oberreuter, 2015). The number of parties acceding to an election and the number of elected parties grows, which results in more difficult forecasts (see Murr, 2011). The five-year election cycles of state elections cause changing party constellations in the fragmented multi-party systems from one election to the next election. The party landscape varies from election to election and new parties appear first on state level e.g. the Greens or the new party AFD, founded in 2013. The blurring lines between the partys' affiliations give rise to changing voting behavior. The number of undecided, swing and tactical voters increased over the last years and voters decide very late for whom to vote, producing political campaigns with "horse race" character, compare Strauss (2007). Polls emulate the "horse race" character ahead of elections, where the current sentiment of the voter behavior is published.

This thesis focuses on forecasts based on poll data from different institutes. The institutes collect data from respondents and their vote intention for whom to vote, if the election was held next Sunday. The question implicates, that the answer is a snapshot of the current opinion of the voters not a specific forecast, as it implies that the election is held the next days. Especially, on state level, there are sometimes only few polls before election date. As polls are the best available measure before election, but are not available on a daily basis, in this thesis forecasts based on polls are provided for every day up to 120 days before election. Shortly before election, the number of polls increases as polls are of greater interest for the public and politicians. Over the last years, polls improved in quality and quantity with a broader variety of institutes and a higher number of polls in advance of elections even on a regional level (see Traugott, 2014). Hence, polls are a good data base to forecast state level elections in Germany.

Forecasts are provided for 13 states, where the AFD reached the 5% threshold since 2014 starting with the election in Saxony and ending with the Northrhine Westphalia state election in May 2017. To forecast single vote shares in the

German multi-party elections the range of methods varies from basic methods like averaging, over nonparametric regression based methods to dynamic linear models and parametric regression based methods. The aim of this thesis is to find forecasting methods, which provide better forecasts than the poll value as benchmark and are close to the real election outcome. Further, as poll data are only conducted at irregular intervals, a regularized time series structure is generated (on the one hand as benchmark and as data base for some methods, which need a regularized, daily time series).

1.2 Structure

This thesis is organized as follows. Chapter 2 gives an overview of the existing voting and electoral systems, with the German case explained in detail. In Section 2.1 the different party systems (one-party, two-party and multi-party systems) and the electoral systems (plurality, majority and proportional systems) are outlined. Two-party systems like the US are easier to forecast because two parties are clearer to differentiate in their issue-orientation and enable voters an easier decision than multi-party systems. Section 2.2 explains the German personalized proportional representation system for the state elections. Subsection 2.2.3 deals with the German 5% threshold for small parties in comparison to other countries and its importance for the composition of the parliament. On the one hand the threshold protects the parliament against fragmentation but also enables small parties to gain a seat in the parliament. The multi-party system in Germany mostly provides five or six elected parties who have to form (sometimes unstable) coalitions of two or three parties.[1] In Section 2.3 the changing party landscape and voting behavior is discussed. The party landscape in Germany changed from the dominance of the "catch all parties" SPD and CDU to a chance for smaller parties to gain over 20% of the votes like the AFD in Saxony-Anhalt or the Greens in Baden Württemberg with over 30% in the last elections on state level. The party landscape on state

[1] E.g. in the state Lower Saxony the Red-Green coalition of the SPD and Greens was dissolved in August 2017.

level is even more volatile than on the national level. Further, the section deals with the changing voting behavior in elections. The number of undecided and non-voters increased over the last decades with decreasing voter turnout rates (with a slight improvement over the last years). In the 60s and 70s only five percent of the voters decided shortly before an election for whom to vote (see Plischke, 2014). According to the last survey of the Allensbach institute, nearly 50 percent of the voters are undecided one month before the national election in September 2017.[2] Also tactical voting (voting for a party who is seen as the most likely winner or with a potential to reach the 5% threshold) and other types of voting behavior are outlined in this section.

Chapter 3 describes the history of polls and their importance in media, public and politics. Potential errors are discussed in Section 3.2 and also the different types of polls used by the several institutes. In Section 3.3 the role of polls in media, public and for forecasting purposes is explained.

In Chapter 4 the existing literature is discussed. Up to now most research is made for two-party systems. The US is the most studied nation followed by UK and France (see Bunker and Bauchowitz, 2015). Section 4.1 provides an overview of the literature with information about the author, studied nation, forecasting objective, method and variables. According to Lewis-Beck and Dassonneville (2015) there are three types of forecasting models (structural, aggregate and synthetic models) described in this section. The German forecasting literature is outlined in Section 4.2. Most research, based on fundamental variables and regressions is used to forecast the vote share of the incumbent or a coalition, compare e.g. Norpoth and Gschwend (2003).

Chapter 5 describes the data preparation and aggregation in this thesis. For the 13 state elections between 2014 and 2017 irregular polling data have to be customized to generate daily polling data to apply on e.g. parametric regression based models. Also an overview of the polling institutes and the development of the polls on state level over the last decades and the year before the relevant elections is given.

[2] http://www.faz.net/aktuell/politik/bundestagswahl/bundestagswahl-unentschlossenheit-auf-hoechstniveau-15163692.html (Last visit: 2017-09-11).

To forecast state elections, several methods are displayed in Chapter 6. A good forecast is determined by the Golden rules of forecasting according to Lewis-Beck (2005), which are discussed in Section 6.1. E.g. election forecasts should be made with an appropriate lead time. In this thesis a lead time between 120 days and one day is chosen to forecast the respective election on state level. The latest poll value as a benchmark method is also discussed in this section. The next Section 6.2 outlines the different methods used to forecast the state elections. The methods are based on state level polls of various institutes. Methods dealing with the irregular times series structure of the polls are outlined in Subsection 6.2.1. These methods are basic ones like the unweighted average or weighted average, followed by regression based nonparametric methods (with local constant and local linear estimators). Parametric regression based methods like innovation state space models need a regularized time series stated in Subsection 6.2.2. Also dynamic linear models are provided as forecasting methods in this subsection. The average window forecast method in accordance to the work of Pesaran, Pick and Pranovich (2013) is applicable for all methods stated above but only used for the autoregressive integrated moving average method. Section 6.3 gives an overview of common evaluation criteria to measure the performance of forecasting methods. Chapter 7 contains the applied forecasting methods for 13 state elections between 2014 and 2017. First, the technical computation details and aims of this work are summarized in Section 7.1. Section 7.2 discusses the performance of the forecasting methods with suitable evaluation criteria, on different levels of aggregation. Further, the best forecasting methods with respect to the level of aggregation and compared to the benchmark method are outlined in detail. Finally, Chapter 8 summarizes the results and concludes.

2 Voting system and voting behavior in Germany

2.1 Overview of electoral systems and party systems

Forecasting elections is an established research field in the US, UK and France, compare Bunker and Bauchowitz (2015), Jérôme (2013), Küntzler (2014), Lewis-Beck and Dassonneville (2015), Lewis-Beck, Nadeau and Bélanger (2004) and Walther (2015). Election forecasting in Germany is in its infancy. One reason for this fact is the electoral system in Germany. Compared to other countries, Germany has a fragmented multi-party system, whereas in the UK or the US two-party systems dominate the elections, as well in France where only two parties are electable in the second round. Forecasting only two parties is much easier than forecasting coalitions with more than two parties as more options to vote for cause more volatility. In the following, an overview of electoral and party systems is presented and the German system is outlined in detail. The aim is to forecast election results of single parties in the German multi-party system, where more than two parties are usually elected in comparison to other countries like the US. The party system greatly influences the forecastability of the parties and the electoral system determines the translation of votes into seats. Figure 2.1 shows an overview of the different party systems, divided in one-party, two-party and multi-party systems (see Reynolds, Reilly and Ellis, 2005). Some countries like Cuba or China have a dictatorial or autocratic system, where only one party can be elected or governs the country.

In two-party systems like the US or the UK, two parties dominate. Other parties exist in the political landscape, but have no political importance e.g. the Green party in the US or the Liberals in the UK (see Rathore, 2012). Multi-party systems often provide more than two parties, e.g. Germany, where 30 parties were running for the last parliamentary election in 2013. Further, often more parties (in Germany four to six parties) are elected into the parliament. With more parties being elected, it is difficult for one party to gain the absolute majority. Hence, a multi-party election provokes a coalition government, which is not known in advance in many cases. In Germany often stable governments are produced, whereas coalition governments in e.g. Italy or Austria remain only a few years and governments in Israel rely on support of extreme minority groups to form a government. In a democracy, two-party and multi-party systems both work successfully. Two-party systems have several advantages, voters only have to choose between two parties with two different schools of thought. There are no problems of forming a coalition government like in multi-party systems. In two-party systems, the elections provide only one winner, which can be forecasted easily, compare Graefe (2014). There vote expectation surveys are used, which ask who will win the election instead of classical polls asking for whom you would vote. Statistically, in two-party systems, the elected party is the governing party and needs over 50% of the votes, meaning that voting for a party has a 50% chance of winning and electing the governing party. Whereas in Germany voting for a certain party out of nearly 30 acceding parties like in NW has a lower chance of electing the latter winner and voters have to think about possible coalitions in advance. Although there is a theoretical chance for every party to gain votes in an election, the system in Germany restricts the possibilities for a party to reach the parliament by the 5% threshold. Besides the party system, the electoral system also plays a big role for election forecasting, discussed in detail in the following.

Figure 2.1: Overview of party system families and examples for these types. Own illustration based on Reynolds, Reilly and Ellis (2005).

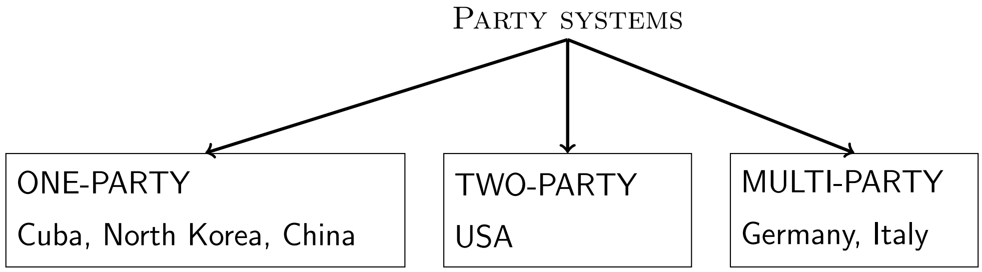
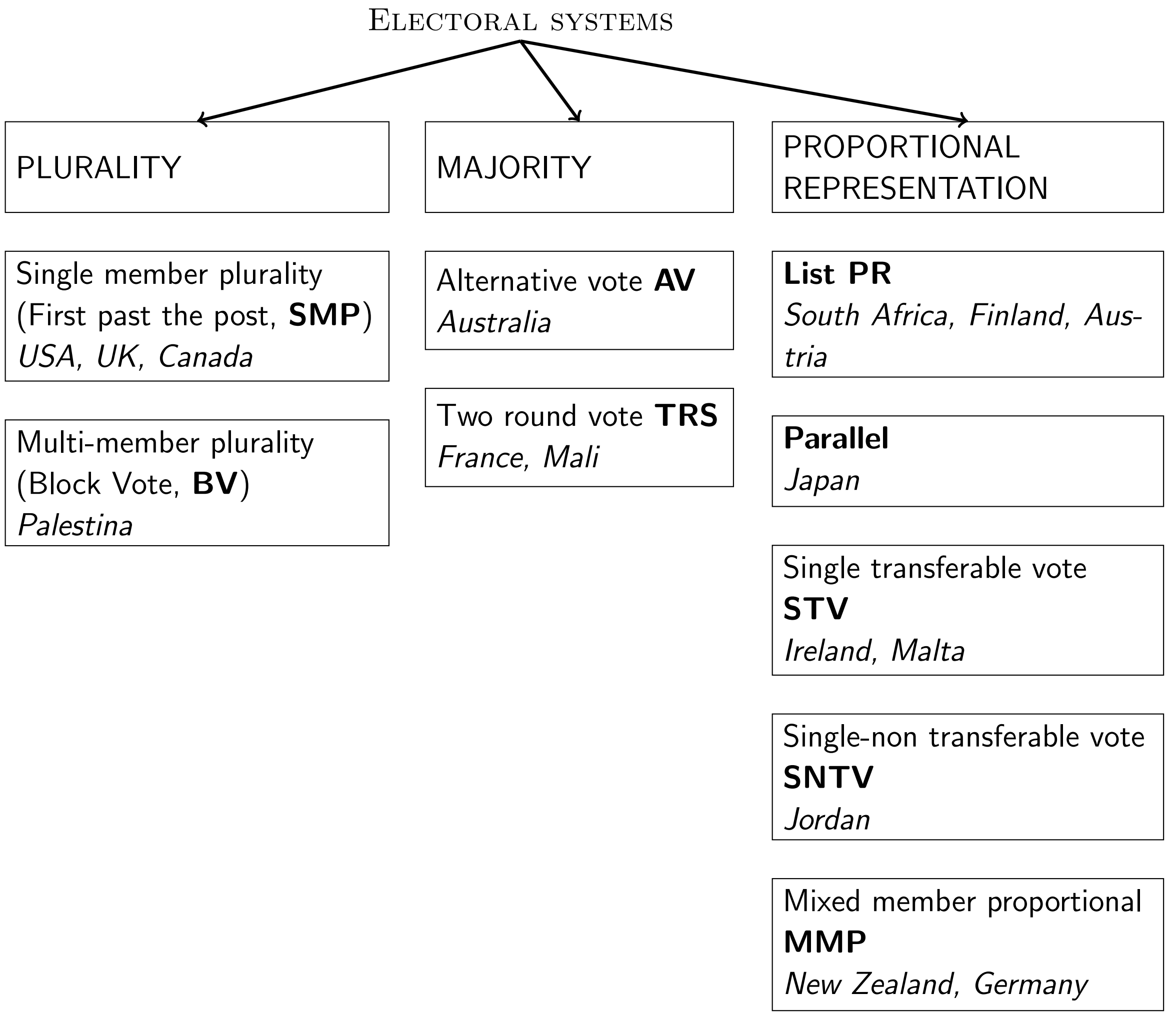


Figure 2.2 provides an overview of the different electoral systems, which can be classified in three major families and nine subgroups inspired by Reynolds, Reilly and Ellis (2005). In the following, the nine subgroups of electoral systems (how votes are translated into seats) are discussed, concerning their advantages and disadvantages. An electoral system concentrates on the design, mechanism and effect of different systems on national, local and supranational level. Besides the translation of votes cast into seats won in an election, the electoral system focuses on the electoral formula, the ballot structure, district magnitude and administrative aspects of elections (see Reynolds, Reilly and Ellis, 2005).

The electoral formula determines, if a system is of majority, plurality, proportional or mixed form. The ballot structure specifies, if a voter makes a single choice or a series of preferences. The administrative aspects of elections influence the distribution of polling places, nomination of candidates and the registration of voters. The electoral system type encourages the way elections are conducted and how voters are registered. Plurality and majority systems produce a two-party system and have a constraining effect on the top candidate of the party, whereas proportional systems provide a greater diversity of parties (multi-party system).

During the 1980s and 1990s there was a general movement towards democratic government and additionally most countries were adding a Proportional Representation (PR) element to their existing plurality system e.g. an Italian referendum in 1993 led to a Mixed member proportional system. Plurality systems are non-

Figure 2.2: Types of electoral systems and subgroups. Illustration based on Reynolds, Reilly and Ellis (2005), p. 28.



proportional electoral systems which are divided in single member plurality (SMP, First past the post) and multi-member plurality (BV, Block Vote). In plurality systems, candidates are elected by plurality not by majority of votes, unlike in majority systems, where candidates need an absolute majority of 50%. These majority systems are splitted in Alternative Vote (AV) and Two Round Models (TRS) like in France. The last family of proportional representation systems provides five (sometimes more) subgroups. The main models are List PR, Single Transferable Vote (STV), Single Non-Transferable Vote (SNTV) and Parallel and Mixed Member Proportional (MMP). In the following, the most important and most common systems are presented with their pros and cons.

Single member plurality (First past the post or SMP) system is the simplest form of plurality systems, where single member districts are used like in the US, Canada, UK and India. This system excludes extremist parties, is simple to understand and to use and provides a clear-cut choice for voters, who only have to choose between two main parties. Furthermore, this system has the advantage that coalitions are the exception and stable governments are provided. Nevertheless, the system has also some disadvantages, especially in terms of discriminating minority groups (e.g. trend to exclude smaller parties, women, ethnic and racial minorities). Third parties often achieve no sufficient level of support and the system generates many wasted votes. A common majority system is the Two Round System used in France and Mali. The election is conducted in two rounds, in the first round all parties participate in the election. If no party gains an absolute majority in the first round, a second round is held a few weeks later. There are two alternatives for the election of the second round. First the majority run off TRS takes place between the two candidates with the highest percentage of votes (e.g. in France the second round is held two weeks after the first one in 2017 between Macron and Le Pen). The second alternative is called the majority plurality TRS, where any candidate who has received more than 12.5% of the votes in the first round, participates in the second round. Regardless, if the candidate receives 50% in the second round, the candidate who gains more votes, wins the election. The advantages of this system are, that the voters can change their mind between the first and the second round and the system enables parties to react to changes in the political landscape. Two round systems are often costly and administratively challenging because of holding two elections and they provide fragmented party systems, which lead to instability and uncertainty. Furthermore, there is often a sharp decline in turnout between the first and the second round.

Proportional representation systems are most common in the world. They are influenced by district magnitude, threshold, list type and apparentement. The district magnitude determines the number of members to be elected in each electoral district. An uneven number of seats works better than an even number and the representation system with the strongest degree of proportionality uses very large districts. The threshold indicates the minimum level of percentage of votes for a party to gain representation. The higher the threshold, the more votes are wasted, but small (extremist) parties have difficulties to reach the threshold. Thresholds in Germany and other countries are described later in Subsection 2.2.3.

Additionally, the list type influences the election system. There are three prevalent types of lists (open, closed and free lists). In a closed list the order of candidates is fixed by the party itself and voters have no influence on the order or selection of the candidates (e.g. in South Africa). In open lists, the voters are free to vote for the party and their preferred candidate. There are several subtypes of open lists varying by country e.g. in Brazil and Finland candidates have no given order, their rank is determined by the number of votes gained in the election. Open lists lead to more extremist candidates e.g. in Kosovo. In free list systems like Ecuador, Switzerland or Luxembourg, the techniques panachage (splitting the votes for more candidates across different parties) and cumulation (assigning more than one vote for a favored candidate) are used. Some countries allow apparentement, this means forming a cartel of small parties to reach high thresholds (often used in Latin America and Israel) (see Reynolds, Reilly and Ellis, 2005). One common system is the Mixed Member Proportional system (MMP, e.g. in New Zealand and Germany), which combines plurality issues with proportional issues.

In MMP systems, the proportional part (List PR) is combined with the plurality part (FPTP). Two votes are splitted, one on a local member (the representative) with the plurality and one on a party, using List PR. This system only leads to few wasted votes and is in favor in terms of accountability, geographic representation and proportionality of translating votes into seats. Regardless of these advantages of the system there are some disadvantages, e.g. the system is very complicated. Also the vote for the party is more important than the vote for the local representative. This leads to changes in the system in regional elections in Germany, where more emphasis is given on the vote for the representative. In Section 2.2, the MMP system is outlined for the German parliamentary election and then discussed in detail for the different German state elections.

2.2 The German multi-party system

2.2.1 Explanation of the system with the example of the German parliamentary election system

As stated in Section 2.1, the German election system is a multi-member proportional system (MMP), called personalized proportional representation. In the history of German elections, the system started as a combination of majority and proportional representation system in 1949 and moved towards the implementation of the 5% threshold in 1956. It proceeded to a modification of the calculation method from D'Hondt to Hare/Niemeyer in 1985 and to the latest change in calculating the mandates in 2009 with the Sainte-Laguë/Schepers method, compare Korte (2009a). An election should be free, equal, secret, direct (without electoral delegates) and universal (18 is the age of electorates and voters), as stated in Article 38 of the German Constitutional Law. The personalized proportional representation system is a combination of the relative majority (first vote) and the proportional representation system (second vote) (see Blumör, Hübner and Maichel, 2011). Every voter has two votes, to elect 598 delegates for the German parliament, 299 directly with the first vote and 299 proportional with the second vote. For the first vote Germany is divided into 299 territorial constituencies. In every constituency the person is elected who has the most votes irrespective of the party membership (relative majority).[3] The 299 other seats of the parliament are allocated by the proportional representation system with the second vote. For example, if a party reaches 40% of the second votes, the party gains at least 40% of the seats in the parliament. There is a rule to except small parties. A party is only elected into the German parliament, if it gains over 5% of the votes with second votes or more than three direct mandates with the first vote. Every winner of the first vote in the constituency gets a seat in the parliament irrespective, if his or her party has reached the 5% threshold.

[3] In the following, six parties are in focus, see Table 2.4 (Greens are abbreviated as GRE, the Left as LIN).

DETERMINATION OF SEATS

The amount of seats reached by every party was calculated according to the Hare/Niemeyer method until 2008, now seats are allocated with the method of Sainte-Laguë/Schepers (see Korte, 2009a). With the method of Sainte-Laguë only second votes, which reach the 5% threshold or three direct mandates are used for allocating seats. The allocation of seats with the Sainte-Laguë method is conducted in four steps:[4]

In a first step, the 598 seats are allocated to the 16 federal states depending on their number of inhabitants. E.g. in the election 2013 there were 74,324,165 inhabitants in Germany, in the state Thuringia 2,154,202 inhabitants. The divisor, calculated by the method of Sainte-Laguë provided a value of 124,050, resulting in a rounded number of 17 seats for Thuringia (2,154,202/124,050$\approx$ 17). All seats of the 16 federal states must sum up to 598.[5] In step two, the second votes in the particular state list are allocated to the seats in this federal state, again with the Sainte-Laguë method. e.g. there were 1,025,123 second votes in Thuringia and 17 seats. The CDU had 477,283 votes and the divisor was 60,000, so the party gained 8 rounded seats. After these two steps, the nationwide minimum number of seats for every party is determined. The minimum number of seats is the maximum number of seats depending on second votes and the won seats per constituency. For every federal state the minimum number of seats for every party is summed up for whole Germany (e.g. if the CDU in Thuringia reaches eight seats through second votes and won nine constituencies, there is one overhang mandate). In Bavaria, the CSU won 45 constituencies and 56 seats through second votes. There are no overhang mandates and the minimum number of seats is 56. Summing up the

[4] See the website of the Bundeswahlleiter for details: https://www.bundeswahlleiter.de/mitteilungen/bundestagswahl-2013/20131009-erlaeuterung-sitzberechnung.html.

[5] The divisor would be exactly 124,287, if the number of all inhabitants in Germany is divided by the number of seats, but then the number of rounded seats would be higher than 598. The Sainte-Laguë method leads to a corrected divisor.

minimum numbers of seats in every federal state per party produces e.g. 242 seats for the CDU and for all parties together 602 seats.

The third step determines the final size of the parliament. Every party should obtain the minimum number of seats and the seats are distributed in proportion to the second votes, gained by every party in all states. Therefore "equalizing mandates" are determined, these display additional seats to compensate the overhang mandates, which are required to guarantee the minimum number of seats and the relation of the second votes between parties.
As seen in Table 2.1, 631 seats are required, in order that every party reaches the appropriate number of second votes per seat, which results in 29 "equalizing mandates", calculated with a divisor from the Sainte-Laguë method. This new method represents the proportion of the parties better than other systems, but provides more mandates. In the following, the election systems of the state elections in Germany are explained in detail.

Table 2.1: Allocation of the number of second votes with a Divisor and the resulting number of seats of the personalized proportional representation system. Illustration based on https://www.bundeswahlleiter.de/dam/jcr/d9a0da9b-f5d1-4043-8452-bf72ed9e86b2/20131009_erl_sitzzuteilung.pdf (Last visit: 2017-04-12).

Party	Minimum number of seats	Second votes	Divisor	Rounded seats
CDU	242	14,921,877		255
SPD	183	11,252,215		193
LIN	60	3,755,699	58,420	64
GRE	61	3,694,057		63
CSU	56	3,243,569		56
Sum	602	36,867,417		631

2.2.2 State election systems in 16 German federal states

Most federal states use the same election system like the parliamentary election system. The states are divided into constituencies like in the parliamentary election but on state level. Table 2.2 indicates an overview of election system, list type, number of votes, seats in parliament and thresholds for the 16 states. [6]

Every election cycle lasts five years in the states (besides Bremen with four years), where the latest referendum in September 2017 confirmed this cycle length. The active right to vote amounts 16 or 18 years, the age of persons eligible to vote (passive right to vote) is 18 years (21 years in Hesse). Most of the federal states (13 out of 16) use the personalized proportional representation system (PPV) like the German parliamentary system.

Only Bremen (HB), Hamburg (HH) and Saarland (SL) use other forms of the proportional system. HB and HH apply a proportional system with open lists (PVo) in contrast to the other states where closed lists are common. Open lists provide more options for voters, as a fixed number of votes is distributed on candidates of different lists or on candidates who are not part of any party. In contrast to closed lists the order of eligible candidates is not determined by the party. In Hamburg, every voter has ten votes, five votes for the candidates of the constituencies and five for the candidates of the state, where every allocation (on candidates or parties) is allowed. In most states the voter has two votes for the parliament, except in Baden Württemberg (BW) and Saarland (SL) the voters have only one vote. In Bremen there is an open list system where the number of votes amounts five per voter and the 83 seats are splitted between the two districts Bremen (68 seats) and Bremerhaven (15 seats). Every voter is allowed to give his or her five votes to a specific person or to a list, but a party is only elected, if it passes the 5% threshold in a district. In Hamburg, the number of votes amounts ten, five for the

[6] List of abbreviations of the federal states:

Abbr.	State	Abbr.	State
BB	Brandenburg	NW	Northrhine Westphalia
BW	Baden Württemberg	RP	Rheinland Palatinate
BY	Bavaria	SH	Schleswig Holstein
HB	Bremen	SL	Saarland
HE	Hesse	SN	Saxony
HH	Hamburg	ST	Saxony-Anhalt
MV	Mecklenburg Western Pomerania	TH	Thuringia
NI	Lower Saxony	WBE	Berlin

candidate on the district list and the other five for the candidates on the state list (both lists are open).

Table 2.2: Overview of state election systems in 16 states with election cycle length, election age, electoral system, number of votes, mandates in the parliament and rules for the 5% threshold. Illustration based on http://www.wahlrecht.de/landtage/index.htm (Last visit: 2017-05-03).

State	Term	Election age active/passive	Electoral system	Number of votes	Mandates (in constituency)	5% threshold
BB	5	16/18∗	PPV	2	88 (44)	Nationally (not for Sorben)
BW	5	18/18	PPV	1	120 (70)	Nationally
BY	5	18/18	PPV	2	180 (92)	Nationally
HB	4	16/18∗	PVo	5	83 (0)	Splitted for Bremen and Bremerhaven
HE	5	18/21	PPV	2	110 (55)	Nationally
HH	5	16/18∗	PVo	10	121 (71)	Nationally
MV	5	18/18	PPV	2	71 (36)	Nationally
NI	5	18/18	PPV	2	135 (87)	Nationally
NW	5	18/18	PPV	2	181 (128)	Nationally
RP	5	18/18	PPV	2	101 (51)	Nationally
SH	5	18/18	PPV	2	120 (60)	Nationally (not for SSW)
SL	5	18/18	PV	1	51 (0)	Nationally
SN	5	18/18	PPV	2	91 (45)	Nationally
ST	5	16/18∗	PPV	2	69 (35)	Nationally
						Nationally
TH	5	18/18	PPV	2	88 (44)	Nationally
WBE	5	18/18	PPV	2	130 (78)	Nationally (including invalid)

∗ new electoral systems: voting is allowed at the age of 16 (in BB since 2012, HB since 2011, HH and ST since 2013)
PPV = personalised proportional voting, PVo = proportional voting with open lists (votes can be given to a state list or single candidates), PV = proportional voting.

As Table 2.2 indicates, there are different sizes of the state parliaments, ranging from 51 mandates in SL to 181 in Northrhine Westphalia (NW). The numbers in rounded brackets in the column "Mandates" show the number of seats in the constituencies. The threshold, which excludes small parties from parliament, amounts 5% in nearly all states, except Schleswig Holstein (SH) and Brandenburg (BB). In Schleswig Holstein, the SSW (Südschleswigsche Wählerverband, the Danish minority) and in Brandenburg the Sorben can reach the parliament without passing the 5% threshold to enable minorities being part of the parliament. In Berlin (WBE) the 5% threshold is calculated due to all casted votes including the invalid votes. The focus here lies on state elections, where the AFD gained the first time representation in the parliament, who are up to now Brandenburg (BB), Baden Württemberg (BW), Bremen (HB), Hamburg (HH), Mecklenburg Western Pomerania (MV), Northrhine Westphalia (NW), Rhineland Palatinate (RP), Schleswig Holstein (SH), Saarland (SL), Saxony (SN), Saxony Anhalt (ST), Thuringia (TH) and Berlin (WBE). All states choose one of the three popular types, Hare/Niemeyer, D'Hondt and Sainte-Laguë of allocating votes into seats. The Sainte-Laguë method was explained in Section 2.2 and is used in BW, HB, HH, NW, RP and SH. Despite, BY, WBE, BB, HE, ST and TH are using the Hare/Niemeyer method and NI, SN and SL use the D'Hondt method.

2.2.3 Five percent threshold in comparison to other countries

The electoral system enables small parties to gain representation but there is also a threshold of 5% to limit the access to the parliament in order to restrict the number of parties. Reynolds, Reilly and Ellis (2005) differentiate between formal and effective/natural thresholds. Legally imposed formal thresholds exist in Germany or New Zealand to restrict access for extremist parties, with an exception if parties win enough seats in a constituency (three in Germany). In contrast to formal thresholds, effective thresholds exist due to mathematical reasons. In Table 2.3 thresholds of different countries including Germany with a 5% threshold are listed. The thresholds in Europe vary from no threshold in the Netherlands or France to a 10% threshold in Turkey.

In Germany, state election thresholds amount 5% with exceptions in SH and BB, see Table 2.3. The 5% threshold in Germany was first introduced for only one state to restrict the fragmentation of the parliament in 1949 (see Probst, 2014). In 1953 this hurdle was extended to whole Germany. A party had to reach 5% in all states, not only in one. In 1956 the rule of three mandates was applied additionally to the 5% threshold. From 1957 to 1983 the number of parties in the German parliament was mostly four.

In Germany, parties reaching under 5% of the second votes or less than three direct candidates are not part of the parliament. An exception is provided, direct candidates of the first vote always gain a seat irrespective if their party is in power due to the second vote.

A threshold is important to enable small parties to gain a seat in the parliament. Further, the threshold influences the voting behavior, e.g. it leads to tactical voting. If a party is near the 5% threshold, voters are more willing to vote for this party. The other possibility is that voters refuse to vote for the party, if there is no perspective that a party reaches five percent of the votes and the votes are wasted. In state elections, small parties often reach the 5% threshold e.g. the Greens in the 1980s in Bremen. In some cases, these parties also gain more than 5% in the following parliamentary elections like the AFD, who was first part of 13 state parliaments and is now part of the national parliament.

Table 2.3: Countries and corresponding thresholds for elections. Own illustration based on http://www.europa-links.eu/fakten/die-wahlverfahren-in-den-landern-europas-199/ Kulke (2010).

Country	Electoral threshold
Austria	4% threshold
Belgium	5%
Bulgaria	4%
Cyprus	1.79%
Czech Republic	5%, 7% party coalitions, 11% more than 3 parties
Denmark	2% or 1 direct mandate
Estonia	More than 5%
Finland	No formal threshold, a lot of parties possible
France	No threshold
Greece	3%
Hungary	5%, 10% for party coalitions and 15% for party groupings
Iceland	5%
Italia	4%, 10% for party groupings
Lativa	5%
Lithuania	5% for parties or 7% for party groupings
Luxembourg	5%-10%
Netherlands	No threshold
Norway	4%
Poland	5% for parties or 8% for party groupings
Romania	5% for parties or 8%
Russia	7%
Sweden	4%
Slovakia	5% for parties or 7% for two-party groupings
Slovenia	4%
Spain	3%
Turkey	10%

2.3 Changing party landscape and voting behavior in Germany

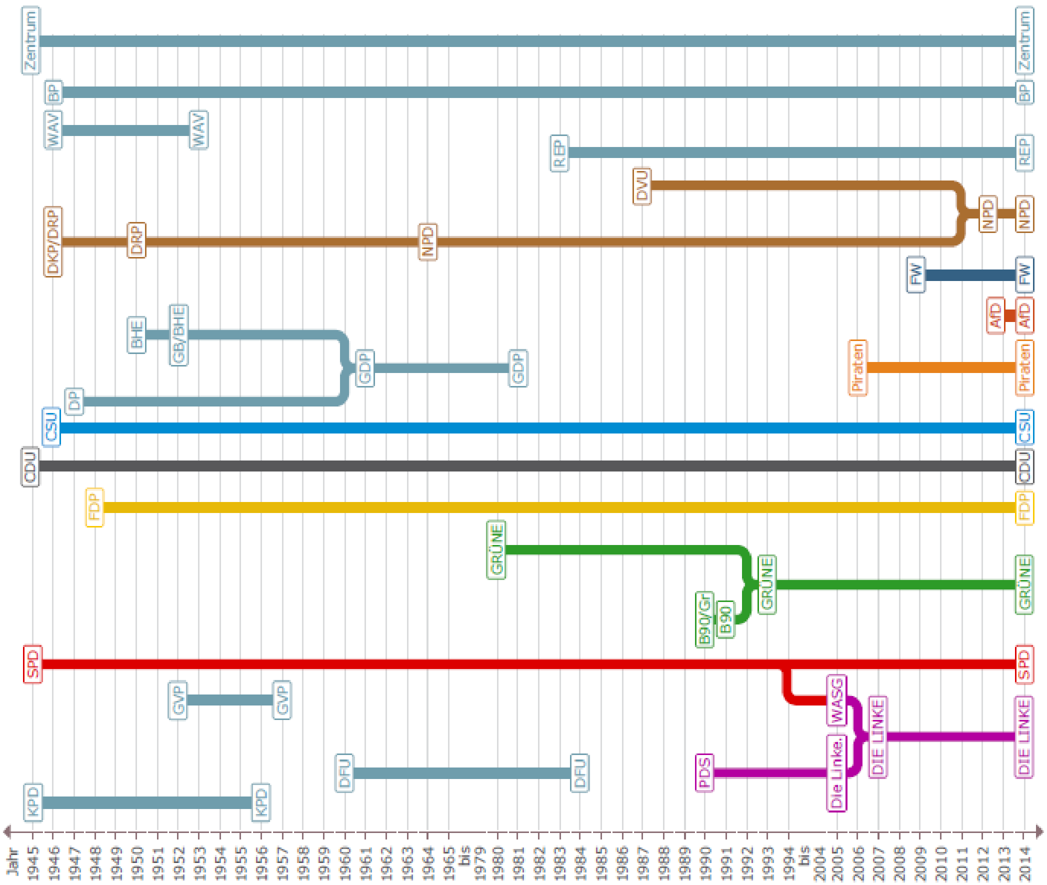
As seen in Subsection 2.2.3 the 5% threshold was introduced to exclude extremist parties from parliament. In the history of elections the voting behavior and party structure varied over the last decades. Especially in state elections small parties emerged at first on a regional level and in state elections before they appear in parliamentary elections. In the following, the history of the parties along with the changes in the party landscape, the second-order elections, their importance and the changing voting behavior are outlined in detail.

2.3.1 History of parties and the resulting changes in the party landscape

In the 1950s, the German multi-party system was highly fragmented with many small parties represented in the parliament compare Figure 2.3, where the development of parties is shown.

With the "economic miracle" in the 1960s, the party system focused on two or three established parties (CDU, SPD and FDP) where the economic situation was attributed to the governing party, compare Blumör, Hübner and Maichel (2011). The two "catch all parties" CDU and SPD gained together 90% of the votes in the 1970s (see Münch and Oberreuter, 2015). In the 1980s, the deconcentration process started with the Greens, who represented new issues (ecological, environmental), which were neglected by the established parties, the so called "popular parties". This led to a four-party system in Germany. After the German reunification 1989, the fragmentation process proceeded and was extended to a five-party system with populist parties like the PDS, WASG or right-wing parties (see Korte, 2009a). With the refugee crisis in Europe, the policy faces new challenges, which new parties exploited to gain votes (see Steinmayr, 2016). The party Front national in France, the FPÖ in Austria or the AFD in Germany benefited from the shift to the right due to the refugee crisis. In the last elections, mostly five parties played a role in the political landscape. In the latest parliamentary election 2017, six parties were elected and the two main parties CDU and SPD only reached about

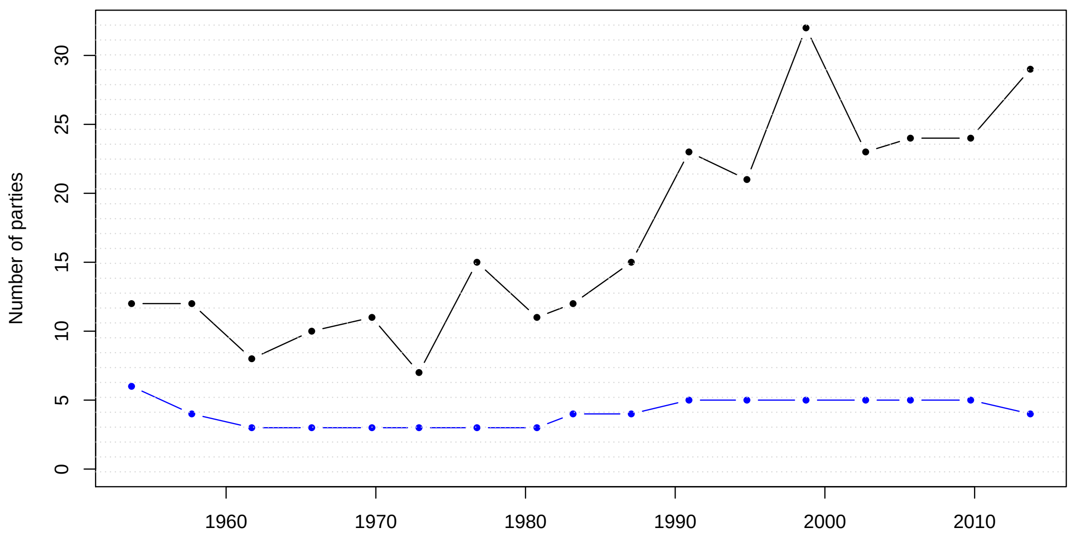
Figure 2.3: Development of most important parties in Germany between 1949 and 2014. Illustration based on http://www.bpb.de/politik/grundfragen/parteien-in-deutschland/138661/entwicklung-des-parteiensystems (Last visit: 2017-06-10).



50% of the votes. Also the number of parties competing in an election increased over the last decades in the parliamentary elections from 14 parties in 1949 to 30 in 2013, see Figure 2.4. The black line indicates the number of parties competing in an election, whereas the blue line shows the number of elected parties passing the 5% threshold. The fragmentation of the party landscape grew steadily and reached its peak in 2000 with more than 30 parties competing.

For state elections, Figure 2.5 indicates a similar picture. For NW, RP, BW, ST, MV, WBE, SL and SH the number of parties acceding to an election (lines with

Figure 2.4: The black line shows the number of parties competing in an election, the blue line the number of elected parties between 1949 and 2013 in parliamentary elections.

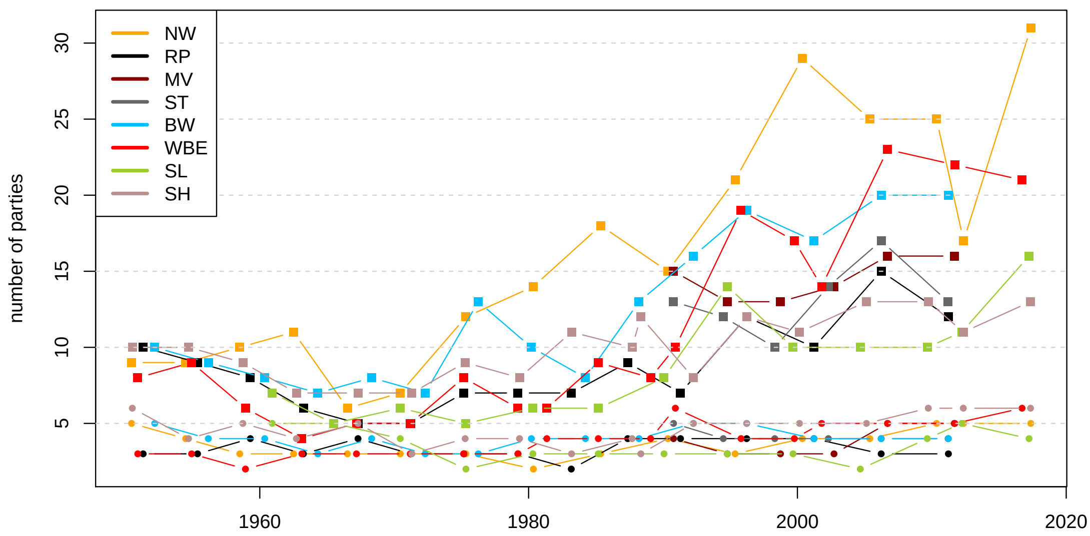


squares) increased up to over 30 in NW in May 2017, the number of parties elected increased to six in the latest election (lines with dots).

Although party identification became less important over the last years, some parties were connected with certain voter groups e.g. the CDU with catholic voters or the SPD with the working class, see Table 2.4 and Korte (2009a). Over the last elections, the CDU and SPD could not be clearly distinguished in their party programs compare Malisch (2013), as seen in the ballot, concerning the permission of the same sex marriage in July 2017. The CDU and SPD were oriented towards the same direction because the sentiment of the voters was gay marriage. As parties orient more on voters and election dates, the party programs and campaign strategy changed over the last elections explained in detail in Subsection 2.3.3.

Table 2.4 shows the acronyms of the parties, the whole name and the orientation of the party, presumed in the population. The parties in Table 2.4 can be divided in established parties like the CDU, SPD and FDP, who were founded at the end of the 1940s and had a clear direction in their party programs (stated in column

Figure 2.5: Number of parties competing in an election (squares) and number of parties elected (dots) in the states (BW, RP, ST, MV, WBE, SL, SH and NW), where the last elections in 2016 and 2017 took place.



"Affiliation"). In history new parties emerged through handling new issues, which were interesting for the voters. Beginning with the Greens in the 1980s picking up ecological issues, the Lefts addressed the problems of social inequality. The AFD is the newest party, founded in 2013 in the shadow of the refugee crisis and Anti-Euro sentiment in Europe.

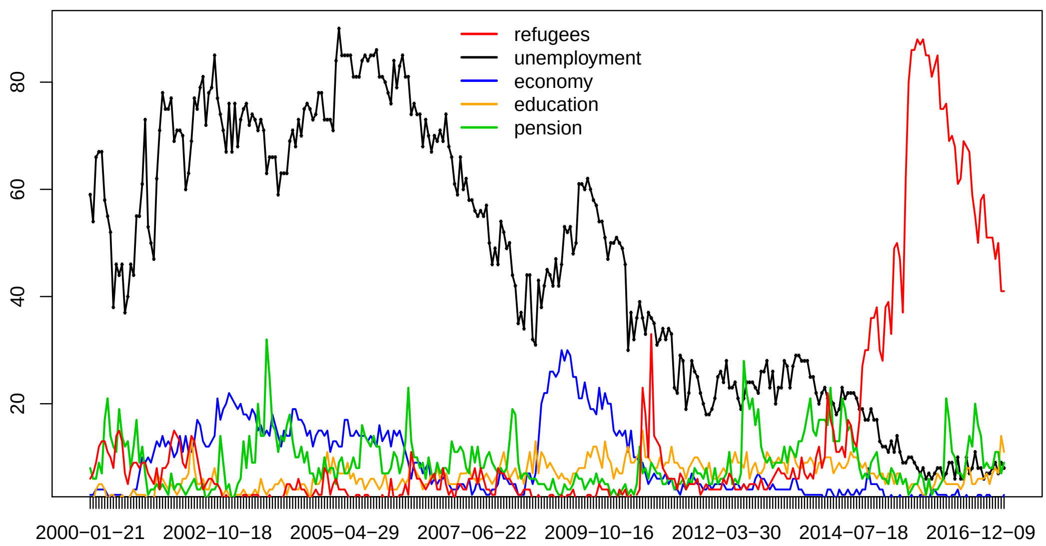
Table 2.4: Acronym, party name and approximate party affiliation. Illustration based on Jérôme (2013, p. 477).

Acronym	Party	Affiliation
AFD	AFD: Alternative for Germany	Anti-euro party
CDU	CDU: Christian Democratic Union of Germany	Center-right
FDP	FDP: Free Democratic Party	Liberal-right
GRE	GRÜNE: Federation 90/The Greens (1980-1987: The Greens)	Ecologist
LIN	LINKE: The left Party.PDS (1990-2002: Party of Democratic Socialism, PDS)	Left-anti-capitalist
SPD	SPD: Social Democratic Party of Germany	Social-democratic

Issue orientation is of growing importance for the voters. The "Forschungsgruppe

Wahlen" asks the respondents several questions, e.g. the vote intention question, for whom to vote for, if the election was held next Sunday. Another question deals with the most important problems in Germany, the respondents are asked what is currently the most important problem for them (choosing between the issues "refugees", "pension", "unemployment", "economy" and "education"), response is given in points e.g. 80 for the refugees, 10 for the unemployment rate. In the 2000s, the unemployment rate (black line) influenced the voting decision most, as can be seen in Figure 2.6, whereas since 2013 the refugees (red line) are in focus.

Figure 2.6: Development of the most important problems between 2000 and 2017 in points, published by Forschungsgruppe Wahlen (FGW). Illustration based on data from the FGW (Forschungsgruppe Wahlen) by order of the ZDF http://www.forschungsgruppe.de/Umfragen/Politbarometer/Langzeitentwicklung_-_Themen_im_Ueberblick/Politik_II/.



Small parties grew in importance over the last decade as they picked up the current issues, which were important for the voters. In state elections, these new parties appeared prior to the parliamentary elections.

2.3.2 Importance of second-order elections for national elections

Most research on elections deals with national elections. There is only sparse research for state elections due to insufficient interest of voters in local elections seen on lower turnout rates in comparison to national elections. Despite, some state elections are of greater interest than others, e.g. Northrhine Westphalia is a state of major importance, due to the higher number of inhabitants. State elections are called "second-order elections" or "barometer elections" which influence first order elections and display the voters sentiment prior the main election (see Anderson and Ward, 1996). Although as Krumpal and Rauhut (2008) tell that 'less is at stake' in second-order elections, they play a big role in the electoral cycle in Germany, especially when they are close to the parliamentary election. In the US some states like Ohio and Nevada are so called "bellwether" states, which show the winner of the national elections in most of the cases, compare Jérôme and Jérôme-Speziari (2011).

Also in Germany state elections are seen as "direction elections" for the national ones and are therefore in focus of this work Krumpal and Rauhut (2008). The voter turnout rates are lower in state elections than in first order elections, but voters are more willing to elect smaller parties (like the AFD), who then appear in the parliamentary election in weakened form. In second-order elections voters behave with abstaining (vote in national elections but not in local), switching (change their voting decision between national and local elections) or partisanship (vote for the same party in national and local elections). Often voters send a signal to a party by voting in local elections, called protest voting or changing their mind due to ideological issues, called sincere voting (see Hobolt and Spoon, 2012). Voters reward the local incumbents for high growth and punish them e.g. for high unemployment rates, especially in local elections (see Khemani, 2001), as they blame the party or candidate for the economic or social situation. Some voters change their minds, if a party puts more weight on issues which are interesting for the voters like the AFD did with the refugee crisis.

The behavior of non-voters in second-order elections is distinguished between selective abstention (abstaining not in every election) and split ticket voting (voting for different parties in different elections), compare Degan and Merlo (2011). In history, regional parties emerged due to situations, e.g. the Greens were elected for the first time in the state election 1979 and represented ecological issues like an anti-nuclear attitude. In Hamburg, the only regional occurring "Schill"-party reached over 19% in 2001 but did not compete for the national election. In 2011 the party "Pirates" became part of the state parliament in Berlin and in three other parliaments. In the last years right wing parties like the AFD in Germany gained in importance due to the refugee crisis in Europe, also in other countries like France (the party of Marine Le Pen reached the second highest share of votes in the first round of the last election). As shown in Section 2.2 the personalized proportional representation system in Germany is a multi-party system, where is little chance for a single party to gain absolute majority, resulting in coalitions, compare Korte (2009a) and Reynolds, Reilly and Ellis (2005). In the following, the changing voting behavior and its influence on the elections is discussed in detail.

2.3.3 Changing voting behavior in Germany

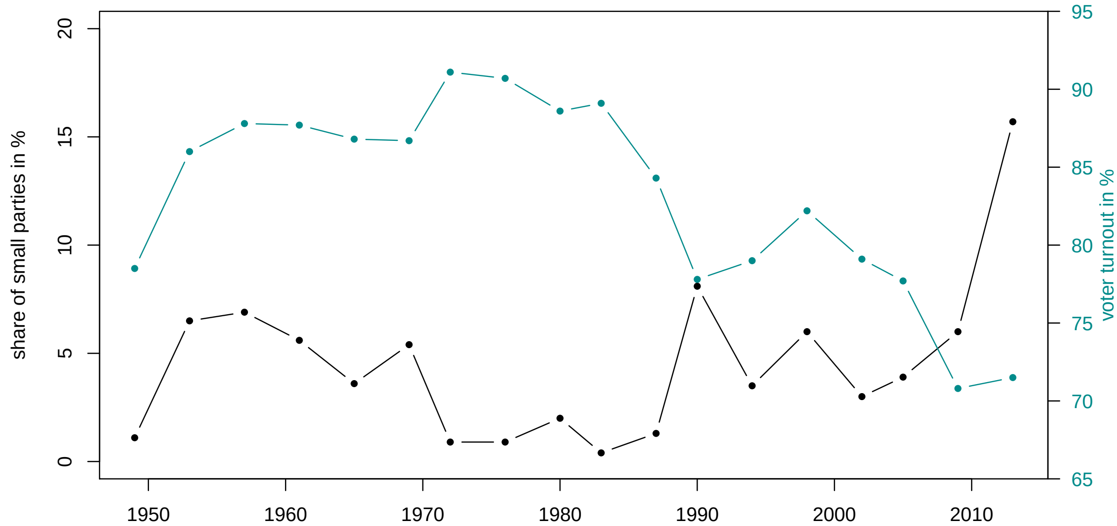
The voting behavior is subject to structural and social changes of parties and voters. The identification with parties dissolved over the last decades and the number of party members shrinked over time. Voters change to non-voters, are more willing to vote for smaller parties or swing to other parties, compare Oberreuter (2012). From the 1950s to 1980s the major part of the vote shares was split between the two popular parties CDU and SPD. With the rise of the Greens in the 1980s other parties had a chance to reach an amount of the votes. The so called "small parties" gained in importance over the last years and led to smaller percents for the two popular parties (see Münch and Oberreuter, 2015). Changes over the last elections are due to changes in party landscape and the voting behavior in the population. The party programs occurred due to political cleavages of the 19th and 20th century which led to the different parties (liberal, christ-democratic and socialist). These cleavages are now modernized and new parties, who deal with new issues, emerge (see Korte, 2009a).

The parties were subject to several further changes since the 1990s and the number of party members decreased steadily. The relationship between society and politics dissolved over the last years, as the party programs are oriented less towards the existing cleavages and not the parties own affiliations. Parties respond to different living environments in the society in order to reach a large amount of voters with several issues (see Oberreuter, 2015). As milieus are diversified in continuously smaller groups and can not be clearly separated, it is difficult for great parties to persuade all voters in comparison to smaller parties. These react to new milieus and interests of the voters, compare Oberreuter (2015). If the existing parties can handle new issues, voters remain, if not, new parties can persuade voters. The party programs are now reacting on the value change in society and base their programs on satisfaction of the voters interests, instead of their own traditions.

Even the behavior of voters is subject to several changes. The party identification became less important over the last decades (seen in declining party-memberships) and the interest of voters in politics changed significantly. Voting for a specific party due to clear conviction of the party and its issues is not of primary importance. Voters change their minds very often during the election cycle and decide very late before an election (see Hillygus, 2011). Even in state elections the behavior of voters is very volatile and they are more willing to vote for new parties. Furthermore, voters punish or reward the government for the current situation which is incorporated by researchers in their models in terms of variables called "government approval", "incumbency", etc. see Chapter 4. By voting for smaller parties, especially the ones below the 5% threshold, votes are wasted.

Also the voter turnout influences the outcome of an election and the composition of the parliament. As seen in Figure 2.7, the voter turnout in parliamentary elections increased since the 1950s to the highest level in the 1970s and decreased until 2013 with some ups and downs as the blue line indicates. The black line shows the vote shares of parties who gained under 5% of the votes and were not part of the parliament, so called "wasted votes". In the election 2013 the smaller parties summed up to 15%. The correlation between the voter turnout and the small parties amounts -0.55 over the last 50 years. This means, the lower the voter turnout rates are, the higher is the vote share of small parties.

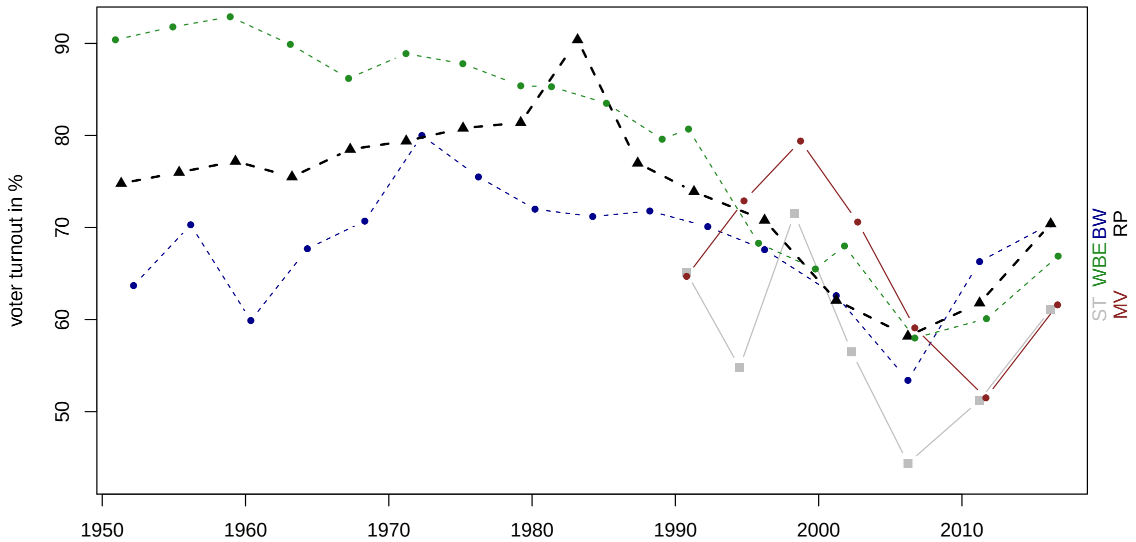
Figure 2.7: The blue line indicates the voter turnout rates of parliamentary elections from 1949 to 2013 and the black line the share of small parties, who gained under 5% representation in the election.



The development of the voter turnout rates in state elections is similar to the development in parliamentary elections. Voter turnout rates for the five states, where elections took place in 2016 are presented in Figure 2.8. The dashed lines show the voter turnout rates for the former West German states Baden Württemberg (BW), Rhineland Palatinate (RP) and Berlin (WBE), whereas the former East German states Saxony Anhalt (ST) and Mecklenburg Western Pomerania (MV) are drawn solid.

On state level, the voter turnout rates in the reported states decreased in the 1990s and 2000s, but increased over the last cycles. In comparison to the turnout rates on the national level the state rates show more variation depending on specific elections. In East German states like Saxony Anhalt, voter turnout rates are often lower than in other states(see Schäfer, 2013). Over the last elections, voter turnout rates increased in all five states with values between 60 and 70 percent (in

Figure 2.8: Development of voter turnout rates in the three states BW, ST, RP (dotted lines) from 1950 to 2017 and the two states MV, WBE (solid lines) from 1990 to 2017.



the last national elections the rates were over 75%.

Figure A.1 in the Appendix shows the voter turnout rates for 13 states, investigated in this work. The rates range from over 90% in the 1970s to the lowest value of 45% in 2013. Generally speaking, the voter turnout rates in the former East states indicate lower rates than e.g. the West German state Rhineland Palatinate (black dotted line). The dotted lines are the rates from 1950 until today (the former West German states). The solid lines range from 1990 to 2017. The dots and triangles display the election dates in the states. There are declining rates in the 1980s, but over the last years, there can be seen a slight increase especially in the 2017 elections.

Figure A.2 in the Appendix shows the voter turnout rates in form of boxplots for every year. Each boxplot includes the range of voter turnout rates of the state elections, which took place in the specific year. The box contains 50% of the values (the start of the box is the 25% quantile, the end of the box the 75% quantile).

The line in the box indicates the median value of the voter turnout rates in the specific year.

As the voter turnout rates decreased, the number of non-voters increased over the last decades. This led to a change in the political landscape. The literature differentiates between four types of non-voters, compare Schäfer (2013). The first type, the artificial non-voter is a voter, who is unable to participate in the election due to illness, vacation, etc. Fundamental non-voters do not vote by conviction, whereas the other type of non-voters (confirming non-voters) do not go to election because they are not confident with the government or the parties. The last type are economic non-voters, who only go to election if they can change the government and influence the economy with their vote. Since voting is not longer seen as a duty for citizens in a democracy, the voter turnout rates decrease. According to Korte (2009a) the voters are sullen with the parties or the government. They are dissatisfied with the political system or with the current social and economic situation. The reasons for abstaining from voting stated above with the four types of non-voters, are divided into several models explaining the voting behavior.

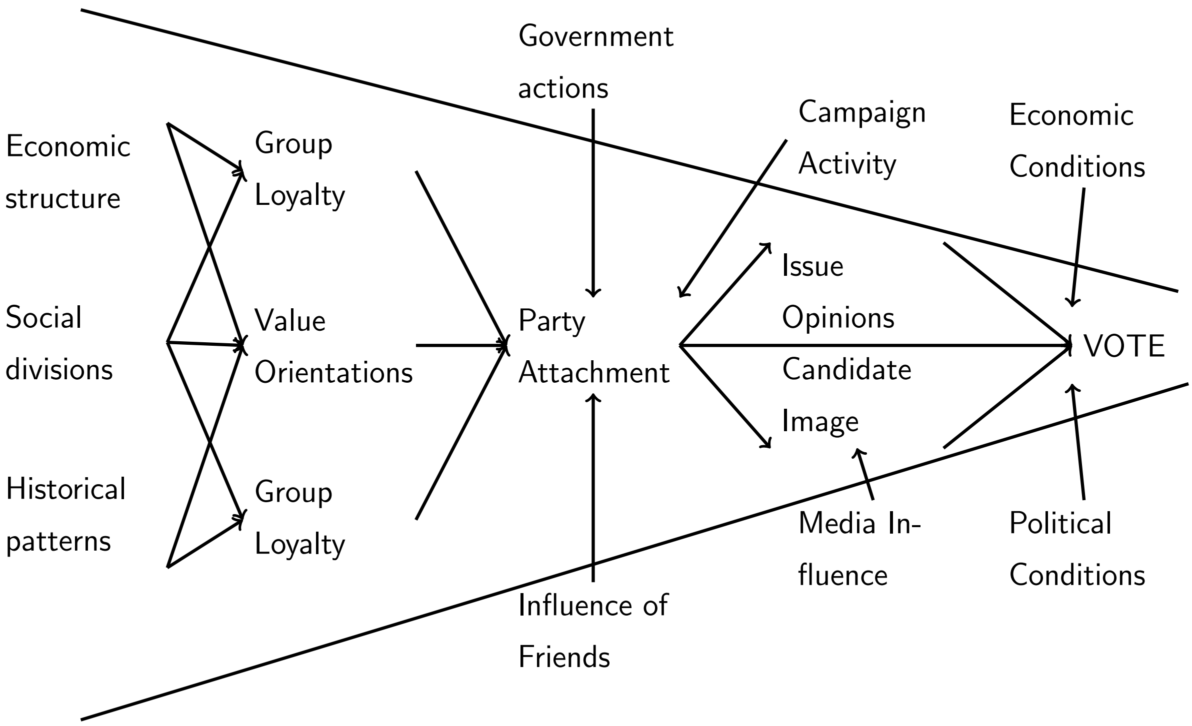
The literature differentiates between three models, the sociological model (Columbia Model), the psychological model (Michigan Model) and the model of rational choice (Rochester Model).[7] The University of Columbia developed the sociological model in the 1940s (see Lazarsfeld, Berelson and Gaudet, 1944). Voter behavior is explained by the membership to social group characteristics as the socioeconomic status, religion or place of residence, which influence the vote decision. To explain voting behavior, the sociological model is based on individual data like milieus and group structures, e.g. income, age, religion, profession, education, etc. (see Korte, 2009b).

The Michigan model was developed by Campbell et al. (1960) as a more dynamic model to explain voting behavior. There are three factors which influence the

[7] http://www.bpb.de/politik/wahlen/bundestagswahlen/62613/theorien-des-waehlerverhaltens?p=all (Last visit: 2017-05-02).

voting behavior. The long term variable party identification, candidate orientation and issue orientation, which are asked in regular polls of the Forschungsgruppe Wahlen. The last approach is called model of rational choice, developed by Downs (1957), the Rochester Model. The voting decision is based on rational choice and voters punish or reward the government and vote for the party who handles their problems best. The Michigan model was extended by Dalton (2013) to the "funnel of causality" presented in Figure 2.9. The long term factors (economic structure, social divisions and historical patterns), influencing the vote are important at the start of the campaign. The closer the election day, short term factors like media, friends, economic and political circumstances influence the voting decision. The funnel is very wide at the beginning and voters behave depending on their long term attitude. Later other factors determine the vote until the voters make their final decision. Today, voting decision takes place ever later in the campaign due to the rising number of short term voters, non-voters, protest voters, swing voters and undecided voters as Plickert (2013) writes. Whereas in the 1960s and 1970s there were only 5% of the voters, who decided very late before the election, compare Plischke (2014).

Figure 2.9: Influence factors of voting decision explained in the funnel of causality. Illustration compare Dalton (2013), p. 184.



As the institute Infratest dimap reveals, the voters decide very late for whom to vote and the parties try to persuade these undecided voters.[8] Voters abstain, if their party has no chance to pass the 5% threshold or they vote tactically for a party to form a coalition like in the 2013 parliamentary election as Niggemeier (2013) writes. The FDP mobilized the voters to give them their second vote for a coalition with the CDU. With growing importance of polls for political decisions and short term campaign effects, polls are in focus of this work to forecast elections results.

[8] see https://www.infratest-dimap.de/service/faqs/ (Last visit: 2017-08-13).

3 History, development and importance of the polls in Germany

In the history of election forecasting, the prevalent type of models are fundamental models based on variables like unemployment rate, Gross Domestic Product or results of the last election. Further details of the methods for election forecasting are presented in Chapter 4. The political landscape and voting behavior see Subsection 2.3.3 became more volatile over the last election cycles, compare Leigh and Wolfers (2006), Münch and Oberreuter (2015) and Wlezien et al. (2013). The volatility of elections can be seen especially on a regional level.

The focus in this work primarily lies on polls as data base, because fundamental variables are not able to cover the short term changes in voting behavior during the campaign. As Oberreuter (2012) writes, the main parties have to deal with different living environments, whereas small parties use current issues for their party programs to mobilize e.g. non-voters or undecided ones and pick up voters from other parties. Furthermore, the polls grow in importance for the policy making process as politicians design their policy at times due to the current sentiment of voters. Parties react on polls, and polls show the reaction of voters on the parties campaigns. According to these reasons, the number of polls increased and also the number of institutes is greater diversified. In the last years the number of regional polls grew and therefore also the forecasts based on state level polls, compare Blumenthal (2014), Christensen and Florence (2008), Holbrook and DeSart (1999) and Traugott (2014). In the following, the history and evolution of polls is outlined.

3.1 History and current significance of polls

Before using polls in conjunction with elections, knowledgeable observers with reputation, political insiders or bellwether states (states which are indicative of the trend in the whole nation) were used to forecast elections (see Hillygus, 2011). In the 1930s media used polls for the first time to show the voters preferences for the US elections, compare Kruke (2014) and Traugott (2014). Journalists published the polls additional to their own opinion and developed the so called "horse race journalism" for the elections. The polling results were reported to show the current sentiment of the voters and to support the campaign of the parties with numbers to increase the tension in the run up to the elections (see Traugott, 2014). Particularly in two-party systems like the US, the question who will win the election is of great interest, as the elected party is the governing one after election. Whereas in multi-party elections, the vote shares of all parties are of interest, in the case of later coalition negotiations.

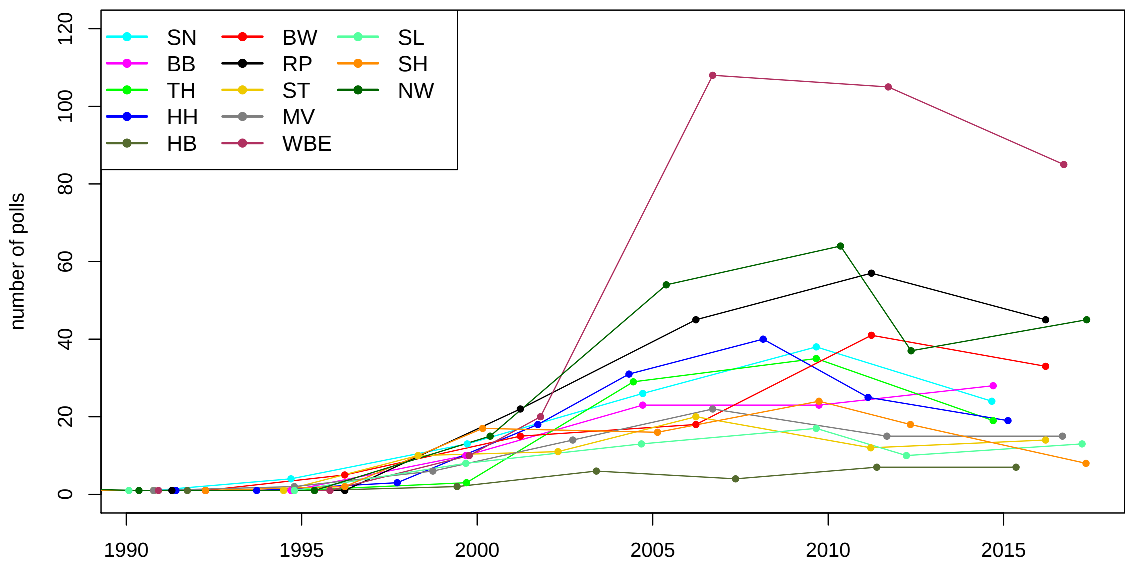
In other countries (UK, Sweden, Australia and Canada) polling started in the mid 1930s and 1940s, initiated by the institute of George Gallup who conducted the first opinion polls in 1936 for the US election. Gallups poll, only based on a few thousands potential voters, beat the Literary Digest extensive poll with millions of respondents.

In Germany, polling institutes were founded after the Second World War, but the first regular polls in the run up of elections were held at first in the 1960s and 1970s. Since the 1980s, the number of polls increased in the US with replacing the expensive face-to-face interview by telephone surveys, compare Blumenthal (2014). In Italy, the number of polls grew in the 1990s (see Castro, 2013). Also in Germany, since the 1990s the polling institutes published more polls due to telephone and online surveys which are cheaper and less time-consuming than face-to-face. Polls are not only published and explained by the media, websites or blogs comment and use polls as well and provide new information sources for voters.

As Figure 3.1 indicates, the number of polls from "1990-01-01" up to "2017-05-14" increased steadily in the run up of the specific elections. The election dates in the states are shown with dots. The number of polls increased extremely in the 2000s due to online surveys in all states. In Berlin (WBE) and Northrhine

Westphalia (NW), the number of polls grew even more, as these states are of special interest for the public. WBE is the state of the highest interest as it is near the government in Berlin and NW is the state with the largest population and so interesting as a potential bellwether state. The SPD vote share decreased in the latest state election in Northrhine Westphalia and now also decreased in the national election. Another reason for the growing number of polls over the last election cycles can be seen in the importance of the media for campaigns, party programs and voter behavior see Oberreuter (2012), who names Germany a "media democracy".

Figure 3.1: Number of polls in 13 states between 1990 and 2017 by the German polling institutes.



Parties often follow the current sentiment, published in polls with their campaign strategy and media evaluates the performance of the government with information from polls, compare Boon (2012). Polling institutes are assigned by media institutes (newspapers, TV channels, etc.) and provide snapshots of the current voter sentiment. As polls are fairly accurate in the run up of an election, they suffer from several errors as Steinmayr (2016) writes. These errors emerge due to sampling mode, different methods of interviewing, different weighting schemes and small sample sizes outlined in detail in Section 3.2.

3.2 Different types of polls and potential errors by conducting polls

Polls are sometimes taken as direct forecast in media and public, compare Miller et al. (2012), but they suffer from bias and failed at several elections e.g. in the German national election 2005 and did not even predict the right winner. This bias (error) occurs for a variety of reasons, explained in the following (see Kasprzyk et al., 2001).

Survey respondents are asked several questions by phone, face-to-face, etc. , the main question concerns the vote intention for whom to vote in an election. The question wording of a vote intention poll can vary, but this question is the one which is published mostly in media and these polls are often called "trial heat polls" (see Wlezien and Erikson, 2004). The question is asked on several points during the campaign and the publication in media generates the style of a race, compare Holbrook and DeSart (1999) and Lock and Gelman (2010). Most polls are vote intention polls, but new research also examines another question, asking about the winner of an election.

These vote expectation surveys ask "who will win the election" and are up to now sometimes used in two-party systems like the US or the UK (see Graefe, 2014) because an explicit winner appears in these elections. In multi-party elections like Germany, vote expectation surveys are a challenging task as the variety of parties often causes several options of coalitions, compare Ganser and Riordan (2015) and Graefe (2016).

Before the 1990s, institutes conducted surveys with face-to-face interviews which were costly in time and money. Institutes switched to surveys by telephone and online, leading to a growing number of polls and new institutes (see Blumenthal, 2014). By conducting a survey, institutes decide about an amount of designs concerning their survey like sampling method, design of the questionnaire (formulation, length, order of the questions, response options), timing and weighting scheme. According to the WAPOR (World Association for Public Opinion Research) a meaningful survey should include the name of the institute, client, date of the poll, type of the interview, question wording, sample size and error.[9] In the following, the weighting methods are explained in detail and the errors caused by the design of the surveys.

WEIGHTING METHODS OF THE INSTITUTES

Institutes are publishing polls relatively regular. These are not the raw polling data. They are weighted with unpublished specific weighting schemes. Institutes usually only publish the "projection" (polls), which are the weighted polls provided for the media. For national elections, the FGW and GMS also provide the "political sentiment", which are the raw polling data without political weighting, but with representative household weight. The real raw polling data, taken by face-to-face, online, etc. are not published. Polls are conducted by an institute, which in a first step randomly selects a representative group of respondents seen in Figure 3.2.[10] In a second step the potential respondents are interviewed with a certain data collection mode (face-to-face, online, etc.).

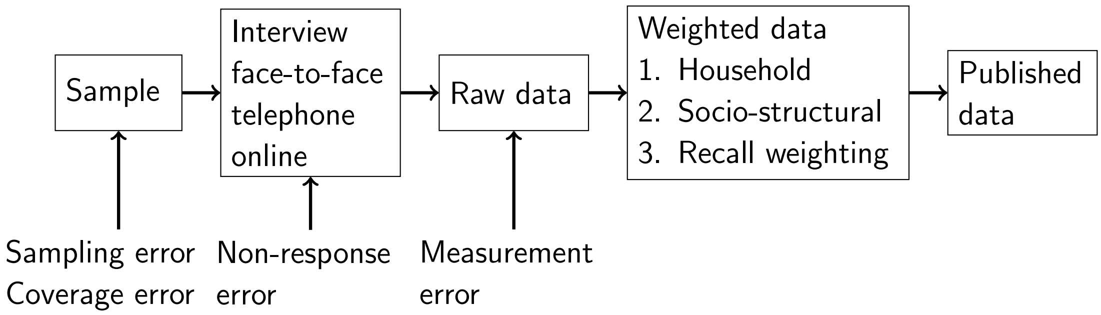
The responses are called raw data which are weighted with the institutes own schemes before they are published. In the last step, seen in Figure 3.2 the institutes correct the raw data with three weighting steps. First, the raw data are weighted by their household type, where single households are given greater weight than multi-member ones (see Jessen, 2014). If households with more than one resident are chosen for an interview, only the one who had recently birthday is selected.[11] Then the data are weighted by socio-structural characteristics e.g. if

[9] http://wapor.org/wp-content/uploads/2011/02/StandardDefinitions2011.pdf

[10] An often used method of selection is called Random Digit Dialing in telephone surveys, where the respondents are chosen randomly according to telephone numbers e.g. the FGW uses this method (http://www.forschungsgruppe.de/Rund_um_die_Meinungsforschung/Glossar_zur_Umfrageforschung/Erklaerungen/ (Last visit: 2017-09-29).

[11] http://www.forschungsgruppe.de/Rund_um_die_Meinungsforschung/Methodik_Po

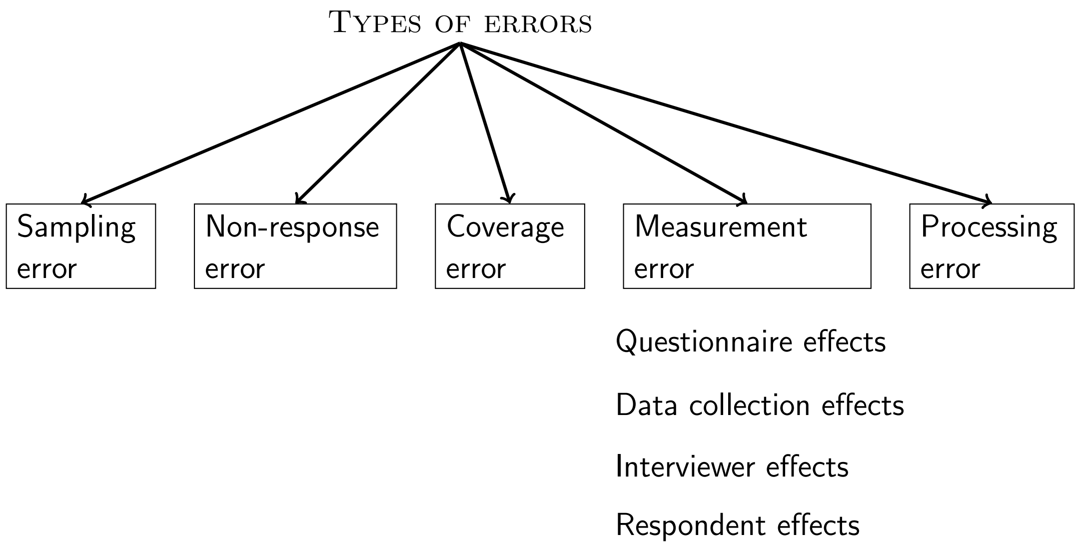
Figure 3.2: Weighting systems of the institutes and potential errors.



some population groups are over- or underrepresented in the sample due to age, profession, religion or other factors to generate a representative sample. Further, the respondents are asked for whom they voted in the last elections referred to as recall weighting. Also campaign events, party identification, tactical voting or unforeseen events are considered in the weighting before they are published. If new parties emerge and gain over 5%, polling institutes calculate their weights by considering these new parties. Some polling institutes published too low values for the AFD in the election in Saxony Anhalt 2016, due to several reasons. Voters did not reveal their real choice for extremist parties or the institutes did not expect thus high values for the AFD and put lower weights on this party.

Irrespective of these institute specific corrections there remain some types of errors stated in Figure 3.3 and mentioned by Gelman, King and Boscardin (1998), Graefe (2013a), Hillygus (2011), Linzer (2013), Schnell and Noack (2014) and Selb and Munzert (2016).

Figure 3.3: Different types of errors in polls. (Note that measurement error is divided in four subgroups.)



TYPES OF ERRORS

The first type of error is called "sampling error", which is the error caused by collecting a representative sample instead of the entire population.

While conducting a survey, interviewers often do not reach the respondents, the respondents refuse the answer on a specific question or refuse being interviewed. This "non-response error" reduces the sample size and increases the variance as Hillygus (2011) writes. The non-response rates in telephone surveys are over 90%, compare Graefe (2016), which is an indicator of the survey quality. The group of potential voters, who are interviewed is called inference population.

Another type of error is the "coverage error", which is caused by the sampling frame also referred to as "frame error", compare Schnell and Noack (2014). An over-coverage results of repeated questioning of the same person, whereas an under-coverage occurs, if persons which are part of the population are not ques-

litbarometer/

tioned (see Jessen, 2014).

Measurement errors are the most common type of errors, caused by four sources (questionnaires, data collection, interviewer and respondent effects). The questionnaire design influences the sample quality and produces errors by question wording, length of question, response category, questionnaire format and open or closed form of questions, compare McDermott and Frankovic (2003). Especially the question for whom would you vote for, if the election was held next Sunday is often of closed format with eight parties standing for selection. Small parties are rarer named by the respondents, as they are not an explicit response category of the question and therefore sometimes underestimated. The Infratest dimap institute provides questions in open format which leads to more neutrality and facilitates voting for small parties.

Measurement error is also caused by data collection mode effects. Interviewers ask respondents in face-to-face interviews per personal interview called "Paper and pencil personal interview method (PAPI)" or computer assisted interview (e.g. with tablet), also referred to as "Computer assisted personal interviewing (CAPI)" (see Kasprzyk et al., 2001). This type of survey is only conducted by Allensbach and is very complex and time intensive (see Brooker and Schaefer, 2017). The most common type of data collection is the telephone interview "Computer assisted telephone interviewing (CATI)", which is less flexible but less expensive than face-to-face.[12] Other types of surveys are self administered mail surveys and diary surveys, which are reported over a specific period of time. In online surveys, no interviewers are needed, but these suffer from low response rates and it is not clear who is the real respondent (see Brooker and Schaefer, 2017). Those are called "Computer assisted self interviewing (CASI)", e.g. vote recognition entry, prepared data entry, touchtone data entry or audio computer-assisted self interview. Interviewer and respondent effects also lead to measurement errors.

Respondent errors are caused by false replies, e.g. the respondents lie for which

[12]https://www.infratest-dimap.de/ueber-uns/was-wir-tun/methoden/telefonerhebungen/(Lastvisit:2017-07-31).

party they tend to vote and if they vote or not. Also the recall question is often answered with the winning party instead of the real elected one as Falter (2014) reveals. Measurement errors resulting from interviewer effects, appear due to errors made by interviewers, e.g. false questioning.

The last type of error is called processing error and is a result of editing or coding imputation after collecting data. Further types of errors are summed up by Brooker and Schaefer (2017), e.g. the "population specific error", which occurs if possible respondents do not own telephones. A "selection error" appears due to errors in selecting the respondents (not randomly selected or too old respondents, who own only fixed telephones). The new types of surveys e.g. online surveys provide other challenges as it is more difficult to prove the respondents required disclosures, concerning his or her age, gender, etc.

3.3 Accuracy and volatility of polls

Polls are "the currency of politics" (Boon, 2012, p. 465) and are carried out by several purposes, as to form the campaign strategy, to forecast the outcome of an election and to understand voting behavior (see Hillygus, 2011). One of the purposes of polls is understanding the behavior of voters on a certain point in the campaign. Also from the perspective of parties, polls are used to plan the further campaign strategy and to react on the voters support e.g. if there is a decline in the vote share for a party a new strategy can be developed. Hence, the Forschungsgruppe Wahlen asks the voters if they are confident with the policy of the government. However, the publication of polls also leads to tactical voting and influence of voters - in particular the undecided ones - in their voting decision. Published polls often influence the voting behavior of the respondents and induce them to vote for the most likely winner ("bandwaggon effect") or to vote for parties out of pity referred to as "underdog effect" (see Jessen, 2014). Hence, the institutes agreed not to publish new polls one week before election date since 1972 but violated this agreement in the 2013 parliamentary election (see Kruke, 2014).

Polls are often used as direct forecasts in the media, compare Kayser and

Leininger (2013), Küntzler (2014) and Jackman (2005). Institutes ask the question "which party would you vote for, if the election was held next Sunday". This vote intention question indicates, that voters respond hypothetical for whom to vote on a specific day before election date. According to the CEO of the German institute Infratest dimap, Michael Kunert, polls are no direct forecast of the election result, but a measure of the current sentiment of the voters.[13] A poll has to be seen as a snapshot of the current opinion and mostly includes the voters reactions on campaign events, compare Armstrong, Green and Graefe (2014), Berg, Nelson and Rietz (2003), Blumenthal (2014), Bunker and Bauchowitz (2015), Kamakura, Mazzon and de Bruyn (2006), Klein (2005) and Mughan (1987).

VOLATILITY AND ACCURACY OF POLLS

As polls suffer from some types of error, they are the best measure and report the current response of voters (see Berg, Nelson and Rietz, 2003). The quality of a poll depends on the timing, institute, sampling method, etc. The closer polls are conducted towards election day, the better are the results, compare Linzer (2013) and Kayser and Leininger (2013). Schnell and Noack (2014) obtain a decreasing amount of error shortly before election. The accuracy of a poll determines the volatility as Coletto and Breguet (2015), Graefe (2014), Küntzler (2014) and Linzer (2013) investigate.

In literature there are various measures to evaluate the quality of polls. The most common one is the "Margin of Error" (MoE), which is denoted as follows,

$$1.96 \cdot \sqrt{\frac{s_h(1-s_h)}{n}} \tag{3.1}$$

(see Jackman (2005)). Where s_h is the poll value on a specific day in the campaign, h are the days before election and n is the sample size of the poll. The notation of

[13]http://www.mdr.de/medien360g/wissen/wahlforschung100.html (Last visit: 2017-07-29).

the polls with s_h is outlined in detail in Chapter 6. An increase in the sample size leads to a gain in statistical precision. The margin of error is the deviation from the poll value. It is calculated as a multiple of the factor 1.96 (for a 95% confidence level) with the sampling error $\sqrt{s_h(1-s_h)/n}$ (see Leigh and Wolfers, 2006). Via the Central Limit Theorem uncertainty of a poll value, calculated with a large sample follows a normal distribution. The 1.96 is obtained by the assumption that the random variable follows a normal distribution and with 0.95 probability, it lies within a 1.96 standard deviation of the mean (see Jackman, 2005). Institutes usually indicate these margin of errors with about three percent, depending on sample size and vote share of the poll.[14] As the margin of error is usually above 1%, the vote shares of parties under 3% are not published in polls. The margin of error varies across the institutes so called "house effects", as there is often a tendency for a specific party in over- or underestimation of poll values. (E.g. the institute Allensbach predicted worse than others in national elections as Küntzler (2014) investigates.) Figure A.3 in the Appendix indicates boxplots for the margin of error of the three institutes, Emind, Forsa and Infratest dimap, which published the most state level polls over the last decades.[15] The boxplots are calculated for the certain parties depending on institute, for all 13 states (investigated in this thesis), 365 days before the respective election dates and with 95% confidence level. A margin of error of e.g. 0.03 means, that there is a 3% deviation in the positive or negative direction from the poll value. The margin of error for the main parties is even higher, than for the smaller ones FDP and AFD. Also differences between the institutes can be found especially for the Left party, where Infratest provides the most accurate results.

In Figure A.4 in the Appendix, the 95% confidence intervals for the six parties for the last state election in Northrhine Westphalia, 365 days before election date ("2017-05-14") are shown. The filled dots indicate the poll dates and values, the

[14] The FGW reports 2% deviation for vote shares of 10% and 3% for vote shares of 40% see http://www.forschungsgruppe.de/Rund_um_die_Meinungsforschung/Methodik_Politbarometer/ (Last visit: 2017-08-11).

[15] Compare Figure 5.1.

corresponding dotted lines show the margin of error in the positive and negative direction. There are greater deviations for the SPD and CDU, who gained about 30 % of the votes in comparison to the smaller parties.

The newest research in US elections aggregates polls from different institutes, investigated by the websites "FiveThirtyEight.com", "pollsters.com", "Princeton-electionconsortium.com" and "RealClearPolitics.com" to reduce volatility as Cuzan, Armstrong and Jones (2005), Gibson and Lewis-Beck (2011) and Hillygus (2011) write. Polls on these sites are weighted by sample size and recentness. Others like Bunker and Bauchowitz (2015) weight the polls with a weighting factor of $k_i/\sqrt{N_i}$, where k is a value, depending on the confidence interval and N the sample size of poll i.

Polls also suffer from errors(biases) (see Section 3.2), which impair the quality and validity of a poll and lead to sometimes very volatile polls, compare Castro (2013). Polls for national elections started in the 1930s in the US and then in other countries, whereas polls for local elections were neglected for a long time (see Hillygus, 2011). In the US, polling on state level basis grew in the 2000s (see "RealClearPolitics.com"). Collecting polls on a state level basis for national elections is a challenging task because not every state is of similar interest for the pollsters e.g. in the US there are great differences in the frequency of polling, compare Hillygus (2011).

In Germany, polls for state elections had a long tradition since the 1960s, but there was mostly only one poll per election cycle shortly before election date, conducted by regional newspapers or the FGW, as Figure 3.1 indicates. Despite, since the 2000s the website "wahlrecht.de" publishes polls from several institutes with a higher frequency. Nevertheless, polls for state elections or on state level basis suffer from higher volatility than national polls. Volatility in state level polls is higher than on national level due to the sometimes smaller sample size in state polls than on the national level, as state polls are often conducted by regional newspapers. Further, the party landscape in states changed from election to election and regional, small parties appear first on state level and reach the 5% threshold easier than on national level. Another reason for the higher volatility lies in the frequency in conducting polls on state level. Sometimes there are only three polls in a year, which leads to high deviations between two polling results, as the voters change their minds very often during the campaign. They do not always converge in accuracy towards the end of the campaign as national ones do (see Hillygus, 2011). Volatility in polls irrespective on national or state level increased due to changing voting behavior explained in detail in Subsection 2.3.3. In particular the increasing number of undecided voters shortly before an election causes higher volatility in polls, compare Castro (2013). Hence, often macroeconomic models were used early in the campaign as the past election results were in the voters minds and the so called "honeymoon effect" is lasting. At the end of the electoral cycle, when the last election was held four or five years ago, the faster changing political landscape and volatile behavior encourages the use of poll based models. Although sometimes in history polls failed before election e.g. in the US 1948 and 1996, in UK 1992, in France 2002, in Italy 2006 and in Germany 2005 as Schnell and Noack (2014) report. Despite these fails, polls are the only way to react to short-term changes in the political landscape, e.g. if new parties arise (in Germany the Pirates and the AFD). These parties can not be recorded so easily with fundamental variables, especially in multi-party elections. In this work, polls from different institutes are taken as data base and state election results are forecasted with different methods. In literature, most models include polls in terms of a popularity index e.g. chancellor support (see Holbrook and DeSart (1999)) and Chapter 4.

4 Literature review

Up to now there is only sparse research on German election forecasting on state level. The history of election forecasting began with Louis Bean, who predicted the US elections by identifying bellwether states in the 1940s, see Jérôme and Jérôme-Speziari (2011) and Lewis-Beck (2005). The national election results of specific bellwether states were close to the national election results for the US and thus taken as forecast for the US elections. In this thesis, state elections are also in focus due to their character as direction elections for the national election in Germany.

In the following, an overview concerning the geographical scope and history of the election forecasting literature is given. Further, the existing literature is divided into three prevalent types of models according to Lewis-Beck and Dassonneville (2015). The last section in this chapter presents forecasting literature on German elections.

4.1 Overview of existing literature on election forecasting

Statistical forecasting models for elections were developed in the middle of the 1970s with the 'Political economy model' of Lewis-Beck and Rice (1984) for US elections. The US elections own the longest history of forecasting,compare Graefe and Armstrong (2014) and Lewis-Beck (2005). Most forecasting research is so far made for the elections in the US, the UK and in France as Bunker and Bauchowitz (2015), Lewis-Beck and Stegmaier (2000), Magalhaes, Aguiar-Conraria and Lewis-Beck (2012), Nadeau, Lewis-Beck and Bélanger (2010) and Walther (2015) report.

Hence, most research deals with national elections for two-party nations or countries with two stage electoral systems. The two-party system is one reason for the interest in the US elections, besides the importance of the US in the political landscape. The relatively clear distinction between only two predominant parties (Democrats and Republicans) makes it easier to forecast the winner of the election. Two-party elections are thus easier to forecast than multi-party elections, where more parties are electable. In multi-party nations frequent changes in party landscape and voting behavior lead to difficulties in forecasting elections (especially with models based on fundamental variables).

OVERVIEW OF EXISTING LITERATURE

Against one's better knowledge, the literature overview, given in Table B.1 in the Appendix, is the most extensive one up to now. The election forecasting literature is reviewed for different countries, methods and variables (in the case of regressions) and contains about 80 models.

Some authors also provide compilations of different election forecasting models (mostly regression based models) (see Campbell (2008), Graefe (2013a), Jiao, Syau and Lee (2006) and Lewis-Beck and Stegmaier (2000)). Campbell (2008) gives an overview of nine forecasting models for US elections, whereas Lewis-Beck and Stegmaier (2000) compare models according to country (US, UK, France and others).

In the literature of election forecasting, the US is the most investigated country followed by the UK, France and other countries, compare Bunker and Bauchowitz (2015) and Magalhaes, Aguiar-Conraria and Lewis-Beck (2012). Table B.1 in the Appendix contains the author(s), the studied nation, the forecasting objective (single party, coalitions, seats, probability of winning, etc.) and the used method. The focus in this table lies on regression based models, therefore the column "forecasting objective" can also be seen as dependent variable. The rightmost column indicates the (independent) variables. Some models, at the end of the table are non-regression models and only obtain information about the authors, studied nation, forecasting objective and used method. The table is sorted in alphabetical order by author with regard to method and studied nation. Research on non-

German elections is listed before research on German elections. Next, regression based models are sorted, followed by other types of models e.g. Bayesian ones in the non-German and German case. The geographical scope ranges from the most studied nation US to the European nations (UK, France, Germany, Austria, Italy, Hungary, Sweden, Ireland, Spain and Netherlands) and other countries like India, Canada and Australia, where election forecasting is a small research field. Further, some researchers provide models for more than one country, e.g. Kamakura, Mazzon and de Bruyn (2006) investigate two stage elections.

The forecasting objective in case of regressions is also the dependent variable. Most authors provide forecasts for single parties, often for the incumbent party in two-party nations. In multi-party elections, forecasts for coalitions are common (see Aichholzer and Willmann (2014), Kayser and Leininger (2013), Küntzler (2014), Norpoth and Gschwend (2010) and Schaffer and Schneider (2005)). Single party forecasts in multi-party elections are difficult to forecast with e.g. political economy models especially for new emerging parties like the AFD or the Pirates (who are near or around the 5% threshold). For Germany, only Jérôme (2013), Selb et al. (2013) and Walther (2015) provide single party forecasts for the multi-party system. Besides the vote share as forecasting objective, some authors examine the seat shares or other measures e.g. Sanders (1995) (popularity, personal expectations), Berlemann, Enkelmann and Kuhlenkasper (2014) (approval rating), Heij and Franses (2011) (probability of vote share) and Kamakura, Mazzon and de Bruyn (2006) (value of a candidate). The most established models are Ordinary Least Squares (OLS) regressions with various independent variables seen in Table B.1.

The "independent variable" can be divided in the types economic, social, political and other variables shown in the table. The predominant regression models are politico-economic ones as Bunker and Bauchowitz (2015) reveal. An example of a politico-economic model is provided by Auberger (2005), who uses the unemployment rate, Gross Domestic Product (GDP) or inflation to forecast the vote share of French elections.

Economic variables play a big role in regression models in form of macroeconomic variables (GDP, inflation, state income, defense spending, interest rate, tax, retail price index) or other economic variables like income level, growth, inequality, poverty and real disposable income. Khemani (2001) forecasts the vote share of the incumbent in India elections with the variables poverty, inequality, state income and growth.

Political variables are included in terms of approval rating (the grade of support for a party), terms in office (incumbency), national identity, opposition, electoral area, party identification, issues, vote support, chancellor support, past presidential vote, political instability, politics and institutions, partisan stronghold, popularity and candidate appeal.

In some cases, a political variable is included in form of historical events, like the German reunification, Wars or political affairs. Examples for including political variables in the model are given by Berlemann, Enkelmann and Kuhlenkasper (2014), who take the Iraq, Vietnam and Afghanistan War into account. Lewis-Beck and Stegmaier (2000) forecast the vote share of the incumbent US party with the variables Vietnam War and the Watergate affair. Most models are solely based on economic and/or political variables especially of the last election. Aichholzer and Willmann (2014), Auberger (2012), Bellucci (2010) and Khemani (2001) provide solely economic models, whereas Berlemann, Enkelmann and Kuhlenkasper (2014), Jérôme, Jérôme and Lewis-Beck (1999), Jérôme and Jérôme-Speziari (2011), Lewis-Beck and Stegmaier (2000) and Lewis-Beck and Tien (1996) investigate politico-economic models. Additionally, social variables are included e.g. age, gender, religion, race, education, social class, personal expectation, political knowledge, member of political organization, income level and vote intention. Andersen and Heath (2003) forecast single vote shares of Canadian, UK and US elections only with social variables. Murr (2011) uses age, gender and education besides other variables to forecast UK elections with a logistic regression. For US elections, Williams et al. (2012) provide a classical OLS regression model, based on social and political variables for single vote share forecasts.

Other variables are included in the models, like biographical information of the candidates compare Armstrong and Graefe (2011), news or the honeymoon effect (the time short after the election is called honeymoon time, where voters are confident with their decision like Leigh and Wolfers (2006) for Australian elections). Further, other elections are included in the models, compare Magalhaes, Aguiar-Conraria and Lewis-Beck (2012) who consider European elections in their models. In Table B.1 in the Appendix, the predominantly models are OLS regression based forecasting models, other types of forecasting models are e.g. logistic regression by Murr (2011) and Nadeau and Lewis-Beck (2001), whereas Whiteley (2008) uses a bivariate regression and Brown, Firth and Payne (1999) a ridge regression. Other types of models are provided by Bafumi, Erikson and Wlezien (2010), who simulate single party vote shares in US elections with the national previous vote and a local component. Cuzan, Armstrong and Jones (2005) provide a PollyVote model, where the vote share of the US incumbent party is forecasted as the average of polls, values from the Delphi method, expert opinions and the IOWA electronic market. Heij and Franses (2011) investigate a binary logit model for elections in the Netherlands and Hobolt and Spoon (2012) a generalized linear model with logit likelihood for the elections of the European Parliament. Klarner (2008) provides a random effects model with Maximum Likelihood regression for US elections. Further, the table contains Bayesian models, compare Jackman (2005) and Linzer (2013) and Dynamic Linear Models compare Strauss (2007) and Walther (2015) or simulation based models like Gordon (2010) and Hanretty (2013). Some research is made with an autoregressive AR(1) model by Berg, Nelson and Rietz (2003), an autoregressive AR(2) model, see Sanders (1995) and ARIFMA models, compare Dolado, Gonzalo and Mayoral (2003) and Lebo and Norpoth (2011).

In the first part of this section, the different types of election forecasting models are discussed with respect to nation, method and variables. Up to this point, a discussion about the data base, which is provided to forecast the vote share, seats, etc. is neglected. In the following, the models are divided according to their static/dynamic character and the type of the data.

THREE PREVALENT TYPES OF FORECASTING MODELS

The existing election forecasting literature seen in Table B.1 can be divided into three prevalent types according to structural, aggregate and synthetic models, compare Lewis-Beck and Dassonneville (2015) and Haupt, Schnurbus and Huber (2017, p. 5).

Structural models preferably apply an OLS regression to forecast the election results. This static approach is based on economic data like macroeconomic variables e.g. unemployment rate, GDP, inflation and social/political data like incumbency, vote support and chancellor support. For French elections, Auberger (2012) and Jérôme, Jérôme and Lewis-Beck (1999) use this static approach, Brown and Chappell (1999), Holbrook and DeSart (1999) and Lewis-Beck and Tien (2012) are examples for US elections. Structural models also apply socioeconomic variables as covariates, which are past values of economic data, which are published only quarterly or monthly, whereas political variables e.g. past election results are available several years before. For a fragmented multi-party system and the changing political landscape this static approach is not appropriate. Especially for multi-party systems, like Germany with changing party landscape, politico-economic variables do not provide sufficient explanatory value for election forecasting (details are provided in the next section).

Due to this, Lewis-Beck and Dassonneville (2015) introduced aggregate models, which solely rely on polling data available shortly before election day. These poll based models range from basic models like using the recent poll value as forecast for German elections like Schnell and Noack (2014), to the weighted average of Italian election polls, compare Hanretty (2013). Gott and Colley (2008) calculate the median of polls one month before election for US elections and Bunker and Bauchowitz (2015) average over pollsters in Chile. More complex models are developed by Jackman (2005) and Walther (2015), who use advanced Bayesian dynamic linear models. Whereas structural models offer static forecasts, aggregate models provide a dynamic perspective.

A combination of aggregate and structural models is called synthetic model. The static character of the structural models is supplemented by the dynamic characteristics of the aggregate models, which are incorporated by polling data. Synthetic models provide advantages due to including short term polling data, which can be updated in case of new available polls. Jérôme, Jérôme-Speziari and Lewis-Beck (2013), Kayser and Leininger (2013), Küntzler (2014), Norpoth and Gschwend (2010), Schäfer (2013) and Selb and Munzert (2016) apply synthetic models to forecast German election results, mostly by including a variable of chancellor support. Over the last years, synthetic models grew in importance as research from Aichholzer and Willmann (2014), Bellucci (2010), Holbrook and DeSart (1999), Lewis-Beck and Tien (1999) and Nadeau, Lewis-Beck and Bélanger (2010) reveals.

In the following, the forecasting literature on German national and state elections is under review.

4.2 Forecasting literature on German elections

The literature on election forecasting concentrates mainly on US elections, seen in Table B.1 in the Appendix, followed by UK and France. Research on German elections grew over the last years, with focus on national elections. At the moment, also aggregate and synthetic models are used to forecast elections, instead of the predominantly structural politico-economic models. The fragmented multi-party system is more challenging to forecast vote shares, than two-party systems. The last part of the table contains research on German elections.

The first forecasting models for German national elections are published by Norpoth and Gschwend (2003) and Gschwend and Norpoth (2005), who provide structural, politico-economic models using OLS regression. As more than two parties are part of the government in multi-party elections, the researchers focus only on the incumbent vote share or the expected coalition, e.g. Gschwend and Norpoth (2005), Kayser and Leininger (2013), Lewis-Beck and Dassonneville (2015) and Norpoth and Gschwend (2010) provide forecasts for the coalition or single incumbent party using OLS regressions. Also Küntzler (2014) forecasts the coalition vote share with a dynamic regression based model.

Whereas Jérôme, Jérôme-Speziari and Lewis-Beck (2013) and Selb and Munzert (2016) investigate vote shares of the main parties CDU, SPD, LIN, GRE, FDP and others, Ganser and Riordan (2015) and Graefe (2015) additionally take the Pirates and the AFD into account.

All research on German elections deals with national elections, only Gaissmaier and Marewski (2011) calculate forecasts for German state elections (in Brandenburg and Northrine-Westphalia) and German national elections, based on different data sets. Averaging and simulating techniques are applied on these data (recognition-based, intention based and wisdom of crowds data). The forecasting objective are large and small parties, but not single vote shares. Up to now there are no other forecasting models, calculating single vote shares for German state elections.

The prevalent type of models are structural ones, but over the last years, the number of synthetic models including poll data increased, compare Küntzler (2014), Lewis-Beck and Dassonneville (2015) and Selb and Munzert (2016). The forecasters use polls to update their models with current data. Graefe (2015) forecasts vote shares based on polls and prediction markets, also Lewis-Beck and Dassonneville (2015) use aggregate, structural and synthetic models, whereas Selb and Munzert (2016) forecast with Bayesian models.

As purely structural models are not suitable to cover the faster changing party landscape with new emerging parties, poll data are used to capture mid or short term developments. In the last German national election 2017, the parliament contains six parties including the FDP, who was not part of the parliament four years ago. Hence, it is difficult to include previous election results in multi-party elections.

Ganser and Riordan (2015) apply data from vote expectation polls, asking about the education, membership to a political organization and other information about the voter to forecast German elections.

In this work, the focus solely lies on aggregate models, where poll data of vote intention polls are used as basis for several models explained in detail in Chapter 6. The next Chapter 5 describes the data structure of polls.

5 Data of German state level polls

This work concentrates on state elections, where the AFD reached at least one seat or rather 5% of the votes and was part of the parliament. Up to now, the AFD is part of 13 state parliaments (and even in the latest national parliament 2017). First, the investigated state elections are illustrated. Section 5.2 deals with the preparation and aggregation of the polling data, which are proposed as data base for the forecasting methods.

5.1 Choice of the investigated German state elections

As Germany is a multi-party system, there are 20 to 30 parties competing in an election, who changed in their importance in history. New parties first appeared in state parliaments or only on regional level and passed the 5% threshold in national elections only in few cases. The Greens first appeared 1979 in the Bremen state parliament and are now an established member of most parliaments. In the state election in Hamburg 2001, the "Schill" party gained nearly 20% of the votes, but remained only in this parliament. Since the 2000s several new parties emerged in the Euro Zone, who focus on anti-European or anti-refugee issues (see Morlok, Poguntke and Zons, 2016). In Germany, the AFD was founded on sixth of February 2013 and first passed the 5% threshold in the state election 2014 in Saxony. Hence, this work investigates the five main parties and the new emerged AFD, who currently play a role in the political landscape. In contrast, Ganser and Riordan (2015) and Graefe (2015) calculate additionally vote shares for the "Pirates", who were founded in 2006 and reached four state parliaments in 2011 and 2012, but are now under the 5% threshold and so far not part of this work.

Table 5.1: Election dates of 13 German state elections.

Date	State
2014-08-31	Saxony
2014-09-14	Brandenburg
2014-09-14	Thuringa
2015-02-15	Hamburg
2015-05-10	Bremen
2016-03-13	Baden-Württemberg
2016-03-13	Rhineland Palatinate
2016-03-13	Saxony-Anhalt
2016-09-04	Mecklenburg Western Pomerania
2016-09-18	Berlin
2017-03-27	Saarland
2017-05-07	Schleswig-Holstein
2017-05-14	Northrhine-Westphalia

Table 5.1 indicates the election dates and the states between 2014 and 2017.[16] The 13 state elections are in between the national cycle 2013-2017 and show the development over this cycle on state level. Elections before 2014 are not investigated due to the missing of the AfD and are not comparable to the 13 chosen elections investigated here. The aim of this work is to forecast the single vote shares of the six parties in these 13 state elections. Most research of multi-party elections in Germany forecasts national coalition vote shares, compare Kayser and Leininger (2013), Küntzler (2014), Norpoth and Gschwend (2010) and Schaffer and Schneider (2005) and the incumbents vote shares, compare Bellucci (2010) and Lewis-Beck and Dassonneville (2015), whereas single vote shares are forecasted by Jérôme, Jérôme-Speziari and Lewis-Beck (2013), Selb and Munzert (2016) and Walther (2015). The latter provide forecasts for the five main parties excluding the AFD. Forecasts on state elections, so called "second-order elections" grew in importance, like Blumenthal (2014) and Christensen and Florence (2008) with an increasing number of polls and research.

[16] The elections of the remaining three states took place in: HE ("2013-09-22"), BY ("2013-09-15") and NI ("2013-02-19").

In Chapter 6, the methods for election forecasting are presented in detail. The methods are applied on poll data, which have to be prepared before using, outlined in the next section.

5.2 Preparation and aggregation of polling data

In Germany, polls were conducted for national and state elections since the 1960s. As stated in Section 3.2, vote intention polls are the predominant type of polls. Vote expectation polls are rather used in the US, as they ask "who will win the election" and therefore are difficult to apply in a multi-party system like Germany. In German elections the vote shares of all parties are of interest, because often coalitions of two or more parties have to be formed.

The question of the vote intention polls is "who would you vote for, if the election was held next Sunday". The polls dealing with this question are in focus of this work and published by several polling institutes in Germany.

The number of polling institutes and polls increased since the 1980s in the US and since the 1990s in Germany due to the new data collection method of telephone and online surveys (see Blumenthal, 2014). As Linzer (2013) reports, over 150 polling institutes exist in the US, in Germany over 100. The six best known institutes are indicated in Table 5.2.[17] The name of the institute, the average number of respondents, the frequency of polls and the type of survey for the national elections are shown.

Most institutes use CATI telephone surveys, the institute INSA online surveys. Only the oldest institute Allensbach interviews the respondents face-to-face and works with the quota method instead of random sampling for choosing representative respondents. Furthermore, results of the Allensbach institute are displayed accurate to a tenth. Other institutes only publish integer results, which indicate

[17]INSA is assigned by BILD, Infratest dimap by ARD, FGW (Forschungsgruppe Wahlen) by RTL, Forsa by Stern
http://www.insa-meinungstrend.de/de/index.php,https://www.forsa.de/methoden/,https://www.infratest-dimap.de/ueber-uns/was-wir-tun/methoden/telefonerhebungen/ (Last visit: 2017-07-31).

Table 5.2: Characteristics of polling institutes for national polls in Germany.

Institute	Number of respondents	Frequency	Type of survey
Infratest	1000-1500	Every 2 weeks	Telephone
Allensbach	1000-2000	1 per month	Face-to-face
FGW	1300	Every 3 weeks	Telephone
Forsa	2500	1 per week	Telephone
Emnid	1500-2500	1 per week	Telephone
INSA	2000	1 per week	Online

errors about two to three percent. The sampling is implemented over a few days or weeks time period e.g. the FGW collects data during the period from Tuesday to Thursday.[18] For national elections, the six institutes in Table 5.2 provide polls on regular dates, displayed in the column "Frequency". The frequency for national polls is mostly weekly and conducted by telephone with Random Digit Dialing.

In contrast to national polls, polls on state level are published at very irregular intervals. Besides the six main institutes, polls for state elections are also published by other institutes, especially regional newspapers provide own polls in the run up to elections.

Figure 5.1 shows the number of polls for the different institutes. The institute Infratest published more than 500 state level polls for the 13 states over the year 1964 until today, followed by the institutes Forsa and Emnid. Allensbach publishes very rarely polls for state elections.

The forecasts in this work are based on poll data for state elections, which are collected from several sources. The FGW publishes data from 1964-2014, available on the website "gesis.org" (GESIS Datenarchiv, Köln. ZA4182 Datenfile Version 2.1.0, doi:10.4232/1.12389), which have to be weighted by the "representation weight". Data of the FGW and of the other institutes from the 1990s until today are published and collected by the website "wahlrecht.de".[19]

[18]http://www.forschungsgruppe.de/Rund_um_die_Meinungsforschung/Glossar_zur _Umfrageforschung/Erklaerungen/Stichprobe (Last visit: 2017-08-31).

[19]Data are collected from:
Forschungsgruppe Wahlen (cumulative data set of state elections from 1962-2004.

Figure 5.1: Barplot of the amount of state level polls for 13 states over the last decades since 1964, published by the ten most common institutes.

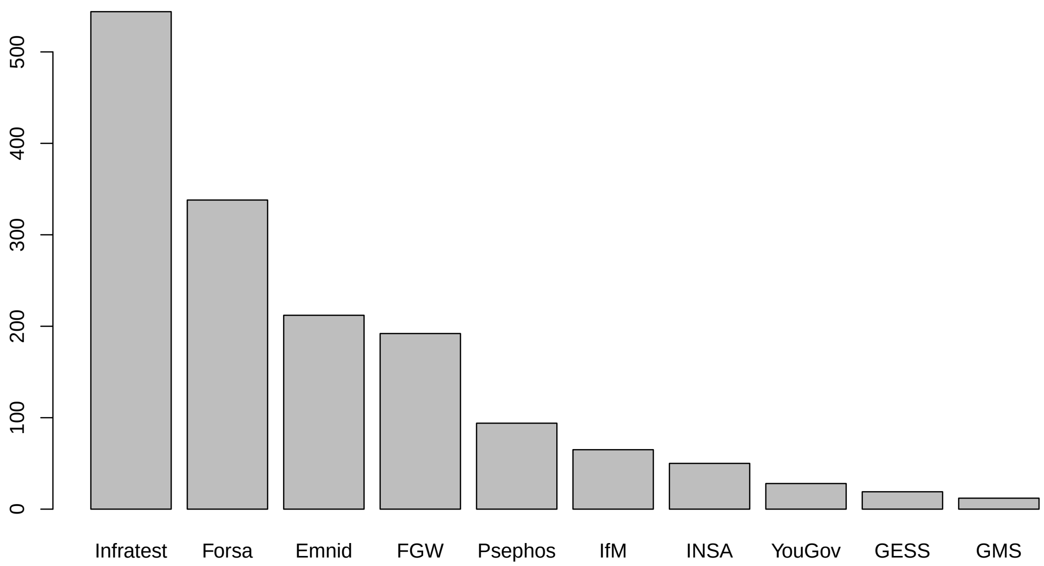


Table 5.3 indicates further details on the number of polls for state elections in Germany per institute (the ten most common ones) and the specific states over the last decades since the 1960s. As seen, the number of polls varies between states and institutes, with the most polls for Berlin (WBE). Polls from other institutes

Scheuch, Wildenmann, Baumert et al. (2015): Landtagswahlen - Integrierter Datensatz 1964-2004. GESIS Datenarchiv, Köln. ZA4182 Datenfile Version 2.1.0, doi:10.4232/1.12389);
Additional for elections after 2004 (Data from Forschungsgruppe Wahlen ZA3992, ZA5383, ZA6592, ZA4399, ZA5625, ZA4870, ZA5935, ZA4745, ZA5627, ZA4866, ZA5372, ZA5936, ZA3990, ZA4867, ZA4868, ZA5623, ZA4511, ZA4864, ZA5934, ZA4396, ZA5622, ZA5644, ZA4401, ZA5626, ZA4394, ZA5381, ZA5643, ZA3993, ZA5379, ZA5642, ZA3994, ZA5377, ZA6591, ZA4403, ZA5624, ZA3991, ZA5375, ZA6593, ZA4404, ZA4405, ZA5629);
Data from several polling institutes e.g. Allensbach, INSA, Forsa, Emnid, YouGov, GMS, Infratest dimap, etc. by Cantow et al. (2017) http://www.wahlrecht.de/umfragen/landtage/index.htm.

are neglected in this table, but included in the data set.

Table 5.3: Number of polls per institute and state over the last decades for 13 states and since 1964.

Institute	SN	BB	TH	HH	HB	BW	RP	ST	MV	WBE	SL	SH	NW
Infratest	27	58	26	34	9	44	82	26	25	72	30	44	67
Forsa	5	16	5	14	2	18	7	6	5	190	6	19	45
FGW	10	10	10	20	15	20	16	13	9	16	15	14	24
Emnid	28	7	3	18	4	12	10	4	30	52	7	3	34
Psephos	0	0	0	36	0	0	43	0	0	0	0	2	13
GMS	0	2	0	2	0	0	0	1	0	0	0	2	5
IfM	27	0	28	0	0	4	0	3	0	0	0	3	0
INSA	3	1	14	1	1	6	5	4	3	3	3	1	5
YouGov	0	0	0	1	0	2	0	0	0	0	0	0	25
GESS	0	1	1	3	0	0	14	0	0	0	0	0	0

In Figure 3.2, potential voters are interviewed by several institutes and the raw data of the sample are weighted by the institutes own weighting schemes. Two institutes FGW and GMS publish the raw data for national polls which are only weighted with socio-cultural weights, referred to as "political sentiment". The other institutes only publish the weighted data called "projection" including tactical voting, party identification, etc.

The polls of the institutes are collected on different modes and weighted with different schemes. The institutes mostly provide projection polls for state level polls. The FGW publishes the raw data, which are only weighted with the representation weight in this thesis. In Table 5.4, the summary statistics of the number of interviewed persons of the 13 states over the last decades is shown. On average, 1015 randomly selected persons responded the questionnaires. For 186 polls no information about the number of respondents is given.

Table 5.4: Descriptive statistics of the number of respondents over the last decades for the investigated states and institutes.

Minimum	1.Quartile	Median	Mean	3.Quartile	Maximum
184	1000	1001	1015	1005	7620

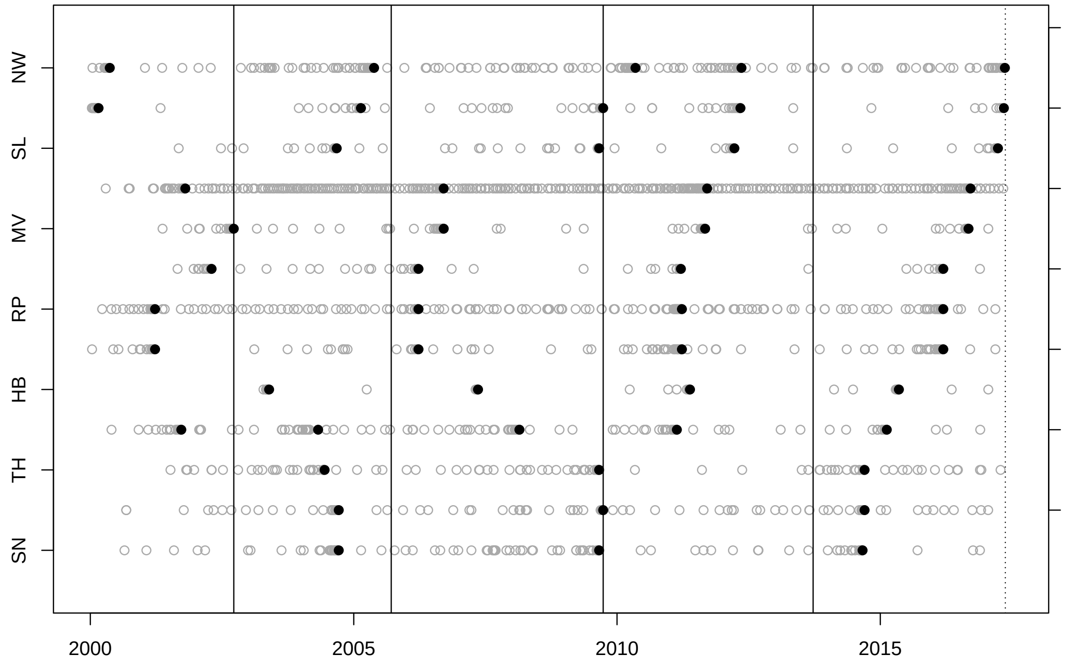
For classical time series models like ARIMA or state space models an equidistant database and sufficient number of observations is needed. Therefore, especially in states like SH, where only few polls are published, the regularized database provides enough observations. For some forecasting methods, a regular time series with equidistant values is needed, but poll data on state level are published very irregularly. Whereas institutes provide polls every week or every second week, see Table 5.2 for national elections, polls for state elections are published in short time distances before the elections, but only a few polls are published in the time period between the last election and the start of the campaign.

Figure 5.2 indicates the irregularly published poll data for the 13 states SN, BB, TH, HH, HB, BW, RP, ST, MV, WBE, SL, SH and NW.[20] The grey dots show the single poll values for the specific states between "2000-01-01" and "2017-05-14" (grey vertical line). The election dates of the state elections are displayed as black dots in the figure. The vertical black lines show the election dates of the national elections ("2002-09-22", "2005-09-18", "2009-09-27", "2013-09-22" and "2017-09-24"). State elections are stretched over the national cycles, e.g. with 13 state elections in the last four-year national cycle (see Haupt, Schnurbus and Huber, 2017). The number of polls increased from the 2000s until today and for some states like WBE, RP and NW there are more polls than for other states. Further, the number of polls shortly before the election date increases, seen in the higher number of grey dots ahead of the black dots. Also the quality of polls improved with higher accuracy, compare Linzer (2013) and Mughan (1987).
As potential voters are interviewed over a few days (mostly four days), the poll date is the day of publishing the polls, not the first or last day the respondents are asked, in contrast to Wlezien and Erikson (2004), who take the middle of the survey period as date for the poll.

To generate regular poll data on a daily basis, the poll value on a specific date is repeated as long as the next value is published, seen in Figure 5.3 on the example

[20] Compare Haupt, Schnurbus and Huber (2017, p. 2) providing this figure for all 16 states.

Figure 5.2: Regional poll dates (gray dots) and regional election dates (black dots) for 13 federal states between 2000 and 2017. National election dates are indicated by vertical lines, the vertical dashed line is the date “2017-05-14”.

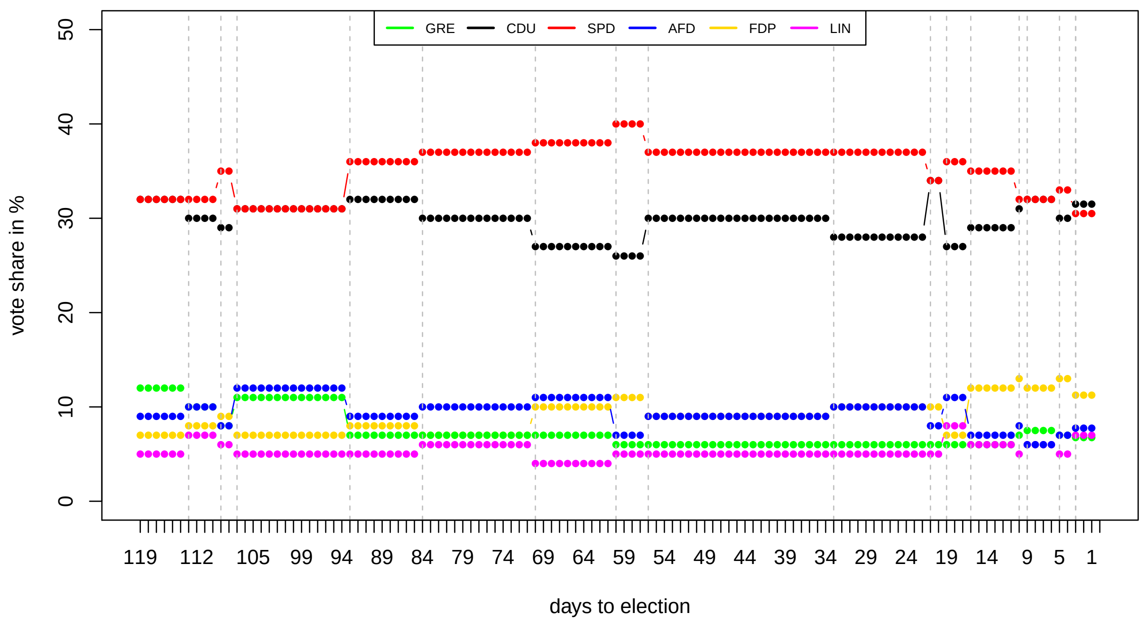


of Northrhine Westphalia. For the six parties the published poll value - the date of the poll is shown by the vertical grey dotted line in days until election - is repeated until a new poll is published.

The first poll value for the data is taken 365 days before the respective election date of the state. If more than one poll is published on a day, the mean of these polls is chosen analogous to Graefe and Armstrong (2013) and Wlezien et al. (2013). In the case of NW, there are six dates, where more than one institute publishes polls and the mean of these polls is calculated. In this figure, only the vote shares of the polls for the six parties with lead time from 120 days until the election date “2017-05-14” are illustrated. The figure also indicates the high volatility even 120 days before election, e.g. the SPD (red points) shows a ten percent range until election. In the Appendix, another Figure A.5 of the last 120 days before the state election "2017-05-07" in Schleswig Holstein is illustrated. There can be seen a type of state which is not of special interest for the public, as only five polls 365 days and three polls 120 days before election were conducted.

Analogous to the Figures 5.3 and A.5, the figures in the Appendix provide information, how to generate a regularized time series of polls for the remaining investigated states of this work. Figure A.6 shows the regularly generated polls for the elections held in 2014 in SN, BB and TH. Figure A.7 indicates the regularized data for the elections in HH and HB in 2015 and Figure A.8 for the state elections in SL, SH and NW in 2017. The regularized data for the elections in the year 2016 in BW, RP, ST, MV and WBE are shown in the work of (Haupt, Schnurbus and Huber, 2017, p. 10).

Figure 5.3: Regularly generated polls for the six parties in the latest NW state election from 120 days up to the election date "2017-05-14". Vertical dotted grey lines show the dates, where polls are published.



Between the states there are differences in the number of polls. The number of polls is increasing shortly before election with increasing interest of the voters. With increasing interest of the voters shortly before election, the number of polls also increases. In Figure A.9 in the Appendix, the number of state level polls is illustrated for the years 1995 up to 2017 for the states Schleswig Holstein and Northrhine Westphalia. To compare a state with less polls like SH (upper part) with a state like NW, where a higher number of polls is published. Grey vertical lines indicate the state election dates in the respective states, the black dots the number of polls. The NW state election is seen as indicator for the national election due to Northrhine Westphalia as Germany's most densely-populated state. Whereas SH is not of special interest for the polling institutes. The number of polls in both states grew from 1995 until today. Shortly before election, shown as vertical black lines, the cumulative number of polls increases to over 60 in the year 2010 in NW and about 20 in SH. In the time after the previous election, there is no or only one poll conducted, see e.g. the last election cycle in Schleswig Holstein. Hence, only data 365 days before election are included in the forecast calculation. The number of polls for the 13 states 365 days before the respective election date are shown in Table 5.5. For SH, there are only five polls available whereas for NW 21 are published in the investigated period. States of major importance are Berlin with 25 polls and BW with 23. In smaller states like Bremen, only six polls one year before are provided.

Table 5.5: Number of published polls 365 days before the respective election for the investigated states.

State	BB	BW	HB	HH	MV	NW	RP	SH	SL	SN	ST	TH	WBE
Polls	11	23	6	12	10	21	18	5	10	14	13	14	25

Sometimes missing data (not availables, NA) are produced by repeating the poll as long as a new poll is published until 365 days before election day. Institutes publish poll values for the parties which suffer from sampling inaccuracies of about 3%, compare https://www.zdf.de/politik/politbarometer/sb-material/methodik-100.html. Therefore some institutes publish a missing value if the vote share of polls is under 3% especially for small parties. To avoid too many missing values by repeating the latest poll value from an institute which published a NA value for a small party, the penultimate value is taken. These replaced values are seen below.[21]

The first columns show the states and the third column the lead days, where the NAs are replaced with the values displayed in column five for the certain party.

[21] E.g. the table below shows the NAs of several lead days of the states and the replaced values.

State	Election date	Lead days	Party	Value instead of NA
BB	2014-09-14	10-1	FDP	0.02
TH	2014-09-14	10-1	FDP	0.02
ST	2016-03-13	365-257	AFD	0.06
ST	2016-03-13	181-100	FDP	0.04
MV	2016-09-04	9-5	FDP	0.03
WBE	2016-09-18	365-355	FDP	0.02
SL	2017-03-27	365-321	FDP	0.04

6 Framework, methods and evaluation criteria for election forecasting

Most models of forecasting elections in literature are fundamental models, also called structural models e.g. by Lewis-Beck and Dassonneville (2015) compare Section 4.1. These types of models have a more static character and use socioeconomic variables like the unemployment rate as covariates. Mostly OLS regression is used to forecast the vote shares of elections in this case. In many cases, polls are included as additional short term information for models based on fundamental variables and on the prior election result, compare Holbrook and DeSart (1999). Fundamental models are difficult to setup for state elections, as the volatility in state elections is even higher than in national elections and the election cycle with five years is often longer than in national elections. The party landscape in states is higher diversified and new parties are usually passing the 5% threshold first, compared to national elections. As the election cycle in state elections lasts usually five years, the last election result as variable might have no great impact on the forecasted vote share. Also issue orientation changed over time. As seen in Figure 2.6, the unemployment rate has been in focus for the voters in the 2000s, but now the refugee crisis and inner security are of primary interest.

For state elections, regional factors are prevalent, e.g. the prime minister or local issues like "Stuttgart21" (the planned train station in BW). The unemployment rate or GDP are mainly available on a national level and play not a big role for state elections. Further, for structural models using fundamental variables, there is no clear issue-orientation in multi-party systems, as can be observed by the blurring borders in the party programs of the major parties, compare Aisch (2013).

Polls as data base improved over the last years in quantity and diversity compare Chapter 5. Data aggregation is introduced due to possible errors of the polls, e.g. aggregating across pollsters depending on the sample size, by giving polls with a smaller sample size a lower weight than ones with greater sample size. Especially in the US this type of aggregation has become common over the last years, compare "FiveThrityEight.com", "RealClearPolitics.com", "Princeton election consortium", "pollster.com" and the references Graefe and Armstrong (2014), Hillygus (2011) and Rothschild (2009). Other researchers aggregate polls depending on "house effects" as some institutes are said to be SPD-friendly like Forsa or CDU-friendly like Allensbach, compare Aisch (2013), Selb and Munzert (2016) and Wlezien and Erikson (2007). Aggregation may improve accuracy of polls, compare Bunker and Bauchowitz (2015) and may balance possible biases and house effects. Solely poll based models (so called aggregate models) are an exception in literature. Nevertheless, over the last years, polls as data base to forecast elections are included in several models , compare Erikson and Wlezien (2008), Selb and Munzert (2016) and Wlezien et al. (2013).

The models applied in this work are based solely on state level polls of different institutes up to 365 days before election, compare Chapter 5. Polls are not taken as forecast, but ahead of the election they are the best available value on irregular dates. By generating a daily time series of polls, explained in Chapter 5 and named here "latest poll value", a benchmark value is created. The major contribution is to investigate different forecasting methods, described in detail in Section 6.2, which are close to the real election result. The minor contribution of this thesis, is to beat the latest poll value in accuracy (see Haupt, Schnurbus and Huber, 2017). In contrast to polls, the forecasts in this work are calculated on a daily basis 120 days before election date. Hence, the forecasts fill the gap between two polls (which are sometimes wide apart each other in terms of time and value).

According to the existing forecasting literature, the statistical models and the latest poll value are evaluated by their forecasting accuracy (measured with e.g. RMSE, compare Hyndman et al. (2002)) and compared to the election result. In the following, the framework of election forecasting with the forecasting objective, lead time and the latest poll value are outlined in detail. Section 6.2 describes the applied methods and the evaluation criteria.

6.1 Framework and objective of election forecasting

Forecasting elections especially for multi-party elections is challenging in the right choice of the data base, method, lead time, variables and the objective (see Walther, 2015). Election forecasting methods are subject to the Golden Rules of forecasting, proposed by Lewis-Beck (2005), which are accuracy, lead time, parsimony, reproducibility.

Parsimony results from the problem, that a small number of elections corresponds to a small sample size. For regressions it is not possible to include many variables due to the sample size. If the sample size is smaller than ten, the degrees of freedom quickly evaporate (see Lewis-Beck and Tien, 1996). The rule of reproducibility means that a method should be easy to understand and implement in time and money.

A method is reproducible, if the included variables are available before election and measured similar over time e.g. in frequency. In this work, the data set is clearly specified as the polls are available on "wahlrecht.de" and updated. The methods are not costly and time intensive and are implemented for different states without changes. An example for violating this rule is the model of Gschwend and Norpoth (2005). They correct the data by deducting the chancellor popularity of Schröder by the amount of the Left party to forecast the vote share of the coalition of SPD and Greens for the 2005 German national election. This forecasting method is difficult to reproduce for another election as the party landscape changes very often. Models in this work are based on polls which provide a sufficient number of data and fulfill the rules of parsimony and reproducibility. If a new party appears or disappears like the AFD, the models can be updated with low computational costs. The third Golden rule requires that the lead time is chosen with an adequate horizon. The maximum lead time is the time between two elections (see Lewis-Beck, 2005). As Table 6.1 indicates, most research is made for US elections with lead times between six months and one day before election day, for UK from nine months to one month by Lewis-Beck, Nadeau and Bélanger (2004) and for UK and Ireland from six to one month by Lewis-Beck and Dassonneville (2015). For Germany, Kayser and Leininger (2013) provide forecasts with lead times from six to two months, Graefe (2015) for 67, 40, 19, 12 and 5 days. Walther (2015) calculates forecasts with one month lead time, Lewis-Beck and Dassonneville (2015) with six months to one month and Norpoth and Gschwend (2003) with three to one month. In this thesis, polls 365 days before election day are used to forecast single vote shares with lead times 120 to 1 day before the relevant election date.

The last golden rule is accuracy, which is commonly measured by RMSE and MAE in most cases, explained in detail in Section 6.3. Lewis-Beck (2005) compares the most common US election forecasting models of e.g. Abramowitz (1996) and Norpoth (2004) in terms of parsimony, lead time, reproducibility, mean absolute error, and a quality index including the four rules.

In forecasting elections, different objectives are in focus, compare Hillygus (2011) and Murr (2011), especially in multi-party elections the choice of the objective is a challenging task. Walther (2015) assumes that there is no clear dependent variable like in two-party systems, because it is sometimes not clear which parties are going to be part of the government. In literature, the most common objective is the specific vote share of a party e.g. Abramowitz (1996), Bellucci (2010), Duquette, Mixon and Cebula (2013) and Holbrook and DeSart (1999) provide forecasts for the incumbent party. Leigh and Wolfers (2006) and Lock and Gelman (2010) investigate only the winner of the election. Whereas others like Dopp (2012), Mughan (1987) and Rigdon, Sauppe and Jacobsen (2015) show the margin of victory. Heij and Franses (2011), Murr (2011) and Rothschild (2009) calculate the probability of a party or candidate winning the election. In the German multi-party system the probability of winning is not investigated in literature. For Germany, Kayser and Leininger (2013), Küntzler (2014) and Norpoth and Gschwend (2010) forecast the coalition or incumbent party. Jérôme, Jérôme-Speziari and Lewis-Beck (2013) and Selb and Munzert (2016) provide forecasts of the five main parties (details

Table 6.1: Applied lead times used in literature ordered by authors and countries.

Author	Lead time in days	Country
Berg, Nelson and Rietz (2003)	84, 56, 28, 14, 7, 1	US
Brown and Chappell (1999)	90, 60, 30, 3	US
Campbell (1996)	60	US
Gott and Colley (2008)	30,...,1	US
Graefe (2013b)	100,...,1	US
Holbrook and DeSart (1999)	60, 30	US
Klarner (2008)	99	US
Lebo and Norpoth (2011)	60	UK
Lewis-Beck, Nadeau and Bélanger (2004)	270, 180, 90	UK
Magalhaes, Aguiar-Conraria and Lewis-Beck (2012)	180	Spain
Mellon and Prosser (2016)	360,..60	UK
Rothschild (2009)	130,...,1	US
Walther (2015)	30	Germany, Sweden
Wang et al. (2015)	45	US
Whiteley et al. (2011)	180	UK
Wlezien et al. (2013)	150, 120, 90, 60, 30	US
Wolfers and Leigh (2001)	360,..1	Australia
Graefe (2015)	67, 40, 19, 12, 5	Germany
Lewis-Beck and Dassonneville (2015)	180, 150, 120, 90,	
60, 30	Germany, Ireland, UK	
Kayser and Leininger (2013)	180, 150, 120, 90, 60	Germany
Norpoth and Gschwend (2003)	90, 30	Germany

in Section 4.2). Kayser and Leininger (2017) summarize the new party AFD in "others" and investigate the other five main parties. In this work, the forecasting objective is the vote share of the five established parties and the new party AFD ("others" are neglected in this work).

FORECASTING OBJECTIVE

The literature provides several forecasting objectives e.g. seats or probability of winning. This thesis focuses on the vote share of the parties.[22] The objective of forecasting state elections is the vote share y, achieved by party p in region r at election e. At election day $h = 0$ the realized vote share is denoted as

$$y_0(p, r, e), \quad p = 1, \dots P, \; r = 1, \dots, R, \; e = 1, \dots, E.$$

In this case $p \,\epsilon\, \{CDU, SPD, FDP, LIN, GRE, AFD\}$,
$r \,\epsilon\, \{SN, BB, TH, HB, HH, BW, RP, ST, MV, WBE, SL, SH, NW\}$ and
$e \,\epsilon\, \{$ see election dates in Table 5.1$\}$. Forecasts are calculated for several lead times based on a time series of polls,

$$s_h(p, r, e), \quad h = H, H-1, \dots 1$$

where H is the number of available polls before election day e, in this case H is 365 days. As stated in Section 6.1, the lead time varies in literature, in this work the lead time is chosen between 120 and 1 day. In the following, the tuple (p, r, e) is omitted for sake of exposition (see Haupt, Schnurbus and Huber, 2017). Depending on a regular or irregular time series of polls, the forecast of y_0 is calculated with several statistical approaches given by

$$\hat{y}_{0|h} = \hat{g}(s_h, \dots, s_H). \tag{6.1}$$

where g is the predictor function in the forecast equation. The notation $\hat{y}_{.|h}$ represents a forecast based on all information available until time period h (see Haupt, Schnurbus and Huber, 2017). In the following, the different statistical approaches are explained in detail, starting with the latest poll value as benchmark.

[22] Notation analogous to Haupt, Schnurbus and Huber (2017).

LATEST POLL VALUE

The latest poll value (Lpv) is often seen as direct forecast in media and public compare Miller et al. (2012), but should be correctly defined as snapshot of the current political sentiment compare Subsection 3.3. Nevertheless, the latest poll value is supposed to be the best available value of the parties vote shares on a specific point during the election cycle after the previous election. The purpose of this work is to take the published poll value (namely the latest poll value) as a benchmark for the applied forecasting methods and to beat it in terms of accuracy see the work of Haupt, Schnurbus and Huber (2017). The statistical methods in this work use polls as data base. Hyndman and Athanasopoulos (2013) suggest the Lpv as simple forecast. A simple forecast in their opinion is the last observed value in a time series. In state elections, a poll value is not published on every day during the campaign, therefore the poll value is repeated until a new poll is published. This resulting latest poll value is taken as a benchmark forecast/value with

$$\widehat{y}_{0|h}^{(bench)} = s_h \tag{6.2}$$

In literature, the Lpv is used as benchmark in aggregate models (see Lewis-Beck and Dassonneville, 2015). Further Gelman and King (1993) compare their forecast at the beginning of the campaign with the available poll value. Sjöberg (2009) provides forecasts of expert groups, journalists and the public and applies the poll value as benchmark. Gibson and Lewis-Beck (2011) use a pool of polls two months ahead of the election and compare it with poll data from the British Election Study. Other research on the latest poll value is published by Schnell and Noack (2014) who use polls 30 up to 1 day before the German election as forecast. Others like Brown and Chappell (1999) and Holbrook and DeSart (1999) include polls as variables in their regression models to forecast vote shares of US elections, whereas Whiteley (2005) does this for UK elections. Mughan (1987) develops an opinion model with the final Gallup poll for the US, one month before election as covariate in a regression. For Germany Selb and Munzert (2016) build a model based on polling data six weeks before election with weights depending on house effects. Also Erikson and Wlezien (2008) see the latest poll value as a snapshot.

If there is more than one poll on a day they calculate the mean of these polls or take the latest one, if there is none on a day. Rothschild (2009) generates a latest poll value by creating a linear regression of the polls and using the trend of the regression as snapshot if there is none poll on a day.

6.2 Methods of election forecasting

In this work, different methods are used to forecast the German state elections in 13 states. The methods can be implemented for all elections, some methods need a regularized structure, explained in the latter Section. There are little differences in the poll databases for the investigated elections in terms of frequencies, but the methods are appropriate for all elections. Polls provide vote shares for the parties on a specific point in time, this value is called the latest poll value. The purpose of forecasting methods is to forecast the vote shares of parties or at least the correct winner of an election. Some forecasters only provide the probability of victory for one candidate in a two-party system see Section 6.1. In the following, the different methods are outlined in detail in Table 6.2. Some methods use the irregular structure of the poll data. These are basic methods like unweighted average (`avg`), weighted average (`avg.wgt`), median (`med`) and nonparametric regression-based methods like the local constant (`np.lc`) or local linear (`np.ll`) estimator.

Some methods need a regular, daily time series of polls. The parametric regression-based methods innovation state space (`iss`) and autoregressive integrated moving average (`arima`) deal with regularly generated daily poll values. Also the dynamic linear models in form of the local level (`dlm`) and the local linear (`dlm_lin`) ones use regularized poll data. The average window forecast (`avew`), which is an application of the work of Pesaran, Pick and Pranovich (2013) can be adopted on irregular or regularized time series (here it is applied on `arima`). Some of the methods are part of the work of Haupt, Schnurbus and Huber (2017) for the five state elections 2016 in BW, RP, ST, MV and WBE. Chapter 7 provides the results of the used methods on the 13 state elections.

Table 6.2: Overview of forecasting methods for state elections used in this work, based on irregular and regular time series. The second column indicates the abbreviation of the methods.

Methods based on irregular time series	Abbreviation
Unweighted average	`avg`
Weighted average	`avg.wgt`
Median	`med`
Nonparametric local constant	`np.lc`
Nonparametric local linear	`np.ll`
Methods based on regularized time series	
Innovation state space	`iss`
Autoregressive integrated moving average	`arima`
Dynamic linear (local level)	`dlm`
Dynamic linear (linear level)	`dlm_lin`
Applicable for all methods above	
Average window forecast	`avew`

6.2.1 Methods using an irregular time series of polls

According to Hyndman and Athanasopoulos (2013), there are simple methods, to perpetuate a time series. Besides taking the last observation, simple averaging methods can be applied. Methods capable of using an irregular time series of polls are the unweighted average (`avg`), the weighted average (`avg.wgt`) and the median (`med`). The first one is the unweighted average, outlined in the following.

Basic methods

UNWEIGHTED AVERAGE

A simple approach to forecast vote shares of polls is to average polls over the last 365 days before election. Averaging polls without weight is denoted as

$$\hat{y}_{0|h} = \frac{1}{H^*} \sum_{j=h}^{H} s_j, \tag{6.3}$$

where $H^* = H - h + 1$, if there is a poll on every day in the time series. The lead time h indicates the days before election date, also called the forecast horizon. Due to the irregular time series structure, H^* is the number of polls to be averaged. The average weights every poll equally and provides a forecast which puts equal weight on polls, irrespective of the sample size and recentness (also suggested by Hyndman and Athanasopoulos (2013)). Cuzan, Armstrong and Jones (2005) calculate in their popular PollyVote approach the average of Delphi, experts, polls and quantitative models with equal weights. Also “wahlumfrage.de”, compare Graefe (2015) computes the simple unweighted average the last 20 days before election of six pollsters in Germany (Forsa, FGW, Allensbach, GMS, Emnid and Infratest). In the case of election forecasting, the unweighted average can produce good forecasts, if the time series changes not much during the campaign or show ups and downs. Hence, in most cases, polls shortly before election generate better results than earlier in the campaign, compare Linzer (2013) and Mughan (1987), especially in state elections, where the number of poll increases shortly before election.

WEIGHTED AVERAGE

Another approach is the weighted average of polls (`avg.wgt`), computed as,

$$\hat{y}_{0|h} = \frac{1}{H^*} \sum_{j=h}^{H} s_j w_j, \quad \text{where } w_j = \frac{H-j}{H} \tag{6.4}$$

The weight w_j is linearly increasing with a decreasing time distance to the next election, as polls closer the election date are supposed to be more informative. A poll 365 days before election date obtains weight zero, while the weight is linearly increasing to one for more recent polls. Most researchers calculate the weighted average depending on sample size, compare Graefe (2015). “Pollitix.de” gives greater weight on polls with larger sample size and polls conducted more recently. Leigh and Wolfers (2006) also weight the polls by sample size and with adjusted variance for the most recent seven days in Australian elections. In the US averaging polls from different institutes by sample size and recentness is a new trend, which is published by “FiveThirtyEight.com” (from Nate Silver), “pollsters.com”, etc. (see Rothschild, 2009).

MEDIAN

The last basic method is calculating the median value of polls denoted as

$$median(s_h, s_{h+1}, \ldots, s_{H^*}). \tag{6.5}$$

Analogous to the unweighted average, the median also does not give recent polls more weight than polls conducted earlier in the campaign. With increasing lead time, the median value does not change much. In literature, Gott and Colley (2008) compute the median over polls of the last month before election. Lemennicier, Lescieux-Katir and Grofman (2010) publish a median voter model for French elections, where the median is used as a benchmark. Lewis-Beck and Dassonneville (2015) estimate their aggregate model by using the median of the German national polls in their function. They choose lead times from six to one month.

Nonparametric regression based methods

Nonparametric regression based forecasts are not common for election forecasting. However, in this work they are applied because they provide a type of weighting observations depending on smoothing parameters, called bandwidth. The bandwidth determines which set of observations is included and determines the weight of either observation. Nonparametric approaches are able to deal with the irregular time series structure of the poll data in contrast to `iss` and `arima`. If the parametric form $m(x, \theta)$ is unknown or cannot be modeled parametrically, nonparametric regression is used. The function $m(x)$ at each point x can be estimated by using nonparametric techniques for data generated by $y_i = m(x_i) + \epsilon_i$ in the notation of Hurvich, Simonoff and Tsai (1998). The unknown smooth function $m(\cdot)$ depends on x_i, the real numbers and ϵ_i, independent random variables with zero mean and variance σ^2 (Hurvich, Simonoff and Tsai, 1998, p. 271).

Nonparametric regression techniques estimate the unknown function $m(x)$ by e.g. a kernel regression or splines. In this work, only kernel regression is applied.

Kernel regression estimates the conditional expectation of Y given x with a weighted filter. Every estimator is determined by the choice of the kernel function $k(\cdot)$, the choice of the bandwidth b and the type of the local estimator. The kernel function determines the weight and imparts smoothness and differentiability on the estimator (Racine, 2008, p. 12). It can be modeled by e.g. three common non-negative symmetric kernel functions seen in Table 6.3 (Fan, 1992).

Table 6.3: Three common used kernel functions.

Uniform kernel:	$k(u) = 1/2$
Epanechnikov kernel:	$k(u) = 3/4 \cdot (1 - u^2)$
Gaussian kernel:	$k(u) = \frac{1}{\sqrt{2\phi}} exp(-\frac{u^2}{2})$

In this work, a second order Gaussian kernel is used. More important than the choice of the kernel function, is the bandwidth, which is a vector of the smoothing parameters. A larger bandwidth depends on more observations and provides a better smoothing with low variance but higher bias. Whereas a smaller bandwidth, depending on less observations provides less smoothing but a higher variance and a low bias (see Racine, 2008).

The bandwidth can be selected by several methods and is specified in advance of the forecast. According to Racine (2008), there are four methods to select the optimal bandwidth, the rule of thumb, plug-in methods, bootstrap and cross-validation methods.

Here, the cross-validation is used to select the optimal bandwidth b and is calculated as,

$$CV(b) = \sum_{i=h}^{H} [y_i - \hat{m}_{-i}(x_i)]^2 \tag{6.6}$$

by Aneiros-Perez, Cao and Vilar-Fernandez (2011), Fan (1992), Hurvich, Simonoff and Tsai (1998), Li and Racine (2004) and Racine and Li (2004). Where $\hat{m}_{-i}$ is an estimator, which only depends on $H - 1$ observations and leaves out the i-th one.

Further, the type of the kernel estimator has to be chosen to estimate $m(x)$. Common estimators are the p-th order polynomial estimator or the Gasser-Müller estimator (see Hurvich, Simonoff and Tsai, 1998). In this thesis, the p-th order local polynomial estimator is used, where the polynom of grade p on every point x is locally weighted adapted. If $p = 0$, the local constant estimator is obtained, for $p = 1$, the local linear case. The following minimization problem in the notation of Hurvich, Simonoff and Tsai (1998),

$$\underset{\beta}{argmin} = \sum_{i=h}^{H} \{y_i - \beta_0 - \cdots - \beta_p(x - x_i)^p\}^2 k(\frac{x - x_i}{b}) \tag{6.7}$$

determines $\hat{\beta}$. Given a kernel weight function $k(u)$, a kernel estimator of $m(x)$ has to be calculated by solving this minimization problem.

Nonparametric local constant regression

The local constant estimator $\hat{\beta}$ is obtained by solving the minimization problem of Equation (6.7), for the case of $p = 0$, minimizing $\sum_{i=h}^{H}(y_i - \beta_0)^2 k(\frac{x - x_i}{b})$ provides

$$\hat{m}_b(x) = \hat{\beta}_0 = \sum_{i=h}^{H} w_i y_i = \sum_{i=h}^{H} \frac{k(\frac{x-x_i}{b})}{\sum_{j=h}^{H} k(\frac{x-x_j}{b})} y_i \tag{6.8}$$

which is the local constant or also called the Nadaraya-Watson estimator. The local constant estimator has the disadvantage, that at the boundary, the kernel smoother exhibits bias. If the bandwidth increases, the estimator simplifies to a constant (see Racine, 2008).

Nonparametric local linear regression

Therefore a local linear kernel estimator could be calculated, which mimics deterministic trending behavior, but does not rule out the possibility, that the process has a mean reverting behavior (as the first partial derivative w.r.t. h, days to election can be equal or close to zero). Again, by solving the minimization problem of Equation (6.7) for the case $p = 1$, the local linear kernel estimator is obtained by Fan (1992),

$$\begin{aligned} \hat{m}_b(x) \quad &= \frac{\sum_{j=h}^{H} k(\frac{x-x_i}{b}) y_i}{\sum_{i=h}^{H} k(\frac{x-x_i}{b})} \\ \text{where } w_i \quad &\equiv k(\frac{x-x_i}{b})[s_{H,2} - (x-x_i)s_{H,1}], \\ \text{with } s_{n,l} \quad &= \sum_{i=h}^{H} k(\frac{x-x_i}{b})(x-x_i)^l, with\ l = 1,2 \end{aligned} \tag{6.9}$$

gives a weighted local average of the observations.

6.2.2 Methods using a regularized time series of polls

Parametric regression based methods

Most research on election forecasting deals with fundamental variables and uses OLS regression compare Chapter 4. In this work, fundamental variables are not in focus for state election forecasts, therefore OLS regression is not part of this work. Instead, time series models are introduced in the following. The parametric regression-based standard times series ARIMA (`arima`) and the nonlinear innovation state space (`iss`) approaches are outlined in detail. They need a regularized time series structure in contrast to the methods displayed in Section 6.2.1. Hence, a regular time series is generated by repeating the poll value as long as a new poll is published, compare Chapter 5. In this work the most common automatic forecasting algorithms ARIMA and exponential smoothing (respectively `iss`) are used for the univariate time series structure (see Hyndman and Khandakar, 2008).

INNOVATION STATE SPACE

The exponential smoothing framework has been developed in the 1950s, but likelihood calculators, prediction intervals and model selection procedures were developed later by Hyndman et al. (2002) and others.[23] The advantages of exponential smoothing lie in its easy application on small and different time series and that it is easy in handling (see Diebold, 2007). Despite, Makridakis and Hibon (2000) point out, there is no improvement in accuracy for more complex models.

The general notation of an ETS-model is ETS(**E**rror, **T**rend, **S**eason). Table B.2 in the Appendix, provides an overview of the 15 possible ETS-models by (Hyndman and Khandakar, 2008). The trend component is differentiated in None (N), Additive (A), Additive damped (Ad), Multiplicative (M) and Multiplicative damped (Md), the Seasonal component in None, Additive and Multiplicative, the error type is additive or multiplicative. According to Hyndman and Athanasopoulos (2013), a time series Y_t can be decomposed in $Y_t = f(S_t, T_t, E_t)$, with $S_t =$ seasonal component, $T_t =$ trend component at period t and $E_t =$ error component. An additive decomposition is $Y_t = (S_t) + T_t + E_t$, a multiplicative decomposition is $Y_t = (S_t) * T_t * E_t$. In this work, the season is not considered, therefore the term S_t can be neglected in the equations above.

For the 13 elections under consideration, the following ETS-models are chosen:

ETS(A,N,N) $=$ Simple exponential smoothing with additive errors
ETS(M,A,N) $=$ Holt Winters with multiplicative errors
ETS(M,N,N) $=$ Simple exponential smoothing with multiplicative errors
ETS(M,Ad,N)$=$ Additive damped trend method with multiplicative errors

Exponential smoothing provides h-step ahead forecasts and also gives greater weight on more recent observations than on earlier ones for a time series $y_1, ..., y_T$ (in the case of this work $s_1, ..., s_T$, the poll values). According to Hyndman and Athanasopoulos (2013), the component form of e.g. ETS(A,N,N) is given by,

$$\begin{aligned} \hat{y}_{t+1|t} &= l_t && \text{forecast equation} \\ l_t &= \alpha y_t + (1-\alpha) \cdot l_{t-1} && \text{smoothing equation} \end{aligned} \tag{6.10}$$

[23] Further information on the history of exponential smoothing can be found in De Gooijer and Hyndman (2006).

where α is the smoothing parameter ranging from zero to one. Higher values of α indicate greater weight on more recent values. The estimate of the level of the time series is denoted by l_t with smoothing parameter α. The other ETS-models include b_t (trend estimate) and a damping parameter ϕ with smoothing parameter β. The ETS-methods of this work are displayed in Table 6.4

Table 6.4: Exponential smoothing methods applied in this work.

ETS(M,N,N)	analogous to ETS(A,N,N) see Equ. (6.10)	
ETS(M,A,N)	$\hat{y}_{t+1\|t} = l_t + hb_t$	Forecast equation
	$l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1})$	Smoothing equation
	$b_t = \beta^*(l_t - l_{(t-1)}) + (1-\beta^*)(b_{(t-1)})$	Trend equation
ETS(M,Ad,N)	$\hat{y}_{t+h\|t} = l_t + \phi_h \cdot b_t$	Forecast equation
	$l_t = \alpha y_t + (1-\alpha)(l_{t-1} + \phi b_{t-1})$	Smoothing equation
	$b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)\phi b_{t-1}$	Trend equation

Exponential smoothing only provides an algorithm for point forecasts (see Hyndman and Khandakar, 2008). Therefore for each ETS-method two innovation state space models with additive and multiplicative errors are employed which provide also prediction intervals. The general state space representation consists of the observation equation and the state equation (one for each state: level, trend, season). According to Hyndman et al. (2002) and Ord, Koehler and Snyder (1997) these equations are denoted as

$$\begin{aligned} &\text{observation equation} \quad && y_t = h(x_{t-1}) + k(x_{t-1}) \cdot \epsilon \\ &\text{state equation} && x_t = f(x_{t-1}) + g(x_{t-1}) \cdot \epsilon \end{aligned} \tag{6.11}$$

Here, y_t is the vote share at time t (also called s_t). Whereas ϵ_t is a Gaussian white noise process with zero mean and variance σ^2, x_{t-1} is the state of the underlying process. Further, $\mu_t = h(x_{t-1})$ then $Y_t = \mu_t + e_t$, as $e_t = k(x_{t-1}) \cdot \epsilon$. A model with additive errors is

$$Y_t = \mu_t + \epsilon_t, \tag{6.12}$$

with $\mu_t = F_{(t-1)+1}$ as one-step ahead forecast at time t-1 and $k(x_{t-1}) = 1$. Whereas a model with multiplicative errors is

$$Y_t = \mu_t + (1 + \epsilon_t), \tag{6.13}$$

with $\mu_t = k(x_{t-1})$ and $\epsilon_t = \frac{e_t}{\mu_t} = \frac{Y_t - \mu_t}{\mu_t}$. Table 6.5 illustrates the state space representation of the models used in this thesis.

Table 6.5: Innovation state space representation.

ETS(A,N,N)	y_t	$= l_{t-1} + \epsilon_t$
	l_t	$= l_{t-1} + \alpha\epsilon_t$
ETS(M,N,N)	y_t	$= l_{t-1} \cdot (1 + \epsilon_t)$
	l_t	$= l_{t-1} \cdot (1 + \alpha\epsilon_t)$
ETS(M,A,N)	y_t	$= (l_{t-1} + b_{t-1}) \cdot (1 + \epsilon_t)$
	l_t	$= (l_{t-1} + b_{t-1}) \cdot (1 + \alpha\epsilon_t)$
	b_t	$= b_{t-1} + \beta(l_{t-1} + b_{t-1})\epsilon_t$
ETS(M,Ad,N)	y_t	$= (l_{t-1} + \phi b_{t-1}) \cdot (1 + \epsilon_t)$
	l_t	$= (l_{t-1} + \phi b_{t-1}) \cdot (1 + \alpha\epsilon_t)$
	b_t	$= \phi b_{t-1} + \beta(l_{t-1} + \phi b_{t-1})\epsilon_t$

The automatic forecasting algorithms are applied on all appropriate models and optimized with Maximum Likelihood Estimation (MLE). In a next step the best model is chosen according to $AIC = L \cdot (\hat{\theta}, \hat{x_0}) + 2q$, compare Hyndman and Khandakar (2008). Innovation state space models allow for nonlinearities in contrast to ARIMA and allow for the calculation of prediction intervals.

AUTOREGRESSIVE INTEGRATED MOVING AVERAGE

A second automatic forecasting algorithm is based on `arima` models (Autoregressive integrated moving average). ARIMA(p,d,q) models transform a non-stationary time series in a stationary time series through d times differencing. ARIMA connects three components, the integrated I(d) part, the autoregressive AR(p) part and the MA(q) component, which assigns the moving average. The ARIMA process in the notation of Hyndman and Khandakar (2008) is given by

$$\phi(B)(1 - B^d)y_t = C + \theta(B)\epsilon_t, \tag{6.14}$$

where ϵ_t is a white noise process with zero mean and variance σ^2. The Back Shift operator is B and $\phi(B)$ and $\theta(B)$ are the polynomials of order p respective q ($\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$).

In literature, time series models for election forecasting are calculated in form of an ARMA(1,1) model for German elections by Schaffer and Schneider (2005), who forecast the vote shares of CDU, SPD, LIN and GRE depending on election stock markets, namely "wahlbörse.de". For the UK case, Lebo and Norpoth (2011) provide an AR(1)-model for the seats and votes for the Conservatives and their major-party opponent. Sanders (1995) also uses an AR(2)-model for UK, whereas Berg, Nelson and Rietz (2003) use an AR(1)-model for prediction markets to forecast the vote shares of US elections.

DYNAMIC LINEAR MODELS

The next statistical approach is the dynamic linear model (`dlm`). Forecasting elections with Dynamic Linear Models (DLM) has become very common in literature over the last years compare Jackman (2005) and Linzer (2013). Most work on DLMs for elections deals with the two-party US case, this work focuses on the German multi-party case for state elections. Dynamic linear models are also a class of state space models. They are able to handle univariate and multivariate time series (see Petris, 2010). They are flexible to capture different data types (e.g. discrete data) and trends (e.g. upward and downward trends of parties in polls). Further, DLMs are models which are updated at each time step and good in forecasting short term forecasts like in the case of this work (120 days to 1 day)(see West and Harrison, 1997). Furthermore, DLMs allow for a number of specifications. In this work, the focus lies on local level and local linear trend models. The DLMs are used in a Bayesian context, where parameters are treated as random variables and described via a probability distribution. The idea of Bayesian statistics is to combine existing knowledge in form of a prior distribution with the likelihood gained from new data. The result is the posterior distribution, which combines the old and the new information. Hence, this distribution can be seen as the new prior until new data is available (see Petris, Petrone and Campagnoli, 2009). As it is difficult to solve the posterior distribution, the e.g. mean of a function can be approximated by simulation techniques like the Monte Carlo method. This method simulates random variables with a Markov Chain, compare Petris, Petrone and Campagnoli (2009). In the case of this work, all DLMs use linear Gaussian distributions. Nevertheless, DLMs can also deal with nonlinear distributions e.g. "exponential family" for modeling a discrete time series. Moreover, non-Gaussian distributions are given in form of Hidden Markov DLMs for modeling structural breaks (see Petris, Petrone and Campagnoli, 2009). Also simple dynamic regression can be modeled with dynamic linear models. Here, a-th order polynomial dynamic linear models are computed. The first order polynomial model is also called "local level" (`dlm`). It is used, if there is no season or trend and the observations are modeled as random fluctuations (see Petris, Petrone and Campagnoli (2009, p. 42)). The fluctuations are modeled around a level, which evolves randomly over time. Additionally, the second order polynomial model "local linear" (`dlm_lin`) is used, which includes a time varying slope (Petris, Petrone and Campagnoli, 2009, p. 43).

The general notation of a DLM by Petris, Petrone and Campagnoli (2009) and West and Harrison (1997) is presented in Equation (6.15). In the state space form the DLM is divided in an observation equation and a state equation:

$$\begin{array}{llcl}
\text{observation equation} & \overset{(m\times 1)}{y_t} = \overset{(m\times p)}{F_t}\,\overset{(p\times 1)}{\theta_t} + \overset{(m\times 1)}{v_t} & & v_t \backsim N_m(0, V_t) \\
\text{state equation} & \overset{(p\times 1)}{\theta_t} = \overset{(p\times p)}{G_t}\,\overset{(p\times 1)}{\theta_{t-1}} + \overset{(p\times 1)}{w_t} & & w_t \backsim N_p(0, W_t)
\end{array} \tag{6.15}$$

A DLM is characterized by the system matrices $\{F_t, G_t, V_t, W_t\}$, according to Petris, Petrone and Campagnoli (2009), where each one can be dependent or independent upon time. The observation or measurement equation describes the evolution of observations y_t through signal θ_t and error v_t. Whereas the state (system, transition) equation describes the evolution of the state vector, using a first order Markov structure. In a first order Markov structure, the state in the future only depends on the current state.

The observable process (Y_t), which only depends on θ_t is determined by a latent process θ_t up to Gaussian random errors. The latent process (unobservable, state) θ_t has a simple dynamics, but does only depend on the previous state θ_{t-1}. According to West and Harrison (1997, p. 101), F_t is the design matrix of known values of independent variables, θ_t (respective θ_{t-1}) is the state or system vector. G_t is the evolution (system, transition or state) matrix as West and Harrison (1997) indicate. The DLM obtains two errors, the observation error v_t and the evolution error w_t. A prior distribution for the initial (pre-sample) state θ_0 is assigned to the model of Petris (2010, p. 2), with

$$\theta_0 \backsim N_p(m_0, C_0), \text{ with mean } m_0 \text{ and variance } C_0. \tag{6.16}$$

The observation and system errors v_t and w_t are independent of the initial information. The aim of DLMs is to estimate the unobserved states or to predict future observations like the vote shares of the polls. For a completely specified DLM (no unknown parameters), Kalman Filtering and smoothing algorithms are used to obtain the means and variances of the conditional distributions of the unobserved states, (see Petris (2010, p. 4)).[24] In the more realistic case, where a DLM is not completely specified, the unknown parameters have to be estimated. In an a-th order polynomial model e.g. the variances W_t and V_t. The model matrices of the DLM contain unknown parameters, stated in a vector ψ, (see Petris, Petrone and Campagnoli (2009, p. 115)). According to Petris (2010, p. 5), there exist two approaches to estimate the unknown parameters of a DLM, Maximum Likelihood Estimation and Bayesian Inference. Estimation of the unknown parameters ψ with MLE causes high standard errors and therefore inaccuracies.[25] This leads to an application of a Bayesian approach, in which posterior distributions can handle inaccuracies, smoothing and forecasting is applied to the values Petris (2010, p. 7).

[24] Filtering and smoothing in this case may lead to unstable issues and not positive semidefinite variance matrices.

[25] Maximum Likelihood Estimation (MLE):
There are n random vectors $Y_1, \dots Y_n$, whose distribution depends on the unknown

In the context of the Bayesian inference, an unknown parameter vector ψ and the states history $\theta_{0:n}$, given $y_{1:n}$ observations is solved by computing the posterior distribution $\pi(\psi, \theta_{0:n}|y_{1:n})$. The implementation and notation in the following is based on Petris, Petrone and Campagnoli (2009). To obtain the posterior distribution with the Bayes rule, MCMC (Markov Chain Monte Carlo) is needed. In Bayesian Inference, the parameters of the posterior distribution are unknown. Hence, simulation Monte Carlo methods, which simulate random variables from a Markov Chain are used to give an approximate posterior distribution (see Petris and Petrone, 2011). As MCMC is prior and model specific, it is no algorithm, that works in all cases, therefore auxiliary tools are needed.[26] These tools are the Gibbs sampler, marginal samplers and hybrid samplers. Gibbs sampler and hybrid sampler (a combination of marginal and Gibbs) generate draws from the joint posterior distribution of the states and parameters (Petris, Petrone and Campagnoli, 2009, p. 132). With the unknown parameter vector ψ and prior distribution $\pi(\psi)$, a sample from the joint posterior of states and parameters is obtained by Equation (6.17) in accordance to Petris, Petrone and Campagnoli (2009),

---

parameterψ. The joint density is

$$p(y_1, ..., y_n; \psi) = \prod_{t=1}^{n} p(y_t|D_{t-1}; \psi)$$

where $p(y_t...)$ is the conditional density of y_t, given the data, (see Petris, Petrone and Campagnoli (2009, p. 116)). The maximum likelihood estimator is denoted as:
$\hat{\psi} = \underset{\psi}{argmax} l(\psi)$, with loglikelihood

$$l(\psi) = -\frac{1}{2}\sum_{t=1}^{n} log|Q_t| - \frac{1}{2}\sum_{t=1}^{n}(y_t - f_t)' Q_t^{-1}(y_t - f_t)$$

[26] Generally speaking, a MCMC algorithm produces random variables with the transition kernel of a irreducible, aperiodic Markov Chain, which has the wanted distribution as stationary distribution π. To find the transition kernel, which has the target distribution as its invariant distribution, a Gibbs sampler is needed (see West and Harrison, 1997). After choosing a starting value, the values of a Markov Chain are simulated with the transition kernel (explained in detail in Table 6.6).

$$\pi(\theta_{0:T}, \psi | y_{1:T}) = \pi(\theta_{0:T} | \psi, y_{1:T}) \pi(\psi | y_{1:T}), (y_{1:T} \text{ means } y_1, y_2, ..., y_T) \quad (6.17)$$

The implementation for the joint posterior is made with the Forward Filtering Backward Sampling, which is outlined in detail in the Table 6.6.

Table 6.6: Forward Filtering Backward Sampling in a Gibbs sampler.

1. Set $\psi = \psi^{(0)}$
2. For the sample $i = 1, ..., N$:
draw $\theta_{0:T}^{(i)}$ from $\pi(\theta_{0:T} \| y_{1:T}, \psi = \psi^{(i-1)}$ by using FFBS
draw $\psi^{(i)}$ from $\pi(\psi \| y_{1,...,T}, \theta_{0:T} = \theta_{0:T}^{(i)})$

Table based on Petris, Petrone and Campagnoli (2009, p. 132).

The first step is initializing the unknown parameter vector ψ. Then for a sample of $1, ..., N$, the $\pi(\theta_{0:T}, \psi | y_{1:T})$ can be obtained by two steps. Drawing the states given the parameters and data from the conditional distribution using FFBS (Forward Filtering Backward Sampling). In the next step, the unknown parameters are drawn from the conditional distribution, the states and the data. In this work, the unknown parameters ψ are the variances and the observations, the data are the poll values of the parties.

APPLICATION ON THE MODELS

The **local level model** is computed, using an a-th order polynomial model with order $a = 1$, the **local linear trend model** with order $a = 2$. The first step of the MCMC Bayesian analysis with the Gibbs sampler and FFBS is drawn with 1000 Gibbs iterations and 100 as burn in, saving every tenth to reduce autocorrelation in MCMC samples. In the Appendix A, Figure A.10 shows the sample means for the evolution variance with 1000 iterations, Figure A.11 the sample means for the observational variance for a local level model in NW state elections.

The joint posterior distribution $\pi(\sigma_\varepsilon^2, \sigma_\xi^2, \theta_{0:n} | y_{1:n})$ is also calculated and the unknown variances d-inverse Gamma priors are assumed as follows (see Petris (2009)),

$$(\sigma_\varepsilon^2)^{-1} \backsim \Gamma(\alpha_\epsilon, \beta_\epsilon)$$
$$(\sigma_\xi^2)^{-1} \backsim \Gamma(\alpha_\xi, \beta_\xi), \text{with } \alpha = \text{shape}, \beta = \text{rate parameter} \tag{6.18}$$

FILTERING, SMOOTHING AND FORECASTING

Given the data $D_t = (y_1, ..., y_t)$, inference is solved by the three possibilities (filtering, smoothing and forecasting).

FILTERING (s=t)

Filtering means estimating the current value of the state vector. With the recursive algorithm, the Kalman Filter, the current best is updated, whenever a new observation is obtained (see Petris, Petrone and Campagnoli (2009, p. 84)). The algorithm denoted by Petris, Petrone and Campagnoli (2009, p. 50) works as follows:

1. Starting with an initial value $\theta_0 \backsim N(m_0, C_0)$
2. One-step predictive density for state θ_t given D_{t-1} : $p(\theta_t|D_{t-1})$
3. One-step predictive density for observations Y_t given D_{t-1} : $p(Y_t|D_{t-1})$
4. The filtering density is obtained by the Bayes rule $p(\theta_t|D_t)$

(6.19)

SMOOTHING (s<t)

The smoothing algorithm is used in retrospective analysis, where the systems behavior is examined, underlying the observations. A backward recursive algorithm is used, starting with a sample $\theta_T|Y_T$ from the filtering density $p(\theta_T|D_T)$ and estimating backwards the history of all states (Petris, Petrone and Campagnoli, 2009, p. 57).

FORECASTING (s>t)

The focus of this work lies on forecasting the vote share of parties. Future values are forecasted, based on prior observations Y_t, e.g. polls respectively the vote shares. The h-step ahead forecast is stated as follows in accordance to Petris, Petrone and Campagnoli (2009, p. 64).

Figure 6.1: Flow of information from the prior observations $Y_1, ..., Y_t$ to the forecasts Y_{t+h}.

$$
\begin{array}{lll}
\text{states} & \theta_t \longrightarrow \theta_{t+1} \longrightarrow \dots & \theta_{t+h} \\
 & | & | \\
\text{observations} & Y_1, ..., Y_t & Y_{t+h}
\end{array}
$$

In Figure 6.1 the $Y_1, ..., Y_t$ observations provide information on the states θ_t, the states provide information on the observations in $t + h$. Due to the Markovian nature of the model, the filtering distribution acts like an initial distribution for the future evolution of the model. In accordance to Petris, Petrone and Campagnoli (2009, p. 64), forecasting is made by the following Equation (6.20) with

$$
\begin{aligned}
a_t(h) &= E(\theta_{t+h}|D_t) \\
R_t(h) &= Var(\theta_{t+h}|D_t) \\
f_t(h) &= E(Y_{t+h}|D_t) \\
Q_t(h) &= Var(Y_{t+h}|D_t) \\
a_t(0) &= m_t \text{ and } R_t(0) = C_t
\end{aligned}
\tag{6.20}
$$

First, start with a sample $\theta_t|D_t \backsim N(m_T, R_T)$. Second, forecast the state θ_{t+h}, given D_t is Gaussian, then forecast the observation Y_{t+h}, given D_t is Gaussian. In the case of election forecasting, the observations are the polls for the six parties and h are the lead days from 120 up to 1 day before the future election day. The starting value is the value $365 + h$ days before election day, where h is the lead time. This value sets the Markov chain in motion.

In literature, dynamic linear models for elections were very popular in the last years. For Chile elections, Bunker and Bauchowitz (2015) provide a dynamic linear model, based on polls. Jackman (2005) uses Bayesian methods to estimate the unknown parameters of the vote shares, the house effects and the day to day variability for Australian elections based on polls. Hence, the MCMC method with Gibbs sampler are outlined in detail. A Bayesian dynamic linear model with shocks modeled by reverse random walks is developed by Linzer (2013). Walther (2015) applies a DLM with a seasonal component (based on house effects and the deviation from the median of the houses) on German and Swedish polls.

AVERAGE WINDOW FORECAST

Additionally to the methods stated above, an average window forecast (`avew`) is provided for the 13 German state elections, according to Jungmittag (2016) and Pesaran, Pick and Pranovich (2013). The method `avew` is calculated for the method `arima`, analogous to Haupt, Schnurbus and Huber (2017), who compute this method for the five state elections in 2016. This approach is suitable for time series with structural breaks, where no exact information about the breaks is needed (see Jungmittag, 2016). In the case of polls, unknown events may appear during the campaign. E.g. positive or negative news, which influence the parties (Haupt, Schnurbus and Huber, 2017, p. 8) or unforeseen events like the refugee crisis, which led to a push or loss for some parties, e.g. the AFD. The average window forecasts can be calculated for each method displayed above, irrespective of the given data base (irregular or regular). The idea of `avew` is, that averaging over estimation windows improves the accuracy of a forecast, compare Jungmittag (2016) and Pesaran and Timmermann (2007). The formula of `avew` is denoted as follows

$$\frac{1}{m}\sum_{i=1}^{m}\hat{y}_{0|h}(w_i) \tag{6.21}$$

with averaging m forecasts for estimation windows $w_1, ..., w_m$. The h-step ahead forecast $\hat{y}_{0|h}(w_i)$ from an estimation window is calculated for the method `arima` in this work. In Table B.3 in the Appendix the expanding estimation windows are displayed (Haupt, Schnurbus and Huber, 2017, p. 8). The size of the expanding estimation windows is ranging from four weeks to 24 weeks, which means that there exist 21 estimation windows with data from 28 up to 168 days.

6.3 Selection of the evaluation criteria to measure the forecasting quality

The poll value in form of the latest poll value Lpv is not only seen as a snapshot of the current opinion, but also as a forecasting tool by Lewis-Beck and Dassonneville (2015). The major contribution of this work is to study several statistical forecasting methods seen in Section 6.2, which are supposed to beat the poll value in accuracy. To assume, which method performs best, different evaluation criteria are under review, which are explained and discussed in the following.

THEORY OF EVALUATION CRITERIA

Armstrong and Collopy (1992) and Hyndman and Koehler (2006) present different out-of-sample accuracy measures to evaluate the forecasting performance of different methods. The purpose of forecasting elections is to find a method, which provides a forecast near the real election result. Hence, the h-step forecast error of a specific method $m \epsilon \{1, ...M\}$, depending on the forecast horizon (lead time h) $h \epsilon \{1, ..., 120\}$ is

$$e_h^{(m)} = \hat{y}_{0|h}^{(m)} - y_0, \tag{6.22}$$

where y_0 is the election result at $h = 0$ and $\hat{y}_{0|h}^{(m)}$, the forecast value depending on h and the method m (see Haupt, Schnurbus and Huber, 2017). If the error is positive, the election result is overestimated, if it is negative, the result is underestimated by the method. The forecasting performance is evaluated by computing the following accuracy measures in Equation (6.23), compare Armstrong and Collopy (1992) and Hyndman and Koehler (2006).[27]

[27]Notation analogous to Haupt, Schnurbus and Huber (2017, p. 8).

$$\begin{array}{lll} \text{Mean squared error} & MSE^{(m)} & = mean({e_h^{(m)}}^2) \\ \text{Root mean squared error} & RMSE^{(m)} & = \sqrt{MSE^{(m)}} \\ \text{Mean absolute error} & MAE^{(m)} & = mean(|e_h^{(m)}|) \\ \text{Median absolute error} & MDAE^{(m)} & = median(|e_h^{(m)}|) \end{array} \tag{6.23}$$

These measures are scale dependent measures which are discussed with their advantages and disadvantages.[28]

The measures, denoted in Equation (6.23) can be averaged with regard to all dimensions of party, region and election, depending on lead time (see Haupt, Schnurbus and Huber, 2017). The effect of different lead times on accuracy should be analyzed, e.g. for the RMSE as follows,

$$RMSE^{(m)} = \left(\frac{1}{H} \sum_{j=h}^{H} {e_h^{(m)}}^2 \right)^{1/2} \tag{6.26}$$

for measuring (out-of-sample) accuracy of method m over the forecast horizon indexed by h (Haupt, Schnurbus and Huber, 2017, p. 8). The performance of an election forecasting method is evaluated by the different accuracy measures. In the following, the choice of the evaluation criteria for the applied methods is discussed.

[28] Scale independent measures compare Hyndman and Koehler (2006) are the following:

$$\begin{array}{ll} \text{Mean absolute percentage error} & MAPE^{m} = mean(|p_h^{(m)}|) \\ \text{Median absolute percentage error} & MDAPE^{m} = median(|p_h^{(m)}|) \\ \text{with} & p_h^{(m)} = \dfrac{100 \cdot e_h^{(m)}}{y_0} \end{array} \tag{6.24}$$

Other accuracy measures are based on relative errors, where each error is divided by the error obtained by using another standard method of forecasting, which are only rarely used,

$$\begin{array}{ll} \text{Mean relative absolute error} & MRAE^{(m)} = mean(|r_h|) \\ \text{Median relative absolute error} & MDRAE^{(m)} = median(|r_h|) \\ \text{with } r_h = \dfrac{e_h^{(m)}}{e_h^{(m*)}} & e_h^{*} = \text{benchmark method.} \end{array} \tag{6.25}$$

CHOICE OF THE APPROPRIATE EVALUATION CRITERIA

The measures can be divided in scale dependent measures (RMSE, MDAE, MAE and MSE) and scale independent measures (e.g. MAPE). Scale dependent measures are often used in literature, compare Armstrong and Collopy (1992), with the RMSE and MSE as the most popular ones. These types of measures are suitable to compare time series with the same scale like in this thesis, where the forecasts are all based on poll data. In contrast to the MSE, the RMSE is on the same scale as the data (Hyndman and Koehler, 2006, p. 682) and therefore easier to interpret. The remaining scale dependent measures MDAE and MAE are less sensitive to outliers in comparison to RMSE and MSE as Hyndman and Koehler (2006) investigate. Scale dependent measures like the MAPE, MDAPE are not suitable, as positive errors are punished more than negative ones and the poll values are only positive and near zero, which can lead to biased results, compare Hyndman and Koehler (2006) and Tashman (2000).

In election forecasting literature, frequently applied methods are RMSE, MAE and MSE. Lewis-Beck (2005), Montgomery, Hollenbach and Ward (2012) and Wolfers and Zitzewitz (2006) measure accuracy with RMSE and MAE. Most forecasters calculate the RMSE and MAE, e.g. Armstrong and Graefe (2011), Bélanger, Lewis-Beck and Nadeau (2005), Berg, Nelson and Rietz (2003), Leigh and Wolfers (2006) and Lockerbie (2008) only the MAE. Strauss (2007) provides the averaged RMSE over 15 elections for the US. Rothschild (2009) and Schnell and Noack (2014) evaluate their forecasts by MSE. Lewis-Beck and Dassonneville (2015) calculate the RMSE, averaged over the national elections and for selected lead times of one to six months. For the German case, Ganser and Riordan (2015) calculate the RMSE and MAPE for their models based on vote expectation polls.

In regression based literature, Aichholzer and Willmann (2014), Graefe (2013b), Kayser and Leininger (2013), Wlezien and Erikson (2004) evaluate their models by the in-sample measure, coefficient of determination. In this work, the four scale dependent measures MSE, RMSE, MAE and MDAE are calculated with different levels of aggregation, to evaluate the performance of the forecasting methods.

7 Empirical application

7.1 Aims and technical computation of the forecasting methods

Polls for state elections are conducted at irregular intervals in contrast to national polls, which usually are published weekly. The quality of polls for state elections varies over the election cycle and between states. As stated in Chapter 5, some states are of greater interest like Northrhine Westphalia or of less interest like Saarland. Nevertheless, similar to polls on all levels, the number of polls increases during the campaign. The state level polls have been of lower interest over the last decades and now gain in importance for several reasons. State elections have a "barometer function" for national elections and in the German case they are stretched over a four-year national election cycle. The 13 state elections investigated in this work, are stretched over the period from "2014-08-31" up to "2017-05-14". National German elections took place between "2013-09-22" and "2017-09-24".[29]

Polls are often seen as direct forecast in media and public (see Miller et al., 2012). These "trial heat polls" own a horse race character especially shortly before election day. However, a poll value is more a snapshot of the current voter sentiment on a specific day during the campaign compare Section 3.3. Polls are not available on a daily basis, but are the best "forecast" during the campaign on a specific date.

[29] The three remaining state elections in BY, NI and HE took place before the national election in "2013-09-22" and are not part of this work.

With the question for "whom to vote for, if the election was held next Sunday", the polls seem to have a forecasting character. The minor contribution of this thesis is to compare the forecasts with the latest poll value as a benchmark method. In contrast to polls, forecasts are obtained on every day before the election. For state level polls sometimes only few polls are published, e.g. in SH only five polls were published in the year before the election. The major contribution of this thesis is to calculate methods of election forecasting based on polls, which provide good forecasts of the real election results.

In summary, the purposes of this thesis are the following:

- As there is no better available measure of the election result during the campaign, the latest poll value is taken as benchmark method.
  $\Rightarrow$ Is there a systematic improvement of the forecasting methods compared to the benchmark method?
- shortly before election, polls generate closer results than early in the campaign (see Linzer, 2013).
  $\Rightarrow$ Is there an improvement of forecasting methods with shorter lead time, analogous to polls?
- The aim of forecasting is to generate a value of the election result in advance of the election, which is close to the real result.
  $\Rightarrow$ Find the best forecasting method, which generates the best forecast of the election result.

In the following, the methods, outlined in Chapter 6 are calculated for 13 states, six parties and different lead times between 120 and 1 day prior the election. First, the computation of the forecasting methods with the software 'R' is explained. Then the forecasting results with respect to different dimensions (region, lead time, parties) are presented in form of tables and graphs. The quality of the forecasting methods is evaluated by common error measures compared to the real election result shown in Section 6.3. Further, the forecasting methods are compared to the latest poll value as benchmark forecast, which is additionally evaluated concerning its quality.

TECHNICAL COMPUTATION OF THE METHODS

The computations are employed with the system 'R' (see R Core Team, 2015).[30] Methods based on irregular time series structure, like nonparametric local constant and local linear methods are implemented with the package 'np' of Racine and Hayfield (2017), Version 0.60-3. To select the optimal bandwidth, the cross validation is used and a second order Gaussian kernel. Some methods are based on a regularized daily time series structure (compare Chapter 5 for generating this time series).

The parametric regression based methods `iss` and `arima` are implemented via the package 'forecast' of Hyndman and Khandakar (2008) Version 8.1. The `ets()`-function automatically chooses the best `iss`-method, depending on the optimization criterion "mse" (mean squared error) of the 15 types of ETS-methods stated in the Appendix in Table B.2. With the function `auto.arima()` of the 'forecast package', an appropriate ARIMA model is selected.[31]

For the dynamic linear models, the implementation in 'R' is made with the 'dlm'-package (Petris, 2009). The unknown parameter, the observational and evolution (error) variances, with assumed independent inverse Gamma prior distributions have to be calculated. Via the function `dlmGibbsDIG`, the first step of the MCMC Bayesian analysis with the Gibbs sampling algorithm and FFBS is implemented to obtain the the posterior distribution of $\theta(\sigma^2_\epsilon, \sigma^2_\xi, \theta_{0:n}|y_{1:n})$. For the local level model, the function `dlmModPoly(a)` with order $a = 1$ is included in the sampling algorithm, for $a = 2$, the local linear trend model is obtained. The function runs with 1000 MCMC iterations. To reduce autocorrelation in the samples, only every tenth is saved, compare Petris and Petrone (2011). The Bayesian character is

[30] The computations in 'R' are explained in detail in a separate 'Read me' document, where the input and results files can also be found.

[31] auto.arima() chooses the best ETS-model with the smallest AIC ($AIC = -2log(L) + 2 \cdot (p + q + P + Q + K)$) (see Hyndman and Khandakar, 2008). In step 2 consider variations from the current model (vary p,q) and exclude or include c from the current model. Then choose the current model with the lowest AIC. Repeat step 2 until no lower AIC can be found. In this work the ARIMA (0,1,2) is chosen as best model $y_t = y_{t-1} + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-1}$.

obtained by running the Gibbs sampler again, when a new poll value is published and the forecast is updated.

7.2 Performance of the forecasting results for 13 state elections

Forecasting elections is an established research field especially in the US. Shortly before the elections, voters, politicians and the media monitor with great interest, who is expected to win the election. The public wants to know in advance who will govern the country. Respectively, in the German multi-party elections, the forecasting objective is the vote share of the single parties, as there are often coalitions of two or three parties. Also small parties are in focus, like the new founded AFD, if these upcoming parties reach the 5% threshold or gain more votes as expected, like the Greens in the last state election in BW. Small parties are important because they minimize the sum of the vote shares or are possible partners in coalitions. In this thesis, 13 state elections between 2014 and 2017 are under review (the latest one in NW 2017-05-14), where the AFD gained more than 5% and thus a seat in the parliament. Due to the changing voting behavior, party programs, party landscape and other issues (for details see Section 3.3), the number of undecided voters shortly before election increased over the last years (see Plickert, 2013). Hence, the data base for the forecasting methods are polls on state level 365 days before the specific election date. A longer period is not needed because the voters decide very late and the number of state level polls, provided more than one year in advance is very low, compare Figure A.9 in the Appendix. Whereas in US elections, the polling institutes start directly after the previous election with new polls, in Germany this is only the case for national elections, in state elections polls are conducted first two years after the last election in most cases.

The ten forecasting methods are based on regularized and irregular time series of polls. A forecast has to be conducted with an adequate forecasting horizon. As stated in Table 6.1, most lead times for elections are one month, two months or three months. In this thesis, forecasts with lead times between 120 days and

1 day are calculated.[32] For ease of understanding and illustration, in the tables of this thesis only forecasts with lead times of 120, 90, 60, 30, 21, 14, 7 and 1 days prior election are presented. Whereas the graphs show mostly forecasts for every day between 120 and 1 day lead time or the average value over 120 days lead time. In the following, the forecasting results are outlined for the 13 state elections depending on their lead time. First, the latest poll value as benchmark is investigated in terms of its performance and quality in order to evaluate its eligibility as a benchmark. Forecast errors are calculated, to give an overview how close the forecasts are to the real election results. Further, the forecasts are compared to the benchmark method `Lpv`. To measure the performance of the methods and see which method performs best, tables and graphs are provided. For this purpose, the performance is calculated with the measures of Section 6.3 RMSE, MSE, MAE and MDAE for different levels of aggregation (party, region and election) and is illustrated in tables and graphs.

7.2.1 Performance of the latest poll value

A contribution of this work is to evaluate the poll quality and performance, which is taken as data base for the forecasting methods. In the following, the poll quality with the new measure `'A'` is under review to see how the polls perform in form of the latest poll value. Further, the performance of this benchmark method `Lpv` is evaluated with common evaluation criteria.

A measure, indicating the accuracy and quality of polls, is 'A_{ijk}' of Martin, Traugott and Kennedy (2005).[33]

[32]The work of Haupt, Schnurbus and Huber (2017) provides forecasts of state elections in 2016 with lead times of 84, 56, 28, 21, 14, 7 and 1 days analogous to Berg, Nelson and Rietz (2003) for `dlm`, `arima`, `Lpv`, `iss` and `avew`.

[33]The 'A_{ijk}' for the two-party case is denoted as follows,

$$A_{ijk} = \log\left(\frac{\frac{r_{ijk}}{d_{ijk}}}{\frac{R_{jk}}{D_{jk}}}\right), \tag{7.1}$$

where the small letters r, d indicate the polls for the Republicans and Democrats on date i in the j-th office (Senator) in the k-th state of the US. The capital letters R, D

A value of exactly zero for 'A_{ijk}' in Equation (7.1) means, that polls and election odds are exactly the same and the ratio is one. If 'A' is smaller than zero, it is biased in the direction of the Democrats (see Martin, Traugott and Kennedy, 2005). For the multi-party case, Wright, Farrar and Russel (2014) expand the measure 'A' by considering the third party in the US. This measure is also used for Italian polls as seen by Callegaro and Gasperoni (2008) and Castro (2013). For the German multi-party case and e.g. party $p = SPD$, 'A' is denoted as follows,

$$A_{eh} = \log \left(\frac{\frac{spd_{eh}}{afd_{eh}+cdu_{eh}+fdp_{eh}+gre_{eh}+lin_{eh}}}{\frac{SPD_{eh}}{AFD_{eh}+CDU_{eh}+FDP_{eh}+GRE_{eh}+LIN_{eh}}} \right) \tag{7.2}$$

Where A_{eh} is a measure for the poll quality in election e and lead time h. In Equation (7.2), the poll values in the numerator are calculated as the interested party SPD divided by the remaining parties. In the denominator, the real election values are calculated as the election result of the SPD divided by the election results of the other parties. If the measure 'A' is near zero, the poll value of the investigated party is equivalent to the election result of the party. Further, a positive value indicates an overestimation of the poll result compared to the real election result. A negative value of 'A' means an underestimation of the poll value compared to the real election result.

Figure A.12 in the Appendix shows the poll quality with measure 'A', averaged over the 13 investigated states and for the six parties. The measure 'A' is illustrated over the lead times 120 up to 1 day before election and for the six parties. For the AFD, the polls are underestimated compared to the real election results (negative values). Other parties like the LIN indicate an overestimation to the real election results. The poll quality varies over the lead time, with higher deviations of the polls from the real election results for lead times of three to one month.

Table 7.1, illustrates the descriptive statistics of the measure 'A' for the six parties and 13 states. The values range from -1.89 to a maximum of 0.74

display the current values.

Table 7.1: Descriptive statistics of the extended measure `'A'` to investigate poll quality for the six parties over investigated states.

Party	Minimum	1.Quartil	Median	Mean	3.Quartil	Maximum
AFD	-1.89	-0.74	-0.39	-0.37	0.00	0.54
CDU	-0.34	0.01	0.10	0.12	0.24	0.67
FDP	-1.37	-0.38	-0.26	-0.23	-0.02	0.70
GRE	-0.34	-0.05	0.07	0.12	0.24	0.77
LIN	-0.51	-0.04	0.06	0.09	0.25	0.67
SPD	-0.48	-0.11	0.05	0.07	0.26	0.74

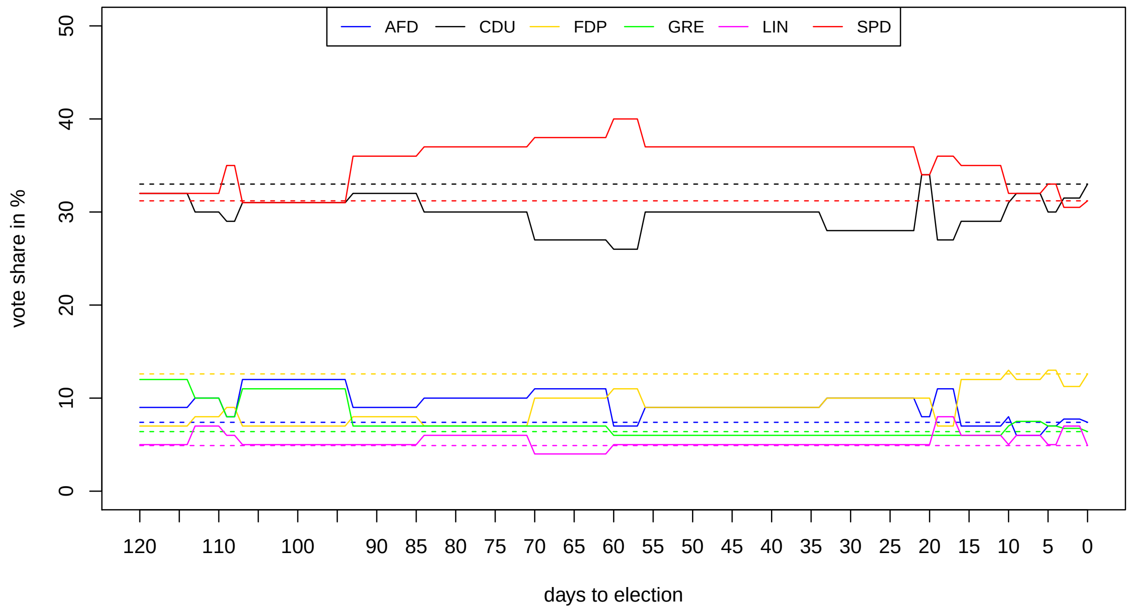
According to the extended measure `'A'`, the polls are mostly underestimated for smaller parties like the FDP or AFD. But putting all together, the measure performs well with median values for all parties between -0.39 and 0.12 near the value zero, which means, that polls and election results are corresponding.

As stated in Section 6.1, the `Lpv` is a snapshot of the current opinion, but often seen as direct forecast in media. In this work, the latest poll value is calculated by repeating the first available poll as long as a new poll is published, to get a value on every day before the election (if there is more than one poll on a day, the mean of these polls is calculated (see Erikson and Wlezien, 2008)). This `Lpv` is taken as benchmark in two forms. On the one hand, the `Lpv` is the poll value, extended on a daily basis and seen as snapshot of the current opinion. On the other hand it can be seen as a simple forecast and benchmark method according to Hyndman and Athanasopoulos (2013).

The other methods are compared to this `Lpv` by measuring the performance of the methods. In the following, the latest poll value is controlled in terms of its performance as a snapshot and respectively as a direct forecast in its benchmark function. As Linzer (2013) reports, the polls improve in accuracy if they are closer to election day. The state election in Northrhine Westphalia is illustrated as an example, because it is the latest election and Northrhine Westphalia is the state with the largest population, used as direction election.

In Figure 7.1, the "forecasted" vote shares for the six parties of the method `Lpv` are calculated for the Northrhine Westphalia state election on "2017-05-14". The last 120 days before election, the vote shares are displayed in solid lines, whereas the real election results are shown as horizontal dotted lines, to see the deviation of the values from the real results. The SPD vote shares were estimated nearly perfect four months before election, then overestimated with almost nine percent. Shortly before election, the Lpv generates very close results.

Figure 7.1: Forecasted vote shares with the method Lpv for the NW state election "2017-05-14" 120 days to 1 day before election. The dotted lines show the real election results on election date.



In Table 7.2, the vote shares of the method Lpv for 120, 90, 60, 30, 21, 14, 7 and 1 day lead time are given for the six parties. The rightmost column indicates the real election results at lead time $h = 0$ (dotted horizontal lines in Figure 7.1).

In Figure A.13 in the Appendix, the Saxony Anhalt state election in "2016-03-13" is shown as an example, where the Lpv is clear off the target, the real election result. The percentage values of the Lpv are drawn with solid lines, the real election results with dotted horizontal lines, to see how good the Lpv performed. The last 20 days before election, the vote shares are near the real election result, but e.g. 100 days before election the AFD is underestimated by nearly 20%. For new emerging parties, polls are more error prone, due to lower experience of the

Table 7.2: Forecasted vote shares with the method Lpv for the NW state election "2017-05-14" for selected lead times before election. The last column displays the real election result.

Party	120	90	60	30	21	14	7	1	0
AFD	9.00	9.00	7.00	10.00	8.00	7.00	6.00	7.75	7.40
CDU	32.00	32.00	26.00	28.00	34.00	29.00	32.00	31.50	33.00
FDP	7.00	8.00	11.00	10.00	10.00	12.00	12.00	11.25	12.60
GRE	12.00	7.00	6.00	6.00	6.00	6.00	7.50	6.75	6.40
LIN	5.00	5.00	5.00	5.00	5.00	6.00	6.00	7.00	4.90
SPD	32.00	36.00	40.00	37.00	34.00	35.00	32.00	30.50	31.20

institutes with new parties.

In Table B.4 in the Appendix, the corresponding vote shares for Saxony Anhalt of the Lpv method for the selected lead times are displayed. The rightmost column with lead time 0 contains the real election results for the state election. Both states ST and NW indicate, that the Lpv generates better results with smaller lead times. In the case of NW, the Lpv values displayed good results between 120 and 90 days, seen in Figure 7.1.

The overall forecasting performance of the benchmark method Lpv for the 13 states is listed in Table B.5 in the Appendix. The table provides forecast errors, calculated as the deviation of the forecasted vote share from the real election result, compare Equation (6.22). In the rightmost column, the real election results for the respective states are given in percent. The table is ordered by date of election, starting with the Saxony state election in 2014, up to the latest state election in Northrhine Westphalia in 2017. Forecast errors are calculated for all lead times between 120 and 1 day, but are only given for the chosen lead times for illustrative purposes. The AFD is mostly underestimated (negative sign), especially 120 days before election in the state elections 2014, 2015 and 2016. Whereas since the elections in MV and WBE in September 2016, a "learning effect" appeared and the AFD was not longer underestimated. The institutes calculated the polls with knowledge of the previous elections and put more weight on the AFD votes.

Putting all together, there is an improvement of the forecast errors with decreasing lead time in most cases. Sometimes ups and downs are observed, the forecast errors decrease until one month before election, than rise again until 14 days before election and then decrease again. In e.g. Saxony Anhalt, the Left party (LIN), see Table B.5 in the Appendix starts with 9.7% forecast error at lead time 120 days, decreasing to 2.7% at 30 days and then rising up to 3.7% 21 days before election. The reasons for these ups and downs can be seen in house effects, as the polls are conducted by several institutes, where some institutes underestimate or overestimate the vote shares. The number of swing voters who change their minds shortly before election also cause these ups and downs. In the following, the forecasting methods are evaluated in terms of which method performs best by calculating the forecast errors and evaluation criteria with respect to region, party and election. Further, the question is answered, if the methods beat the latest poll value in accuracy.

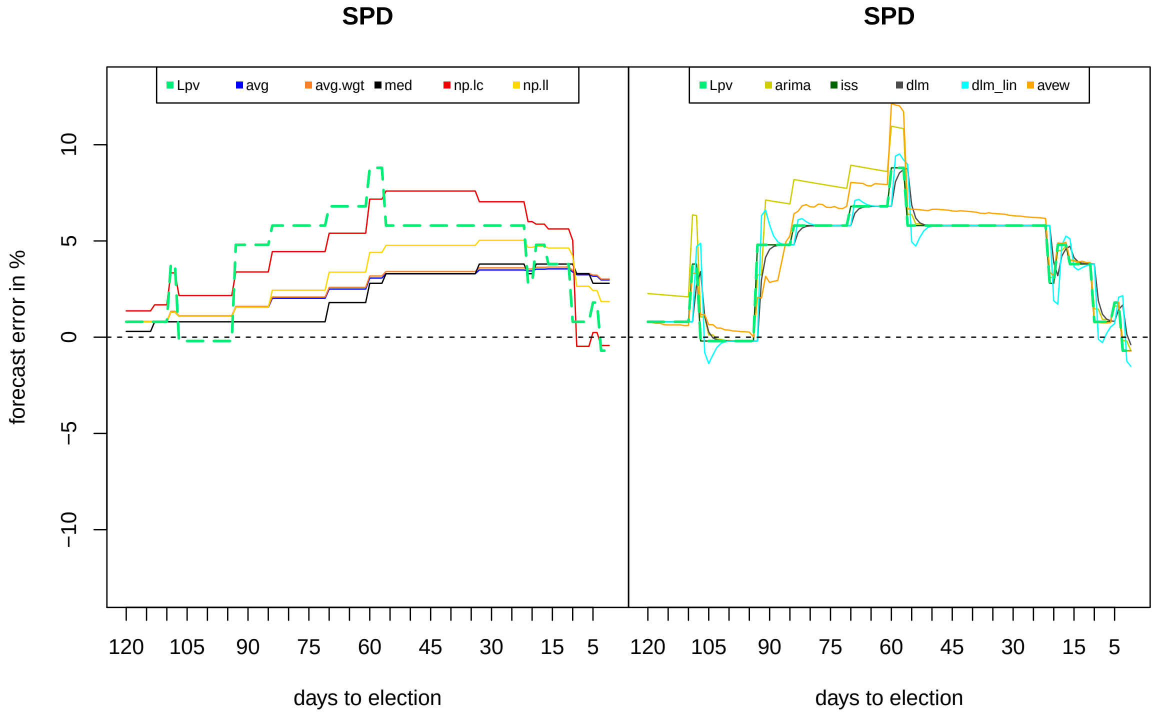
7.2.2 Performance of the forecasting methods for different levels of aggregation

Forecasting results of the SPD for Northrhine Westphalia and different methods

In Figure 7.2, the forecast errors of the ten methods and the forecast errors for the benchmark `Lpv` are shown by the example of the latest election in NW on "2017-05-14" for the SPD. On the left side of the graph, the forecasting methods based on an irregular times series are displayed (`avg`, `avg.wgt`, `med`, `np.lc` and `np.ll`). The right side outlines the forecasting methods based on a regularized time series structure (`arima`, `iss`, `dlm`, `dlm_lin` and `avew`). The green dotted line on each side shows the forecast errors for the benchmark method `Lpv` from 120 up to 1 day before election.

Table 7.3 lists the corresponding forecast errors in percent with lead times 120, 90, 60, 30, 21, 14, 7 and 1 day before election for all ten methods and the `Lpv`. As seen in the Table, there is an improvement of the forecast depending on lead time (decreasing forecast errors). Further, the latest poll value is outperformed by at least one method, especially in the time of 90 days to 30 days prior the election

Figure 7.2: Forecast errors of the methods for the party SPD in the NW state election 120 days to 1 day before election. Green dotted line in each part of the figure shows the forecast error of `Lpv`.



(the green dotted line in Figure 7.2 is above the other methods with higher forecast errors in the left part of the figure).

To measure the specific performance of the forecasting methods, evaluation criteria are applied.[34] The forecasting performance of the methods is calculated on different levels of aggregation with respect to party, region, election and year (see Haupt, Schnurbus and Huber, 2017). Moreover, the performance is measured with different evaluation criteria (RMSE, MAE, MDAE and MSE).

[34]The performance of the methods `arima`, `avew`, `dlm`, `dlm_lin` and `iss` is outlined in detail in the work of Haupt, Schnurbus and Huber (2017) for the five state elections 2016 with the criteria RMSE.

Table 7.3: Forecast errors of the different methods for the SPD in the Northrhine Westphalia state election for selected lead times seen in the columns.

Method	120	90	60	30	21	14	7	1
Lpv	-0.70	5.80	6.80	4.80	-0.20	-0.20	0.80	0.80
avg	0.80	1.55	3.07	3.49	3.44	3.55	3.24	2.97
avg.wgt	0.82	1.60	3.18	3.61	3.55	3.64	3.31	3.02
med	0.30	0.80	2.80	3.80	3.30	3.80	3.30	2.80
np.lc	0.80	1.55	4.41	5.03	4.67	4.63	2.64	1.85
np.ll	1.37	3.39	7.17	7.04	6.00	5.63	-0.48	-0.43
arima	-0.69	5.80	8.61	7.12	-0.17	1.05	2.16	2.27
iss	-0.70	5.80	6.80	4.80	-0.20	-0.20	0.80	0.80
dlm	-0.39	5.80	6.80	4.15	-0.20	3.41	0.80	0.80
dlm_lin	-1.51	5.80	6.80	6.60	-0.21	4.88	0.80	0.80
avew	-0.72	6.34	7.93	3.16	0.37	1.17	0.64	0.77

PERFORMANCE WITH RESPECT TO REGION

Table 7.4 shows the RMSEs per party for the different methods, computed by averaging over the 13 states with the selected lead times. The rightmost column in the table shows the average over all lead times (120, 119,...,1 days) and all regions for the respective party and method. At first glance, it can be seen, that the forecasting accuracy, measured by the RMSE improves with decreasing lead times for all methods. The latest poll value is outperformed by at least one forecasting method on a specific lead time. During the 120 days before election, the Lpv is outperformed by up to 0.73 percentage points in case of the AFD for 60 days lead time. On average, the Lpv displays good results, but is outperformed by e.g. avew for the AFD by 0.32 percentage points or by the method arima for the CDU by 0.14 percentage points. The average RMSEs range from 2.15% to nearly 7% for some methods. FDP, GRE and LIN display the lowest (best) average RMSEs over all methods between 2% and 4%.

Table 7.4: RMSE computed by averaging over all investigated regions for all parties, selected lead times and methods. The last column displays the average RMSE over all lead times.

Party	Method	120	90	60	30	21	14	7	1	Avg.
	med	7.71	7.32	7.13	5.86	5.15	4.75	4.16	3.92	6.54
	avg	7.70	6.95	6.74	6.09	5.70	5.22	4.72	4.36	6.52
	avg.wgt	8.72	7.45	7.22	6.20	5.61	5.13	4.63	4.27	6.85
	np.lc	8.47	7.00	6.84	5.75	5.10	4.01	3.57	3.32	6.38
	np.ll	8.84	5.42	4.12	3.54	3.31	3.35	3.17	2.84	5.11
AFD	arima	7.25	6.18	4.58	4.08	3.61	3.86	2.90	2.63	4.97
	dlm	6.78	4.90	4.90	4.31	3.82	3.72	2.96	2.38	5.03
	dlm_lin	6.80	4.89	5.10	4.33	3.78	3.71	2.97	2.66	5.04
	iss	6.78	5.10	4.85	4.29	3.75	3.94	2.98	2.64	5.02
	Lpv	6.78	4.90	4.85	4.27	3.73	3.95	2.98	2.64	4.99
	avew	6.72	4.18	4.33	4.02	3.69	3.84	2.95	2.33	4.67
	med	6.07	6.07	5.76	5.37	4.92	4.58	4.11	3.88	5.62
	avg	6.15	6.18	5.91	5.50	5.12	4.73	4.27	3.98	5.70
	avg.wgt	6.09	6.22	5.91	5.45	5.06	4.66	4.19	3.90	5.66
	np.lc	5.89	6.11	5.87	4.83	4.42	3.80	3.29	2.86	5.30
	np.ll	9.70	9.36	8.32	3.75	3.31	3.99	2.34	2.32	7.05
CDU	arima	5.69	5.58	4.58	4.12	3.08	3.35	2.08	2.56	4.66
	dlm	5.68	5.69	5.28	4.38	3.58	3.59	2.18	2.45	4.82
	dlm_lin	5.64	5.69	5.27	4.48	3.47	3.63	2.06	2.97	4.85
	iss	5.68	5.69	5.20	4.28	3.27	3.40	2.15	2.57	4.80
	Lpv	5.68	5.69	5.20	4.28	3.27	3.40	2.15	2.57	4.80
	avew	5.66	5.78	5.31	4.05	3.34	3.28	2.11	2.53	4.72
	med	2.41	2.43	2.52	2.65	2.50	2.35	2.04	1.94	2.46
	avg	2.49	2.54	2.49	2.36	2.25	2.10	1.88	1.77	2.40
	avg.wgt	2.32	2.54	2.48	2.35	2.28	2.08	1.85	1.75	2.37
	np.lc	2.25	2.51	2.48	2.27	2.15	1.93	1.65	1.29	2.29
	np.ll	4.75	4.27	4.06	2.28	2.10	2.00	1.27	1.28	3.47
FDP	arima	2.60	2.55	2.22	1.89	1.66	1.34	1.18	1.23	2.23
	dlm	2.59	2.55	2.36	1.81	1.67	1.46	1.18	1.16	2.26
	dlm_lin	2.59	2.52	2.37	1.64	1.69	1.51	1.21	1.28	2.28
	iss	2.59	2.54	2.25	1.92	1.68	1.32	1.19	1.23	2.25
	Lpv	2.59	2.54	2.25	1.94	1.68	1.32	1.19	1.23	2.25
	avew	2.65	2.60	2.29	1.93	1.67	1.35	1.16	1.14	2.37
	med	2.75	2.64	2.53	2.11	1.97	1.67	1.56	1.55	2.34

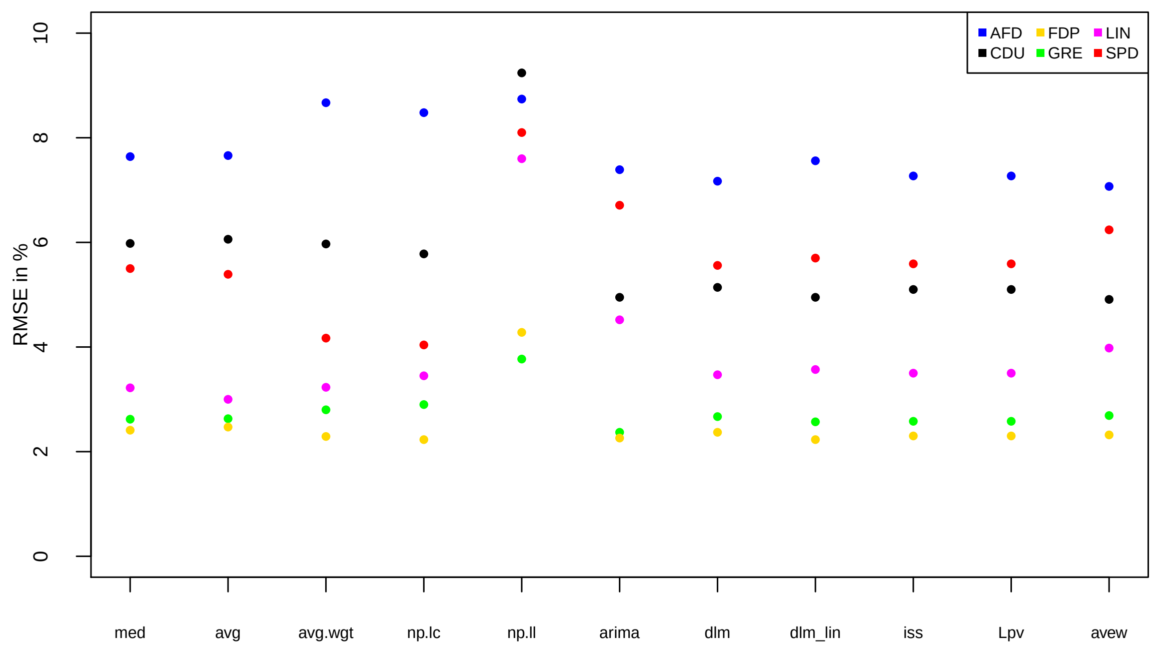
Table 7.4 – continued from previous page

Party	Method	120	90	60	30	21	14	7	1	Avg.
	avg	2.71	2.53	2.37	2.13	2.07	1.88	1.65	1.54	2.31
	avg.wgt	2.89	2.61	2.44	2.18	2.11	1.86	1.62	1.51	2.37
	np.lc	2.93	2.71	2.49	1.98	1.93	1.73	1.30	1.21	2.35
	np.ll	4.54	3.43	3.17	1.80	1.71	4.90	0.93	1.04	3.08
GRE	arima	2.62	2.29	4.80	1.67	1.29	2.05	0.92	1.00	2.16
	dlm	2.94	2.47	1.99	1.54	1.39	1.70	0.82	0.83	2.16
	dlm_lin	2.94	2.44	1.99	1.32	1.31	1.83	1.09	1.11	2.21
	iss	2.94	2.44	1.93	1.79	1.38	2.03	0.85	1.00	2.15
	Lpv	2.94	2.46	1.93	1.85	1.38	2.03	0.85	1.00	2.17
	avew	3.12	2.61	1.98	1.87	1.37	1.99	0.92	0.81	2.88
	med	3.24	3.02	2.90	2.57	2.57	2.24	1.90	1.75	2.81
	avg	3.00	2.94	2.92	2.66	2.64	2.41	2.14	1.96	2.77
	avg.wgt	3.24	3.03	3.01	2.72	2.63	2.39	2.11	1.94	2.85
	np.lc	3.46	3.01	3.04	2.73	2.64	2.41	1.84	1.65	2.88
	np.ll	7.69	3.58	3.43	2.09	2.41	3.28	1.55	1.51	4.02
LIN	arima	4.66	3.15	2.93	2.13	2.57	1.67	1.46	1.78	2.99
	dlm	3.53	2.63	2.89	1.94	2.36	1.90	1.49	1.54	2.55
	dlm_lin	3.53	2.67	2.86	1.99	2.71	1.78	1.58	2.15	2.62
	iss	3.53	2.66	2.73	2.07	2.52	1.73	1.46	1.79	2.58
	Lpv	3.53	2.66	2.73	2.07	2.52	1.73	1.46	1.79	2.58
	avew	4.17	2.79	2.74	2.21	2.37	1.71	1.42	1.61	2.75
	med	7.71	7.32	7.13	5.86	5.15	4.75	4.16	3.92	6.54
	avg	7.70	6.95	6.74	6.09	5.70	5.22	4.72	4.36	6.52
	avg.wgt	8.72	7.45	7.22	6.20	5.61	5.13	4.63	4.27	6.85
	np.lc	8.47	7.00	6.84	5.75	5.10	4.01	3.57	3.32	6.38
	np.ll	8.84	5.42	4.12	3.54	3.31	3.35	3.17	2.84	5.11
SPD	arima	7.28	6.18	4.58	4.08	3.61	3.86	2.90	2.63	5.00
	dlm	6.78	4.90	4.90	4.31	3.82	3.72	2.96	2.38	5.03
	dlm_lin	6.80	4.89	5.10	4.33	3.78	3.71	2.97	2.66	5.04
	iss	6.78	5.10	4.85	4.29	3.75	3.94	2.98	2.64	5.02
	Lpv	6.78	5.13	4.85	4.27	3.73	3.95	2.98	2.64	5.03
	avew	6.72	4.18	4.33	4.02	3.69	3.84	2.95	2.33	4.67

Figure 7.3 depicts the values of the rightmost column of Table 7.4, the average RMSEs over all regions and lead times (120 to 1 day) for the respective forecast-

ing method and party.[35] As the figure shows, the `arima` and `avew` method both outperform the `Lpv` for the party CDU. Also for the SPD, `avew` is clearly better on average. Considering the performance of the methods on specific lead times and not on average, the methods provide improvements in comparison to the `Lpv` of 0 to 0.73 percentage points, especially within three to one months before election. Analogous to the polls, there is an improvement in accuracy with decreasing lead time by the other forecasting methods.

Figure 7.3: RMSE (in percent) computed by averaging over all regions for all parties and lead times per method. The points indicate the values for the parties of the last column in Table 7.4.



When the results are aggregated over only few states, Figure 7.4 provides the

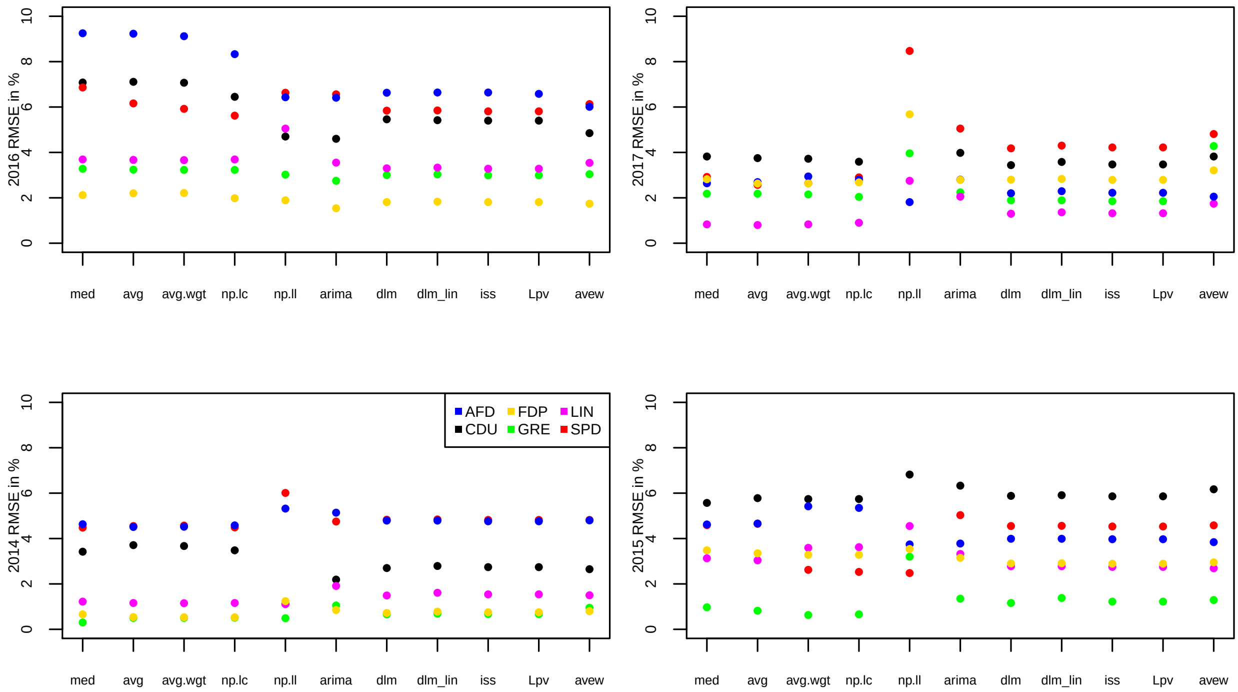
[35]The order of the methods in the figure is analogous to the order of the methods in the table, `Lpv` is the penultimate one in the figure and is written with large initial letter as benchmark method.

results. The RMSEs are computed by averaging over the states SN, TH and BB for 2014, over HB and HH in 2015, over BW, RP, ST, MV and WBE in 2016 and over SL, SH and NW in 2017. The aggregation is made for the different methods, lead times (120 to 1) and parties, but in contrast to Figure 7.3 not over all 13 states. The four graphs show the average RMSEs over lead times between 120 and 1 day of the forecasting methods and the benchmark `Lpv` for the single parties. Figure 7.4 indicates differences in the average RMSEs with respect to year and method. The average RMSEs increased over the years 2014 (bottomleft graph) and 2015 (bottomright graph), to the highest values ranging from 2% to 10% in 2016 (topleft graph). In the years 2014 to 2016 the AFD grew in importance. In 2016, the AFD was underestimated in most polls during the campaign and therefore higher forecast errors with corresponding higher RMSEs are obtained. In 2014, the forecasting methods depending on regularized poll dates display the best results for the FDP, GRE and LIN in contrast to the `Lpv`, for the CDU, `arima` is the best method. Also in 2015, the basic and nonparametric methods performed better for GRE, SPD and CDU than `Lpv`. The highest average RMSEs are provided in 2016, where the AFD was underestimated by the polling institutes. For more details, concerning the performance of the methods in 2016 see the work of Haupt, Schnurbus and Huber (2017). In 2017, the averaged RMSEs improve and indicate smaller values for the RMSEs than in the years before especially for the AFD due to adjustments of the institutes for the values of the AFD.

Figure A.14 in the Appendix illustrates the average MDAEs for the years 2014 to 2017 analogous to Figure 7.4. The averaged MDAEs are lower than the RMSEs, because the MDAE is less sensitive to outliers than the RMSE (see Hyndman and Koehler, 2006). In contrast to the RMSEs with 0 to 10%, the average MDAEs range between 0 and 8% for the specific years and over all lead times and parties. Like the RMSEs, the MDAEs reveal better values over the years 2014 to 2017, calculating the averages over all lead times.

To sum up, averaging over regions provides RMSEs per party and different methods, which improve with decreasing lead time. The `Lpv` as benchmark is outperformed by up to 0.73 percentage points in case of the e.g. AfD with 60

Figure 7.4: RMSE (in percent) computed by averaging over regions depending on year. The RMSEs calculated in 2014 contain the elections in SN, BB and TH, in 2015 HB and HH, in 2016 BW, RP, ST, MV and WBE and in 2017 SL, SH and NW. Points display the average values for every party over all lead times per method.



days lead time. Most methods show average RMSEs of 2-7%, with only 2-4% for the GRE, LIN and FDP. By summing up certain regions by year, the highest RMSEs are displayed in 2016 and the lowest in 2017. Mostly the methods `arima` and `avew` performed best by averaging over regions.

PERFORMANCE WITH RESPECT TO PARTY

The next level of aggregation is calculated with respect to parties. Figure 7.5 shows the RMSEs computed by averaging over six parties AFD, CDU, FDP, GRE, LIN and SPD for the methods, all lead times between 120 and 1 days and for the regions SL, SH and NW. The elections took place in 2017, beginning with

SL in March 2017 and ending with NW in "2017-05-14". On the left side of the figure are the RMSEs of the forecasting methods, based on a regularized time series (`avg`, `avg.wgt`, `med`, `np.ll` and `np.lc`), respectively for each state. On the right side of the figure, forecasting methods based on the irregular poll data (`iss`, `arima`, `dlm`, `dlm_lin` and `avew`) are shown. To compare the performance of the methods not only with each other, but also with the benchmark `Lpv`, the dotted blue line in each graph of Figure 7.5 displays the performance of the `Lpv`. The figure indicates, that the performance of the `Lpv` sometimes outperforms the other methods for specific lead times (when the blue dotted line is below the other ones). In elections, where only few polls are published during the campaign, like in Schleswig Holstein, the forecasting methods outperform the `Lpv` in particular on days, where no poll is published. In the graph, these dates can be seen, where no change is in the lines e.g. between 60 and 30 days before election in the case of Schleswig Holstein. Whereas the last graph in the figure shows the performance of the methods and the `Lpv` for NW state elections (where the most polls are conducted), leading to more variation in the performance of the methods and the benchmark method.

Figure 7.5: RMSEs computed by averaging over all parties in the states Saarland, Schleswig Holstein and Northrhine Westphalia 2017 for every method and all lead times. The benchmark `Lpv` is shown as dotted blue line.

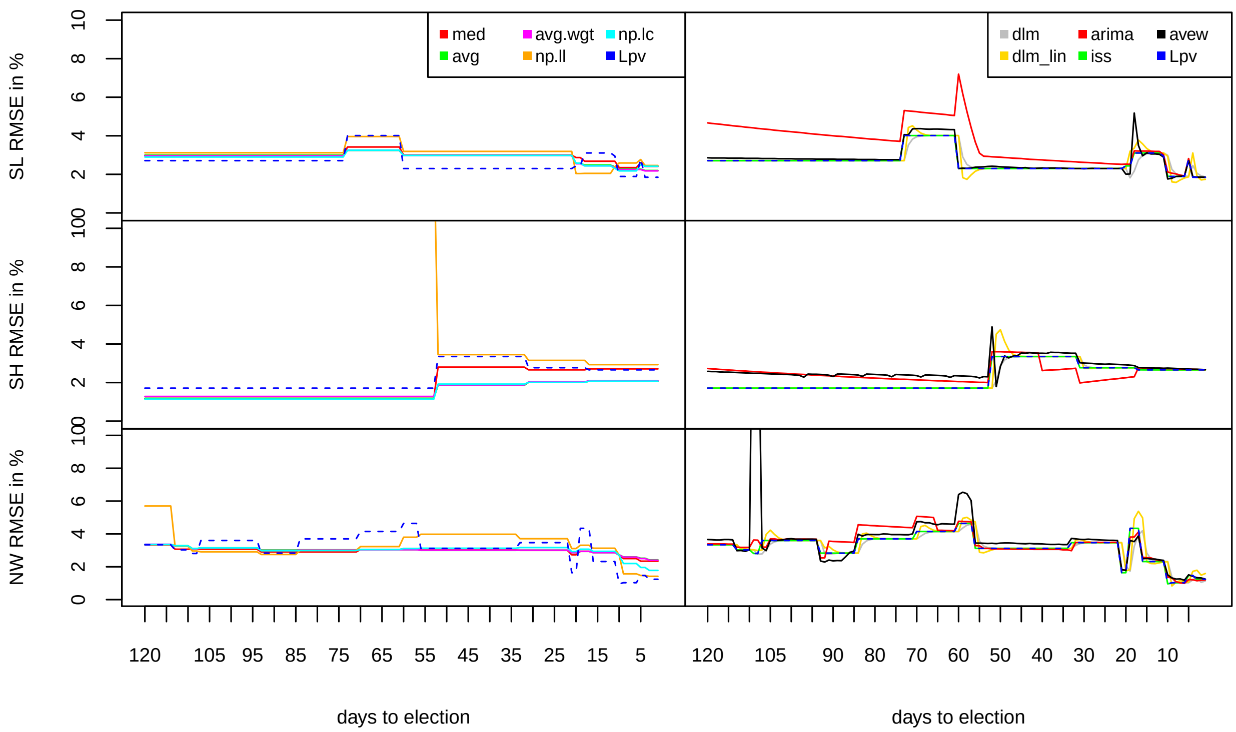


Table 7.5 presents the corresponding values to the Figure 7.5. The forecasting accuracy of all methods is measured with the RMSE over all parties and for all lead times between 120 and 1 day for the specific regions SL, SH and NW. The table illustrates the RMSEs for selected lead times and for the average RMSEs in the rightmost column. In the case of Schleswig Holstein, the `Lpv` is outperformed by the basic methods `avg` and `avg.wgt` and the nonparametric method `np.lc` with up to 0.76 percentage points at lead time 60 days. The relative improvement of the `avg` method in comparison to the benchmark amounts to over 50% for the average RMSE. In Northrhine Westphalia, the relative improvement of the method `med` amounts to 11%, compared to the benchmark. Especially for the lead times between 60 days and 30 days, the relative improvements of the basic methods

compared to the Lpv is up to 50%.

In Table B.6 in the Appendix, the aggregate results of the RMSE for the remaining states (except SL, SH and NW), ordered by their election dates are displayed.[36] RMSEs are computed by averaging over all parties, for all lead times and all regions (only results for selected lead times and the average values over all lead times are indicated). However, RMSEs of the states (besides the exceptions of the 2016 elections) range between 2 and 4% on average and show lower values shortly before election. Compared to other accuracy measures for German elections, e.g. for German elections by, Lewis-Beck and Dassonneville (2015), who compute an average RMSE over 5% four months before election, the RMSEs in this work perform remarkably well. 120 days before election date, the RMSEs range between two and four percent for all state elections, except the 2016 elections.

Analogous to the results for SL, SH and NW at least one method exists, which outperforms the benchmark for a specific lead time in accuracy. The average RMSEs over all lead times in the state Saxony Anhalt amount over 6%, because the AFD was underestimated by the polling institutes which led to high forecast errors and thus high RMSEs. The same is observed for BW, RP and MV, which provide average RMSEs of over 4%. The relative improvements of the average RMSEs of the methods compared to the benchmark in the single states range between 0% in BB, HB and HH and over 30% in WBE. During the campaign e.g. in Saxony, a relative improvement of the method `np.ll` compared to Lpv of over 70% is observed.

In conclusion, averaging over six parties for the methods, lead timea dn region, provide the following results. In states with only few polls like SH, the methods outperform the Lpv on days where no polls are published and leads to an information gain of up to 0.76 percentage points for e.g. the method `np.lc`. In comparison to other forecasting literature for German national elections like Lewis-Beck and Dassonneville (2015), the RMSEs range between 2-4% compared to over 5% in the work of Lewis-Beck. There is no general best method by aggregating the RMSE

[36] In the work of Haupt, Schnurbus and Huber (2017), the performance of the methods in the elections in BW, RP, ST, MV and WBE is outlined in detail.

with respect to the parties. At most, the `np.ll` and `avew` method are the best methods compared by their average RMSEs (`np.ll` is best in SN, BB, HH and WBE, `avew` in RP and ST).

PERFORMANCE WITH RESPECT TO PARTY AND REGION

The last level of aggregation to measure the performance of the forecasts is calculated with respect to party and region. Hence, not only the RMSE is taken as accuracy measure, but also the MDAE, MAE and MSE. In Table 7.6, the performance measures RMSE, MDAE, MAE and MSE are computed by averaging over six parties and 13 states, which means averaging over 78 values for the different methods.[37] The rightmost column displays the averages computed over 120 to 1 day lead time (for illustrative purposes, only selected lead times are shown in the table). The average over 120 days lead times shows, that there are relative improvements of 0% (MSE) to 3% (MAE) of the methods compared to the `Lpv`. Hence, there is no method which performs best, averaged over all parties, regions and lead times. During the campaign, there are improvements of 0 up to 0.46 percentage points by observing the MDAE as performance measure, which means relative gains of up to 25% by using the methods `np.ll` and `arima` instead of `Lpv`. Also MAE displays relative gains of up to 8% and absolute gains of 0.22 percentage points for the method `np.ll`. In some cases, MSE and RMSE show slight improvements of the methods compared to the `Lpv` of up to 12 percent.

[37] Others like Lewis-Beck and Dassonneville (2015) average over only eight or nine observations.

Table 7.5: RMSEs computed by averaging over all parties in the states Saarland, Schleswig Holstein and Northrhine Westphalia 2017 for every method and selected lead times. The rightmost column shows the average RMSE per method and state over all lead times from 120 days up to 1 day.

State	Method	120	90	60	30	21	14	7	1	Avg.
	med	2.99	2.99	3.00	3.00	3.00	2.68	2.35	2.41	2.98
	avg	2.99	2.99	2.98	2.98	2.98	2.48	2.28	2.20	2.92
	avg.wgt	2.96	2.96	2.98	2.98	2.98	2.45	2.25	2.18	2.91
	np.lc	2.90	2.90	2.98	2.98	2.98	2.44	2.18	2.42	2.89
	np.ll	3.12	3.12	3.19	3.19	3.19	2.05	2.59	2.46	3.13
SL	arima	4.67	4.00	7.20	2.62	2.52	3.20	1.96	1.85	3.78
	dlm	2.71	2.71	4.01	2.30	2.30	3.10	1.94	1.87	2.72
	dlm_lin	2.71	2.71	4.01	2.30	2.30	3.20	1.71	1.74	2.77
	iss	2.71	2.71	2.30	2.30	2.30	3.11	1.89	1.85	2.73
	Lpv	2.71	2.71	2.30	2.30	2.30	3.11	1.89	1.85	2.73
	avew	2.86	2.79	2.30	2.29	2.31	3.06	1.90	1.85	2.84
	med	1.18	1.18	1.18	2.66	2.66	2.71	2.71	2.71	2.01
	avg	1.18	1.18	1.18	2.01	2.01	2.07	2.07	2.07	1.57
	avg.wgt	1.28	1.28	1.28	2.03	2.03	2.10	2.10	2.10	1.66
	np.lc	1.14	1.14	1.14	2.01	2.01	2.07	2.07	2.07	1.60
	np.ll	14.84	14.84	14.84	3.15	3.15	2.93	2.93	2.93	10.94
SH	arima	2.73	2.34	2.05	2.01	2.23	2.75	2.70	2.67	2.55
	dlm	1.71	1.71	1.71	2.91	2.77	2.66	2.66	2.66	2.32
	dlm_lin	1.71	1.71	1.71	2.73	2.77	2.68	2.66	2.66	2.39
	iss	1.71	1.71	1.71	2.77	2.77	2.66	2.66	2.66	2.35
	Lpv	1.71	1.71	1.71	2.77	2.77	2.66	2.66	2.66	2.35
	avew	2.58	2.31	2.35	3.01	2.93	2.76	2.72	2.67	2.74
	med	3.34	2.90	3.10	3.00	2.73	2.90	2.50	2.34	2.97
	avg	3.36	3.02	3.04	3.01	2.86	2.86	2.60	2.42	3.02
	avg.wgt	3.37	3.00	3.04	3.01	2.86	2.86	2.59	2.40	3.01
	np.lc	3.36	2.98	3.11	3.17	2.93	2.94	2.20	1.78	3.03
	np.ll	5.70	2.75	3.80	3.71	3.11	3.13	1.57	1.42	3.45
NW	arima	3.39	3.54	4.77	3.48	1.80	2.54	1.02	1.20	3.51
	dlm	3.34	2.80	4.15	3.45	3.47	2.47	1.01	1.14	3.29
	dlm_lin	3.34	3.03	4.15	3.58	3.47	2.20	1.16	1.59	3.40
	iss	3.34	2.83	4.64	3.47	1.64	2.32	1.03	1.24	3.31
	Lpv	3.34	2.83	4.64	3.47	1.64	2.32	1.03	1.24	3.31
	avew	3.66	2.36	6.40	3.66	1.83	2.47	1.26	1.25	4.61

Table 7.6: Performance measures RMSE, MAE, MDAE and MSE computed by averaging over all parties and regions, displayed for selected lead times. The rightmost column shows the average performance over all lead times.

Measure	Method	120	90	60	30	21	14	7	1	Avg.
	med	5.06	4.88	4.77	4.32	3.97	3.65	3.22	3.03	4.52
	avg	5.01	4.71	4.60	4.28	4.04	3.73	3.37	3.14	4.45
	avg.wgt	5.05	4.75	4.63	4.25	4.05	3.69	3.32	3.09	4.44
	np.lc	4.94	4.60	4.51	3.97	3.75	3.18	2.74	2.51	4.21
	np.ll	7.58	6.03	5.41	3.40	3.21	3.73	2.32	2.17	5.11
RMSE	arima	5.31	4.57	4.39	3.49	3.00	2.94	2.19	2.14	4.02
	dlm	4.77	4.03	4.04	3.53	3.20	2.92	2.17	2.04	3.85
	dlm_lin	4.78	4.09	4.09	3.55	3.15	2.91	2.23	2.34	3.91
	iss	4.78	4.12	4.00	3.57	3.07	2.97	2.20	2.15	3.87
	Lpv	4.78	4.08	4.00	3.57	3.07	2.98	2.20	2.15	3.87
	avew	5.05	4.03	4.06	3.51	3.11	2.86	2.18	2.06	3.95
	med	3.10	3.00	2.80	2.70	2.75	2.45	1.95	1.90	2.70
	avg	2.87	2.97	2.78	2.91	2.62	2.41	2.23	1.92	2.77
	avg.wgt	3.34	2.98	3.02	2.92	2.71	2.37	2.10	1.89	2.85
	np.lc	2.80	2.80	2.70	2.55	2.44	1.87	1.57	1.51	2.41
	np.ll	2.97	2.87	2.81	1.84	1.53	1.29	0.97	1.09	2.39
MDAE	arima	2.29	2.60	2.40	2.18	1.61	1.55	1.10	1.16	2.16
	dlm	2.55	2.50	2.45	2.10	1.90	1.55	1.15	0.97	2.20
	dlm_lin	2.35	2.50	2.55	2.05	1.85	1.60	1.20	1.43	2.20
	iss	2.50	2.55	2.35	2.30	1.70	1.55	1.15	1.20	2.20
	Lpv	2.50	2.50	2.35	2.30	1.70	1.55	1.15	1.20	2.20
	avew	2.37	2.61	2.43	2.33	1.72	1.54	1.11	1.04	2.21
	med	3.75	3.64	3.59	3.38	3.17	2.86	2.49	2.36	3.42
	avg	3.76	3.60	3.53	3.37	3.19	2.92	2.62	2.46	3.41
	avg.wgt	3.80	3.64	3.56	3.33	3.21	2.88	2.58	2.42	3.40
	np.lc	3.62	3.46	3.42	3.06	2.92	2.49	2.10	1.91	3.16
	np.ll	5.14	4.17	3.68	2.57	2.44	2.52	1.69	1.61	3.38
MAE	arima	3.87	3.44	3.24	2.72	2.30	2.25	1.60	1.64	2.99
	dlm	3.43	3.16	3.21	2.71	2.50	2.23	1.59	1.53	2.83
	dlm_lin	3.43	3.16	3.26	2.73	2.43	2.22	1.66	1.79	2.95
	iss	3.43	3.22	3.08	2.80	2.36	2.25	1.61	1.64	2.93
	Lpv	3.43	3.16	3.08	2.79	2.35	2.25	1.61	1.64	2.92
	avew	3.55	3.10	3.10	2.79	2.40	2.18	1.58	1.55	2.91

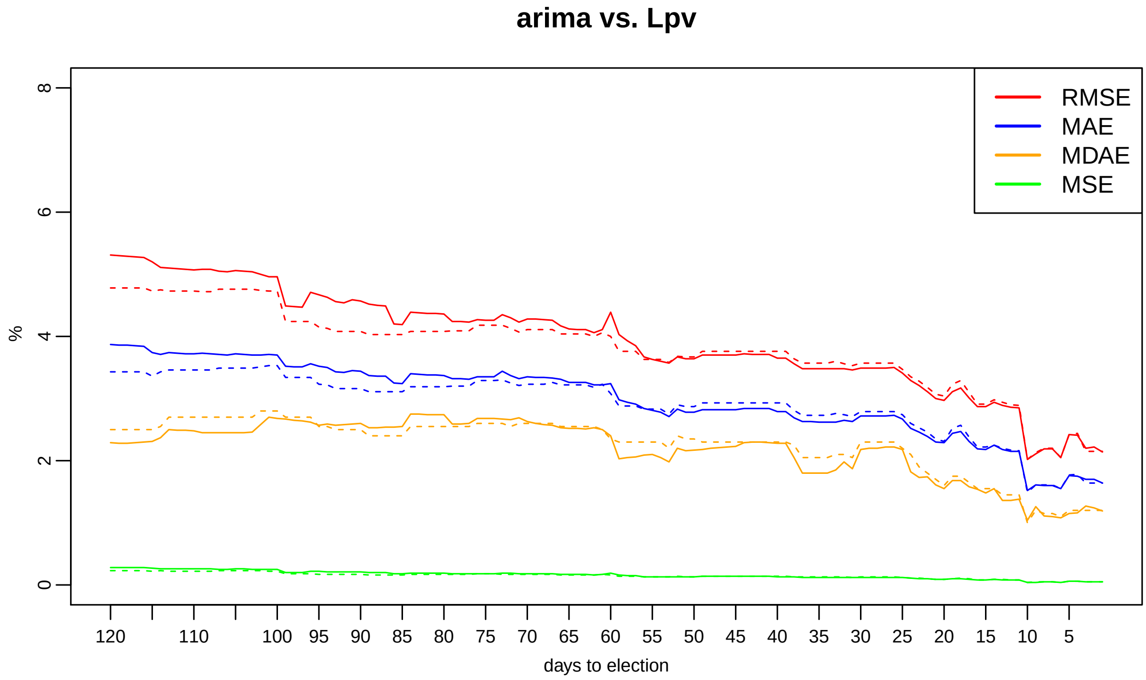
Table 7.6 – continued from previous page

Measure	Method	120	90	60	30	21	14	7	1	Avg.
	`med`	0.26	0.24	0.23	0.19	0.16	0.13	0.10	0.09	0.20
	`avg`	0.25	0.22	0.21	0.18	0.16	0.14	0.11	0.10	0.20
	`avg.wgt`	0.26	0.23	0.21	0.18	0.16	0.14	0.11	0.10	0.20
	`np.lc`	0.24	0.21	0.20	0.16	0.14	0.10	0.08	0.06	0.18
	`np.ll`	0.57	0.36	0.29	0.12	0.10	0.14	0.05	0.05	0.26
MSE	`arima`	0.28	0.21	0.19	0.12	0.09	0.09	0.05	0.05	0.16
	`dlm`	0.23	0.17	0.16	0.12	0.10	0.09	0.05	0.04	0.15
	`dlm_lin`	0.23	0.17	0.17	0.13	0.10	0.08	0.05	0.05	0.15
	`iss`	0.23	0.17	0.16	0.13	0.09	0.09	0.05	0.05	0.15
	`Lpv`	0.23	0.17	0.16	0.13	0.09	0.09	0.05	0.05	0.15
	`avew`	0.25	0.16	0.16	0.12	0.10	0.08	0.05	0.04	0.16

Figure 7.6 illustrates the performance, computed with RMSE, MAE, MDAE and MSE of the method `arima` compared to the benchmark, averaged over all regions and parties. The solid lines indicate the respective performance measures of the `arima` method, the dotted lines the ones for the benchmark. First, the accuracy of both methods improves with decreasing lead time. All measures, stated in Figure 7.6, are scale dependent measures, which means, that the scale of the accuracy measure depends on the scale of the data, compare Hyndman and Koehler (2006). The RMSE is more sensitive to outliers than the MDAE and MAE, seen in the peak at lead time 60 in the red line (RMSE), in contrast to the blue line (MAE), where the small peak is hardly not observable. The yellow lines of MDAE indicate, that the solid line (`arima`) is often below the dotted line (`Lpv`), where the `arima` method often outperforms the benchmark. The differences of both methods are small in absolute numbers, but in relative numbers there are gains of up to 7% for some lead times.

Figure A.15 in the Appendix shows the performance of the four accuracy measures for the method `np.ll` compared to the benchmark. Analogous to the Figure 7.6 the values are averaged over all regions and all parties for lead times between 120 and 1 day before election. The absolute gains of the `np.ll` method compared to `Lpv` are up to 0.46 percentage points with the measure MDAE (yellow line), which means 25% relative improvement. The sensitivity of the RMSEs to outliers

Figure 7.6: Comparison of the method np.ll with the benchmark Lpv according to different accuracy measures MDAE, RMSE, MAE and MSE. Values are computed by averaging over all parties and regions. Dotted lines show the values for Lpv, the solid lines the values for arima.



is observed for the lead times 120 to 60 days, where the MDAE obtains less deviation of the np.ll to the benchmark than the RMSE. Especially the extreme outlier between lead time 20 and 10 in the figure is hardly not observable for the MDAE in contrast to the RMSE.

Finally, the performance of the methods is measured by averaging over parties and regions. By investigating the performance, not only the RMSE, but also the MSE, MDAE and MAE are used. The RMSE provides relative gains of up to 25% by using np.ll or arima instead of Lpv. Overall, there is no clear best method by averaging over parties and regions.

Although, there are only small gains in absolute values, compared to the latest poll value. The methods are easy to implement and provide relative gains of up to 50% for some methods and averaging options. Compared to other forecasters stated in the Section above, the methods in this work show lower RMSEs and some methods hit the correct vote share on the first decimal place like in the BW election for the Greens. Using these forecasting methods apart from the `Lpv` gives an additional information to the voters and can even smooth the biased poll results (caused by some types of errors).

8 Conclusion

State elections are in between the national election cycle and stretched over the four-year cycle. The penultimate German national election took place in 2013, the next one took place in September 2017. State elections in this work were conducted in the time period between 2014 and 2017 and fill the gap between two national cycles. They provide a good orientation for the national elections in terms of the current sentiment of the voters and are direction elections for the national ones, especially the state elections in densely-populated states like Northrhine Westphalia. In particular, a state election shows the real voting behavior, e.g. how many people really vote for the right wing parties as many voters do not admit to vote for certain parties. The importance of state elections grew over the last years. Hence, state elections provide information about the current party landscape and show emerging parties e.g. the AFD who was not part of the national government in 2013. Whereas national elections and polls were in focus until today, now state level polls and elections are coming more in forefront. Polls especially in the run up of elections are an important tool to show the current sentiment of the voters opinion. Over the last years, the polling industry grew and provided a larger number of polls than in the years before.

Polls are a snapshot of the current sentiment and are often taken as direct forecast by media and public. In this thesis, polls are seen as data base for statistical forecasts in aggregate models and as a benchmark in form of the "latest poll value". Polls on state level are published on irregular dates and reveal differences in the number of polls between states. I propose a closer look at the poll values and provide statistical forecasts of vote shares based on the irregular time series of polls as well as forecasts based on generated regular daily time series of polls. The forecasts are computed for the single vote shares in the German multi-party system and are conducted for 13 out of 16 state elections.

Difficulties in forecasting single vote shares are obtained by the changing voting behavior, seen in variations in polls during the campaign due to the high number of undecided voters. Further, the changing party landscape is demonstrated by losses of votes for the "catch all parties" and fragmented parliaments with five or six different parties. Polls are the only way to detect these short term changes in contrast to fundamental models, which deal rather with long term variables. The forecast errors and poll quality, measured with the `'A'` for the benchmark method `Lpv` support these changes. The first finding concerns the performance of the `Lpv`. The benchmark improves over lead time and shows a good poll quality for the main parties, measured by the measure `'A'`. In most states, this measure indicates an underestimation of the polls compared to the real election results for the AFD. The `Lpv` as benchmark fills a gap between two published polls by repeating the last available poll until a new poll appears. Nevertheless, the `Lpv` provides a daily value of a poll, the time between the last available poll and the next one can be remarkably great and also the deviations of the values.

Hence, state elections in this thesis are forecasted with different methods and evaluated with the measures RMSE, MAE, MDAE and MSE with different levels of aggregation and with lead times from 120 to 1 day before election. The methods show an improvement over lead time analogous to polls. One task of this work is to find methods, which beat the `Lpv` in accuracy. In the findings of Chapter 7 there can be found relative improvements of up to 50% compared to the `Lpv` in some cases and for specific lead times. The performance measures are calculated with respect to different levels of aggregation e.g. per party or region. No general best method exists to forecast the vote shares in state elections. Some methods perform better in specific states e.g. the method `np.ll` performs best in Saxony Anhalt and Berlin by computing the average RMSE over all parties and lead times. Whereas in Thuringia, the weighted average `avg.wgt`, the unweighted average `avg` and median `med` perform better than other methods and the benchmark `Lpv`. By calculating forecasts of state elections for daily lead times between 120 and 1 day before election, the gap of publishing polls at irregular intervals is filled. Forecasting methods especially in states with very few polls like SH, provide better RMSEs compared to the benchmark method on lead times, where no poll is published. The performance measures are computed for all states and all lead times. To control for "learning effects" of the polls and forecasting methods over elections, the performance of forecasting methods is evaluated for specific years. Averaging over all lead times and specific regions depending on year indicates higher values for the RMSE in 2016, where state elections in BW, RP, ST, MV and WBE took place. In 2017, for SL, SH and NW state elections, lower values of average RMSE are computed, as the polls provided lower levels of underestimation of the AFD. The improvement in forecasting performance from the year 2016 to 2017 in state elections appeared due to an improvement in polls. As polls are the data base in this thesis, the forecasting methods also improve in their forecasting performance. Compared to other election forecasting models, the forecasts in this work perform well with values for the RMSE between 2-4% in most cases, besides the elections in 2016, which display higher values of the RMSE.

Over the last years, polls gained in importance as data base in aggregate models, as they improved in quantity and quality. As state elections and polls on state level detect trends especially for new parties, they grow in importance as direction elections for the national elections. The AFD is up to now part of 13 out of 16 state parliaments. In contrast to the Pirates in 2006, who only appeared in two state parliaments, the AFD is now established and gained over 5% in the state elections in 2014, 2015, 2016 and 2017. Hence, state level polls and their forecasts are therefore important for national elections to see if a party becomes an established party with a seat in the government on state and then national level.

Extensions of this work can be made in terms of variations of the forecasting models and by including error corrections due to learning effects from election to election. The forecasting methods can be extended by choosing variations of the used models e.g. trying alternatives of the dynamic linear model.

A Appendix-Graphs

Figure A.1: Voter turnout rates of the West German states between 1950 and 2017 (dotted lines) and East German states between 1990 and 2017 (solid lines). The points and triangles indicate the election dates.

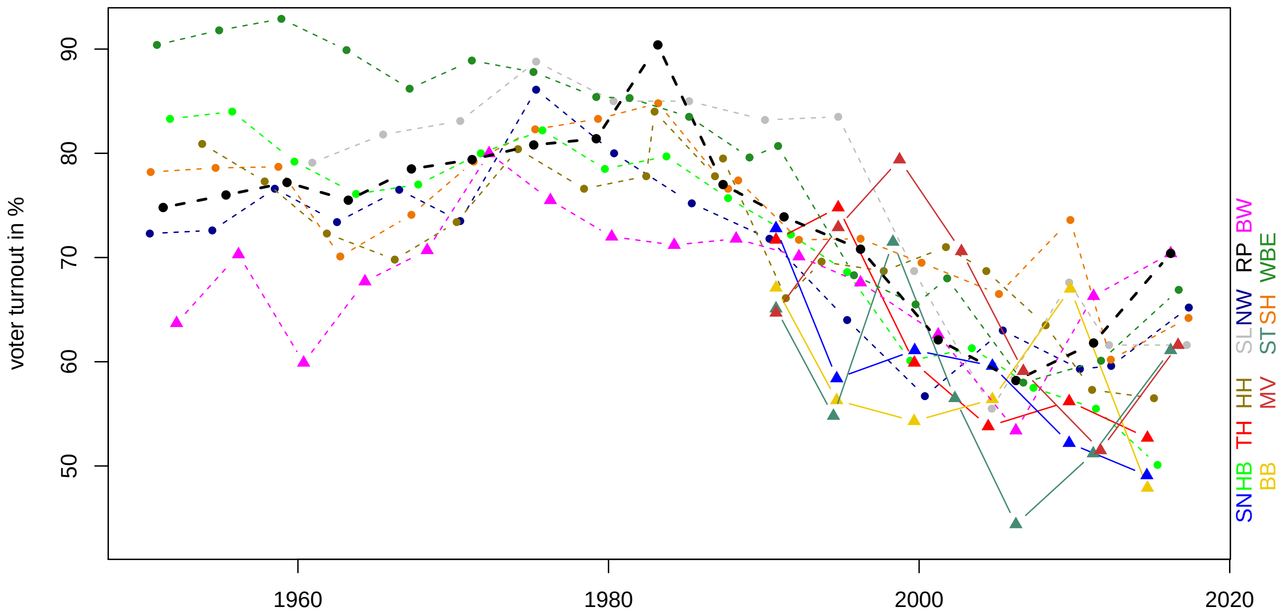


Figure A.2: Boxplots of the voter turnout rates for the 13 investigated states between 1950 and 2017, displayed for every year. After an increase in the 1970s, there was a sharp decline in the 1980s.

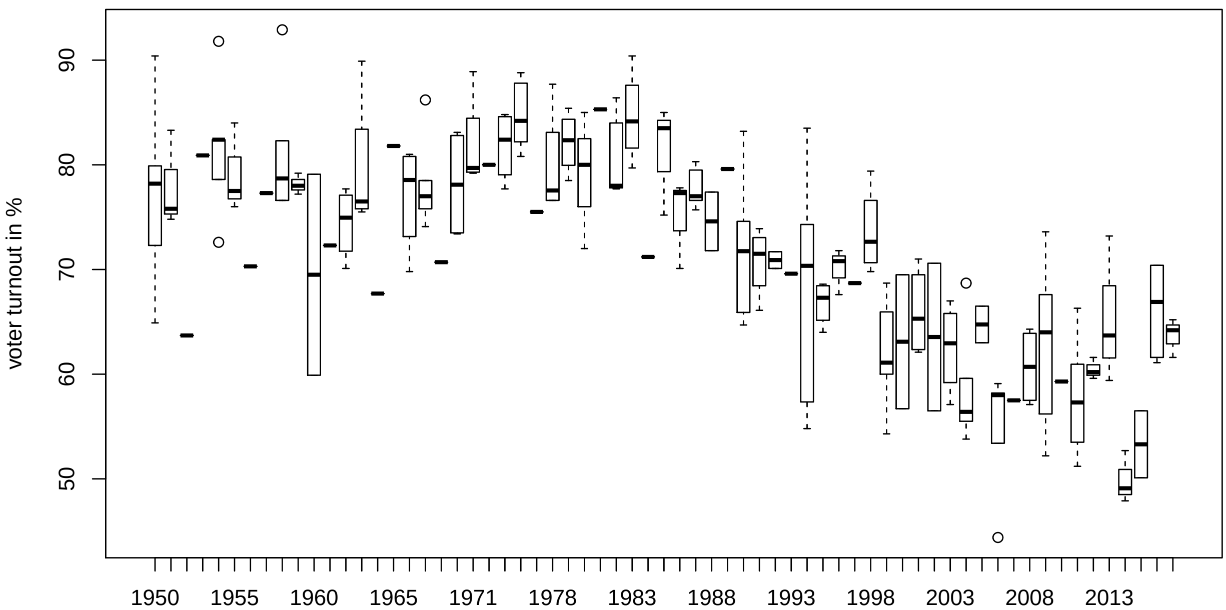


Figure A.3: Boxplots of the margin of errors calculated with the factor of the 95% confidence interval for the three institutes (Emnid, Forsa, Infratest dimap), depending on party across all 13 states (365 days before the respective election date).

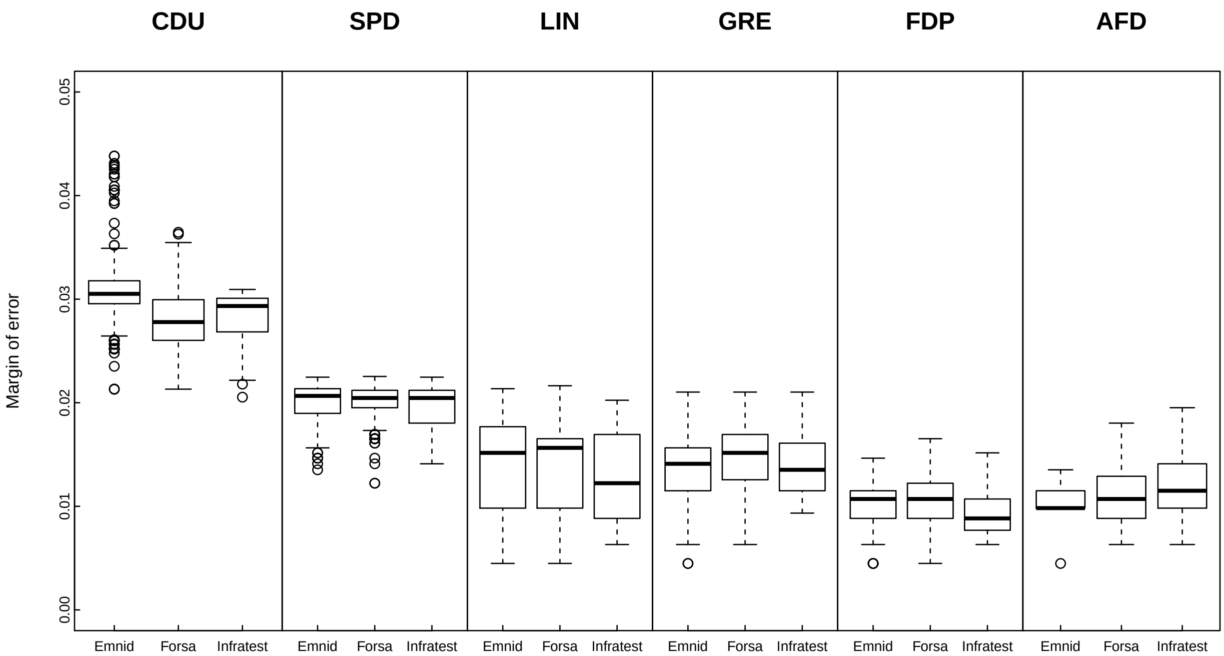


Figure A.4: 95% confidence intervals of the margin of errors for polls of the six parties 365 days before the state election in NW on 2017-05-14. Dots indicate the poll dates and values, lines indicate the interval width.

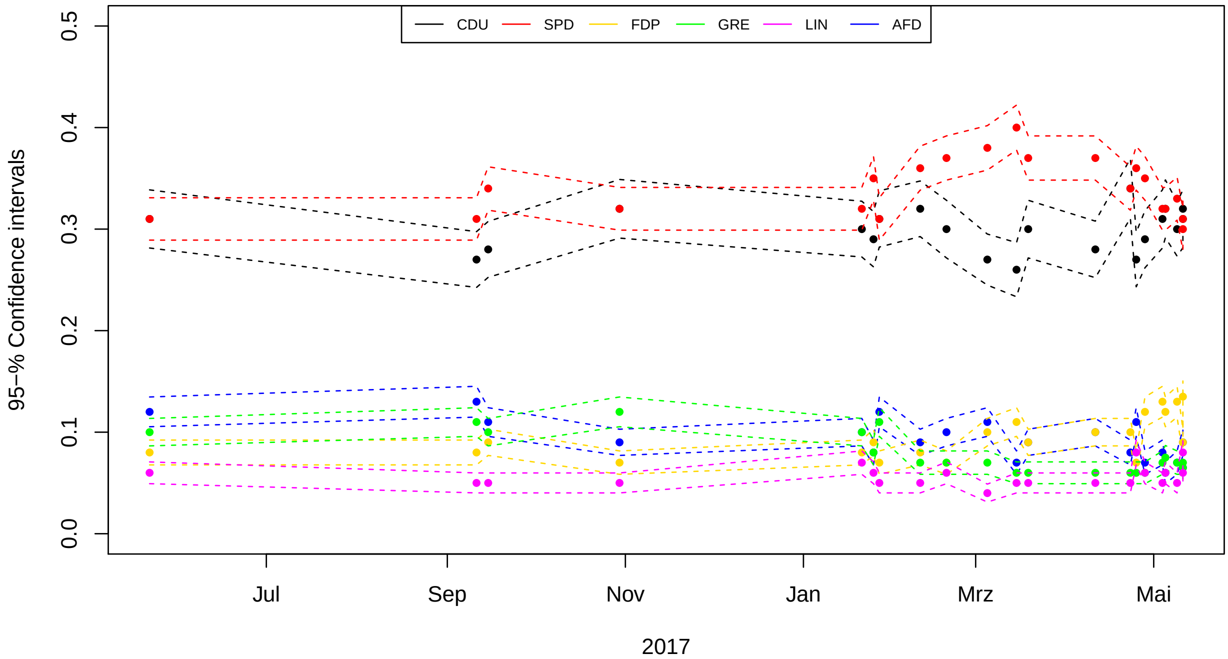


Figure A.5: Regularly generated vote shares of polls for the six parties in the SH state election on "2017-05-07". Vertical grey lines show the dates, where polls are published.

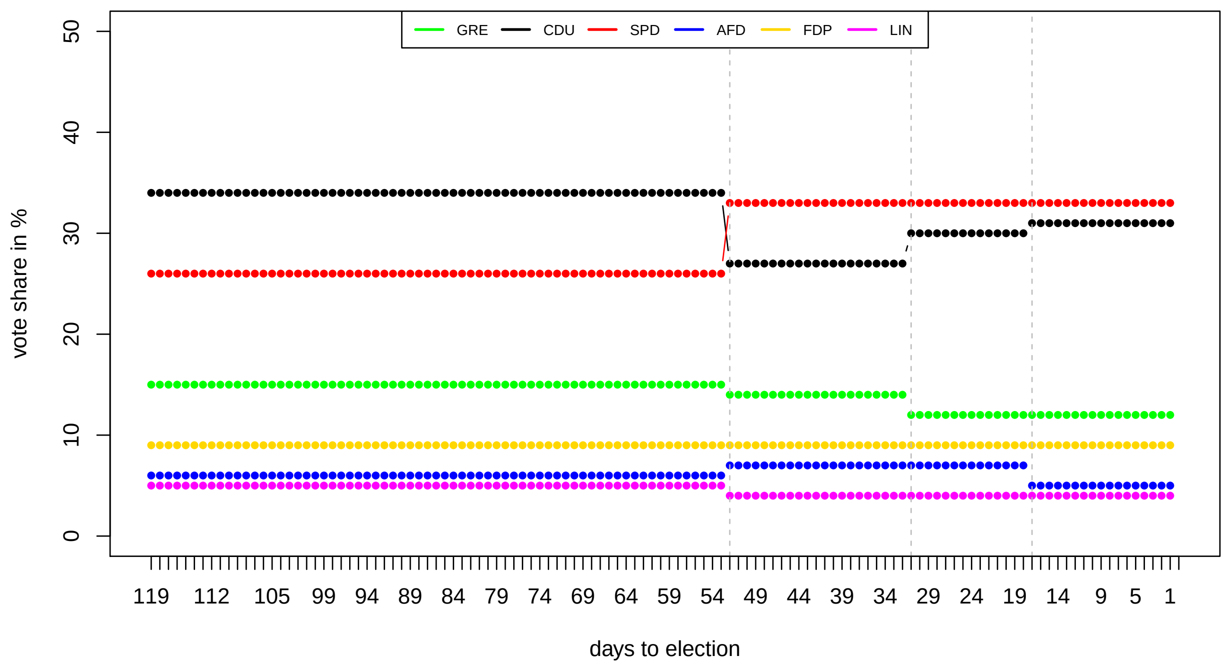


Figure A.6: Regularly generated vote shares of polls for the six parties in the SN, BB, TH state elections 2014. Vertical grey lines show the dates, where polls are published, the asterisks indicate the election outcome on the respective election dates.

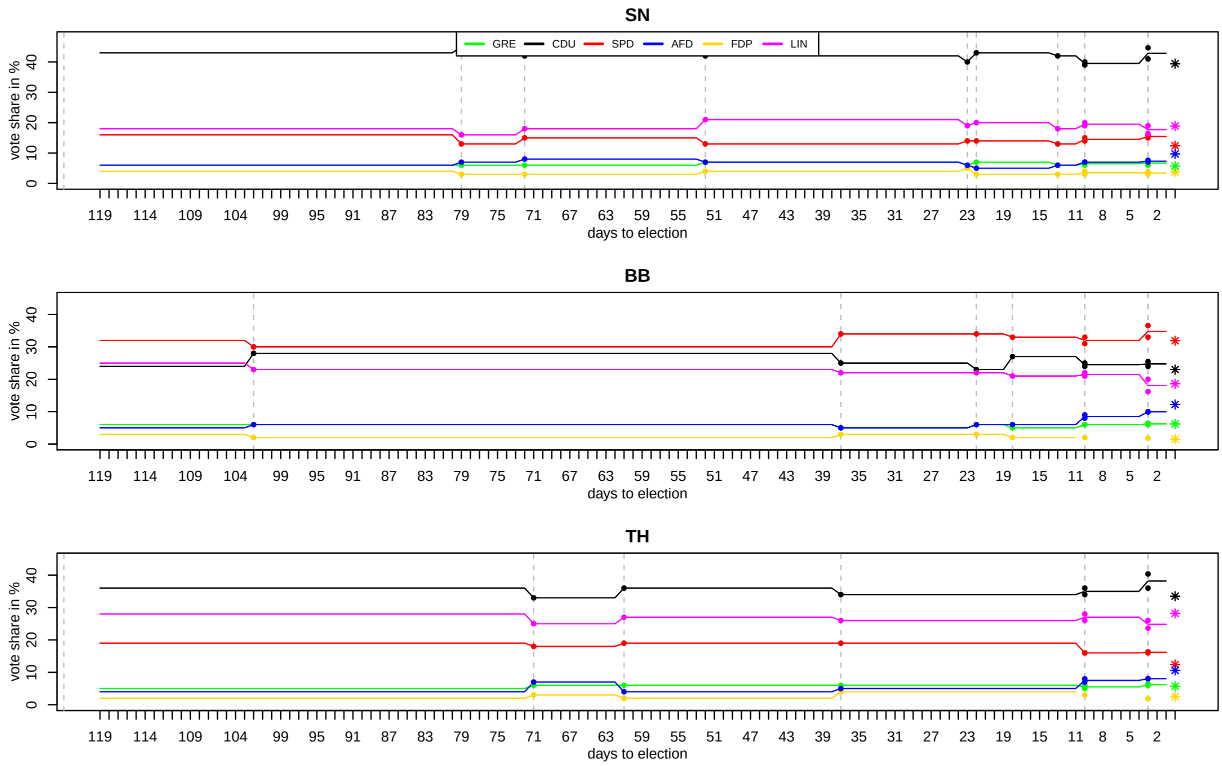


Figure A.7: Regularly generated vote share of polls for the six parties in the HH and HB state elections 2015. Vertical grey lines show the dates, where polls are published, the asterisks indicate the election outcome on the respective election dates.

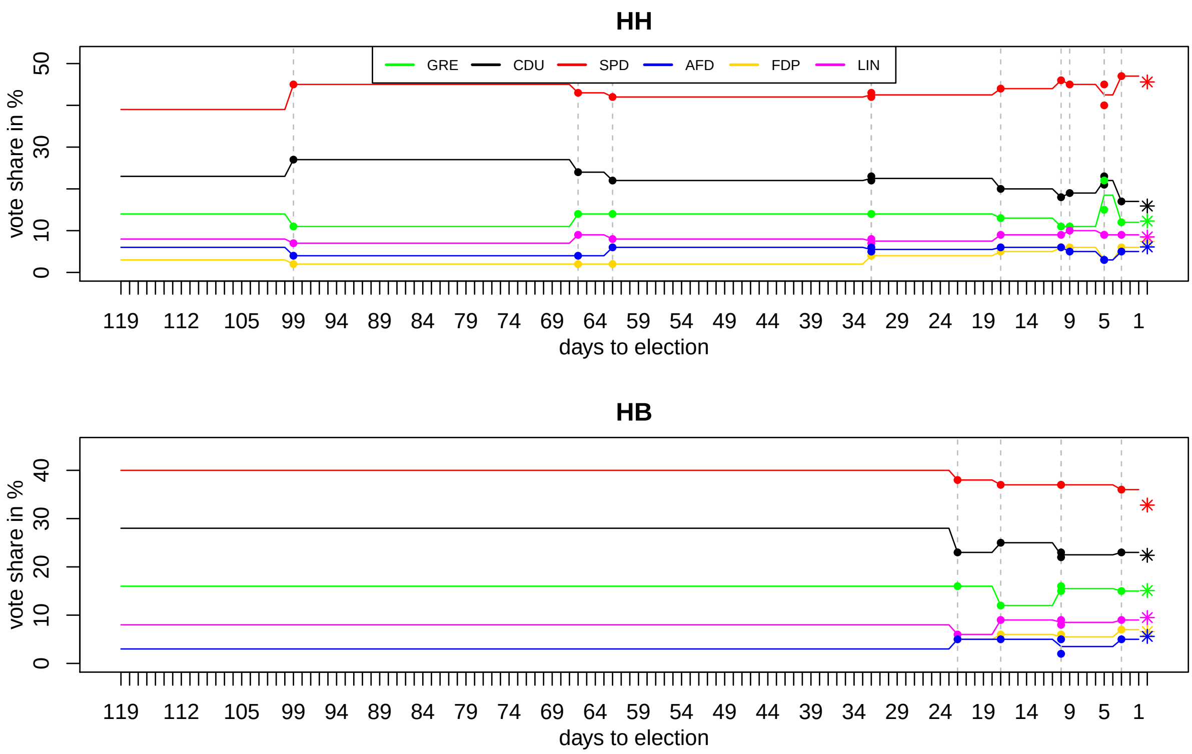


Figure A.8: Regularly generated vote share of polls for the six parties in the SL, SH, NW state elections 2017. Vertical grey lines show the dates, where polls are published, the asterisks indicate the election outcome on the respective election dates.

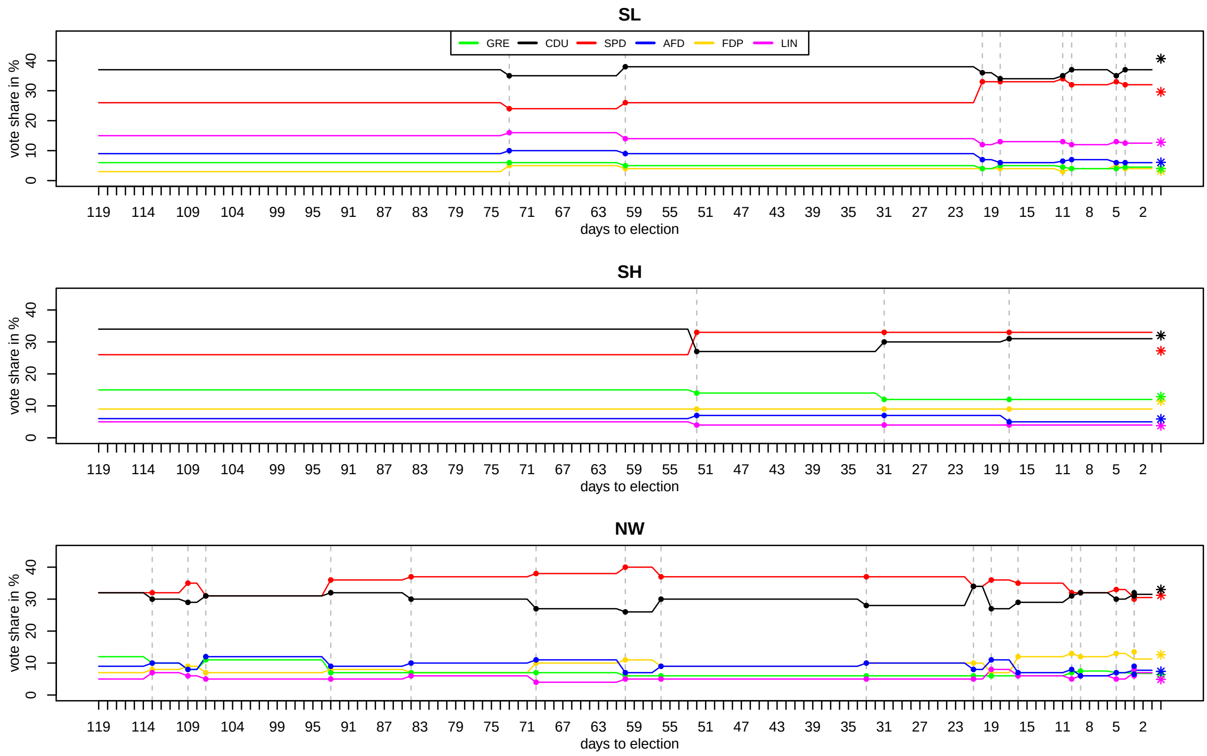


Figure A.9: Number of state level polls in SH and NW before the elections between 1995 and 2017 (black dots). Grey vertical lines indicate the election dates.

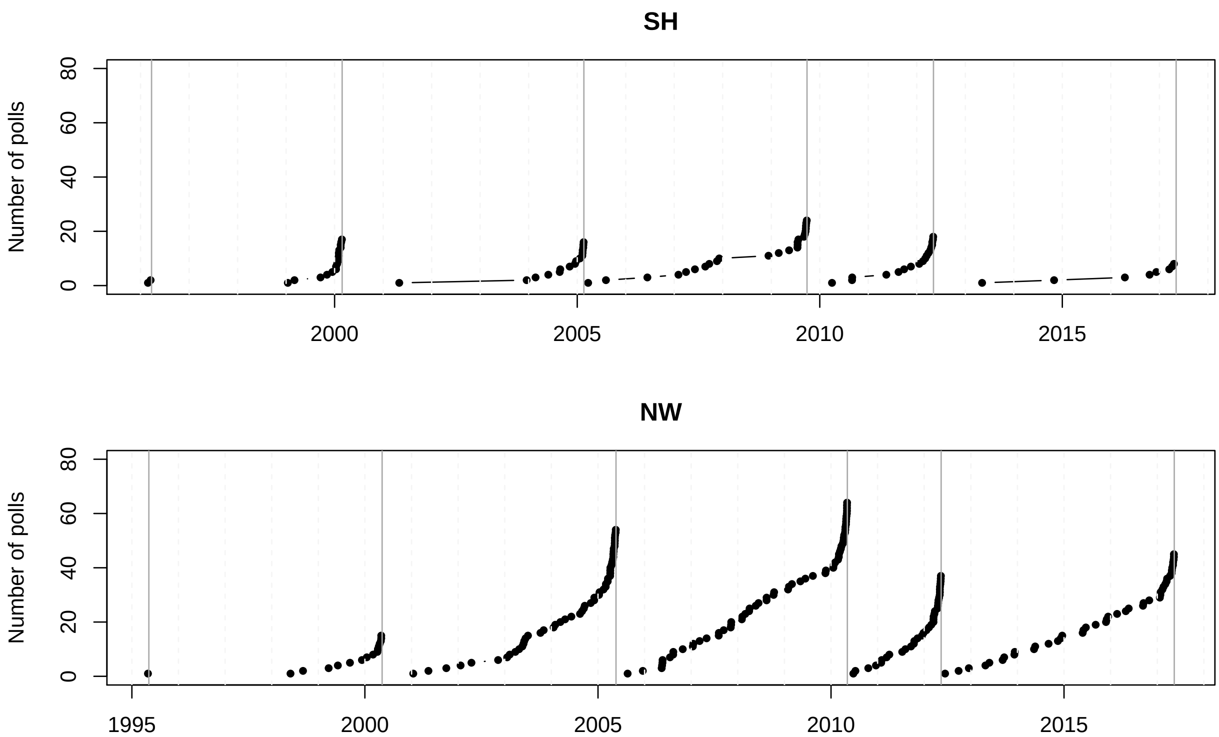


Figure A.10: Running sample means of the evolution variance of a local level DLM for the NW state election, with 1000 MCMC iterations.

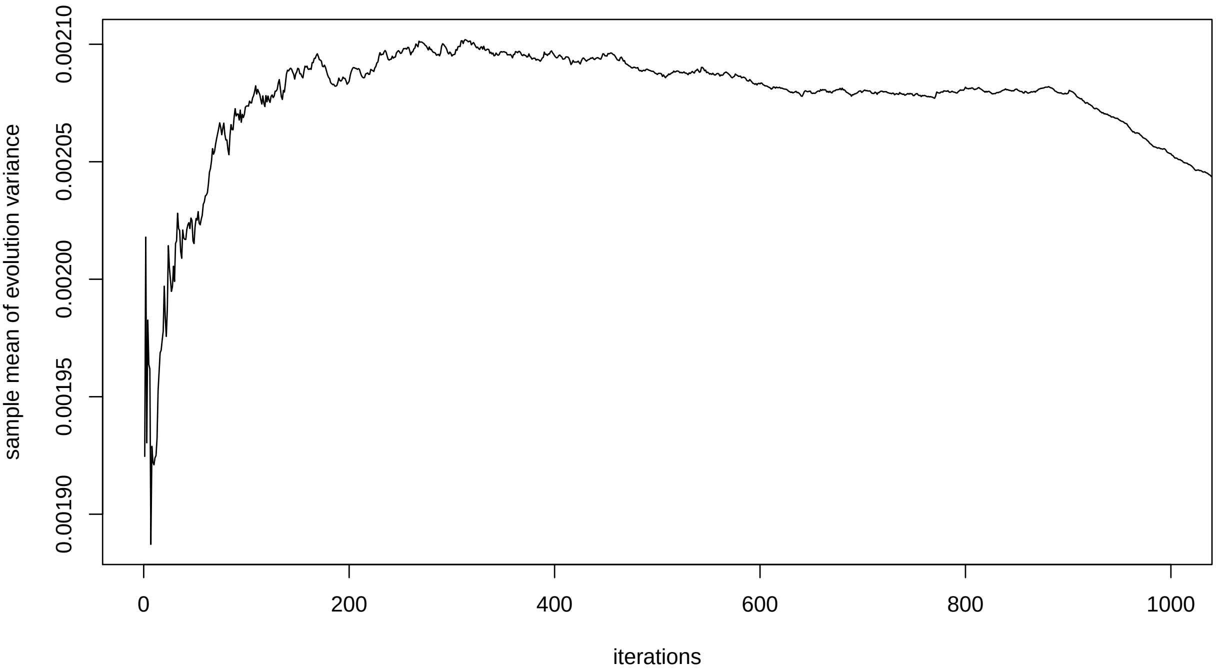


Figure A.11: Running sample means of the observational variance of a local level DLM for the NW state election, with 1000 MCMC iterations.

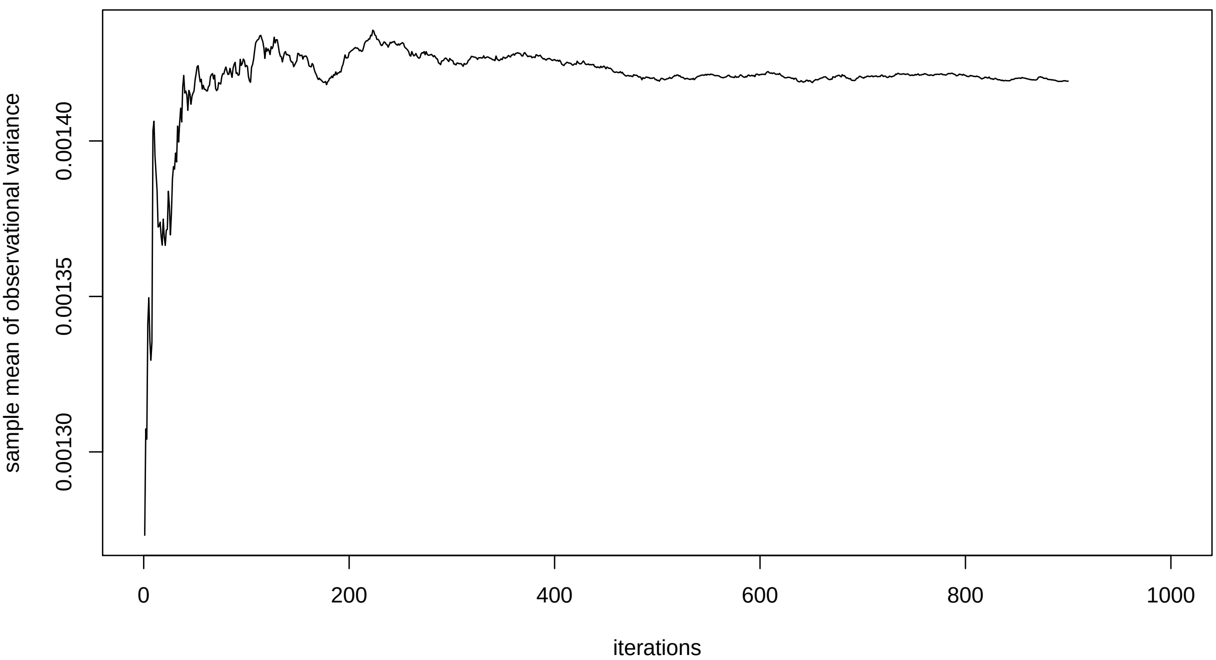


Figure A.12: Poll quality measured with the extended '`A`' for the investigated states and the six parties. The horizontal line indicates, that polls and election outcomes are equivalent.

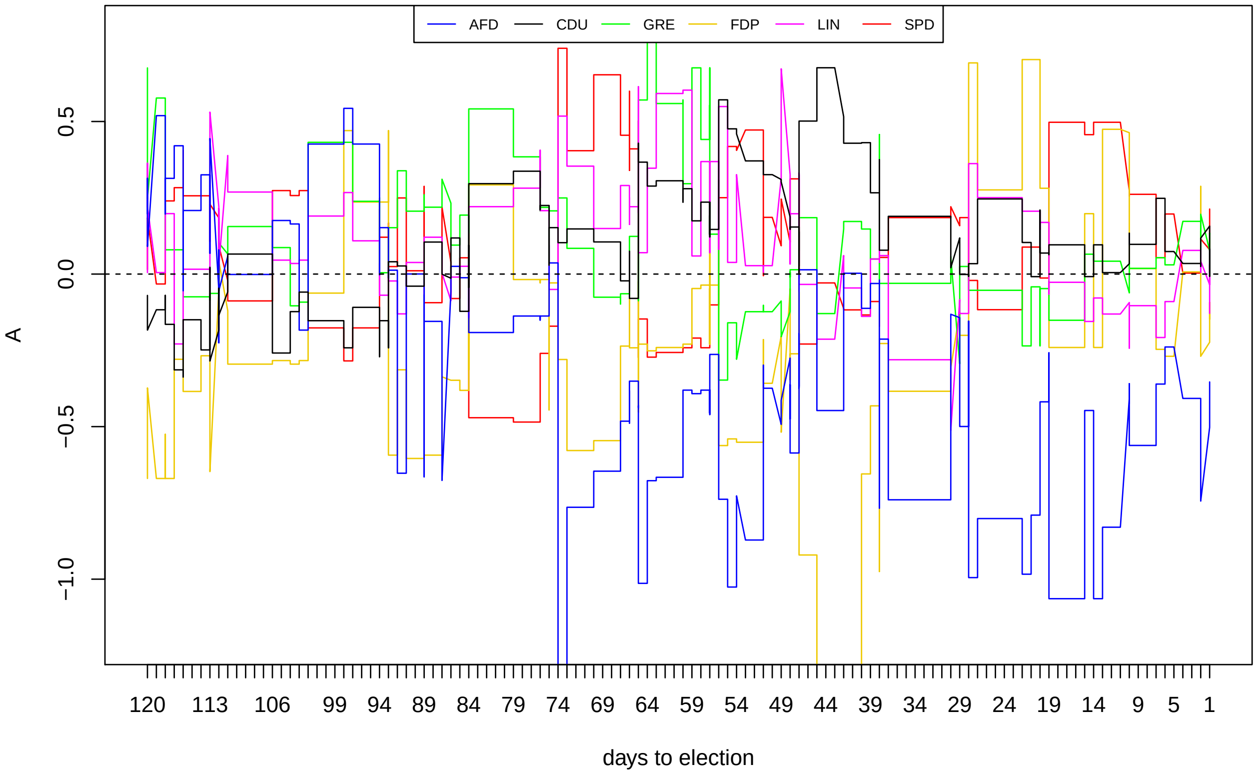


Figure A.13: Forecasts with the method `Lpv` and real election results in the ST state election “2016-03-13”. Dotted lines indicate the real outcomes for the election, solid lines the forecast values for 120 to 1 day lead time.

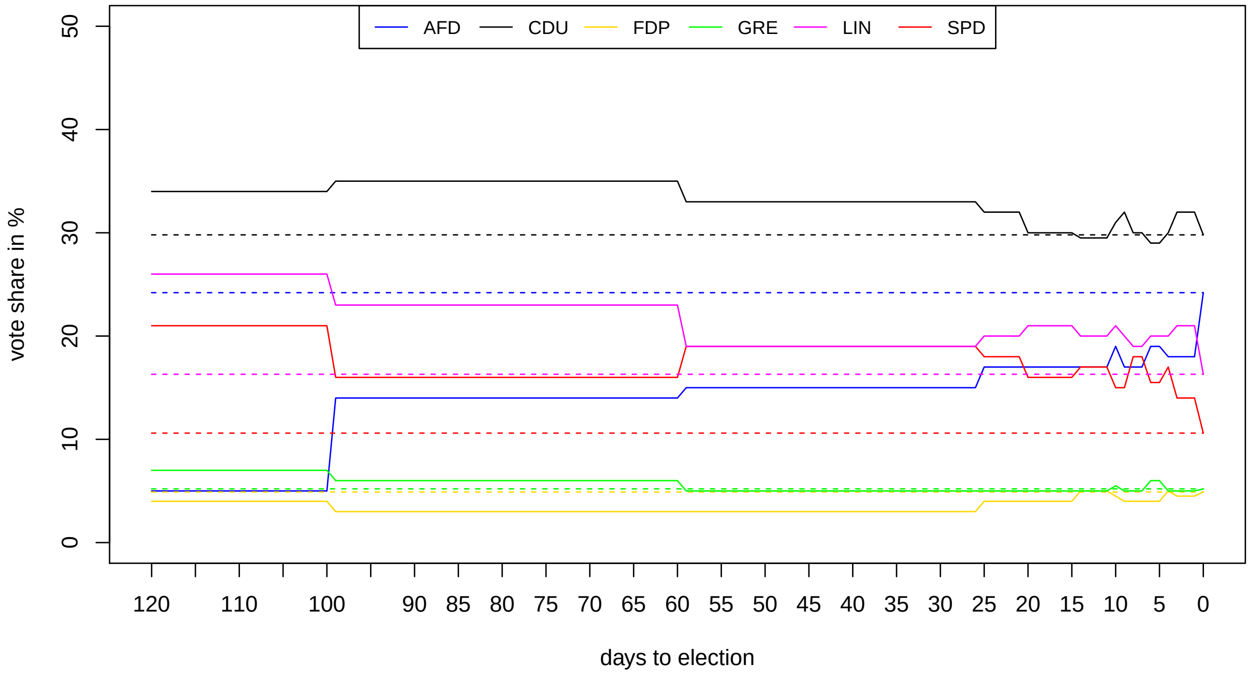


Figure A.14: MDAE (in percent) computed by averaging over regions depending on year. The MDAEs calculated in 2014 contain the elections in SN, BB and TH, in 2015 HB and HH, in 2016 BW, RP, ST, MV and WBE and in 2017 SL, SH and NW. Points display the average values for every party over all lead times per method.

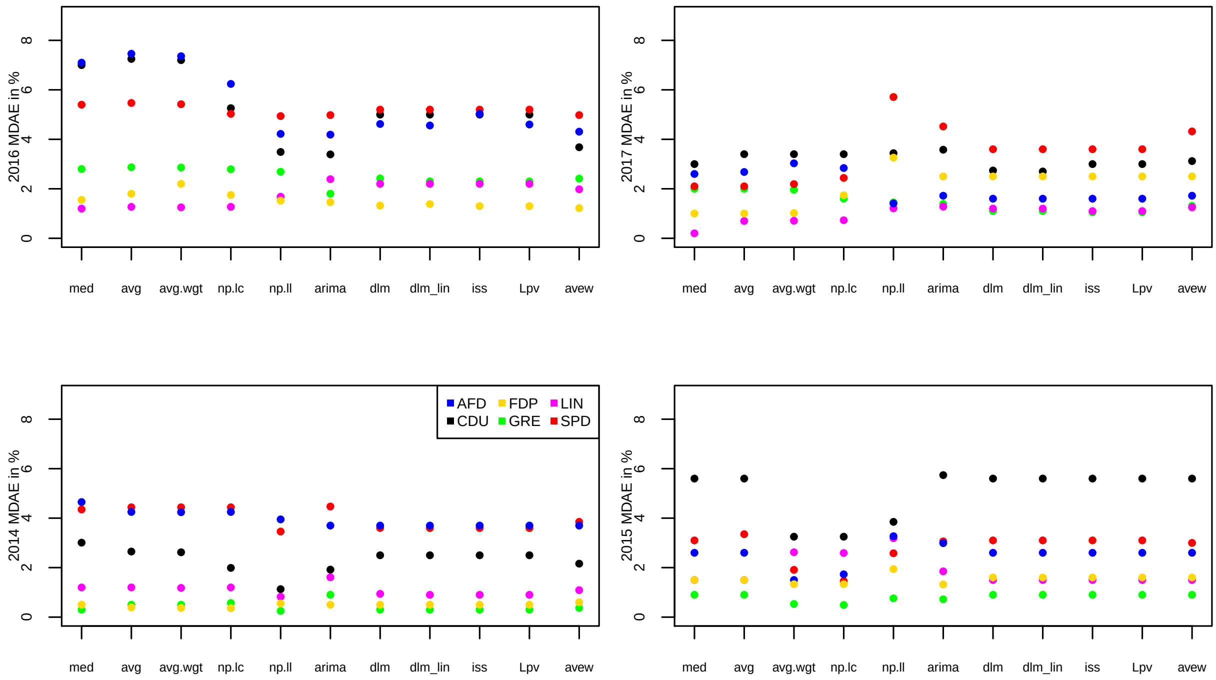
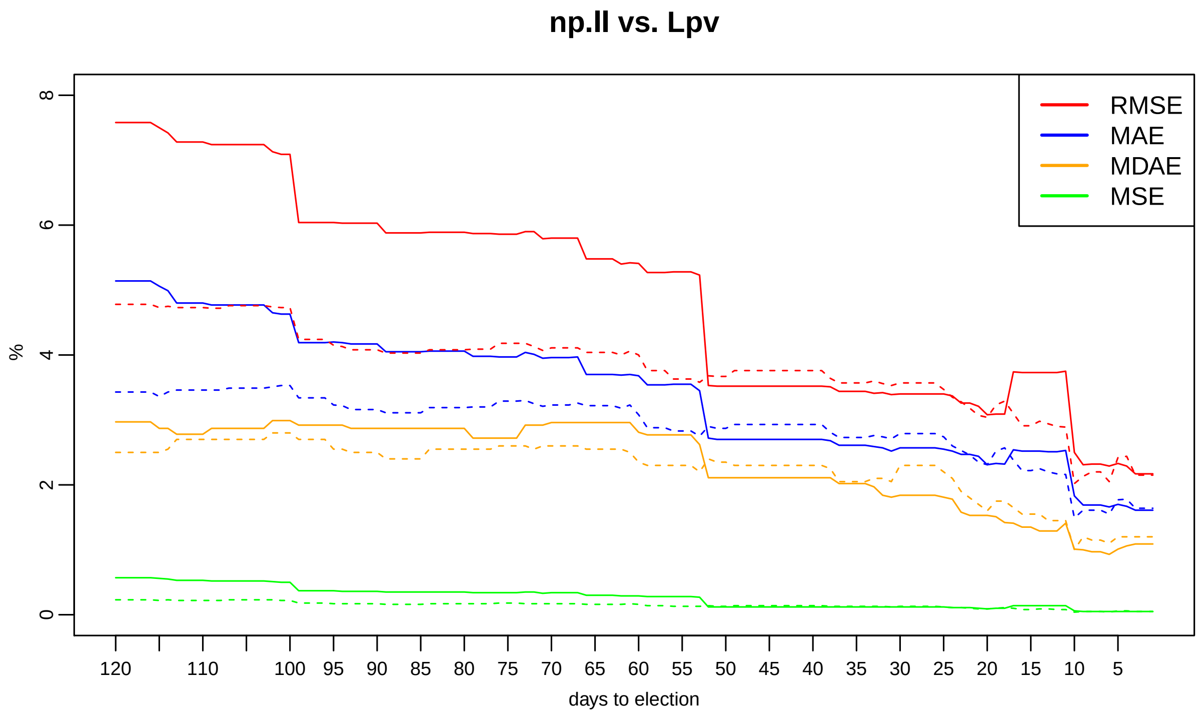


Figure A.15: Comparison of the method `np.ll` with the benchmark `Lpv` according to different accuracy measures MDAE, RMSE, MAE and MSE. Values are computed by averaging over all parties and regions. Dotted lines show the values for `Lpv`, the solid lines the values for `np.ll`.



B Appendix-Tables

Table B.1: Overview of existing election forecasting literature ordered by author, method and investigated country. The rightmost column displays the independent variables in the case of an regression, or simply variables for other methods.

Author	Country	Forecasting objective	Method	(Independent) variables
Abramowitz (1996)	US	Incumbent single vote share	OLS Regression	Approval rating, GDP, terms in office
Abramowitz (2008)	US	Incumbent single vote share	OLS Regression	GDP, terms in office, Incumbent
Aichholzer and Willmann (2014)	Austria	Coalition vote share	OLS Regression	Unemployment, incumbency, unemployment x incumbency
Andersen and Heath (2003)	US,UK, Canada	Single vote share of left, right	OLS Regression	Age, gender, religion, race, national identity, education, social class
Armstrong and Graefe (2011)	US	Incumbent single vote share	OLS Regression	Biographical index
Auberger (2012)	France	Single vote share of left, right	OLS Regression	Unemployment, GDP, inflation
Bélanger, Lewis-Beck and Nadeau (2005)	UK	Incumbent single vote share	OLS Regression	Inflation, public approval, terms in office

Table B.1 – continued from previous page

Author	Country	Forecasting objective	Method	(Independent) variables
Bélanger, Nadeau and Lewis-Beck (2010)	UK	Single vote share	OLS Regression	Opposition, leader approval
Bélanger and Soroka (2012)	US, UK, Canada	Incumbent single vote share	OLS Regression	Approval, economy, time
Bellucci (2010)	Italy	Incumbent single vote share	OLS Regression	Government approval, economic sentiment, time in office, electoral area
Brown and Chappell (1999)	US	Democrats single vote share	OLS Regression	Days before election, poll share, democrats x days before election
Cuzan (2012)	US	Democrats single vote share	OLS Regression	Fiscal, growth, all news, duration
Duquette, Mixon and Cebula (2013)	US	Incumbent single vote share	OLS Regression	Tenure, leadership
Graefe (2013b)	US	Incumbent single vote share	OLS Regression	Party identification, issue, leadership
Graefe (2014)	US	Incumbent single vote share	OLS Regression	Survey respondents to win
Holbrook and DeSart (1999)	US	Democrats single vote share	OLS Regression	Prior vote, September polls
Hummel and Rothschild (2014)	US	Democrats incumbent vote share	OLS Regression	Presidential approval, state outcomes, home state

Table B.1 – continued from previous page

Author	Country	Forecasting objective	Method	(Independent) variables
Jérôme, Jérôme and Lewis-Beck (1999)	France	Single vote share	OLS Regression	Past presidential vote, unemployment change, regional ideology, political instability
Jérôme and Jérôme-Speziari (2011)	US	Incumbent single vote share	OLS Regression	Economics, incumbent credibility, partisan stronghold, politics and institutions
Khemani (2001)	India state and national	Incumbent single vote share	OLS Regression	Local and national growth, poverty, inequality, state income
Leigh and Wolfers (2006)	Australia	Incumbent single vote share	OLS Regression	Unemployment, inflation, honeymoon, wage, GDP growth
Lemennicier, Lescieux-Katir and Grofman (2010)	France	Single vote share	OLS Regression	Distance of the right wing leader to the overall median vote
Lewis-Beck, Bélanger and Fauvelle-Aymar (2008)	US, UK, France	Single vote share	OLS Regression	Economy, popularity, previous vote
Lewis-Beck, Nadeau and Bélanger (2004)	UK	Incumbent single vote share	OLS Regression	Inflation, public approval, number of terms in office
Lewis-Beck and Rice (1984)	US	Incumbent single vote share	OLS Regression	President, GNP
Lewis-Beck and Skalaban (1989)	US	Incumbent single vote share	OLS Regression	Partisan identification, interest, media, politics
Lewis-Beck and Stegmaier (2000)	US	Incumbent vote choice	OLS Regression	Popularity, inflation, unemployment, Vietnam, Watergate

Table B.1 – continued from previous page

Author	Country	Forecasting objective	Method	(Independent) variables
Lewis-Beck and Tien (1996)	US	Incumbent single vote share	OLS Regression	Economics, party strength, candidate appeal
Lewis-Beck and Tien (2011)	US	Incumbent single vote share	OLS Regression	Growth, popularity
Lockerbie (2008)	US	Single vote share	OLS Regression	Time in White House, own situation
Magalhaes, Aguiar-Conraria and Lewis-Beck (2012)	Spain	Incumbent single vote share	OLS Regression	European elections, inflation, unemployment rate
Norpoth (2004)	US	Incumbent single vote share	OLS Regression	Incumbent, opposition candidate, electoral cycle, partisanship
Sanders (1995)	UK	Popularity	OLS Regression	Popularity, personal expenditure, tax
		Personal expectation	AR(2)	Interest rate, inflation, tax, personal expectation
Stegmaier and Lewis-Beck (2009)	Hungary	Single vote share	OLS Regression	Vote intention, unemployment rate
Whiteley (2005)	UK	Single seat share	OLS Regression	Party vote in polls, seats of parties
Williams et al. (2012)	US	Single vote share	OLS Regression	Age, race, involvement, leadership, party affiliate charisma
Wlezien and Erikson (2004)	US	Incumbent single vote share	OLS Regression	Cumulative growth, presidential approval, trial heat poll results
Wlezien et al. (2013)	UK	Single vote share	OLS Regression	Vote intention in polls

Table B.1 – continued from previous page

Author	Country	Forecasting objective	Method	(Independent) variables
Bafumi, Erikson and Wlezien (2010)	US	Democrats single vote share	Simulation	Specification prediction, local component
Berg, Nelson and Rietz (2003)	US	Single vote share	AR(1)	
Berlemann, Enkelmann and Kuhlenkasper (2014)	US	Approval vote share	Semiparametric additive mixed	Inflation, unemployment, government consumption, time in office, Vietnam, Afghanistan, Iraq
Brown, Firth and Payne (1999)	UK	Single vote share	Ridge Regression	
Brooks, Dodson and Hotchkiss (2010)	US	Incumbent single vote share	Logistic regression	Defense spending, social (refugee, race, gender), income
Cuzan, Armstrong and Jones (2005)	US	Incumbent single vote share	Pollyvote	Polls, IOWA, Delphi, experts
Dolado, Gonzalo and Mayoral (2003)	Spain	Single parties	ARIFMA	
Graefe (2013a)	US	Incumbent single vote share	Multiple regression, equal weights model	
Graefe and Armstrong (2011)	US	Republican, Democrats	Biographical index	

Table B.1 – continued from previous page

Author	Country	Forecasting objective	Method	(Independent) variables
Graefe and Armstrong (2012)	US	Incumbent single vote share	Big issue voting model regression	Vote support
Gott and Colley (2008)	US	Single vote share	Median of polls	
Gordon (2010)	US	Republican expert ratings	Bayesian Simulation	Expert ratings
Hanretty (2013)	Italy	Single vote share	Simulation	
Heij and Franses (2011)	Netherlands	Vote share probability	Binary logit model	Voting attitude, previous voting, political interest
Hobolt and Spoon (2012)	European Parliament	Multinomial partisanship abstention switching	Generalized linear model with logit likelihood	Income level, predictions (partisanship, knowledge, education, female, age) political system variance
Jackman (2005)	Australia	Coalition vote share	Bayesian	
Kamakura, Mazzon and de Bruyn (2006)	Two stage elections	Single vote share value of candidate	Nested Logit Factor Model	Known demographic unobserved deviations in voter preferences
Klarner (2008)	US	Incumbent single vote share	Random effects, ML regression	Lagged vote, incumbency, approval, income growth, vote intention
Lebo and Norpoth (2011)	UK	Single vote share	ARIFMA	Approval, major, Falkland war, inflation, unemployment rate
Linzer (2013)	US	Incumbent single vote share	Bayesian	

Table B.1 – continued from previous page

Author	Country	Forecasting objective	Method	(Independent) variables
Mughan (1987)	UK	Single vote share government, alliance, opposition	Incremental	Party vote in preceding
			Opinion polling	Party's opinion poll in month before
			Economic model	Unemployment rate, GDP
Murr (2011)	UK	Single vote share	Logistic regression	Task difficulty, education, attention, age, female, income
Nadeau and Lewis-Beck (2001)	US	Incumbent single vote share	Logistic regression	National Business Index, real disposable income, incumbent party ID, race
Nadeau, Lewis-Beck and Bélanger (2010)	France	Single vote share prediction explanation	Multi equation regression	Vote=popularity, popularity= unemployment, term in office, approval
Strauss (2007)	US	Incumbent single vote	DLM	
Walther (2015)	Sweden, Germany	Single vote share	DLM	
Wang et al. (2015)	US	XBox Data depending on polls		
Whiteley (2008)	UK	Incumbent single vote share	Bivariate regression	Retail price index, public approval, numbers of terms in office
Ganser and Riordan (2015)	Germany	Single vote share	OLS Regression	Education, political knowledge, intention to participate, member of political organization,time before election

Table B.1 – continued from previous page

Author	Country	Forecasting objective	Method	(Independent) variables
Gschwend and Norpoth (2005)	Germany	Coalition vote share	OLS Regression	Long term party identification, chancellor support, term in office
Jérôme, Jérôme-Speziari and Lewis-Beck (2013)	Germany	Single vote share SPD, CDU, FDP, Grüne, Linke, others	OLS Regression	Unemployment rate, chancellor support, reunification
Kayser and Leininger (2013)	Germany	Coalition vote share	OLS Regression	Previous vote share, party ID, benchmark growth, log term in office
Norpoth and Gschwend (2003)	Germany	Coalition vote share	OLS Regression	Party support, chancellor support, term in office
Norpoth and Gschwend (2010)	Germany	Incumbent single vote share	OLS Regression	Long term partisanship, chancellor support, time in office
Gaissmaier and Marewski (2011)	BB, NW state elections Germany	Large vs. small parties	Intention based, recognition based, wisdom of crowds	
Graefe (2015)	Germany	Single vote shares CDU, SPD, LIN, FDP, GRE, AFD, Pirates, others	Polls, prediction markets PollyVote	
Küntzler (2014)	Germany	coalition vote share	Regression dynamic model	Long term partisanship, popularity of chancellor, tenure of government in office

Table B.1 – continued from previous page

Author	Country	Forecasting objective	Method	(Independent) variables
Lewis-Beck and Dassonneville (2015)	UK, Germany, Ireland	Incumbent single vote share	Regression Polling Synthetic	Government approval, GDP Median poll Government approval, GDP, polling
Selb and Munzert (2016)	Germany	Single vote share CDU, SPD, Linke, Grüne, FDP, others	Bayesian	Poll, party level, party institute

Table B.2: Summary of existing exponential smoothing methods.

		Season		
		N	A	M
	Trend	(None)	(Additive)	(Multiplicative)
N	(None)	N,N	N,A	N,M
A	(Additive)	A,N	A,A	A,M
Ad	(Additive damped)	Ad,N	Ad,A	Ad,M
M	(Multiplicative)	M,N	M,A	M,M
Md	(Multiplicative damped)	Md,N	Md,A	Md,M

Illustration from Hyndman and Khandakar (2008, p. 2).

Table B.3: Data base for expanding estimation windows used for `avew` method in this work.

w_i	Size of est. window	Data
w_1	4 weeks	$\{h, h+1, \ldots, h+27, h+28\}$
w_2	5 weeks	$\{h, h+1, \ldots, h+34, h+35\}$
w_3	6 weeks	$\{h, h+1, \ldots, h+41, h+42\}$
$\vdots$	$\vdots$	$\vdots$
w_{21}	24 weeks	$\{h, h+1, \ldots, h+167, h+168\}$

Illustration from Haupt, Schnurbus and Huber (2017, p. 8).

Table B.4: Forecasted vote shares with the method `Lpv` for the ST state election "2016-03-13" for selected lead times before election. The last column displays the real election result on lead time h=0.

Party	120	90	60	30	21	14	7	1	0
AFD	5.00	14.00	14.00	15.00	17.00	17.00	17.00	18.00	24.20
CDU	34.00	35.00	35.00	33.00	32.00	29.50	30.00	32.00	29.80
FDP	4.00	3.00	3.00	3.00	4.00	5.00	4.00	4.50	4.90
GRE	7.00	6.00	6.00	5.00	5.00	5.00	5.00	5.00	5.20
LIN	26.00	23.00	23.00	19.00	20.00	20.00	19.00	21.00	16.30
SPD	21.00	16.00	16.00	19.00	18.00	17.00	18.00	14.00	10.60

Table B.5: Forecast errors of the different methods for all parties and all states for selected lead times and the method `Lpv`. The rightmost column indicates the results on election day with lead time h=0.

State	Party	120	90	60	30	21	14	7	1	0
	AFD	-3.70	-3.70	-1.70	-2.70	-4.70	-4.70	-2.70	-2.39	9.70
	CDU	3.60	3.60	2.60	2.60	3.60	3.60	0.10	3.43	39.40
SN	FDP	0.20	0.20	-0.80	0.20	-0.80	-0.80	-0.30	-0.37	3.80
	GRE	0.30	0.30	0.30	1.30	1.30	1.30	0.80	0.99	5.70
	LIN	-0.90	-0.90	-0.90	2.10	1.10	1.10	0.60	-1.18	18.90
	SPD	3.60	3.60	2.60	0.60	1.60	1.60	2.10	3.04	12.40
	AFD	-7.20	-6.20	-6.20	-7.20	-6.20	-6.20	-3.70	-2.24	12.20
	CDU	1.00	5.00	5.00	2.00	0.00	4.00	1.50	1.75	23.00
BB	FDP	1.50	0.50	0.50	1.50	1.50	0.50	0.50	0.50	1.50
	GRE	-0.20	-0.20	-0.20	-1.20	-0.20	-1.20	-0.20	0.02	6.20
	LIN	6.40	4.40	4.40	3.40	3.40	2.40	2.90	-0.51	18.60
	SPD	0.10	-1.90	-1.90	2.10	2.10	1.10	0.10	2.89	31.90
	AFD	-6.60	-6.60	-6.60	-5.60	-5.60	-5.60	-3.10	-2.56	10.60
	CDU	2.50	2.50	2.50	0.50	0.50	0.50	1.50	4.68	33.50
TH	FDP	-0.50	-0.50	-0.50	1.50	1.50	1.50	1.50	1.50	2.50
	GRE	-0.70	-0.70	0.30	0.30	0.30	0.30	-0.20	0.51	5.70
	LIN	-0.20	-0.20	-1.20	-2.20	-2.20	-2.20	-1.20	-3.38	28.20
	SPD	6.60	6.60	6.60	6.60	6.60	6.60	3.60	3.77	12.40
	AFD	-2.60	-2.60	-2.60	-2.60	-0.60	-0.60	-2.10	-0.60	5.60
	CDU	5.60	5.60	5.60	5.60	0.60	2.60	0.10	0.60	22.40
HB	FDP	-1.60	-1.60	-1.60	-1.60	-1.60	-0.60	-1.10	0.40	6.60
	GRE	0.90	0.90	0.90	0.90	0.90	-3.10	0.40	-0.10	15.10
	LIN	-1.50	-1.50	-1.50	-1.50	-3.50	-0.50	-1.00	-0.50	9.50
	SPD	7.20	7.20	7.20	7.20	5.20	4.20	4.20	3.20	32.80
	AFD	-0.10	-2.10	-0.10	-0.60	-0.60	-0.10	-1.10	-1.10	6.10
	CDU	7.10	11.10	6.10	6.60	6.60	4.10	3.10	1.10	15.90
HH	FDP	-4.40	-5.40	-5.40	-3.40	-3.40	-2.40	-1.40	-1.40	7.40
	GRE	1.70	-1.30	1.70	1.70	1.70	0.70	-1.30	-0.30	12.30
	LIN	-0.50	-1.50	-0.50	-1.00	-1.00	0.50	1.50	0.50	8.50
	SPD	-6.60	-0.60	-3.60	-3.10	-3.10	-1.60	-0.60	1.40	45.60

Table B.5 – continued from previous page

State	Party	120	90	60	30	21	14	7	1	0
	AFD	-7.10	-7.10	-8.10	-4.60	-3.10	-6.10	-4.10	-4.10	15.10
	CDU	13.00	10.00	8.00	6.10	4.00	3.00	3.00	2.50	27.00
BW	FDP	-3.30	-3.30	-3.30	-3.20	-0.30	-1.80	-1.30	-1.30	8.30
	GRE	-6.30	-5.30	-2.30	-4.20	-2.30	0.20	1.70	1.70	30.30
	LIN	2.10	1.10	0.10	2.60	1.10	0.60	1.10	1.10	2.90
	SPD	3.30	5.30	6.30	2.90	1.30	3.80	0.30	0.30	12.70
	AFD	-6.60	-5.60	-5.60	-3.60	-3.60	-4.10	-2.60	-3.60	12.60
	CDU	9.20	7.20	7.20	5.20	5.20	3.20	3.20	3.20	31.80
RP	FDP	-2.20	-1.20	-1.20	-0.20	-0.20	-0.20	-0.20	-1.20	6.20
	GRE	2.70	3.70	3.70	2.70	2.70	4.70	0.70	1.70	5.30
	LIN	2.20	2.20	2.20	1.20	1.20	1.20	1.20	0.20	2.80
	SPD	-6.20	-5.20	-5.20	-5.20	-5.20	-3.70	-2.20	-1.20	36.20
	AFD	-19.20	-10.20	-10.20	-9.20	-7.20	-7.20	-7.20	-6.20	24.20
	CDU	4.20	5.20	5.20	3.20	2.20	-0.30	0.20	2.20	29.80
ST	FDP	-0.90	-1.90	-1.90	-1.90	-0.90	0.10	-0.90	-0.40	4.90
	GRE	1.80	0.80	0.80	-0.20	-0.20	-0.20	-0.20	-0.20	5.20
	LIN	9.70	6.70	6.70	2.70	3.70	3.70	2.70	4.70	16.30
	SPD	10.40	5.40	5.40	8.40	7.40	6.40	7.40	3.40	10.60
	AFD	-2.80	-2.80	-1.80	-1.80	-1.80	-1.80	0.20	1.20	20.80
	CDU	5.00	5.00	6.00	6.00	4.00	4.00	3.00	3.00	19.00
MV	FDP	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	3.00
	GRE	3.20	3.20	2.20	2.20	1.20	1.20	1.20	1.20	4.80
	LIN	2.80	2.80	3.80	3.80	5.80	2.80	-0.20	-0.20	13.20
	SPD	-8.60	-8.60	-8.60	-8.60	-6.60	-4.60	-2.60	-2.60	30.60
	AFD	0.80	0.80	-1.20	0.80	0.80	0.80	-0.20	-1.20	14.20
	CDU	1.40	0.40	2.40	2.40	2.40	-0.60	1.40	-0.60	17.60
WBE	FDP	-2.70	-2.70	-2.70	-1.70	-1.70	-1.70	-1.70	-1.70	6.70
	GRE	2.80	3.80	3.80	1.80	1.80	3.80	-0.20	1.80	15.20
	LIN	0.40	1.40	2.40	0.40	0.40	1.40	-1.60	-0.60	15.60
	SPD	1.40	1.40	-0.60	-0.60	-0.60	2.40	2.40	2.40	21.60
	AFD	2.90	2.90	2.90	2.90	2.90	-0.10	0.90	-0.10	6.10
	CDU	-3.70	-3.70	-2.70	-2.70	-2.70	-6.70	-3.70	-3.70	40.70
SL	FDP	-0.20	-0.20	0.80	0.80	0.80	0.80	0.80	0.80	3.20
	GRE	2.00	2.00	1.00	1.00	1.00	1.00	0.00	0.50	4.00
	LIN	2.20	2.20	1.20	1.20	1.20	0.20	-0.80	-0.30	12.80
	SPD	-3.60	-3.60	-3.60	-3.60	-3.60	3.40	2.40	2.40	29.60

Table B.5 – continued from previous page

State	Party	120	90	60	30	21	14	7	1	0
SH	AFD	0.10	0.10	0.10	1.10	1.10	-0.90	-0.90	-0.90	5.90
	CDU	2.00	2.00	2.00	-2.00	-2.00	-1.00	-1.00	-1.00	32.00
	FDP	-2.50	-2.50	-2.50	-2.50	-2.50	-2.50	-2.50	-2.50	11.50
	GRE	2.10	2.10	2.10	-0.90	-0.90	-0.90	-0.90	-0.90	12.90
	LIN	1.20	1.20	1.20	0.20	0.20	0.20	0.20	0.20	3.80
	SPD	-1.20	-1.20	-1.20	5.80	5.80	5.80	5.80	5.80	27.20
NW	AFD	1.60	1.60	-0.40	2.60	0.60	-0.40	-1.40	0.35	7.40
	CDU	-1.00	-1.00	-7.00	-5.00	1.00	-4.00	-1.00	-1.50	33.00
	FDP	-5.60	-4.60	-1.60	-2.60	-2.60	-0.60	-0.60	-1.35	12.60
	GRE	5.60	0.60	-0.40	-0.40	-0.40	-0.40	1.10	0.35	6.40
	LIN	0.10	0.10	0.10	0.10	0.10	1.10	1.10	2.10	4.90
	SPD	0.80	4.80	8.80	5.80	2.80	3.80	0.80	-0.70	31.20

Table B.6: RMSEs computed by averaging over all parties for all states, every method and selected lead times. The rightmost column indicates the average RMSE over 120 lead times per state and method.

State	Method	120	90	60	30	21	14	7	1	Avg.
	`med`	2.80	2.80	2.61	2.18	2.41	2.41	2.33	2.21	2.56
	`avg`	3.05	3.05	2.74	2.53	2.37	2.37	2.19	2.19	2.74
	`avg.wgt`	3.03	3.03	2.71	2.50	2.34	2.34	2.16	2.16	2.71
	`np.lc`	3.05	3.05	2.74	2.18	1.94	1.94	1.76	1.82	2.60
	`np.ll`	2.36	2.36	1.02	1.09	1.57	1.57	1.52	1.56	1.71
SN	`arima`	2.81	2.66	1.61	1.86	2.60	2.61	1.47	2.20	2.26
	`dlm`	2.61	2.60	1.73	1.85	2.23	2.62	1.46	2.05	2.22
	`dlm_lin`	2.55	2.60	1.73	1.85	2.89	2.62	1.51	2.67	2.27
	`iss`	2.60	2.60	1.73	1.85	2.61	2.62	1.46	2.21	2.23
	`Lpv`	2.60	2.60	1.73	1.85	2.62	2.62	1.46	2.21	2.23
	`avew`	2.46	2.25	1.41	1.83	2.70	2.62	1.48	2.21	2.02
	`med`	4.31	4.05	4.05	3.59	3.48	3.44	3.22	3.07	3.91
	`avg`	4.22	4.04	4.04	3.92	3.74	3.65	3.45	3.18	3.97
	`avg.wgt`	4.21	4.03	4.03	3.90	3.71	3.62	3.41	3.13	3.95
	`np.lc`	4.14	4.04	4.04	3.92	3.60	3.43	3.12	2.78	3.92
	`np.ll`	5.36	3.41	3.41	3.16	3.01	3.00	2.66	2.32	3.61
BB	`arima`	4.53	3.73	3.74	3.49	3.05	3.13	1.99	1.68	3.65
	`dlm`	4.00	3.80	3.80	3.55	3.22	3.23	2.07	1.61	3.64
	`dlm_lin`	4.00	3.80	3.80	3.55	2.98	3.31	1.84	2.03	3.65
	`iss`	4.00	3.80	3.80	3.55	3.07	3.24	2.03	1.68	3.62
	`Lpv`	4.00	3.80	3.80	3.55	3.07	3.24	2.03	1.68	3.62
	`avew`	3.98	4.01	3.90	3.52	3.03	3.23	2.06	1.65	3.70
	`med`	3.43	3.43	3.43	3.52	3.52	3.52	3.34	3.38	3.44
	`avg`	3.39	3.39	3.39	3.41	3.41	3.41	3.28	3.21	3.38
	`avg.wgt`	3.41	3.41	3.40	3.42	3.42	3.42	3.29	3.21	3.39
	`np.lc`	3.48	3.48	3.39	3.41	3.41	3.41	3.28	3.21	3.41
	`np.ll`	5.36	5.36	4.00	3.79	3.79	3.79	2.90	2.61	4.47
TH	`arima`	4.44	4.18	3.89	3.63	3.69	3.69	2.26	3.08	3.83
	`dlm`	3.93	3.96	3.54	3.70	3.70	3.70	2.25	2.92	3.74
	`dlm_lin`	4.10	3.96	4.41	3.70	3.70	3.70	1.91	3.52	3.76
	`iss`	3.96	3.96	3.98	3.70	3.70	3.70	2.18	3.07	3.73
	`Lpv`	3.96	3.96	3.98	3.70	3.70	3.70	2.18	3.07	3.73
	`avew`	4.38	4.08	3.97	3.85	3.81	3.78	2.40	3.05	3.87

Table B.6 - continued from previous page.

State	Method	120	90	60	30	21	14	7	1	Avg.
HB	med	4.31	4.31	4.31	4.31	3.17	2.53	2.22	1.84	4.00
	avg	4.31	4.31	4.31	4.31	3.17	2.76	2.52	2.27	4.03
	avg.wgt	4.31	4.21	4.12	4.10	3.67	2.68	2.45	2.20	3.78
	np.lc	4.31	4.23	4.11	4.10	3.69	2.26	2.52	2.27	3.78
	np.ll	4.31	4.00	4.00	4.21	3.13	8.85	2.00	1.41	5.50
	arima	5.39	4.91	4.53	4.23	2.74	2.31	2.02	1.36	4.37
	dlm	3.99	3.99	3.99	3.99	2.99	2.38	2.00	1.44	3.75
	dlm_lin	3.99	3.99	3.99	3.99	2.65	2.61	2.14	1.27	3.76
	iss	3.99	3.99	3.99	3.99	2.69	2.41	2.02	1.38	3.73
	Lpv	3.99	3.99	3.99	3.99	2.69	2.41	2.02	1.38	3.73
	avew	3.99	3.99	3.99	3.99	3.82	2.23	2.03	1.36	3.74
HH	med	4.41	4.51	4.10	3.93	3.93	3.70	3.20	3.14	4.17
	avg	4.41	4.51	4.20	4.03	4.03	3.70	3.05	2.86	4.20
	avg.wgt	4.41	4.54	4.20	4.02	4.02	3.67	3.00	2.81	4.16
	np.lc	4.41	4.58	4.20	4.03	4.03	3.69	2.98	2.04	4.17
	np.ll	4.41	5.34	3.78	3.88	3.88	3.28	1.18	1.55	4.09
	arima	4.82	5.64	3.73	3.34	3.35	2.00	1.59	1.18	4.33
	dlm	4.41	5.18	3.76	3.42	3.39	2.13	1.64	1.22	4.13
	dlm_lin	4.41	5.18	3.60	3.33	3.39	1.84	1.80	1.45	4.17
	iss	4.41	5.18	3.71	3.39	3.39	2.08	1.69	1.06	4.12
	Lpv	4.41	5.18	3.71	3.39	3.39	2.08	1.69	1.06	4.12
	avew	4.43	5.83	3.16	3.31	3.32	1.90	1.55	1.03	4.25
BW	med	7.48	6.97	6.79	5.75	5.35	4.63	4.33	4.04	6.32
	avg	7.39	7.02	6.82	5.62	5.39	4.89	4.51	4.08	6.32
	avg.wgt	7.39	7.00	6.79	5.54	5.30	4.80	4.41	3.98	6.28
	np.lc	7.39	6.97	6.67	4.00	3.57	3.03	2.41	2.10	5.65
	np.ll	7.61	6.22	4.30	2.67	2.17	2.08	1.73	1.71	4.39
	arima	6.20	5.19	5.05	3.73	2.02	3.14	2.27	2.19	4.31
	dlm	6.90	6.04	5.56	3.68	2.45	2.66	2.23	2.14	4.90
	dlm_lin	6.90	6.04	5.56	3.68	2.07	2.63	2.40	2.37	4.89
	iss	6.90	6.04	5.56	4.07	2.38	3.24	2.30	2.20	4.85
	Lpv	6.90	6.04	5.56	4.11	2.38	3.27	2.30	2.20	4.86
	avew	6.94	5.64	5.23	3.80	1.86	2.95	2.24	2.16	4.51

Table B.6 - continued from previous page.

State	Method	120	90	60	30	21	14	7	1	Avg.
RP	med	5.84	5.43	5.43	4.60	4.60	4.27	3.51	3.47	5.12
	avg	5.76	5.34	5.34	4.69	4.69	4.32	3.99	3.89	5.08
	avg.wgt	5.76	5.33	5.33	4.65	4.65	4.28	3.94	3.84	5.05
	np.lc	5.76	5.09	5.09	3.67	3.67	3.35	3.01	2.90	4.62
	np.ll	6.16	4.52	4.52	3.14	3.14	2.79	2.47	2.33	4.06
	arima	5.47	4.31	4.44	3.38	3.44	3.19	1.92	2.18	3.94
	dlm	5.53	4.66	4.66	3.47	3.55	3.12	2.08	1.88	4.22
	dlm_lin	5.53	4.64	4.66	3.66	3.55	3.23	1.76	2.21	4.22
	iss	5.53	4.66	4.66	3.55	3.55	3.27	1.99	2.20	4.19
	Lpv	5.53	4.66	4.66	3.55	3.55	3.27	1.99	2.20	4.19
	avew	5.43	4.38	4.41	3.34	3.39	3.11	1.94	1.88	3.90
ST	med	9.44	9.25	9.25	7.62	6.01	5.36	4.52	4.32	8.14
	avg	9.44	8.21	8.21	7.47	6.87	6.06	5.38	4.94	7.83
	avg.wgt	10.33	8.16	8.16	7.39	6.77	5.96	5.29	4.86	7.81
	np.lc	10.59	8.00	8.00	7.47	6.87	4.84	4.19	3.93	7.71
	np.ll	15.51	4.96	4.96	4.19	4.14	3.91	3.73	3.64	6.86
	arima	10.31	7.65	5.23	5.07	4.31	4.05	4.29	3.52	6.34
	dlm	9.94	5.91	5.91	5.42	4.60	4.16	4.12	3.34	6.46
	dlm_lin	9.94	5.91	5.91	5.42	4.51	4.16	4.57	3.71	6.41
	iss	9.94	5.91	5.91	5.42	4.58	4.22	4.37	3.59	6.39
	Lpv	9.94	5.91	5.91	5.42	4.58	4.22	4.37	3.59	6.38
	avew	10.35	5.15	5.29	5.02	4.23	4.04	4.25	3.31	6.26
MV	med	5.86	5.86	5.30	5.30	5.12	4.46	3.35	2.75	5.37
	avg	5.75	5.75	5.37	5.37	5.04	4.67	3.88	3.32	5.37
	avg.wgt	5.71	5.71	5.33	5.33	4.99	4.61	3.81	3.25	5.34
	np.lc	5.40	5.40	5.37	5.37	5.04	4.37	3.19	2.70	5.16
	np.ll	7.36	7.36	5.17	5.17	4.23	3.56	2.25	1.77	6.00
	arima	6.36	5.52	5.19	4.86	4.10	2.83	1.68	1.77	5.00
	dlm	4.58	4.58	4.70	4.70	4.10	2.93	1.74	1.72	4.37
	dlm_lin	4.58	4.58	4.70	4.70	3.91	2.68	1.59	1.87	4.36
	iss	4.58	4.58	4.70	4.70	4.04	2.88	1.70	1.76	4.35
	Lpv	4.58	4.58	4.70	4.70	4.04	2.88	1.70	1.76	4.35
	avew	5.92	4.78	4.58	4.59	4.15	2.19	1.32	1.72	4.51

Table B.6 - continued from previous page.

State	Method	120	90	60	30	21	14	7	1	Avg.
	med	4.47	4.10	3.85	3.60	3.60	3.36	3.26	2.65	3.87
	avg	4.09	3.81	3.47	3.21	3.21	3.11	2.89	2.74	3.57
	avg.wgt	4.03	3.75	3.40	3.13	3.13	3.02	2.79	2.64	3.49
	np.lc	2.32	2.30	2.21	2.07	2.07	2.00	1.60	1.49	2.31
	np.ll	2.01	2.14	2.08	1.84	1.84	1.84	0.91	0.97	1.96
WBE	arima	3.29	2.93	2.52	1.60	1.53	2.11	1.44	1.52	2.67
	dlm	1.82	2.10	2.42	1.43	1.47	2.08	1.36	1.02	2.56
	dlm_lin	1.82	2.11	2.42	1.71	1.47	2.10	1.90	1.79	2.77
	iss	1.82	2.94	2.42	1.63	1.58	2.08	1.49	1.53	2.73
	Lpv	1.82	2.10	2.42	1.47	1.47	2.08	1.49	1.53	2.61
	avew	2.55	2.19	2.79	1.73	1.47	1.98	1.32	1.02	2.55

Bibliography

Abramowitz, A. I. (1996): Bill and Als excellent adventure. American Politics Quarterly, 24 No. 4, 434–442.

Abramowitz, A. I. (2008): Forecasting the 2008 presidential election with the Time-for-Change Model. PS: Political Science & Politics 41 No. 4.

Aichholzer, J. and Willmann, J. (2014): Forecasting Austrian national elections: The Grand Coalition model. International Journal of Forecasting, 30 No. 1, 55–64.

Aisch, G. (2013): Wie tendenziös sind Wahlumfragen. ⟨URL: http://www.zeit.de/politik/deutschland/2013-09/wahlumfragen-parteil ichkeit-bundestagswahl(2013-09-22)⟩.

Andersen, R. and Heath, A. (2003): Social identities and political cleavages: The role of political context. Journal of the Royal Statistical Society Series A, 166 No. 3, 301–327.

Anderson, C. and Ward, D. (1996): Barometer elections in comparative perspective. Electoral Studies, 15 No. 4, 447–460.

Aneiros-Perez, G. and Cao, R. and Vilar-Fernandez, J. (2011): Functional methods for time series prediction: A nonparametric approach. Journal of Forecasting, 30 No. 4, 377–392.

Armstrong, J. and Graefe, A. (2011): Predicting elections from biographical information about candidates: A test of the index method. Journal of Business Research, 64 No. 7, 699–706.

Armstrong, S. and Collopy, F. (1992): Error measures for generalizing about forecasting methods: Empirical comparisons. International Journal of Forecasting, 8 No. 1, 69–80.

Armstrong, J. and Green, K. and Graefe, A. (2014): Golden Rule of forecasting: Be conservative., Working Paper.

Auberger, A. (2005): Forecasts of the 2004 French European election. Swiss Political Science Review, 11 No. 3, 61–78.

Auberger, A. (2012): Voting and economic factors in French elections for the European Parliament. Public Choice, 153 No. 3, 329–340.

Bafumi, J. and Erikson, R. and Wlezien, C. (2010): Forecasting house seats from generic congressional polls: The 2010 midterm election. Political Science and Politics, 43 No. 4, 633–636.

Bélanger, E. and Lewis-Beck, M. S. and Nadeau, R. (2005): A Political Economy Forecast for the 2005 British General Election. The British Journal of Politics and International Relations, 7 No. 2, 191–198.

Bélanger, E. and Nadeau, R. and Lewis-Beck, M. S. (2010): Forecasting the Vote for a Third Party: The British Liberals, 1974-2005. The British Journal of Politics and International Relations, 12 No. 4, 634–643.

Bélanger, E. and Soroka, S. (2012): Campaigns and the prediction of election outcomes: Can historical and campaign-period prediction models be combined? Electoral Studies, 31 No. 4, 702–714.

Bellucci, P. (2010): Election cycles and electoral forecasting in Italy, 1994-2008. International Journal of Forecasting, 26 No. 1, 54–67.

Berg, J. and Nelson, F. and Rietz, T. (2003): Accuracy and forecast standard error of prediction markets., Working Paper.

Berlemann, M. and Enkelmann, S. and Kuhlenkasper, T. (2014): Unraveling the relationship between presidential approval and the economy: A multidimensional semiparametric approach. Journal of Applied Econometrics, 30 No. 3, 468–486.

Blumenthal, M. (2014): Polls, forecasts, and aggregators. PS: Political Science & Politics, 47 No. 2, 297–300.

Blumör, F. and Hübner, E. and Maichel, A. (2011): Die politische Ordnung in Deutschland. 13th edition. Bayerische Landeszentrale für politische Bildung.

Boon, M. (2012): Predicting elections: A 'Wisdom of Crowds' approach. International Journal of Market Research, 54 No. 4, 465–483.

Brooker, R. and Schaefer, T. (2017): Methods of measuring public opinion..

Brooks, C. and Dodson, K. and Hotchkiss, N. (2010): National security issues and US presidential elections, 1992-2008. Social Science Research, 39 No. 4, 518–526.

Brown, L. and Chappell, H. Jr. (1999): Forecasting presidential elections using history and polls. International Journal of Forecasting, 15 No. 2, 127–135.

Brown, P. J. and Firth, D. and Payne, C. D. (1999): Forecasting on British election night 1997. Journal of the Royal Statistical Society Series A, 162 No. 2, 211–226.

Bunker, K. and Bauchowitz, S. (2015): Electoral Forecasting and Public Opinion Tracking in Latin America: An Application to Chile., Working Paper.

Callegaro, M. and Gasperoni, G. (2008): Accuracy of pre-election polls for the 2006 Italian parliamentary election: Too close to call. International Journal of Public Opinion Research, 20 No. 2, 148–170.

Campbell, J. E. (1996): Polls and votes. The trial-heat presidential election forecasting model, certainty, and political campaigns. American Politics Quarterly, 24 No. 4, 408–433.

Campbell, J. E. (2008): Editor's introduction: Forecasting the 2008 national elections. PS: Political Science & Politics, 41 No. 4, 679–682.

Campbell, A. and Converse, P. E. and Miller, W. E. and Stokes, D. E. (1960): The American voter. New York: John Wiley and Sons.

Castro, G. (2013): Polling and acccuracy measure: Evidence from the Italian case., Working Paper.

Christensen, W. and Florence, L. (2008): Predicting presidential and other multistage election outcomes using state-level pre-election polls. The American Statistican, 62 No. 1, 1–10.

Coletto, D. and Breguet, B. (2015): The accuracy of public polls in provincial elections. Canadian Political Science Review, 9 No. 1, 41–54.

Cuzan, A. (2012): Forecasting the 2012 presidential election with the fiscal model., Working Paper.

Cuzan, A. and Armstrong, S. and Jones, R. (2005): Combining methods to forecasting the 2004 presidential election to pollyvote., Working Paper.

Dalton, R. J. (2013): Citizen politics. Public opinion and political parties in advanced industrial democracies. SAGE, ISBN 978–1–4522–0300–3.

De Gooijer, J.G. and Hyndman, R. (2006): 25 years of time series forecasting. International Journal of Forecasting, 22 No. 3, 443–473.

Degan, A. and Merlo, A. (2011): A structural model of turnout and voting in multiple elections. Journal of the European Economic Association, 9 No. 2, 209–245.

Diebold, F. (2007): Elements of forecasting. Thomson, ISBN 978–0–324–32359–7.

Dolado, J. and Gonzalo, J. and Mayoral, L. (2003): Long-Range dependence in Spanish political opinion poll series. Journal of Applied Econometrics, 18 No. 2, 137–155.

Dopp, K. (2012): Vote miscount or poll response bias? Estimating the causes of the discrepancies between polls and reported vote shares., Working Paper.

Downs, A. (1957): An Economic Theory of Political Action in a Democracy. Journal of Political Economy, 65 No. 2, 135–150.

Duquette, C. M. and Mixon, F. G. and Cebula, R. J. (2013): The impact of legislative tenure and seniority on general election success: Econometric evidence from U.S. house races. Atlantic Economic Journal, 41 No. 2, 161–172.

Erikson, R. and Wlezien, C. (2008): Are political markets really superior to polls as election predictors? Public Opinion Quarterly, 72 No. 2, 190–215.

Falter, J.; Falter, J. and Schoen, H., editor (2014): Handbuch Wahlforschung. 2nd edition. Wiesbaden: Springer VS, ISBN 978–3–658–05163–1.

Fan, J. (1992): Design-adaptive Nonparametric Regression. Journal of the American Statistical Association, 87 No. 420, 998–1004.

Gaissmaier, W. and Marewski, J. (2011): Forecasting elections with mere recognition from small, lousy samples: A comparison of collective recognition, wisdom of crowds, and representative polls. Judgement and Decision Making, 6 No. 1, 73–88.

Ganser, C. and Riordan, P. (2015): Vote expectations at the next level. Trying to predict vote shares in the 2013 German federal election by polling expectations. Electoral Studies, 40, 115–126.

Gelman, A. and King, G. (1993): Why are american presidential election campaign polls so variable when votes are so predictable? British Journal of Political Science, 23 No. 4, 409–451.

Gelman, A. and King, G. and Boscardin, W. (1998): Estimating the probability of events that have never occured: When is your vote decisive? Journal of the American Statistical Society, 93 No. 441, 1–9.

Gibson, R. and Lewis-Beck, M. S. (2011): Methodologies of election forecasting: Calling the 2010 UK "hung parliament". Electoral Studies, 30 No. 2, 247–249.

Gordon, S. (2010): Using expert assessments to forecast the 2010 House election., Working Paper.

Gott, R. and Colley, W. (2008): Median statistics in polling. Mathematical and Computer Modelling, 48 No. 9-10, 1396–1408.

Graefe, A. (2013a): Improving forecasts using equally weighted predictors., Working Paper.

Graefe, A. (2013b): Issue and leader voting in U.S. presidential elections., Working Paper.

Graefe, A. (2014): Accuracy of vote expectation surveys in forecasting elections., Working Paper.

Graefe, A. (2015): German election forecasting: Comparing and combining methods for 2013. German Politics, 24 No. 2, 195–204.

Graefe, A. (2016): Forecasting proportional representation elecctions from non-representative expectation surveys. Electoral Studies, 42, 222–228.

Graefe, A. and Armstrong, J. S. (2011): Who should be nominated to run in the 2012 presidential election? Long-term forecasts based on candidates' biographies., Working Paper.

Graefe, A. and Armstrong, J. S. (2012): Predicting elections from the most important issue: A test of the take-the-best heuristic. Journal of Behavioral Decision Making, 25 No. 1, 41–48.

Graefe, A. and Armstrong, J. S. (2013): Forecasting elections from voters' perceptions of candidates' ability to handle issues. Journal of Behavioral Decision Making, 26 No. 3, 295–303.

Graefe, A. and Armstrong, J. S. (2014): Accuracy of combined forecasts for the 2012 presidential election: The PollyVote., Working Paper.

Gschwend, T. and Norpoth, H. (2005): Prognosemodell auf dem Prüfstand: Die Bundestagswahl 2005. Politische Vierteljahresschrift, 46 No. 4, 682–688.

Hanretty, C. (2013): The 2013 Italian election: A poll-based forecast., Working Paper.

Haupt, H. and Schnurbus, J. and Huber, S. (2017): Dynamic poll-based forecasts for German regional elections., Working Paper.

Heij, C. and Franses, P. H. (2011): Correcting for survey effects in pre-election polls. Statistica Neerlandica, 65 No. 3, 352–370.

Hillygus, D.S. (2011): The evolution of election polling in the United States. Public Opinion Quarterly, 75 No. 5, 962–981.

Hobolt, S. B. and Spoon, J.-J. (2012): Motivating the European voter: Parties, issues and campaigns in European Parliament elections. European Journal of Political Research, 51 No. 6, 701–727.

Holbrook, T. and DeSart, J. (1999): Using state polls to forecast presidential election outcomes in the American states. International Journal of Forecasting, 15 No. 2, 137–142.

Hummel, P. and Rothschild, D. (2014): Fundamental models for forecasting elections at the state level. Electoral Studies, 35, 123–139.

Hurvich, C. M. and Simonoff, J. S. and Tsai, Chih-Ling (1998): Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. Journal of the Royal Statistical Society Series B, 60 No. 2, 271 – 293.

Hyndman, R. and Khandakar, Y. (2008): Automatic time series forecasting: The forecast package for R. Journal of Statistical Software 27 No. 3.

Hyndman, R. and Koehler, A. (2006): Another look at measures of forecast accuracy. International Journal of Forecasting, 22 No. 4, 679–688.

Hyndman, R.J. and Koehler, A.B. and Snyder, R.D. and Grose, S. (2002): A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting, 18 No. 3, 439–454.

Hyndman, R. and Athanasopoulos, G. (2013): Forecasting Principles and Practice. OTexts ⟨URL: http://otexts.org/fpp/⟩.

Jackman, S. (2005): Pooling the polls over an election campaign. Australian Journal of Political Science, 40 No. 4, 499–517.

Jérôme, B. (2013): Editor's introduction for forecasting the 2013 German elections. PS: Political Science & Politics, 46 No. 3, 477–478.

Jérôme, B. and Jérôme, V. and Lewis-Beck, M. S. (1999): Polls fail in France: forecasts of the 1997 legislative election. International Journal of Forecasting, 15 No. 2, 163–174.

Jérôme, B. and Jérôme-Speziari, V. and Lewis-Beck, M. S. (2013): A political-economy forecast for the 2013 German elections: Who to rule with Angela Merkel? PS: Political Science & Politics, 46 No. 3, 479–480.

Jérôme, B. and Jérôme-Speziari, V. (2011): Forecasting the 2012 US Presidential election., Working Paper.

Jessen, A. (2014): Was steckt hinter den Zahlen? Methoden der Demoskopie. ⟨URL: http://www.bpb.de/apuz/192965/was-steckt-hinter-den-zahlen-methoden-der-demoskopie?p=all⟩.

Jiao, Y. and Syau, Y. and Lee, E. S. (2006): Fuzzy adaptive network in presidential elections. Mathematical and Computer Modelling, 43 No. 3-4, 244–253.

Jungmittag, A. (2016): Combination of forecasts across estimation windows: An application to air travel demand. Journal of Forecasting, 35, 373–380.

Kamakura, A. and Mazzon, J. and Bruyn, A. de (2006): Modeling voter choice to predict the final outcome of two-stage elections. International Journal of Forecasting, 22 No. 4, 689–706.

Kasprzyk, D. and Moriarity, C. and McMillen, M. and Harris-Kojetin, B. (2001): Measuring and reporting sources of error in surveys. ⟨URL: https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/04/spwp31.pdf⟩.

Kayser, M. and Leininger, A. (2013): A benchmarking forecast of the 2013 Bundestag Election., Working Paper.

Kayser, M. and Leininger, A. (2017): A Länder-based forecast of the German Bundestag Election. Political Science and Politics, 50 No. 3, 689–692.

Khemani, S. (2001): Decentralization and accountability: Are voters more vigilant in local than in national elections?, Working Paper.

Klarner, C. (2008): Forecasting the 2008 U.S. house, senate and presidential elections at the district and state level. PS: Political Science & Politics, 41 No. 4, 723–728.

Klein, M. (2005): Die "Zauberformel". Über das erfolgreiche Scheitern des Prognosemodells von Gschwend und Norpoth bei der Bundestagswahl 2005. Politische Vierteljahresschrift, 46 No. 4, 689–691.

Korte, K.-R. (2009a): Das personalisierte Verhältniswahlrecht. ⟨URL: http://www.bpb.de/politik/wahlen/bundestagswahlen/62517/das-wahlsystem⟩.

Korte, K.-R. (2009b): Theorien des Wählerverhaltens. ⟨URL: http://www.bpb.de/politik/wahlen/bundestagswahlen/62613/theorien-des-waehlerverhaltens?p=all⟩.

Kruke, A. (2014): Fragen über Fragen: Zur Geschichte der politischen Umfrage. ⟨URL: http://www.bpb.de/apuz/192960/fragen-ueber-fragen-zur-geschichte-der-politischen-umfrage?p=all⟩.

Krumpal, I. and Rauhut, H. (2008): Zum Ausmaß der bundespolitischen Parteipolitikverflechtung bei Landtagswahlen in Deutschland 1996 bis 2000. Zeitschrift für Parlamentsfragen, 39 No. 2, 249–269.

Kulke, R. (2010): Die Wahlverfahren in den Ländern Europas. ⟨URL: http://www.europa-links.eu/fakten/die-wahlverfahren-in-den-landern-europas-199/⟩.

Küntzler, T. (2014): Using bayesian data combination of fundamental variable-based forecasts and poll-based forecasts to predict the 2013 German election., Working Paper.

Lazarsfeld, P. and Berelson, B. and Gaudet, H. (1944): The People's Choice. How the voter makes up his mind in a presidential campaign. Duell, Sloan and Pearce. New York.

Lebo, M. and Norpoth, H. (2011): Yes, Prime Minister: The key to forecasting British elections. Electoral Studies, 30 No. 2, 258–263.

Leigh, A. and Wolfers, J. (2006): Competing approaches to forecasting elections: Economic models, opinion polling and prediction markets., Working Paper.

Lemennicier, B. and Lescieux-Katir, H. and Grofman, B. (2010): The 2007 French Presidential election. Canadian Journal of Political Science, 43 No. 1, 137–161.

Lewis-Beck, M. S. (2005): Election forecasting: Principles and practice. The British Journal of Politics & International Relations, 7 No. 2, 145–164.

Lewis-Beck, M. S. and Bélanger, E. and Fauvelle-Aymar, C. (2008): Forecasting the 2007 French Presidential election: Ségolène Royal and the Iowa Model. French Politics, 6 No. 2, 106–115.

Lewis-Beck, M. S. and Dassonneville, R. (2015): Forecasting elections in Europe: Synthetic models. Research & Politics, 2 No. 1, 1–11.

Lewis-Beck, M. S. and Nadeau, R. and Bélanger, E. (2004): General election forecasts in the United Kingdom: A political economy model. Electoral Studies, 23 No. 2, 279–290.

Lewis-Beck, M. S. and Rice, T. (1984): Forecasting presidential elections: A comparison of naive models. Political Behavior, 6 No. 1, 9–21.

Lewis-Beck, M. S. and Skalaban, A. (1989): Citizen forecasting: Can voters See into the future? British Journal of Political Science, 19 No. 1, 146–153.

Lewis-Beck, M. S. and Tien, C. (1999): Voters as forecasters: A micromodel of election prediction. International Journal of Forecasting, 15 No. 2, 175–184.

Lewis-Beck, M. S. and Tien, C. (2012): Election forecasting for turbulent times. PS: Political Science & Politics, 45 No. 4, 625–629.

Lewis-Beck, M.S. and Stegmaier, M. (2000): Economic determinants of electoral outcomes. Annual Review Political Science, 3 No. 1, 183–219.

Lewis-Beck, M.S. and Tien, C. (1996): The future in forecasting: Prospective presidential models. American Politics Quarterly, 24 No. 4, 468–491.

Lewis-Beck, M.S. and Tien, C. (2011): Election forecasting theory, practice, Japan., Working Paper.

Li, Q. and Racine, J. (2004): Cross-validated local linear nonparametric regression. Statistica Sinica, 14 No. 2, 485–512.

Linzer, D. (2013): Dynamic bayesian forecasting of Presidential elections in the states. Journal of the American Statistical Association, 109 No. 501, 124–134.

Lock, K. and Gelman, A. (2010): Bayesian combination of state polls and election forecasts., Working Paper.

Lockerbie, B. (2008): Election forecasting: The future of the presidency and the house. PS: Political Science & Politics, 41 No. 4, 713–716.

Magalhaes, P. and Aguiar-Conraria, L. and Lewis-Beck, M. (2012): Forecasting Spanish elections. International Journal of Forecasting, 28 No. 4, 769–776.

Makridakis, S. and Hibon, M. (2000): The M3-Competition: Results, conclusions and implications. International Journal of Forecasting, 16 No. 4, 451–476.

Malisch, R. (2013): Hat Deutschland die Wahl? Smart Investor, 9, 16–21.

Martin, E. and Traugott, M. and Kennedy, C. (2005): A review and proposal for a new measure of poll accuracy. Public Opinion Quarterly, 69 No. 3, 342–369.

McDermott, M. and Frankovic, K. (2003): The polls - review horserace polling and survey method effects: An analysis of the 2000 campaign. Public Opinion Quarterly, 67 No. 2, 244–264.

Mellon, J. and Prosser, C. (2016): Correcting Polling Error in Over-Engaged Samples Using Representative Turnout Weighting..

Miller, M. and Wang, G. and Kulkari, S. and Poor, V. (2012): Citizen forecasts of the 2008 U.S. presidential election. Politics and Policy, 40 No. 6, 1019–1052.

Montgomery, J. and Hollenbach, F. and Ward, M. (2012): Improving predictions using Ensemble Bayesian Model Averaging. Political Analysis, 20, 271–291.

Morlok, M. and Poguntke, T. and Zons, G. ; Tsatsos, D. and von Alemann, U. and Morlok, M. and Poguntke, T. and Zons, G. , editor (2016): Etablierungschancen neuer Parteien. Nomos, 51, ISBN 978–3–8487–3216–6.

Mughan, A. (1987): General election forecasting in Britain: A comparison of three simple models. Electoral Studies, 6 No. 3, 195–207.

Münch, Ursula and Oberreuter, Heinrich (2015): Die neue Offenheit: Wahlverhalten und Regierungsoptionen im Kontext der Bundestagswahl 2013. 1st edition. Frankfurt am Main: Campus, ISBN 3593501619.

Murr, A. E. (2011): "Wisdom of crowds"? A decentralised election forecasting model that uses citizens' local expectations. Electoral Studies, 30 No. 4, 771–783.

Nadeau, R. and Lewis-Beck, M. S. (2001): National economic voting in U.S. presidential elections. The Journal of Politics, 63 No. 1, 159–181.

Nadeau, R. and Lewis-Beck, M. S. and Bélanger, E. (2010): Electoral forecasting in France: A multi-equation solution. International Journal of Forecasting, 26 No. 1, 11–18.

Niggemeier, S. (2013): Verwählt. ⟨URL: http://www.spiegel.de/spiegel/print/d-91056800.html(2013-02-18)⟩.

Norpoth, H. (2004): From primary to general election: A forecast of the presidential vote. PS: Political Science & Politics, 37 No. 4, 737–740.

Norpoth, H. and Gschwend, T. (2003): Politbarometer und Wahlprognosen: Die Kanzlerfrage. In **Wüst, A., editor:** Politbarometer. Opladen: Leske and Budrich, ISBN 3–8100–4051–7, 109–123.

Norpoth, H. and Gschwend, T. (2010): The chancellor model: Forecasting German elections. International Journal of Forecasting, 26 No. 1, 42–53.

Oberreuter, H. (2012): Do elections matter? Zeitschrift für Politik, 59 No. 2, 169–184.

Oberreuter, H. (2015): Sekundär, aber unersetzbar. Die Politische Meinung, 60 No. 530, 30–34.

Ord, J. K. and Koehler, A. B. and Snyder, R. D. (1997): Estimation and prediction for a class of dynamic nonlinear statistical models. Journal of the American Statistical Association, 92 No. 440, 1621–1629.

Pesaran, H. and Pick, A. and Pranovich, M. (2013): Optimal forecasts in the presence of structural breaks. Journal of Econometrics, 177 No. 2, 134–152.

Pesaran, H. and Timmermann, A. (2007): Selection of estimation window in the presence of structural breaks. Journal of Econometrics, 137, 134–161.

Petris, G. (2009): Bayesian and Likelihood Analysis of Dynamic Linear Models. ⟨URL: https://cran.r-project.org/web/packages/dlm/vignettes/dlm.pdf⟩.

Petris, G. (2010): An R package for dynamic linear models. Journal of Statistical Software, 36 No. 12 ⟨URL: http://www.jstatsoft.org/v36/i12/⟩.

Petris, G. and Petrone, S. (2011): State Space models in R. Journal of Statistical Software, 41 No. 4 ⟨URL: http://www.jstatsoft.org/⟩.

Petris, G. and Petrone, S. and Campagnoli, P. (2009): Dynamic linear models with R. 1st edition. Springer.

Plickert, P. (2013): Das schwierige Geschäft mit den Umfragen. ⟨URL: http://www.faz.net/aktuell/wirtschaft/meinungsforschung-das-schwierige-geschaeft-mit-den-umfragen-12541293.html(2013-08-21⟩.

Plischke, T. (2014): Fällt die Wahlentscheidung immer später? Die Entwicklung des Zeitpunkts der Wahlentscheidung bei den Bundestagswahlen 1969 bis 2009. Politische Vierteljahresschrift, 55 No. 1, 118–144.

Probst, L. (2014): Die Fünfprozenthürde im deutschen Wahlsystem. ⟨URL: http://www.bpb.de/politik/wahlen/bundestagswahlen/175680/die-fuenfprozenthuerde⟩.

R Core Team (2015): R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2015 ⟨URL: https://www.R-project.org/⟩.

Racine, J. (2008): Nonparametric Econometrics: A Primer. Foundations and Trends in Econometrics, 3 No. 1, 1–88.

Racine, J. and Hayfield, T. (2017): Nonparametric kernel smoothing methods for mixed data types. ⟨URL: https://github.com/JeffreyRacine/R-Package-np⟩.

Racine, J. and Li, Q. (2004): Nonparametric estimation of regression functions with both categorial and continuous data. Journal of Econometrics, 119 No. 1, 99–130.

Rathore, A. (2012): Get complete information on the proportional representation (PR). ⟨URL: http://www.preservearticles.com/2012031026112/get-complete-information-on-the-proportional-representation-pr.html⟩.

Reynolds, A. and Reilly, B. and Ellis, A. (2005): Electoral system design: The new international IDEA Handbook. International Institute for Democracy and Electoral Assistance, ISBN 91–85391–18–2.

Rigdon, S. and Sauppe, J. and Jacobsen, S. (2015): Forecasting the 2012 and 2014 elections using Bayesian prediction and optimization. SAGE Open, 5 No. 2, 1–16.

Rothschild, D. (2009): Forecasting elections: Comparing prediction markets, polls and their biases. Public Opinion Quarterly, 73 No. 5, 895–916.

Sanders, D. (1995): Forecasting political preferences and election outcomes in the UK: experiences, problems and prospects for the next general election. Electoral Studies, 14 No. 3, 251–272.

Schäfer, A. (2013): Wahlbeteiligung und Nichtwähler. Aus Politik und Zeitgeschichte, 48-49, 39–46.

Schaffer, L.-M. and Schneider, G. (2005): Die Prognosegüte von Wahlbörsen und Meinungsumfragen zur Bundestagswahl 2005. Politische Vierteljahresschrift, 46 No. 4, 674–681.

Schnell, R. and Noack, M. (2014): The Accuracy of pre-election polling of German general elections. Methods, Data, Analyses, 8 No. 1, 5–24.

Selb, P. and Herrmann, M. and Munzert, S. and Schübel, T. and Shikano, S. (2013): Forecasting runoff elections using candidate evaluations from first round exit polls. International Journal of Forecasting, 29 No. 4, 541–547.

Selb, P. and Munzert, S. (2016): Forecasting the 2013 German Bundestag election using many polls and historical election results. German Politics, 25 No. 1, 73–83.

Sjöberg, L. (2009): Are all crowds equally wise? A comparison of political election forecasts by experts and the public. Journal of Forecasting, 28 No. 1, 1–18.

Stegmaier, M. and Lewis-Beck, M. S. (2009): Learning the economic vote the Hungarian forecasts 1998-2010. Politics and Policy, 37 No. 4, 769–780.

Steinmayr, A. (2016): Exposure to refugees and voting for the far-right (unexpected) Results from Austria., Working Paper.

Strauss, A. (2007): Florida or Ohio? Forecasting presidential state outcomes using reverse random walks. ⟨URL: http://www.mindlessphilosopher.net/princeton/strauss_reverse%20random%20walk.pdf⟩.

Tashman, L. (2000): Out-of-sample tests of forecasting accuracy: An analysis and review. International Journal of Forecasting, 16 No. 4, 437–450.

Traugott, M. (2014): Public opinion polls and election forecasting. PS: Political Science & Politics, 47 No. 2, 342–345.

Walther, D. (2015): Picking the winner(s): Forecasting elections in multiparty systems. Electoral Studies, 40, 1–13.

Wang, W. and Rothschild, D. and Goel, S. and Gelman, A. (2015): Forecasting elections with non representative polls. International Journal of Forecasting, 31 No. 3, 980–991.

West, M. and Harrison, J. (1997): Bayesian forecasting and dynamic models. 2nd edition. New York: Springer, ISBN 0–387–94725–6.

Whiteley, P. (2005): Forecasting seats from votes in British general elections. The British Journal of Politics and International Relations, 7 No. 2, 165–173.

Whiteley, P. (2008): Evaluating rival forecasting models of the 2005 general election in Britain - An encompassing experiment. Electoral Studies, 27 No. 4, 581–588.

Whiteley, P. and Sanders, D. and Stewart, M. and Clarke, H. (2011): Aggregate level forecasting of the 2010 general election in Britain: The Seats-Votes model. Electoral Studies, 30 No. 2, 278–283.

Williams, E. A. and Pillai, R. and Deptula, B. and Lowe, K. B. (2012): The effects of crisis, cynicism about change, and value congruence on perceptions of authentic leadership and attributed charisma in the 2008 presidential election. The Leadership Quarterly, 23 No. 3, 324–341.

Wlezien, C. and Erikson, R. (2004): The fundamentals, the polls, and the presidential vote. PS: Political Science & Politics, 37 No. 4, 747–751.

Wlezien, C. and Erikson, R. (2007): The horse race: What polls reveal as the election campaign unfolds. International Journal of Public Opinion Research, 19 No. 1, 74–88.

Wlezien, C. and Jennings, W. and Fisher, S. and Ford, R. and Pickup, M. (2013): Polls and the Vote in Britain. Political Studies, 61 No. 51, 66–91.

Wolfers, J. and Leigh, A. (2001): Three tools for forecasting federal elections: Lessons from 2001., Working Paper.

Wolfers, J. and Zitzewitz, E. (2006): Prediction markets in theory and practice., Working Paper.

Wright, M. and Farrar, D. and Russel, D. (2014): Polling accuracy in a multiparty election. International Journal of Public Opinion Research, 26 No. 1, 113–124.