

Group Identities and Parliamentary Debates*

Jon H. Fiva[†]

Oda Nedregård[‡]

Henning Øien[§]

February 1, 2025

Abstract

In list-based proportional representation systems, the election of individual politicians depends on their rank position on the ballot. Party organizations therefore play a crucial role in determining which candidates will be elected, and incumbents have incentives to remain popular among party elites. What role do politicians' group identities play in such a setting? We study this question using legislative speech from the Norwegian Parliament and recently developed techniques for measuring group differences in high-dimensional choices. Across the background characteristics we consider — gender, age, urbanicity, and class background — we document substantial differences in speech, even when comparing legislators from the same political party and policy committee. This suggests that even in party-centered environments, politicians' social ties and group identities matter for the policymaking process.

Keywords: political parties, representation, text analysis, penalized logistic regression

*The paper was previously circulated under the title “Polarization in Parliamentary Speech.” We thank Elliott Ash, Jack Blumenau, Ali Cirone, Gary Cox, Jens Olav Dahlgaard, Elias Dinas, Ben Geys, Orit Kedar, Yvonne Markaki, Maria Olsson, Carlo Prato, Carlos Sanz, Jesse Shapiro, Dan Smith, Martin Søyland, Janne Tukiainen, Galina Zudenkova, and the referees for their valuable comments and suggestions. We are also grateful to Tuva Værøy for assistance with data collection. Fiva acknowledges financial support from the Norwegian Research Council (grant no. 314079).

[†]Department of Economics, BI Norwegian Business School. E-mail: jon.h.fiva@bi.no.

[‡]Department of Economics, Sciences Po. E-mail: oda.nedregard@sciencespo.fr.

[§]Norwegian Institute of Public Health and Department of Health Management and Health Economics, University of Oslo. E-mail: henning.oien@medisin.uio.no.

1. Introduction

In closed-list proportional representation systems, voters cast their ballots for political parties rather than individual candidates, and candidates are elected according to their rank on the party ballot.¹ In such systems, party elites employ various strategies to discipline rank-and-file incumbents, resulting in high levels of legislative cohesion (see, e.g., Cirone, Cox and Fiva, 2021). Does the significant power parties hold to enforce party discipline in a closed-list proportional representation system render politicians' background characteristics irrelevant?

We examine the extent to which politicians' social ties and group identities influence the policymaking process within Norway's closed-list proportional representation system. We rely on floor speeches between 1981 and 2021, along with individual-level background characteristics, as a means to gain insight into legislators' policy preferences. In candidate-centered electoral environments, such as the United States, roll-call voting can be used to measure individual legislators' policy preferences (e.g., Poole and Rosenthal, 2000). However, in most parliamentary systems, voting is tightly controlled through party discipline, meaning that legislators vote with their party possibly not because of their personal policy preferences, but sometimes in spite of them (Bailer et al., 2022; Proksch and Slapin, 2012; Schwarz, Traber and Benoit, 2017). Floor speeches provide valuable insight into individual politicians' contributions to policymaking, reflecting the full legislative process, from policy formation to implementation and justification.

To quantify differences in speech patterns by legislators with different background characteristics, we build on the work of Gentzkow, Shapiro and Taddy (2019). Their method is ideal for our purposes since it allows us to compare bivariate characteristics while taking into account covariates, and controlling for small sample bias.² While

¹Catalinac and Motolinia (2021) identify 47 countries worldwide that use closed-list proportional representation, including many Western European democracies.

²Because parties might allocate legislators strategically to committees based on their descriptive backgrounds, we include committee fixed effects in our analyses. Previous work has shown, for example, that women are assigned disproportionately to committees that focus on women's issues and social

Gentzkow, Shapiro and Taddy (2019) study differences in political speech *across* the two parties in the United States Congress, we are primarily interested in quantifying differences across legislators belonging to the *same political party*.

As a benchmark to assess within-party divergence, we first quantify divergence *across* the political blocs (i.e., political polarization) that dominate Norwegian politics; the left-leaning social democratic camp vs. the right-leaning conservative camp. We document substantial bloc divergence in legislative speech, which is increasing over our sample period. We find similar, but somewhat smaller, differences across the four background dimensions that we consider. In the second half of our sample period (2002–2021), we estimate that the probability that a neutral observer correctly guesses the speaker’s bloc affiliation is about 67 percent after hearing a one-minute speech. For the four background dimensions that we consider, a one-minute speech gives a neutral observer about 56 – 63 percent chance of correctly guessing the relevant background characteristic. In most parliamentary sessions, we can rule out that the observed differences in speech patterns are driven by random variations for all background characteristics.

After establishing the existence of speech divergence between legislators with opposing background characteristics, we examine how these differences manifest by scrutinizing the most divergent words between background pairs. For example, we find that rural legislators focus on topics, such as regional policies, farming, and sparsely populated areas, while urban legislators prioritize city-specific issues like crime and urban development. Similarly, women devote more attention to family and welfare policies in their speeches than their male colleagues.³ These divisions align with the policy preferences of voters that share their background characteristics.

issues and are often underrepresented on committees that deal with economics or foreign affairs (see, e.g., Michelle Heath, Schwindt-Bayer and Taylor-Robinson, 2005; Pansardi and Vercesi, 2016). We find a similar pattern in Norway. Søyland and Høyland (2021) find that legislators in the Norwegian parliament primarily participate in debates related to their respective committees.

³Several scholars have previously documented similar general differences in legislative speech between men and women. See, for example, Bäck, Debus and Müller (2014); Lippmann (2022); Clayton, Josefsson and Wang (2017); Osborn and Mendez (2010); Blumenau (2021) for studies of Sweden, France, Uganda, the United States, and the United Kingdom, respectively. Less is known about the other background characteristics that we consider (Gulzar, 2021).

A key reason for focusing on legislative debate is that it serves as an “observable product of party politics” (Laver, 2021, p.26). Legislative speeches reveal legislators’ intentions and priorities before formal laws are enacted and shed light on inter-party divisions, intra-party disagreements, and opposition to policy compromises (e.g., Baumgartner, Breunig and Grossman, 2019; Martin and Vanberg, 2008; Ivanusch, 2024; Itzkovitch-Malka et al., 2024). While the canonical spatial model of multiparty competition assumes that parties can be treated “as if” they are unitary actors—either due to legislator sorting into party groups with similar preferences (Krehbiel, 1993) or because party leaders wield control over key rewards and punishments to enforce strong party discipline (Cox and McCubbins, 2005)—our analysis challenges this assumption. Specifically, we find evidence suggesting that parties are unable to fully discipline their rank-and-file members, and that descriptive representation matters for the policymaking process. This result is particularly noteworthy in the Norwegian context, where party leaders have “both the ability and the incentive to select only speakers who they expect to stick to the official party line” (Laver, 2021, p.27). An alternative interpretation of our findings is that parties may intentionally allow politicians’ individual preferences to surface. If voters hold diverse policy preferences, a vote-maximizing party might strategically encourage politicians with particular background characteristics to publicly advocate for these issues.

Our study contributes to the literature in two ways. First, we contribute to the broader literature on how disparities in *descriptive representation* influence a group’s *substantive representation*?⁴ Existing evidence from candidate-centered electoral environments suggest that politicians’ social ties and group identities matter for policy outcomes, (see, e.g., Chattopadhyay and Duflo, 2004; Hyytinen et al., 2018; Pande, 2003).⁵ Less is known about the importance of group affiliations in party-centered electoral environments, where parties have strong tools to discipline their elected officials.

⁴In her seminal work, Pitkin (1967) distinguishes between legislators who are descriptively similar and thereby “stand for” their constituents, and those who substantively “act for” constituents by promoting issues of concern to that group.

⁵The findings in this literature are however, not unequivocal, (see, e.g., Ferreira and Gyourko, 2014; Gagliarducci and Paserman, 2012).

Previous studies on substantive representation have often focused on specific underrepresented groups.⁶ This fragmentation limits our understanding of whether representation mechanisms are consistent across groups. We address this gap by adopting a comprehensive approach, simultaneously analyzing four key background characteristics.

Second, we contribute to an emerging literature on intraparty politics in list-based electoral systems. Much of this literature focuses on how parties allocate nominations and valuable positions to their members (e.g., Buisseret et al., 2022; Cox et al., 2021; Folke, Persson and Rickne, 2016; Fujiwara and Sanz, 2020; Meriläinen and Tukiainen, 2018) and how this shapes party stability (e.g., Buisseret and Prato, 2022; Cirone, Cox and Fiva, 2021; Matakos et al., 2018). Cirone, Cox and Fiva (2021), for example, argue that party elites create career paths within the party partly because they want to increase legislative cohesion. Empirically, it is complicated to quantify the extent to which party leaders control their rank-and-file members. Roll-call votes in parliamentary systems suffer from a number of problems that prevent them from forming a reliable basis for estimating individual legislators’ ideal points (Goet, 2019; Hug, 2010; Peterson and Spirling, 2018; Schwarz, Traber and Benoit, 2017). We believe that our study, based on legislative speech, delivers important insights into intraparty dynamics in party-centered environments.

2. Empirical case: Norway 1981–2021

2.1 *Election system*

Norwegian national elections are held every fourth year in September (1981, 1985, ..., 2021) using closed-list proportional representation.⁷ Each parliamentary session starts in the first week of October and normally ends in the third week of June. Below, we define

⁶Bailer et al. (2022) provide a notable counterexample, examining how gender, migrant status, social class, and age influence parliamentary questions in the German Bundestag. Their findings suggest that descriptive representation fosters substantive representation, particularly early in legislators’ careers.

⁷An unusual constitutional feature is the lack of any provision for early parliamentary dissolution. Norway and Switzerland are the only Western European countries without this “safety valve” (Strøm and Swindle, 2002).

parliamentary sessions by the year in which they ended.

Seats are allocated in two rounds. First, regular seats are allocated at the district level using the Modified Sainte-Laguë method. Second, adjustment seats are given to parties that are underrepresented nationally after the first-tier seats have been allocated, provided that those parties reach an electoral threshold of 4% of the national vote count (Fiva and Smith, 2017). Candidate nominations and rank positions are formally determined by party conventions at the electoral district level. In our sample period (1981–2021), the parliament consisted of 155–169 members (MPs).

2.2 *Two-bloc politics*

The Norwegian party system has been shaped by political cleavages that originated from economic, geographical, and cultural factors (Rokkan, 1967). The policy space has, however, been well represented by a left-right dimension, where the main political divide went between the left-leaning social democratic bloc and the right-leaning conservative bloc (Strøm, 2022). The left-wing bloc traditionally consisted of two parties, the *Socialist Left Party* (SV) and the *Labour Party* (A). The right-wing bloc has been more fragmented and consists of the *Conservatives* (H), the *Liberals* (V), *Christian Democrats* (KrF), and the *Progress Party* (FrP). The *Centre Party* (Sp) has traditionally sided with the conservative bloc but switched sides in 2005 when they formed a government together with A and SV.⁸

Two-bloc politics has significantly shaped the formation of cabinets in Norway and contributed to the prevalence of minority governments. Norway ranks among the world's parliamentary democracies with one of the highest rates of minority governments (Strøm, 2022). In our 40-year sample period, Norway was governed by minority governments for about 29 years (see Table A.1). Despite being one of the countries with the highest prevalence of minority governments, Norway's democracy remains stable, likely because

⁸In addition to the seven main parties mentioned, four minor parties have been represented by a single MP in one or more election periods during our sample period (*Future for Finnmark*, the *Coastal Party*, the *Green Party*, and the *Red Party*).

it is historically rooted in a consensus-based political culture and has relatively high levels of political trust. The political consensus culture in Norway has likely emerged because of its relatively egalitarian nature. The presence of “cross-cutting cleavages” has created enough common ground across groups with potentially conflicting interests to avoid deep-seated conflicts (Lipset and Rokkan, 1967; Wollebæk, Brekke and Fladmoe, 2022). This, together with most elites supporting the basic institutions, has created an environment that accepts policymaking through compromise. However, in later years, there are concerns that controversial issues, such as immigration, identity politics, and urbanization, have increased divergence across groups (Gulbrandsen and Engelstad, 2013; Wollebæk, Brekke and Fladmoe, 2022).

2.3 Party discipline

Voting against one’s party on a whipped vote is the ultimate act of defiance (Proksch and Slapin, 2015). Like in most parliamentary systems, intraparty cohesiveness in roll-call voting is extremely high in Norway. In the 2017-2021 election period, for example, the seven main parties were individually united in 96% of cases.⁹ Generally, parties only allow legislators to break party ranks on issues of strong constituency interest (e.g., roads) or moral beliefs (e.g., abortion), and only when they do not threaten the standing of the government (Rasch, 1999).

2.4 Parliamentary committees

There are currently 12 standing committees (*fagkomiteer*) in the Norwegian Parliament. These have responsibility for the majority of parliamentary proceedings. The committee size varies from 11 to 18 members. Representatives are proportionally assigned to each committee, according to their party’s size in the parliament. The exception is *The Standing Committee of Scrutiny and Constitutional Affairs*, where all parties are represented.

⁹The sample includes roll-call votes recorded by the electronic voting device of the *Storting*, and therefore excludes unanimous and some near-unanimous decisions. Panel A of Figure A.1 illustrates that when a party is *not* united, it is typically because a small fraction of legislators broke with the party line. Panel B of Figure A.1 shows that party discipline has been stable at a high level for decades.

Figure A.2 shows the number of legislators from each party by the policy area of the standing committee.¹⁰ The figure shows that the largest parties in each political bloc, the Labour Party and the Conservatives, are typically represented by several legislators in all policy areas.¹¹ Most of the other parties, however, are typically represented by a single legislator in each policy area. Due to their relatively small size and the confidential nature of their proceedings, parliamentary committees foster an environment conducive to legislative compromise (Strøm, 2022).

2.5 *Parliamentary speech*

There are strict behavioral rules in the Norwegian Parliament. All speeches must be addressed to the parliamentary president and should strictly concern the matter that is discussed. The tone should be formal and the audience is not allowed to call out or demonstrate other forms of rowdy disapproval or agreement. In contrast to many other parliamentary systems, such as Finland’s (Simola, Nieminen and Tukiainen, 2023) and the United Kingdom’s (Proksch and Slapin, 2012), the speech length is strictly regulated by parliamentary rules in the Norwegian Parliament.¹²

In Norway’s party-centered political environment, parties possess both explicit and implicit authority to distribute floor time. Søyland and Høyland (2021) document that committee assignments and leadership positions within the committee responsible for preparing the issue under discussion, are particularly important for the allocation of speaking time. This indicates that parliamentary debates function as a platform for committee members to update the plenary on committee discussions. Given that MPs

¹⁰We choose to describe groupings by policy area rather than committee name, because some committees change names, merge, or split during our sample period (see Table A.2). In most instances, each policy area captures a single committee, but there are a few exceptions. For example, “Foreign Affairs” and “Defence” were two separate committees in the 1993–2009 period.

¹¹Figure A.3 plots seat shares over time for each of the main parties and an “other” category.

¹²The first speech of an ordinary debate is restricted to 15 minutes, while the second and third speeches are restricted to 10 and 3 minutes. Minor comments should not exceed 1 minute. Accounts by cabinet ministers should not exceed 1 hour. If the account is followed by a debate, one MP from each party is allowed 5 minutes to comment. MPs may speak only twice per topic unless exceptions are made by the parliamentary president. There are different ways in which MPs can ask questions of cabinet members, as explained by Søyland (2022). In *Interpellations*, ministers have one month to prepare responses to written questions. In *Oral Question Hour*, MPs pose short questions for immediate answers.

primarily engage in debates stemming from their respective committees, it is important to control for committee fixed effects in our analyses that follow below.

3. Data

3.1 *Four decades of parliamentary debates*

In this paper, we analyze plenary speeches in the Norwegian Parliament from October 1981 to June 2021 (N=623,863). Our data comes from the *Norwegian Parliamentary Debates Dataset*, which includes all speeches since 1945 (Fiva, Nedregård and Øien, 2025).

From March 2004 our data set includes time stamps. Figure A.4 shows the speech length distribution for this sub-sample. Speech length is measured as the time from the start of one speech to the beginning of the next, slightly overstating the actual speech length. The rules of conduct are evident in the empirical distribution of speech length, with distinct spikes just above the one-, three-, five-, and ten-minute marks.

We drop speeches by presidents and vice presidents of the Parliament (164,125 observations), since these contain formalities and parliamentary proceedings that are of little relevance to the question we are studying. Additionally, we remove individuals who have served as cabinet members at any point during a given election period, as their roles differ from those of regular MPs; for instance, they are responsible for providing parliamentary accounts and addressing questions from MPs (see footnote 12), as well as deputy MPs (115,433 observations). Deputy MPs substitute for MPs who are promoted to the cabinet or for some other reason are prevented from serving.

There are two official forms of written Norwegian, *bokmål* and *nynorsk*.¹³ The linguistic differences between the two languages might lead the model to predict a separation of speeches that, in reality, are equal in terms of topic and meaning. To address this problem, we exclude all speakers that predominantly use the minority language *nynorsk*,

¹³*Nynorsk* is used by a minority of the Norwegian population. Approximately 14 percent of Norwegian pupils had *nynorsk* as their main language in primary education in 2019/2020 (*Grunnskolen Informasjonssystem* <https://gsi.udir.no/>), and there are 15 percent *nynorsk* speeches in our data

defined as having a majority of their speeches in this form (38,186 observations), and individual *nynorsk* speeches (714 observations).

Lastly, we exclude party-independent MPs and MPs from minor parties (4,206 observations).¹⁴ This is to restrict our analysis to parties that have a clear bloc affiliation and existed for the entire time period that we are studying. After aggregating to session level, we are left with 4,896 MP-session observations (610 unique MPs).

Before standardizing the features, we eliminate the names of all MPs and cabinet members from parliamentary speeches. We then lemmatize all words to allow several versions of a word to be analyzed as one using the Oslo-Bergen tagger (Johannessen et al., 2012). Lemmatization is better than stemming at discriminating between words with different meanings, and hence yields a more accurate image of speech. To reduce the number of features to something manageable, a common first step is to strip out elements of the raw text other than words. In line with this convention, we remove all punctuation, numbers, symbols and parentheses. We also remove party names and party acronyms. Next, we retain words that occur more than ten times in at least one parliamentary session, spoken in at least ten unique speaker-sessions, and spoken at least twenty times across all sessions. We remove a set of stop words (Table A.11) and procedural words (Table A.12) which appear frequently but is not informative about the differences in speech that we aim to measure.¹⁵ Appendix B shows how feature engineering and lemmatization affect two example speeches. Because compound words are quite common in Norwegian, e.g., *velferdsstat* meaning “welfare state”, we rely on single words or unigrams as input into our analyses. After pre-processing, we are left with a parliamentary vocabulary of 16,595 unique lemmas. Table A.3 shows that, on average, a legislator speaks 45 times, and utters 4,348 words in a session (after pre-processing).

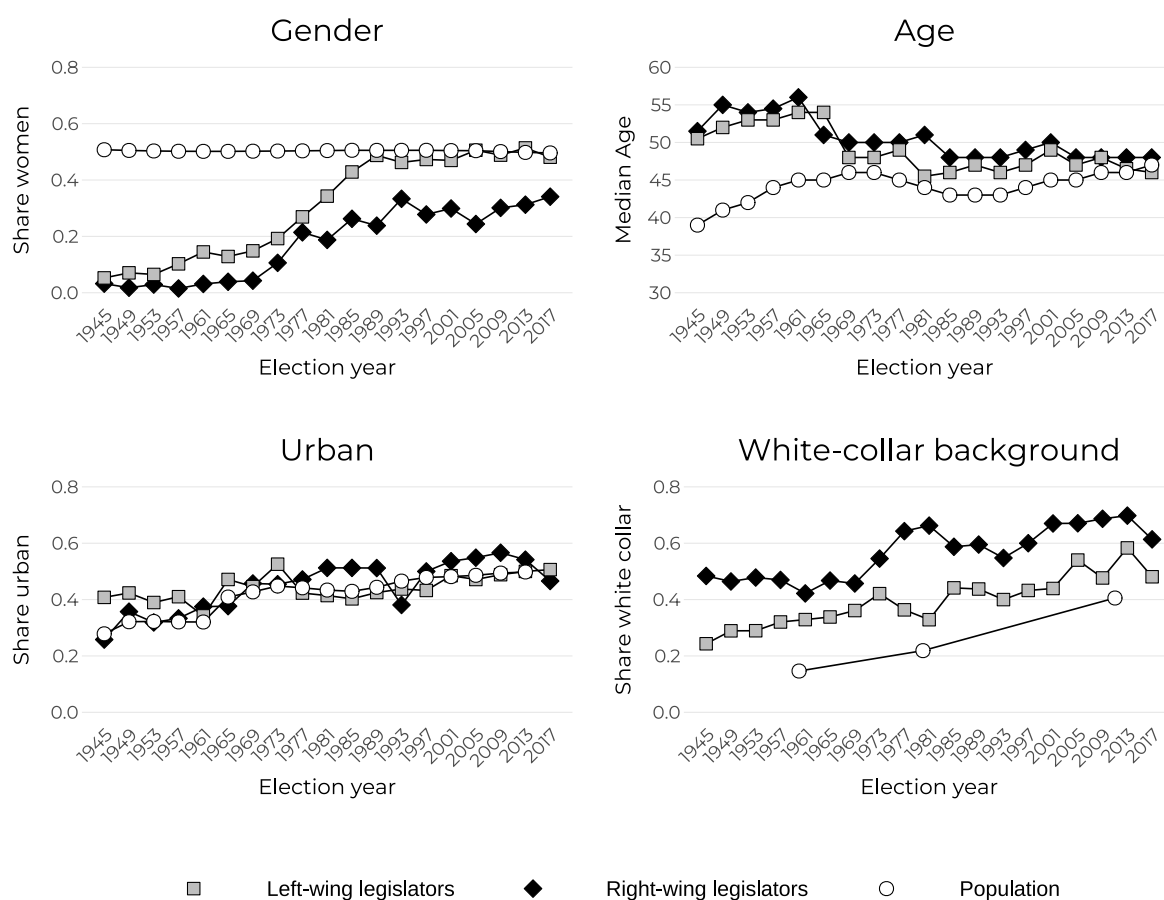
¹⁴We also remove one MP that did not belong to a committee in one election period.

¹⁵We also standardize a common linguistic class marker that is not related to substantive policy differences by changing all words that starts with ‘fra’ to starting with ‘fre’.

3.2 Characteristics of politicians

As mentioned above, we consider four background characteristics; gender, urbanicity, age, and class background. Most of this information is readily available from electoral lists and is organized by Fiva and Smith (2017). We manually supplement this data using biographies whenever necessary.

Figure 1: Descriptive representation over time



Note: The figure shows how background characteristics of left-wing legislators (gray squares), right-wing legislators (black diamonds), and the population (white circles) evolved over the 1945-2017 period. The left-wing parties are SV, A, and SP (2005-2017). The right-wing parties are V, KrF, H, FrP, and SP (1945-2001). The top-left panel plots the fraction of women in the legislature (population) by election year. The top-right panel plots the median age of legislators (median age of citizens, eighteen years and older) by election year. The bottom-left panel plots the fraction of legislators (citizens) residing in municipalities with “town” status (using data from Fiva and Smith (2017)). The bottom-right panel plots the fraction of legislators whose father held a white-collar occupation (ISCO codes 1, 2, 3, 4, 5, and a residual category including self-employed and capitalists) (using biographical information from the Archive of Politicians) by election year. For this panel, we do not have a complete population counterpart readily available. Instead, we rely on numbers from Modalsli (2017), which are based on father-son pairs identified in Norwegian censuses (1960 with fathers observed in 1910; 1980 with fathers observed in 1960; and 2011 with fathers observed in 1980).

Figure 1 displays how descriptive representation in the Norwegian Parliament has evolved since 1945 for each political bloc. The top-left panel shows that the fraction of female legislators in the left-wing bloc increased dramatically during the 1970s and 1980s. Since about 1990, there is close to gender parity within the left-wing bloc. The fraction of women legislators in the right-wing bloc has increased more modestly during our sample period. In total, women are still underrepresented in Parliament today.

An understudied aspect of political representation is the difference in age of the citizens and legislators (Gulzar, 2021). The top-right panel shows that during the period from 1945 to 1965, the median age of legislators elected in each bloc was approximately ten years higher than the median age of the adult population. However, in the subsequent election years, these age gaps gradually decreased. As of 2017, the median age of left-wing legislators was 46, right-wing legislators had a median age of 48, and the median age of the overall population stood at 47.

To distinguish between politicians from urban and rural areas, we follow Fiva and Smith (2017) and code legislators (citizens) residing in a municipality with historical town status as urban.¹⁶ The bottom-left panel of Figure 1 shows that urban areas are represented in parliament in about proportion to their size in the population for both blocs. This is likely driven both by the electoral law, which distributes seats *across* districts partly based on their population size, and parties' strong tendency to geographically balance their ticket *within* the district (Fiva, Halse and Smith, 2021).

To measure class background, we rely on information about fathers' occupation from the *Archive of Politicians* at the Norwegian Centre for Research Data.¹⁷ Using these biographical data, we classify fathers' occupations using ISCO-08 (International Standard Classification of Occupations) codes. In our main analysis, we measure class background

¹⁶Municipalities are traditionally designated as towns by the government, and identified by the second digit in their municipality identifier number being "0".

¹⁷We use father's occupation – the traditional proxy for social background (Buis, 2013; Marks, 2008) – because the vast majority (85 percent) of MPs elected in our sample period is born between 1915 and 1970, a period in which fathers were the primary earners and mothers' primary role was to manage the household. Another alternative measure of class background could be based on politicians' own pre-office occupation. However, because most MPs come from white-collar jobs, we quickly run into problems with statistical precision when analyzing these data based on the methods presented below.

using two broad categories: Politicians whose father held a “blue-collar occupation” (ISCO codes 6–9; including farmers) and politicians whose father held a “white-collar occupation” (ISCO codes 1–5; including a residual “other” category).

The bottom-right panel of Figure 1 contrasts the development in legislators’ social background with corresponding numbers from Modalsli (2017), based on father-son pairs identified in Norwegian population censuses (1960 with fathers observed in 1910; 1980 with fathers observed in 1960; and 2011 with fathers observed in 1980). There is a positive trend in the share of fathers having a white-collar occupation for both left-wing legislators, right-wing legislators, and the general (prime-age male) population. The substantial level difference between the three curves suggests that even in the comparatively egalitarian case of Norway, elected politicians are a privileged elite.

Table A.4 shows that the background characteristics overlap only to a modest degree. The pairwise correlations vary from -0.106 to 0.182 . The most notable distinctions indicate that white-collar legislators tend to be somewhat younger and more urban than their blue-collar counterparts. There are no clear systematic associations between politicians’ gender, age, and urbanicity.

In Table A.5 we present descriptive statistics for the speech data by legislator background characteristics. On average, women speak somewhat less than men. We find similar average differences for the other background characteristics: the young speak slightly more than the old; urban legislators speak slightly more than rural legislators; and legislators with white-collar backgrounds speak slightly more than legislators with blue-collar backgrounds.

4. Methods

As mentioned above, we build on the work of Gentzkow, Shapiro and Taddy (2019) to quantify differences in speech patterns by legislators with different background characteristics. In this section, we explain how we proceed.

4.1 Measurement and estimation of political divergence

Gentzkow, Shapiro and Taddy (2019) specify a multinomial model of speech where choice probabilities, $\mathbf{q}_t^{C(i)}(\mathbf{x}_{it})$, defined over the J available words for speaker i in session t , vary by party.¹⁸ In our case, $C^k(i)$ denotes the group affiliation of characteristic k for individual i where $k \in \{\text{Bloc, Gender, Age, Urban, Background}\}$, and we let $A_t^k = \{i : C^k(i) = A, m_{it} > 0\}$ and $B_t^k = \{i : C^k(i) = B, m_{it} > 0\}$ denote the set of active speakers in indicator and reference category, respectively, where m_{it} is total number of words spoken by individual i in session t . Our set of indicator categories for our five characteristics is $A \in \{\text{Right-wing, Female, Old, Urban, White-collar}\}$ and our set of reference categories is $B \in \{\text{Left-wing, Male, Young, Rural, Blue-collar}\}$. To estimate the choice probabilities for each background characteristic, we run separate multinomial logistic regressions. The model is represented by the following equations:

$$q_{jt}^{C^k(i)}(\mathbf{x}_{it}) = \frac{e^{u_{ijt}^k}}{\sum_l e^{u_{ilt}^k}}, u_{ijt}^k = \alpha_{jt}^k + \mathbf{x}'_{it} \boldsymbol{\gamma}_{jt}^k + \phi_{jt}^k \mathbf{1}_{i \in A_t^k}. \quad (1)$$

Here α_{jt}^k is word j 's baseline popularity for characteristic k at time t , $\boldsymbol{\gamma}_{jt}^k$ is a K -vector capturing the effects of \mathbf{x}_{it} on the propensity to use word j for characteristic k in session t , and our coefficient of interest ϕ_{jt}^k measures the effect of characteristic k affiliation on the propensity to use word j in session t . Because we are interested in measuring whether legislative speech is distinguishable according to legislators' background characteristics within political party over time, we include party and session fixed effects in \mathbf{x}_{it} . In addition, we include committee fixed effects to rule out that potential differences are driven by allocation of committee membership.¹⁹

¹⁸Speech by legislator i in session t is a J -dimensional vector of word counts, \mathbf{c}_{it} . The j -th element in \mathbf{c}_{it} , c_{ijt} , is the number of times legislator i used word j in session t . This model assumes that speeches are generated as independent draws from a multinomial distribution. Specifically, the speech \mathbf{c}_{it} is based on a vector of choice probabilities $\mathbf{q}_t^{C(i)} = \{q_{t1}, \dots, q_{tJ}\}^{C(i)}$, where $q_{tj}^{C(i)}$ is the probability of choosing word j in session t by legislator i . Given the total number of words $m_{it} = \sum_j c_{ijt}$ spoken by i in session t , the model can be summarized as follows: $\mathbf{c}_{it} \sim \text{Multinomial}(m_{it}, \mathbf{q}_t^{C(i)})$. In this formulation, legislator i selects a total of m_{it} words according to the probabilities specified by $\mathbf{q}_t^{C(i)}$.

¹⁹If a politician switches committees during a parliamentary session, we use the committee where he/she spent most days to construct committee fixed effects.

To quantify the divergence between $q_t^{A^k}$ and $q_t^{B^k}$, we use the expected posterior at \mathbf{x} that a neutral observer identifies group k affiliation after hearing one word:

$$\boldsymbol{\pi}_t(\mathbf{x}) = \frac{1}{2} \mathbf{q}_t^{A^k}(\mathbf{x}) \cdot \boldsymbol{\rho}_t(\mathbf{x}) + \frac{1}{2} \mathbf{q}_t^{B^k}(\mathbf{x}) \cdot (\mathbf{1} - \boldsymbol{\rho}_t(\mathbf{x})), \quad (2)$$

where $\rho_{jt}(x) = \frac{q_{jt}^{A^k}(x)}{q_{jt}^{A^k}(x) + q_{jt}^{B^k}(x)}$ is the posterior probability a neutral observer assigns to an individual being part of A^k after hearing word j . This measure reflects the expected posterior probability that an observer with a neutral prior correctly identifies the speaker’s group k affiliation after hearing a single word. For a neutral observer, with probability $\frac{1}{2}q_{jt}^{A^k}$, the posterior probability of being correct is $\rho_{jt}(x)$, and with probability $\frac{1}{2}q_{jt}^{B^k}$, the posterior probability of being correct is $1 - \rho_{jt}(x)$. The expected posterior of being correct for word j is $\frac{1}{2}q_{jt}^{A^k} \rho_{jt}(x) + \frac{1}{2}q_{jt}^{B^k} (1 - \rho_{jt}(x))$. By summing over all j words, we obtain the expected posterior of being correct for an “average” word at \mathbf{x} , as shown in Equation (2). The average political divergence, $\bar{\pi}_t$, which is the measure we report in the results section, is the average of Equation (2) across the values of \mathbf{x} .

The value of $\bar{\pi}_t$, which ranges from 0.5 to 1, represents the probability of correctly identifying the speaker based on an average word. In the absence of any information, the probability of a correct guess is 0.5. A $\bar{\pi}_t$ value exceeding 0.5 indicates that the average word contains information to distinguish whether the speaker belongs to group A or group B. This suggests that differences in the legislators’ backgrounds significantly influence their speech patterns, thereby shaping the policymaking process.

To identify which words are important in distinguishing speech by group A^k and B^k , we follow Gentzkow, Shapiro and Taddy (2019) and measure the change in the expected posterior at \mathbf{x}_{it} that a speaker belongs to group A^k after removing word j from the vocabulary:

$$\zeta_{jt}(\mathbf{x}_{it}) = \frac{1}{2} - \frac{1}{2} \sum_{s \neq j}^J \left(\frac{q_s^{A^k}(\mathbf{x}_{it})}{1 - q_j^{A^k}(\mathbf{x}_{it})} \rho_s + \frac{q_s^{B^k}(\mathbf{x}_{it})}{1 - q_j^{B^k}(\mathbf{x}_{it})} \rho_s \right). \quad (3)$$

The divergence of word j is the average of $\zeta_{jt}(\mathbf{x}_{it})$ across all active speakers i . This measures the change in the expected posterior of the speaker belonging to A^k , because

for a neutral observer, the posterior that the speaker belongs to A^k is equal to $\frac{1}{2}$, and the corrected posterior after removing word j is given by the right-most term in Equation (3).²⁰ This measure of divergence has both direction and magnitude. Intuitively, if the removal of the word j from the vocabulary results in the corrected expected posterior falling below $\frac{1}{2}$, we have removed a word predicting that the speaker belongs to A^k . The more positive the difference between $\frac{1}{2}$ and the corrected posterior, the more divergent the word is in favor of A^k . If we remove a word and the corrected posterior is above $\frac{1}{2}$, we have removed a word predicting the speaker belongs to B^k , and the more negative the difference in Equation (3), the more divergent the word is in favor of B^k . More divergent words pull $\bar{\pi}_t$ above 0.5 and towards 1, indicating a stronger predictive power of words for identifying the speaker’s group affiliation.

4.2 Estimation

The model is estimated using distributed multinomial regression (*dmr*), developed by Taddy (2015). This method approximates the logit likelihood in Equation (1) with J -independent Poisson likelihoods. In high-dimensional speech data, traditional estimation methods are computationally infeasible. *dmr* assumes that word counts are independently Poisson distributed with mean $e^{\log(m_{it})+u_{ijt}}$, where m_{it} is the speech length of speaker i in session t . This transforms the negative log-likelihood for each word j into:

$$l(\alpha_{jt}, \gamma_{jt}, \phi_{jt}) = \sum_t \sum_i \left[m_{it} \exp(\alpha_{jt} + x'_{it} \gamma_{jt} + \phi_{jt} \cdot 1_{i \in A_t^k}) - c_{ijt} (\alpha_{jt} + x'_{it} \gamma_{jt} + \phi_{jt} \cdot 1_{i \in A_t^k}) \right] \quad (4)$$

The estimation is done by fitting J -independent Poisson regressions, providing a computational advantage as the regressions can be done entirely in parallel.

To address small sample bias, a Lasso penalty is included on the coefficient of interest.

The estimation is performed by solving the following minimization problem:

²⁰Define $\rho_j^{A^k}$ as the posterior probability that a speaker belongs to A^k . The expected posterior over words for a neutral observer, ignoring \mathbf{x}_{it} for simplicity, is $1/2 \sum_{j=1}^J [q_j^{A^k} + q_j^{B^k}] \rho_j = 1/2 \sum_{j=1}^J [q_j^{A^k} + q_j^{B^k}] \frac{q_j^{A^k}}{q_j^{A^k} + q_j^{B^k}} = 1/2 \sum_{j=1}^J q_j^{A^k} = 1/2$

$$\hat{\alpha}_{jt}, \hat{\gamma}_{jt}, \hat{\phi}_{jt} = \arg \min_{\alpha_{jt}, \gamma_{jt}, \phi_{jt}} \left[l(\alpha_{jt}, \gamma_{jt}, \phi_{jt}) + N \sum_t [\psi(|\alpha_{jt}| + \|\gamma_{jt}\|_1) + \lambda_j |\phi_{jt}|] \right] \quad (5)$$

The Lasso penalty, $\lambda_j |\phi_{jt}|$, shrinks coefficients, producing a sparse solution and reducing small sample bias. The optimal λ_j is determined by minimizing a corrected version of Akaike’s information criterion, starting with a high λ_j and gradually decreasing it. A constant penalty $\psi = 10^{-5}$ is applied to other coefficients to ensure convergence.

4.3 *Magnitudes*

The average divergence $\bar{\pi}_t$ represents the posterior that a neutral observer assigns to a speaker’s true identity after hearing *a single word*. However, political speeches usually last for several minutes. How does the probability of correctly guessing a speaker’s characteristic change with the number of words? We quantify the informativeness of speech by speech length and session using Monte Carlo simulations. For each speaker i and session t , we draw words according to the estimated choice probabilities in Equation (2) and compute the average divergence.

4.4 *Validation*

As previously mentioned, the number of words a legislator could choose is large relative to the total amount of speech we observe. As a consequence, many words are said mostly by one type of legislator purely by chance. The estimator we use controls for this bias by applying a lasso penalty (see equation (5)) but is still biased in finite samples.

To quantify this bias, we rely on a permutation test by randomly reassigning $C^k(i) \in \{A^k, B^k\}$ and re-estimating the model 100 times. In these series, $q_t^{A^k} = q_t^{B^k}$ by construction, so the true value of $\boldsymbol{\pi}_t(\mathbf{x})$ is 0.5. The deviation from 0.5 provides a valid measure of finite-sample bias under the permutation.

5. Results

5.1 *Across-bloc divergence*

The left-hand panel of Figure 2 illustrates the average divergence (defined by Equation (2)) across party blocs for each year in our sample period. The gray shaded area represents the average divergence in hypothetical data in which each speaker’s party bloc is randomly assigned with the probability that the speaker is right-wing (left-wing). The upper and lower bounds on the light gray shaded area correspond to the 5th and the 95th percentiles of the placebo distributions. The dark gray shaded area represents the corresponding 10th and 90th percentiles. For each year in our sample period, the observed divergence consistently falls outside the placebo distribution, providing strong statistical evidence of political polarization in legislative speech.

In the first two decades of our sample period, the estimated $\bar{\pi}$ falls in the range 0.503–0.505. The initial twenty years reflect a time when Norway, like numerous Western nations, experienced a political shift towards the right. Interestingly, the left-to-right divergence in parliamentary speech does not seem to increase in this period. This may be explained by the strategic response of the Labor Party to the right-wing wave. The Labor Party, by far the most influential party in post-war Norway, pivoted its politics towards a more market-friendly direction, emphasizing individual rights and freedoms (Njølstad, 2023; Sejersted, 2023). For instance, it was a minority labor government (Brundtland III), with support from the right, that signed the agreement to make Norway part of the European Economic Area (EEA) in 1992, resulting in a significant liberalization of the Norwegian economy (Sejersted, 2023). The EEA agreement and the subsequent 1994 referendum on EU membership, were the defining issues of the 1990s (Notaker, 2023). However, as illustrated in Figure 2, these issues did not change the left-to-right divergence in Norwegian politics, likely because the Conservative and Labor Party, the two dominant forces on either side of the political spectrum, held similar positions on these matters.

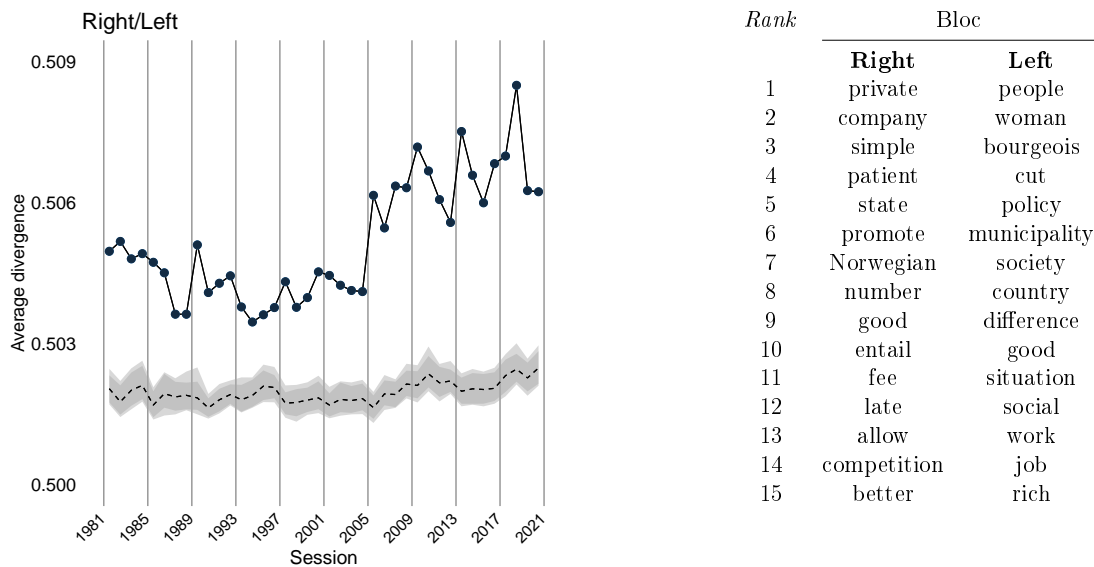
In the second half of our sample period, a rising trend in across-bloc divergence emerges, seemingly starting after the 2005 election.²¹ This aligns well with the literature documenting the growing importance of issue competition in Western Europe, i.e. competition on which issues should dominate the political agenda (e.g., Green-Pedersen, 2007, 2019). This is also reflected in the 15 most divergent words for each bloc across the full sample period, which we show in the right-hand panel of Figure 2 (Appendix A.6 gives the top-fifty). The top words suggest that parties do not compete solely by offering different policy positions (as in, e.g., Downs (1957)), but also by highlighting distinct issues – preferably issues they ‘own’; these are issues for which a majority of the electorate traditionally perceives them as competent and capable in addressing (Abou-Chadi, 2016; Petrocik, 1996). Right-leaning legislators appear concerned about the performance of the private sector, taxes, and national identity. Our model identifies “private”, “company”, “competition”, “fee” and “Norwegian” as key right-leaning words. Left-leaning legislators emphasize redistribution and employment, and tend to use terms like “people”, “society”, “work”, and “rich”.

When interpreting Figure 2, the reader should keep in mind that $\bar{\pi}$ is the posterior that a neutral observer expects to assign to a speaker’s true party bloc after hearing a *single word* from the vocabulary (defined in section 3.1). As a rough comparison, Gentzkow, Shapiro and Taddy (2019) report a $\bar{\pi}$ of about 0.502 – 0.504 for the 1870–1990 period in the United States Congress, which later increases to about 0.510. Simola, Nieminen and Tukiainen (2023) find lower levels of divergence in the Finnish Parliament. For the period after 2000, they find a $\bar{\pi}$ of about 0.502 – 0.504.²² In section 5.3, we consider the

²¹In the 2005 election, a center-left coalition (Stoltenberg II) dominated by the Labor Party won the parliamentary majority. In the election campaign, and as the leading party of the center-left coalition, the Labor Party reverted to more traditional, at least rhetorically, left-wing issues. This reversion is probably a response to the spectacular loss of Stoltenberg’s first government, which can be considered to be the peak of right-leaning Labor governments in Norway. Stoltenberg I implemented a series of significant reforms, including privatization, tax and pension system overhaul, and the introduction of performance measures and corporatization in public management. However, these reforms were unpopular with voters, resulting in the Labor Party’s parliamentary share plummeting from 39% to 26% in the 2001 election, its worst performance since the interwar period.

²²There are several reasons why our estimates are not directly comparable to the ones reported by Gentzkow, Shapiro and Taddy (2019) and Simola, Nieminen and Tukiainen (2023). For example, both of these studies rely on bigrams rather than unigrams.

Figure 2: Across-bloc divergence over time



Note: In the left-hand panel the black points correspond to the average bloc divergence of speech for each session in the period 1981–2020 after controlling for legislators’ committee assignment. The black vertical lines indicate elections. The gray shaded area represents the average divergence in hypothetical data in which each speaker’s party bloc is randomly assigned with a probability that the speaker is right-wing. We construct 100 hypothetical data sets and compute the average divergence in each session. The upper and lower bounds of the light gray shaded area correspond to the 5th and the 95th highest scores across the placebo distributions. The dark gray shaded area represents the corresponding 10th and 90th highest scores. The dashed line corresponds to the mean divergence for each session across the distribution of placebo estimates. In the right-hand panel, we provide a list of the 15 words with the highest relative utility for each bloc.

informativeness of legislative speech by speech length.

5.2 Within-party divergence

Figure 3 illustrates within-party political divergence in parliament over time across four background dimensions: gender, age, urbanicity, and class background. To aid in the interpretation and validation of our findings, Table 3 lists the top 15 most divergent words for each background dimension, and a more comprehensive listing of the top 50 most divergent words for each dimension can be found in Tables A.7 to A.10.

The estimated within-party divergences across all dimensions are somewhat below the across-bloc divergence established in Figure 2. However, we can typically rule out that observed differences in speech are driven by random variations since observed differences (black points) fall outside the placebo distribution (gray shaded area). Figure A.5 contrast

the distribution of actual estimates (black points in Figure 3) to the distribution of mean placebo estimates (black dashed line in Figure 3) for each background characteristics. There is minimal overlap between the distributions for gender and age, while some overlap is noticeable for the two other background characteristics.

Gender

According to the literature review by O'Brien and Piscopo (2019, p.54), scholars typically categorize women's interest as those (i) issues that directly affect women *as women* (e.g., reproductive health), (ii) issues connected to women's traditional role as caregivers (e.g., children), and (iii) issues tied to the social sphere more broadly (e.g., health care and education). Consistent with this categorization, data from the *Norwegian National Election Survey* (1981–2017) reveal that female voters focus more on child care, education, and welfare policies (see Appendix C). Table 3 indicates that these policy issues are mirrored by female legislators in parliament. For example, the word that most clearly separates women from men legislators (belonging to the same policy committee and party) is “children”, followed by “woman” and “work” (see also Table A.7). These gender differences aligns with the findings of Lippmann (2022), who study lawmaking in the French parliament, and Baskaran and Hessami (2025) which examines council meeting discussions in Bavaria.

Age

The existing empirical evidence on the effect of legislators' age is relatively scarce. Poole (2007) found that US Congress members' voting records demonstrate a high degree of continuity across a politician's lifetime suggesting that age is not important. However, two recent papers that use regression discontinuity designs and data from Germany and Japan, respectively, come to the opposite conclusion (politicians' age matter for policy) (Baskaran, Hessami and Schirner, 2023; McClean, 2021). The upward trend in the top-right panel of Figure 3 suggest that the age dimension have become increasingly im-

portant over time in the Norwegian parliament. Table 3 reveal that young legislators talk more about childcare (e.g., “kindergarten”) and schooling (e.g., “school”), while their older counterparts appear to prioritize health-related topics (e.g., “treatment”, “patient”) (see Appendix Table A.8). We see no evidence, based on the top-fifty lists, that older legislators talk more about pensions, as one might have expected from survey evidence (Appendix C).

Urban-Rural

During the initial half of our sample period, we find no clear evidence that the rural-urban status of a legislator affects parliamentary speech. In the bottom-left panel of Figure 3 the divergence estimates frequently position in the upper range, but they do not consistently exceed the placebo distribution. However, in the latter half our sample period, the divergence estimates are trending upwards.²³ At the end of the sample period, our estimated divergence is comparable to the one we find for other background characteristics. The increasing urban-rural divide could be a second-order manifestation of deeper demographic and cultural divides, as observed in other European countries (Gallego et al., 2016; Maxwell, 2019; Rodríguez-Pose, 2018).

Knutsen (2004) argues that rural populations tend to support parties addressing agrarian concerns, decentralization, economic support for smaller municipalities, and welfare policies targeting rural communities. The interests that separate rural from urban interests are mirrored in parliamentary speeches. Across our full sample period, we observe that rural legislators frequently discuss regional policy (e.g., “municipality”, “district”), farming (e.g., “agriculture”), and sparsely populated areas (e.g., “Finnmark” and “Hedmark”). In short, rural legislators appear to talk about topics that resonate with the concerns of rural voters (see Appendix C). In contrast, urban legislators often address issues encountered in cities (e.g., “city”, “police”, and “prison”), as shown in Table 3.

²³Huijsmans and Rodden (2025) similarly find that urban-rural electoral divides in Norway have increased in the same period and document analogous patterns in other European countries.

Class

We find that politicians whose fathers have a white-collar occupation tend to talk more about children and education (e.g., “children”, “kindergarten” and “school”), while politicians with blue-collar backgrounds tend to talk more about farming (e.g., “agriculture” and “farmer”) and defense-related topics (e.g., “defense”) . These findings complement related studies from the United States and Latin America, which typically finds that working-class politicians tend to propose and cosponsor bills that are more leftist on labor, economic, and redistributive issues (Carnes and Lupu, 2023). In the context of party-centered political systems, the impact of legislators’ occupational and social backgrounds remains less explored. However, a notable study by O’Grady (2019) using speech data from the United Kingdom, reveals that the Labour Party’s shift towards career politicians from working-class MPs significantly diluted the advocacy for working-class voters’ interests.

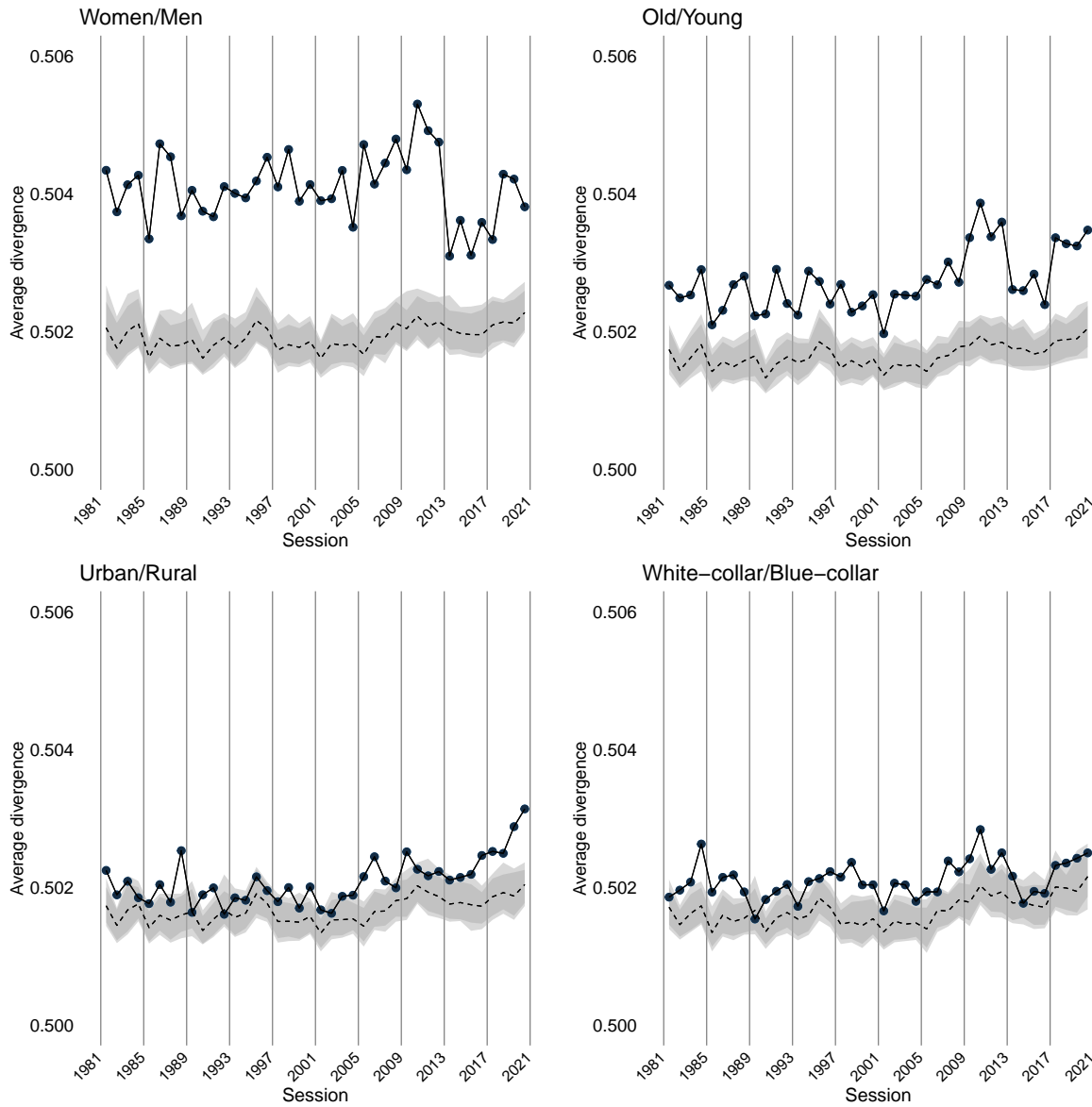
5.3 *Magnitudes*

In Figure 4, we plot the average expected posterior across two time periods, by the length of speech, for each dimension we consider.²⁴ In the first half of our sample period (1982–2001), we find that the probability that a neutral observer correctly guesses the speaker’s bloc affiliation is about 62 (78) percent after hearing one minute (three minutes) of speech.²⁵ In line with Figure 2, we observe that it becomes easier to predict a legislator’s bloc affiliation over time. In the second half of our sample period (2002–2021), we estimate that the probability that a neutral observer correctly guesses the speaker’s bloc affiliation is about 67 (83) percent after hearing one minute (three minutes) of speech. When interpreting these numbers one should keep in mind that parliamentary speeches might reflect more cohesion within a party than actually exists because party leaders make it

²⁴In Figure A.6 we show the results for individual sessions.

²⁵After pre-processing our data, the median number of words per minute of speech is 37.14 . To calculate the median number of words per minute of speech, we use data from March 2004 onwards, which, as previously mentioned, includes a time stamp for each speech. Before pre-processing, one minute of speech is 150 words (see section 3.1 and Appendix B).

Figure 3: Within-party divergence over time



Note: This figure displays divergence in legislative speech for four dimensions (given in the sub-panel headings) in the period 1982-2021, controlling for party and the legislator's committee assignment. The models are estimated separately for each background characteristic. The black points correspond to the average divergence of speech in each session and the black vertical lines indicate elections. The gray shaded area represents the average divergence in hypothetical data in which each speaker's identity is randomly assigned. We construct 100 hypothetical data sets and compute the average divergence in each session. The dashed line corresponds to the mean divergence for each session across the distribution of placebo estimates. The upper and lower bounds of the light gray shaded area correspond to the 5th and the 95th highest scores across the placebo distributions. The dark gray shaded area represents the corresponding 10th and 90th highest scores.

Table 1: Most divergent words for each background dimension

| <i>Rank</i> | Gender | | Age | |
|-------------|---------------|--------------|------------------|--------------|
| | Women | Men | Old | Young |
| 1 | children | Norwegian | Nordic | day |
| 2 | woman | relationship | cooperation | policy |
| 3 | work | political | mention | select |
| 4 | measures | lie | Norwegian kroner | Norway |
| 5 | school | policy | million | people |
| 6 | increase | fall | emphasize | trade |
| 7 | parent | that is | very | chamber |
| 8 | family | expression | municipality | relationship |
| 9 | young | post | director | job |
| 10 | education | director | number | ensure |
| 11 | child welfare | context | patient | tie |
| 12 | kindergarten | belief | state | children |
| 13 | municipality | of course | area | kindergarten |
| 14 | people | foundation | company | Norwegian |
| 15 | competence | Norway | country | school |

| <i>Rank</i> | Urbanicity | | Father's occupation | |
|-------------|--------------------|------------------|---------------------|-------------|
| | Urban | Rural | White | Blue |
| 1 | Norwegian | municipality | children | defense |
| 2 | that is | good | policy | good |
| 3 | city | agriculture | Norway | vision |
| 4 | international | relationship | that is | patient |
| 5 | separate | district | police | agriculture |
| 6 | country | defense | state | director |
| 7 | Norway | Finmark | woman | area |
| 8 | Bergen | Akershus | school | district |
| 9 | ensure | Norwegian kroner | kindergarten | light |
| 10 | police | county | municipality | attempt |
| 11 | problem | Hedmark | human | if |
| 12 | tie | million | tie | of course |
| 13 | political | register | cash support | day |
| 14 | prison | children | Bergen | Hedmark |
| 15 | Minister of Health | Nordland | political | farming |

Note: This table displays the 15 most divergent words of legislative speech for four dimensions: gender, age, urbanicity, and class. Each model is estimated separately and controls for the speaker's party and committee assignment.

difficult for rebels to express their views on the floor (Proksch and Slapin, 2015).

We estimate that the probability of correctly guessing the speaker’s gender, age, urbanicity status, or occupational background category is close to, but slightly below, the estimated across-bloc divergence during the first half of our sample period. A one-minute speech gives a neutral observer approximately 56 – 63 percent chance of correctly identifying the relevant background characteristic.

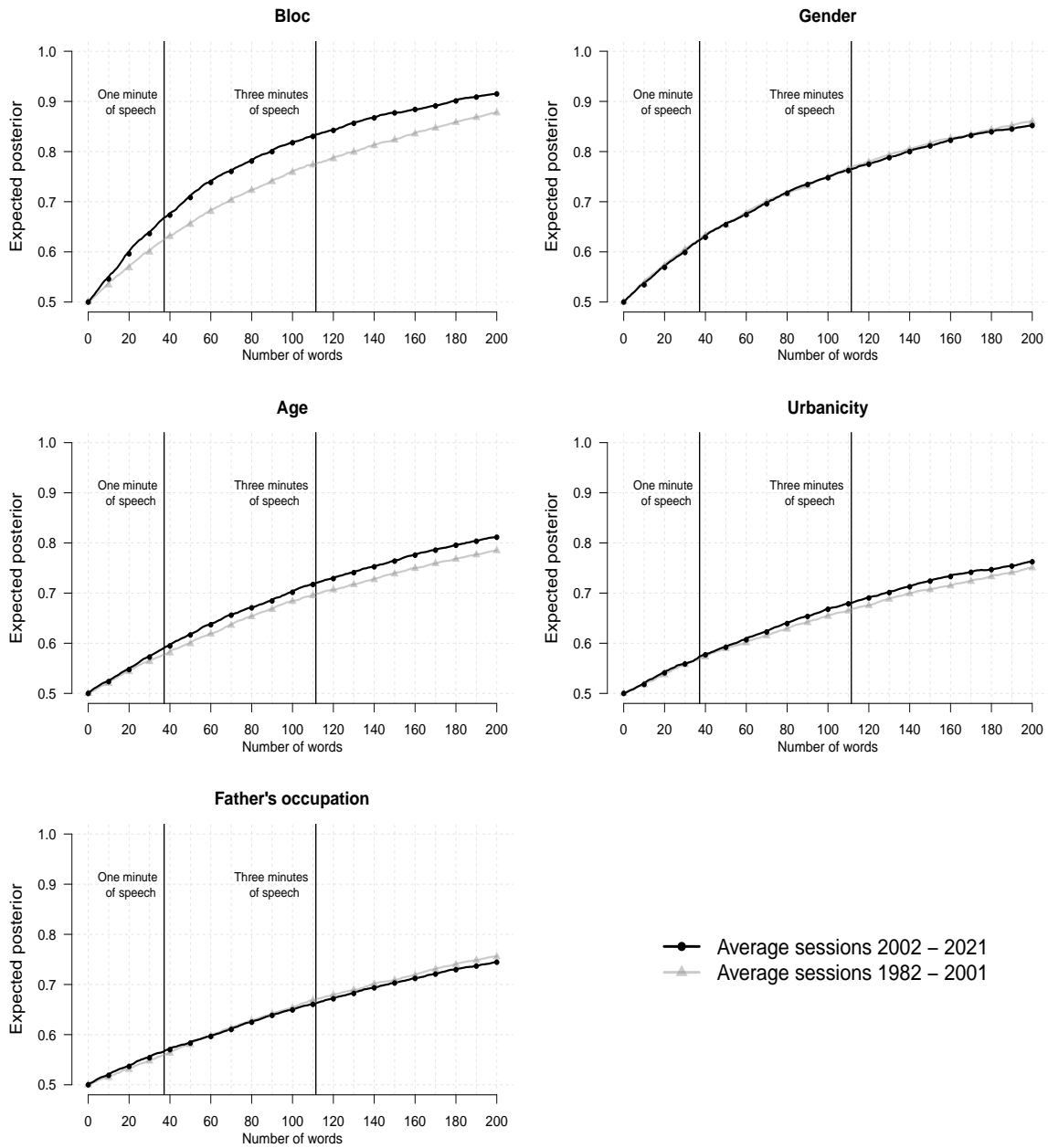
5.4 *Sensitivity checks*

Our baseline analyses suggest that legislator background characteristics matter for legislative speech. Next, we explore the sensitivity of this result to various modelling choices.

First, we repeat our baseline analyses without controlling for party affiliation. In other words, we compare legislators with different background characteristics (from the same committee) both across and within parties. Even though candidate background characteristics vary across political blocs (Figure 1), we find that the estimated background divergence change only slightly when party fixed effects are omitted (Figure A.7). When compared with our baseline results, we find that it is somewhat easier to guess the group identity of speakers, especially towards the end of our sample period, when party fixed effects are excluded. In the case of gender, our estimates are almost unaltered throughout our sample period (i.e. it is about equally easy to distinguish between men and women belonging to the same party as it is to distinguish between men and women belonging to different parties). This suggests that political parties do not strongly orchestrate parliamentary speech.

Next, we explore the role of standing committees in the parliament. Individual committees are unlikely to be mirror images of the parliament both because of demand and supply forces. On the demand side, parties might keep an eye on the background characteristics of legislators when arranging the composition of committees. On the supply side, politicians with different backgrounds may be interested in different policy areas and may therefore self-select into different committees.

Figure 4: Political divergence by speech length



Note: The figure shows average divergence as a function of the number of words after data pre-processing (see section 3.1 for details). The expected posterior is calculated by drawing 200 words for each speaker i and session t , given characteristics x_{it} using the estimated choice probabilities. The expected posterior as a function of the number of words is calculated as follows: Within each session, we calculate the average probability that a neutral observer assigns to the speaker's true identity after hearing the first word. Then we use this probability as the prior, when we calculate the posterior probability after hearing an additional word. This is continued until 200 words are spoken. The average lines in the figure are found by taking the average over the session-specific, expected posterior. The vertical line marks the median number of words per minute of speech. This number is calculated by dividing the number of words by the number of minutes for each speaker-session observation and taking the median of this ratio across observations. The median one-minute speech is calculated using data from the 2004 session, which contains the starting time of each speech.

We find a substantial imbalance in the composition of committees regarding gender (Figure A.8), but less so for the other background characteristics (Figure A.9 – A.11).²⁶ For example, women are underrepresented in the *Committee of Finance* and overrepresented in the *Committee of Family and Cultural Affairs* throughout our sample period. This imbalance is particularly strong in the 2005–2009 election period. In these parliamentary sessions, the *Committee of Finance* and the *Committee of Family and Cultural Affairs* consisted of 14% and 75% women, respectively.

Our analyses control for committee assignment via fixed effects. In other words, we are comparing differences in speech patterns across women and men, who belong to the same committee and party. If we drop the committee fixed effects, we reach qualitatively the same conclusions as in our baseline analysis. As one might expect, the differences in the estimates for gender divergence occur when committee assignment is least balanced across genders. The most salient difference is, however, that the placebo distributions widen without committee fixed effects (Figure A.13).

6. Conclusion

Political parties can be conceptualized as a team of candidates “tethered by a rubber band to the ideology espoused by the parties whose label they run on” (Grofman, 2008). In the context of the closed-list system we examine, the rubber band is not expected to be very elastic. Indeed, roll-call vote data from Norway suggest that political elites limit within-party divergence to a minimum. Nevertheless, we argue that it would be premature to conclude that party leaders have complete control over their rank-and-file legislators. Four decades of legislative speech data indicate that legislators tend to advocate for the policy preferences of voters who share their background characteristics, in line with the commitment issues highlighted in the citizen-candidate framework (Alesina, 1988; Besley

²⁶Figures A.8 – A.11 document the development of group representation in committees by parliamentary session. Figure A.12 displays group representation in committees relative to their share in Parliament for our entire sample period.

and Coate, 1997; Osborne and Slivinski, 1996).

In contrast to much of the existing literature on substantive representation, which tends to focus on a single characteristic, we take a comprehensive approach by concurrently considering four key background traits. For all background characteristics, including those that are mostly unobservable to voters (such as class background), we consistently find that estimates of within-party divergence fall (i) below our benchmark estimates of across-bloc divergence but (ii) above placebo estimates. In addition, we find that the inclusion of party fixed effects does not substantially change how easy it is to separate background pairs. Together, these findings suggest that the results are not driven by party elites strategically assigning speeches to speakers based on their descriptive characteristics. Instead, our results show that even in party-centered environments legislators' group identities matter for what they bring to the table.

References

- Abou-Chadi, Tarik. 2016. “Niche party success and mainstream party policy shifts—how green and radical right parties differ in their impact.” *British Journal of Political Science* 46(2):417–436.
- Alesina, Alberto. 1988. “Credibility and Policy Convergence in a Two-Party System with Rational Voters.” *American Economic Review* 78:796–806.
- Bäck, Hanna, Marc Debus and Jochen Müller. 2014. “Who Takes the Parliamentary Floor? The Role of Gender in Speech-making in the Swedish Riksdag.” *Political Research Quarterly* 67(3):504–518.
- Bailer, Stefanie, Christian Breunig, Nathalie Giger and Andreas M. Wüst. 2022. “The Diminishing Value of Representing the Disadvantaged: Between Group Representation and Individual Career Paths.” *British Journal of Political Science* 52(2):535–552.
- Baskaran, Thushyanthan and Zohal Hessami. 2025. “Women in Political Bodies as Policymakers.” *Review of Economics and Statistics* forthcoming.
- Baskaran, Thushyanthan, Zohal Hessami and Sebastian Schirner. 2023. “Can Young Politicians Influence Policy in a Gerontocracy?” Working paper, Ruhr-University Bochum.
- Baumgartner, Frank R, Christian Breunig and Emiliano Grossman. 2019. *Comparative policy agendas: Theory, tools, data*. Oxford University Press.
- Besley, Timothy and Stephen Coate. 1997. “An Economic Model of Representative Democracy.” *Quarterly Journal of Economics* 112(1):85–114.
- Blumenau, Jack. 2021. “The Effects of Female Leadership on Women’s Voice in Political Debate.” *British Journal of Political Science* 51(2):750–771.
- Buis, Maarten L. 2013. “The composition of family background: The influence of the economic and cultural resources of both parents on the offspring’s educational attainment in the Netherlands between 1939 and 1991.” *European Sociological Review* 29(3):593–602.
- Buisseret, Peter and Carlo Prato. 2022. “Competing Principals? Legislative Representation in List Proportional Representation Systems.” *American Journal of Political Science* 66(1):156–170.
- Buisseret, Peter, Olle Folke, Carlo Prato and Johanna Rickne. 2022. “Party Nomination Strategies in List Proportional Representation Systems.” *American Journal of Political Science* 66(3):714–729.
- Carnes, Nicholas and Noam Lupu. 2023. “The economic backgrounds of politicians.” *Annual Review of Political Science* 26:253–270.

- Catalinac, Amy and Lucia Motolinia. 2021. “Why Geographically-Targeted Spending Under Closed-List Proportional Representation Favors Marginal Districts.” *Electoral Studies* 71:102329.
- Chattopadhyay, Raghavendra and Esther Duflo. 2004. “Women as Policy Makers: Evidence from a Randomized Policy Experiment in India.” *Econometrica* 72(5):1409–1443.
- Cirone, Alexandra, Gary W. Cox and Jon H. Fiva. 2021. “Seniority-Based Nominations and Political Careers.” *American Political Science Review* 115(1):234–251.
- Clayton, Amanda, Cecilia Josefsson and Vibeke Wang. 2017. “Quotas and Women’s Substantive Representation: Evidence from a Content Analysis of Ugandan Plenary Debates.” *Politics & Gender* 13(2):276–304.
- Cox, Gary W., Jon H. Fiva, Daniel M. Smith and Rune J. Sørensen. 2021. “Moral Hazard in Electoral Teams: List Rank and Campaign Effort.” *Journal of Public Economics* 200.
- Cox, Gary W. and Matthew D. McCubbins. 2005. *Setting the Agenda: Responsible Party Government in the U.S. House of Representatives*. New York: Cambridge University Press.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York, NY: Harper and Row.
- Ferreira, Fernando and Joseph Gyourko. 2014. “Does Gender Matter for Political Leadership? The Case of US mayors.” *Journal of Public Economics* 112:24–39.
- Fiva, Jon H., Askill H. Halse and Daniel M. Smith. 2021. “Local Representation and Voter Mobilization in Closed-list Proportional Representation Systems.” *Quarterly Journal of Political Science* 16(2):185–213.
- Fiva, Jon H. and Daniel M. Smith. 2017. “Norwegian Parliamentary Elections, 1906-2013: Representation and Turnout Across Four Electoral Systems.” *West European Politics* 40(6):1373–1391.
- Fiva, Jon H, Oda Nedregård and Henning Øien. 2025. “The Norwegian Parliamentary Debates Dataset.” *Scientific Data* 12, 4.
- Folke, Olle, Torsten Persson and Johanna Rickne. 2016. “The Primary Effect: Preference Votes and Political Promotions.” *American Political Science Review* 110(3):559–578.
- Fujiwara, Thomas and Carlos Sanz. 2020. “Rank Effects in Bargaining: Evidence from Government Formation.” *The Review of Economic Studies* 87(3):1261–1295.
- Gagliarducci, Stefano and M Daniele Paserman. 2012. “Gender interactions within hierarchies: evidence from the political arena.” *The Review of Economic Studies* 79(3):1021–1052.
- Gallego, Aina, Franz Buscha, Patrick Sturgis and Daniel Oberski. 2016. “Places and preferences: A longitudinal analysis of self-selection and contextual effects.” *British Journal of Political Science* 46(3):529–550.

- Gentzkow, Matthew, Jesse M Shapiro and Matt Taddy. 2019. “Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech.” *Econometrica* 87(4):1307–1340.
- Goet, Niels D. 2019. “Measuring Polarization with Text Analysis: Evidence from the UK House of Commons, 1811–2015.” *Political Analysis* 27(4):518–539.
- Green-Pedersen, Christoffer. 2007. “The Growing Importance of Issue Competition: The Changing Nature of Party Competition in Western Europe.” *Political Studies* 55(3):607–628.
- Green-Pedersen, Christoffer. 2019. *The reshaping of West European party politics: Agenda-setting and party competition in comparative perspective*. Comparative Politics.
- Grofman, Bernard. 2008. 102 The Impact of Electoral Laws on Political Parties. In *The Oxford Handbook of Political Economy*. Oxford University Press.
- Gulbrandsen, Trygve and Fredrik Engelstad. 2013. Elite consensus on the Norwegian welfare state model. In *Norway in Transition*. Routledge pp. 194–214.
- Gulzar, Saad. 2021. “Who Enters Politics and Why?” *Annual Review of Political Science* 24:253–275.
- Hug, Simon. 2010. “Selection Effects in Roll Call Votes.” *British Journal of Political Science* 40(1):225–235.
- Huijismans, Twan and Jonathan Rodden. 2025. “The Great Global Divider? A Comparison of Urban-Rural Partisan Polarization in Western Democracies.” *Comparative Political Studies* 58(2):261–290.
- Hyytinen, Ari, Jaakko Meriläinen, Tuukka Saarimaa, Otto Toivanen and Janne Tukiainen. 2018. “Public employees as politicians: Evidence from close elections.” *American Political Science Review* 112(1):68–81.
- Itzkovitch-Malka, Reut, Guy Mor, Odelia Oshri and Shaul Shenhav. 2024. “Talking representation: How legislators re-establish responsiveness in cases of representational deficits.” *European Journal of Political Research* 63(3):950–972.
- Ivanusch, Christoph. 2024. “Issue competition in parliamentary speeches? A computer-based content analysis of legislative debates in the Austrian Nationalrat.” *Legislative Studies Quarterly* 49(1):203–221.
- Johannessen, Janne Bondi, Kristin Hagen, André Lynum and Anders Nøklestad. 2012. OBT+ stat. A Combined Rule-based and Statistical Tagger. In *Exploring Newspaper Language. Corpus compilation and research based on the Norwegian Newspaper Corpus*, ed. Gisle Andersen. John Benjamins Publishing Company pp. 51–65.
- Knutsen, Oddbjørn. 2004. *Urban-Rural Residence*. London: Palgrave Macmillan UK pp. 132–158.

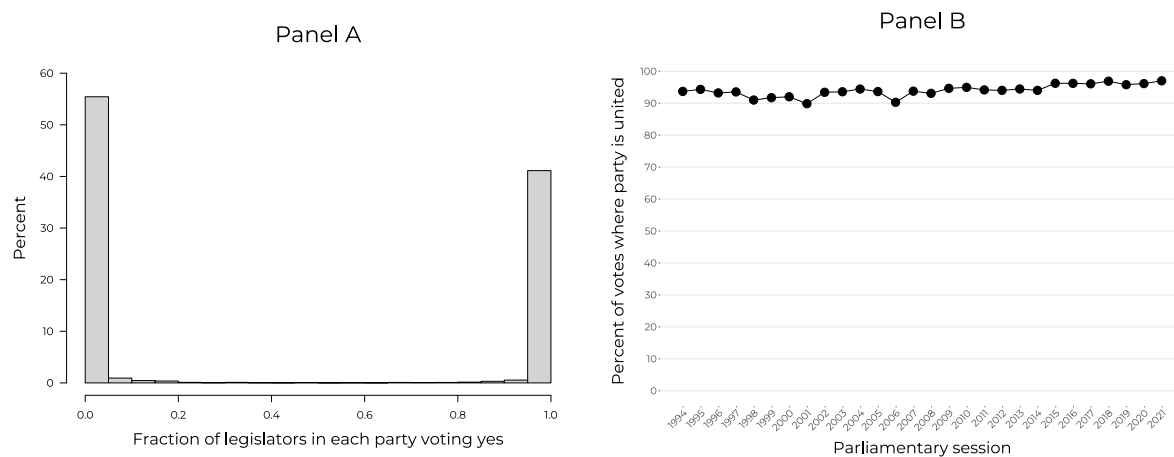
- Krehbiel, Keith. 1993. "Where's the Party?" *British Journal of Political Science* 23(2):235–266.
- Laver, Michael. 2021. Analyzing the Politics of Legislative Debate. In *The Politics of Legislative Debates*, ed. Hanna Back, Marc Debus and Jorge M Fernandes. Oxford: Oxford University Press chapter 2, pp. 21–33.
- Lippmann, Quentin. 2022. "Gender and lawmaking in times of quotas." *Journal of Public Economics* 207:104610.
- Lipset, Seymour Martin and Stein Rokkan. 1967. *Party Systems and Voter Alignments: Cross-national Perspectives*. Toronto: The Free Press.
- Marks, Gary Neil. 2008. "Are father's or mother's socioeconomic characteristics more important influences on student performance? Recent international evidence." *Social Indicators Research* 85:293–309.
- Martin, Lanny W and Georg Vanberg. 2008. "Coalition government and political communication." *Political Research Quarterly* 61(3):502–516.
- Matakos, Konstantinos, Riikka Savolainen, Orestis Troumpounis, Janne Tukiainen and Dimitrios Xefteris. 2018. "Electoral institutions and intraparty cohesion." *VATT Institute for Economic Research Working Papers* 109.
- Maxwell, Rahsaan. 2019. "Cosmopolitan immigration attitudes in large European cities: Contextual or compositional effects?" *American Political Science Review* 113(2):456–474.
- McClellan, Charles T. 2021. "Does the Underrepresentation of Young People in Political Institutions Matter for Social Spending?" Working paper, University of Michigan.
- Meriläinen, Jaakko and Janne Tukiainen. 2018. "Rank Effects in Political Promotions." *Public Choice* 177(1-2):87–109.
- Michelle Heath, Roseanna, Leslie A. Schwindt-Bayer and Michelle M. Taylor-Robinson. 2005. "Women on the Sidelines: Women's Representation on Committees in Latin American Legislatures." *American Journal of Political Science* 49(2):420–436.
- Modalsli, Jørgen. 2017. "Intergenerational mobility in Norway, 1865–2011." *The Scandinavian Journal of Economics* 119(1):34–71.
- Njølstad, Olav. 2023. "Arbeiderpartiets fornyelse på 1980-tallet." <https://www.norgeshistorie.no/oljealder-og-overflod/1913-arbeiderpartiets-fornyelse-pa-1980-tallet.html>. Retrieved 30. oktober 2023 from [norgeshistorie.no](https://www.norgeshistorie.no).
- Notaker, Hallvard. 2023. "Mer marked, svakere stat." <https://www.norgeshistorie.no/oljealder-og-overflod/1904-mer-marked-svakere-stat.html>. Retrieved 30. oktober 2023 from [norgeshistorie.no](https://www.norgeshistorie.no).

- Osborn, Tracy and Jeanette Morehouse Mendez. 2010. "Speaking as Women: Women and Floor Speeches in the Senate." *Journal of Women, Politics & Policy* 31(1):1–21.
- Osborne, Martin J. and Al Slivinski. 1996. "A Model of Political Competition with Citizen-Candidates." *The Quarterly Journal of Economics* 111(1):65–96.
- O'Brien, Diana Z. and Jennifer M. Piscopo. 2019. The Impact of Women in Parliament. In *The Palgrave Handbook of Women's Political Rights*. Springer pp. 53–72.
- O'Grady, Tom. 2019. "Careerists Versus Coal-Miners: Welfare Reforms and the Substantive Representation of Social Groups in the British Labour Party." *Comparative Political Studies* 52(4):544–578.
- Pande, Rohini. 2003. "Can Mandated Political Representation Increase Policy Influence for Disadvantaged Minorities? Theory and Evidence from India." *American Economic Review* 93(4):1132–1151.
- Pansardi, Pamela and Michelangelo Vercesi. 2016. "Party Gate-Keeping and Women's Appointment to Parliamentary Committees: Evidence from the Italian Case." *Parliamentary Affairs* 70(1):62–83.
- Peterson, Andrew and Arthur Spirling. 2018. "Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems." *Political Analysis* 26(1):120–128.
- Petrocik, John R. 1996. "Issue Ownership in Presidential Elections, with a 1980 Case Study." *American Journal of Political Science* 40(3):825–850.
- Pitkin, Hanna F. 1967. *The Concept of Representation*. Berkeley, CA: University of California Press.
- Poole, Keith T. 2007. "Changing Minds? Not in Congress!" *Public Choice* 131(3-4):435–451.
- Poole, Keith T and Howard Rosenthal. 2000. *Congress: A political-economic history of roll call voting*. Oxford University Press on Demand.
- Proksch, Sven-Oliver and Jonathan B. Slapin. 2012. "Institutional Foundations of Legislative Speech." *American Journal of Political Science* 56(3):520–537.
- Proksch, Sven-Oliver and Jonathan B. Slapin. 2015. *The Politics of Parliamentary Debate: Parties, Rebels, and Representation*. Cambridge University Press.
- Rasch, Bjørn Erik. 1999. Electoral Systems, Parliamentary Committees, and Party Discipline: the Norwegian Storting in a Comparative Perspective. In *Party Discipline and Parliamentary Government*, ed. David M. Farrell Shaun Bowler and Richard S. Katz. Ohio State University Press Columbus pp. 121–140.
- Rodríguez-Pose, Andrés. 2018. "The revenge of the places that don't matter (and what to do about it)." *Cambridge journal of regions, economy and society* 11(1):189–209.

- Rokkan, Stein. 1967. *Geography, Religion and Social Class: Crosscutting Cleveages in Norwegian Politics*. In *Party Systems and Voter Alignments*. New York: Free Press.
- Schwarz, Daniel, Denise Traber and Kenneth Benoit. 2017. “Estimating Intra-Party Preferences: Comparing Speeches to Votes.” *Political Science Research and Methods* 5(2):379–396.
- Sejersted, Francis. 2023. “Norsk historie fra 1970 til 1990.” https://snl.no/norsk_historie_fra_1970_til_1990. Retrieved 30. oktober 2023 from *Store norske leksikon*.
- Simola, Salla, Jeremias Nieminen and Janne Tukiainen. 2023. “A Century of Partisanship in Finnish Political Speech.” Aboa Centre for Economics Discussion paper No. 160.
- Strøm, Kaare and Stephen M Swindle. 2002. “Strategic parliamentary dissolution.” *American Political Science Review* 96(3):575–591.
- Strøm, Kaare W. 2022. *Norway: A Land of Minority Governments*. In *Minority Governments in Comparative Perspective*. Oxford University Press.
- Søyland, Martin. 2022. “Party Control and Responsiveness: How MPs Use Variation in Lower-Level Institutional Design as an Electoral Responsiveness Mechanism.” Proceedings of the Digital Parliamentary Data in Action Workshop.
- Søyland, Martin G. and Bjørn Høyland. 2021. *Norway: Committee-Membership Matters, Party Loyalty Decides*. In *The Politics of Legislative Debate*, ed. Hanna Back, Marc Debus and Jorge M Fernandes. Oxford: Oxford University Press chapter 31, pp. 634–650.
- Taddy, Matt. 2015. “Distributed Multinomial Regression.” *The Annals of Applied Statistics* 9(3):1394–1414.
- Wollebæk, Dag, Jan-Paul Brekke and Audun Fladmoe. 2022. “Polarization in a consensual multi-party democracy – attitudes toward immigration in Norway.” *Journal of Elections, Public Opinion and Parties* pp. 1–24.

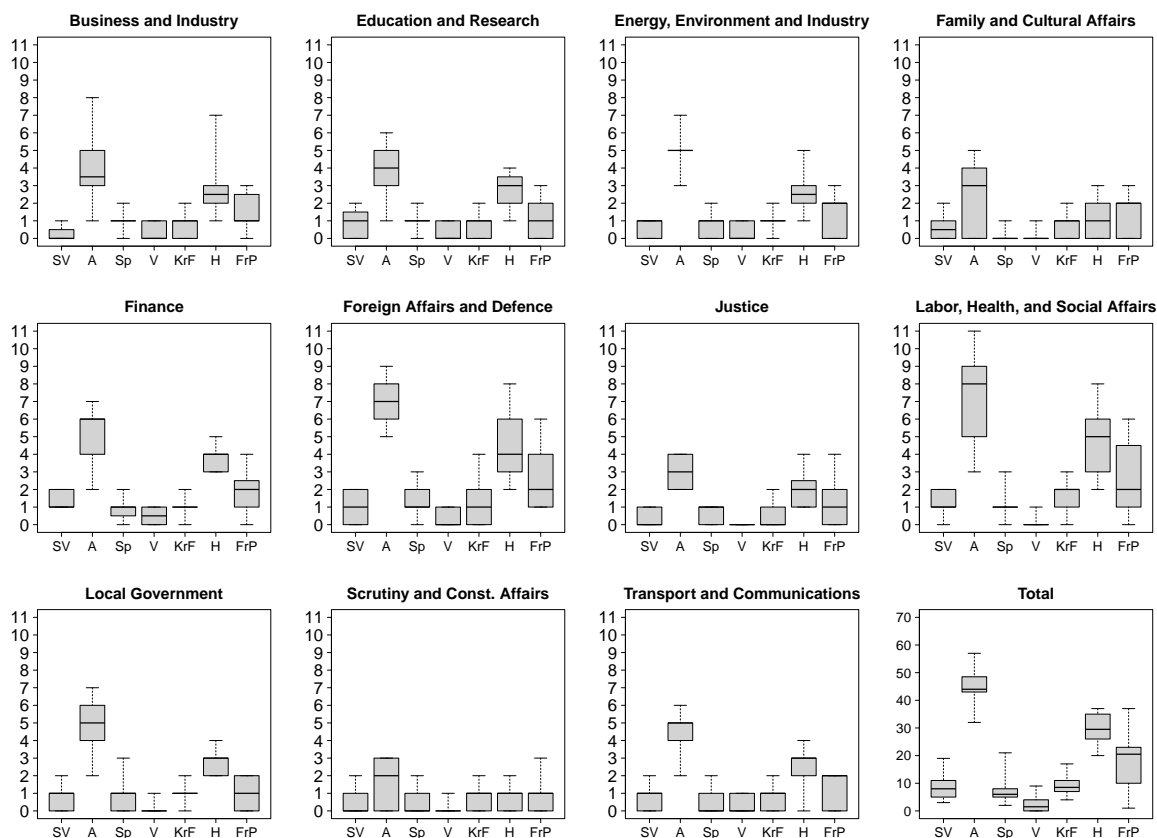
For Online Publication: Appendix A

Figure A.1: Party discipline measured by roll-call votes



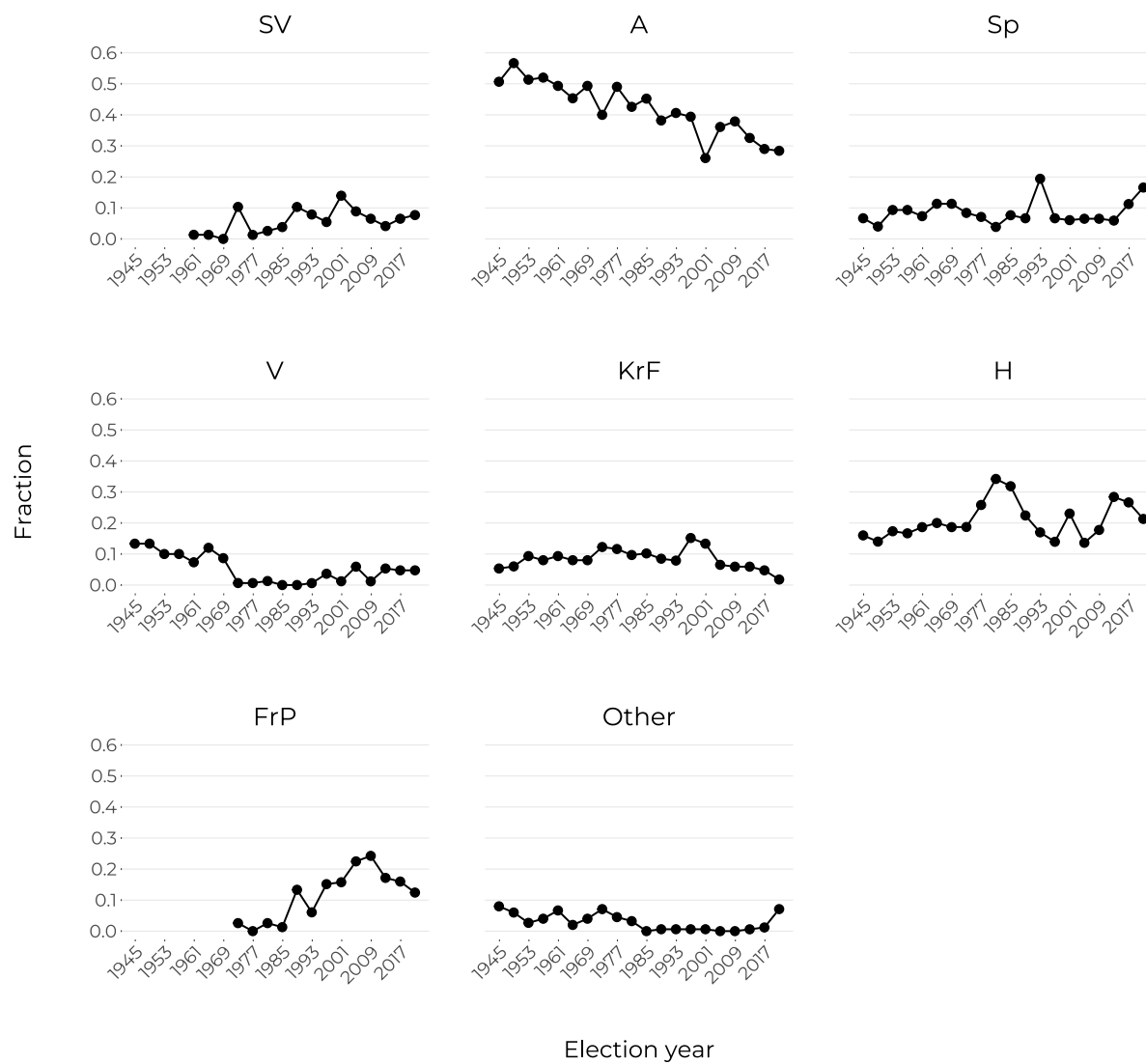
Note: Panel A shows the percent of legislators of each party voting yes to a proposal. Panel B shows the fraction of votes where a party is united by parliamentary session. The sample includes roll-call votes recorded by the electronic voting device of the Storting, excluding unanimous and near-unanimous decisions, in the parliamentary sessions 1994-2021 (October 1993-June 2021), i.e. seven full election periods. The unit of observation is the party-vote ($N = 152,562$). The main parties are: the Socialist Peoples' Party/Socialist Left Party (SV), the Labour Party (A), the Centre Party (Sp), the Christian Democrats (KrF), the Liberal Party (V), the Conservative Party (H), and the Progress Party (FrP). The data stem from Norwegian Agency for Shared Services in Education and Research (Sikt): <https://polsys.sikt.no/storting/voteringsarkivet>.

Figure A.2: Number of legislators from each party by policy area



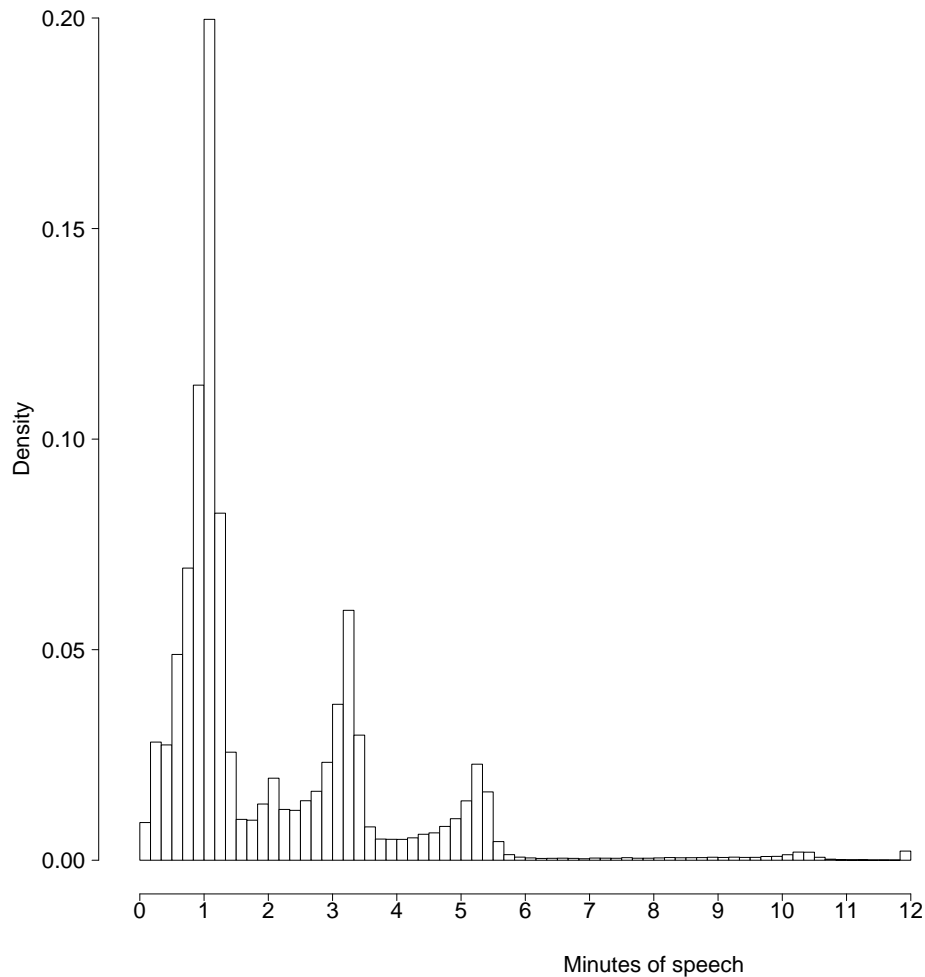
Note: This box-and-whisker plot shows the number of legislators from each main party by policy areas in the 1982-2021 period. Whiskers indicate the minimum and maximum values for each party. The main parties, ordered ideologically from “left” to “right” in the figure, are the Socialist Left Party (SV), the Labour Party (A), the Centre Party (Sp), the Christian Democrats (KrF), the Liberal Party (V), the Conservative Party (H), and the Progress Party (FrP). Because some committees change names, merge, or split during the sample period we present descriptive statistics by policy area, rather than by individual committee. Table A.2 provides an overview of the committee structure in the Norwegian Parliament, and how committees map into policy areas. If a politician switches committees during a parliamentary session, we use the committee where he/she spent most days. The committee compositions reflect the compositions in our sample (e.g., excluding Nynorsk speakers) and hence deviates somewhat from the general committee composition in the Norwegian parliament.

Figure A.3: Parties' seat shares by election year



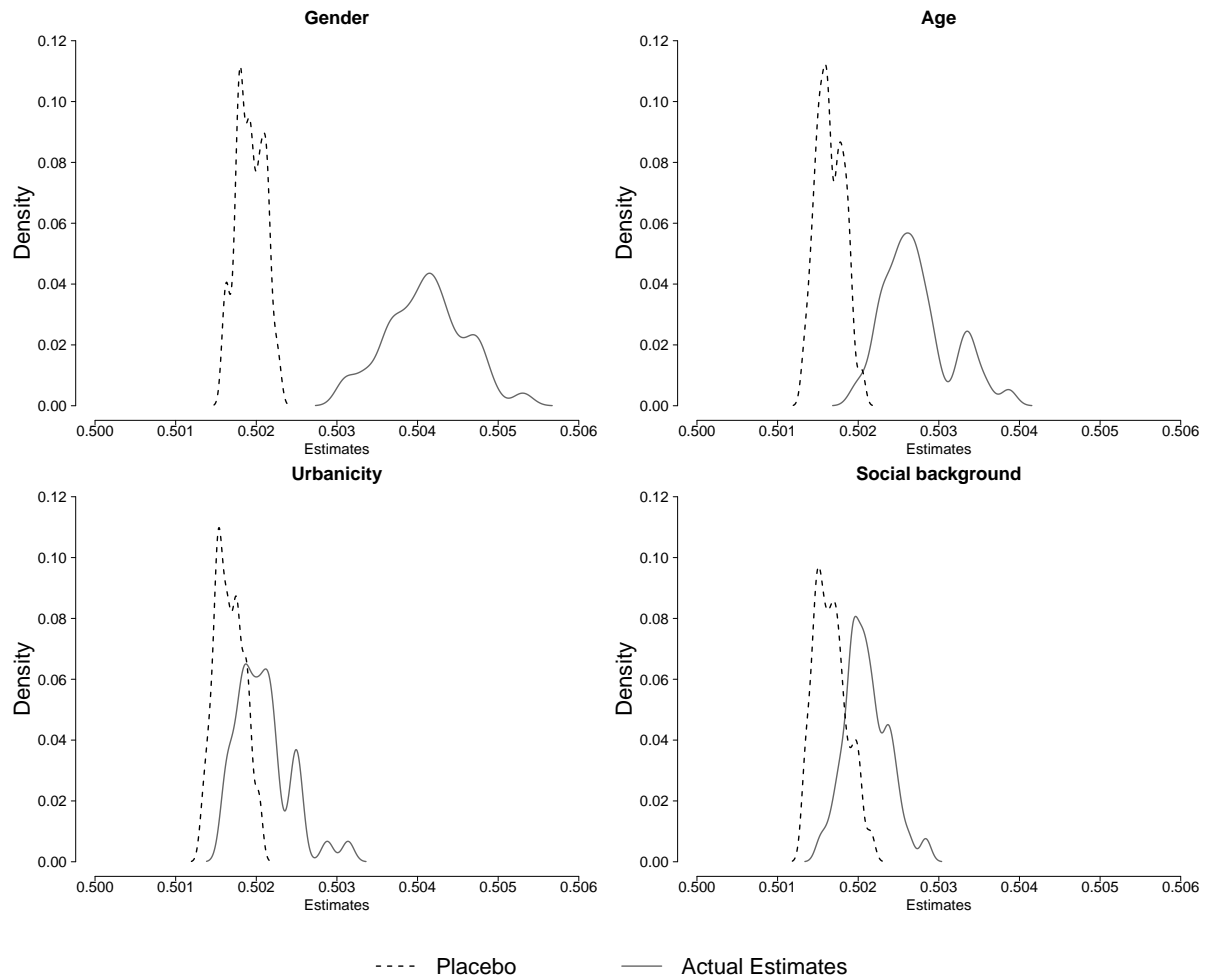
Note: Figure shows the parties' seat shares by election year. The main parties are: the Socialist Peoples' Party/Socialist Left Party (SV), the Labour Party (A), the Centre Party (Sp), the Christian Democrats (KrF), the Liberal Party (V), the Conservative Party (H), and the Progress Party (FrP). "Other" is a residual category for non-main parties.

Figure A.4: Empirical distribution of speech length



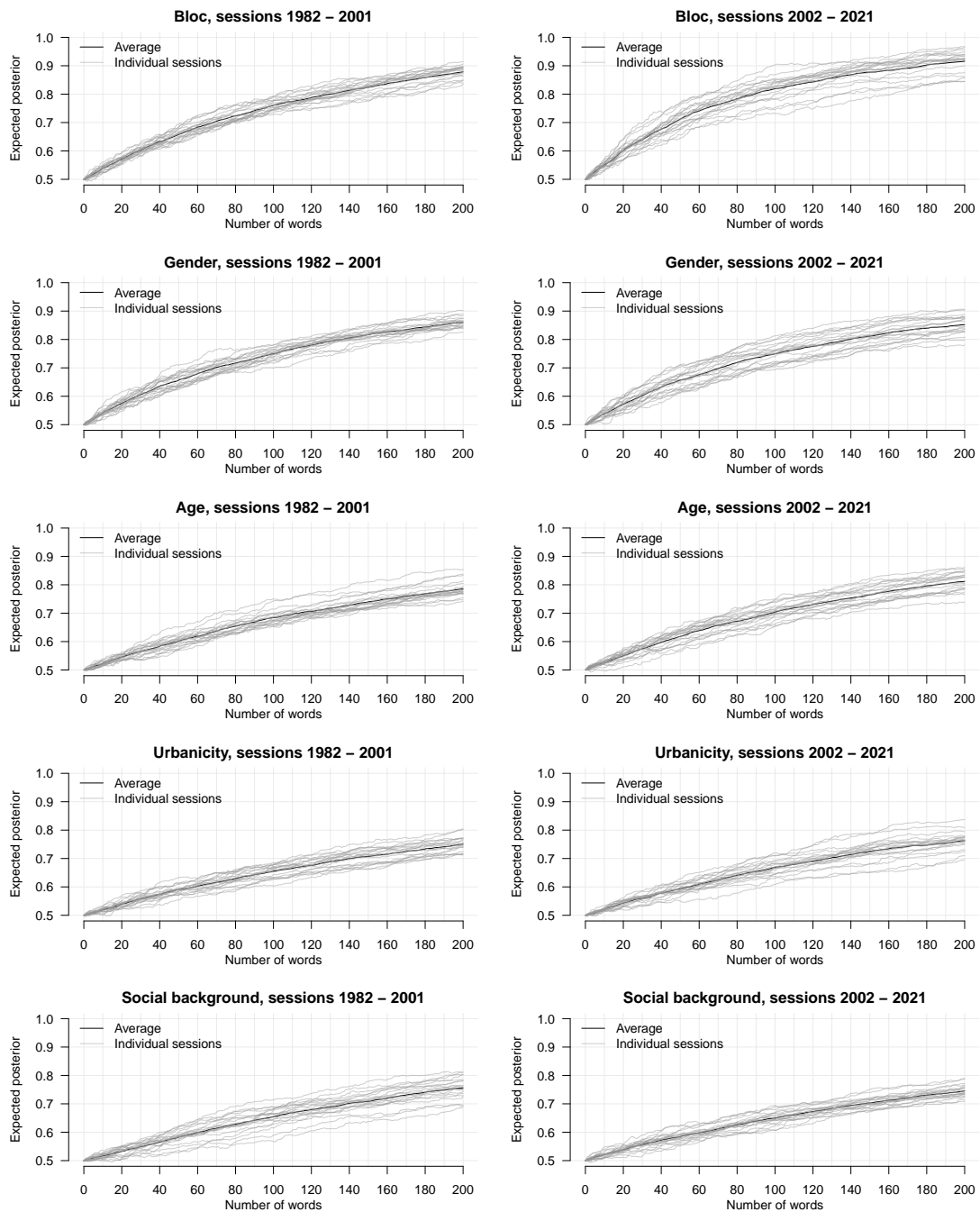
Note: The figure shows the distribution of speech length after and including the 2004 session. In this period, the data includes the timestamp at the beginning of every speech. Speech length is measured as the number of minutes from the start of the speech to the beginning of the next speech. Measured speech length slightly overstates the actual speech length by the amount of time it takes for the next speaker to start. However, because the parliamentary rules of conduct strictly control debate length, the time between speakers is generally minimal. This is clearly seen in the distribution. Speeches are clustered just above the one, three, and five-minute mark. The bin size is set to ten seconds. 40 percent of the speeches do not have a timestamp. These speeches are mostly very short speeches about parliamentary procedure. We discard them. Speech length measurement is nested within days. Therefore, there will be an overestimation of the last speech and interim speeches on days with multiple debates. We truncate the speech length at 12 minutes to mitigate the overestimation of speech length and improve the figure's readability.

Figure A.5: Comparison of actual estimates and mean placebo estimates for each election year



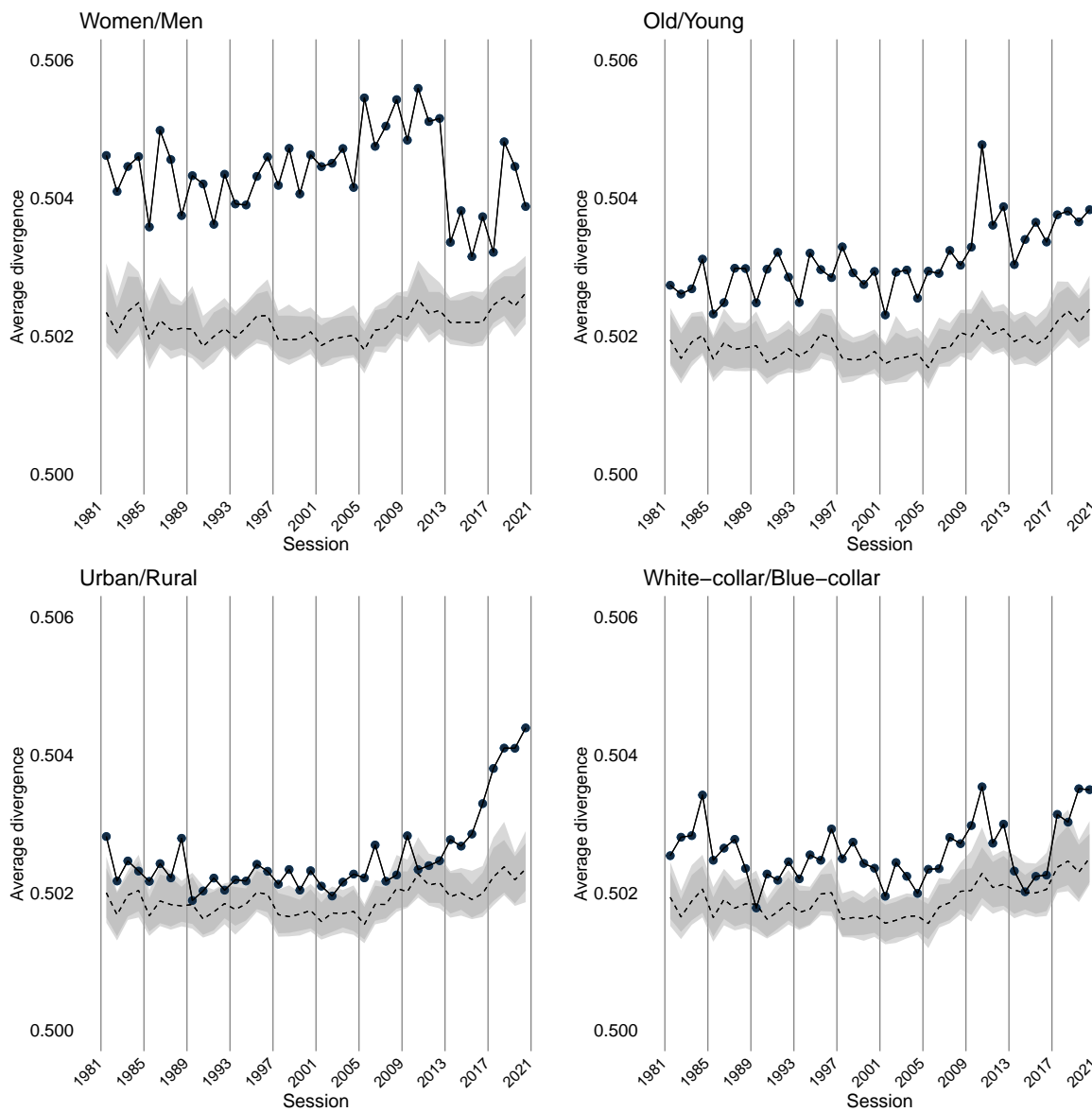
Note: This figure shows density plots of divergence estimates over the period 1982–2021. The solid gray line represents the density of yearly estimates using MPs' true group affiliation. The dashed gray line represents the session averages of placebo estimates, where MPs' background affiliations are randomized. The y-axis reflects the density of observations.

Figure A.6: Political divergence by speech length for individual sessions



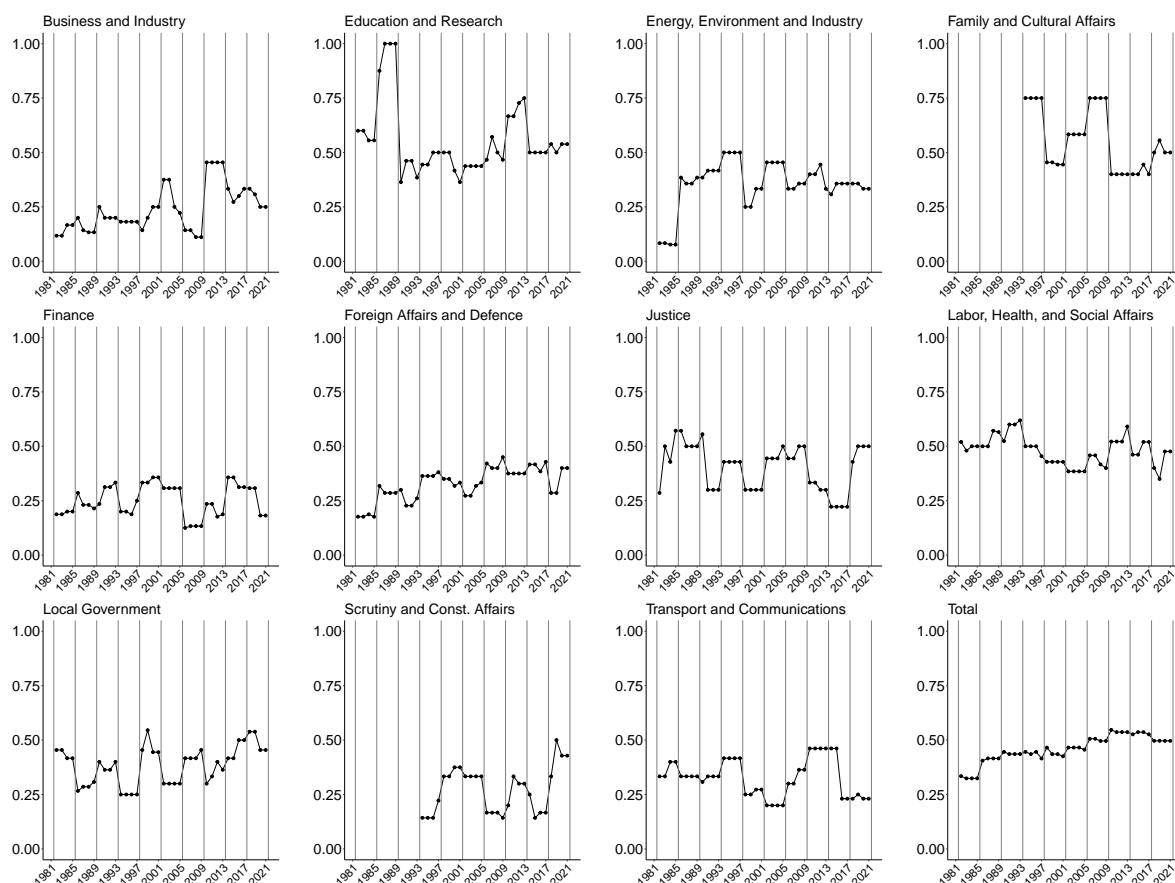
Note: The figure shows average gender, age, urbanicity, and social background divergence as a function of number of words for individual sessions and an average across all sessions. The expected posterior is computed by drawing 200 words for each speaker i and session t , given characteristics x_{it} using the estimated choice probabilities. The expected posterior is calculated in the following way: Within each session, we calculate the average probability that a neutral observer assigns to the speaker's true identity after hearing the first word. Then we use this probability as the prior when we calculate the posterior probability after hearing an additional word. This is continued until 200 words are spoken. The average lines in the figure are found by taking the average across session-specific expected posterior.

Figure A.7: Analysis of background divergence when controlling for committee assignment, but not party



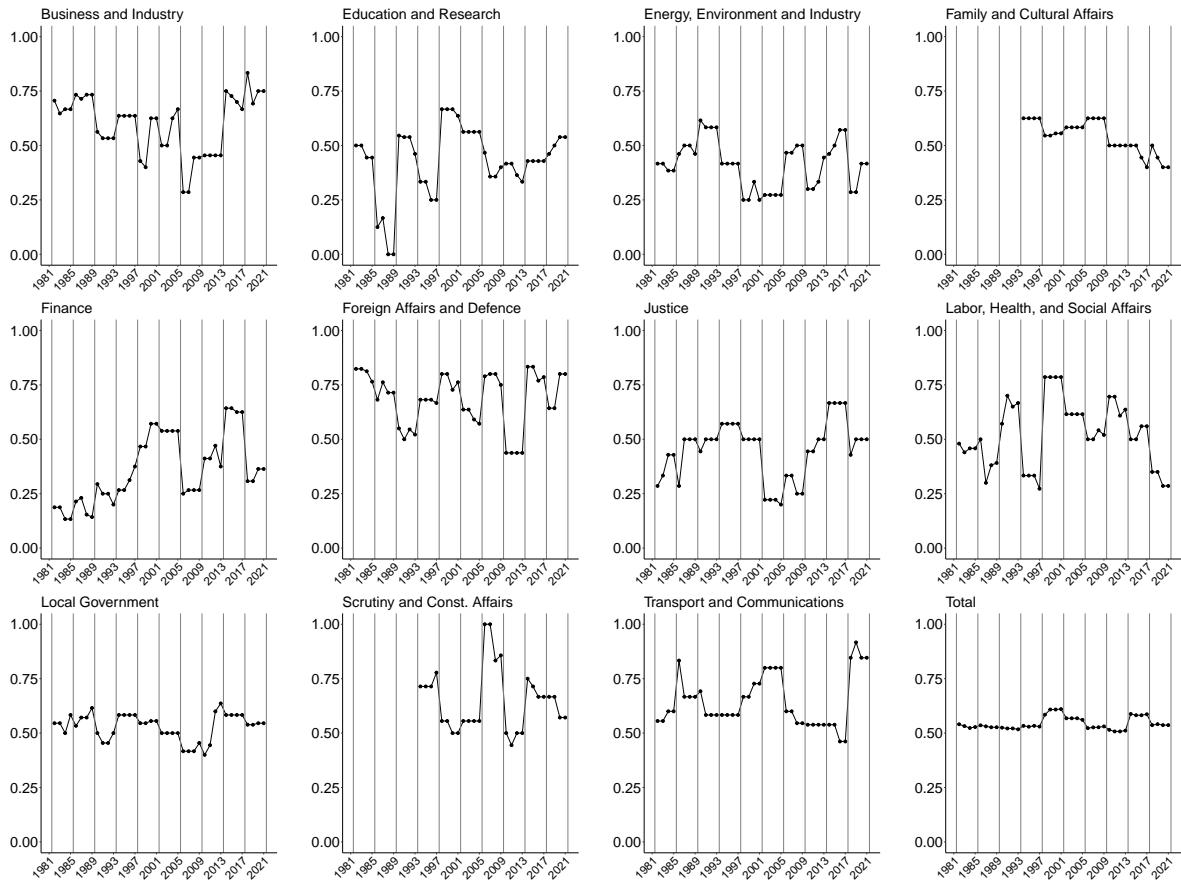
Note: This figure displays divergence of legislative speech for four dimensions (given in the sub-panel headings) in the period 1982-2021, controlling for the legislator's committee assignment. The models are estimated separately for each background characteristic. Speeches from legislators who do not sit in one of the 12 main Parliamentary committees are disregarded. The black points correspond to the average divergence of speech in each session and the black vertical lines indicate elections. The gray shaded area represents the average divergence in hypothetical data in which each speaker's identity is randomly assigned. We construct 100 hypothetical data sets and compute the average in each session. The dashed line corresponds to the mean for each session across the distribution of placebo estimates. The upper and lower bounds of the light gray shaded area correspond to the 5th and the 95th highest scores across the placebo distributions. The dark gray shaded area represents the corresponding 10th and 90th highest scores.

Figure A.8: Fraction of female politicians by policy area and parliamentary session



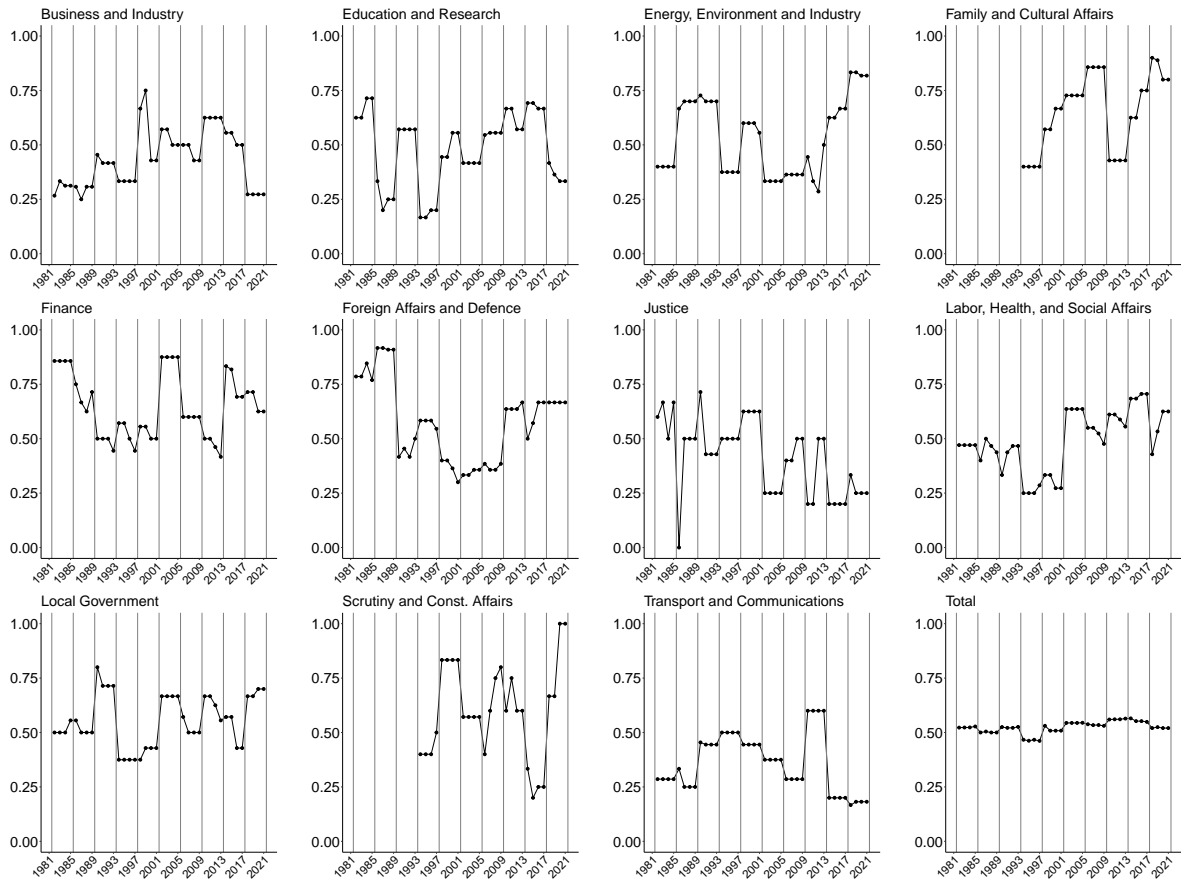
Note: The figures show the fraction of female politicians, by policy area and parliamentary session. The black vertical lines indicate elections. If a politician switches committees during a parliamentary session, we use the committee where he/she spent most days. For a detailed description of which committees are contained in each policy area, see Table A.2. The committee compositions reflect the compositions in our sample (e.g., excluding Nynorsk speakers) and hence deviates somewhat from the general committee composition in the Norwegian parliament.

Figure A.9: Fraction of old politicians by policy area and parliamentary session



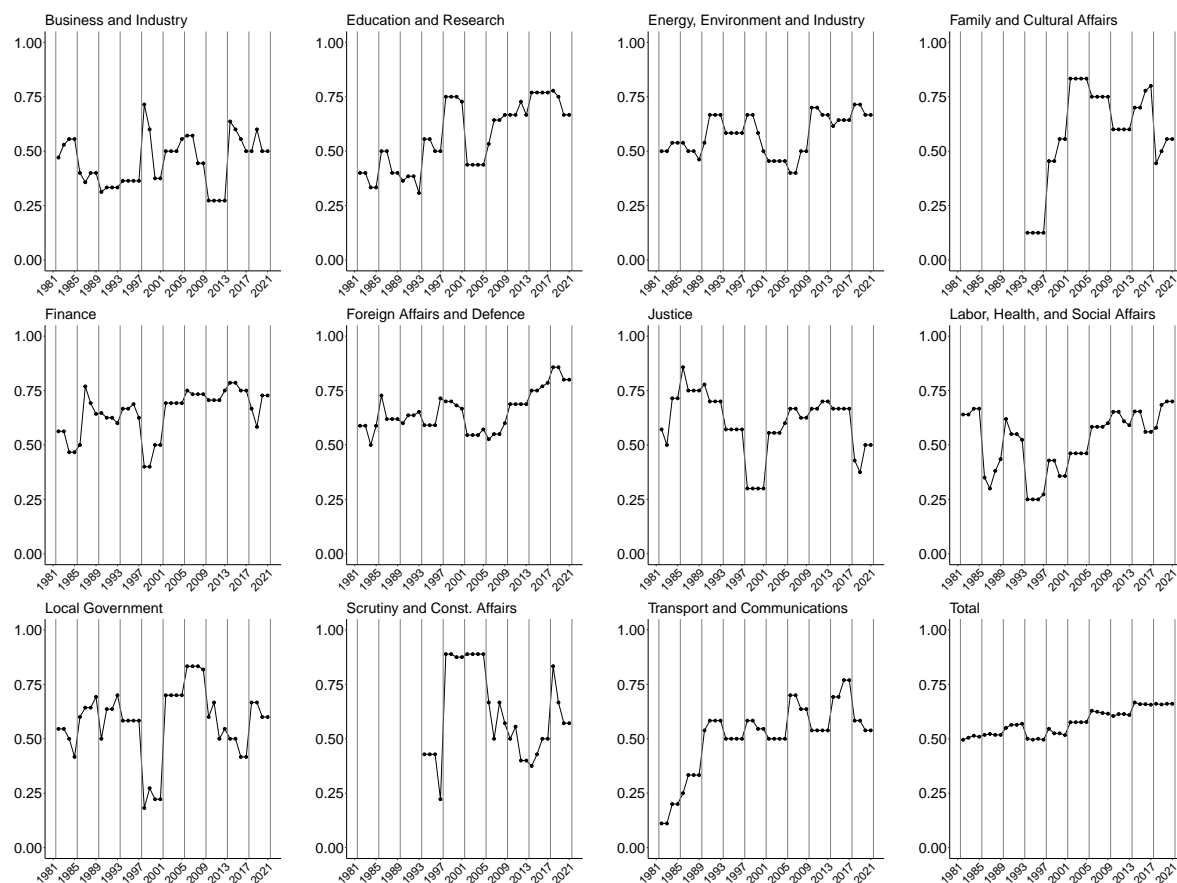
Note: The figures show the fraction of politicians above 47 years old, by policy area and parliamentary session. The black vertical lines indicate elections. If a politician switches committees during a parliamentary session, we use the committee where he/she spent most days. For a detailed description of which committees are contained in each policy area, see Table A.2. The committee compositions reflect the compositions in our sample (e.g., excluding Nynorsk speakers) and hence deviates somewhat from the general committee composition in the Norwegian parliament.

Figure A.10: Fraction of urban politicians by policy area and parliamentary session



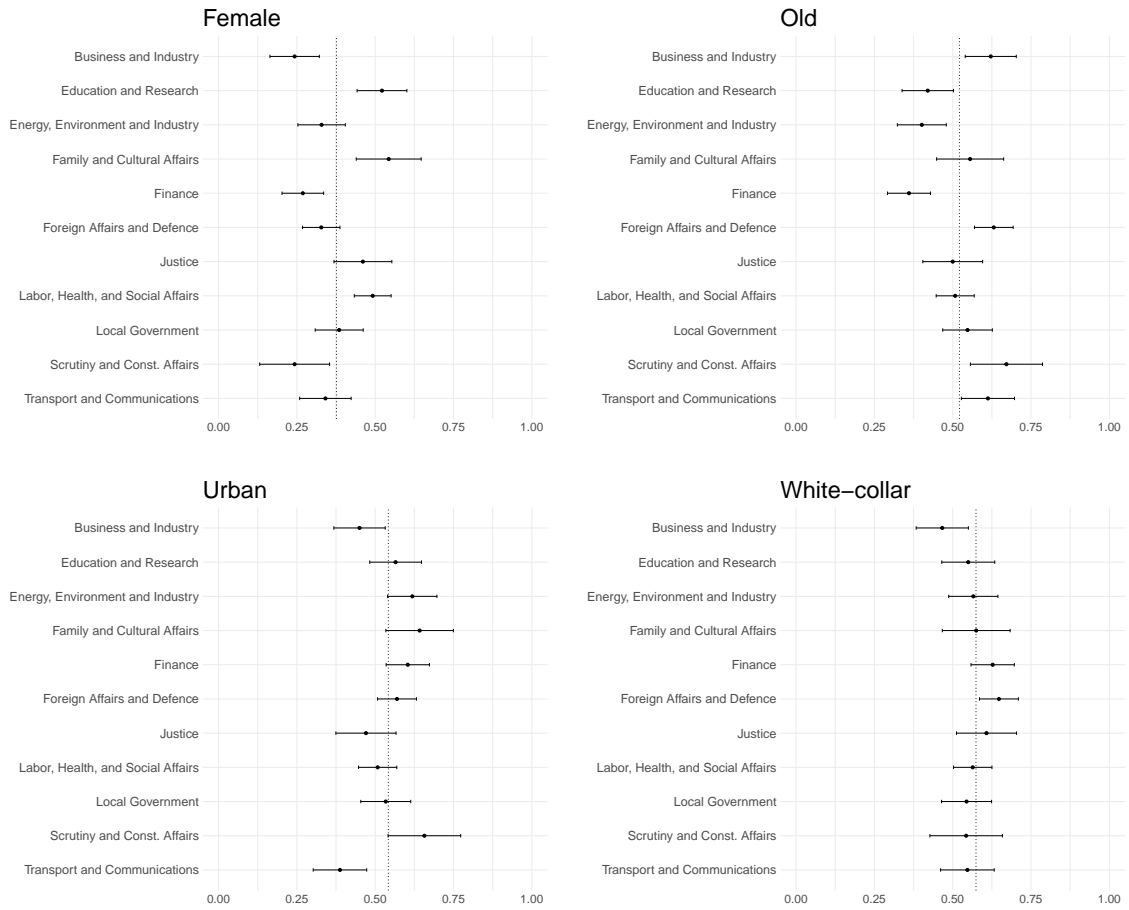
Note: The figures show the fraction of politicians living in a municipality with a city status, by policy area and parliamentary session. The black vertical lines indicate elections. If a politician switches committees during a parliamentary session, we use the committee where he/she spent most days. For a detailed description of which committees are contained in each policy area, see Table A.2. The committee compositions reflect the compositions in our sample (e.g., excluding Nymorsk speakers) and hence deviates somewhat from the general committee composition in the Norwegian parliament.

Figure A.11: Fraction of politicians with white-collar background by policy area and parliamentary session



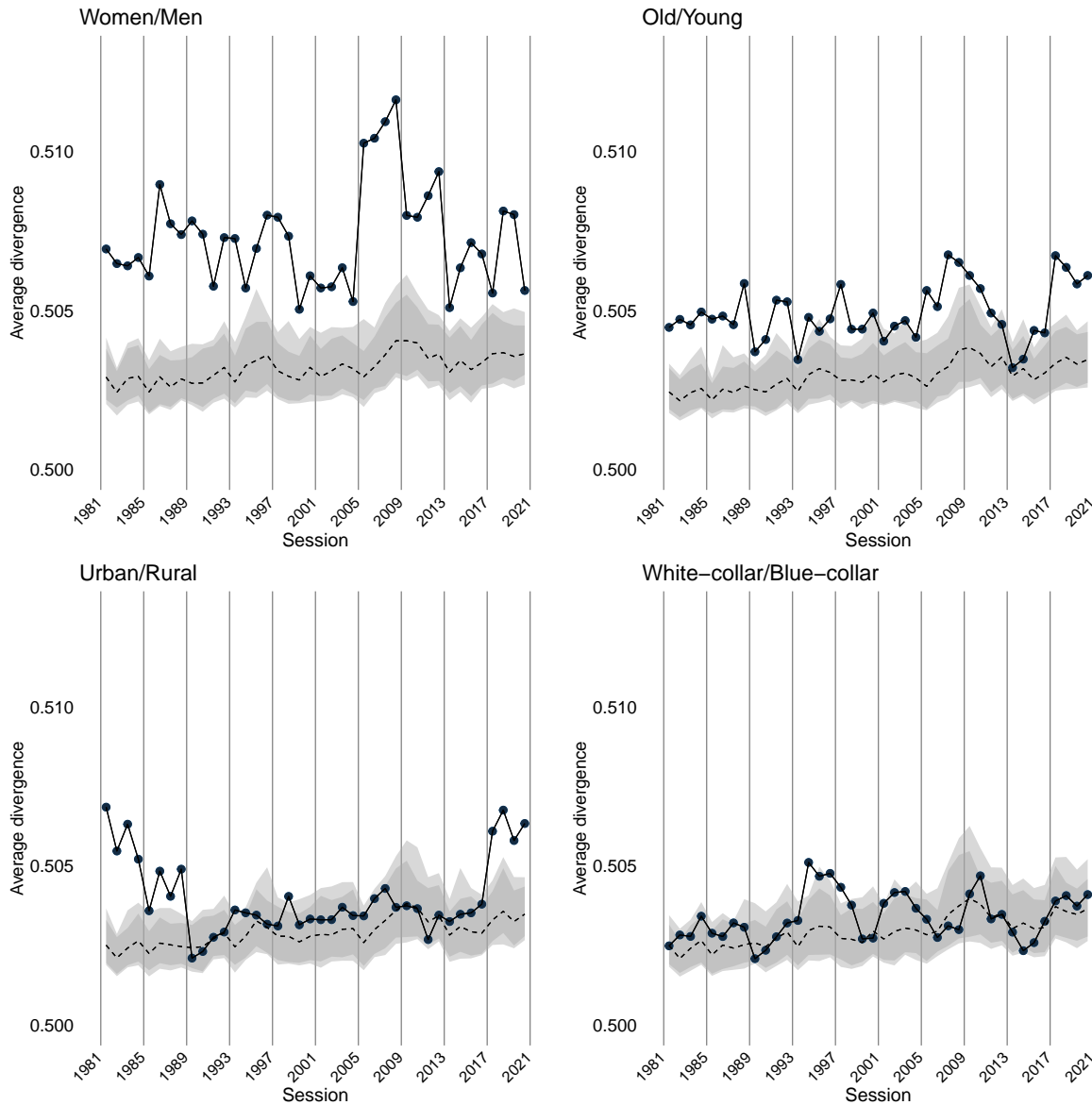
Note: The figures show the fraction of politicians whose fathers worked in white-collar jobs, by policy area and parliamentary session. The black vertical lines indicate elections. If a politician switches committees during a parliamentary session, we use the committee where he/she spent most days. For a detailed description of which committees are contained in each policy area, see Table A.2. The committee compositions reflect the compositions in our sample (e.g., excluding Nynorsk speakers) and hence deviates somewhat from the general committee composition in the Norwegian parliament.

Figure A.12: Committee representation



Note: This figure displays the representation of MPs by committee and background characteristic. The vertical bars represent the mean fraction of females, old, urban, and white-collar MPs in Parliament from 1981 – 2021. Values that exceed the vertical bar suggest that MPs with the given characteristic is overrepresented in the committee relative to their representation in Parliament. Values that are lower than the bar suggest that MPs with the group characteristic are underrepresented. The horizontal lines represent 95 percent confidence intervals. If a politician switches committees during a parliamentary session, we use the committee where he/she spent most days. For a detailed description of which committees are contained in each policy area, see Table A.2.

Figure A.13: Analysis of background divergence when controlling for party affiliation, but not committee



Note: This figure displays divergence in legislative speech for four dimensions (given in the sub-panel headings) in the period 1982-2021. The models are estimated separately for each background characteristic. The black points correspond to the average divergence of speech in each session and the black vertical lines indicate elections. The gray shaded area represents the average divergence in hypothetical data in which each speaker's identity is randomly assigned. We construct 100 hypothetical data sets and compute the average in each session. The dashed line corresponds to the mean for each session across the distribution of placebo estimates. The upper and lower bounds of the light gray shaded area correspond to the 5th and the 95th highest scores across the placebo distributions. The dark gray shaded area represents the corresponding 10th and 90th highest scores.

Table A.1: Norway's governments 1981–2021

| Time period | Prime minister | Parties | Parl. basis | Appointment reason | Resignation reason |
|---------------------|-------------------------|----------------|-------------|-----------------------|-----------------------|
| Oct 1981 – Jun 1983 | Kåre Willoch (H) | H | Minority | General elections | - |
| Jun 1983 – Sep 1985 | Kåre Willoch (H) | H, KrF, Sp | Majority | Government expansion | - |
| Sep 1985 – May 1986 | Kåre Willoch (H) | H, KrF, Sp | Minority | - | Government crisis |
| May 1986 – Oct 1989 | Gro H. Brundtland (A) | A | Minority | Government crisis | General elections |
| Oct 1989 – Nov 1990 | Jan P. Syse (H) | H, KrF, Sp | Minority | General elections | Government crisis |
| Nov 1990 – Oct 1996 | Gro H. Brundtland (A) | A | Minority | Government crisis | Change prime minister |
| Oct 1996 – Oct 1997 | Thorbjørn Jagland (A) | A | Minority | Change prime minister | General elections |
| Oct 1997 – Mar 2000 | Kjell M. Bondevik (KrF) | KrF, Sp, V | Minority | General elections | Government crisis |
| Mar 2000 – Oct 2001 | Jens Stoltenberg (A) | A | Minority | Government crisis | General elections |
| Oct 2001 – Oct 2005 | Kjell M. Bondevik (KrF) | KrF, H, V | Minority | General elections | General elections |
| Oct 2005 – Oct 2013 | Jens Stoltenberg (A) | A, SV, Sp | Majority | General elections | General elections |
| Oct 2013 – Jan 2018 | Erna Solberg (H) | H, FrP | Minority | General elections | - |
| Jan 2018 – Jan 2019 | Erna Solberg (H) | H, FrP, V | Minority | Government expansion | - |
| Jan 2019 – Jan 2020 | Erna Solberg (H) | H, FrP, V, KrF | Majority | Government expansion | - |
| Jan 2020 – Oct 2021 | Erna Solberg (H) | H, V, KrF | Minority | Government reduction | - |

Note: The parties are Socialist Left Party (SV), the Labour Party (A), the Centre Party (Sp), the Christian Democrats (KrF), the Liberal Party (V), the Conservative Party (H), and the Progress Party (FrP). Source www.regjeringen.no.

Table A.2: Committee structure in the Norwegian Parliament, 1981–2021

| Policy area | English committee name | Norwegian committee name | Time period |
|-------------------------------------|-------------------------------------|--|-------------|
| Business and Primary Industry | Agriculture | Landbrukskomiteen | 1981–1993 |
| Business and Primary Industry | Maritime and Fisheries | Sjøfart og fiskerikomiteen | 1981–1993 |
| Business and Primary Industry | Business | Næringskomiteen | 1993–2021 |
| Education and Research | Church and Education | Kirke- og undervisningskomiteen | 1981–1993 |
| Education and Research | Church, Education and Research | Kirke-, utdannings- og forskningskom. | 1993–2017 |
| Education and Research | Education and Research | Utdannings- og forskningskomiteen | 2017–2021 |
| Energy, Environment and Industry | Energy and Industry | Energi- og industrikomiteen | 1981–1993 |
| Energy, Environment and Industry | Energy and the Environment | Energi- og miljøkomiteen | 1993–2021 |
| Family and Cultural Affairs | Family, Culture and Admin. | Familie-, kultur- og administrasjonskom. | 1993–2005 |
| Family and Cultural Affairs | Family and Culture | Familie- og kulturkomiteen | 2005–2021 |
| Finance | Finance | Finanskomiteen | 1981–2021 |
| Foreign Affairs and Defence | Foreign Affairs and Constitution | Utenriks- og konstitusjonskomiteen | 1981–1993 |
| Foreign Affairs and Defence | Defence | Forsvarskomiteen | 1981–2009 |
| Foreign Affairs and Defence | Foreign Affairs | Utenrikskomiteen | 1993–2009 |
| Foreign Affairs and Defence | Foreign Affairs and Defence | Utenriks- og forsvarskomiteen | 2009–2021 |
| Justice | Justice | Justiskomiteen | 1981–2021 |
| Labor, Health, and Social Affairs | Social Affairs | Sosialkomiteen | 1981–2005 |
| Labor, Health, and Social Affairs | Consumer and Administration | Forbruker- og administrasjonskomiteen | 1981–1993 |
| Labor, Health, and Social Affairs | Labor and Social Affairs | Arbeids- og sosialkomiteen | 2005–2021 |
| Labor, Health, and Social Affairs | Health and Social Affairs | Helse- og omsorgskomiteen | 2005–2021 |
| Local Government | Local Govern. and Environ. Protect. | Kommunal- og miljøvernkomiteen | 1981–1993 |
| Local Government | Local Government | Kommunalkomiteen | 1993–2005 |
| Local Government | Local Govern. and Administration | Kommunal- og forvaltningskomiteen | 2005–2021 |
| Scrutiny and Constitutional Affairs | Scrutiny and Constitutional Affairs | Kontroll- og konstitusjonskomiteen | 1993–2021 |
| Transport and Communications | Transport | Samferdselskomiteen | 1981–2005 |
| Transport and Communications | Transport and Communications | Transport- og kommunikasjonskomiteen | 2005–2021 |

Note: This table shows the committee structure in the Norwegian Parliament in our sample period (1981–2021). Our baseline analysis includes committee fixed effects. The aggregation of committees to policy areas is used for descriptive analyses only (e.g., Figure A.2). Data source: www.regjeringen.no.

Table A.3: Summary statistics by speaker-session

| <i>Variable</i> | Median | Mean | St.Dev. | Min | Max | N |
|--------------------|---------------|-------------|----------------|------------|------------|----------|
| Number of words | 3585 | 4348.23 | 3139.54 | 113 | 32579 | 4896 |
| Number of speeches | 35 | 45.24 | 36.81 | 1 | 327 | 4896 |
| Words per speech | 100.84 | 108.2 | 38.31 | 26.9 | 397 | 4896 |
| Minutes of speech | 69.32 | 86.42 | 64.53 | 1 | 547.27 | 2194 |
| Words per minute | 37.14 | 37.1 | 5.09 | 15.31 | 54.04 | 2194 |

Note: This table shows the median, mean, standard deviation, minimum, and maximum for the variables listed in Column 1. The last column shows the number of speaker-session observations. There are 4,896 speaker-session observations, and 610 unique speakers in the period 1982–2021. Speech length is calculated using data after and including the 2004 session. We calculate speech duration by subtracting the current speech’s start time from the next one’s start time. This method will exaggerate the duration of the last speech of debates. There, we set any speech time calculated above 2 hours as missing, as it’s likely to be an exaggerated figure rather than an accurate representation of the speaker’s time. The summary statistics for minutes of speech and words per minute are based on 2,194 speaker-sessions, and 336 unique speakers. All variables are based on the sample after data pre-processing (see section 3.1 for details.)

Table A.4: Correlation matrix for politicians' background characteristics

| | Right-wing | Female | Old | Urban | White-collar |
|--------------|------------|--------|--------|-------|--------------|
| Right-wing | 1 | | | | |
| Female | -0.084 | 1 | | | |
| Old | 0.032 | -0.005 | 1 | | |
| Urban | 0.037 | 0.01 | -0.075 | 1 | |
| White-collar | 0.139 | 0.027 | -0.106 | 0.182 | 1 |

Note: This table shows the pairwise correlations for the dummy variables describing politicians' background. The correlation matrix is based on 4,896 MP-session observations.

Table A.5: Summary statistics by speaker characteristics.

| Variable | Median | Mean | St.Dev. | Min | Max | N | Median | Mean | St.Dev. | Min | Max | N |
|--------------------|--------|--------|---------|-------|---------|--------|--------|--------|---------|-------|---------|--------|
| <i>Bloc</i> | Right | | | | | | Left | | | | | |
| Number of words | 3898.0 | 4625.0 | 3190.7 | 160.0 | 32579.0 | 2619 | 3212.0 | 4029.9 | 3049.4 | 113.0 | 24808.0 | 2277 |
| Number of speeches | 39.0 | 47.7 | 35.5 | 1.0 | 327.0 | 2619 | 31.0 | 42.4 | 38.0 | 1.0 | 326.0 | 2277 |
| Words per speech | 100.6 | 107.1 | 37.1 | 26.9 | 397.0 | 2619 | 101.3 | 109.5 | 39.6 | 28.2 | 394.5 | 2277 |
| Minutes of speech | 71.5 | 86.6 | 61.3 | 1.0 | 547.3 | 1199 | 68.4 | 86.2 | 68.2 | 4.2 | 532.4 | 995 |
| Words per minute | 36.9 | 37.0 | 5.3 | 15.3 | 54.0 | 1199 | 37.5 | 37.2 | 4.8 | 21.2 | 51.4 | 995 |
| <i>Gender</i> | Women | | | | | | Men | | | | | |
| Number of words | 3230.0 | 3948.0 | 2929.8 | 134.0 | 25984.0 | 1811 | 3770.0 | 4583.2 | 3233.9 | 113.0 | 32579.0 | 3085 |
| Number of speeches | 30.0 | 40.5 | 35.8 | 1.0 | 326.0 | 1811 | 38.0 | 48.0 | 37.1 | 1.0 | 327.0 | 3085 |
| Words per speech | 103.7 | 111.6 | 39.8 | 28.3 | 394.5 | 1811 | 99.4 | 106.2 | 37.3 | 26.9 | 397.0 | 3085 |
| Minutes of speech | 65.5 | 83.0 | 66.1 | 3.1 | 547.3 | 865 | 72.5 | 88.6 | 63.4 | 1.0 | 427.9 | 1329 |
| Words per minute | 38.2 | 38.1 | 4.9 | 17.4 | 53.0 | 865 | 36.3 | 36.5 | 5.1 | 15.3 | 54.0 | 1329 |
| <i>Age</i> | Old | | | | | | Young | | | | | |
| Number of words | 3421.0 | 4129.7 | 3002.9 | 113.0 | 24808.0 | 2666.0 | 3813.0 | 4609.5 | 3277.1 | 134.0 | 32579.0 | 2230 |
| Number of speeches | 32.0 | 42.7 | 35.9 | 1.0 | 326.0 | 2666.0 | 39.0 | 48.3 | 37.7 | 1.0 | 327.0 | 2230 |
| Words per speech | 101.4 | 109.2 | 39.7 | 26.9 | 397.0 | 2666.0 | 100.1 | 107.0 | 36.5 | 34.0 | 316.2 | 2230 |
| Minutes of speech | 72.5 | 88.5 | 66.7 | 1.7 | 532.4 | 1199.0 | 66.9 | 83.9 | 61.8 | 1.0 | 547.3 | 995 |
| Words per minute | 36.3 | 36.6 | 5.3 | 17.4 | 54.0 | 1199.0 | 37.9 | 37.7 | 4.8 | 15.3 | 53.7 | 995 |
| <i>Urbanicity</i> | Urban | | | | | | Rural | | | | | |
| Number of words | 3616.0 | 4469.9 | 3378.1 | 113.0 | 32579.0 | 2572 | 3549.0 | 4213.6 | 2847.1 | 136.0 | 24808.0 | 2324 |
| Number of speeches | 36.0 | 46.3 | 39.1 | 1.0 | 327.0 | 2572 | 35.0 | 44.0 | 34.0 | 1.0 | 247.0 | 2324 |
| Words per speech | 100.9 | 109.0 | 40.2 | 28.2 | 397.0 | 2572 | 100.8 | 107.3 | 36.1 | 26.9 | 349.3 | 2324 |
| Minutes of speech | 68.2 | 86.2 | 66.3 | 1.7 | 547.3 | 1190 | 70.8 | 86.7 | 62.4 | 1.0 | 427.9 | 1004 |
| Words per minute | 37.3 | 37.3 | 5.2 | 15.3 | 54.0 | 1190 | 37.0 | 36.9 | 5.0 | 17.4 | 53.5 | 1004 |
| <i>Occupation</i> | White | | | | | | Blue | | | | | |
| Number of words | 3707.0 | 4498.8 | 3278.3 | 113.0 | 32579.0 | 2791.0 | 3395.0 | 4149.1 | 2949.2 | 134.0 | 25984.0 | 2053.0 |
| Number of speeches | 37.0 | 46.8 | 36.8 | 1.0 | 327.0 | 2791.0 | 32.0 | 43.0 | 36.9 | 1.0 | 326.0 | 2053.0 |
| Words per speech | 99.7 | 107.2 | 38.3 | 28.2 | 397.0 | 2791.0 | 103.6 | 110.2 | 38.3 | 26.9 | 349.3 | 2053.0 |
| Minutes of speech | 66.9 | 81.7 | 55.1 | 1.0 | 383.6 | 1333.0 | 72.5 | 94.2 | 77.9 | 4.3 | 547.3 | 809.0 |
| Words per minute | 37.5 | 37.3 | 5.3 | 15.3 | 54.0 | 1333.0 | 36.4 | 36.7 | 4.7 | 23.5 | 53.5 | 809.0 |

Note: This table shows summary statistics across overlapping characteristics. It shows the median, mean, standard deviation, minimum, and maximum for the variables listed in Column 1. There are 4,896 speaker-session observations, and 610 unique speakers in the period 1982–2021. Speech length is calculated using data after and including the 2004 session. The summary statistics for minutes of speech and words per minute are based on 2,194 speaker-sessions, and 336 unique speakers. All variables are based on the sample after data pre-processing (see section 3.1 for details.)

Table A.6: Most divergent words across blocs

| Rank | Most divergent words for right | | | | Most divergent words for left | | | |
|------|--------------------------------|-------------|--------|-------|-------------------------------|----------------|--------|-------|
| | Norwegian | English | #Right | #Left | Norwegian | English | #Right | #Left |
| 1 | privat | private | 57 | 42 | folk | people | 68 | 101 |
| 2 | bedrift | company | 58 | 46 | kvinne | woman | 23 | 38 |
| 3 | enkelt | simple | 64 | 49 | borgerlig | bourgeois | 7 | 23 |
| 4 | pasient | patient | 30 | 17 | kutte | cut | 15 | 29 |
| 5 | stat | state | 83 | 73 | politikk | policy | 70 | 82 |
| 6 | fremme | promote | 59 | 52 | kommune | municipality | 118 | 143 |
| 7 | norsk | Norwegian | 203 | 188 | samfunn | society | 57 | 67 |
| 8 | nummer | number | 50 | 39 | land | country | 167 | 176 |
| 9 | god | good | 245 | 242 | forskjell | difference | 23 | 34 |
| 10 | medføre | entail | 20 | 12 | bra | good | 31 | 38 |
| 11 | avgift | fee | 23 | 16 | situasjon | situation | 79 | 89 |
| 12 | sen | late | 41 | 33 | sosial | social | 20 | 30 |
| 13 | tillate | allow | 21 | 14 | arbeid | work | 105 | 118 |
| 14 | konkurransen | competition | 20 | 13 | jobb | job | 38 | 49 |
| 15 | bedre | better | 90 | 87 | rik | rich | 8 | 15 |
| 16 | redusere | reduce | 61 | 54 | internasjonal | international | 49 | 51 |
| 17 | prosjekt | project | 42 | 36 | ungdom | youth | 14 | 22 |
| 18 | gjærne | glad | 40 | 32 | ansatt | employee | 23 | 33 |
| 19 | skatt | treasure | 20 | 16 | penge | money | 62 | 69 |
| 20 | familie | family | 22 | 19 | utdanning | education | 28 | 33 |
| 21 | person | person | 24 | 19 | fylke | county | 23 | 35 |
| 22 | fall | fall | 58 | 50 | skattelette | tax relief | 3 | 8 |
| 23 | innen | within | 28 | 23 | Trøndelag | Trøndelag | 13 | 20 |
| 24 | håpe | hope | 38 | 33 | politisk | political | 74 | 79 |
| 25 | benytte | use | 20 | 14 | industri | industry | 23 | 31 |
| 26 | fremlegge | present | 9 | 5 | krig | war | 8 | 13 |
| 27 | utfordring | challenge | 54 | 48 | understreke | emphasize | 45 | 49 |
| 28 | skatte | tax | 7 | 3 | tiltak | measures | 79 | 88 |
| 29 | skape | create | 64 | 62 | kutt | cut | 5 | 10 |
| 30 | vekst | growth | 27 | 23 | satsing | initiative | 25 | 29 |
| 31 | øke | increase | 150 | 150 | rettferdig | fair | 6 | 11 |
| 32 | nevne | mention | 52 | 46 | forsvar | defense | 33 | 34 |
| 33 | løsning | solution | 47 | 42 | nødt | have to | 12 | 17 |
| 34 | kostnad | cost | 21 | 17 | militær | military | 10 | 13 |
| 35 | foreligge | exist | 21 | 17 | åra | years | 1 | 5 |
| 36 | forskjellig | different | 24 | 19 | skatteutt | tax cut | 1 | 6 |
| 37 | uttrykk | expression | 31 | 26 | fin | fine | 8 | 13 |
| 38 | nettopp | precisely | 54 | 48 | sjølsagt | naturally | 2 | 7 |
| 39 | system | system | 33 | 27 | privatisering | privatization | 3 | 7 |
| 40 | politiker | politician | 23 | 17 | slag | champion | 18 | 23 |
| 41 | frivillig | voluntary | 16 | 12 | unge | young | 4 | 9 |
| 42 | klar | clear | 122 | 116 | Nordland | Nordland | 4 | 9 |
| 43 | selsvagt | of course | 25 | 20 | mann | man | 11 | 14 |
| 44 | borger | citizen | 6 | 2 | ulik | separate | 49 | 54 |
| 45 | skattebetaler | taxpayer | 5 | 2 | alvorlig | serious | 29 | 34 |
| 46 | forelder | parent | 20 | 16 | syn | vision | 43 | 45 |
| 47 | bilist | motorist | 5 | 2 | barnehage | kindergarten | 17 | 22 |
| 48 | næring | industry | 35 | 29 | arbeidsledighet | unemployment | 10 | 16 |
| 49 | sektor | sector | 30 | 28 | fordeling | distribution | 9 | 13 |
| 50 | basere | base | 20 | 16 | atomvåpen | nuclear weapon | 3 | 7 |

Note: This table shows the 50 most divergent words by bloc from a model which includes parliamentary committee and parliamentary session fixed effects. For each word, we also report the number of occurrences per 100,000 words in the raw data (before feature selection and without covariate adjustment). The word counts are normalized by the number of words spoken by the each group.

Table A.7: Most divergent words across gender

| Rank | Most divergent words for women | | | | Most divergent words for men | | | |
|------|--------------------------------|----------------|--------|------|------------------------------|------------------------|--------|------|
| | Norwegian | English | #Women | #Men | Norwegian | English | #Women | #Men |
| 1 | barn | children | 142 | 47 | norsk | Norwegian | 165 | 213 |
| 2 | kvinne | woman | 61 | 14 | forhold | relationship | 106 | 126 |
| 3 | arbeid | work | 131 | 100 | politisk | political | 66 | 81 |
| 4 | tiltak | measures | 101 | 74 | ligge | lie | 75 | 86 |
| 5 | skole | school | 101 | 53 | politikk | policy | 67 | 80 |
| 6 | øke | increase | 161 | 144 | fall | fall | 45 | 59 |
| 7 | forelder | parent | 31 | 12 | altså | that is | 81 | 97 |
| 8 | familie | family | 31 | 15 | uttrykk | expression | 20 | 33 |
| 9 | ung | young | 35 | 15 | innlegg | post | 51 | 61 |
| 10 | utdanning | education | 47 | 21 | direktør | director | 11 | 30 |
| 11 | barnevern | child welfare | 22 | 6 | sammenheng | context | 40 | 51 |
| 12 | barnehage | kindergarten | 33 | 12 | oppfatning | belief | 13 | 25 |
| 13 | kommune | municipality | 163 | 112 | selvfølgelig | of course | 28 | 38 |
| 14 | folk | people | 95 | 76 | grunnlag | foundation | 37 | 48 |
| 15 | kompetanse | competence | 39 | 21 | Norge | Norway | 170 | 180 |
| 16 | tilbud | offer | 52 | 31 | klar | clear | 109 | 125 |
| 17 | oppta | admit | 52 | 40 | interessant | interesting | 19 | 28 |
| 18 | jobb | job | 54 | 37 | selvsagt | of course | 17 | 26 |
| 19 | lide | suffer | 1 | 1 | standpunkt | position | 10 | 18 |
| 20 | ungdom | youth | 25 | 14 | poeng | point | 11 | 17 |
| 21 | mann | man | 19 | 9 | situasjon | situation | 75 | 87 |
| 22 | rettighet | right | 25 | 14 | utgangspunkt | starting point | 25 | 30 |
| 23 | gruppe | group | 38 | 28 | peke | point | 28 | 33 |
| 24 | vold | violence | 16 | 6 | problem | problem | 106 | 109 |
| 25 | psykisk | mental | 20 | 8 | konklusjon | conclusion | 9 | 15 |
| 26 | fremdeles | still | 12 | 7 | milliard | billion | 27 | 39 |
| 27 | opplæring | training | 18 | 8 | fornuftig | reasonable | 10 | 18 |
| 28 | jente | girl | 9 | 3 | registrere | register | 20 | 27 |
| 29 | likestilling | equality | 12 | 4 | punkt | point | 20 | 26 |
| 30 | unge | young | 12 | 3 | realitet | reality | 12 | 20 |
| 31 | lærer | teacher | 33 | 14 | foreligge | exist | 14 | 22 |
| 32 | forebygge | prevent | 17 | 9 | forsøke | attempt | 13 | 19 |
| 33 | jobbe | work | 24 | 14 | inne | inside | 28 | 33 |
| 34 | glad | glad | 50 | 39 | kr | Norwegian kroner | 116 | 131 |
| 35 | omsorg | care | 16 | 8 | forstå | understand | 30 | 36 |
| 36 | ansvar | responsibility | 77 | 61 | rimelig | inexpensive | 13 | 19 |
| 37 | hverdag | everyday | 12 | 7 | kritikk | criticism | 12 | 17 |
| 38 | fengsel | prison | 11 | 6 | dag | day | 274 | 252 |
| 39 | bistand | aid | 17 | 12 | riktig | right | 50 | 59 |
| 40 | helse | health | 37 | 21 | skape | create | 60 | 65 |
| 41 | voksen | adult | 11 | 4 | opplegg | program | 14 | 18 |
| 42 | sykdom | disease | 13 | 6 | særdeles | very | 4 | 7 |
| 43 | overgrep | abuse | 10 | 5 | anledning | opportunity | 21 | 27 |
| 44 | pasient | patient | 33 | 20 | enig | agree | 67 | 69 |
| 45 | mor | tender | 7 | 2 | ut | out | 179 | 183 |
| 46 | handel | trade | 47 | 33 | diskutere | discuss | 25 | 26 |
| 47 | kontantstøtte | cash support | 9 | 4 | NRK | Norwegian Broadcasting | 8 | 9 |
| 48 | land | country | 175 | 169 | prinsipiell | principle | 10 | 15 |
| 49 | minister | minister | 9 | 5 | tydelig | clear | 26 | 26 |
| 50 | menneskerettighet | human right | 13 | 8 | diskusjon | discussion | 16 | 18 |

Note: This table shows the 50 most divergent words by gender from our baseline model which includes political party, parliamentary committee, and parliamentary session fixed effects. For each word, we also report the number of occurrences per 100,000 words in the raw data (before feature selection and without covariate adjustment). The word counts are normalized by the number of words spoken by the each group.

Table A.8: Most divergent words across age groups

| Most divergent words for old | | | | | Most divergent words for young | | | |
|------------------------------|-------------------|----------------------|------|--------|--------------------------------|--------------------|------|--------|
| Rank | Norwegian | English | #Old | #Young | Norwegian | English | #Old | #Young |
| 1 | nordisk | Nordic | 26 | 15 | dag | day | 247 | 272 |
| 2 | samarbeid | cooperation | 54 | 39 | politikk | policy | 63 | 88 |
| 3 | nevne | mention | 56 | 43 | velge | select | 42 | 52 |
| 4 | kr | Norwegian kroner | 128 | 123 | Norge | Norway | 167 | 186 |
| 5 | mill | million | 74 | 62 | folk | people | 75 | 90 |
| 6 | understreke | emphasize | 53 | 41 | handel | trade | 34 | 42 |
| 7 | svært | very | 63 | 56 | sal | chamber | 30 | 38 |
| 8 | kommune | municipality | 136 | 122 | forhold | relationship | 117 | 122 |
| 9 | direktør | director | 24 | 24 | jobb | job | 39 | 46 |
| 10 | nummer | number | 50 | 40 | sørge | ensure | 29 | 39 |
| 11 | pasient | patient | 30 | 18 | knytte | tie | 36 | 39 |
| 12 | stat | state | 81 | 77 | barn | children | 76 | 81 |
| 13 | område | area | 103 | 98 | barnehage | kindergarten | 15 | 23 |
| 14 | bedrift | company | 52 | 54 | norsk | Norwegian | 190 | 204 |
| 15 | land | country | 175 | 167 | skole | school | 64 | 74 |
| 16 | utvikle | develop | 32 | 26 | politisk | political | 73 | 79 |
| 17 | internasjonal | international | 52 | 48 | klare | clear | 14 | 20 |
| 18 | prosjekt | project | 43 | 34 | klar | clear | 117 | 122 |
| 19 | uttrykk | expression | 31 | 26 | utfordring | challenge | 50 | 53 |
| 20 | Østfold | Østfold | 7 | 4 | konkret | specific | 27 | 33 |
| 21 | virksomhet | business | 34 | 29 | kutte | cut | 17 | 25 |
| 22 | forsvar | defense | 44 | 22 | type | type | 23 | 28 |
| 23 | greie | straightforward | 9 | 7 | oppleve | experience | 34 | 39 |
| 24 | fylkeskommune | regional government | 23 | 17 | bidra | contribute | 64 | 68 |
| 25 | arbeidsplass | workplace | 39 | 39 | penge | money | 60 | 70 |
| 26 | kirke | church | 15 | 12 | diskutere | discuss | 23 | 29 |
| 27 | norden | Nordics | 7 | 4 | stille | exhibit | 38 | 43 |
| 28 | behandling | treatment | 65 | 54 | opplegg | program | 14 | 20 |
| 29 | forbindelse | connection | 52 | 47 | politi | police | 22 | 26 |
| 30 | europaråd | The European Council | 5 | 2 | jobbe | work | 15 | 20 |
| 31 | grunnlag | foundation | 47 | 41 | lys | light | 25 | 28 |
| 32 | felle | trap | 32 | 28 | eneste | only | 23 | 31 |
| 33 | peke | point | 34 | 28 | sentrumsparti | center party | 5 | 5 |
| 34 | regional | regional | 15 | 10 | samferdselsminister | Transport Minister | 8 | 8 |
| 35 | sjøfolk | seamen | 5 | 3 | borgerlig | bourgeois | 11 | 18 |
| 36 | ordning | scheme | 50 | 46 | unge | young | 5 | 8 |
| 37 | innen | within | 28 | 23 | nettopp | precisely | 49 | 54 |
| 38 | god | good | 252 | 234 | satsing | initiative | 26 | 28 |
| 39 | mening | opinion | 28 | 23 | registere | register | 23 | 26 |
| 40 | pst | percent | 82 | 84 | regjeringsparti | government party | 30 | 34 |
| 41 | avgift | fee | 17 | 22 | øke | increase | 144 | 155 |
| 42 | statlig | state | 35 | 31 | poeng | point | 13 | 18 |
| 43 | naturligvis | of course | 6 | 4 | sektor | sector | 28 | 31 |
| 44 | utbygging | development | 22 | 19 | bistand | aid | 13 | 14 |
| 45 | etablere | establish | 24 | 19 | vekst | growth | 22 | 29 |
| 46 | NATO | NATO | 17 | 10 | kontantstøtte | cash support | 5 | 6 |
| 47 | religionspolitikk | religious policy | 1 | 1 | lur | clever | 8 | 12 |
| 48 | kraft | power | 18 | 19 | utslipp | emission | 9 | 15 |
| 49 | skipsfart | shipping | 5 | 3 | bolig | residence | 12 | 16 |
| 50 | fylke | county | 30 | 26 | uke | week | 17 | 21 |

Note: This table shows the 50 most divergent words by age groups from our baseline model which includes political party, parliamentary committee, and parliamentary session fixed effects. For each word, we also report the number of occurrences per 100,000 words in the raw data (before feature selection and without covariate adjustment). The word counts are normalized by the number of words spoken by the each group.

Table A.9: Most divergent words across urbanicity

| Rank | Most divergent words for urban | | | | Most divergent words for rural | | | |
|------|--------------------------------|------------------------|--------|--------|--------------------------------|----------------------|--------|--------|
| | Norwegian | English | #Urban | #Rural | Norwegian | English | #Urban | #Rural |
| 1 | norsk | Norwegian | 205 | 187 | kommune | municipality | 118 | 142 |
| 2 | altså | that is | 98 | 84 | god | good | 235 | 254 |
| 3 | by | city | 19 | 16 | landbruk | agriculture | 12 | 25 |
| 4 | internasjonal | international | 54 | 45 | forhold | relationship | 114 | 125 |
| 5 | ulik | separate | 53 | 49 | distrikt | district | 18 | 30 |
| 6 | land | country | 176 | 166 | forsvar | defense | 30 | 37 |
| 7 | Norge | Norway | 184 | 168 | Finnmark | Finnmark | 8 | 17 |
| 8 | Bergen | Bergen | 10 | 8 | Akershus | Akershus | 3 | 8 |
| 9 | sørge | ensure | 36 | 32 | kr | Norwegian kroner | 118 | 135 |
| 10 | politi | police | 23 | 25 | fylke | county | 22 | 35 |
| 11 | problem | problem | 112 | 104 | Hedmark | Hedmark | 2 | 7 |
| 12 | knytte | tie | 38 | 37 | mill | million | 62 | 76 |
| 13 | politisk | political | 82 | 69 | registre | register | 22 | 28 |
| 14 | fengsel | prison | 9 | 7 | barn | children | 84 | 73 |
| 15 | helseminister | Minister of Health | 10 | 8 | Nordland | Nordland | 4 | 9 |
| 16 | direktør | director | 28 | 19 | jordbruk | farming | 4 | 10 |
| 17 | europaråd | The European Council | 5 | 2 | område | area | 98 | 104 |
| 18 | menneskerettighet | human right | 12 | 6 | skog | forest | 5 | 10 |
| 19 | dreie | turn | 20 | 16 | tilbud | offer | 35 | 41 |
| 20 | Drammen | Drammen | 3 | 2 | næring | industry | 27 | 39 |
| 21 | Vestfold | Vestfold | 4 | 3 | familie | family | 21 | 20 |
| 22 | handel | trade | 40 | 35 | bidra | contribute | 65 | 67 |
| 23 | skip | ship | 6 | 6 | eiendom | property | 6 | 8 |
| 24 | nettopp | precisely | 54 | 49 | samisk | Sami | 6 | 8 |
| 25 | flyktning | refugee | 11 | 8 | ressurs | resource | 33 | 39 |
| 26 | understreke | emphasize | 48 | 46 | landbrukspolitikk | agricultural policy | 4 | 8 |
| 27 | bydel | district | 3 | 1 | oppgave | task | 38 | 41 |
| 28 | tiltak | measures | 82 | 85 | Russland | Russia | 8 | 8 |
| 29 | menneske | human | 49 | 43 | bonde | farmer | 7 | 10 |
| 30 | helsevesen | Health service | 12 | 10 | fylkeskommune | regional government | 17 | 24 |
| 31 | innvandrere | immigrant | 6 | 4 | år | year | 226 | 234 |
| 32 | fattig | poor | 14 | 10 | nordisk | Nordic | 21 | 20 |
| 33 | justisminister | Justice | 13 | 11 | psykisk | mental | 11 | 13 |
| 34 | innsatt | inmate | 3 | 3 | jernbane | railroad | 9 | 18 |
| 35 | asylsøker | asylum seeker | 7 | 5 | sentrumsparti | center party | 5 | 5 |
| 36 | NRK | Norwegian Broadcasting | 10 | 7 | forsvarsminister | Defense Minister | 7 | 7 |
| 37 | sykehus | hospital | 23 | 22 | fisker | fisherman | 4 | 8 |
| 38 | omstilling | changeover | 12 | 12 | skape | create | 63 | 63 |
| 39 | politikk | policy | 79 | 71 | forelder | parent | 19 | 18 |
| 40 | barnevern | child welfare | 14 | 8 | lån | loan | 8 | 10 |
| 41 | Gro Ruddalen | Gro Ruddalen | 2 | 1 | bistand | aid | 14 | 13 |
| 42 | Statoil | Statoil | 11 | 8 | arbeidsplass | workplace | 37 | 41 |
| 43 | industri | industry | 28 | 26 | dag | day | 255 | 264 |
| 44 | telemark | telemark | 3 | 3 | ungdom | youth | 17 | 18 |
| 45 | utgangspunkt | starting point | 30 | 27 | eiendomsskatt | property tax | 3 | 3 |
| 46 | kriminalitet | crime | 11 | 11 | tjeneste | service | 28 | 31 |
| 47 | pensjonist | retire | 8 | 6 | miljøvernminister | Environment Minister | 6 | 8 |
| 48 | kriminell | criminal | 7 | 5 | mat | food | 7 | 11 |
| 49 | trondheim | Trondheim | 7 | 6 | samferdselsminister | Transport Minister | 6 | 10 |
| 50 | sektor | sector | 31 | 28 | landbruksminister | Agriculture Minister | 3 | 6 |

Note: This table shows the 50 most divergent words by urbanicity status from our baseline model which includes political party, parliamentary committee, and parliamentary session fixed effects. For each word, we also report the number of occurrences per 100,000 words in the raw data (before feature selection and without covariate adjustment). The word counts are normalized by the number of words spoken by the each group.

Table A.10: Most divergent words across father's occupation

| Rank | Most divergent words for white-collar | | | | Most divergent words for blue-collar | | | |
|------|---------------------------------------|----------------------|--------|--------|--------------------------------------|----------------------|--------|--------|
| | Norwegian | English | #White | #Other | Norwegian | English | #White | #Other |
| 1 | barn | children | 81 | 74 | forsvar | defense | 33 | 35 |
| 2 | politikk | policy | 81 | 67 | god | good | 238 | 248 |
| 3 | Norge | Norway | 187 | 160 | syn | vision | 39 | 52 |
| 4 | altså | that is | 99 | 82 | pasient | patient | 21 | 28 |
| 5 | politi | police | 26 | 21 | landbruk | agriculture | 14 | 23 |
| 6 | stat | state | 82 | 75 | direktør | director | 23 | 25 |
| 7 | kvinne | woman | 30 | 29 | område | area | 96 | 108 |
| 8 | skole | school | 71 | 63 | distrikt | district | 19 | 29 |
| 9 | barnehage | kindergarten | 19 | 17 | lys | light | 26 | 28 |
| 10 | kommune | municipality | 121 | 139 | forsøke | attempt | 17 | 18 |
| 11 | menneske | human | 47 | 45 | dersom | if | 42 | 47 |
| 12 | knytte | tie | 38 | 37 | naturligvis | of course | 4 | 7 |
| 13 | kontantstøtte | cash support | 6 | 5 | dag | day | 256 | 263 |
| 14 | Bergen | Bergen | 10 | 7 | Hedmark | Hedmark | 3 | 6 |
| 15 | politisk | political | 80 | 70 | jordbruk | agriculture | 5 | 9 |
| 16 | bedrift | company | 55 | 50 | svært | very | 58 | 63 |
| 17 | type | type | 27 | 23 | skog | forest | 5 | 9 |
| 18 | flyktning | refugee | 10 | 8 | næring | industry | 28 | 38 |
| 19 | samisk | Sami | 6 | 7 | bonde | farmer | 7 | 10 |
| 20 | forelder | parent | 19 | 17 | mat | food | 7 | 11 |
| 21 | kirke | church | 14 | 13 | penge | money | 66 | 62 |
| 22 | Akershus | Akershus | 6 | 4 | arbeidsplass | workplace | 37 | 41 |
| 23 | Israel | Israel | 5 | 2 | NATO | NATO | 16 | 11 |
| 24 | peke | point | 31 | 31 | minister | minister | 5 | 7 |
| 25 | gruppe | group | 31 | 31 | postverk | mailing | 2 | 4 |
| 26 | ansvar | responsibility | 65 | 68 | landbrukspolitikk | agricultural policy | 5 | 8 |
| 27 | sektor | sector | 31 | 28 | sammenheng | context | 46 | 50 |
| 28 | asylsøker | asylum seeker | 7 | 5 | sjølsagt | naturally | 3 | 7 |
| 29 | jernbaneverk | railroad | 3 | 2 | innen | within | 25 | 27 |
| 30 | skip | ship | 6 | 6 | NSB | NSB | 7 | 9 |
| 31 | Husbanken | The Housing Bank | 5 | 5 | greie | straightforward | 7 | 9 |
| 32 | gasskraftverk | plant | 7 | 6 | rovdyr | predator | 2 | 4 |
| 33 | gjerne | glad | 40 | 32 | prosjekt | project | 39 | 39 |
| 34 | hvorvidt | whether | 10 | 8 | soldat | soldier | 4 | 4 |
| 35 | milliard | billion | 37 | 33 | Svalbard | Svalbard | 4 | 5 |
| 36 | samferdselsminister | Transport Minister | 8 | 7 | opplegg | program | 15 | 20 |
| 37 | avis | newspaper | 10 | 9 | Nordland | Nordland | 5 | 9 |
| 38 | miljøvernminister | Environment Minister | 7 | 6 | heimevern | home defense | 2 | 3 |
| 39 | privat | private | 53 | 46 | situasjon | situation | 80 | 89 |
| 40 | transportplan | transport plan | 6 | 5 | skogbruk | forestry | 2 | 4 |
| 41 | utviklingsland | developing countries | 6 | 4 | helseminister | Minister of Health | 7 | 11 |
| 42 | sjøfolk | seamen | 4 | 4 | understreke | emphasize | 42 | 54 |
| 43 | statlig | state | 33 | 33 | europaråd | The European Council | 3 | 4 |
| 44 | selsvagt | of course | 25 | 20 | Statoil | Statoil | 10 | 10 |
| 45 | fiskeriminister | Fisheries Minister | 4 | 6 | svaer | answer | 31 | 34 |
| 46 | telenor | telenor | 4 | 4 | nevne | mention | 48 | 53 |
| 47 | barnevern | child welfare | 11 | 13 | afghanistan | Afghanistan | 6 | 5 |
| 48 | arbeidsløshet | unemployment | 3 | 2 | hær | army | 3 | 3 |
| 49 | øsavtalen | the EEA Agreement | 12 | 10 | jobb | job | 41 | 44 |
| 50 | lærer | teacher | 22 | 16 | åra | years | 2 | 5 |

Note: This table shows the 50 most divergent words by occupational background from our baseline model which includes political party, parliamentary committee, and parliamentary session fixed effects. For each word, we also report the number of occurrences per 100,000 words in the raw data (before feature selection and without covariate adjustment). The word counts are normalized by the number of words spoken by the each group.

Table A.11: List of stop words

å adj åh al aldri ale ålein all alle aller allerede allereie allikevel alltid andre angå annan annen annerledes annleis anté åpenbar årsak AS at åtti autentisera autentisere av avstå avtaler både bak bare begynne begynnelse behald behalde beho beholde beklager bemerk berøre berre bestemme bety betydeleg betydelig betydning bind biol bla blant bli blokka boblen bortanfor borte bortenfor bortse bre breitt bruk bruke burda burde byrja byrjar byrjing c ca co d da dårleg dårlig dato de dei deira del den denne der deriv dere deres deretter derfor derfra derfrå deri derimot dermed dertil derved desse dessutan dessuten dessverre det dette din disse diverse dog døme driv du dv e effekt eg egen egentlig ei eie eigen eigentleg ein eit eksempel elle eller en ene enhver enn ennå enno enten etc etter etterpå f få faktisk fare fastsetja fastsette fekk fem femte ferd ferdig ff finne fire fleire flere fò følgde følge følgjande følgje før forbi fordi føre førehand foreslå forestilling forfalle forfell forferdeleg forferdelig forhånd forresten forskning forskning først fortelja fortelle fortsatt fra frå fram framleis framsyning frem fru g gå gammal gammel gang ganske gi gjekk gjeld gjemme gjennom gjer gjera gjø gjorda gjorde gjøre glipp gong gøynde grad grei grunn h ha hå hå ham han handling hans hei heim hel helde helle helst hen henhold hennar henne hennes her heretter herr herved hes hhv hit hitten hjelp hjem ho holde høve hovudsakelig hovudsakleg hun hundre hva hvem hver hvilken hvis hvor hvordan hvoretter hvorfor hvorfra hvori hvorledes hvorpå hvorved i id idet igjen ikke ikkje imidlertid imot indeks informasjon ingen ingenting initiativet inn innanfare innhald inneheld inneholde innenfor innfall innover istedenfor itd j ja jeg jo kanskje kg kjemme kjenne km komma komme kor korlei kort krus kun kunna kunne kunngjera kunngjøre kva kvar kven kvifare l la lag lang legge leggja leite lét ligne like likevel liknande linje lite liten løp løpe m man mang måte måtte me med medan meget meir meir mellom mellomtid men mene mg middel midla million min ml mogleg mohandes mot møter motsetning mulig muligens mye mykje n na nå nær namn når navn ned nedover nei nemleg nemlig neppe neste nesten ni nitti no nødvendig nødvendigvis noe noen nok noko nokon nokre normal notater notere null ny nyleg nylig nytteverdi nyttig o oftast ofte og òg også ok om omtrent ons ønske ønskjer opp openbere opp oppfinnelse oppnå oppstå ord ovanfare ovenfor over overalt på påverkar påvirke per plassere pleie potensiell primær prøve q que ram rask raude reduction ref refs relativ resultat resulterande resultere rød rund s særleg særlig saman same samme sammen samsvar samtidig sann sånn sannsynleg sannsynlegvis sannsynlig sannsynligvis se seg sei seia sein seinare sek seksjon sekund sel selv sende setja si sid sidan side sikker sin sist sistnemnde sistnevnt sjå sjøl sjølv sju skje skjønn skulle slekt slik slut snar snill som spesiell spesifisera spesifisere spørja spørre sprang spring stad stade stand statsråd sted sterk stole stopp stoppe stor sup syne syv t ta tak takk takke tall te tegn tek teke tidlegare tidlig til tilbake tilgjengeleg tilgjengelig tillegg tilstede tilstrekkeleg tilstrekkelig tilsvarande tilsvare tilsynelatande tilsynelatende tips to trengre tro tru tusen tvers tykk tykkjer uansett umiddelbar un under unnta up utan utanfare ute utelate uten utenfor v vær være var vår ved veg vei veit vel veldig velkommen vellykka vellykket vær vera verd verden verdi verkeleg vesentleg vesentlig vi via vid vidare viktig vilja vill ville villig virke virkelig vis vise viss vite vol vore w www x

Note: This table lists the stop words that are excluded in the word selection.

Table A.12: List of procedural words

aksjonskanalen anmodningsvedtak åpning avgjerd avgjerdedyktigheit avstemming avstemning bebude beslutning beslutningsdyktighet bevilgningsreglement bli budsjett budsjettavtale budsjettbehandling budsjettinnstilling daddelvedtak dagsorden debatt delegera delegere demokrati departement direktorat dobbeltstemme dok dokument drøfting drøftingar eøs-avtale etat eu fagkomité finansdebatt finansinnstilling finanstale fleirtal fleirtalsregjering flertall flertallsregjering folkeavstemming folkeavstemning folkestyre folkesuverenitet forberede førebuaende forfatning forhandling forhandlingar forholdstallsval forholdstalsval forklaring forretningsorden forslag forslagsrett fraksjon fraksjonsleder fraksjonsleiar fullmaktslov grunnlov grunnlovsforslag gruppestyre heimfallsrett hemmeleg hemmelig heimfallsrett høgtideleg høring høringsbrev høyring høringsbrev høytidelig i initiativdebatt innberetning innmelding innpiskar innpiske innstilling interpellasjon investitur kabinettsspørsmål koalitionsregjering komite komité kong konstituering konstitusjon konsultasjon kontrassegnatur kritikkvedtak lagting laus led leiar lobbyisme løgsovanmerkningslov lovbehandling lovmerknad lovvedtak løyvingsreglement maktfordeling mandat manntal manntall melding mindretallsregjering mindretallsregjering mistillitsforslag møte nasjonalbudsjett negativ nominasjon norges nou odelsting odelstingsmelding odelstingsproposisjon offentlig offentlig ombudsmann ombudsmann opinion opning opplysningsplikt oppmodningsvedtak opposisjon organisasjonskanal otmeld otrpp på parlamentarisk parlamentarisme parti partigruppe partiprogram plenum prerogativ president presidentsk presidentskap presselosje prop proposisjon protokoll redegjørelse referatsak referendum regjering regjeringskrise regjeringsutnemning regjeringsutnevning replikk representant representantforslag revidere riksrett riksrevisjon sak saksordførar saksordfører sammentred sanksjon sedvane sesjon settepresident skriftleg skriftlig slott sperregrensning spørjetime spørjetimespørsmål spørretime spørretimespørsmål spørsmål statsadministrasjon statsbudsjett statsforvaltning statsminister statsråd stemmeplikt stemmerett stemmeseddel stemmesetel stmeld storting stortingsmelding stortingsperiode stortingsproposisjon stortingsvedtak strpp talarrekkefølgje talerrekkefølge til ting trontale trontaledebatt utbyting utgreingar utjæmningsmandat utjævningsmandat utredning valgliste valgnerlag valgordning valliste valnederlag valordning vararepresentant varsel vedlagd vedlegge vedtak votering

Note: This table lists the procedural words that are excluded in the word selection. The words in the dictionary are obtained from <https://www.stortinget.no/no/Stottemeny/Ordbok/>.

For Online Publication: Appendix B

To illustrate the feature selection process described in section 3.1 we provide two examples. First, consider the following speech by Per Sandberg (Progressive Party) held on April 24, 2013.²⁷ This speech lasted 64 seconds and consisted of 147 words. After pre-processing, we are left with 57 words indicated in boldface in the excerpt below:

*Jeg blir litt **overrasket** når **justisministeren** **påpeker** at man har **reduisert køen**. Ja, det er **enkelt** å **reduere køen** når man **slipper de kriminelle ut** og gir dem strafferabatt, for da blir det **ledig kapasitet** til å **fylle opp** med nye. Det er også **betenkelig** at **justisministeren** **peker** på **økt kapasitet** i nordregionen fordi man **åpner** for **elektronisk soning**. Det er jo ikke en **økt kapasitet** i fengselsinstitusjonene, det er bare en **økt kapasitet** gjennom å **øke antall kriminelle** som er ute på **åpen soning**. **Situasjonen i fengslene** begynner å bli **alvorlig**. Det **pekes** på at man i mye større grad må bruke verneveste og **skjold** i hverdagen. Man **finner** i større grad **ulike våpen**, **hjemmesnekrede våpen**, i **fengslene**. **Situasjonen** er **alvorlig**. Så mitt siste spørsmål til **ministeren** er: Vil **ministeren** nå **garantere** for at det ikke skjer **uheldige episoder** på grunn av **nedbemanning** i **norske fengsler**?*

As explained in section 3.1, we lemmatize all words to allow several versions of a word to be analyzed as one. Here is the Sandberg speech after pre-processing and lemmatization:

overraske justisminister påpeke redusere kø enkel redusere kø slipp kriminell ut ledig kapasitet fyller betenkelig justisminister peke øke kapasitet åpne elektronisk soning øke kapasitet øke kapasitet øke antall kriminell åpen soning situasjon fengsel alvorlig peke skjold hverdag finn ulik våpen våpen fengsel situasjon alvorlig minister minister garantere uheldig episode nedbemanning norsk fengsel

Translated to English:

surprise "minister of justice" "point out" reduce queue simple reduce queue let criminal out free capacity fill questionable "minister of justice" point increase capacity open electronic sentencing increase capacity increase capacity increase number of criminal open sentence situation prison serious point

²⁷A video of this speech is available at <https://www.stortinget.no/no/Hva-skjer-pa-Stortinget/Videoarkiv/Arkiv-TV-sendinger/?mbid=/2013/H264-full/Storting/04/24/Stortinget-20130424-095410.mp4&msid=862&meid=9419>.

shield everyday find different weapon prison situation serious minister minister guarantees unfortunate episode downsizing Norwegian prison

As a second example, we use a 44-second-long speech from Nikolai Astrup (Conservatives) from May 22, 2013.²⁸ This speech consisted of 108 words before pre-processing. After pre-processing, we are left with 20 words, which we again indicate in boldface in the excerpt below:

*Den **vitenskapelige** debatten er viktig, også innenfor **klimaområdet**, og den må **fortsette** – det er mye vi ikke vet. Men **like** viktig som debatten om **vitenskapen** er debatten om hva vi faktisk gjør av **tiltak**. Fremskrittspartiet viste under forhandlingene om **klimaforliket** i Stortinget at de var villige til å være med på langt mer **ambisiøse tiltak** for å **redusere klimagassutslippene** i Norge enn det regjeringen selv hadde lagt frem. Er det litt **pinlig** for SV og regjeringen at Fremskrittspartiet var villig til å gå lenger enn det regjeringen selv var villig til for å **redusere klimagassutslippene** i Norge? Og: Var det derfor regjeringen **valgte** å **utestenge** Fremskrittspartiet fra **klimaforlikene**?*

After pre-processing and lemmatization:

vitenskapelig klimaområde fortsette lik vitenskap tiltak klimaforlik ambisiøs tiltak redusere klimagassutslipp Norge pinlig redusere klimagassutslipp Norge velge utestenge klimaforlik

Translated to English:

scientific “climate area” continue “the same” science measure “climate settlement” ambitious measure reduce “greenhouse gas emissions” Norway embarrassing reduce “greenhouse gas emissions” Norway choose to exclude “climate settlement”

²⁸A video of this speech is available at <https://www.stortinget.no/nn/kva-skjer-pa-stortinget/videoarkiv/Arkiv-TV-sendinger/?mbid=/2013/H264-full/Storting/05/22/Stortinget-20130522-095507.mp4&msid=3313&meid=9433>

For Online Publication: Appendix C

To quantify key issues that separate voters, we rely on ten waves of the *Norwegian National Election Survey* (1981-2017), and estimate linear probability models of the type:

$$D_i = \eta_{t[i]} + \sum_J \beta_j Policy_i^j + u_i \quad (6)$$

where D_i is a bivariate self-reported characteristic of respondent i (e.g., $D_i = 1$ for left-wing voters; $D_i = 0$ for right-wing voters). $Policy_i^j$ is a set of dummy variables capturing whether survey respondent i mentions policy area j when asked to “name one or two issues that were particularly important to your vote”. We categorize voters’ responses into 21 policy areas using keyword searches (see Table C.1).²⁹ The β_j ’s measures the extent to which respondents that mention policy area j are more likely to have a particular characteristic (e.g., self-identify as a left-wing voter) than other respondents, when holding mentions of other policy areas constant. $\eta_{t[i]}$ is a set of survey wave fixed effects ensuring that we compare respondents with different background characteristic that are surveyed in the same year.

The top-left panel of Figure C.1 clearly shows the contrasting priorities of left- and right-wing voters regarding key policy issues. For example, mentioning ‘redistribution’ as a key issue is associated with a 23 percentage point higher likelihood of a respondent supporting a left-wing party, while a focus on ‘business’ concerns corresponds to a 32 percentage point decrease in the likelihood that the respondent supports a left-wing party (all else equal). The overall pattern in the top-left panel of Figure C.1 reflects well-established differences in political priorities: left-wing voters are predominantly concerned with redistribution, employment, and the welfare state, whereas right-wing voters prioritize the private sector, taxation, family issues, abortion, and immigration.

There are also considerable differences across voter background characteristics, al-

²⁹For example, the policy area “employment” includes responses mentioning “employment” and “work.” The top three policy areas are health (13%), elderly care (11%), and education (11%). 24% of participants did not cite specific issues or did not respond, and 14% focused on government formation or other aspects not in our policy categories.

Table C.1: Classification of policy areas based on reported voter preferences: 1981-2017.

| Policy area | Keywords | Keywords in English | Percent of Respondents |
|-----------------|--|--|------------------------|
| abortion | abort | abortion | 3 |
| agriculture | landbruk, jordbruk, fiskeri | agriculture, farming, fishery | 1 |
| business | bedrift, handel, næring | company, trade, business | 2 |
| children | barn | children | 4 |
| defense | militær, forsvar | military, defense | 1 |
| economy | økonomi | economy | 3 |
| education | utdanning, skole/skule, student | education, school, student | 11 |
| elder care | eldre | elderly | 11 |
| employment | arbeid, sysselsetting | work, employment | 7 |
| environment | miljø | environment | 8 |
| EU | eu, eec, eøs | eu, eec, eea | 0 |
| family | familie | family | 4 |
| health care | helse, sykehus/sjukehus | health, hospital | 13 |
| immigration | innvandring, innvandrere, asyl | immigration, immigrants, asylum | 5 |
| Norway | norsk, Norge | Norwegian, Norway | 6 |
| pension | pensjon | pension | 2 |
| redistribution | utjevning, fordeling | equalization, distribution | 2 |
| regional policy | desentralisering, distrikt | decentralization, district | 2 |
| taxation | skatt, avgift, bompenger | tax, fee, toll | 10 |
| transport | samferdsel, kollektiv, jernbane, vei/veg | transport, public transport, railway, road | 2 |
| welfare | velferd, trygd | welfare, social security | 2 |
| other issues | | | 14 |
| no issue | | | 24 |

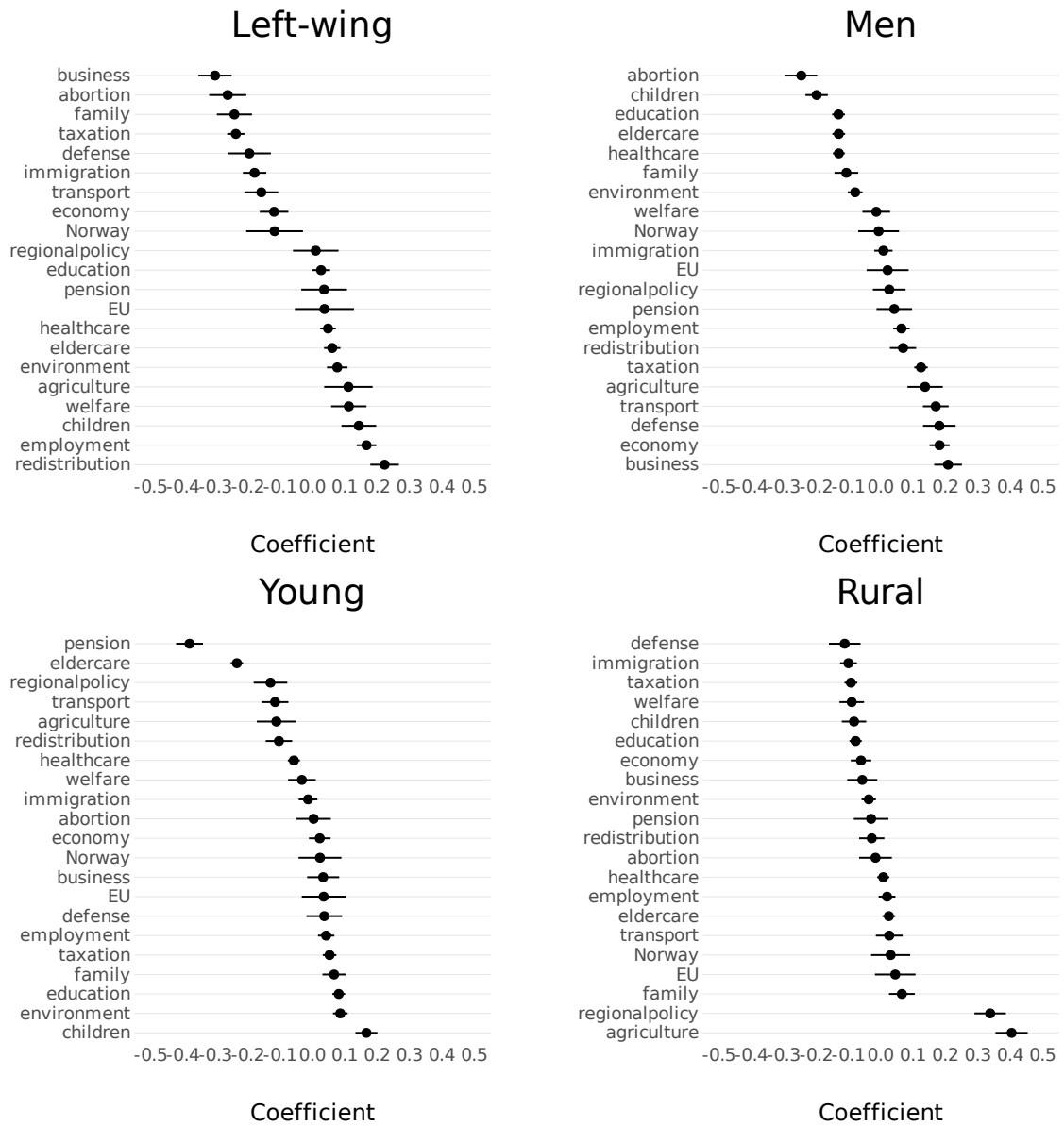
Note: This table categorizes various policy areas as mentioned by survey respondents when asked to highlight one or two significant issues influencing their vote. The data is taken from ten waves of the Norwegian National Election Survey (1981-2017) (N = 20,303). The first column lists the policy areas. The subsequent column provides corresponding keywords used to identify mentions of these policy areas in survey responses. The third column provides the English translation of the keywords. The final column represents the percentage of respondents mentioning each policy area (the sum of which will exceed 100% as respondents can reference multiple areas).

though, naturally, they are less pronounced than for party affiliation.³⁰ Men, relative to women, are concerned with many of the same issues as survey respondents affiliated with right-wing parties. The main exception is abortion, which female respondents are much more likely to mention as one of their two key policy areas. Women seem to care relatively more about welfare policies ('children', 'elder care', 'health care', and 'education').

While the young appear to care particularly about children and education, the old are relatively more concerned about pensions and elder care. Voters residing in rural areas differ markedly from voters residing in urban areas when it comes to regional policy and agriculture, while many of the other differences are more muted. The survey does not include direct questions about class background (or parents' occupation).

³⁰The R^2 in the party affiliation specification is 0.10. For the background characteristics specifications, the R^2 values are 0.08, 0.08, and 0.05.

Figure C.1: Voter preferences measured in surveys: 1981–2017



Note: This figure reports coefficients and corresponding 95% confidence intervals for four linear probability models. The dependent variables are equal to one if the survey respondent is (i) affiliated with one of the left-wing parties (SV, AP, or SP (2005–2017)) ($N = 11,012$), (ii) male ($N = 20,303$), (iii) young (< 47 years) ($N = 20,303$), and (iv) residing in municipality without town status (“rural”) ($N = 19,444$). The independent variables are dummies capturing whether the survey respondent mentions the policy area when asked the following question “Can you name one or two issues which had a particular influence on the way you voted?”. Survey wave fixed effects are included, but not reported. Data from the National Election Survey 1981–2017.

For Online Publication: Appendix D

Please note that in this paper we use data from the Norwegian Election Survey in Appendix C. As is customary, we will submit all programs needed for replication if our paper is accepted for publication. However, we are not authorized to provide the original datasets for confidentiality reasons. We will collaborate with researchers interested in replicating the results in Appendix C by providing them all the necessary information on how to obtain the data, in particular by facilitating their access to the institutions that are the original depositories of the data. For all other analyses, we will submit all data and programs needed for replication if our paper is accepted for publication.