Does a district-vote matter for the behavior of politicians? A textual analysis of parliamentary speeches.[*]

Andreas Born[†] Aljoscha Janssen[‡]

May 6, 2019

Abstract

In most democracies members of parliament are either elected over a party list or by a district. We use a discontinuity in the German parliamentary system to investigate the causal effect of a district-election on an MP's conformity with her party-line. A district-election does not affect roll call voting behavior causally, possibly due to overall high adherence to party voting. Analyzing the parliamentary speeches of each MP allows us to overcome the high party discipline with regard to parliamentary voting. Using textual analysis and machine learning techniques, we create two measures of closeness of an MP's speeches to her party. We find that district-elected members of parliament do not differ, in terms of speeches, from those of their party-peers who have been elected through closed party lists. However, both speeches and voting correlate with district characteristics suggesting that district-elections allow districts to select more similar politicians.

Keywords: Party-line; Textual Analysis; Regression Discontinuity; Parliamentary Speeches; Voting

[†]Stockholm School of Economics, andreas.born@phdstudent.hhs.se, Stockholm School of Economics, P.O Box 6501, SE-113 83 Stockholm

[‡]Stockholm School of Economics, aljoscha.janssen@phdstudent.hhs.se, Stockholm School of Economics, P.O Box 6501, SE-113 83 Stockholm

1 Introduction

Plurality voting and proportional representation are the two most common electoral systems through which western democracies determine their members of parliament. The choice of system is relevant not only for political process but influences economic outcomes of the respective country (Lizzeri and Persico, 2001; Persson, Roland, and Tabellini, 1997). For example research suggests that public spending, rent seeking, size of the government, and redistribution vary to a large degree with the electoral system (Milesi-Ferretti, Perotti, and Rostagno, 2002; Mueller, 1997; Persson and Tabellini, 1999, 2004). In particular, previous literature predicts that district elected members of the parliament (MPs) are acting in favor of their district whereas closed party list elected MPs are more aligned to the general interest of a party (Carey, 2007; Depauw and Martin, 2009).[1]

In a mixed member proportional system MPs who are elected through a party list coexists with MPs who are elected via a district vote. In this paper, we make use of a discontinuity in the German parliamentary system to analyze if MPs in the German parliament differ in the extend to which they deviate from their party's line in their roll call voting and parliamentary speeches depending on the way they entered parliament. We use text analysis and machine learning to investigate if parliamentary speeches of district and list elected MPs differ systematically. These methods allow us to assess speeches in a reproducible and objective way that does not rely human coding of language. We base this analysis on two criteria. The type of words and word-combinations used in the speeches and the similarity of the language in the speeches to that of the respective party manifesto.

Our analysis makes use of the fact that most German politicians run both in a district and on a party list at the same time (dual candidacy strategy), however enter parliament only over one of these. This allows us to employ a regression discontinuity design to quantify the causal effect of a district election on the conformity to the party-line in roll-call elections as well as parliamentary speeches. To our knowledge the article is the first that (1) analyze speeches of German in MPs incorporating recent information retrieval methods, (2) use machine learning methods to predict party affiliations from parliamentary speeches and (3) use a regression discontinuity approach to evaluate politicians behavior in terms of roll-call votings and speeches.

Using data from three legislative periods (2005-2017), we show that district and list elected candidates do not differ significantly at the margin in terms of voting. Thus the data does not support the hypothesis that winning a district mandate causes MPs to deviate more or less from their party's voting. Yet, data on voting is limited, the majority of votes are no roll-call votes, and additionally party discipline is very high.[2]

---

[1]The choice of political system itself is of course not random and might reflect underlying voter preferences or reflect previous institutions that have been in place in a country see for example Alesina, Glaeser, and Sacerdote (2001) and Acemoglu, Johnson, and Robinson (2001).

[2]The average share of MPs voting with their party is around 90%, even if one counts an abstention as a deviation if the party votes positive or negatively.

Hence, voting might not express an MP's actual actions behind the scenes. To address this concern we analyze parliamentary speeches. Here too we find that MPs who won a district marginally and would otherwise have been elected over a party list do not differ in their parliamentary speeches compared to MPs which marginally lost a district and enter parliament over a party list. Neither do they use a wording closer or further away from the party manifestos. In conclusion we do not find evidence that winning a district has a causal effect on deviations from the party-line either in terms of roll-call votes or in terms of speeches further away from the party average or party manifestos. This result is robust to several specifications of the discontinuity regression.

Deviations in voting and speeches do however correlate with economic and social characteristics of the district the politician has been elected in. Thus speeches do differ systematically which suggests that they are informative about the policy position of the according member of parliament. Especially the share of unemployed and inhabitants without a secondary degree correlate highly positively with roll call vote deviations and negatively with the similarity of speeches to party average and party manifestos. We conclude that there is no causal effect of a district election on voting or speeches but different districts elect politicians who are different to begin with.

A possible alternative explanation for the non-result and confound to our analysis could be that district elected MPs experience faster career progression (Folke, Persson, and Rickne, 2016; Meriläinen and Tukiainen, 2018) within a party which incentives conformity with the party-line. This might off-set the effect of a district mandate at the margin.[3] We explore this relationship by investigating if district elected MPs have a higher probability to increase their positioning in the closed party lists over election periods. Exploiting the same discontinuity, we find that district elected MPs do not experience better list positions than list elected MPs. The finding suggests that career progression due to winning a district is not confounding the results.

We rationalize the results in a stylized framework based on electoral competition in the flavor of Downs (1957) between the major two parties (here conservatives and social democrats).[4] The intuition is that candidates who follow a dual candidacy strategy face two objectives. On the one hand, politicians face distribution of voters in their district. On the other hand, national parties decide about the party-line based on accumulated distributions of voters across all districts.[5] A politician in a district has to decide between focusing on the median voter in their district which may increase the probability of election

---

[3]The question of rank effects has also been discussed in political science, see (Crisp, Olivella, Malecki, and Sher, 2013), Anagol and Fujiwara (2016), or André, Depauw, Shugart, and Chytilek (2017).

[4]For simplicity, our framework relies on the median voter theorem. Previous literature has shown that this is an abstraction of reality, for example studies on probabilistic voting models can explain policy divergence between parties, i.e. Lindbeck and Weibull (1987) or Lindbeck and Weibull (1993). Aragones and Palfrey (2002) and Carrillo and Castanheira (2008) provide alternative rationalizations for deviations from median voter policies. Gerber and Lewis (2004) provide an overview over this literature. Our framework adapts the model of Kittsteiner and Eyster (2007) to the particular situation in the German parliament.

[5]This is adapted to the institutional background of the German parliamentary system.

or to be close to the median voter on the national level to maximize within party career progression and increase in the list position. The basic setup explains observations of our empirical analysis: Winning a district is not directly associated with a higher position on the party list. In districts in which the national median voter is closely aligned to the median voter of a district, politicians position themselves close to the party-line and their opponent. In those districts, we neither observe deviations nor large voting margins in elections. However, in districts that are very different, we may see politicians who deviate from the party-line to focus on the district election and those who may deviate less to concentrate on the list position and party. Therefore we observe higher vote margins and higher deviation in districts very different from the national average. We show that low margins are associated with low deviations in districts close to the average. Further, districts very different from the average in terms of their socioeconomic characteristics are correlated with higher margins and higher deviation.

Previous research in the political science literature has investigated differences between district and list elected candidates. Within the German federal parliament Sieberer (2010); Neuhäuser, Mischler, Ruxton, and Neuhäuser (2013); Ohmura (2014); Sieberer (2015) explore relations between list and direct elected MPs and voting behavior. Sieberer (2010) compares roll call voting of district and list MPs between 2005 and 2009. The author finds that district MPs are significantly more likely to deviate from a party-line than MPs that are elected via a party list. Neuhäuser et al. (2013) confirm Sieberer (2010) results for the same period of time. Further, they show that the results are robust when controlling for MPs characteristics. The authors argue that directly elected MPs reelection probability is less dependent on the party. Therefore, direct MPs are less reliant on their standing within a party. Sieberer (2015) extends his previous finding by analyzing MPs explanation of votes. He shows that district MPs are more likely to voice reservation to the party-line. However, Sieberer (2015) also concludes that dual candidates (MPs which have been candidates on the list as well as in a district) do not differ in their explanation of vote. Ohmura (2014) challenges Sieberer (2010) and Neuhäuser et al. (2013) by investigating not only how district and list MPs differ but also if dual candidacy plays an important role. Ohmura (2014) analyzes roll call votes from MPs between 2002 and 2013 and shows that dual candidate MPs do not significantly differ in their deviation rate to the party-line. However, Ohmura (2014) shows that pure district MPs are more likely to deviate from the party-line. He argues that the candidacy strategy (i.e. pure district or dual), as well as re-election probability instead of the mandate itself, determines the voting of MPs. Finally, Sieberer (2014) confirms this results analyzing deviation of party-line by MPs with a dual candidacy strategy between 1949 and 2013.

Our research extends the literature in several dimensions. We employ a discontinuity design to evaluate the causal effect of mandate type on the deviation to the party-line with respect to voting and speeches in parliament. That allows us to draw conclusions about the causal effect of a district election. First, we establish that being voted into partliament by a district instead of a party list has no causal effect on voting in roll-call

votes. This paper's main contribution is that we analyze parliamentary speeches. These can serve to inform about MP's actions in committees behind closed doors and inform about MP-behavior in non roll-call votes, which constitute the majority of all votes in the German Parliament (Sieberer, Saalfeld, Ohmura, Bergmann, and Bailer, 2018). The use of speeches also addresses the concern that roll-call votes are a selected sample of votes (Carrubba, Gabel, and Hug, 2008). Parliamentary speeches are one of the most visible activities through which MPs express their policy positions. For example the works of Maltzman and Sigelman (1996); Proksch and Slapin (2012) underline the importance speeches for MPs and political parties.

The analysis of parliamentary speeches connects to a growing literature of textual analysis of political speeches based on machine learning mechanisms, see for example Grimmer and Stewart (2013), Martin and Vanberg (2008), and Quinn, Monroe, Colaresi, Crespin, and Radev (2010). Note especially Gentzkow, Kelly, and Taddy (2017) for a review of methodologies and different usages in economics.[6] Using these mechanisms provides us with a objective way of comparing speeches compared to an approach that involves hand-coding speeches according to some criteria. Moreover, this method can readily be extended to new data or different projects.

We use approaches of computational linguistics and computer science that uses text documents such as speeches to analyze MP's positions. Our textual analysis is related to previous work from Laver, Benoit, and Garry (2003) and Slapin and Proksch (2008) who estimate political positions using the word frequencies in party manifestos. In this paper we make use of the term frequency of MP's speeches to estimate distance measure between speeches. Further we use machine learning methods to evaluate how well one can predict party positions from speeches. The connection of machine learning and methods of causal inference is part of an emerging econometric literature (see Athey, 2018 as well as Mullainathan and Spiess, 2017).

We also relate to a literature of the effect of political competition. Examples include Strömberg (2008) who evaluates how competition affects campaign spending in US presidential elections or Dal Bó, Dal Bó, and Snyder (2009) who shows that increased political competition is associated with a lower probability of political dynasties. Using German data Bernecker (2014) shows that district opposition party MPs who expect a tight race are less absent in parliaments beforehand. We show that MPs do not only face competition in their district but also competition within a party may play an important role.

The remainder of the paper is organized as follows. In section 2 we describe the electoral system in Germany. Section 3 summarizes the predictions from theory. In section 4 we present our data sources, data manipulation and information retrieval methods. Section 5 shows descriptive statistics and section 6 describes our identification strategy. In section 7 we present the results, and rationalize these in a model in section 8. Section

---

[6]Gentzkow et al. (2017) describe most of the methodologies used in this paper.

9 concludes.

2   Institutional Background

The German federal parliament is a mixed-member proportional (MMP) representation. Voters of the parliament have two votes. The so called 'first vote' is the district component of the mixed-member proportional representation system. Currently, Germany has 299 districts, and in each district, electors may vote for a distinct candidate. First-past-the-post voting is used, that means for each district one politician enters parliament.

The distribution of the currently seats/members of the parliament (MP) to the different parties is determined by a proportional vote for a party list, called the 'second vote'. More specifically, voters of each state elect one party list. Given the 16 German states the proportion of parties is approximately equal to the shares of second votes aggregated over all states.[7] The direct mandates replaces a list mandate that has been determined by the party lists and second votes.[8]

Candidates on the party lists and direct candidates are not mutually exclusive. Indeed many candidates are on a party list as well as candidates in a district. Manow (2013) shows that a quarter of MPs are pure list or pure district candidates while the others are list as well as direct candidates. Appointment of candidates on the lists and the district is subject to rules. The party lists for each state have to be determined by secret election within parties. Direct candidates are either selected in elections within parties on the district level or have at least 200 signatures from voters within a district.

Although every MP in the German parliament has a free mandate and is formally not bound by a specific party-line, the mixed-member proportional representation has the intention to implement alternative incentive structure for direct MPs. Therefore direct candidates should represent the interests of a district which may deviate from the interest of a national wide party-line (see for example Scarrow, 2001). Nevertheless, the fact that the majority of MPs are dual candidates (list as well as district) such that a defeat of a candidate within a district is not necessarily connected to a loss of the membership within the parliament.

3   Predictions

In this paper we test three hypothesis we derived from previous literature. Our main research question investigates the relationship between mandate types and deviation of the party-line both with respect to voting and to speeches. Hypothesis 3 investigates

---

[7]In detail, the distribution of seats in parliament is determined by the Sainte-Lague method. Note, that due to the possibility of overhang seats (in case the number of direct mandates is larger than the number of seats coming from the second votes), as well as adjustment seats to reduce the possibility of strategic voting, the size of the German parliament may vary between election periods.

[8]Should a party win more direct seats than it won party seats, other parties receive overhang-seats keeping representation relative to the party vote-shares.

a potential confounding effect, namely that politicians who win a district gain influence influence in their party (possibly inducing them to act more in accordance with their party-line).

Hypothesis 1. *Members of parliament elected over a district-vote are more likely to deviate from the party-line in parliamentary roll-call votes than MPs elected over a party-list.*

Hypothesis 2. *The speeches of members of parliament elected over a district-vote differ from the speeches of MPs elected over a party-list. The wording of the speeches have (i) a higher distance from the speeches of their party members, and (ii) a higher distance from the manifesto of their party.*

Hypothesis 3. *Politicians winning a district-election marginally, are more likely to have a better (lower) list position in the next election cycle as compared to politicians loosing a district-election marginally.*

4 Data

We connect four data sources which cover a period from October 2005 to October 2017. That means we have data from three elections of the federal parliament and three legislation periods.[9] Firstly, we received data about the MPs and their direct vote shares within a district if applicable from a German nongovernmental organization (*Abgeordnetenwatch.de*). Secondly, we use data from MP's roll call voting[10] which is publicly available from the German parliament (*Bundestag*). We connect the data with the list positions from elected and non-elected candidates, and socio-economic statistics on the district level which we received from the federal electoral management body (*Bundeswahlleiter*)[11]. Thirdly, we receive party manifests for all parties represented by MPs for the three legislation periods byLehmann, Matthieß, Merz, Regel, and Werner (2017). Finally, we use parliament protocols from the German parliament (*Bundestag*).

---

[9]The first legislation period (the 16th of the German parliament since establishment in 1949) has existed between October 2005 and October 2009. The government has been a coalition of conservatives (CDU/CSU fraction) and socialdemocrats (SPD). The second legislation period lasted between October 2009 and October 2013 and conservatives led a government coalition with the liberal party. Finally, our last legislation period in the sample ranges from October 2013 and October 2017. Conservatives and socialdemocrats formed the government jointly.

[10]Note that we do not consider secret votings within the parliament. Note that secret votings are uncommon and solely used for personnel decisions (i.e. the voting for the chancellor).

[11]Note that for the election in September 2017 we collected the list positions manually from party websites on the state level.

4.1 Information Retrieval from Speeches

To measure the differences between speeches, we use the parliamentary protocols. We consider each verbal speech within the plenary hall as well as each registered interposed question of MPs. Note that we exclude the president, as well as the vice presidents (one for each party) of the German federal parliament, as they are leading and controlling the debates and formal procedures, which deviates from the role of other MPs. For similar reasons we exclude speeches of the ministers and federal chancellor of the government. We use conventional methods of information retrieval.[12] To convert text into a quantifiable measure, for each sentence we separate all words spoken. We stem each word to obtain the word's root form and then exclude most common words of the German language (*commonly called stopwords by the literature*). We use the stemmed words to, for each text, create a vector that with the count of words the text contains.[13] We then weight each element of the vector by the term frequency as well as the inverse document frequency of that word. This is a standard transformation with the purpose to decrease the relevance of words used by almost all MPs as well as words that are very rarely used (i.e. just at a single occasion) which would otherwise be prone to over-fitting. As a result, the presented information retrieval methods give us a vector of weighted term frequencies for each MP. Finally, we can include any combinations of subsequent words in the methodology above. That means a term refers to any single word used and not excluded by the methods above as well as the combination of any two (or more) subsequent words (after stemming and excluding stopwords).

We use these vectors to build two measures of similarity of the MPs' speeches to their respective party. Our first approach is to determine how similar the words an MP uses in her speeches are to the words used in the other speeches of her party in terms of the distance of the words used. We compute the cosine similarity of an MP's vector of term frequencies to the average vector of the MP's own party. Our second approach captures how close a member of parliament's speeches align to his or her party's policy position as measured by the party's election manifesto. We use the party-manifestos to build a measure of how well an MP's party-affiliation can be predicted based on her speeches.

To get a measure of distance of an MP's speech to all other speeches of the same party, we use the cosine similarity. In detail, let A and B be the term frequency vector of two MPs for in an election period. The cosine similarity is then defined as

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|_2 \|\mathbf{B}\|_2} = \frac{\sum\limits_{i=1}^{n} A_i B_i}{\sqrt{\sum\limits_{i=1}^{n} A_i^2} \sqrt{\sum\limits_{i=1}^{n} B_i^2}},$$

---

[12]Our approach is closely related to text analyses in the literature. For a detailed discussion and a guideline see Gentzkow et al. (2017).

[13]This results in a vector that has the size of all distinct words occurring in any speech. Naturally, for a single speech most entries in such vector are zero.

which ranges from 0 to 1. A higher value refers to a higher similarity between the two documents. Given our data we calculate for the cosine similarity for a MP in an election period to all other MPs within the same party and the same election period. Afterwards, we take the average of all cosine similarities. We interpret the final value as the average similarity of one MP to all other party members within a period. Again, the measure takes values between 0 and 1, where higher values refers to a high similarity to the speeches of all other party members.

For the second measure we make use of the election manifestos parties write before the election of the federal parliament.[14] We stem and vectorize the manifestos in the same way as the speeches and use them to build a measure of how close an MP's language resembles the language of her party's manifesto compared to the other manifestos. We measure the closeness to the party's manifesto by the probability with which a speech can be predicted as belonging to the MP's party by a classifier trained on all sentences of all parties manifestos. In other words, we measure how close an MP talks to it's own party's issues and policy standpoints relative to the other parties as defined in the parties' manifestos. Hence, if getting elected in a district makes MPs deviate further from their party-line, we should observe a downward break in the predictability of speeches by MPs who are just below the discontinuity at which MPs just won or lost a district election. Contrasting to the first measure, this can be regarded as an inter-party measure, as it captures the distance of speeches to the own party's manifesto *relative* to the other parties' manifestos. In the following we describe this strategy more in detail.

To create the measure, we train a classifier on the sentences of the election manifestos. We use the trained classifiers to predict for each MP which party she belongs to, based on her speeches. Then we take the predicted probability of the party the MP actually belongs to as a measure for the closeness of this MP's speeches to the her party's manifesto.

Thus the classifier we use has to be able to solve multinomial prediction problem with discrete features. In addition we need a classifier that yields not only predictions but a probability score which takes values from 0 to 1, where 1 is highest similarity (100% predicted probability) and 0 the lowest similarity (0% predicted probability).[15] We found that a penalized logistic regression does the best job as classifier in this situation. This is probably due to the sparseness and high-dimensional data as well as the fact that the data weuse for predictions (speeches) differs considerably from the train data (manifestos), we chose the hyper-parameters as to maximize accuracy of the predictions. We provide technical details of the prediction problem and our solution in the *Appendix*.[16]

---

[14]The according elections were in 2005, 2009, 2013.

[15]This not a trivial classification problem as there are 5 different parties and moreover we train the classifier on substantially different texts than we make predictions about both in terms of written vs. spoken language and the length of the texts. Our goal are not perfect predictions but to create a measure of manifesto closeness for each single MP, hence variation is to a certain extend is wanted to make the measure meaningful.

[16]Note that we also present robustness checks using other machine learning mechanisms.

5 Descriptive Analysis

Table 1 shows summary statistics of the number of MPs, the number of district elected MPs and the average district vote share. Note that each column refers to one of the three election period where the 16th legislature period covers the months between October 2005 and October 2009, the 17th period the month between October 2009 and October 2013 and the 18th period the covers the month between October 2013 and October 2017. We start by investigating the number of MPs and the number of district elected MPs within each of the three election periods.[17] The two big parties, conservatives (CDU/CSU[18]) and socialdemocrats (SPD), are the majority of MPs in the German parliament. These two parties win almost all direct mandates. While the conservatives have increased the number of district MPs during the last three periods the socialdemocrats lost a substantial part of district MPs. Investigating the district vote shares and comparing the district vote shares of all parties to the shares of the conservatives and socialdemocrats we see that the conservative party has a stable district vote share over 40 percent while the social democratic party lost district votes.

[Table 1 about here.]

Table 2 describes the variables describing politician's behavior. Using roll call data we measure the party-line as the majority vote within a party. In general, a MP has the option to agree, refuse or abstain a vote. Furthermore, a MP could generally do not attend a voting within the parliament. We do not consider MPs which are not attending and treat these observations as missing values. In order to evaluate the party-line we take the majority vote (agreement, refusal or abstention) of a party. In case an MPs votes against the majority of a party he deviates from the party-line. Comparing the different election periods in the first segment of Table 2 we see that period 16 is characterized by higher deviations than period 17 and 18. Comparing the two major parties, one observes that deviation for conservatives is less common than for social-democrats.

The second segment describes the average similarity of politician's speeches to other party members measured by the cosine similarity. A higher value describes a higher similarity. The third segment shows the average closeness of speeches to the manifestos predicted with the multinomial logistic regression. Note that you find summary statistics of an additional robustness check using a neural network in the appendix.

[Table 2 about here.]

---

[17]Note that the number of MPs is not stable as some MPs leave and some enter the parliament during the legislature period.

[18]The CDU/CSU is a party union where the CSU operates solely in the federal state Bavaria. The CDU/CSU can be seen as the main conservative party within Germany. The SPD is the social democratic party of Germany.

In Appendix 10.2 we provide additional insights about the list position and changes over election periods, as well as graphical investigations of the relation of district vote shares, roll call deviations, cosine similarities and manifesto closeness.

6   Identification Strategy

To provide first evidence regarding our hypotheses, we use an ordinary least squares regression in which we regress the outcome variable on a dummy indicating whether a member of parliament entered over a district or over a list. This is meant to give first evidence and possibly to confirm findings from previous papers. The following equation details this regression.

$$y_{p,i,t} = \alpha \; d_{p,i,t}^{won} + \beta \; Z_{p,i,t} + \varepsilon_{p,i,t}, \tag{1}$$

where $y_{p,i,t}$ is the outcome variable, $d_{p,i,t}^{won}$ is a dummy variable that takes the value one when a politician won a district, $Z_{p,i,t}$ is a vector of control variables, and $\varepsilon_{p,i,t}$ is the robust standard error. The basic least square regression may result in a biased estimator for several reasons. First, MPs who won a district election with a high margin may differ systematically to those MPs in the parliament who lost a district. For example, MPs who won the district with a high margin may have different career outlooks within a party. Such unobservable characteristic may influence the behavior of MPs. Formally we expect that the regressor is correlated with the unobservables and $E[\varepsilon|d^{won}, Z] \neq 0$, such that our estimates are not unbiased.

To investigate our hypotheses and avoid biased estimates we employ a regression discontinuity design as the main identification strategy. The forcing variable in this design is the relative margin by which a member of parliament won or lost a district election[19]. We denote this variable as $m_{p,i,t}$, where p is the member of parliament, i the district, and t the election period. We restrict our sample to politicians who a) entered the German federal parliament over a list or b) entered over a district-mandate, but had a list position sufficiently high, such that they would have entered over the list if they had not won their district. This restriction ensures that we avoid any selection bias which could arise from only observing winners but not losers of an election.

We argue that close enough to the cut-off where a politician just won or lost the district election, and hence entered the federal parliament with a district- or a list-mandate the assignment of the vote-share is quasi random. That means that the identifying assumption is that politicians below the cut-off do not differ in any relevant dimension from politicians above the cut-off. Our RDD specification is as below,

$$y_{p,i,t} = \alpha \; d_{p,i,t}^{won} + \beta \; f(m_{p,i,t}) + \varepsilon_{p,i,t}, \tag{2}$$

---

[19]The margin by which the according member won/lost divided by the vote-share of the first and second placed competitor.

where $y_{p,i,t}$ is the outcome variable, $d_{p,i,t}^{won}$ is a dummy whether a politician won a district or not, $f(m_{p,i,t})$ are different polynomials of the forcing variable described above, and $\varepsilon_{p,i,t}$ is the robust standard error. We limit the observations in the regression in two ways. Firstly, only members of parliament are included who have or could have entered the parliament through their party list. Secondly, we limit the regression to observations with the forcing variable within a certain margin around zero.

7 Results

7.1 Deviation from the Party Voting

In this subsection we present the results of the analysis regarding the deviation of different politicians from the parties voting line in parliamentary roll-call votes. Table 3 presents the results of an ordinary least squares regression and Table 4 the results of the regression discontinuity analysis.

Specification (1) in Table 3 shows the results of a naive regression as specified in Equation 1. Specifications (3) and (4) limit the sample to observations from members of parliament who either did enter over a party list or would have entered over a party list if they had lost their district election. This prevents a selection bias as we cannot observe the (counterfactual) votes of candidates who lost in a district election but did not enter parliament through a list either. Specification (2) and (4) add state, party, and session fixed effects as controls. This is relevant as the districts members of the different parties win are not evenly spread over all states and also the average percentage of votes that deviate from the party-line is different across parties (e.g. mostly higher for opposition parties) and across time. Specification (4) thus includes both fixed effects and prevents selection effects and is hence our preferred specification. As the table shows, the coefficient is not stable across the regressions and does change sign once the regression includes fixed effects, yet it is small in all regressions. This suggests care when interpreting the results. Our preferred specification suggests that the more politicians receive winning a district election are 1.85 percentage points more likely to deviate in roll-call votes.

[Table 3 about here.]

To investigate a causal relation between a district election and deviations in roll-call votes in parliament, we use the discontinuity at the cutoff where a candidate marginally lost or won their district election. Table 4 displays the results of the discontinuity regression as specified in Equation 2. In specification (1) only the averages are compared on a bandwidth of 5 percentage points around zero. Specification (2) adds the linear margin of the vote share on a bandwidth of 10 percentage points around zero. Specification (3) adds a second order polynomial and increases the bandwidth to 50. Finally specification (4) adds a third order polynomial and increases the margin even further to 50 percentage

points.[20] The bandwidth of the last two specifications is such that the majority of all observations are used, and only the tails are cut-off to avoid a bias due to over-fitting at the borders.

[Table 4 about here.]

Notably, the coefficients are positive in all specifications and larger than in the OLS, however, none or weakly significant. This does not allow to draw strong conclusions from the specification[21].

We do not find sufficient evidence that MPs who entered parliament over a district are more prone to deviate from the party-line in their voting, compared to MPs who entered parliament over a party list. Moreover deviations in general are very low.

Finding 1 *Entering parliament over a district election instead of their party's list does not cause MPs to vote against their party-line more often.*

7.2 Speeches

Next we present the results regarding the analysis of the deviation from the party average in the different speeches. We will present the results for the described two measures, the distance of the words used in the speech to the party average. And the likelihood with which a predictor calibrated on the election manifestos of all parties predicts the speech as coming from the according party of the MP. Table 5 and 7 show the results of an ordinary OLS regression and Table 6 and 8 the results of the regression discontinuity design.

[Table 5 about here.]

Specification (1) in Table 5 and 7 show the results of a completely naive regression as specified in Equation 1. Specification (3) and (4) restrict the sample to observations from members of parliament who either entered over a party list or entered over a district-election, but would have entered over a party list if they had lost their in district. This prevents a selection bias as we are lacking the contractual votes of candidates who lost in a district election but did not enter parliament through a list either. Specification (2) and (4) add state, party, and session fixed effects as controls. This is relevant as the districts winners of the different parties are not evenly spread over all states and also the average percentage of votes that deviate from the party-line is different across parties (e.g. mostly higher for opposition parties) and across time.

Table 5 shows that an OLS regression predicts a significantly negative correlation of winning a district mandate and the similarity of speeches in parliament with speeches of the own party. However, adding controls increases the coefficients such that they are positive and insignificant. Yet the size of the coefficient is small compared to the average within party similarity (see Table 2).

---

[20]Altering the bandwidth slightly for either specification does not change the result.

[21]Figure 5 in the appendix displays the discontinuity graphically. Once fitted by a second and once by a third order polynomial. Visual inspection shows a slight but not very strong jump at the margin.

[Table 6 about here.]

In contrast, Table 6 shows that, at the margin of just winning or loosing a district, there is no significant effect of winning a district on the closeness of an MP's speeches to her party peer's speeches. The coefficients are all small and positive however far from being significant. This suggests that there is no discontinuity in the cosine similarity of speeches at the margin of just being elected or not in a district.

The second measure we investigate is the manifesto-similarity of speeches, which measures how close an MP's speeches are towards their own party's manifesto relative to the other parties' manifestos. Figure 3 shows a fitted polynomial and the binned sample average of the speech similarity measure around the voting cut-off. Visual inspection of the figure supports the conclusion that there is no discontinuity. The graph shows a small increase at the cut-off, however this difference is marginal and well within the variation of the bins.

[Table 7 about here.]

Table 7 shows the identical OLS regressions for the predictability measure. Here specifications (1) and (3) are significantly negative, however as for the similarity measure the coefficient turns positive and insignificant once controlling for the party, state, and session in (2) and (4).

Table 8 then shows the result of the regression discontinuity regression for the predictability measure. Here all coefficients are small and negative and non-significant. Hence here as well the data suggests that there is no discontinuity in speech behavior according to the predictability measure.[22]

[Table 8 about here.]

Finding 2 *Entering parliament over a district election instead of their party's list does not cause MPs to use a wording (i) further away from the speeches of their party peers or (ii) further away from their party's manifesto relative to the other manifestos.*

7.3 Future List Position

In this subsection we present the results of the analysis the effect of winning a district election on the future position on the party list. We define the change of a list position as of a MP from the position in election t to the next election, i.e. $\Delta_t = P_t - P_{t+1}$. A MP with a positive δ_t means that a politician is in a better position to get reelected over

---

[22]Figure 4 in the appendix shows a fitted polynomial and the binned sample average of predictability around the voting cutoff. Visual inspection of the figure seems to confirm the conclusion that there is no discontinuity as the imposed cut-off is marginal.

the list.[23] Table 9 presents the results of an ordinary least squares regression and Table 10 the results of the regression discontinuity analysis.

Specification (1) in Table 9 shows the results of a simplistic regression as specified in Equation 1. Specification (3) and (4) limit the sample to observations from members of parliament who either entered over a party list or entered over a district-election, but would have entered over a party list if they had lost their in district. This prevents a selection bias as we are lacking the contractual votes of candidates who lost in a district election but did not enter parliament through a list either. Specification (2) and (4) add state, party, and session fixed effects as controls. This is relevant as the districts members of the different parties win are not evenly spread over all states and also the average percentage of votes that deviate from the party-line is different across parties (e.g. mostly higher for opposition parties) and across time.

[Table 9 about here.]

Investigating the results, an OLS regression does not suggest a strong relationship. The coefficient is insignificant in the first three specifications and even changes sign. Only specification (4) suggests a negative relationship, which would mean that winning a district correlates with a worse position on a list.

This result is not supported by the analysis of the potential discontinuity shown in Table 10. The table shows that all coefficients are non-significant, which suggests that no relationship exists between marginally winning or loosing a district and the future list position.[24]

[Table 10 about here.]

Finding 3 *Entering parliament over a district election instead of their party's list does not cause MPs to have a higher list position in the election for the following legislation period.*

8 A Simple Model of the Median Voter

This section introduces a stylized model that rationalizes the (non-)result of the empirical analysis in a simplified environment. We further show some descriptive relationships that support the model's implications.

Suppose a country has n districts, indexed by d. Each district is characterized with socio-economic characteristics C^d. Further, voter i in district d is characterized by her

---

[23]On average politicians increase their list positions over time, see the summary statistics in Appendix 10.2.

[24]Figure 10 in the appendix shows the potential discontinuity graphically with a 2nd and 3rd order polynomial fitted. Visual inspection of the figure confirms the result of Table 10 that there no discontinuity in the future list position exists.

preference about political policy that can be represented one-dimensional on the unit interval $x_i^d \in [0, 1]$. Therefore we can place all voter-preferences along a one-dimensional political spectrum. Each individual voter has single-peaked preferences decreasing in the distance between the voter's optimal point and the actual policy in place $U_i(|x_i - x|)$, where $\partial U_i(|x_i - x|)/\partial |x_i - x| < 0$. The distribution of the voters in district d on the interval is described by $F^d(x)$. The aggregate distribution on the national level sum of distributions of all districts, $G(x) = \sum_d F^d(x)$.

Within this simple environment we consider a situation where two parties *left* and *right* (l and r) compete on the national level as well as in each district $d \in \{1, n\}$. The parties compete on a national level for the proportion of seats in the parliament while candidates on a district compete for one seat. We assume that candidates have a dual candidacy strategy, so they compete for election both on a closed party list as well as in their a district. On the national level the two parties $P \in \{l, r\}$ try to attract the median voter, x_M, to maximize their probability of election,[25] hence announce the median voter's preferred policy as their campaign platform.[26]

In each district two candidates of the two parties l^d and r^d are competing for election. Additionally, each of them might also enter over the party list, such that each politician has to balance two parts of his objective function. In their district, candidates are facing a distribution of voters preferences $F^d(x)$ that potentially differs from the national distribution $G(x)$. A candidates individual district-election probability $p_i^d(\Delta x_i^d)$ is a decreasing function in the distance of the candidate's policy platform from the district's median voter Δx_i^d. Hence, to maximize the probability of winning the district, a candidate should choose the district median policy x_M^d. However, a candidate also has an incentive to minimize the distance to his party's platform (we will also refer to as party-line). Denote the individual probability of getting elected over a party list as $p_i^P(\Delta x_i^P)$, which is a decreasing function of the distance between the candidates policy platform and the party-line, Δx_i^P. The intuition is that the party determines the candidates list position on the national level and a policy platform closer to the party median is advantageous for a MP's chances of being placed high on the list. This assumption is backed-up by the results of Section 7, where we show that winning a district is not associated with an increase in the party list. We expect that the trade-off between motives is dependent on the individual expectation of success within a district. For example, a left-leaning candidate who runs in a right-leaning district is more likely to focus on the party-line instead of the district.

This basic setup can explain the non-result at the discontinuity of district candidates. The argument is exemplified in Figure 1. Distribution G in the figure depicts the overall distribution of voters X on the national level. The three distributions F^d for $d = \{1, 2, 3\}$

---

[25] For a general discussion of assumption of the median voer model see for example Gruber (2005).

[26] Here politicians and parties always follow through on their campaign platform post election. For example Austen-Smith and Banks (1989); Aragonès, Postlewaite, and Palfrey (2007); Corazzini, Kube, Maréchal, and Nicolo (2014); Born (2018) study situations in which promises are costly to break but not completely binding.

depict *sample* distributions of three different districts with the same variance but unequal means. Suppose all four distributions are normally distributed such that x_M is the median voter of the distribution G. The distribution of F^2 has the same median as G such that $x_M^2 = x_M$. In district $d = 2$ both candidates l^2 and r^2 have an incentive to act according to $x_M^2 = x_M$. In other words, in districts which are close to the median voter on the national level, politicians do not deviate from the party-line for two reasons: The median voter in the district is the same as on the national level which also coincides with the party-line. As a result, the model predicts differences between winners and losers of such central district to be characterized by very close margins. Hence, there should not be a difference between winners or losers in these districts due to the alignment of the district and national median policy.

In comparison, consider district $d = 1$ (the same argument is holds for opposite leaning district $d = 3$). Here the median voter is x_M^1. That means, the district is left leaning compared to the national level and candidates face the trade off between the median voter of their district and their party-line (the national level median). A candidate chooses his party platform as to maximize the sum of the probability to win the district-election and to be elected over the party list $p_i^d + p_i^P$. Unlike in a central district, there is a trade off between the optimal policy platform for the district and the party election probabilities. For the choice of policy platform the individual effect of a candidate matters. A candidate who has a relatively higher chance of winning the district (due to individual specific effects) has a higher return of choosing his policy platform close to the median policy of the district. Vice versa, the weaker of both candidates has a higher return of choosing a platform close to the party-line. Hence, both candidates will deviate in their choice of policy platform in districts $d = 1, 3$. As a result the theory predicts higher winning margins as well as a higher diversion from the party policy of winners in these districts.

[Figure 1 about here.]

We proceed to test the implications of the stylized model. First of all, the model predicts that the differences in marginal votes should be small in districts that are close to the national median. Within those districts behavior of politicians is also similar to the party-line. This prediction rationalizes that we do not find evidence of behavioral differences at the discontinuity. On the flip side, our model predicts that districts that are different from the national level on the political spectrum, are characterized by (1) higher margins between winner and loser and (2) politicians who deviate further from their party's position.

We do not observe the exact political orientation of districts but try to evaluate both hypothesis by showing correlations between district socioeconomic statistics and the voting margins (between winner and loser) as well as the behavior of MPs. The assumption is that the socioeconomic characteristics are correlated with the actual political median position of a district. To test both hypothesis we look on correlations on the sample of politicians of the two major parties CDU (conservatives) as well as SPD

(social democrats). We evaluate if the margin between winner and loser for politician p at legislature period t in district d, $margin_{p,d,t}$ is correlated with key socioeconomic characteristics collected in a vector $C_{d,t}$. The vector includes *the population density, the share of male in population, the share of German citizens, the share of population older than 60 years old, the unemployment share, the share of cars per 1000 inhabitants, the share of population without a lowest secondary degree.*[27] We also investigate the correlation of $C_{d,t}$ with the behavior of politicians $y_{p,d,t}$. We try therefore to approximate if the ordering of a politician is correlated with specific characteristics.

[Table 11 about here.]

Note first that the Pearson correlation coefficients are for all MPs from the conservative and social-democratic parties. We investigate the correlation with the absolute value of the marginal vote share, i.e. the absolute value of the second to the first and first to second in terms of votes. The intuition is that we would like to explore the correlation of the district characteristics with marginal vote share across party.

First, we turn to the hypothesis that the characteristics may be correlated with the marginal differences within a district. One observes that all but one off the characteristics $C_{d,t}$ are correlated with the marginal vote share. However, the strongest it is with the share of cars, the unemployment share and the share of population without secondary degree. In districts with more cars, lower unemployment, higher share of males, higher share of German citizens and lower share of population without a secondary degree the vote differential between the two candidates is larger. We cannot reject the model prediction that districts deviating from the national level on the political spectrum are characterized by (1) higher margins between winner and loser.

Second, we explore the correlation between the deviation of the party-line and the district characteristics. The model predicts that deviation should be higher the more different a district is from the national median. We observe that especially those characteristics associated for social disadvantages are correlated with roll call deviations. MPs deviate more from the party-line the higher the unemployment and the higher the share of population without secondary degree. Considering the speeches we use the cosine similarity. The higher share of males is correlated with less deviations. Next we evaluating the correlation of the speeches with the district characteristics. Here we see a similar pattern across the cosine similarity, and the closeness measures to the party manifestos predicted with machine learning methods. A higher unemployment share as well as a higher share of population without a secondary degree is correlated with more deviation from the party-line. We conclude that especially in socially disadvantaged districts MPs deviate from the party-line. This observation is in line with the predictions in the model.

---

[27]In Appendix 10.2 (Table 13) we provide some key summary statistics.

9 Discussion

This paper investigates whether winning a district seat causes a change in an MP's adherence to a party-line in three dimensions among the members of the German mixed-member parliament. Firstly, we investigate whether MPs who are elected by a district deviate more often from their party-line in roll-call votes than MPs who enter parliament over a party list. Secondly, we evaluate whether the speeches of these MPs differ more from those of their peers and the manifesto of their party. Thirdly, we rule out that winning a district seat increases an MP's position on a party list in future elections which could possibly confound the results of the first two questions we analyze.

We don't find evidence that MPs who marginally won a district election deviate more often from the party-line in roll-call votes or speeches than their peers who marginally lost a district election. In other words, entering parliament through over a district does not cause MPs to deviate further from their party-line. Our findings suggest that a mixed-member parliamentary system binds all MPs, even MPs elected by a district, to a strong party discipline as it is common for systems with relative representation. In contrast, district elected MPs do not have the same degree of accountability to their district as their counterparts in a majoritarian parliamentary system does.

In relation to research that demonstrates how semi-open lists promote careers of politicians who score many personal votes, we show that the same type of career progression does not occur for politicians who win their district election in a mixed-member parliament. A possible reason for this is that in each district a party only has a single candidate. An explanation for the absence of the effect in German system is that here every party only sends one candidate into a municipality election whereas an open or semi-open list setting allows competition among members of the same party. Hence, unlike in an open list election, a candidate in a district cannot demonstrate her popularity relative of other members within the same party.

Finally, the correlation of speeches with district characteristics demonstrates that districts with different socio-economic characteristics select different MPs. Hence, an advantage of the mixed-member parliemantary system seems to be that it allows the parliament to reflect the economic interests of heterogeneous districts.

10 Appendix

10.1 Manifesto-closeness Measure

10.1.1 Description of Methodology

Multinomial Logistic Regression:

In the main specification we train multinomial logistic regression with L2 regularization to classify speeches based on the party manifestos. A logistic regression yields likelihoods instead of absolute predictions without additional transformation that would be necessary with other algorithms such as Naive Bayes classifier, Random Forest classifier, or a Support Vector Machines.[28] We controlled that the prediction of the logistic classifier does not perform significantly worse than any of the other mentioned classifiers as measured by the count of speeches for with which it assigns the highest likelihood to the correct party. We use cross-validation to find the optimal height of the penalty term, maximal number of features, whether to include 4-,3-, and 2-grams, and whether to weight by tf-idf. The set of parameters that leads to the best internal prediction values is a penalty parameter of 1, to include 1-, 2-, 3-, 4-grams in the analysis, to include tf-idf weighting, and to set the maximal amount of features to 100,000. After training the classifier on the party manifestos, we use the classifier to categorize the merged speeches each MP held over an entire period of office. The classifier yields a prediction of the party member-ship of the according mp. In particular we use the probability the classifier assigns to the MP's actual party as a measure of similarity between the speeches of the MP and the MP's party. That means a high predicted probability of the actual party of an MP indicates that the MPs speeches are closer to the party manifesto of his or her own party than the manifestos of other parties. This measure will take values from 0 to 1, where 1 is highest similarity (100% predicted probability) and 0 the lowest similarity (0% predicted probability).

After training the classifier on the party manifestos, we use the classifier to categorize the merged speeches each MP held over an entire period of office. The classifier yields a prediction of the party member-ship of the according MP. In particular we use the probability the classifier assigns to the MP's actual party as a measure of similarity between the speeches of the MP and the MP's party. That means a high predicted probability of the actual party of an MP indicates that the MPs speeches are closer to the party manifesto of his or her own party than the manifestos of other parties.

10.1.2 Robustness

To ensure that the results do not hinge on the particularities of the manifestos that might be quite different to speeches, as a robustness check, we use the same logistic classifier and train it on speeches instead of manifestos in a leave-one-out fashion. That means, for

---

[28]The probabilities by these other algorithms need to be calibrated on a share of the data not used for training in order to produce evenly distributed probabilities, see e.g.Zadrozny and Elkan (2002).

each MP we fit the logistic classifier on all speeches but the ones from the MP in question. We use the so fitted classifier to predict the party that MP belongs to and then repeat the process for the next MP. This methodology gives a higher share of correct predictions (around 80%) compared to the usage of manifestos.

Table 11 provides the result of the robustness check of Table 7. Notable without controls the coefficient is positive and significantly different from zero. Controlling for state, party, and session fixed effects hwoever brings the coefficient close to and not significantly different from zero.

[Table 12 about here.]

Table 12 provides a robustness check of Table 8 by displaying the RDD estimates of the relationship of marginally winning vs. loosing a district mandate on the manifesto-closeness of the according MP's speeches. Noticeable, as in the main analysis, none of the coefficients is significantly different from zero.

[Table 13 about here.]

Table 12 provides a robustness check of Table 8 by displaying the RDD estimates of the relationship of marginally winning vs. loosing a district mandate on the manifesto-closeness of the according MP's speeches. Noticeable, as in the main analysis, none of the coefficients is significantly different from zero.

10.2 Additional Descriptive Analysis

The first part of our research investigates the impact of a district mandate on the probability of deviation from the party line in roll call votes. Further, we explore the relationship between the deviation in voting and the development of the list position. We start with evaluating correlations of variables with the deviation from the party line. To give a graphical impression about possible correlations Figure **??** shows the district vote margin to the share of deviations from an individual MPs. Note firstly, that the vote margin is defined as the direct vote share of district candidacy MPs to the closest opponent. Further, we treat each election period independently such that one observation corresponds to an district candidacy MP within one election period. The graph includes a linear least square regression line as well as the 95% confidence band. We see a slightly negative relationship between district margin and the share deviations. District candidacy MPs with a higher district margin are correlated with lower deviations. However, the relationship is not strong.

List Places

We also present some summary statistics about the list place changes between consecutive periods. We calculate the list place change of those MPs which are part of a state lists.

In detail, let P_t be the list place at time t where a lower position gives a higher possibility to enter the parliament in case of no direct mandate. The list change is defined as $\Delta_t = P_t - P_{t+1}$. Note that we can only observe MPs which are part of a list within the next election period. In general MPs increase their list position (get a better list place). This effect is stronger for the conservatives than for the socialdemocrats. Furthermore, the effect is stronger for district elected MP. However, the increase is mostly due to those district MPs from the socialdemocrats. The list place changes socialdemocratic, district elected candidates is lower and even negative from the 17th to the 18th period as well as the 18th to the 19th election period..

[Table 14 about here.]

Correlations with Socioeconomic Characteristics

This subsection shows summary statistics of socioeconomic variables across all three legislation periods.

[Table 15 about here.]

10.3   Additional Results

Supporting the analysis in Section 7.2 the following figures show fitted polynomials and sample averages of the RDD estimations of the section.

Figure 5 shows a fitted polynomial and the binned sample average of voting-deviations from the party-line. Visual inspection shows a slight but not very strong jump at the margin.

[Figure 2 about here.]

Figure 3 shows a fitted polynomial and the binned sample average of the cosine distance of an MP's speeches to the party average. The displayed cutoff indicates whether an MP won or lost the vote. Visual inspection of the figure confirms the conclusion that there is no discontinuity.

[Figure 3 about here.]

Figure 4 shows a fitted polynomial and the binned sample average of manifesto-closeness of speeches around the voting cutoff. Visual inspection of the figure confirms the conclusion that there is no discontinuity.

[Figure 4 about here.]

Figure 10 shows a fitted polynomial and the binned sample average of the future list positions. Visual inspection of the figure confirms the result of Table 10 that there no discontinuity in the future list position exists.

[Figure 5 about here.]

References

ACEMOGLU, D., S. JOHNSON, AND J. A. ROBINSON (2001): "The colonial origins of comparative development: An empirical investigation," *American economic review*, 91, 1369–1401.

ALESINA, A., E. GLAESER, AND B. SACERDOTE (2001): "Why doesn't the US have a European-style welfare system?" *National bureau of economic research.*

ANAGOL, S. AND T. FUJIWARA (2016): "The runner-up effect," *Journal of Political Economy*, 124, 927–991.

ANDRÉ, A., S. DEPAUW, M. S. SHUGART, AND R. CHYTILEK (2017): "Party nomination strategies in flexible-list systems: Do preference votes matter?" *Party Politics*, 23, 589–600.

ARAGONES, E. AND T. R. PALFREY (2002): "Mixed equilibrium in a Downsian model with a favored candidate," *Journal of Economic Theory*, 103, 131–161.

ARAGONÈS, E., A. POSTLEWAITE, AND T. PALFREY (2007): "Political reputations and campaign promises," *Journal of the European Economic Association*, 5, 846–884.

ATHEY, S. (2018): "The impact of machine learning on economics," in *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press.

AUSTEN-SMITH, D. AND J. S. BANKS (1989): "Electoral accountability and incumbency," in *Models of Strategic Choice in Politics*, University of Michigan Press.

BERNECKER, A. (2014): "Do politicians shirk when reelection is certain? Evidence from the German parliament," *European Journal of Political Economy*, 36, 55–70.

BORN, A. (2018): "Promise Competition," *working paper.*

CAREY, J. M. (2007): "Competing principals, political institutions, and party unity in legislative voting," *American Journal of Political Science*, 51, 92–107.

CARRILLO, J. D. AND M. CASTANHEIRA (2008): "Information and strategic political polarisation," *The Economic Journal*, 118, 845–874.

CARRUBBA, C., M. GABEL, AND S. HUG (2008): "Legislative voting behavior, seen and unseen: A theory of roll-call vote selection," *Legislative Studies Quarterly*, 33, 543–572.

CORAZZINI, L., S. KUBE, M. A. MARÉCHAL, AND A. NICOLO (2014): "Elections and deceptions: an experimental study on the behavioral effects of democracy," *American Journal of Political Science*, 58, 579–592.

CRISP, B. F., S. OLIVELLA, M. MALECKI, AND M. SHER (2013): "Vote-earning strategies in flexible list systems: Seats at the price of unity," *Electoral Studies*, 32, 658–669.

DAL BÓ, E., P. DAL BÓ, AND J. SNYDER (2009): "Political dynasties," *The Review of Economic Studies*, 76, 115–142.

DEPAUW, S. AND S. MARTIN (2009): "Legislative party discipline and cohesion in comparative perspective," *Intra-party politics and coalition governments*, 103–120.

DOWNS, A. (1957): "An economic theory of political action in a democracy," *Journal of political economy*, 65, 135–150.

FOLKE, O., T. PERSSON, AND J. RICKNE (2016): "The primary effect: Preference votes and political promotions," *American Political Science Review*, 110, 559–578.

GENTZKOW, M., B. T. KELLY, AND M. TADDY (2017): "Text as data," Tech. rep., National Bureau of Economic Research.

GERBER, E. R. AND J. B. LEWIS (2004): "Beyond the median: Voter preferences, district heterogeneity, and political representation," *Journal of Political Economy*, 112, 1364–1383.

GRIMMER, J. AND B. M. STEWART (2013): "Text as data: The promise and pitfalls of automatic content analysis methods for political texts," *Political analysis*, 21, 267–297.

GRUBER, J. (2005): *Public finance and public policy*, Macmillan.

KITTSTEINER, T. AND E. EYSTER (2007): "Party platforms in electoral competition with heterogeneous constituencies," *Theoretical Economics*, 2, 41–70.

LAVER, M., K. BENOIT, AND J. GARRY (2003): "Extracting policy positions from political texts using words as data," *American Political Science Review*, 97, 311–331.

LEHMANN, P., T. MATTHIESS, N. MERZ, S. REGEL, AND A. WERNER (2017): "Manifesto Corpus. Version: 2017-2," *Berlin: WZB Berlin Social Science Center*.

LINDBECK, A. AND J. W. WEIBULL (1987): "Balanced-budget redistribution as the outcome of political competition," *Public choice*, 52, 273–297.

——— (1993): "A model of political equilibrium in a representative democracy," *Journal of public Economics*, 51, 195–209.

LIZZERI, A. AND N. PERSICO (2001): "The Provision of Public Goods under Alternative Electoral Incentives," *American Economic Review*, 91, 225–239.

MALTZMAN, F. AND L. SIGELMAN (1996): "The politics of talk: Unconstrained floor time in the US House of Representatives," *The Journal of Politics*, 58, 819–830.

MANOW, P. (2013): "Mixed rules, different roles? An analysis of the typical pathways into the Bundestag and of MPs' parliamentary behaviour," *The Journal of Legislative Studies*, 19, 287–308.

MARTIN, L. W. AND G. VANBERG (2008): "Coalition government and political communication," *Political Research Quarterly*, 61, 502–516.

MERILÄINEN, J. AND J. TUKIAINEN (2018): "Rank effects in political promotions," *Public Choice*, 177, 87–109.

MILESI-FERRETTI, G. M., R. PEROTTI, AND M. ROSTAGNO (2002): "Electoral Systems and Public Spending*," *The Quarterly Journal of Economics*, 117, 609–657.

MUELLER, D. C. (1997): *Perspectives on public choice: a handbook*, Cambridge University Press.

MULLAINATHAN, S. AND J. SPIESS (2017): "Machine learning: an applied econometric approach," *Journal of Economic Perspectives*, 31, 87–106.

NEUHÄUSER, A., M. MISCHLER, G. D. RUXTON, AND M. NEUHÄUSER (2013): "Gründe für von der Fraktionsdisziplin abweichendes Abstimmungsverhalten bei Bundestagsabgeordneten," *AStA Wirtschafts-und Sozialstatistisches Archiv*, 7, 91–99.

OHMURA, T. (2014): "When Your Name is on the List, It is Time to Party: The Candidacy Divide in a Mixed-Member Proportional System," *Representation*, 50, 69–82.

PERSSON, T., G. ROLAND, AND G. TABELLINI (1997): "Separation of powers and political accountability," *The Quarterly Journal of Economics*, 112, 1163–1202.

PERSSON, T. AND G. TABELLINI (1999): "The size and scope of government:: Comparative politics with rational politicians," *European Economic Review*, 43, 699–735.

——— (2004): "Constitutional rules and fiscal policy outcomes," *American Economic Review*, 94, 25–45.

PROKSCH, S.-O. AND J. B. SLAPIN (2012): "Institutional foundations of legislative speech," *American Journal of Political Science*, 56, 520–537.

QUINN, K. M., B. L. MONROE, M. COLARESI, M. H. CRESPIN, AND D. R. RADEV (2010): "How to analyze political attention with minimal assumptions and costs," *American Journal of Political Science*, 54, 209–228.

SCARROW, S. E. (2001): "Germany: The mixed-member system as a political compromise," *Mixed-Member Electoral Systems: The Best of Both Worlds*, 55–69.

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

——— (2014): "Do Single Member Districts Increase Voting Dissent Under Mixed Electoral Systems? A Longitudinal Study of Voting Behavior in the German Bundestag, 1949-2013," .

——— (2015): "Using MP statements to explain voting behaviour in the German Bundestag: An individual level test of the Competing Principals Theory," *Party Politics*, 21, 284–294.

SIEBERER, U., T. SAALFELD, T. OHMURA, H. BERGMANN, AND S. BAILER (2018): "Roll-Call Votes in the German Bundestag: A New Dataset, 1949–2013," *British Journal of Political Science*, 1–9.

SLAPIN, J. B. AND S.-O. PROKSCH (2008): "A scaling model for estimating time-series party positions from texts," *American Journal of Political Science*, 52, 705–722.

STRÖMBERG, D. (2008): "How the Electoral College influences campaigns and policy: the probability of being Florida," *American Economic Review*, 98, 769–807.

ZADROZNY, B. AND C. ELKAN (2002): "Transforming Classifier Scores into Accurate Multiclass Probability Estimates," in *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, KDD '02, 694–699.

Figure 1: Example of Voter Distributions in Different Districts
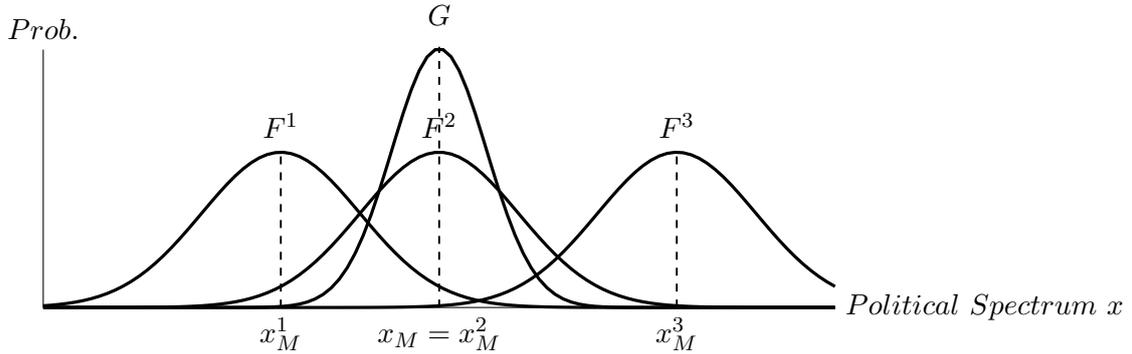
The graph presents an example of possible voter distributions across the political spectrum x. F^1, F^2 and F^3 present three different distributions of voter preferences in districts whereas G present the distribution on the national level. Note, that the G is an aggregation but this graph just examplifies this relation. All distributions are normal. In this example, the median voters are located at x_M^1, x_M^2 and x_M^3. The national median voter is located at the same position as the median voter of district two.

Figure 2: Graphical analysis of the effect of a district vote on deviations.
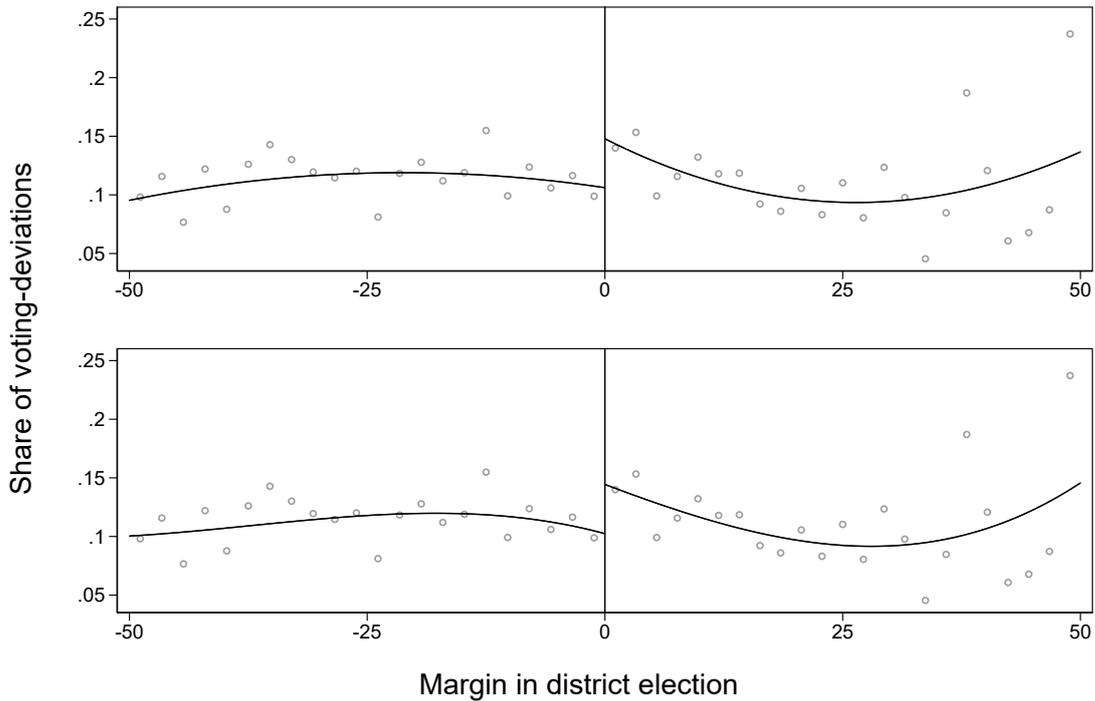
The graphs show binned averages of MPs' share of voting deviations as a function of the win/loss margin to being the district-vote winner. The bins are created to variance evenly spaced using a polynomial regression. The lines in the graphs are second- (top) third-order (bottom) polynomials fitted to the observations.

Figure 3: Graphical analysis, winning a district mandate on the cosine distance to party average.

Margin in district election
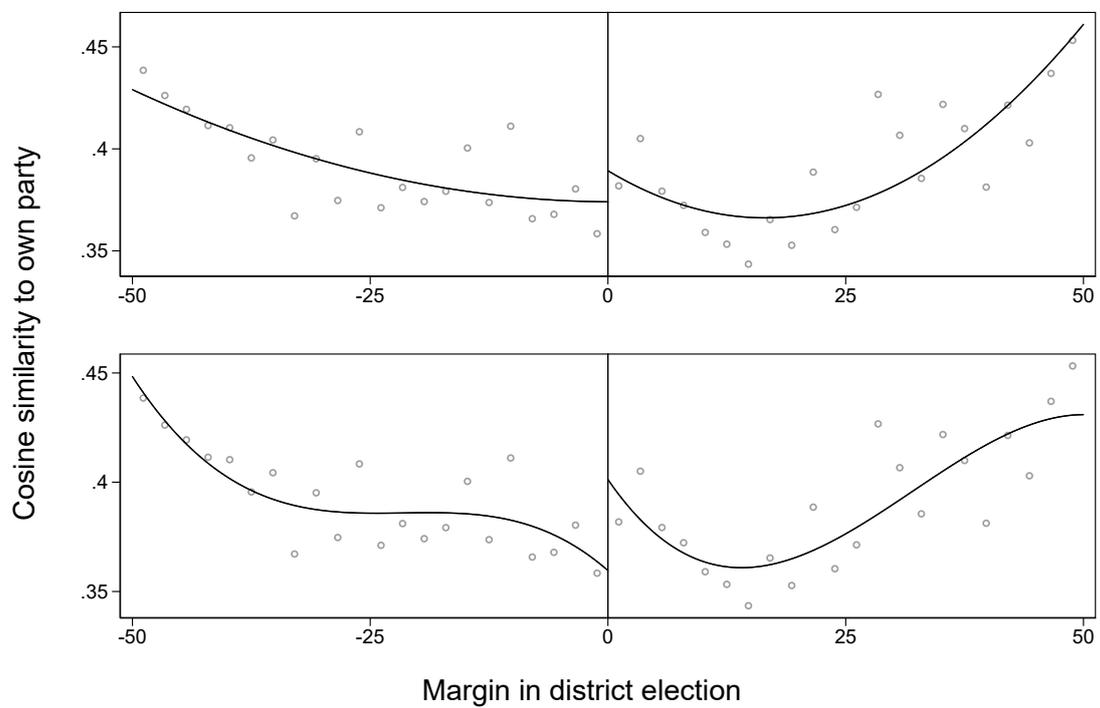
The graphs show binned averages of MPs' cosine similarity in their speeches as a function of the win/loss margin to being the district-vote winner. The bins are created to variance evenly spaced using a polynomial regression. The lines in the graphs are second- (top) third-order (bottom) polynomials fitted to the observations.

Figure 4: Graphical analysis, district vote on manifesto-closeness of speeches.

The graphs show binned averages of MPs' speeches' manifesto-closeness as a function of the win/loss margin to being the district-vote winner. The bins are created to variance evenly spaced using a polynomial regression. The lines in the graphs are second- (top) third-order (bottom) polynomials fitted to the observations.

Figure 5: Graphical analysis effect district vote on future list positions.

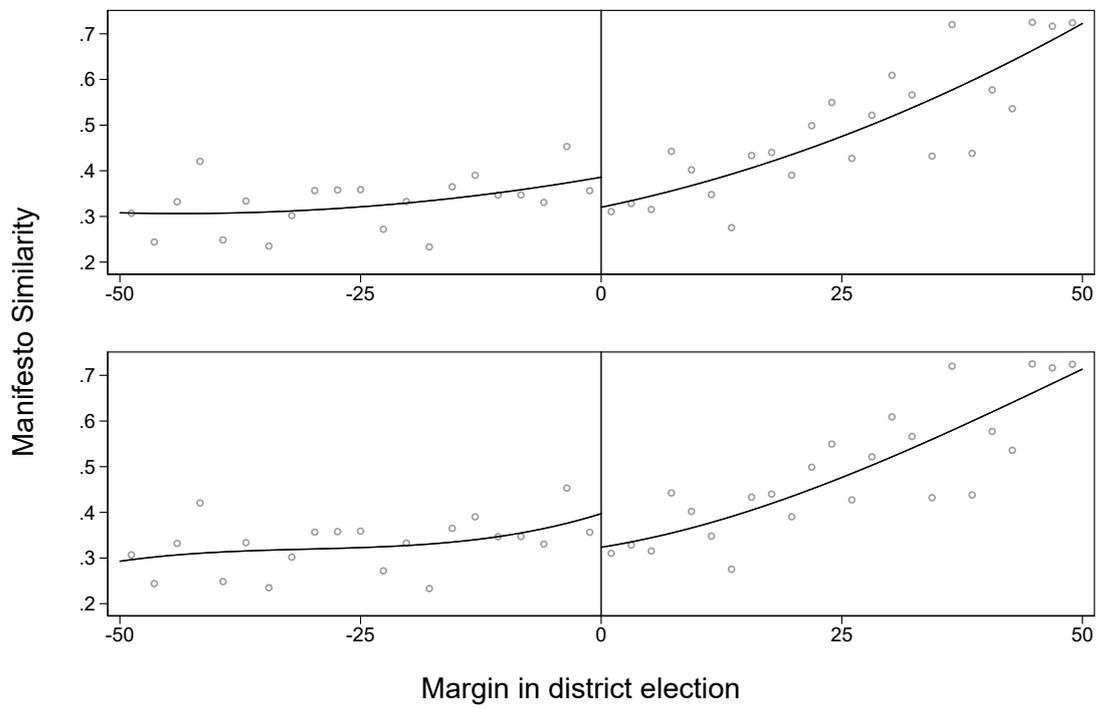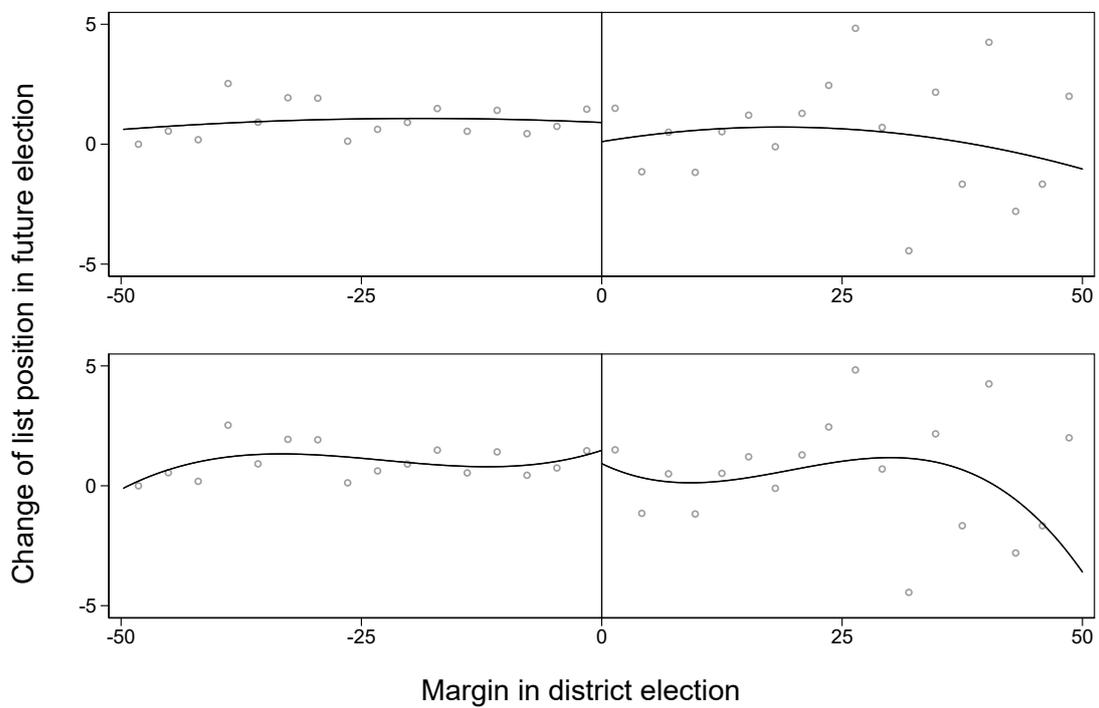*The graphs show binned averages of MPs' future list position as a function of the win/loss margin to being the district-vote winner. The bins are created to variance evenly spaced using a polynomial regression. The lines in the graphs are second- (top) third-order (bottom) polynomials fitted to the observations.*

Table 1: Summary Statistics of Parliament and their Members

Summary Variables	Period 16	Period 17	Period 18
No of MPs	640	653	658
No of MPs, Conservatives	238	245	323
No of MPs, Socialdemocrats	228	154	205
No of Distict MPs	299	299	299
No of Distict MPs, Conservatives	150	218	236
No of Distict MPs, Socialdemocrats	145	64	58
Avg. District Vote Share, Conservatives	43.31	41.37	45.72
	(9.96)	(7.35)	(8.33)
Avg. District Vote Share, Socialdemocrats	41.1	31.69	32.26
	(8.77)	(8.62)	(8.94)

Each statistic is divided across into the three different election periods. We show statistics for all MPs and those MPs of the two major parties (Conservatives: CDU/CSU fraction and Socialdemocrats: SPD). Note, that the number if MPs is fixed, however, some MPs exit and other enter within election periods. The average district vote share is the average vote share across those MPs that have been candidates in a district. Standard deviations are shown in parentheses.

Table 2: Summary Statistics of MP's Parliamentary Behavior

Summary	Period 16	Period 17	Period 18
Roll Call Deviation			
Avg Voting Deviation	0.3	0.1	0.11
	(0.26)	(0.1)	(0.12)
Avg Voting Deviation, Conservatives	0.16	0.06	0.09
	(0.11)	(0.08)	(0.1)
Avg Voting Deviation, Socialdemocrats	0.15	0.14	0.13
	(0.1)	(0.11)	(0.13)
Cosine Similarity			
Avg Cosine Similarity	0.39	0.43	0.39
	(0.09)	(0.08)	(0.08)
Avg Cosine Similarity, Conservatives	0.35	0.42	0.35
	(0.08)	(0.08)	(0.08)
Avg Cosine Similarity, Socialdemocrats	0.36	0.42	0.39
	(0.07)	(0.07)	(0.07)
Manifesto Closeness			
Avg Manifesto Closeness	0.37	0.49	0.39
	(0.38)	(0.37)	(0.22)
Avg Manifesto Closeness, Conservatives	0.43	0.66	0.5
	(0.33)	(0.29)	(0.2)
Avg Manifesto Closeness, Socialdemocrats	0.04	0.11	0.24
	(0.11)	(0.16)	(0.13)

Each statistic is divided across into the three different election periods. We show statistics for all MPs and those MPs of the two major parties (Conservatives: CDU/CSU fraction and Socialdemocrats: SPD). The first segment describes the average deviation of politicians from the party majority. The second segment describes the cosine similarity of speeches. The third segment shows the average closeness of speeches to the party manifestos. All outcomes lay between between zero and one. For the roll call deviation a value of one means a deviation from the party majority in all voting. For the cosine similarity a value of 1 means that the speeches of a candidate is the same to party colleagues. For the Manifesto closeness a one and can be interpreted as the predictability by the machine learning method (Logistic Regression). A MP's party is perfectly predictable from his speeches and the closeness to the party manifesto. Standard deviations are shown in parentheses.

Table 3: OLS Winning a District Election on Voting against the party-line.

	(1) Simple	(2) with controls	(3) No selection	(4) with controls
Won district	-0.0222***	0.00993	-0.0120*	0.0185**
	(0.00520)	(0.00625)	(0.00728)	(0.00840)
Observations	1,931	1,865	1,423	1,423
R-squared	0.009	0.111	0.002	0.109
State FE		YES		YES
Party FE		YES		YES
Session FE		YES		YES

Robust standard errors in parenthesis *** p<0.01 ** p<0.05 * p<0.1

Linear Least Square regression. One observation corresponds to a politician within one election period. The dependent variable is percentage of deviations in roll-call votes measured in comparison to the majority vote within the party. Wondistrict is a dummy that takes the value one if a politician won their district. State fixed effects are fixed effects for the 16 German federal states, Party fixed effects for the party of a MP, and Session fixed effects for an election period. Models (1) and (2) are models without selection where two adds the fixed effects as controls. Models (3) and (4) select the sample such that we only consider MPs who either entered the parliament over a party list or entered over a district-election, but would have entered over a party list if they had lost their district.

Table 4: Regression Discontinuity Design: Winning a District Election on voting against the party-line.

	(1) Average BW 5	(2) Linear BW 10	(3) 2nd-Pol BW 40	(4) 3rd-Pol BW 40
Won district	0.0453	0.0592	0.0450*	0.0537
Observations	161	309	931	931
Robust Std. Error	0.0574	0.0581	0.0320	0.0404
Robust p-value	0.120	0.109	0.0940	0.120

Robust standard errors *** p<0.01 ** p<0.05 * p<0.1

Regression Discontinuity Design. One observation corresponds to a politician within one election period. The dependent variable is percentage of deviations in roll-call votes measured in comparison to the majority vote within the party. Model (1) only the average are compared on a bandwidth of 5 percentage points around a winning margin of zero. Model (2) ads the linear margin of the voteshare on a bandwidth of 10 percentage points around zero. Model (3) adds a second order polynomial and increases the bandwidth to 50. Finally Model (4) adds a third order polynomial and increases the margin even further to 50 percentage points.

Table 5: OLS Winning a District Election on the Cosine Distance to Party Average.

	(1) Simple	(2) with controls	(3) No selection	(4) with controls
Won district	-0.0347***	0.00654	-0.0328***	0.00752
dummy	(0.00391)	(0.00434)	(0.00531)	(0.00548)
Observations	1,882	1,818	1,392	1,392
R-squared	0.040	0.315	0.027	0.344
State FE		YES		YES
Party FE		YES		YES
Session FE		YES		YES

Robust standard errors in parenthesis *** p<0.01 ** p<0.05 * p<0.1

Linear Least Square regression. One observation corresponds to a politician within one election period. The dependent variable is the cosine similarity measure of a politician where 1 corresponds two an identical similarity between an individual politicians and other party MPs. Wondistrict is a dummy that takes the value one if a politician won their district. State fixed effects are fixed effects for the 16 German federal states, Party fixed effects for the party of a MP, and Session fixed effects for an election period. Models (1) and (2) are models without selection where two adds the fixed effects as controls. Models (3) and (4) select the sample such that we only consider MPs who either entered the parliament over a party list or entered over a district-election, but would have entered over a party list if they had lost their district.

Table 6: Regression Discontinuity Design: Winning district election on the Cosine Distance to Party Average.

	(1) Average BW 5	(2) Linear BW 10	(3) 2nd-Pol BW 40	(4) 3rd-Pol BW 40
Won district 0.0243	0.0309	0.0266*	0.0445	
Observations	152	296	914	914
Robust Std. Error	0.0264	0.0269	0.0175	0.0207
Robust p-value	0.161	0.204	0.0108	0.0562

Robust standard errors *** p<0.01 ** p<0.05 * p<0.1

Regression Discontinuity Design. One observation corresponds to a politician within one election period. The dependent variable is the cosine similarity measure of a politician where 1 corresponds two an identical similarity between an individual politicians and other party MPs. Model (1) only the average are compared on a bandwidth of 5 percentage points around a winning margin of zero. Model (2) ads the linear margin of the voteshare on a bandwidth of 10 percentage points around zero. Model (3) adds a second order polynomial and increases the bandwidth to 50. Finally Model (4) adds a third order polynomial and increases the margin even further to 50 percentage points.

Table 7: OLS Winning a District Election on Manifesto-closeness of Speeches.

	(1) Simple	(2) with controls	(3) No selection	(4) with controls
Won district	-0.00932	-0.0129	0.0202	-0.00006
dummy	(0.0153)	(0.0145)	(0.0195)	(0.0180)
Observations	1,882	1,818	1,392	1,392
R-squared	0.000	0.386	0.001	0.366
State FE		YES		YES
Party FE		YES		YES
Session FE		YES		YES

Robust standard errors in parenthesis *** p<0.01 ** p<0.05 * p<0.1

Linear Least Square regression. One observation corresponds to a politician within one election period. The dependent variable is the predictability score of a MP to his party manifesto. A one corresponds to a perfect predictability. Wondistrict is a dummy that takes the value one if a politician won their district. State fixed effects are fixed effects for the 16 German federal states, Party fixed effects for the party of a MP, and Session fixed effects for an election period. Models (1) and (2) are models without selection where two adds the fixed effects as controls. Models (3) and (4) select the sample such that we only consider MPs who either entered the parliament over a party list or entered over a district-election, but would have entered over a party list if they had lost their district.

Table 8: Regression Discontinuity Design: Winning a District Election on Manifesto-closeness of Speeches.

	(1) Average BW 5	(2) Linear BW 10	(3) 2nd-Pol BW 40	(4) 3rd-Pol BW 40
Won district	-0.0527	-0.0744	-0.0708	-0.0610
Observations	152	296	914	914
Robust Std. Error	0.101	0.104	0.0699	0.0826
Robust p-value	0.847	0.864	0.383	0.460

Robust standard errors *** p<0.01 ** p<0.05 * p<0.1

Regression Discontinuity Design. One observation corresponds to a politician within one election period. The dependent variable is the predictability score of a MP to his party manifesto. A one corresponds to a perfect predictability. Model (1) only the average are compared on a bandwidth of 5 percentage points around a winning margin of zero. Model (2) ads the linear margin of the voteshare on a bandwidth of 10 percentage points around zero. Model (3) adds a second order polynomial and increases the bandwidth to 50. Finally Model (4) adds a third order polynomial and increases the margin even further to 50 percentage points.

Table 9: OLS of Winning a District Election on List Position in next Election.

	(1) Simple	(2) with controls	(3) No selection	(4) with controls
Won district	0.0799	-0.344	-0.783	-1.155**
	(0.353)	(0.396)	(0.481)	(0.516)
Observations	1,249	1,249	1,003	1,003
R-squared	0.000	0.044	0.005	0.084
State FE		YES		YES
Party FE		YES		YES
Session FE		YES		YES

Robust standard errors in parenthesis *** p<0.01 ** p<0.05 * p<0.1

Linear Least Square regression. One observation corresponds to a politician within one election period. The dependent variable is the list change of a MP to the next period, i.e. $\Delta_t = P_t - P_{t+1}$. A MP with a positive Δ_t means that a politician is in a better position to get reelected over the list. Wondistrict is a dummy that takes the value one if a politician won their district. State fixed effects are fixed effects for the 16 German federal states, Party fixed effects for the party of a MP, and Session fixed effects for an election period. Models (1) and (2) are models without selection where two adds the fixed effects as controls. Models (3) and (4) select the sample such that we only consider MPs who either entered the parliament over a party list or entered over a district-election, but would have entered over a party list if they had lost their district.

Table 10: RDD of Winning a District Election on List Position in next Election.

	(1) Average BW 5	(2) Linear BW 10	(3) 2nd-Pol BW 50	(4) 3rd-Pol BW 50
Won district	0.0857	-0.0252	0.432	0.405
Observations	173	327	424	940
Robust Std. Error	1.963	1.999	1.292	1.409
Robust p-value	0.786	0.767	0.653	0.569

Robust standard errors *** p<0.01 ** p<0.05 * p<0.1

Regression Discontinuity Design. One observation corresponds to a politician within one election period. The dependent variable is the list change of a MP to the next period, i.e. $\Delta_t = P_t - P_{t+1}$. A MP with a positive Δ_t means that a politician is in a better position to get reelected over the list. Model (1) only the average are compared on a bandwidth of 5 percentage points around a winning margin of zero. Model (2) ads the linear margin of the voteshare on a bandwidth of 10 percentage points around zero. Model (3) adds a second order polynomial and increases the bandwidth to 50. Finally Model (4) adds a third order polynomial and increases the margin even further to 50 percentage points.

	Margin	Roll Call Dev.	Cosine Sim	Menifesto Closeness
Pop. density (Persons per sqm)	-0.215***	0.044	-0.011	-0.066*
Share male	0.250***	-0.102***	-0.006	0.049
Share German citizens	0.137***	-0.054*	0.000	0.065*
Share older than 60 years	0.018	-0.077**	-0.036	0.072**
Cars per thousand pop.	0.382***	-0.005	-0.034	0.070*
Unemployment share	-0.326***	0.090***	-0.142***	-0.170***
Share without secondary degree	-0.184***	0.099***	-0.078*	-0.125***

Each row corresponds to different socioeconomic characteristics of a district. One observation is the politician and the district of nomination. We calculate the Pearson correlation coefficients to the absolute marginal vote share, the roll call voting deviation, the cosine similarity, the closeness to the manifestos predicted by the multinomial logistic regression.

Table 11: OLS of Winning a District Vote on Manifesto-closeness (Robustness).

| | (1) | (2) | (3) | (4) |
	Simple	with controls	No selection	with controls
Won district	0.149***	0.00516	0.162***	0.00961
dummy	(0.0118)	(0.0138)	(0.0147)	(0.0168)
Observations	1,882	1,818	1,392	1,392
R-squared	0.076	0.281	0.068	0.283
State FE		YES		YES
Party FE		YES		YES
Session FE		YES		YES

Robust standard errors in parenthesis *** p<0.01 ** p<0.05 * p<0.1

Linear Least Square regression. One observation corresponds to a politician within one election period. The dependent variable is the predictability score of a MP to his party manifesto. A one corresponds to a perfect predictability. Wondistrict is a dummy that takes the value one if a politician won their district. State fixed effects are fixed effects for the 16 German federal states, Party fixed effects for the party of a MP, and Session fixed effects for an election period. Models (1) and (2) are models without selection where two adds the fixed effects as controls. Models (3) and (4) select the sample such that we only consider MPs who either entered the parliament over a party list or entered over a district-election, but would have entered over a party list if they had lost their district.

Table 12: RDD of Winning a District Vote on Manifesto-closeness (Robustness).

	(1) Average BW 5	(2) Linear BW 10	(3) 2nd-Pol BW 40	(4) 3rd-Pol BW 40
Won district	0.0462	0.0641	0.00985	0.0434
Observations	152	296	914	914
Robust Std. Error	0.103	0.105	0.0627	0.0765
Robust p-value	0.203	0.208	0.490	0.241

Robust standard errors *** p<0.01 ** p<0.05 * p<0.1

Regression Discontinuity Design. One observation corresponds to a politician within one election period. The dependent variable is the predictability score of a MP to his party manifesto. A one corresponds to a perfect predictability. Model (1) only the average are compared on a bandwidth of 5 percentage points around a winning margin of zero. Model (2) ads the linear margin of the voteshare on a bandwidth of 10 percentage points around zero. Model (3) adds a second order polynomial and increases the bandwidth to 50. Finally Model (4) adds a third order polynomial and increases the margin even further to 50 percentage points.

Appendix	Per 16	Per 17	Per 18
List Place Change	0.84	0.38	1.33
	(5.31)	(4.79)	(4.95)
List Place Change, Conservatives	1.32	0.61	2.52
	(5.92)	(6.25)	(5.75)
List Place Change, Socialdemocrats	1.02	0.3	0.57
	(6.69)	(6.19)	(5.04)
District MP, List Place Change	1.06	0.25	1.33
	(7.54)	(7.08)	(6.58)
District MP, List Place Change, Conservatives	1.55	0.64	2.08
	(7.5)	(6.71)	(6.22)
No of Distict MPs, Socialdemocrats	0.83	-0.79	-1.16
	(7.7)	(8.79)	(7.45)

Each statistic is divided across into the three different election periods. We explore the average of the list place changes defined as $\Delta_t = P_t - P_{t+1}$ such that a positive value means that a politicians increase its list position. We show statistics for all MPs and those MPs of the two major parties (Conservatives: CDU/CSU fraction and Socialdemocrats: SPD). In the second part of the table we restrict the sample to those MPs that are elected through a district. Standard deviations are reported in parentheses.

Table 13: Summary Statistics of the Socioeconomic Statistics

	Mean	SD
Population density (Persons per sqm)	887.04	1413.81
Share Male	0.49	0.01
Share German citizens	0.91	0.05
Share older than 60 years	0.10	0.02
Cars per thousand inhabitants	628.73	94.06
Unemployment share, percent	8.94	4.46
Share without secondary degree, percent	7.61	2.51

The summary statistics shows the averages and standard deviations of socioeconomic variables across all districts across the three election periods.