1-2022

Does a district mandate matter for the behavior of politicians? An analysis of roll-call votes and parliamentary speeches

Andreas BORN
Spotify, Sweden

Aljoscha JANSSEN
Singapore Management University, ajanssen@smu.edu.sg

Citation

BORN, Andreas and JANSSEN, Aljoscha. Does a district mandate matter for the behavior of politicians? An analysis of roll-call votes and parliamentary speeches. (2022). *European Journal of Political Economy*. 71, 1-28.

Available at: https://ink.library.smu.edu.sg/soe_research/2532

European Journal of Political Economy xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Political Economy

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

Does a district mandate matter for the behavior of politicians? An analysis of roll-call votes and parliamentary speeches ☆

Andreas Born [a], Aljoscha Janssen [b,c,*]

[a] Spotify, Sweden
[b] Singapore Management University, School of Economics, 90 Stamford Road, 17890, Singapore
[c] IFN, Sweden

ARTICLE INFO

Keywords:
Party line
Textual analysis
Regression discontinuity
Parliamentary speeches
Voting

ABSTRACT

In most democracies, members of parliament (MPs) are elected either through a party list or by a district. We use a discontinuity in the German electoral system to investigate the causal effect of a district election on an MP's conformity with the party line. A district election does not affect roll-call voting behavior causally, possibly due to overall high adherence to party-line voting. Analyzing the parliamentary speeches of each MP allows us to overcome the high party-line 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 the party line. 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 candidates rather than changing MPs' behavior.

1. Introduction

Plurality voting and proportional representation are the two most common electoral systems through which western democracies determine their members of parliament (MPs). The choice of system is relevant not only for political process but also for its influence on the economic outcomes of the country (Lizzeri and Persico, 2001; Persson et al., 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 et al., 2002; Mueller, 1997; Persson and Tabellini, 1999, 2004). In particular, previous literature predicts that district-elected MPs act in favor of their district whereas MPs elected on a closed party list are more closely aligned to the general interest of a party (Carey, 2007; Depauw and Martin, 2009).[1]

This article examines if the behavior of district-elected and party-list-elected MPs differs. In a mixed-member proportional system, MPs who are elected through a party list coexist with MPs who are elected via a district vote. 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) but enter parliament through only one of these. This allows us to employ a regression discontinuity design to quantify the causal effect of a district

☆ We thank Yan Chen, Richard Friberg, Erik Lindqvist, Jaakko Meriläinen, Torsten Persson, and David Strömberg as well as seminar participants at the Stockholm School of Economics for helpful comments. We also thank *Abgeordnetenwatch.de* for the data access. Financial support from the Jan Wallander and Tom Hedelius Foundation, Sweden is gratefully acknowledged.

* Corresponding author at: Singapore Management University, School of Economics, 90 Stamford Road, 17890, Singapore.
E-mail addresses: aborn@spotify.com (A. Born), ajanssen@smu.edu.sg (A. Janssen).

[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 et al. (2001) and Acemoglu et al. (2001).

https://doi.org/10.1016/j.ejpoleco.2021.102070
Received 4 June 2020; Received in revised form 22 June 2021; Accepted 25 June 2021
Available online 29 June 2021

Please cite this article as: Andreas Born, Aljoscha Janssen, *European Journal of Political Economy*, https://doi.org/10.1016/j.ejpoleco.2021.102070

election on the conformity to the party line in roll-call elections as well as parliamentary speeches. Considering parliamentary speeches, we use text analysis and machine learning to investigate whether 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 on 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 other speeches of the party.

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 do 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 are limited, the majority of votes are not roll-call votes, and party discipline is very high.[2] 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 with MPs who marginally lost a district and enter parliament over a party list. Neither do they use wording closer or further away from the party's. Thus, 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 being further away from the speeches of that party or the party's manifesto. This result is robust to several specifications of the discontinuity regression.

It may be possible that our non-result of differences in speeches is due to a selective choice of parliamentary speakers. Indeed, Proksch and Slapin (2012) show that the party leadership controls speakers. Also, Bäck and Debus (2017) show that parties selectively exclude speakers on economic topics from districts with worse economic situations. The selection of speakers could ensure a speech close to the party line, thereby explaining the similar speeches. That concern is somewhat dispelled when considering that a deviation in a speech can be much less obvious than a deviation in voting. A deviation in speech can be done, for example, by carefully selecting the topic to speak on or by avoiding a contentious issue by omission. On the empirical side, deviations in voting and speeches do correlate with economic and social characteristics of the district in which the politician was elected. Thus speeches do differ systematically, which suggests that they are informative about the policy position of the MP. In particular, the share of unemployed people and the share of inhabitants without a secondary degree correlate highly positively with roll-call vote deviations and negatively with the similarity of speeches to party speeches. Note that our observation does not necessarily contradict the finding of Bäck and Debus (2017), as we do not incorporate the speaking time or frequency of speeches. We conclude that there is no causal effect of a district election on an MP's voting or speeches; instead, our results suggest that different districts elect politicians who are already different before their election.

We rule out a possible alternative explanation for the non-result that could confound our analysis, which could be that district-elected MPs experience faster career progression (Folke et al., 2016; Meriläinen and Tukiainen, 2018) within a party, which incentivizes conformity with the party line.[3] This might offset the effect of a district mandate at the margin.[4] We explore this relationship by investigating whether 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).[5] The intuition is that candidates who follow a dual candidacy strategy face two objectives. On the one hand, politicians face a 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.[6] A politician in a district has to decide between focusing on the median voter in the district, which may increase the probability of election, or focusing on the median voter on the national level to maximize within-party career progression and achieve a higher position on the list. 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 observe neither 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 et al. (2013), Ohmura (2014), and Sieberer (2015) explore relations between list and direct elected MPs and voting behavior.[7] Sieberer (2010) compares roll-call voting of district and list MPs

[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 positively or negatively.

[3] Relatedly, Maaser and Stratmann (2018) show that the mandate type affects an MP's sorting in political committees.

[4] The question of rank effects has also been discussed in political science; see Crisp et al. (2013), Anagol and Fujiwara (2016), or André et al. (2017).

[5] 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; see 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 of this literature. Our framework adapts the model of Kittsteiner and Eyster (2007) to the particular situation in the German parliament.

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

[7] Note that also other differences between list and district-elected MPs have been studied. For example, Potrafke et al. (2020) show that district-elected MPs are on average more physically attractive.

between 2005 and 2009. The author finds that district MPs are significantly more likely to deviate from a party line than MPs who are elected via a party list. Neuhäuser et al. (2013) confirm the findings of Sieberer (2010) 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' probability of reelection 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 reservations about the party line. However, Sieberer (2015) also concludes that dual candidates (MPs who were candidates on the list as well as in a district) do not differ in their explanation of votes. Ohmura (2014) challenges Sieberer (2010) and Neuhäuser et al. (2013) by investigating not only how district and list MPs differ but also whether 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 rate of deviation from the party line. However, Ohmura (2014) shows that pure district MPs are more likely to deviate from the party line. She argues that the candidacy strategy (i.e., pure district or dual) and reelection probability, instead of the mandate itself, determine the voting of MPs. Sieberer (2014) confirms these results by analyzing deviations from the party line by MPs with a dual candidacy strategy between 1949 and 2013. Recently, Sieberer and Ohmura (2019) use roll-call votes between 1949 and 2013 to show that MPs are more likely to deviate from a party line if they rely on a district. Thereby the authors support evidence from Stoffel (2014).

Our research extends the literature in several dimensions. We employ a discontinuity design to evaluate the causal effect of mandate type on the deviation from the party line in voting and speeches in parliament. Thereby, we add to recent work in political science that uses the regression discontinuity design in election outcomes to establish causality.[8] The regression discontinuity design allows us to draw conclusions about the causal effect of a district election. We establish that being voted into parliament by a district instead of via a party list has no causal effect on voting in roll-call votes.

Due to our discontinuity design, we analyze roll-call votes and speeches of MPs that win or lose a district marginally but are elected through the party list. Thus, we are testing a generic effect of winning a district on the behavior of MPs, i.e., the effect of winning a district rather than being elected from the list, independent of overall electoral marginality. We do not contradict previous findings, as our sample of interest does not include MPs with a large vote margin in a district and those dependent on winning a district. However, our analysis shows that we do not find a generic district effect.

An additional contribution is that we analyze parliamentary speeches in Germany. These can potentially inform about MPs' 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 et al., 2018). The use of speeches also addresses the concern that roll-call votes are a selected sample of votes (Carrubba et al., 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) and Proksch and Slapin (2012) underline the importance of speeches for MPs and political parties. However, also, speeches are not an unbiased overview of all preferences within parties. Party leaders have an incentive to protect the party label and therefore control speakers and individual speeches (Proksch and Slapin, 2012). Within Germany, speeches have been analyzed in political science and computer science literature. Bernauer and Bräuninger (2009) show that there is substantial intra-party heterogeneity in parliamentary speeches. Proksch et al. (2019) provide a sentiment approach that recovers government-opposition dynamics. Similar to our study, Bäck and Debus (2017) compare district-elected MPs' speeches. They find that district-elected MPs deliver fewer speeches in economic debates if the economic situation in their district is worse. We therefore add to the literature of German parliamentary speeches by connecting two approaches. First, we use speeches as a measure of intra-party differences. Second, we use the measure as an outcome variable within a discontinuity design to evaluate whether a district election affects conformity.

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 et al. (2010). Note especially Gentzkow et al. (2017) for a review of methodologies and different usages in economics.[9] Using these mechanisms provides us with a objective way of comparing speeches rather than an approach that involves hand-coding speeches according to some criteria. Moreover, this method can readily be extended to new data.

We use approaches from computational linguistics and computer science that examine text documents such as speeches to analyze MPs' positions. Our textual analysis is related to previous work from Laver et al. (2003) and Slapin and Proksch (2008), who estimate political positions using word frequencies in party manifestos. Further, Biessmann et al. (2016), Diermeier et al. (2012), and Peterson and Spirling (2018) use machine learning methods to predict party affiliation. In this paper we make use of term frequency in MPs' speeches to estimate the distance measure between speeches. Further, we use machine learning methods to evaluate how well one can predict party positions from speeches. In comparison to previous literature, we use the estimates as an outcome within our regression discontinuity analysis. The connection of machine learning and methods of causal inference is part of an emerging econometric literature (see, e.g., Athey, 2018, as well as Mullainathan and Spiess, 2017).

Our study also relates to the literature on the effect of political competition. Examples include Strömberg (2008), who evaluates how competition affects campaign spending in U.S. presidential elections, and Dal Bó et al. (2009), who show 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 from parliament beforehand. We show that MPs' competition not only in their district but also within their party may play an important role.

[8] See, for example, Hyytinen et al. (2018) for an overview of recent approaches using regression discontinuity design as well as an empirical test. In the German context, Ade et al. (2014) and Freier (2015) study the incumbency effects using a regression discontinuity.

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

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 and our 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 we rationalize the results in a model in Section 8. Section 9 concludes.

2. Institutional background

The German federal parliament is a mixed-member proportional 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, meaning that for each district one politician enters parliament.

The distribution of the current seats/MPs 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.[10] The direct mandate replaces a list mandate that was determined by the party lists and second votes.[11]

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 candidates as well as direct candidates. Given the political landscape in Germany during our period of study, in most cases the social democratic (SPD) and conservative (CDU/CSU) candidates have a realistic chance to win a district.[12] In some states, such as Bavaria, MPs of the leading party can only enter the parliament through the district as the number of districts exceeds the share of seats allowed by second votes. Appointment of candidates on the lists and in the district is subject to rules. The party lists for each state are 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 not formally bound by a specific party line, the mixed-member proportional representation has the potential to create an alternative incentive structure for direct MPs. Direct candidates may therefore represent the interests of a district, which may deviate from the interest of a nationwide party line (see for example Scarrow, 2001). Nevertheless, the fact that the majority of MPs are dual candidates (list as well as district) means that candidates defeated within a district do not necessarily lose their membership within the parliament.

3. Predictions

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

We start with the analysis of roll-call votes. Previous literature shows (1) that pure district candidates are more likely to deviate from the party line (Ohmura, 2014), (2) that dual candidates (MPs who were candidates on a party list and in a district) do not differ from other candidates in roll-call votes (Sieberer, 2015), and (3) that MPs who depend on district election deviate from the party line in roll-call votes (Sieberer and Ohmura, 2019). We want to test whether we observe a generic effect of winning a district. To establish causality, we employ a regression discontinuity design. Thus, we compare the roll-call votes of MPs who won a district marginally but would have entered the parliament through a party list if they had lost to those who marginally lost a district and entered the parliament through a party list.

The mandate effect could potentially lead in both directions. On the one hand, a district candidate could represent the interests of a district and cultivate a personal vote. Thereby, a district-elected candidate could deviate more often from the party line. Indeed such an effect would align with the intention of a direct mandate, which allows an MP to act more independently from their party. On the other hand, there is also the possibility that list MPs are more likely to deviate from the party line. For example, list MPs could see themselves as responsible for the party's functional interests, such as a stronger orientation toward interest groups within the party (Klingemann and Wessels, 2003), which could lead to more deviation from the party average.

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

[10] In detail, the distribution of seats in parliament is determined by the Sainte–Laguë 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 of seats to reduce the possibility of strategic voting, the size of the German parliament may vary between election periods.

[11] Should a party win more direct seats than party seats, other parties receive seats that compensate ("Ausgleichsmandate") for the overhang seats of the large parties, keeping representation relative to the parties' vote shares.

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

Besides roll-call votes, we use parliamentary speeches. These speeches may differ from roll-call votes for several reasons. First of all, the roll-call votes available are a selection of all votes in the parliament. Roll-call voting takes place if sufficiently many MPs (or a party group) request the procedure (Sieberer et al., 2020). Therefore, it is possible that roll-call votes only take place for issues where deviation from a party line is less or more likely (Ainsley et al., 2020).[13] We do not observe votes that are not roll-call votes. Further, in most parliamentary systems, including the German one, votes are to some extent controlled by the party leadership, including internal party test votes or previous reporting to party leadership in case of deviation (Laver et al., 2003; Proksch and Slapin, 2010; Schwarz et al., 2015; Stecker, 2015). In short, roll-call votes are subject to strict party discipline.

We argue that speeches can serve as a more sensitive measure of individual MPs' deviation from their party line than votes. Unlike in votes, deviation in speeches is not necessarily clear cut. For example, MPs could omit mentioning party-specific terms, emphasize a position favorable to their district, or express views that are not formally the party's, all while avoiding being in direct conflict with their party's position. The topic an MP chooses to give a speech about is already the first step in avoiding or highlighting a specific aspect of a party's position. In a nutshell, speeches give MPs far more possibilities to deviate to varying degrees from the party line, to show their preferences or signal their position to their districts, without potential punishment from the party (Schwarz et al., 2015). Despite the advantage that speeches offer a more sensitive measure, we acknowledge that speeches also suffer from a selection effect. Proksch and Slapin (2012) show that party leaders in Germany have an incentive to protect the party label. In Germany, speakers can be selected according to the preferences of party leadership. Further, punishment for speeches deviating from the party line is possible. Even though speeches could still suffer from a bias, they offer a more sensitive measure.[14] We therefore also evaluate the same hypothesis using speeches.

Besides, we acknowledge that roll-call votes and speeches may address different issues. For example, roll-call votes often concern foreign policy or immigration, while speeches relate to environmental policy. In this case, both measures would be unrelated.

Hypothesis 2. The speeches of members of parliament elected through a district vote differ from the speeches of MPs elected through a party list. The wording of the speeches has (i) a higher cosine distance from the speeches of their party group members and (ii) a lower level of predictability as speeches of their party.

Hypothesis 3 investigates a potential confounding effect, namely that politicians who win a district gain influence in their party, possibly inducing them to act more in accordance with their party line.

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 with politicians losing a district election marginally.

4. Data

We connect four data sources that 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. First, we received data about the MPs and their direct vote shares within a district if applicable from a German nongovernmental organization (*Abgeordnetenwatch.de*). Second, we use data from MPs' roll-call voting, which is publicly available from the German parliament (*Bundestag*).[15] We connect the data with the list positions from elected and non-elected candidates and with socioeconomic statistics on the district level, which we received from the federal electoral management body (*Bundeswahlleiter*).[16] Third, we receive party manifests for all parties represented by MPs for the three legislation periods from Lehmann et al. (2017). Finally, we use parliament protocols from the German parliament (*Bundestag*).

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 responsible for 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.[17] To convert text into a quantifiable measure, for each sentence we separate all words spoken. We stem and lemmatize each word to obtain the word's root form and then exclude most common words of the German language (commonly called stopwords in the literature). For each text, we use the stemmed words to create a vector showing the count of words in the text.[18] We use these vectors to build three measures of

[13] MPs could request voting in cases with especially high or especially low support of all parliamentary members. International literature shows a selection bias in roll-call votes (Carrubba et al., 2006; Hug, 2010; VanDoren, 1990).

[14] In the Appendix, we compare speeches and roll-call votes and show that speeches offer more variety around the party line.

[15] Note that we solely consider roll-call voting. Other votes are not observed. As roll-call votes take place only if requested by a sufficiently large group of MPs, those votes account for only a small number of votes in total (Sieberer et al., 2020).

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

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

[18] This results in a vector whose length equals the total number of distinct words occurring in any speech. Naturally, for a single speech most entries in such a vector are zero.

similarity of the MPs' speeches to those of their party. Two of these we present here; the third we present as a robustness check in the Appendix.

Our first approach is to determine how similar the words used in an MP's speeches are to the words used in the other speeches of the party in terms of the distance between the words used. Specifically, 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 uses a predictive approach instead. For each MP we train a predictor on all other speeches available in the respective parliamentary session, to predict which party an MP is aligned with. The reasoning here is that an MP whose own party can be predicted with a high likelihood is more aligned with their own party, whereas a low likelihood signals an emphasis on topics and positions of the other party — hence a larger deviation from their own party. As a robustness check, we measure how close a member of parliament's speeches align to the party's official policy position as measured by the party's election manifesto. Here, we use the party manifesto instead of other speeches within the party to build a measure of how well an MP's party affiliation can be predicted from the MP's speeches. We elaborate more on this measure in the Appendix.

The first measure uses the vectors of stemmed and filtered words and weights 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 of decreasing the relevance of words used by almost all MPs as well as words that are very rarely used, 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. Based on these we also 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). This helps to capture negations, for example.

To get a measure of the distance of an MP's speech from 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 in an election period. The cosine similarity is then defined as

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

which ranges from 0 to 1. A higher value refers to a higher similarity between the two documents. Given our data, we calculate the cosine similarity of a MP in an election period to all other MPs within the same party and the same election period. We then 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 refer to higher similarity to the speeches of all other party members.

To create the second measure, we train a classifier on the sentences of all speeches in that legislative session, except those of the individual MP to assess. We use the trained classifiers to predict which party the MP belongs to, based on their speeches. We use the predicted probability of the party the MP actually belongs to as a measure of the closeness of this MP's speeches to the party's speeches, relative to the other party's speeches. The classifier we use has to be able to solve a multinomial prediction problem with discrete features. In addition, we need a classifier that yields not only predictions but a probability score that takes values from 0 to 1, where 1 is the highest similarity (100% predicted probability) and 0 the lowest similarity (0% predicted probability). We found that an L2-regularized logistic regression does an excellent job as a classifier in this situation.[19] This is likely due to the sparseness and high dimensionality of the data, as well as the fact that the data are used for predictions. Moreover, some of these other methods require larger amounts of data to work well. We choose the hyper-parameters to maximize the accuracy of the predictions. We provide technical details of the prediction problem and our solution, as well as an assessment of the most predictive words, in the Appendix. With this method, we can predict most affiliations correctly, which makes us confident that the model is reasonable and performs well.

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 periods, where the 16th legislature period covers the months between October 2005 and October 2009, the 17th period covers the months between October 2009 and October 2013, and the 18th period covers the months between October 2013 and October 2017. The two big parties, the conservatives (CDU/CSU) and social democrats (SPD), account for the majority of MPs in the German parliament. These two parties win almost all direct mandates.

Table 2 presents the variables describing MPs' 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 from a vote. Furthermore, a MP has the option not to attend a vote within the parliament. We do not consider MPs who are not attending, and we treat these observations as missing values.[20] In order to evaluate the party line, we take the majority vote (agreement, refusal, or abstention) of a party. In cases where an MP votes against the majority of a party, the MP is counted as deviating from the party line.[21]

[19] It does much better than classifiers based on linear regressions, random forests, vector support machines, and gradient boosted trees. It also does better than a dense neural network. For the reasons of higher simplicity and higher ease of interpretation, we prefer a logistic regression.

[20] However, we acknowledge that non-attendance may not be random. Bernecker (2014) shows that MPs facing a secure reelection show less presence in parliament. Gavoille (2018) shows similar results for the French parliament. Arnold et al. (2014) show that absence is not related to outside earnings.

[21] Note that we follow the literature that describes abstention as a weak form of disagreement with the party line; see Becher and Sieberer (2008). In the Appendix we show robustness when abstentions are not treated as weak deviations from the party line when the majority of the party chooses agreement or refusal.

Table 1
Summary statistics of parliaments and their members.

Summary variables	Date of German Federal Election		
	2005/09	2009/09	2013/09
No. of MPs	640	653	658
No. of MPs, Conservatives	238	245	323
No. of MPs, Social democrats	228	154	205
No. of District MPs	299	299	299
No. of District MPs, Conservatives	150	218	236
No. of District MPs, Social democrats	145	64	58
Avg. District Vote Share, Conservatives	43.31	41.37	45.72
	(9.96)	(7.35)	(8.33)
Avg. District Vote Share, Social democrats	41.1	31.69	32.26
	(8.77)	(8.62)	(8.94)

Each statistic is divided across the three different election periods. We show statistics for all MPs and for MPs of the two major parties (conservative CDU/CSU faction and social democratic SPD). Note that the number of MPs is fixed; however, some MPs exit and others 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 MPs' parliamentary behavior.

Summary variables	Date of German Federal Election		
	2005/09	2009/09	2013/09
Roll call deviation			
Avg. Voting deviation	0.14	0.1	0.11
	(0.11)	(0.1)	(0.12)
Avg. Voting deviation, Conservatives	0.16	0.06	0.09
	(0.11)	(0.08)	(0.1)
Avg. Voting deviation, Social democrats	0.15	0.14	0.13
	(0.1)	(0.11)	(0.13)
Cosine similarity			
Avg. Cosine similarity	0.39	0.42	0.38
	(0.09)	(0.07)	(0.08)
Avg. Cosine similarity, Conservatives	0.35	0.41	0.35
	(0.08)	(0.08)	(0.08)
Avg. Cosine similarity, Social democrats	0.36	0.42	0.39
	(0.07)	(0.07)	(0.07)
Predictability measure			
Avg. Predictability measure	0.62	0.63	0.7
	(0.26)	(0.28)	(0.26)
Avg. Predictability measure, Conservatives	0.76	0.79	0.81
	(0.19)	(0.19)	(0.18)
Avg. Predictability measure, Social democrats	0.54	0.64	0.61
	(0.27)	(0.22)	(0.27)

Each statistic is divided across the three different election periods. We show statistics for all MPs and those MPs of the two major parties (conservative CDU/CSU faction and social democratic 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 predictability of party affiliation. All outcomes lie between zero and one. For the roll-call deviation, a value of 1 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 are the same as those of party colleagues. For the predictability measure, a value of 1 can be interpreted as meaning that an MP's party is perfectly predictable from the MP's speeches. Standard deviations are shown in parentheses.

The second segment describes the average similarity of politician's speeches to those of other party members, measured by the cosine similarity. A higher value describes a higher similarity. The third segment shows the average predictability measure created by predicting the party affiliation of a politician based on their speeches.

6. Identification strategy

To provide initial 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 through a district election or through a list. This is meant to give preliminary 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 1 when a politician won a district, $Z_{p,i,t}$ is a vector of control variables, and $\varepsilon_{p,i,t}$ is the error term. The basic least squares regression may result in a biased estimator for several

reasons. First, MPs who won a district election with a high margin may differ systematically from 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 characteristics 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.[22] We denote this variable as $m_{p,i,t}$, where p is the member of parliament, i is the district, and t is 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 that they would have entered over the list if they had not won their district. This restriction ensures that we avoid any selection bias that could arise from observing only winners and not losers of an election.

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

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

where $y_{p,i,t}$ is the outcome variable, $d^{won}_{p,i,t}$ is a dummy for 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. First, we include only those MPs who entered or could have entered the parliament through their party list. Second, we limit the regression to observations with the forcing variable within a certain margin around zero.

7. Results

7.1. Deviation from party-line voting

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

Specification (1) in Table 3 shows the results of a naive regression as specified in Eq. (1). Specifications (3) and (4) limit the sample to observations from members of parliament who either entered on a party list or would have entered on a party list if they had lost their district election. This prevents selection bias, as we cannot observe the (counterfactual) parliamentary votes of candidates who lost a district election and did not enter parliament through a list either. Specifications (2) and (4) add state, party, and session fixed effects as controls. This is relevant as the districts that 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 both includes 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 changes sign once the regression includes fixed effects, yet it is small in all regressions. This suggests care is needed when interpreting the results. Our preferred specification suggests that politicians who receive more votes when they win a district election are 1.85 percentage points more likely to deviate in roll-call votes. As on average 0.1188 of votes deviate from the party line, winning a district increases the probability of deviation by 15.57%.

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 Eq. (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 40.[23] Finally, specification (4) adds a third-order polynomial and increases the margin even further to 40 percentage points.[24] 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.

Notably, the coefficients are positive in all specifications and larger than in the OLS; however, only one is weakly significant. This does not allow us to draw strong conclusions from the specification.[25]

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.

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.*

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

[23] Note that we increase the bandwidth to show robustness to the result in the first model.

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

[25] Fig. F.1 in the Appendix displays the discontinuity graphically, fitted once by a second-order polynomial and once by a third-order polynomial. Visual inspection shows a slight but not very strong jump at the margin.

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	−0.0216***	0.0102	−0.0114	0.0189**
	(0.00520)	(0.00625)	(0.00728)	(0.00840)
Observations	1931	1865	1423	1423
R-squared	0.009	0.110	0.002	0.108
State FE		YES		YES
Party FE		YES		YES
Session FE		YES		YES

Robust standard errors in parentheses *** p<0.01 ** p<0.05 * p<0.1.
Linear least squares regression. One observation corresponds to a politician within one election period. The dependent variable is the fraction of deviations in roll-call votes measured in comparison to the majority vote within the party. $Wondistrict$ is a dummy that takes the value 1 if a politician won their district. $State$ fixed effects are fixed effects for the 16 German federal states, $party$ fixed effects are for the party of an MP, and $session$ fixed effects are for an election period. Models (1) and (2) are models without selection, where model (2) adds the fixed effects as controls. Models (3) and (4) select the sample such that we only consider MPs who either entered the parliament on a party list or entered through a district election but would have entered on 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	0.0448	0.0583	0.0448*	0.0530
Observations	161	309	931	931
Robust Std. Error	0.0575	0.0581	0.0321	0.0405
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 fraction of deviations in roll-call votes measured in comparison to the majority vote within the party. In Model (1) only the average is compared on a bandwidth of 5 percentage points around a winning margin of zero. Model (2) adds the linear margin of the vote share on a bandwidth of 10 percentage points around zero. Model (3) adds a second-order polynomial and increases the bandwidth to 40. Finally, Model (4) adds a third-order polynomial and increases the margin even further to 40 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	−0.0343***	0.00502	−0.0316***	0.00691
	(0.00381)	(0.00427)	(0.00520)	(0.00538)
Observations	1882	1818	1392	1392
R-squared	0.041	0.309	0.027	0.336
State FE		YES		YES
Party FE		YES		YES
Session FE		YES		YES

Robust standard errors in parentheses *** p<0.01 ** p<0.05 * p<0.1.
Linear least squares 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 to an identical similarity between an individual politician and other MPs in the party. $Wondistrict$ is a dummy that takes the value 1 if a politician won their district. $State$ fixed effects are fixed effects for the 16 German federal states, $party$ fixed effects are for the party of a MP, and $session$ fixed effects are for an election period. Models (1) and (2) are models without selection, where model (2) adds the fixed effects as controls. Models (3) and (4) select the sample such that we only consider MPs who either entered the parliament on a party list or entered through a district election but would have entered on a party list if they had lost their district.

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 two described measures: the distance of the words used in the speech from the party average, and the likelihood with which a predictor calibrated on the speeches of all other MPs predicts the speech as coming from the MP's own party. These results are backed up by a robustness check that uses the parties' election manifestos instead of speeches to make the prediction (see the robustness section). Tables 5 and 7 show the results of an OLS regression, and Tables 6 and 8 show the results of the regression discontinuity design.

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	0.0222	0.0288	0.0328**	0.0390
Observations	152	296	820	820
Robust Std. Error	0.0271	0.0276	0.0190	0.0228
Robust *p*-value	0.176	0.219	0.0395	0.155

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 to an identical similarity between an individual politician and other MPs in the party. In Model (1), only the averages are compared on a bandwidth of 5 percentage points around a winning margin of zero. Model (2) adds the linear margin of the vote share on a bandwidth of 10 percentage points around zero. Model (3) adds a second-order polynomial and increases the bandwidth to 40. Finally, Model (4) adds a third-order polynomial and increases the margin even further to 40 percentage points.

Table 7
OLS Winning a district election on predictability of speeches.

	(1) Simple	(2) With controls	(3) No selection	(4) With controls
Won district	0.149***	0.00516	0.162***	0.00961
	(0.0118)	(0.0138)	(0.0147)	(0.0168)
Observations	1882	1818	1392	1392
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 squares regression. One observation corresponds to a politician within one election period. The dependent variable is the predictability score of a MP to belong to their own party. A value of 1 corresponds to a perfect predictability. $Won district$ *is a dummy that takes the value 1 if a politician won their district.* $State$ *fixed effects are fixed effects for the 16 German federal states,* $party$ *are fixed effects for the party of an MP, and* $session$ *are fixed effects for an election period. Models (1) and (2) are models without selection, where model (2) adds the fixed effects as controls. Models (3) and (4) select the sample such that we only consider MPs who either entered the parliament on a party list or entered through a district election but would have entered on a party list if they had lost their district.*

Specification (1) in Tables 5 and 7 shows the results of a completely naive regression as specified in Eq. (1). Specifications (3) and (4) restrict the sample to observations from members of parliament who either entered on a party list or entered through a district election but would have entered on a party list if they had lost their district. This prevents selection bias, as we lack the (counterfactual) votes of candidates who lost in a district election and did not enter parliament through a list either. Specifications (2) and (4) add state, party, and session fixed effects as controls. This is relevant as the district winners from 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 an MP's speeches in parliament with speeches of the MP's own party. However, adding controls increases the coefficients such that they are positive and nonsignificant. Yet the size of the coefficient is small compared to the average within-party similarity (see Table 2).

In contrast, Table 6 shows that, at the margin of just winning or losing a district, there is no significant effect of winning a district on the closeness of an MP's speeches to the speeches of the MP's party peers. The coefficients are all small and positive but 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 elected in a district.

The second measure we investigate is the predictability of speeches, which measures how close an MP's speeches are to their own party's speeches relative to the other parties' speeches.[26] Table 7 shows the identical OLS regressions for the predictability measure. Here specifications (1) and (3) are significantly positive; however, the similarity measure turns nonsignificant after controlling for the party, state, and session in specifications (2) and (4).

[26] Fig. F.2 in the Appendix shows a fitted polynomial and the binned sample average of the speech similarity measure around the voting cutoff. Visual inspection of the figure supports the conclusion that there is no discontinuity. The graph shows a small increase at the cutoff; however, this difference is marginal and well within the variation of the bins.

Table 8
Regression discontinuity design: Winning a district election on predictability of speeches.

	(1) Average BW 5	(2) Linear BW 10	(3) 2nd-Pol BW 40	(4) 3rd-Pol BW 40
Won district	0.0462	0.0641	0.0193	0.0707
Observations	152	296	820	820
Robust Std. Error	0.103	0.105	0.0693	0.0844
Robust *p*-value	0.203	0.208	0.308	0.223

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 an MP to their own party. A value of 1 corresponds to a perfect predictability. In Model (1), only the average is compared on a bandwidth of 5 percentage points around a winning margin of zero. Model (2) adds the linear margin of the vote share on a bandwidth of 10 percentage points around zero. Model (3) adds a second-order polynomial and increases the bandwidth to 40. Finally, Model (4) adds a third-order polynomial and increases the margin even further to 40 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	1249	1249	1003	1003
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 parentheses *** p<0.01 ** p<0.05 * p<0.1.
Linear least squares regression. One observation corresponds to a politician within one election period. The dependent variable is the list position change of an MP to the next period, i.e., $\Delta_t = P_t - P_{t+1}$. A positive Δ_t means that the MP is in a better position to get reelected over the list. $Won district$ is a dummy that takes the value 1 if a politician won their district. $State$ fixed effects are fixed effects for the 16 German federal states, $party$ fixed effects are for the party of a MP, and $session$ fixed effects are for an election period. Models (1) and (2) are models without selection, where model (2) adds the fixed effects as controls. Models (3) and (4) select the sample such that we only consider MPs who either entered the parliament on a party list or entered through a district election but would have entered on a party list if they had lost their district.

Table 8 then shows the result of the regression discontinuity regression for the predictability measure. Here all coefficients are small and nonsignificant. Hence here as well the data suggest that there is no discontinuity in speech behavior according to the predictability measure.[27]

Finding 2. *Entering parliament through a district election instead of on their party's list does not cause MPs to use wording (i) further away from the speeches of their party peers or (ii) harder to classify as belonging to their own party based on the party's speeches relative to the other parties' speeches.*

7.3. Future list position

In this subsection we present the results of the analysis of the effect of winning a district election on the MP's future position on the party list. We define the change of a list position of an MP as the change from the position in election t to the next election, i.e., $\Delta_t = P_t - P_{t+1}$. A positive δ_t means that the MP is in a better position to get reelected on the list.[28] Table 9 presents the results of an ordinary least squares regression, and Table 10 presents the results of the regression discontinuity analysis.

Specification (1) in Table 9 shows the results of a simplistic regression as specified in Eq. (1). Specifications (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 district. This prevents a selection bias as we are lacking 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 that 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.

[27] Fig. F.3 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 cutoff is marginal.

[28] On average, politicians increase their list positions over time; see the summary statistics in Appendix E.

Table 10
Regression discontinuity design of winning a district election on list position in next election.

	(1) Average BW 5	(2) Linear BW 10	(3) 2nd-Pol BW 40	(4) 3rd-Pol BW 40
Won district	0.0141	−0.00518	0.369	0.562
Observations	187	355	450	945
Robust Std. Error	1.808	1.841	1.170	1.436
Robust *p*-value	0.831	0.813	0.627	0.639

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 position change of an MP to the next period, i.e., $\Delta_i = P_i - P_{i+1}$. A positive Δ_i means that an MP is in a better position to get reelected over the list. In Model (1), only the average is compared on a bandwidth of 5 percentage points around a winning margin of zero. Model (2) adds the linear margin of the vote share on a bandwidth of 10 percentage points around zero. Model (3) adds a second-order polynomial and increases the bandwidth to 40. Finally, Model (4) adds a third-order polynomial and increases the margin even further to 40 percentage points.

Investigating the results, we find that the OLS regression does not suggest a strong relationship. The coefficient is nonsignificant 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 nonsignificant, which suggests that no relationship exists between marginally winning or losing a district and the future list position.[29]

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. The following model is based on a median voter model where two politicians compete for a district vote. At the same time, the politicians face the possibility of winning a list position by acting in favor of the national party. The trade-off of winning a district and acting in favor of a party could explain our result.

We acknowledge that the model presents one explanation, which is not necessary unique. Further, the model is based on assumptions. In particular, we model the political competition on the district level using the median voter model. We therefore do not model the representative democracy and parties. Instead, we focus on the district component of the German voting system and model the candidates' incentive to satisfy a party line that is exogenous. Also, when districts are considered, the median voter model is related to important assumptions. We assume that only two candidates run for a district seat, while in reality most parties have a candidate.[30] We defend the assumption with the observation that the majority of voters vote only for the two leading parties when considering the district. Further, we assume that voters base their votes on single issues or a single policy dimension. In addition to competition in a district, we assume that candidates face the possibility of being elected on the party list. In the model, candidates can influence their list position by behaving in accordance with the national party line. Thereby, the model assumes that the position on a list is directly associated with a candidate's behavior in relation to the national party line. Note that the party lists in Germany are decided on the state level rather than the national or district level. While the party line on the state level could deviate from the national party line, our model treats both as equivalent. Thus we assume that national and state party lists are sufficiently similar such that the behavior of a candidate close to a national party line is also related to a better position on the party list created at the state level.

Suppose a country has n districts, indexed by d. Each district is characterized with socioeconomic characteristics C^d. Further, voter i in district d is characterized by a policy preference that can be represented one-dimensionally on the unit interval $x_i^d \in [0, 1]$. Therefore we can place all voter preferences along a one-dimensional political spectrum.[31] 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 is the sum of distributions of all districts, $G(x) = \sum_d F^d(x)$.

[29] Table 10 in the Appendix shows the potential discontinuity graphically, fitted by second- and third-order polynomials. Visual inspection of the figure confirms the result of Table 10 that there exists no discontinuity in the future list position.

[30] Note further that, in reality, the likelihood of winning is often not only a function of politicians' behavior, as shown in this model. Instead, we observe districts where the election of a specific party member is almost certain. We do not model the variation in the likelihood of winning a district. Instead, we solely assume that a voter has certain preferences on a unit interval that could be satisfied by an MP's behavior.

[31] We do not model other aspects of the voters' utility function such as the valence characteristic (Stoffel, 2014).

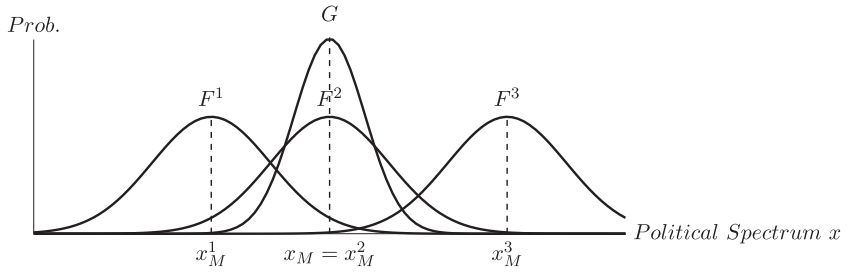

Fig. 1. Example of voter distributions in different districts. *The graph presents an example of possible voter distributions across the political spectrum x. F^1, F^2, and F^3 present three different distributions of voter preferences in districts whereas G presents the distribution on the national level. Note that G is an aggregation, but this graph just exemplifies 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 2.*

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 in a district compete for one seat. We assume that candidates have a dual candidacy strategy, so they compete for election on a closed party list as well as in 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,[32] and hence announce the median voter's preferred policy as their campaign platform.[33]

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 parliament on the party list, such that each politician has to balance two parts of an objective function. In their district, candidates face a distribution of voters' preferences $F^d(x)$ that potentially differs from the national distribution $G(x)$. A candidate's individual district-election probability $p_i^d(\Delta x_i^d)$ is a decreasing function of 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 the party's platform (which we also refer to as the 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 candidate's policy platform and the party line, Δx_i^P. The intuition is that the party determines the candidate's list position and a policy platform closer to the party median is advantageous for an MP's chances of being placed high on the list.[34] In Section 7 we show that winning a district is not itself associated with an increase in position on 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 Fig. 1. Distribution G in the figure depicts the overall distribution of voters X on the national level. The three distributions F^d for $d = \{1, 2, 3\}$ 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 that 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 districts 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 holds for an 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 a trade-off between the median voter of their district and their party line (the national-level median). Candidates choose their party platform so as to maximize the sum of the probability of winning the district-election and the probability of being elected on 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 from choosing a policy platform close to the median policy of the district. Vice versa, the weaker of the two candidates has a higher return from 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 for winners in these districts.

[32] For a general discussion of assumption of the median voter model, see for example Gruber (2005).

[33] In our stylized model politicians and parties always follow through on their campaign platform after the election. Austen-Smith and Banks (1989), Aragonès et al. (2007), Corazzini et al. (2014), and Born (2018), for example, study situations in which promises are costly to break but not completely binding.

[34] Note that the party list in Germany is created at the state level. We assume that the state party line is sufficiently close to the national party line such that the correlation between the closeness of a candidate's behavior to the national party line and a better list position on the state party list is positive.

Table 11
Correlation with socioeconomic characteristics.

	Margin	Roll Call Dev.	Cosine Sim
Pop. density (persons per square mile)	−0.215***	0.044	−0.011
Share male	0.250***	−0.102***	−0.006
Share German citizens	0.137***	−0.054*	0.000
Share older than 60 years	0.018	−0.077**	−0.036
Cars per thousand pop.	0.382***	−0.005	−0.034
Unemployment share	−0.326***	0.090***	−0.142***
Share without secondary degree	−0.184***	0.099***	−0.078*

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, and the cosine similarity of speeches.

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 hypotheses 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 hypotheses, we look at correlations for the sample of politicians of the two major parties, the CDU/CSU (conservatives) and 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 males in the population, the share of German citizens, the share of the population older than 60 years, the unemployment share, the share of cars per 1000 inhabitants, and the share of population without a lowest secondary degree.[35] 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 (see Table 11).

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 the marginal vote share across parties.

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 of the characteristics $C_{d,t}$ are correlated with the marginal vote share. However, the strongest is with the share of cars, the unemployment share, and the share of the population without a secondary degree. In districts with more cars, lower unemployment, a higher share of males, a higher share of German citizens, and lower share of the 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 deviation from 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 with social disadvantages are correlated with deviation in roll-call votes. MPs deviate more from the party line the higher the unemployment share and the higher the share of the population without a secondary degree. Considering the speeches we use the cosine similarity. The higher share of males is correlated with less deviation. Next we evaluate the correlation of the speeches with the district characteristics. A higher unemployment share as well as a higher share of the 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.

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. First, 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 on a party list. Second, we evaluate whether the speeches of these MPs differ more from those of their peers. Third, 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 do not 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 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,

[35] In (Table E.2) we provide some key summary statistics.

Table A.1
OLS Winning a district election on voting against the party line (without abstention).

	(1) Simple	(2) With controls	(3) No selection	(4) With controls
Won district	−0.00639***	0.000441	−0.0102***	−0.00211
	(0.00192)	(0.00225)	(0.00209)	(0.00228)
Observations	1927	1861	1419	1419
R-squared	0.006	0.146	0.015	0.172
State FE		YES		YES
Party FE		YES		YES
Session FE		YES		YES

Robust standard errors in parentheses *** p<0.01 ** p<0.05 * p<0.1.
Linear least squares regression. One observation corresponds to a politician within one election period. The dependent variable is the fraction of deviations in roll-call votes measured in comparison to the majority vote within the party. Within this regression we do not code abstention as deviations. $Won district$ *is a dummy that takes the value 1 if a politician won their district.* $State$ *fixed effects are fixed effects for the 16 German federal states,* $party$ *fixed effects are for the party of a MP, and* $session$ *fixed effects are for an election period. Models (1) and (2) are models without selection, where model (2) adds the fixed effects as controls. Models (3) and (4) select the sample such that we only consider MPs who either entered the parliament on a party list or entered through a district election but would have entered on a party list if they had lost their district.*

even MPs elected by a district, to a strong party discipline as is common in 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 have.

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. An explanation for the absence of the effect in the German system is that here every party sends only one candidate into a district election, whereas an open or semi-open list setting allows competition among members of the same party. Hence, unlike in an open list election, district candidates cannot demonstrate their popularity relative to other members within the same party.

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

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Robustness of roll-call deviation measure

When evaluating the effect of a mandate type on roll-call deviations, we code abstentions (from a vote) of MPs as deviation if the majority of a party does not abstain. We believe that abstention is an important weak deviation from a party line if the majority assents or dissents. Previous studies of roll-call voting indeed often interpret abstention as a deviation from the party line (Becher and Sieberer, 2008). However, in this section we show robustness for two different measures of deviation from the party line: (1) we do not consider abstention as a deviation but as a missing vote, and (2) we only consider assent and dissent votes (“yes” and “no” votes). Both deviation measures do not change the interpretation of the analysis, i.e., we do not find an effect on the margin.

First of all, the average deviation for an MP decreases from 11.9% (when considering abstention as dissent) to 2.8% (when not considering abstention) to 2.1% (when only considering “yes” and “no” votes). We then proceed to redo our analysis, i.e., we use an ordinary least squares analysis as well as the regression discontinuity design. In both cases, we regress the new deviation measures on a dummy indicating whether a member of parliament entered through a district or on a list.

Considering the first measure of not coding an abstention as a dissent, we present the results in Table A.1 for the OLS regression and Table A.2 for the regression discontinuity design. Results are comparable to the one in the main paper. The only difference in the OLS regression is that even with controls and the selection of MPs who entered the parliament on a party list or entered through a district election but would have entered over a party list if they had lost their district, we do not observe a positive effect of a district win on deviations. As in the main paper, the regression discontinuity design does not show any significant coefficient. We observe similar results when considering only “strict deviations”, i.e., a dissent when the majority of a party assents or vice versa (see Tables A.3 and A.4).

Table A.2
Regression discontinuity design: Winning a district election on voting against the party line (without abstention).

	(1) Average BW 5	(2) Linear BW 10	(3) 2nd-Pol BW 40	(4) 3rd-Pol BW 40
Won district	0.0111	0.0171	0.00910	0.0107
Observations	160	308	930	930
Robust Std. Error	0.0299	0.0301	0.0138	0.0187
Robust *p*-value	0.359	0.338	0.438	0.363

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 fraction of deviations in roll-call votes measured in comparison to the majority vote within the party. Within this regression we do not code abstention as deviations. In Model (1), only the average is compared on a bandwidth of 5 percentage points around a winning margin of zero. Model (2) adds the linear margin of the vote share on a bandwidth of 10 percentage points around zero. Model (3) adds a second-order polynomial and increases the bandwidth to 40. Finally, Model (4) adds a third-order polynomial and increases the margin even further to 40 percentage points.

Table A.3
OLS Winning a district election on strict voting against the party line.

	(1) Simple	(2) With controls	(3) No selection	(4) With controls
Won district	−0.00323**	−9.18e−05	−0.00612***	−0.00167
	(0.00163)	(0.00185)	(0.00179)	(0.00198)
Observations	1927	1861	1419	1419
R-squared	0.002	0.132	0.008	0.154
State FE		YES		YES
Party FE		YES		YES
Session FE		YES		YES

Robust standard errors in parentheses *** p<0.01 ** p<0.05 * p<0.1.
Linear least squares regression. One observation corresponds to a politician within one election period. The dependent variable is the fraction of deviations in roll-call votes measured in comparison to the majority vote within the party. Within this regression we only consider yes and no votes. $Wondistrict$ is a dummy that takes the value 1 if a politician won their district. $State$ fixed effects are fixed effects for the 16 German federal states, $party$ fixed effects are for the party of a MP, and $session$ fixed effects are for an election period. Models (1) and (2) are models without selection, where model (2) adds the fixed effects as controls. Models (3) and (4) select the sample such that we only consider MPs who either entered the parliament on a party list or entered through a district election but would have entered on a party list if they had lost their district.

Table A.4
Regression discontinuity design: Winning a district election on strict 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	0.00502	0.00672	0.00501	0.00321
Observations	160	308	930	930
Robust Std. Error	0.0194	0.0195	0.00937	0.0125
Robust *p*-value	0.719	0.610	0.732	0.674

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 fraction of deviations in roll-call votes measured in comparison to the majority vote within the party. Within this regression we only consider yes and no votes. In Model (1) only the average is compared on a bandwidth of 5 percentage points around a winning margin of zero. Model (2) adds the linear margin of the vote share on a bandwidth of 10 percentage points around zero. Model (3) adds a second-order polynomial and increases the bandwidth to 40. Finally, Model (4) adds a third-order polynomial and increases the margin even further to 40 percentage points.

Appendix B. Robustness of cosine similarity

In the following, we turn to the cosine similarity measures and show how the simple similarity measure relates to MPs' roll-call votes. The similarity measure solely uses word frequency to show the distance of an MP's speech to the speeches of other MPs of the same party.

Distribution:

First, we look at the distribution of the cosine similarity. Similar to the roll-call votes, where the majority of MPs never or only seldom deviate from the party line, we expect that the similarity measures follow a skewed distribution, with a large fraction of MPs that show a high similarity and a few MPs' speeches that are very different from the speeches of other party members. We further argue that speeches offer MPs the possibility to please a district without a strong deviation from a party line such as in roll-call

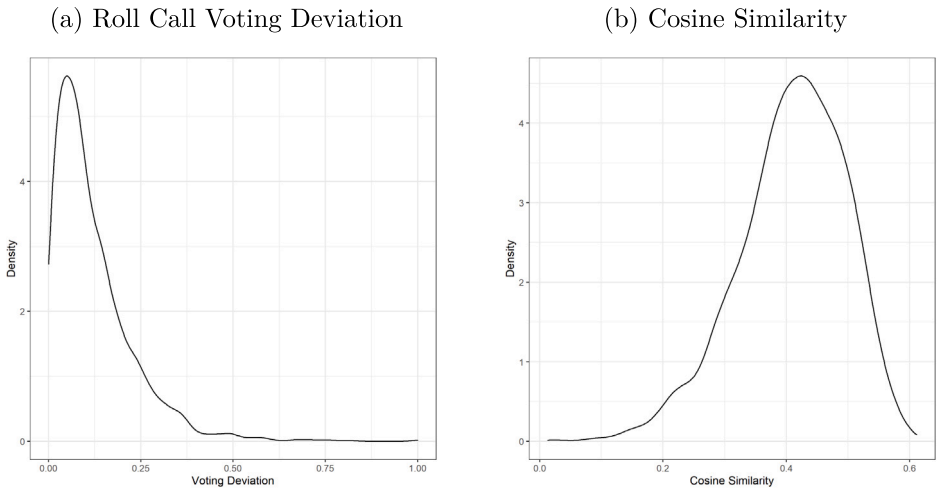

Fig. B.1. Density. *The two figures illustrate two densities. Subfigure (a) shows roll-call voting deviations from the party line of all MPs across all periods. Subfigure (b) presents cosine similarities of speeches (between 0 and 1, with 1 meaning identical speeches) of all MPs across all periods.*

votes. While a deviation from a party line in roll-call votes could result in a less favorable position within a party, an MP has the possibility to choose words and sentences that may deviate slightly from the central party position and thereby satisfy a district. We therefore expect another difference in the distribution. Compared to the distribution of roll-call vote deviations, we expect that a higher share of cosine similarities will deviate from the mean. The intuition is that a higher share of MPs deviate from the party line by choosing their own words.

Fig. B.1 shows the density of roll-call vote deviations (Fig. B.1(a)) as well as speech similarities for all politicians across all periods (Fig. B.1(b)). Both distributions are skewed. Only a few MPs deviate from the party line, and only a few MPs show low similarity in their speeches compared to speeches of fellow party members. We believe that the similarity of both is the first indication that the measure for speeches is suitable. We further observe that the cosine similarity density function is wider than the density function of deviations in roll-call votes, meaning that a higher share of MPs deviate from the mean. This observation satisfies our expectations that MPs' speeches often differ only slightly from those of party members.

Roll-Call Voting and Cosine Similarity:

In a second step, we show details on the correlation between deviations of roll-call votes and the cosine similarity of MPs' speeches. Despite the differences described earlier, we expect that roll-call votes and cosine similarities are correlated. We evaluate the correlation using the following regression model, where i corresponds to an MP, p corresponds to a party, and t corresponds to an election period:

$$Deviation_{p,i,t} = \beta Similarity_{p,i,t} + \rho_p + \xi_t + \gamma_i + \epsilon_{p,i,t},$$

where $Deviation$ is the roll-call deviation and $Similarity$ is the cosine similarity measure. We include different fixed effects. When including party (ρ_p) and period (ξ_t) fixed effects, we compare correlation for MPs of the same party in the same period. When including MP fixed effects (γ_i), we compare deviations in roll-call vote and speech similarity of the same MP across the three different periods.

Table B.1 shows the results of the regression model. Model (1) includes party fixed effects, model (2) adds period fixed effects, and model (3) shows results with MP fixed effects. In all models, we observe a negative correlation between the cosine similarity measure and the roll-call deviations. A higher speech similarity leads to a lower number of deviations in roll-call votes. Interestingly, this correlation is strong when adding MP fixed effects, using the variation only on the MP level across periods. Therefore an MP with more deviation in roll-call votes also uses more different words during speeches. The co-movement across periods for a specific MP is a further indicator that validates the measure of the cosine similarity.

Cosine Similarity to Speeches of Different Parties:

We extend the approach by evaluating the cosine similarities of MPs' speeches to the speeches of MPs by different parties. Therefore, we extend the cosine similarities and do not only evaluate the average cosine similarity to the MP's own party members but also build a measure in relation to party members for each competing party. Note that the cosine similarities to the MP's own and different parties are not directly comparable as the measure is dependent on the absolute number of words and corresponding frequencies. Therefore, it is possible that the cosine similarities of a conservative candidate to conservative party members are smaller than the similarities to green party members as the conservative party members use more unique words. However, we can use the variation across members of the same party, periods, and MP to examine if MPs have more distance to the speeches of their own party when having a higher cosine similarity to other parties. Further, we use the same variation to evaluate if a higher roll-call deviation is related to a higher cosine similarity to other parties' speeches.

Table B.1
Correlation of roll call deviations and cosine similarity.

	Deviations in roll call voting		
	(1)	(2)	(3)
Cosine similarity	−0.134 ***	−0.107 **	−0.209 ***
	(0.014)	(0.044)	(0.054)
Constant	0.132***		
	(0.005)		
Party FE	Yes	Yes	No
Period FE	No	Yes	No
MP FE	No	No	Yes
N	10,030	10,030	10,030
R^2	0.078	0.096	0.817

Robust standard errors in parentheses *** p<0.01 ** p<0.05 * p<0.1.
The table shows regression results of deviations in roll-call voting and the cosine similarity of MPs' speeches to party members. One observation corresponds to an MP within a period. The outcome variable is the deviation from the party line in roll-call voting within a period. The cosine similarity is the euclidean distance measure that shows the average distance of words used compared to other members of a party. We indicate if party, period, or MP fixed effects are included.

In the initial step, we compare if the cosine similarity to the MP's own party is correlated with cosine similarities of competing parties and/or parties of the coalition. We hypothesize that MPs with speeches closer to competing parties should have lower cosine similarities to their own party. We evaluate the relations across the conservative and social democratic parties separately.[36] Considering one party at a time, we evaluate the following regression model:

$$Similarity_{p,i,t} = \beta \mathbf{Similarity}_{-\mathbf{p,i,t}} + \varepsilon_{p,i,t},$$

where $Similarity_{p,i,t}$ is the cosine similarity measure of speeches of politician i to speeches of members of the MP's own party p in period t. $\mathbf{Similarity}_{-\mathbf{p,i,t}}$ is a vector of cosine similarities to other parties, except the MP's own party p. Thus, β measures the correlation between the cosine similarities of a politician.

We present the results of the regression in Table B.3. We observe a positive correlation of the cosine similarity with the social democrats and the CSU for the conservative party, but a negative correlation with the green and the liberal party. Controlling for other similarities, a higher cosine similarity with the social democratic party members is correlated to a higher similarity to members of the MP's own party. A higher similarity to the green party relates to a lower similarity to the MP's own party. We argue that the observations are intuitive as politicians who choose words closer to those of the more distant parties tend to deviate from members of their own party's speeches. When considering the social democratic party results, the cosine similarities are correlated to members of the conservative party, the green party, and the left party. We solely observe a very negative correlation with the cosine similarity of the liberal party. Also, these correlations are in line with differences in political opinions that translate into the cosine similarities. Finally, we consider the CSU, the Bavarian part of the conservative party. We observe very similar results for the remaining part of the conservative party. The major difference is that the cosine similarities to the MP's own party are correlated with cosine similarities to the liberal party members.

Additionally, we evaluate if a higher cosine similarity to other parties relates to more roll-call deviations from the MP's own party line. Again, we show regression evidence for the following model separately for the two major parties, i.e., social democrats and conservatives:

$$Deviation_{p,i,t} = \beta \mathbf{Similarity}_{\mathbf{p,i,t}} + \varepsilon_{p,i,t},$$

where $Deviation_{p,i,t}$ shows the roll-call deviation of MP i of party p at t and $\mathbf{Similarity}_{\mathbf{p,i,t}}$ is a vector of all cosine similarity measures except the similarity to the MP's own party. Further, ξ_t are time fixed effects. We show the results in Table B.2. Model (1) refers to the sample of CDU MPs, Model (2) refers to the MPs from the social democratic party, and Model (3) refers to CSU MPs. Considering MPs of the CDU, we find that after controlling for other party similarities, a higher similarity with social democrats relates to a lower deviation. This observation could be related to the fact that both parties are classified in the political center (with two periods of a coalition in our sample), and the conservative party's MPs may tend to deviate to more conservative or right-wing positions. Indeed, we observe a positive correlation to the similarity measures of the CSU, the Bavarian component of the CDU. The CSU's position tends toward a more conservative political spectrum. A higher cosine similarity with the CSU is related to more deviations in roll-call votes. For the SPD candidates, the observation differs. We observe a positive correlation of the CDU and the green party, meaning a higher cosine similarity correlates with deviation of roll-call votes. We also observe that a higher cosine similarity to the FDP (a liberal party in the parliament) is negatively correlated with the roll-call votes' deviations. While we rationalize the results of a positive correlation of the CDU and the green party with the fact that the SPD members seem to deviate toward either the conservative or green party lines, we cannot explain the relation to the FDP. Finally, for CSU party members, only a positive cosine similarity to the CDU relates to less deviation. All other results are nonsignificantly different from zero.

[36] Note that we distinguish between the CDU and the CSU to allow for heterogeneity and evaluate similarity between CDU and CSU.

Table B.2
Correlation of cosine similarity between parties.

	Cosine similarity own party		
	CDU (1)	SPD (2)	CSU (3)
Cosine similarity SPD	0.724***		0.449***
	(0.054)		(0.122)
Cosine similarity CDU		0.649***	0.381***
		(0.037)	(0.087)
Cosine similarity CSU	0.551***	0.266***	
	(0.053)	(0.051)	
Cosine similarity GRUENE	−0.129 **	0.358***	−0.175 *
	(0.054)	(0.053)	(0.105)
Cosine similarity FDP	−0.174 ***	−0.515 ***	0.639***
	(0.064)	(0.041)	(0.100)
Cosine similarity DIE LINKE	−0.010	0.241***	−0.272 ***
	(0.043)	(0.030)	(0.072)
Constant	0.008***	0.007***	0.023***
	(0.002)	(0.002)	(0.003)
N	1844	2579	456
R^2	0.990	0.993	0.993

Robust standard errors in parentheses *** p<0.01 ** p<0.05 * p<0.1.
The table shows regression results of the cosine similarity of MPs' speeches to members of the same party and the cosine similarity measure of other parties. One observation corresponds to an MP within a period. The outcome variable is the average cosine similarity of a MP's speech to the speeches of other members of the same party. The other cosine similarity measures is the euclidean distance of a MP's speech to the speeches of members of other parties. GRUENE refers to the green party, FDP to the liberal party, and DIELINKE to the left party (democratic socialist).

Appendix C. Policy analysis

One concern with our analysis of political speeches is that we do not differentiate between speeches according to their content or policy topics. Therefore, it is possible that an MP who happens to be an expert on a highly specialized policy differs from the other members of the party because this MP's speeches do use different words. In this robustness check, we perform a heterogeneity analysis of speeches that involve other policies. Thereby, this section works as a robustness check as we show robustness when we filter speeches by policy area.

To show robustness, we consider ten common policy areas: foreign, education, development, migration, regional, environmental, European, economic, health, and social policy. For each policy area, we filter individual speeches and reduce the set of speeches to those that include at least one word associated with a specific policy area. The specific German words before lemmatization and stemming are presented in Table C.1. We created the identifying words according to our best knowledge, before analysis. However, the approach is robust when we use only a random subset of the words. Nevertheless, we acknowledge that our reduction of speeches is not perfect, and some speeches that are not about the relevant policy area could remain part of the analysis. The requirement of including at least one word associated with a policy area reduces the number of speeches by at least half, resulting in less than 5% of all speeches for some policy areas. After filtering the speeches, we calculate the cosine similarity of speeches from an MP within a period to all party members in the same way we do for the main analysis. Using these values, we conduct the same regression discontinuity analysis. Therefore, we use the model in Eq. (2) to show regression evidence if winning a district is connected to the higher or lower similarity of speeches to the speeches of party members when considering speeches on specific policy issues.

Fig. C.1 shows the coefficients for winning a district of each separate regression when considering the regression discontinuity and including the linear margin of the vote share on a bandwidth of 10 percentage points around zero as a regressor. Note that the results are similar when comparing only the average effect of winning a district on bandwidth of 5 percentage points, or when including higher polynomials of the vote share as controls. The results show that we do not observe any significant effect, independent of the policy area. Therefore, the robustness check confirms our null result. However, we argue that the point estimates further show that the cosine similarity is an appropriate measure. For example, the results show that winning a district is associated with a lower similarity to the party members for the regional policy area. Indeed, this effect is the most negative, compared to all other policy areas. In comparison, we see a nonsignificant positive effect, meaning a higher similarity to the party line when considering social or health policy. Therefore, we see some but nonsignificant effects of higher distance to the party line when winning a district for a policy area that could be important for the district's view.

Table B.3
Correlation of roll call deviations and cosine similarity of all parties.

	Deviations in roll-call voting		
	CDU (1)	SPD (2)	CSU (3)
Cosine similarity CDU	−1.395 ***	1.897**	−4.151 ***
	(0.455)	(0.803)	(1.431)
Cosine Similarity CSU	2.673***	0.374	3.047
	(0.755)	(0.809)	(2.046)
Cosine similarity SPD	−1.876 **	−3.181 ***	1.847
	(0.823)	(1.075)	(1.963)
Cosine similarity GRUENE	0.661	2.810***	−0.238
	(0.526)	(0.955)	(1.815)
Cosine similarity FDP	0.177	−2.615 ***	0.181
	(0.551)	(0.946)	(2.248)
Cosine similarity DIE LINKE	−0.579	0.489	−1.216
	(0.369)	(0.488)	(1.090)
Constant	0.128***	0.216***	0.205*
	(0.029)	(0.040)	(0.117)
N	1838	2579	456
R^2	0.148	0.058	0.146

Robust standard errors in parenthesis *** p<0.01 ** p<0.05 * p<0.1.
The table shows regression results of deviations in roll call voting and cosine similarity of MPs' speeches to party members of the same and other parties. One observation corresponds to an MP within a period. The outcome variable is the deviation from the party line in roll-call voting within a period. The cosine similarity is the euclidean distance measure that shows the average distance of words used compared to members of a party. GRUENE refers to the green party, FDP to the liberal party, and DIELINKE to the left party (democratic socialist).

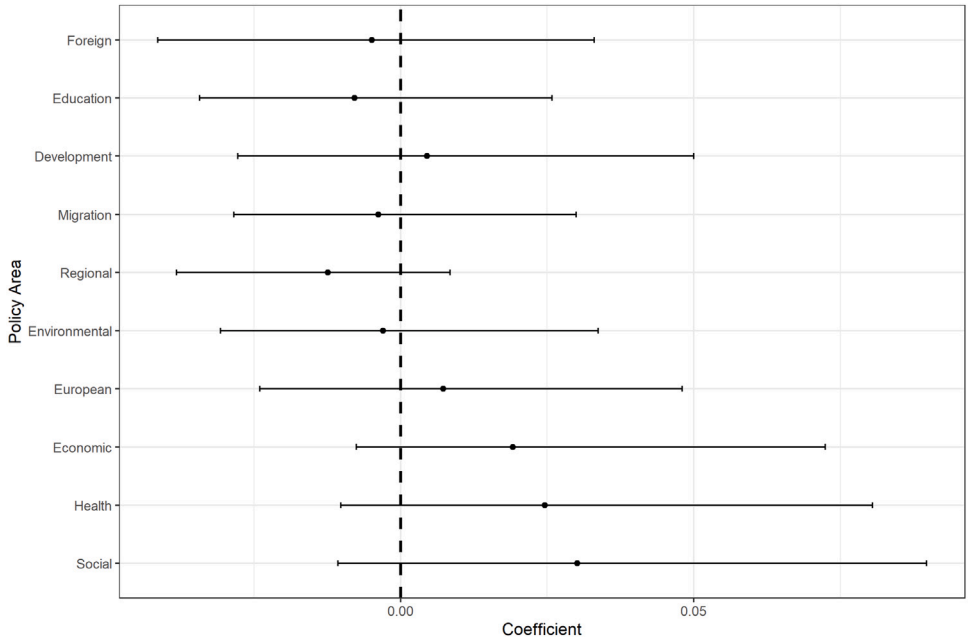

Fig. C.1. Policy-specific regression discontinuity analysis. *The figure shows coefficient estimates from separate regression discontinuity analyses when only considering speeches in one specific policy area. Each coefficient shows the effect of a district win on the cosine similarity. The models evaluate the effect considering a margin of 10 percentage points. We further include a linear margin of the vote share. The effects are robust when not including additional regressors, including higher-order polynomials, or when considering a larger bandwidth.*

Table C.1
Identification of policies in speeches.

Policy	Identification words
sozialpolitik	sozialpolitik, sozial, benachteiligt, arm, kranke, unterst\`utzung, sozialversicherung, krankheit, alter, arbeitslosigkeit, pflegebed´urftig, sozialhilfe, wohngeld, familienpolitik, mieterschutz, sozialfürsorge, wohlfahrtsstaat, hartz
gesundheitspolitik	gesundheitspolitik, gesundheit, gesundheitsystem, `arzte, apotheker, pharmaindustrie, pharmazeutika, krankheitsfall, gesundheitsversorgung, krankenkassen, gesundheitssektor, krankenversicherern, gesundheitsreform, gesundheitspr´amie
wirtschaftspolitik	wirtschaftspolitik, wirtschaft, unternehmen, wirtschaftssystem, wirtschaftssubjekt, volkswirtschaft, `okonomik, wettbewerb, struktur, infrastruktur, geld, konjunktur, handel, markt, w´ahrung, fiskalpolitik, nachfrage, angebot, wachstum, marktwirtschaft
europapolitik	europapolitik, europa, eu, europarat, europ`aische, beitrittskandidat, beitrittsl´ander, beitrittskandidat
entwicklungspolitik	entwicklungspolitik, entwicklung, entwicklungsl`andern, entwicklungsr´uckstand, entwicklungshilfe, nichtregierungsorganisationen, ngo, armut, armutsbekämpfung, weltbank, mikrofinanz, entschuldung, menschenrechte
außenpolitik	außenpolitik, sicherheitspolitik, international, beziehungen, stabilit`at, kontinuit´at, weltlage, nahostkonflikt, krieg, waffen, botschafter, milit`arisch, milit´ar, nato, staatenbund, konflikt, staatsgrenze, abkommen, friedensprozess, supranational
umweltpolitik	umweltpolitik, umwelt, natur, umweltprobleme, umweltbewusstsein, umweltprotest, umweltbewegung, atomkraft, landwirtschaft, industriepolitik, umweltverschmutzung, kohle, klima, klimawandel, klimagipfel, meeresverschmutzung, atomausstieg, umweltvertraeglichkeit, erwärmung
regionalpolitik	regionalpolitik, strukturpolitik, gebietsk`orperschaft, wahlkreis, region, standortpolitik, region, kommune, regionalentwicklung, wirtschaftsf´orderung, wirtschaftskraft, selbstverwaltung, fachkr`aftebedarf, industriefl´ache, gemeinde, st`adtetag, st´adte
bildungspolitik	bildungspolitik, bildung, schule, l`anderebene, l´ander, schule, lehrer, kultusministerien, lehrplan, hochschule, universit`aten, studiengeb´uhren, hochschulbildung, professoren, betreuung, föderalismus, chancengleichheit, herkunft, schulferien, schultypen, gymnasium, hauptschule, gesamtschule, lehrplan, schulsystem
migration	migration, ausl`ander, staatenlos, flucht, fl´uchtling, asyl, asylpolitik, asylantrag, migrationspolitik, dublin, integration, einwanderung, fl`uchtlingspolitik, ausl´anderpolitik, zuwanderung, aufenthalt, aufenthaltsstatus, einb`urgerung, zuwanderung, sprachf´orderung, integrationsgesetz, masseneinwanderung

The table presents words we use for identifying policy areas of speeches. The first column presents the policy area (in German). The second column shows German words that we use to associate a speech with the policy area.

Appendix D. Party-closeness measure

In the main specification we train a multinomial logistic regression with L2 regularization on the speeches of all other MPs to predict every MP's party affiliation based on the speeches. This means that for each MP we fit the logistic classifier on all speeches but the ones from the MP in question in a leave-one-out fashion. We use the classifier fitted in this way to predict the party that the MP belongs to and then repeat the process for the next MP. With this methodology we can correctly predict the (known to us) party affiliation of around 80% of MPs. Importantly, the goal is not to make correct predictions, but to compare the 'certainty' of the model. That means we do not need perfect predictions, but the predictions have to be reasonably good for us to be confident that these are not only random noise. As the model assigns the correct party the highest probability in around 80% of cases, we are confident that the model is reasonably accurate.

D.1. Description of methodology

To do the classification we use an L2-regularized logistic regression. One advantage of a logistic regression over some other regressions is that it yields likelihoods instead of absolute predictions without the additional transformation that would be necessary with other algorithms such as naive Bayes classifier, random forest classifier, or support vector machines.[37] We verified 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 which it assigns the highest likelihood to the correct party. We use cross-validation to find the optimal height of the penalty term; the 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 to use 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 all speeches except those of the MP to predict for, we use the classifier to make predictions for the merged speeches given by the MP in question over an entire period in office. The classifier yields a prediction of the party membership of the MP in question. We use the probability that the classifier assigns the MP's actual party as a measure of similarity between the MP's

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

Table D.1
Summary statistics for classifying the party using party manifestos.

Summary variables	2005/09	2009/09	2013/09
Predictability from Manifestos:			
Avg predictability measure	0.34	0.49	0.38
	(0.23)	(0.37)	(0.24)
Avg predictability measure, Conservatives	0.48	0.76	0.56
	(0.21)	(0.26)	(0.19)
Avg predictability measure, Social democrats	0.17	0.13	0.18
	(0.14)	(0.19)	(0.11)

The statistic is divided across the three different election periods. We show statistics for all MPs and for MPs of the two major parties (conservative CDU/CSU and social democratic SPD). The third segment shows the average predictability of party affiliation when using the party manifestos as a classifier. All outcomes lie between zero and one. For the predictability measure, a value of 1 can be interpreted as meaning that an MP's party is perfectly predictable from the MP's speeches. Standard deviations are shown in parentheses.

speeches and those of the MP's party. This measure will take values from 0 to 1, where 1 represents the highest similarity (100% predicted probability) and 0 represents the lowest similarity (0% predicted probability).

D.2. Robustness: Manifestos

As a robustness measure we make use of the election manifestos parties write before the election of the federal parliament. We stem and vectorize the manifestos in the same way as the speeches and use them to build a measure of how closely an MP's language resembles the language of their party's manifesto compared to the manifestos of other parties. 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 closely an MP's language matches the MP'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. In contrast to the first measure, this can be regarded as an inter-party measure, as it captures the distance of speeches to the MP's own party's manifesto relative to the other parties' manifestos. In the following we describe this strategy more in detail.

In the robustness specification we also train a multinomial logistic regression with L2 regularization to classify speeches based on the party manifestos to then predict an MP's party affiliation. The purpose of this is to ensure that the results do not hinge on the particularities of speeches when predicting the origin of other speeches.

After training the classifier on the party manifestos, we use the classifier to categorize the merged speeches given by each MP over an entire period in office. The classifier yields a prediction of the party membership of the MP in question. In particular, we use the probability that the classifier assigns the MP's actual party as a measure of similarity between the MP's speeches and those of the MP's party. A high predicted probability of the actual party of an MP indicates that the MP's speeches are closer to the party manifesto of the MP's own party than to the manifestos of other parties.

First, we show average values of the prediction of the party membership in Table D.1. We then turn to the regression evidence. We start with the naive OLS regression, equivalent to the methodology in the main part of the paper. Table D.2 provides the results of the robustness check for Table 7. Notably, without controls the coefficient is positive and significantly different from zero. Controlling for state, party, and session fixed effects, however, makes the coefficient close to and not significantly different from zero.

Table D.3 provides a robustness check of Table 8 by displaying the regression discontinuity design estimates of the relationship of marginally winning vs. losing a district election on the closeness of the MP's speeches to the party manifesto. Noticeably, as in the main analysis, none of the coefficients is significantly different from zero.

D.3. The most predictable words

To evaluate if the classifier has an intuitive interpretation, D.4 shows the most predictive words that distinguish between MPs of the conservative and social democratic parties.[38] The coefficients are based on the logistic regression of classifying the party manifestos. The results show that the words that distinguish conservatives from social democrats are intuitive. For example, naming one's own party is highly predicted for both major parties. Further, "extremism" or "bureaucracy" is related to the conservatives, while "employees" and "welfare state" are words connected to social democrats. Overall, the results suggest that the logistic regression selects comprehensible words to classify speeches.

[38] Note that we exclude all stopwords.

Table D.2
OLS of winning a district vote on manifesto closeness (Robustness).

	(1) Simple	(2) With controls	(3) No selection	(4) With controls
Won district	0.136***	0.0201	0.160***	0.0205
	(0.0135)	(0.0128)	(0.0175)	(0.0159)
Observations	1882	1818	1392	1392
R-squared	0.052	0.443	0.060	0.405
State FE		YES		YES
Party FE		YES		YES
Session FE		YES		YES

Robust standard errors in parentheses *** p<0.01 ** p<0.05 * p<0.1.
Linear least squares regression. One observation corresponds to a politician within one election period. The dependent variable is the predictability score of an MP to the party manifesto. A value of 1 corresponds to a perfect predictability. $Wondistrict$ is a dummy that takes the value 1 if a politician won their district. $State$ fixed effects are fixed effects for the 16 German federal states, $party$ fixed effects are for the party of a MP, and $session$ fixed effects are for an election period. Models (1) and (2) are models without selection, where model (2) adds the fixed effects as controls. Models (3) and (4) select the sample such that we only consider MPs who either entered the parliament on a party list or entered through a district election but would have entered on a party list if they had lost their district.

Table D.3
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	−0.0373	−0.0775	−0.0436	−0.0539
Observations	152	296	820	820
Robust Std. Error	0.0981	0.101	0.0721	0.0861
Robust *p*-value	0.769	0.758	0.455	0.384

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 in relation to the party manifesto. A value of 1 corresponds to a perfect predictability. In Model (1) only the average is compared on a bandwidth of 5 percentage points around a winning margin of zero. Model (2) adds the linear margin of the vote share on a bandwidth of 10 percentage points around zero. Model (3) adds a second-order polynomial and increases the bandwidth to 40. Finally, Model (4) adds a third-order polynomial and increases the margin even further to 40 percentage points.

Appendix E. 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 by evaluating correlations of variables with the deviation from the party line. To give a graphical impression of possible correlations, Fig. E.1 shows the district vote margin in relation to the share of deviations from individual MPs. Note, first, 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 a district candidacy MP within one election period. The graph includes a linear least squares regression line as well as the 95% confidence band. We see a slightly negative relationship between the 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 position changes

We also present some summary statistics about the list position changes between consecutive periods. We calculate the list position change of those MPs that are part of party lists. In detail, let P_t be the list position at time t, where a lower position gives a higher possibility of entering 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 who 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 social democrats. Furthermore, the effect is stronger for district-elected MPs. However, the increase is mostly due to those district MPs from the social democratic party. The list place changes of social democratic, district-elected candidates are lower and even negative from the 17th to the 18th election period as well as from the 18th to the 19th election period (see Table E.1).

Correlations with socioeconomic characteristics

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

Table D.4
Most predictive words of classifier based on party manifestos.

Wording	German wording	CDU	SPD
Employee (female)	arbeitnehmerinn	−4.09	3.94
Consumer (female)	verbraucherinn	−5.45	1.70
Federal chancellor	bundeskanzl	−2.06	5.08
Christian	christlich	4.74	−2.10
Extremist	extremist	4.64	−1.58
Red	rot	4.40	−1.81
Governed by the Union-Parties	unionsgefuhrt	4.53	−1.47
Citizen (female)	burgerinn	−2.63	3.05
Extremism	extremismus	3.55	−2.05
Welfare state	sozialstaat	−3.10	2.23
Attractive	attraktiv	3.19	−2.11
Health premiums	gesundheitsprami	3.99	−1.27
Social democracy	sozialdemokrati	−1.33	3.71
Grow	aufwachs	−1.02	3.93
Demographic change	bevolkerungswandel	3.56	−1.34
Peace power	friedensmacht	−1.37	3.49
Resettler	aussiedl	3.58	−1.17
Participation	mitbestimm	−2.95	1.63
Chancellor	kanzl	−0.97	3.61
Berlin	berlin	3.15	−1.43
Employee insurance	arbeitsversicher	−1.26	3.25
Government	regier	−3.30	1.19
Immigration history	zuwanderungsgeschicht	3.49	−1.00
Follow	folgt	−1.49	3.00

The table shows the coefficients of the most predictive words that distinguish between MPs of the conservative and social democratic party. The coefficients are based on the logistic regression based on classifying the party manifestos. The ordering of the words is based on the absolute difference between the coefficients of a conservative MP (reported in column three) and a social-democrat (reported in column four).

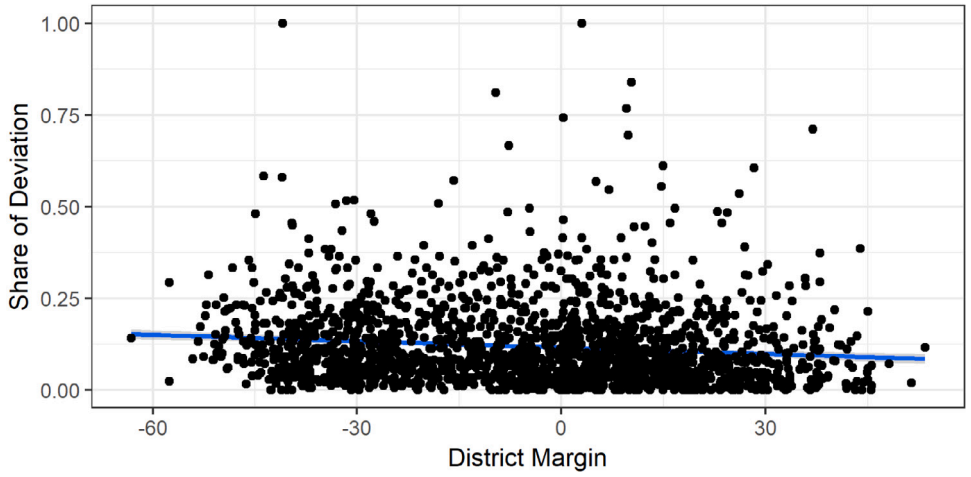

Fig. E.1. District vote share and deviation from party line in roll call voting. *District vote margin of district candidacy MPs to closes opponent compared to the share of deviations to party line within one election period. The graph includes a linear least square regression line as well as the 95% confidence band.*

Appendix F. Additional results

Supporting the analysis in Section 7.2, the following figures show fitted polynomials and sample averages of the regression discontinuity design estimations of the section.

Fig. F.1 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.

Fig. F.2 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.

Fig. F.3 shows a fitted polynomial and the binned sample average of predictability measures of speeches around the voting cutoff. Visual inspection of the figure confirms the conclusion that there is no discontinuity.

Table E.1
Descriptive statistics of list place changes.

	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 District 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 E.2
Summary statistics of the socioeconomic variables.

	Mean	SD
Population density (Persons per square mile)	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 show the averages and standard deviations of socioeconomic variables across all districts across the three election periods.

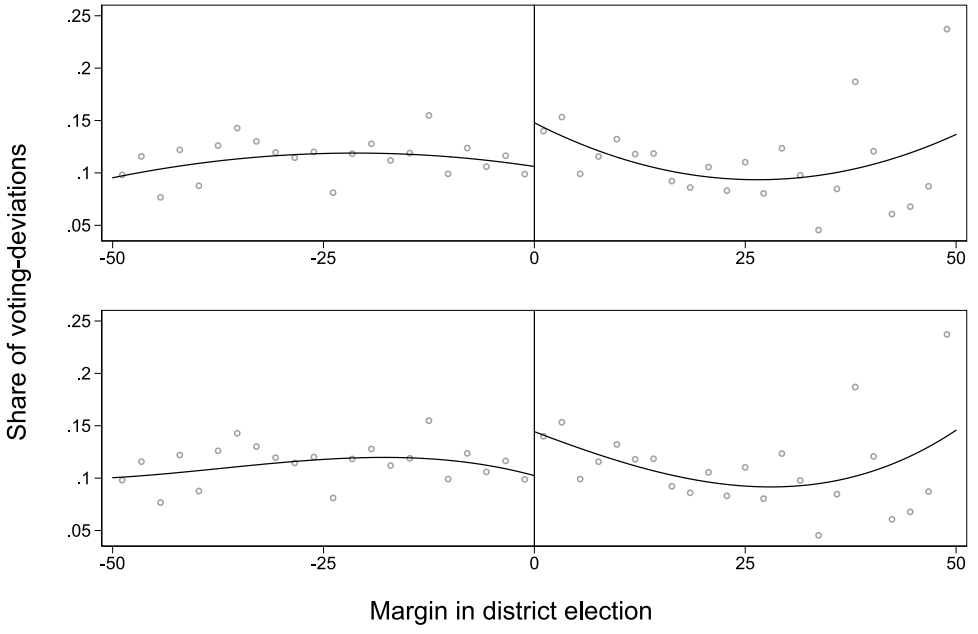

Fig. F.1. Graphical analysis of the effect of a district vote on deviations. *The graphs show binned averages of MPs' share of voting deviations as a function of the win/loss margin to being the district vote winner. The bins are created to be evenly spaced using a polynomial regression. The lines in the graphs are second-order (top) and third-order (bottom) polynomials fitted to the observations.*

Fig. F.4 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 is no discontinuity in the future list position.

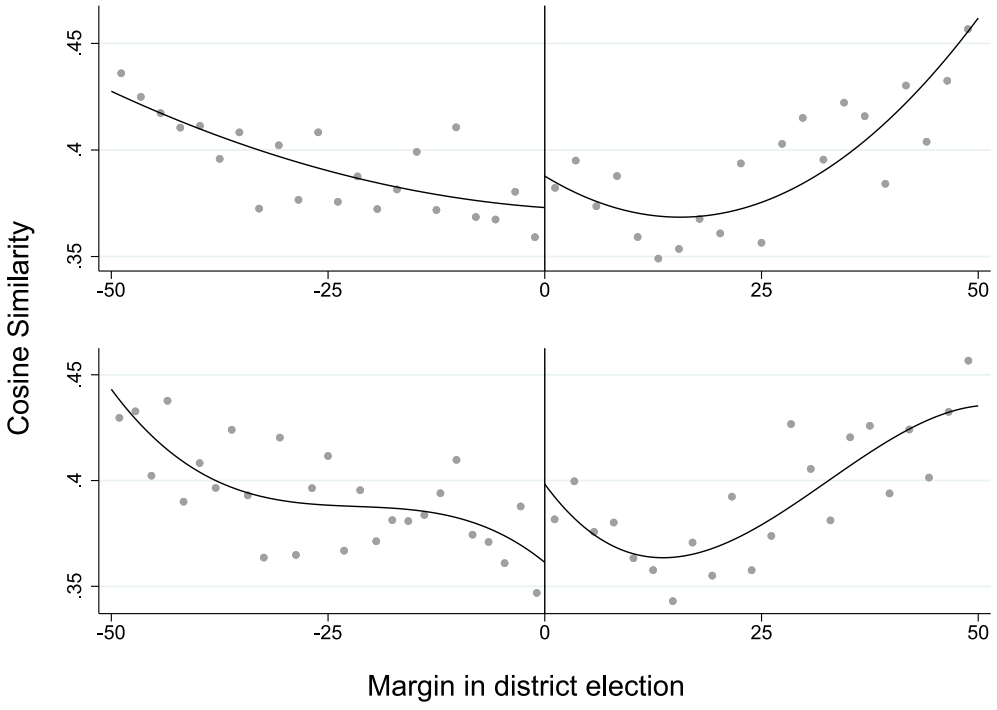

Fig. F.2. Graphical analysis, winning a district mandate on the cosine distance to party average. *The graphs show binned averages of MPs' cosine similarity in their speeches as a function of the win/loss margin in the district vote. The bins are created to be evenly spaced using a polynomial regression. The lines in the graphs are second-order (top) and third-order (bottom) polynomials fitted to the observations.*

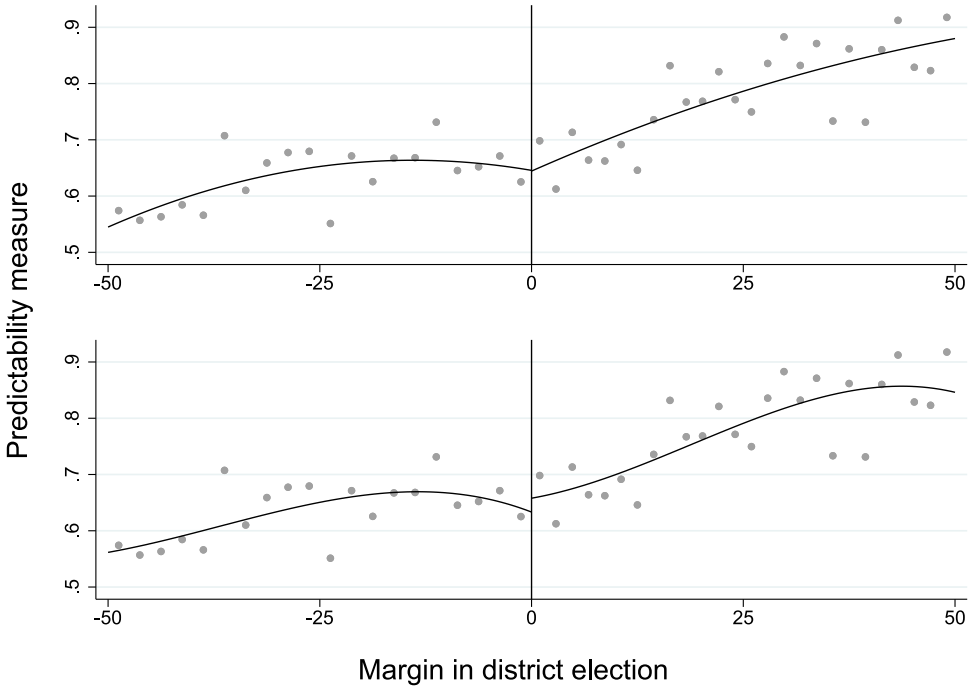

Fig. F.3. Graphical analysis, district vote on predictability of speeches. *The graphs show binned averages of MPs' speeches' predictability measures as a function of the win/loss margin in the district vote. The bins are created to be evenly spaced using a polynomial regression. The lines in the graphs are second-order (top) and third-order (bottom) polynomials fitted to the observations.*

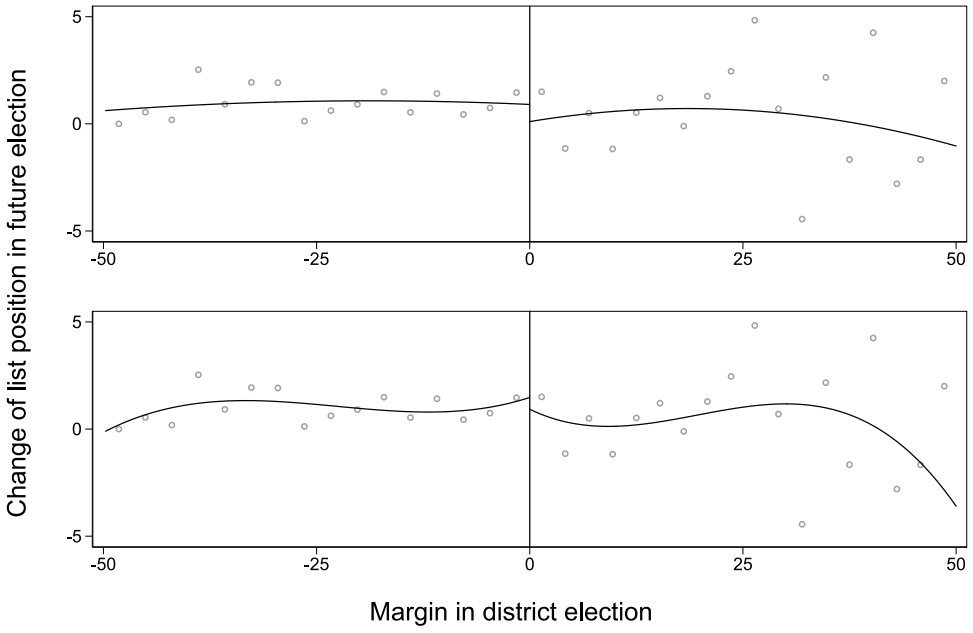

Fig. F.4. Graphical analysis effect district vote on future list positions. *The graphs show binned averages of MPs' future list position as a function of the win/loss margin in the district vote. The bins are created to be evenly spaced using a polynomial regression. The lines in the graphs are second-order (top) and third-order (bottom) polynomials fitted to the observations.*

References

Acemoglu, Daron, Johnson, Simon, Robinson, James A., 2001. The colonial origins of comparative development: An empirical investigation. Amer. Econ. Rev. 91 (5), 1369–1401.

Ade, Florian, Freier, Ronny, Odendahl, Christian, 2014. Incumbency effects in government and opposition: Evidence from Germany. Eur. J. Political Econ. 36, 117–134.

Ainsley, Cairlin, Carrbura, Clifford J, Crisp, Brian F, Demirkaya, Betul, Gabel, Matthew J, Hadzic, Dino, 2020. Roll-call vote selection: Implications for the study of legislative politics. Am. Polit. Sci. Rev. 1–16.

Alesina, Alberto, Glaeser, Edward, Sacerdote, Bruce, 2001. Why Doesn't the US Have a European-Style Welfare System? National Bureau of Economic Research.

Anagol, Santosh, Fujiwara, Thomas, 2016. The runner-up effect. J. Polit. Econ. 124 (4), 927–991.

André, Audrey, Depauw, Sam, Shugart, Matthew S., Chytilek, Roman, 2017. Party nomination strategies in flexible-list systems: Do preference votes matter? Party Polit. 23 (5), 589–600.

Aragones, Enriqueta, Palfrey, Thomas R., 2002. Mixed equilibrium in a Downsian model with a favored candidate. J. Econom. Theory 103 (1), 131–161.

Aragonès, Enriqueta, Postlewaite, Andrew, Palfrey, Thomas, 2007. Political reputations and campaign promises. J. Eur. Econom. Assoc. 5 (4), 846–884.

Arnold, Felix, Kauder, Björn, Potrafke, Niklas, 2014. Outside earnings, absence, and activity: Evidence from German parliamentarians. Eur. J. Political Econ. 36, 147–157.

Athey, Susan, 2018. The impact of machine learning on economics. In: The Economics of Artificial Intelligence: An Agenda. University of Chicago Press.

Austen-Smith, David, Banks, Jeffrey S., 1989. Electoral accountability and incumbency. In: Models of Strategic Choice in Politics. University of Michigan Press.

Bäck, Hanna, Debus, Marc, 2017. Representing the region on the floor: Socioeconomic characteristics of electoral districts and legislative speechmaking. Parliam. Aff. 71 (1), 73–102.

Becher, Michael, Sieberer, Ulrich, 2008. Discipline, electoral rules and defection in the Bundestag, 1983–94. German Polit. 17 (3), 293–304.

Bernauer, Julian, Bräuninger, Thomas, 2009. Intra-party preference heterogeneity and faction membership in the 15th German Bundestag: a computational text analysis of parliamentary speeches. German Polit. 18 (3), 385–402.

Bernecker, Andreas, 2014. Do politicians shirk when reelection is certain? Evidence from the German parliament. Eur. J. Political Econ. 36, 55–70.

Biessmann, Felix, Lehmann, Pola, Kirsch, Daniel, Schelter, Sebastian, 2016. Predicting political party affiliation from text. In: International Conference on the Advances in Computational Analysis of Political Text. PolText.

Born, Andreas, 2018. Promise competition. In: Working Paper.

Carey, John M., 2007. Competing principals, political institutions, and party unity in legislative voting. Am. J. Polit. Sci. 51 (1), 92–107.

Carrillo, Juan D., Castanheira, Micael, 2008. Information and strategic political polarisation. Econ. J. 118 (530), 845–874.

Carrubba, Clifford, Gabel, Matthew, Hug, Simon, 2008. Legislative voting behavior, seen and unseen: A theory of roll-call vote selection. Legislative Stud. Q. 33 (4), 543–572.

Carrubba, Clifford J, Gabel, Matthew, Murrah, Lacey, Clough, Ryan, Montgomery, Elizabeth, Schambach, Rebecca, 2006. Off the record: Unrecorded legislative votes, selection bias and roll-call vote analysis. Br. J. Polit. Sci. 691–704.

Corazzini, Luca, Kube, Sebastian, Maréchal, Michel André, Nicolo, Antonio, 2014. Elections and deceptions: an experimental study on the behavioral effects of democracy. Am. J. Polit. Sci. 58 (3), 579–592.

Crisp, Brian F, Olivella, Santiago, Malecki, Michael, Sher, Mindy, 2013. Vote-earning strategies in flexible list systems: Seats at the price of unity. Elect. Stud. 32 (4), 658–669.

Dal Bó, Ernesto, Dal Bó, Pedro, Snyder, Jason, 2009. Political dynasties. Rev. Econom. Stud. 76 (1), 115–142.

Depauw, Sam, Martin, Shane, 2009. Legislative party discipline and cohesion in comparative perspective. Intra-Party Polit. Coalition Gov. 103–120.

Diermeier, Daniel, Godbout, Jean-François, Yu, Bei, Kaufmann, Stefan, 2012. Language and ideology in Congress. Br. J. Polit. Sci. 42 (1), 31–55.

Downs, Anthony, 1957. An economic theory of political action in a democracy. J. Polit. Econ. 65 (2), 135–150.
Folke, Olle, Persson, Torsten, Rickne, Johanna, 2016. The primary effect: Preference votes and political promotions. Am. Polit. Sci. Rev. 110 (3), 559–578.
Freier, Ronny, 2015. The mayor's advantage: Causal evidence on incumbency effects in German mayoral elections. Eur. J. Political Econ. 40, 16–30.
Gavoille, Nicolas, 2018. Who are the 'ghost'MPs? Evidence from the French parliament. Eur. J. Political Econ. 53, 134–148.
Gentzkow, Matthew, Kelly, Bryan T., Taddy, Matt, 2017. Text as Data. Technical Report, National Bureau of Economic Research.
Gerber, Elisabeth R., Lewis, Jeffrey B., 2004. Beyond the median: Voter preferences, district heterogeneity, and political representation. J. Polit. Econ. 112 (6), 1364–1383.
Grimmer, Justin, Stewart, Brandon M., 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Polit. Anal. 21 (3), 267–297.
Gruber, Jonathan, 2005. Public Finance and Public Policy. Macmillan.
Hug, Simon, 2010. Selection effects in roll call votes. Br. J. Polit. Sci. 40 (1), 225–235.
Hyytinen, Ari, Meriläinen, Jaakko, Saarimaa, Tuukka, Toivanen, Otto, Tukiainen, Janne, 2018. When does regression discontinuity design work? Evidence from random election outcomes. Quant. Econ. 9 (2), 1019–1051.
Kittsteiner, Thomas, Eyster, Erik, 2007. Party platforms in electoral competition with heterogeneous constituencies. Theor. Econ. 2 (1), 41–70.
Klingemann, Hans-Dieter, Wessels, Bernhard, 2003. The political consequences of Germany's mixed-member system: Personalization at the grass roots? In: Shugart, Matthew, Wattenberg, Martin P. (Eds.), Mixed-Member Electoral Systems : The Best of Both Worlds? Oxford University Press, Oxford.
Laver, Michael, Benoit, Kenneth, Garry, John, 2003. Extracting policy positions from political texts using words as data. Am. Polit. Sci. Rev. 97 (2), 311–331.
Lehmann, Pola, Matthieß, Theres, Merz, Nicolas, Regel, Sven, Werner, Annika, 2017. Manifesto corpus. Version: 2017–2. In: WZB Berlin Social Science Center. WZB Berlin Social Science Center, Berlin.
Lindbeck, Assar, Weibull, Jörgen W., 1987. Balanced-budget redistribution as the outcome of political competition. Publ. Choice 52 (3), 273–297.
Lindbeck, Assar, Weibull, Jörgen W., 1993. A model of political equilibrium in a representative democracy. J. Publ. Econ. 51 (2), 195–209.
Lizzeri, Alessandro, Persico, Nicola, 2001. The provision of public goods under alternative electoral incentives. Amer. Econ. Rev. 91 (1), 225–239. http://dx.doi.org/10.1257/aer.91.1.225, URL http://www.aeaweb.org/articles?id=10.1257/aer.91.1.225.
Maaser, Nicola, Stratmann, Thomas, 2018. Election rules, legislators' incentives, and policy outcomes: Evidence from the mixed member system in Germany. Eur. J. Political Econ. 54, 227–239.
Maltzman, Forrest, Sigelman, Lee, 1996. The politics of talk: Unconstrained floor time in the US House of Representatives. J. Polit. 58 (3), 819–830.
Manow, Philip, 2013. Mixed rules, different roles? An analysis of the typical pathways into the Bundestag and of MPs' parliamentary behaviour. J. Legislative Stud. 19 (3), 287–308.
Martin, Lanny W., Vanberg, Georg, 2008. Coalition government and political communication. Polit. Res. Q. 61 (3), 502–516.
Meriläinen, Jaakko, Tukiainen, Janne, 2018. Rank effects in political promotions. Publ. Choice 177, 87–109.
Milesi-Ferretti, Gian Maria, Perotti, Roberto, Rostagno, Massimo, 2002. Electoral systems and public spending*. Q. J. Econ. (ISSN: 0033-5533) 117 (2), 609–657. http://dx.doi.org/10.1162/003355302753650346.
Mueller, Dennis C., 1997. Perspectives on Public Choice: A Handbook. Cambridge University Press.
Mullainathan, Sendhil, Spiess, Jann, 2017. Machine learning: an applied econometric approach. J. Econ. Perspect. 31 (2), 87–106.
Neuhäuser, Alice, Mischler, Maike, Ruxton, Graeme D, Neuhäuser, Markus, 2013. Gründe für von der Fraktionsdisziplin abweichendes Abstimmungsverhalten bei Bundestagsabgeordneten. AStA Wirtschafts-Und Sozialstatistisches Archiv 7 (1–2), 91–99.
Ohmura, Tamaki, 2014. When your name is on the list, it is time to party: The candidacy divide in a mixed-member proportional system. Representation 50 (1), 69–82.
Persson, Torsten, Roland, Gerard, Tabellini, Guido, 1997. Separation of powers and political accountability. Q. J. Econ. 112 (4), 1163–1202.
Persson, Torsten, Tabellini, Guido, 1999. The size and scope of government:: Comparative politics with rational politicians. Eur. Econ. Rev. 43 (4–6), 699–735.
Persson, Torsten, Tabellini, Guido, 2004. Constitutional rules and fiscal policy outcomes. Amer. Econ. Rev. 94 (1), 25–45.
Peterson, Andrew, Spirling, Arthur, 2018. Classification accuracy as a substantive quantity of interest: Measuring polarization in westminster systems. Polit. Anal. 26 (1), 120–128.
Potrafke, Niklas, Rösch, Marcus, Ursprung, Heinrich, 2020. Election systems, the "beauty premium" in politics, and the beauty of dissent. Eur. J. Political Econ. 101900.
Proksch, Sven-Oliver, Lowe, Will, Wäckerle, Jens, Soroka, Stuart, 2019. Multilingual sentiment analysis: A new approach to measuring conflict in legislative speeches. Legislative Stud. Q. 44 (1), 97–131.
Proksch, Sven-Oliver, Slapin, Jonathan B., 2010. Position taking in European parliament speeches. Br. J. Polit. Sci. 587–611.
Proksch, Sven-Oliver, Slapin, Jonathan B., 2012. Institutional foundations of legislative speech. Am. J. Polit. Sci. 56 (3), 520–537.
Quinn, Kevin M, Monroe, Burt L, Colaresi, Michael, Crespin, Michael H, Radev, Dragomir R, 2010. How to analyze political attention with minimal assumptions and costs. Am. J. Polit. Sci. 54 (1), 209–228.
Scarrow, Susan E., 2001. Germany: The mixed-member system as a political compromise. Mixed-Member Electoral Systems: The Best of Both Worlds 55–69.
Schwarz, Daniel, Traber, Denise, Benoit, Kenneth, 2015. Estimating intra-party preferences: comparing speeches to votes. Polit. Sci. Res. Methods.
Sieberer, Ulrich, 2010. Behavioral consequences of mixed electoral systems: Deviating voting behavior of district and list MPs in the German Bundestag. Elect. Stud. 29 (3), 484–496.
Sieberer, Ulrich, 2014. Do single member districts increase voting dissent under mixed electoral systems? A longitudinal study of voting behavior in the German Bundestag, 1949-2013.
Sieberer, Ulrich, 2015. Using MP statements to explain voting behaviour in the German Bundestag: An individual level test of the Competing Principals Theory. Party Polit. 21 (2), 284–294.
Sieberer, Ulrich, Ohmura, Tamaki, 2019. Mandate type, electoral safety, and defections from the party line: The conditional mandate divide in the German Bundestag, 1949–2013. Party Polit. 1354068819881637.
Sieberer, Ulrich, Saalfeld, Thomas, Ohmura, Tamaki, Bergmann, Henning, Bailer, Stefanie, 2018. Roll-call votes in the German Bundestag: A new dataset, 1949–2013. Br. J. Polit. Sci. 1–9. http://dx.doi.org/10.1017/S0007123418000406.
Sieberer, Ulrich, Saalfeld, Thomas, Ohmura, Tamaki, Bergmann, Henning, Bailer, Stefanie, 2020. Roll-call votes in the German Bundestag: A new dataset, 1949–2013. Br. J. Polit. Sci. 50 (3), 1137–1145.
Slapin, Jonathan B., Proksch, Sven-Oliver, 2008. A scaling model for estimating time-series party positions from texts. Am. J. Polit. Sci. 52 (3), 705–722.
Stecker, Christian, 2015. How effects on party unity vary across votes. Party Polit. 21 (5), 791–802.
Stoffel, Michael F., 2014. MP behavior in mixed-member electoral systems. Elect. Stud. 35, 78–87.
Strömberg, David, 2008. How the Electoral College influences campaigns and policy: the probability of being Florida. Amer. Econ. Rev. 98 (3), 769–807.
VanDoren, Peter M., 1990. Can we learn the causes of congressional decisions from roll-call data? Legislative Stud. Q. 311–340.
Zadrozny, Bianca, Elkan, Charles, 2002. Transforming classifier scores into accurate multiclass probability estimates. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '02, ACM, New York, NY, USA, ISBN: 1-58113-567-X, pp. 694–699. http://dx.doi.org/10.1145/775047.775151, URL http://doi.acm.org/10.1145/775047.775151.