



Kandidatafhandling
i Statskundskab

Elitepolarisering i Danmark

Et maskinlæringsstudie af polariseringen i folketalingstaler og
lovafstemninger

Niels Frederik Lange

Under vejledning af Frederik Hjorth

September 2021

Antal tegn: 167.425



Niels Frederik Lange

Elitepolarisering i Danmark

Kandidatgrad i Statskundskab, september 2021

Vejleder: Frederik Hjort

Antal tegn: 167.425

Københavns Universitet

Det Samfundsvidenskabelige Fakultet

Institut for Statskundskab

Øster Farimagsgade 5

1353 København K



Abstract

Increased elite polarization is often mentioned as a factor with the potential to undermine the social and democratic cohesiveness of Denmark. Recently, political commentators and experts seem to suppose that elite polarization has increased. This is despite the fact that Denmark has long been considered a country with a consensus-driven political culture. However, we do not know much about the state of elite polarization in Denmark today and how it compares to elite polarization in previous periods of Danish politics. In this thesis, I examine how elite polarization in parliamentary speeches has evolved over the past decades by conducting a descriptive case study of parliamentary speeches in the Danish parliament (Folketinget) from 1990 to 2020.

I conceptualize elite polarization in political speech as a dimension of elite polarization that is distinct from polarization in legislative voting and ideological polarization. To measure elite polarization in speeches, I rely on a novel approach that utilizes the performance indicator of a supervised machine learning model to measure differences in political speech. The results show a slight increase in elite polarization of political speech in Denmark from 1990 to 2020 with a varying level of polarization across time. Unlike the elite polarization of legislative voting, the polarization in political speech has not increased more for value-based policies than redistributive policies. I conclude that Denmark has experienced periods of relatively high levels of elite polarization in political speech, but today, the polarization is comparable to the mid-1990s. Finally, the results indicate that the differences between elite polarization in legislative voting and speeches are conceptual and empirical.

The fact that Denmark has experienced periods of increasing elite polarization in political speech raises questions of its potential impact on mass polarization. Additionally, the empirical differences between polarization in political speech and polarization of legislative voting suggest that politicians may appear disagreeable in debates without this affecting their legislative decisions. This can potentially make it difficult for voters to assess which politicians who represent them the best. In sum, the overall findings call for further research that can help understand both the drivers of increased polarization in political speech and its potential impacts on voters.

Indhold

1	Introduktion	8
1.1	Uddybning af problemstillingen og specialets bidrag	10
1.2	Afgrænsning	11
1.3	Specialets opbygning	13
2	Litteraturgennemgang	14
2.1	Konceptualisering og måling af elitepolarisering	14
2.2	Elitepolarisering i forskellige parlamentariske kontekster	16
2.3	Polariseringen mellem danske partier	18
3	Teori og hypoteser	19
3.1	Sprogbrug: En særskilt dimension af elitepolarisering	20
3.2	Den generelle udvikling i elitepolarisering	23
3.3	Elitepolariseringen for værdi- og fordelingspolitiske emner	24
3.4	Forskelle mellem polariseringen i sprogbrug og lovafstemninger	25
4	Metode og data	27
4.1	Undersøgellesdesign: Et deskriptivt casestudie	28
4.2	Data: Folketingstaler og -afstemninger	31
4.3	Tekst-som-data og superviseret maskinlæring	36
4.4	Operationalisering af elitepolarisering i sprogbrug	44
4.5	Operationalisering af elitepolarisering i parlamentariske afstemninger	47
4.6	Validering af målet for elitepolarisering i sprogbrug	49
4.7	Analysestrategi	51
5	Analyse	55
5.1	Den generelle hypotese	56
5.2	Værdi- og fordelingshypotesen	62
5.3	Forskelshypotesen	71
6	Konklusion	73

7	Diskussion	75
7.1	Elitepolarisering i sprogbrug uden for en dansk kontekst	75
7.2	Teoretiske og metodiske implikationer	77
7.3	Empiriske implikationer: Konsekvenser af elitepolarisering	81
	Litteratur	83
	Appendiks	88

Figurer

4.1	Gennemsnitligt antal ord per talestykke over tid	33
4.2	Antallet af talestykker over tid	34
4.3	Antallet af talestykker per parti	35
4.4	Proces for træning og test af superviserede maskinlæringsmodeller	44
4.5	Fordelingen af talestykker på værdi- og fordelingspolitiske emner	53
5.1	Udviklingen i elitepolarisering i sprogbrug mellem blokke	57
5.2	Udviklingen i elitepolarisering i sprogbrug mellem enkelte partier og resten af Folketinget	59
5.3	Udviklingen i elitepolarisering i sprogbrug mellem blokke på tværs af værdi- og fordelingspolitiske emner	63
5.4	Udviklingen i elitepolarisering i sprogbrug mellem enkelte partier og resten af Folketinget på tværs af værdi- og fordelingspolitiske emner	69
5.5	Udviklingen i elitepolarisering af sprogbrug og parlamentariske afstemninger	72
5.6	Korrelation mellem elitepolarisering i sprogbrug og parlamentariske afstemninger	73
7.1	Udviklingen i brugen af "skattestop" og "frit valg"	79
B.1	ROC-kurver	6
B.2	Udviklingen i elitepolarisering i sprogbrug mellem blokke (RFC)	8
B.3	Udviklingen i elitepolarisering i sprogbrug mellem enkelte partier og resten af Folketinget (RFC)	9
B.4	Udviklingen i elitepolarisering i sprogbrug mellem blokke på tværs af værdi- og fordelingspolitiske emner (RFC)	10
B.5	Udviklingen i elitepolarisering i sprogbrug mellem enkelte partier og resten af Folketinget på tværs af værdi- og fordelingspolitiske emner (RFC) . .	11
D.1	Prædikterede sandsynligheder for alle taler	19
D.2	Prædikterede sandsynligheder for velfærdspolitiske taler	20
D.3	Prædikterede sandsynligheder for skattepolitiske taler	21
D.4	Prædikterede sandsynligheder for udlændingepolitiske taler	22
D.5	Prædikterede sandsynligheder for klimapolitiske taler	23
D.6	Korrelation mellem mål for elitepolarisering i sprogbrug og ideologisk elitepolarisering	24

Tabeller

4.1	Variable i målet for polarisering af parlamentariske afstemninger	48
4.2	Valideringsframework	49
4.3	Opsummering af analysestrategi	55
5.1	De mest frekvente ”røde” og ”blå” ord	66
A.1	Oversigt over variable i datasættet	2
B.1	Hyperparametre i grid search	4
C.1	Fordelings- og værdipolitiske ordbøger	13
D.1	De mest ”røde” taler	15
D.2	De mest ”blå” taler	17
E.1	Oversigt over regeringer fra 1990 til i dag	25

1 Introduktion

Selvom uenighed og konkurrence er et grundvilkår for et velfungerende demokrati, så er det også centralt, at politikere er i stand til at indgå kompromisser om samfundets værdimæssige og økonomiske fundament. I denne sammenhæng er Danmark ingen undtagelse. Alligevel synes der at være en bred opfattelse af, at der over de senere år har været en stigende grad af polarisering blandt politikere - med afgørende politisk og social betydning (Dinesen, 2019). Uenigheden mellem politikerne ses blandt andet i politiske diskussioner af regeringens håndtering af COVID-19, hvor oppositionen har anklaget regeringen for at holde Folketinget *ud i strakt arm*. Coronakrisen er en særegen situation, men det er beskyldninger om *blokpolitik* ikke (Kurrild-Klitgaard et al., 2006). Kampen mellem Folketingets partier har eksisteret længe. Spørgsmålet er imidlertid om forskellene og splittelsen er blevet større end tidligere. Tydelige forskelle mellem partier kan hjælpe vælgerne med at stemme på det parti, som bedst repræsenterer deres egne politiske overbevisninger, men det har også nogle konsekvenser. Det kan begrænse samarbejdsmulighederne mellem partierne, skubbe vælgerne mod de ideologiske poler og reducere den politiske tillid (Skytte, 2020).

Udviklingen i elitepolarisering er velbelyst i den amerikanske forskningslitteratur, og der er konsensus om, at der har været en stigende polarisering blandt den politiske elite over de senere år (Abramowitz og Saunders, 2008; Levendusky, 2009; McCarty, 2019). Begrebet elitepolarisering er dog flertydigt, og flere studier peger på, at stigningen i elitepolarisering er sket langs flere dimensioner (Gentzkow et al., 2019). De amerikanske studier har primært fokuseret på den ideologiske polarisering af eliten forstået som forskellene mellem politikernes ideologiske positioner. Den ideologiske polarisering er typisk blevet målt med afsæt i politikernes stemmeafgivning ved lovfastemninger (McCarty, 2019; Poole og Rosenthal, 1985). Nyere studier har dog også undersøgt polariseringen i politikernes sprogbrug via tekst-som-data analyser af politiske taler. Det er på den baggrund blevet vist, at der både er sket en stigning i den ideologiske polarisering og polariseringen af elitens sprogbrug over de senere år i USA (Gentzkow et al., 2019). Polarisationen af politikernes sprogbrug har dog været mere markant, mens den ideologiske polarisering ikke er forskellig fra tidligere perioders polarisering. (Gentzkow et al., 2019).

I litteraturen er den stigende elitepolarisering blevet koblet sammen med en række forskellige politiske konsekvenser for demokratiet. På den ene side kan elitepolarisering

bidrage til, at elitens politiske positioner bliver mere tydelige for vælgerne, hvilket gør det lettere for vælgerne at skelne mellem partierne og dermed holde partierne ansvarlige for deres politik (Hetherington, 2009). På den anden side kan en stigende elitepolarisering også reducere muligheden for samarbejde på tværs af blokke og partier, hvilket potentielt kan lede til situationer med ubeslutsomhed og politisk dødvande. I USA er det endvidere blevet demonstreret, at polariseringen af politikeres ideologiske positioner kan øge polariseringen blandt vælgerne og føre til, at vælgerne synes dårligere om politiske opponenter (affektiv polarisering) (Skytte, 2020). Samtidig er en hårdere og mere respektløs tone også blevet koblet til en stigende grad af affektiv polarisering og øget politisk mistillid. Affektiv polarisering hos vælgerne er derudover blevet kædet sammen med en tendens til forfordeling af partifæller i situationer, hvor partitilhørsforhold ellers synes irrelevant, og mindsket sympati for partimodstandere (Iyengar et al., 2012; Iyengar og Westwood, 2015; Levendusky, 2010). En stigende grad af elitepolarisering har ikke kun en positiv betydning for demokratiet. Det kan også have problematiske effekter, som kan være skadende for sammenhængskraften i et samfund.

Uden for den amerikanske kontekst er elitepolarisering mindre velbeskrevet, og i Danmark er der et fåtal af studier, som afdækker udviklingen i elitepolarisering. Fælles for studierne er dog, at der ikke er indikationer på en stigende grad af elitepolarisering (Christiansen, 2019b). Blokpolitik og snævre flertal er ikke den typiske situation, og samarbejdet i Folketinget sker både mellem de store midterpartier og ved inddragelse af yderpartierne (Kurrild-Klitgaard et al., 2006). Billedet af Danmark som et konsensus land ser derfor fortsat ud til at stå (Lijphart, 1999). Der er imidlertid indikationer på, at der er sket en stigning i polariseringen mellem danske politikere på det værdipolitiske område (Christiansen, 2019b). De danske studier har primært fokuseret på polarisering af afstemninger i Folketinget. Polariseringen af partiernes parlamentariske afstemningsadfærd afspejler dog kun en dimension af de forskelle, som eksisterer mellem politikere. En anden central dimension er som nævnt det politiske sprogbrug, hvor der synes at være konsensus om, at tonen og konkurrencen mellem danske partier er intensiveret over de senere år. Spørgsmålet er imidlertid, om der faktisk er sket en polarisering af elitens sprogbrug i Danmark, og om denne udvikling er anderledes end det, som vi har set for polariseringen af afstemninger. Det er netop dette spørgsmål, som jeg vil beskæftige mig med i mit speciale, hvor jeg undersøger følgende problemstilling:

Hvordan har polariseringen i sprogbrug udviklet sig blandt danske politikere over de seneste årtier, og er udviklingen i polariseringen af politikernes sprogbrug forskellig fra udviklingen i polariseringen af politikernes parlamentariske afstemningsadfærd?

I de følgende afsnit vil jeg først uddybe min problemstilling med fokus på, hvordan den skrives sig ind i den eksisterende litteratur om elitepolarisering. Dernæst beskriver jeg kort mine afgrænsninger i specialet. Endelig gennemgår jeg specialets opbygning, og hvordan de enkelte afsnit bidrager til at besvare min problemstilling.

1.1 Uddybning af problemstillingen og specialets bidrag

Min problemstilling er todelt; det primære fokus er en undersøgelse af den sproglige elitepolarisering i Danmark. Den første del af problemstillingen har til formål at afdække udviklingen i polariseringen af danske politikeres sprogbrug over tid, hvilket endnu ikke er undersøgt i en dansk kontekst. Det er også baggrunden for, at jeg har valgt at holde problemstillingen bred, da det giver mulighed for at belyse forskellige aspekter af et ubeskrevet fænomen. Den anden del af problemstillingen fokuserer på at sammenholde udviklingen i elitepolarisering i sprogbrug med elitepolarisering af afstemningsadfærd. Det bidrager til at give et mere nuanceret billede af den sproglige polarisering, herunder om den adskiller sig fra polariseringen af afstemninger. Endelig har jeg med problemstillingen også et klart deskriptiv fokus. Dette er et bevidst valg, som har forskellige implikationer. Jeg sætter særskilt fokus på denne beslutning i afsnit 4.

Med specialet bidrager jeg til studiet af elitepolarisering på tre forskellige måder. For det første leverer jeg et empirisk bidrag til studiet af elitepolarisering i Danmark. Det skal ses i lyset af, at elitepolarisering er et relativt underbelyst fænomen i dansk politologi (Dinesen, 2019). Ved at afdække den empiriske udvikling i elitepolarisering i sprogbrug og sammenholde udviklingen med polariseringen af parlamentariske afstemninger bidrager jeg med ny viden om elitepolariseringen i Danmark. Det skal dog tilføjes, at der tidligere er foretaget undersøgelser af elitepolarisering i Danmark, men disse studier har primært fokuseret på polariseringen af parlamentariske afstemninger (Christiansen, 2019b; Kurrild-Klitgaard et al., 2006). Ved også at fokusere på elitepolariseringen i sprogbrug tilføjer min undersøgelse altså viden om en dimension af elitepolarisering, der endnu

ikke er afdækket i en dansk kontekst.

For det andet leverer jeg også et metodisk bidrag til litteraturen om elitepolarisering. Det metodiske bidrag skal ses i lyset af, at jeg anvender en maskinlæringsbaseret tilgang til målingen af elitepolarisering i sprogbrug. Den maskinlæringsbaserede tilgang er ikke unik, da den er inspireret af lignende amerikanske og britiske studier (Gentzkow et al., 2019; Goet, 2019; Peterson og Spirling, 2018). Der er dog tale om relativt få studier, hvorfor mit fokus på den danske kontekst bidrager med at validere det maskinlæringsbaserede mål for elitepolarisering uden for en britisk og amerikansk kontekst. Derudover foreslår jeg med specialet også en ny tilgang til at anvende målet i en flerpartikontekst, hvilket uddybes nærmere i afsnit 4.4.

For det tredje bidrager jeg også konceptuelt til studiet af elitepolarisering. Det konceptuelle bidrag er koblet til skellet mellem elitepolarisering i sprogbrug, elitepolarisering af afstemningsadfærd og ideologisk elitepolarisering. Flere af de øvrige studier, som undersøger udviklingen i elitepolarisering med afsæt i politiske taler, argumenterer for, at deres mål indfanger ideologisk elitepolarisering (Goet, 2019; Peterson og Spirling, 2018). Imidlertid observerer Gentzkow et al. (2019) i en undersøgelse af polariseringen i politiske taler, at der er en empirisk forskel på udviklingen i ideologisk og sproglig polarisering. Jeg vil i forlængelse af denne observation argumentere for, at forskellen ikke alene er empirisk, men derimod også konceptuel. Det konceptuelle bidrag uddybes nærmere i afsnit 3.1 og 7.2, hvor jeg udfolder mit argument om, at elitepolarisering i sprogbrug bedst forstås som en særskilt dimension af elitepolarisering.

1.2 Afgrænsning

Polarisering er et bredt begreb, som i en politologisk sammenhæng både favner forskelle mellem vælgere og den politiske elite. Det er derfor relevant at foretage en række afgrænsninger. Først og fremmest fokuserer jeg på elitepolarisering. Det betyder, at jeg afgrænser mig fra at undersøge polariseringen blandt vælgerne, selvom de to fænomener ofte studeres i relation til hinanden (Abramowitz og Saunders, 2008; Hetherington, 2009; Levendusky, 2009). Det skyldes blandt andet, at elitepolarisering ofte er blevet betragtet som en mulig forklaring på polarisering mellem vælgere. Samtidig skal afgrænsningen også ses i lyset af en mere generel afgrænsning fra at undersøge forklaringer på og effekter af elitepolarisering. Jeg har i stedet valgt et deskriptivt fokus, da jeg mener, at det er

centralt først at afdække omfanget af elitepolarisering før, at man undersøger årsagerne og sammenhængene, der ligger bag fænomenet. Jeg vil dog forholde mig til potentielle forklaringer på den observerede udvikling i elitepolarisering som led i diskussion af specialets resultater, da formålet med diskussionen blandt andet er at give et muligt udgangspunkt for fremtidig forskning.

En anden central afgrænsning relaterer sig til min beslutning om at fokusere undersøgelsen på polariseringen blandt folketingspolitikere. Dette fokus betyder, at jeg alene undersøger en relativt lille del af eliten i dansk politik. For eksempel kan der argumenteres for, at eliten også omfatter meningsdannere, politiske eksperter og journalister med flere (McCarty, 2019). Beslutningen om at afgrænse undersøgelsen til folketingspolitikere hviler på en betragtning om, at det netop er folketingspolitikere, der træffer de mest afgørende politiske valg i det danske samfund. Samtidig sikrer dette fokus også en lighed til øvrige studier af elitepolarisering, hvor det netop er politikerne i den lovgivende forsamling, som er i fokus (Fiva et al., 2021; Peterson og Spirling, 2018; Poole og Rosenthal, 1985). Det styrker muligheden for at sammenholde mine resultater med resultaterne i den øvrige litteratur på området.

Endelig afgrænser jeg mig til at undersøge polariseringen i sprogbrug i folketingstaler. Det betyder, at jeg ikke undersøger polariseringen i sprogbrug på sociale medier eller i massemedierne. En bekymring kan i den forbindelse være, at polariseringen i folketingstaler er mindre end polariseringen i sprogbrug på øvrige medier, da der er nogle relativt faste rammer for denne type kommunikation. Derudover er det langt fra alle folketingstaler, som når ud til vælgerne. Polarisationen i folketingstaler har altså potentielt en mindre indvirkning på vælgernes opfattelser sammenlignet med den sproglige polarisering i for eksempel TV-debatter, og derfor har politikerne potentielt også et mindre incitament til at anvende et polariseret sprog. Jeg vil dog argumentere for, at det ikke udgør et problem. Derimod giver det anledning til at betragte folketingstaler som en least-likely case for polarisering i sprogbrug. Viser der sig med andre ord at være et polariseret sprogbrug i folketingstaler er det sandsynligt, at det også gælder folketingspolitikeres øvrige kommunikation.

1.3 Specialets opbygning

Specialet består af syv hovedafsnit. Efter det indledende afsnit følger afsnit 2, hvor jeg gennemgår den eksisterende litteratur om elitepolarisering. Formålet med afsnittet er at give et overblik over den eksisterende litteratur om elitepolarisering. Jeg har særligt fokus på, hvordan tidligere studier har konceptualiseret og målt elitepolarisering - både i en dansk og udenlandsk kontekst. I afsnit 3 definerer jeg undersøgelsens kernebegreb *elitepolarisering i sprogbrug*. Jeg argumenterer i den forbindelse for, at begrebet bedst forstås som en særskilt dimension af elitepolarisering, som er forskellig fra ideologisk elitepolarisering og elitepolarisering i parlamentariske afstemninger. I det samme afsnit præsenterer jeg mine hypoteser, som er udledt med afsæt i den eksisterende litteratur om elitepolarisering.

Afsnit 4 udgør specialets metode- og dataafsnit. Det primære formål med afsnittet er præsentere, hvordan jeg tester specialets hypoteser. Derfor præsenterer jeg specialets forskningsdesign, der består af et deskriptivt casestudie, hvor Danmark betragtes som en least-likely case for elitepolarisering. En central del af afsnittet er også at præsentere den maskinlæringsbaserede tilgang til måling af elitepolarisering sprogbrug, da det er en relativ ny metodisk tilgang inden for feltet. I afsnit 5 gennemgår jeg specialets resultater. Her fokuserer jeg både på den generelle udvikling i elitepolarisering i sprogbrug og udviklingen inden for værdi- og fordelingspolitiske emner. Derudover sammenligner jeg også udviklingen i elitepolarisering af sprogbrug med elitepolarisering af afstemninger.

I afsnit 6 konkluderer jeg på undersøgelsens resultater. Min undersøgelse viser, at der generelt har været en stigende tendens for elitepolariseringen i sprogbrug over de seneste årtier. Polariseringen er højest omkring 2001-2012, mens den fra 2013 falder til et niveau, som er sammenligneligt med start 1990'erne. Samtidig viser resultaterne også, at der ikke er klare forskelle på udviklingen i elitepolarisering af sprogbrug inden for forskellige værdi- og fordelingspolitiske emner. Endelig viser mine resultater, at der er en empirisk forskel på elitepolarisering i sprogbrug og elitepolariseringen af afstemninger. Det understøtter mit argument om, at elitepolarisering i sprogbrug bedst forstås som en særskilt dimension af elitepolarisering. I afsnit 7 diskuterer jeg de bredere implikationer af mine resultater og den eksterne validitet af min undersøgelse. Afsnittet indeholder en række anbefalinger til den fremtidige forskning om elitepolarisering.

2 Litteraturgennemgang

Det primære formål med specialet er at beskrive og forstå den historiske udvikling i polariseringen mellem partierne i Folketinget. Mit fokus er på elitepolariseringen i sprogbrug, som jeg undersøger med udgangspunkt i et relativt nyt maskinlæringsbaseret mål. Et centralt argument i mit speciale er, at polarisering i sprogbrug bedst forstås som sin egen dimension af elitepolarisering. En forudsætning for at forstå dette argument er at forstå, hvordan den eksisterende litteratur har konceptualiseret elitepolarisering, og hvordan det er blevet målt i forskellige parlamentariske kontekster. I det følgende afsnit præsenterer jeg derfor først, hvordan elitepolarisering, som et generelt fænomen, er blevet konceptualiseret og målt. Dernæst ser jeg nærmere på de udfordringer, som litteraturen identificerer i forhold til at måle elitepolarisering i forskellige parlamentariske kontekster. Endelig gennemgår jeg den eksisterende litteratur om elitepolarisering i Danmark. Det giver et grundlag for at forstå, hvordan udviklingen i elitepolarisering hidtil er blevet beskrevet i en dansk kontekst.

2.1 Konceptualisering og måling af elitepolarisering

Begrebet elitepolarisering har sine rødder i Downs (1957) og Sartori (1976) spatiale model for partisystemer, hvor vælgere og partier er fordelt på et ideologisk højre-venstre kontinuum. Inden for rammerne af det ideologiske kontinuum kan vælgere afgive deres stemme til det parti, som bedst muligt afspejler deres egen placering (Downs, 1957). Modellen giver med andre ord et framework til at forstå konkurrencen mellem partier. Med afsæt i den spatiale model forventer Downs (1957), at topartisystemer vil konvergere mod den ideologiske midte, mens flerpartisystemer i større grad vil opleve ideologisk spredning mellem partierne (Downs, 1957). På baggrund af det samme framework skelner Sartori (1976) mellem centripetale og centrifugale kræfter, som henholdsvis trækker partier mod eller væk fra midten og dermed skaber politiske systemer med enten polariseret pluralisme eller moderat pluralisme. Med udgangspunkt i en spatial forståelse af partiers ideologiske placering kan elitepolarisering konceptualiseres som forskellene mellem partiers fordeling på et kontinuum, som enten repræsenterer ideologiske positioner eller positioner på et bestemt politikområde. Selvom dette muligvis er en simplificering af elitepolarisering som fænomen, så udgør det et enkelt udgangspunkt for at forstå polarisering, der også er bredt

anvendt i litteraturen om polarisering (Barber og McCarty, 2015; Dalton, 2008; McCarty et al., 2006; Peterson og Spirling, 2018; Robison og Mullinix, 2016; Theriault, 2008).

En fundamental forudsætning for at forstå og beskrive udviklingen i elitepolarisering er at have et mål, som indfanger essensen af begrebet. I litteraturen er der forskellige tilgange til at måle begrebet. Overordnet set omfatter de mest udbredte metoder afstemningsbaserede mål (også kaldet 'roll-call voting'), spørgeskemaer blandt vælgere eller eksperter og indholdsanalyser af partiprogrammer (Christiansen, 2019b; Dalton, 2008; Gabel og Huber, 2000; Goet, 2019; Hetherington, 2009; Huber og Inglehart, 1995; Peterson og Spirling, 2018; Poole og Rosenthal, 1985). De forskellige mål er forbundet med forskellige fordele og ulemper, som blandt andet relaterer sig til mængden af tilgængelig data og de politiske institutioner i et givent land (Peterson og Spirling, 2018). Det er centralt at overveje, hvordan valget af et specifikt mål er hensigtsmæssigt i forhold til at indfange essensen af elitepolarisering i den politiske kontekst, som en undersøgelse foretages i.

Historisk set har afstemningsbaserede tilgange været meget udbredte i studiet af elitepolarisering. For eksempel er der i den amerikanske litteratur flere studier, som har anvendt afstemningsbaserede mål. Det er på den baggrund veletableret, at der er sket en stigning i elitepolariseringen i USA (Aldrich og Rohde, 1997; Dalton, 2008; Hetherington, 2009; McCarty et al., 2006; Poole og Rosenthal, 1985; Schlesinger, 1985). De amerikanske resultater har primært fokuseret på Kongressen, hvor elitepolariseringen er blevet dokumenteret med en række forskellige målingstilgange (Brewer, 2005; Layman og Carsey, 2002). I de seneste årtier er der blandt andet set en stigning i andelen af afstemninger, hvor størstedelen af et parti stemmer mod størstedelen af det andet parti, samt en stigning i partisamstemmighed, hvor størstedelen af alle politikere stemmer sammen med deres eget parti (Bond og Fleisher, 2000; Brewer, 2005; Layman og Carsey, 2002). Ligeledes er der med afsæt i afstemningsbaserede idealpunktsestimater for politikeres ideologisk placering (DW-Nominate score) blevet målt en større ideologisk afstand mellem partierne samtidig med, at de enkelte politikere i partierne er blevet mere enige (Brewer, 2005; Poole og Rosenthal, 1985). De afstemningsbaserede mål tegner altså et klart billede af, at der sket en stigning i elitepolariseringen i USA over de senere år.

2.2 Elitepolarisering i forskellige parlamentariske kontekster

Størstedelen af den eksisterende litteratur om elitepolarisering er fokuseret på topartisystemer, mens litteraturen om polarisering i flerpartisystemer er mere begrænset. Dette skal ses i lyset af, at litteraturen omkring polarisering har et amerikansk afsæt (Dinesen, 2019). Alligevel er der en række eksempler på studier af elitepolarisering, som er foretaget i parlamentariske kontekster, der er forskellige fra den amerikanske (Christiansen, 2019b; Dalton, 2008; Fiva et al., 2021; Goet, 2019; Green et al., 2020; Merkley et al., 2020; Peterson og Spirling, 2018). En af de centrale udfordringer for denne litteratur har været at konstruere et polariseringsmål, som kan indfange udviklingen i elitepolarisering og samtidig tage højde for de idiosynkratiske dynamikker, som præger forskellige parti- og valg-systemer. Selvom det umiddelbart synes ligetil at adoptere de afstemningsbaserede mål fra USA, så er der en risiko for, at disse mål ignorerer en række vigtige ikke-ideologiske incitamenter. For eksempel kan individuelle politikeres stemmeafgivning være påvirket af strategiske overvejelser om at skjule interne politiske stridigheder i partiet eller den politiske blok (Bräuninger et al., 2016). Det gælder både to- og flerpartisystemer, men udfordringerne er særligt udtalte i flerpartisystemer, hvor den underliggende stemmeadfærd i de lovgivende forsamlinger er særligt påvirkede af ikke-ideologiske incitamenter (Müller og Strøm, 1999).

En central faktor, der påvirker muligheden for at måle elitepolarisering i forskellige parlamentariske kontekster, er valgsystemet. I den sammenhæng er en interessant pointe, at forskellige valgsystemer indebærer forskellige mekanismer, der påvirker den måde, som elitepolarisering kommer til udtryk på. For eksempel peger et britisk studie på, at 'first-past-the-post'-systemet gør de politiske partiers mulighed for at blive genvalgt afhængig af enkelte partimedlemmers omdømme (Proksch og Slapin, 2012). Principielt set kunne det lede en til at tro, at det giver plads til ideologisk forskellighed, men det politiske system i Storbritannien er også præget af en stærk partidisciplin. En række studier peger på, at den ideologiske forskellighed primært afspejler sig i parlamentariske taler, mens partidisciplinen særligt præger parlamentariske afstemninger (Goet, 2019; Kam, 2009; Peterson og Spirling, 2018). I proportionale valgsystemer, som det norske og danske, er der generelt mulighed for at anvende forskellige værktøjer til at disciplinere partimedlemmer (Cirone et al., 2021; Kam, 2009). For eksempel kan partieliter belønne loyale partimedlemmer med fordelagtige placeringer på valglister eller vigtige ordførerskaber, mens

illooyalitet kan straffes med dårlige valglisteplaceringer og ingen ordførerskaber. Alligevel finder Fiva et al. (2021), at individuelle partimedlemmer kan forme politik i particentrede kontekster som den norske. Samlet set understreger litteraturen, at det ikke alene er afgørende at overveje antallet af partier, når man ønsker at måle elitepolarisering, det er også værdifuldt at se på de incitamenter, som karakteriserer bestemte parlamentariske kontekster.

Der er gjort flere forskellige forsøg på at håndtere de udfordringer, som præger målingen af elitepolarisering i parlamentariske kontekster, som adskiller sig fra den amerikanske. Traditionelt set har man anvendt ekspertspørgeskemaer og spørgeskemaer blandt vælgerne til at placere forskellige partier på en ideologisk højre-venstre skala (Christiansen, 2019b; Gabel og Huber, 2000; Kam, 2009; Layman og Carsey, 2002). Sådanne spørgeskemaer er modsat afstemningsbaserede mål ikke påvirkede af politikernes egne strategiske overvejelser. Imidlertid er et problem ved denne tilgang, at der sjældent er konsistent data tilbage i tid eller på tværs af lande, og det begrænser muligheden for at kortlægge udviklingen i polarisering (Peterson og Spirling, 2018). En anden målingstilgang er baseret på det såkaldte Comparative Manifesto Data Project (CMP), hvor forskellige partiers ideologiske position estimeres på baggrund af en indholdskodning af deres valgprogrammer (Christiansen, 2019b; Gabel og Huber, 2000). En fordel ved denne tilgang er, at det er muligt at estimere partiers ideologiske positioner tilbage i tid. Denne tilgang er dog udfordret af, at der i visse tilfælde er store problemer med kvaliteten af data (Hansen, 2008). Selvom spørgeskema- og manifestbaserede tilgange overkommer nogle af de problematikker, som præger afstemningsbaserede mål i flerpartisystemer, så åbner det på samme tid døren op for andre målingsudfordringer.

I nyere studier af elitepolarisering har parlamentariske taler vist sig som et lovende alternativ til de mere traditionelle måder at analysere polarisering på. Det skyldes i særdeleshed, at taler ikke i samme grad er påvirket af en stærk partidisciplin (Peterson og Spirling, 2018). Endvidere har fremkomsten af automatiserede tekst-som-data metoder og adgangen til historiske parlamentstaler også styrket muligheden for at undersøge elitepolarisering tilbage i tid på en effektiv og valid måde (Gentzkow et al., 2019; Goet, 2019; Peterson og Spirling, 2018). Eksempelvis anvender Peterson og Spirling (2018) maskinlæringsmodeller til at måle udviklingen i elitepolarisering i det britiske 'House of Commons' baseret på parlamentstaler for perioden 1935 til 2013. Studiet viser, at

moderne britisk politik er omtrent lige så polariseret, som det var i midten af 60'erne. I et lignende studie af taler fra den amerikanske kongres tegner der sig et billede af, at elitepolariseringen i USA er steget markant i de senere år (Gentzkow et al., 2019). Udviklingstendenserne i det talebaserede polariseringsmål er dog ret forskellige fra et afstemningsbaseret mål, og polariseringen i taler er særligt steget i de senere år (Gentzkow et al., 2019). Umiddelbart leverer det talebaserede elitepolariseringsmål en løsning på de udfordringer, der præger målingen af elitepolarisering uden for USA. Imidlertid rejser afvigelserne mellem polariseringsudviklingen i tale- og afstemningsbaserede mål også et spørgsmål om, hvorvidt de to mål faktisk måler det samme underliggende fænomen, eller om polariseringen i sprogbrug udgør et empirisk og konceptuelt særskilt fænomen. Dette spørgsmål udfoldes nærmere i afsnit 3.1.

2.3 Polariseringen mellem danske partier

Litteraturen om polarisering i Danmark er begrænset, og der er få studier, som har fokuseret specifikt på elitepolarisering. I en undersøgelse af dansk blokpolitik for perioden 1953-2004 viser Kurrild-Klitgaard et al. (2006), at de fire gamle partier i Folketinget har vedtaget størstedelen af alle lovforslag i samstemmighed. Andelen af lovforslag vedtaget ved samstemmighed falder sjældent under 75 procent, og generelt set er Folketinget kendetegnet ved et relativt stort samarbejde på tværs af partier (Damgaard og Svensson, 1989; Kurrild-Klitgaard et al., 2006). I forlængelse heraf finder Christiansen (2019b) i et studie af elitepolariseringen af afstemningsadfærd, at der ikke er tegn på øget polariseringen mellem de politiske partier i Folketinget over de seneste år. Det er tilfældet, selvom der er sket en øget fragmentering af det danske partisystem. Imidlertid ses der i en vis udstrækning en tendens til stigende polarisering for værdipolitiske spørgsmål (Christiansen, 2019b). Til trods for det relativt lave antal af danske elitepolariseringsstudier, så er de overordnede indikationer, at polariseringen mellem danske politikere og partier er marginal.

De eksisterende studier af udviklingen i elitepolarisering i Danmark er baseret på politikeres stemmeadfærd, ekspertspørgeskemaer og CMP (Christiansen, 2019b; Damgaard og Svensson, 1989; Kurrild-Klitgaard et al., 2006). Det er således endnu ikke undersøgt, hvordan elitepolariseringen har udviklet sig, når der tages udgangspunkt i et mål, som er baseret på tekst-som-data-analyse og parlamentariske taler. Som tidligere fremhævet pe-

ger elitepolariseringslitteraturen på, at parlamentstaler er underlagt andre incitamenter og strategiske overvejelser end afstemningsbaserede mål. Det er blevet brugt som et argument for, at talebaserede mål er velegnede til undersøgelser i parlamentariske kontekster, hvor stemmeadfærd kan være stærkt påvirket af faktorer såsom stærk partidisciplin (Peterson og Spirling, 2018). Der lægges dog vægt på, at det talebaserede mål grundlæggende indfanger ideologiske forskelle mellem politikere, og dermed måler ideologisk polarisering (Goet, 2019). Et amerikansk studie fremhæver imidlertid, at et talebaseret mål reflekterer partiforskelle i værdier, mål og innovation i politisk retorik, der er anderledes end forskellene mellem partiernes ideologiske positioner på bestemt lovgivning (Gentzkow et al., 2019). Ved at introducere et nyt talebaseret elitepolariseringsmål sigter jeg altså ikke alene efter at overkomme nogle af de udfordringer, som har præget tidligere anvendte polariseringsmål i den danske elitepolariseringslitteratur. Jeg sigter efter at bidrage med et nyt perspektiv på elitepolarisering i Danmark, der kan vise potentielle forskelle i de værdier, mål og retorisk overvejelser, som kommer til udtryk i folketalingstaler.

3 Teori og hypoteser

I dette afsnit præsenterer jeg de hypoteser, som vil være omdrejningspunktet for min analyse af udviklingen i elitepolarisering i Danmark. Et vigtigt første skridt er i den sammenhæng at klargøre, hvilken forståelse af polarisering, der ligger til grund for min undersøgelse. Derfor vil jeg først definere begrebet elitepolarisering i sprogbrug. Jeg argumenterer for, at polarisering i sprogbrug adskiller sig konceptuelt fra polarisering i parlamentariske afstemninger og ideologisk polarisering. Dernæst præsenterer jeg de tre hypoteser, som primært udledes med afsæt i den eksisterende litteratur om elitepolarisering i Danmark. Den første hypotese vedrører den generelle udvikling i elitepolariseringen i sprogbrug på tværs af politiske emner. Den anden hypotese sigter mod at nuancere den generelle udvikling ved at se på variationen i polarisering inden for forskellige værdi- og fordelingspolitiske emner. Endelig tester jeg med den tredje hypotese, om der er forskel på udviklingen i elitepolarisering af sprogbrug og elitepolarisering i parlamentariske afstemninger.

3.1 Sprogbrug: En særskilt dimension af elitepolarisering

Som begreb er polarisering ofte blevet anvendt bredt til at beskrive splittelse og forskelle mellem sociale, kulturelle og etniske identiteter (McCarty, 2019). På et abstrakt plan kan polarisering således forstås *som en udvikling, hvor forskellen mellem grupper bliver større over tid*. I den samfundsvidenskabelige litteratur er der ofte blevet fokuseret på polarisering i forhold til et relativt begrænset antal af politiske fænomener. Disse fænomener omfatter blandt andet ideologiske orienteringer og politiske præferencer på bestemte politikområder (McCarty, 2019). En konsekvens heraf er, at polarisering ikke altid kan forstås entydigt, da forskellen mellem politikere og vælgere kan udvikle sig langs flere forskellige dimensioner (Skytte, 2020). Dette er ikke uvæsentligt, da forskellige dimensioner af polarisering ikke nødvendigvis udvikler sig ens over tid. For eksempel kan der være sket en stigning i den ideologiske polarisering blandt politikere og vælgere, selvom polariseringen for så vidt angår et bestemt politisk emne er faldet. Eksemplet er abstrakt, og ideologiske positioner er selvfølgelig nært beslægtede med positioner på bestemte politikområder. Ikke desto mindre understreger det vigtigheden af at være bevidst omkring hvilke dimensioner af polarisering, der undersøges.

Jeg fokuserer primært på polariseringen i politisk sprogbrug, som jeg vil argumentere for udgør en særskilt dimension af polarisering. Specialet skriver sig dermed op i mod en række øvrige studier, der måler elitepolarisering med udgangspunkt i politiske taler. Som fremhævet i litteraturgennemgangen betragtes polariseringen i sprogbrug i disse studier ikke som en særskilt dimension af elitepolarisering. Taler bliver derimod set som en datakilde, der kan anvendes til at måle ideologisk elitepolarisering (Goet, 2019; Peterson og Spirling, 2018). Heri ligger en antagelse om, at politisk tale indeholder informationer om ideologiske og politiske præferencer, samt at politisk sprogbrug generelt er informeret af politiske overvejelser (Goet, 2019; Peterson og Spirling, 2018). Disse studier lægger sig således tæt op af øvrige elitepolariseringsstudier, hvor der for eksempel anvendes mål baseret på parlamentariske afstemninger eller lignende mål til at indfange ideologisk elitepolarisering. Der eksisterer en forståelse af, at man måler det samme underliggende fænomen med de forskellige mål (se for eksempel (Barber og McCarty, 2015; Peterson og Spirling, 2018)). Jeg vil dog fastholde, at polariseringen i sprogbrug bedre kan forstås som en særskilt dimension af polarisering.

Politiske taler er underlagt andre incitamenter og begrænsninger end traditionelle

mål for elitepolarisering (Gentzkow et al., 2019). Afstemningsbaserede mål er påvirkede af specifikke parlamentariske forhold såsom antallet af partier, valgsystemet, strategiske overvejelser og kompromisdannelse. Derimod reflekterer parlamentariske taler potentielt forskelle i værdier, mål og retoriske teknikker, som er forskellige fra partiers politiske positioner på bestemte lovgivningsspørgsmål (Gentzkow et al., 2019). I min optik er det derfor centralt at understrege, at forskellen mellem afstemninger og taler ikke alene er afgrænset til, hvordan politikernes adfærd i henholdsvis afstemninger og taler er styret af forskellige incitament. Parlamentariske taler indfanger forskelle i partiers sprogbrug, valg af retoriske virkemidler og kommunikationsstrategier. Dermed indfanger talerne en dimension af polarisering, der for det første ikke kommer til udtryk i afstemninger, og for det andet afspejler mere end blot partimedlemmers ideologiske eller politiske position.

Politiske partier differentierer dem selv gennem deres kommunikation, og politikerne bruger bestemte termer og vendinger til at diskutere de samme emner. Et eksempel fra den amerikanske kontekst ses i debatten om immigration, hvor republikanere refererer til *illegal aliens*, mens demokrater taler om *undocumented workers* (Gentzkow et al., 2019). I en dansk kontekst ses sådanne forskelle i debatten på boligområdet, hvor den tidligere VLAk-regering ofte talte om *ghettoer*. Derimod har den nuværende regering ændret fokus, og VLAk-regeringens ghettoer omtales i stedet som *udsatte boligområder*. Denne beslutning er bevidst, og boligminister Kaare Dybvad fremhæver begrebsændringen som et forsøg på ikke at stemple de mennesker, som bor i de udsatte områder - alligevel fastholder den Socialdemokratiske regering deres støtte til VLAk-regeringens ghettopakke (Ritzau, 2019). Eksemplet viser, at partier er villige til at differentiere deres sprogbrug, selvom de i bund og grund støtter de samme politiske beslutninger og den samme lovgivning, som partier de står i opposition til. Forskelle i politisk sprogbrug eksisterer altså på trods af, at partier faktisk stemmer for den samme lovgivning. Polarisationen mellem partier givet deres politiske sprogbrug kan dermed indfange en dimension af polarisering, der ikke nødvendigvis kommer til udtryk i partiernes lovgivningsadfærd.

Der er udover rent konceptuelle overvejelser også empiriske indikationer på, at elitepolarisering i sprogbrug udgør en særskilt dimension af elitepolarisering i sprogbrug. Det ses blandt andet i Gentzkow et al. (2019) studie af polarisationen i amerikanske kongrestaler, hvor en sammenligning af polarisering målt ved henholdsvis taler og afstemninger over tid viser to forskellige udviklingstendenser. Polarisationen i afstemninger er steget, men

er på et niveau, som svarer til starten af det 20. århundrede. Polariseringen i sprogbrug er derimod steget over hele perioden fra slutningen af 1800-tallet til i dag og er på et niveau, som er markant højere end tidligere (Gentzkow et al., 2019). Studiet viser et sammenfald mellem stigninger i polariseringen af sprogbrug samt indførelsen af C-SPAN's dækning af alle taler i kongressen og det republikanske partis dagsorden vedrørende *contract with America*, som førte til et øget fokus på politisk kommunikation (Gentzkow et al., 2019). Selvom der alene er tale om tidlige sammenfald, hvor der ikke er foretaget statistiske sammenhængstest, så sandsynliggør resultaterne, at polariseringen af sprogbrug også empirisk set relaterer sig til bestemte retoriske og kommunikative udviklinger. Der er altså også empiriske pejlinger på, at politiske taler indfanger andet end partimedlemmers ideologiske standpunkter og parlamentariske afstemningsadfærd.

Samlet set er der flere grunde til at betragte polarisering i sprogbrug som en særskilt dimension af elitepolarisering, der er forskellig fra polariseringen af politikeres afstemningsadfærd og ideologiske positioner. Derfor har jeg også valgt at konceptualisere de to centrale begreber fra min problemformulering forskelligt. Elitepolarisering i sprogbrug defineres *som en udvikling, hvor forskellen mellem politiske partiers sprogbrug bliver større over tid*, mens elitepolarisering i parlamentariske afstemninger defineres *som en udvikling, hvor uenigheden mellem partierne i parlamentariske afstemninger bliver større over tid*. Først og fremmest indebærer disse definitioner en inddeling af folketingsmedlemmer (eliten) i partier, hvilken er det typiske udgangspunkt for studier af elitepolarisering (Dalton, 2008; Peterson og Spirling, 2018; Poole og Rosenthal, 1985). Dernæst indebærer denne definition også en antagelse om, at mindre eller faldende polarisering viser sig ved en udvikling, hvor forskellen mellem politiske partiers sprogbrug eller stemmeafgivning bliver mindre over tid. Endelig kan det tilføjes, at forskellen mellem politiske partiers sprogbrug og afstemningsadfærd kan blive større af både interne og eksterne grunde. Forskellen mellem partier kan således stige, hvis partimedlemmerne i et parti for eksempel begynder at tale mere konsistent og generelt har et ensartet politisk sprogbrug (internt). Samtidigt kan forskellen også stige, hvis partimedlemmerne i forskellige partier begynder tale mere forskelligt fra hinanden, selvom de enkelte partier ikke nødvendigvis er blevet mere konsistente i deres politiske sprogbrug. Det samme gælder afstemninger. Definitionerne, herunder forståelserne af hvad der udgør mere eller mindre polarisering, vil være udgangspunktet for de efterfølgende afsnit, hvor jeg præsenterer mine hypoteser.

3.2 Den generelle udvikling i elitepolarisering

Der eksisterer ikke meget viden om elitepolarisering i Danmark, og det er dermed svært at danne sig klare forventninger om, hvordan polariseringen har udviklet sig. De enkelte fund fra danske studier tyder på, at polariseringen af eliten er begrænset (Christiansen, 2019b; Kurrild-Klitgaard et al., 2006). En af de ting, der er blevet peget på som en forklarende faktor, er den stærke tilbøjelighed til konsensus i dansk politik (Lijphart, 1999). Det synes derfor sandsynligt, at polariseringen af den danske elite, modsat den udvikling vi har set i USA, ikke er steget over de seneste årtier. Alligevel er det værd at fremhæve, at der i en periode har været en styrkelse af fløjpartierne i dansk politik, hvor vi har set partier som Dansk Folkeparti og Enhedslisten vinde frem. Det kunne lede en til at forvente en stigende grad af polarisering i Danmark, men Christiansen (2019b) fremhæver, at yderpartierne har bevæget sig tættere på midten af dansk politik, og det har medført en faldende elitepolarisering over tid. Det er dog centralt at fremhæve, at de tidligere danske studier ikke undersøger polariseringen i sprogbrug, men i stedet fokuserer på polariseringen af parlamentariske afstemninger. Selvom det er nærliggende at forvente en lignende udvikling for polariseringen i sprogbrug, så er der som fremhævet grund til at betragte polariseringen i sprogbrug som en særskilt dimension af elitepolarisering, som følger en særskilt udviklingstendens.

Polariseringen i sprogbrug reflekterer potentielt forskelle i partiers mål, værdier og kommunikation, som er forskellig fra partiernes ideologiske placering såvel som deres adfærd i parlamentariske afstemninger. I USA er den stigende grad af polarisering i sprogbrug over de senere årtier blevet kædet sammen med innovation i politisk tale, overtalelseteknikker og udviklingen i mediedækningen (Gentzkow et al., 2019). Den generelle professionalisering af politik er ikke alene et amerikansk fænomen. I en undersøgelse af udviklingen i det danske partisystem over de seneste 25 år fremhæver Green-Pedersen og Kosiara-Pedersen (2020), at der er sket en professionalisering og centralisering af partiororganisationer, politikformulering og kampagneføring. Det er blandt andet blevet muliggjort af, at partiernes indtægter også er steget markant i samme periode (Green-Pedersen og Kosiara-Pedersen, 2020). Der er således grund til at forvente, at det også vil afspejle sig i udviklingen for elitepolarisering i sprogbrug i Danmark. Derfor forventer jeg på trods af den faldende polarisering af danske partiers parlamentariske afstemningsadfærd, at *elitepolariseringen i sprogbrug er steget igennem de seneste årtier* (den generelle hypotese).

Den generelle hypotese bidrager til at forstå, hvordan elitepolariseringen af sprogbrug i Danmark har udviklet sig i helt generelle termer på tværs af politiske problemstillinger. Derudover tjener testen af hypotesen også et andet formål. Ved at undersøge den generelle polarisering mellem danske partier etableres en baseline, som specialets øvrige analyser kan holdes op mod. Det er særligt væsentligt for undersøgelsen af værdi- og fordelingshypotesen, hvor jeg nuancerer udviklingen i den sproglige polarisering og ser nærmere på tendenserne inden for særskilte politiske emner. Resultaterne fra disse analyser kan holdes op mod baselinetendensen fra den generelle hypotese. Det giver mig mulighed for at se nærmere på hvilke politiske emner, der er særligt polariserende i dansk politik. I det følgende vil jeg præsentere mine konkrete forventninger til hvilke politiske emner, der er særligt polariserende for eliten af dansk politik.

3.3 Elitepolariseringen for værdi- og fordelingspolitiske emner

I Danmark er politisk konkurrence typisk blevet forstået ud fra to forskellige dimensioner - en fordelingspolitisk dimension og en værdipolitisk dimension. Den fordelingspolitiske dimension vedrører centrale politikområder såsom skatte-, indkomst- og velfærdspolitik og har typisk rummet forhold, der viser modsætningerne mellem arbejderklassen og borgerskabet (Andersen og Borre, 2003; Inglehart, 2008). Den værdipolitiske dimension omfatter derimod politikområder såsom klima-, kultur og udlændingepolitik og har siden 1990'erne vist sig som en afgørende dimension i konkurrencen mellem partierne (Andersen og Borre, 2003). Begge dimensioner er typisk blevet forstået spatialt, hvor partier og vælgere er placeret på en venstre-/højreskala (Andersen og Borre, 2003). Det er dog værd at bemærke, at partiers placering på én dimension ikke nødvendigvis medfører en tilsvarende placering på den anden dimension. For eksempel kan Socialdemokratiet placeres til venstre på skalaen for fordelingspolitiske spørgsmål, mens partiet i større grad har rykket sig til højre på visse værdipolitiske områder. Det er derfor forventeligt, at konkurrencen og polariseringen mellem partier vil være forskellig alt efter, om der fokuseres på en fordelings- eller værdipolitisk dimension.

I Christiansen (2019b) studie af elitepolarisering observeres der fra 1998-2015 en stigning i polariseringen inden for værdipolitiske emner, mens der ikke er tegn på polarisering inden for fordelingspolitiske emner. Som jeg tidligere har argumenteret for, er der dog ikke grund til at forvente, at polariseringen i sprogbrug og polariseringen af afstemnings-

adfærd udvikler sig ensartet. Alligevel er det ikke utænkeligt, at visse politiske tendenser afspejler sig i begge dimensioner af elitepolarisering. Her tænker jeg navnligt på, at der gennem de seneste årtier har det været et øget fokus værdipolitiske spørgsmål i dansk politik, mens fordelingspolitiske emner er trådt mere i baggrunden (Bille, 2006, 2014; Green-Pedersen og Kosiara-Pedersen, 2020). Særligt udlændingepolitikken er trådt frem på den politiske scene og har spillet en afgørende rolle siden Dansk Folkepartis oprettelse i 1995. Der peges endda på, at spørgsmålet om indvandring udgør en ny politisk skillelinje i dansk politik (Thomsen, 2017). Værdipolitikken er således ikke kun blevet tildelt mere opmærksomhed, men der er også i stor grad stærke og forskellige holdninger til emnet, som splitter både vælgere og politikere. Jeg forventer derfor, at *elitepolariseringen i sprogbrug er steget mere inden for værdipolitiske emner end fordelingspolitiske emner igennem de seneste årtier* (værdi- og fordelingshypotesen)

Forventningen i værdi- og fordelingshypotesen afspejler den empiriske udvikling i elitepolariseringen af danske partiers afstemningsadfærd. Dette er tilfældet på trods af mit argument om, at der med afstemninger og sprogbrug er tale om to forskellige dimensioner af elitepolarisering, der ikke nødvendigvis følger den samme udvikling. Det skal ses i lyset af, at de to dimensioner af elitepolarisering på trods af deres forskelle forventes at være påvirket af de samme større tendenser i dansk politik, herunder øget fokus på og uenighed om værdipolitiske emner. Imidlertid mener jeg fortsat, at der tale om to forskellige dimensioner af elitepolarisering, der over en længere periode sandsynligvis udvikler sig forskelligt. I det næste afsnit formulerer jeg derfor en konkret hypotese, som fokuserer på forskellen mellem de to dimensioner.

3.4 Forskelle mellem polariseringen i sprogbrug og lovaftemninger

Udover den faktiske udvikling i elitepolarisering i Danmark vil jeg med nærværende speciale også fokusere på, om polariseringen i sprogbrug er forskellig fra polariseringen af parlamentarisk afstemningsadfærd. Som udgangspunkt forventer jeg, at polariseringen har udviklet sig forskelligt for de to dimensioner af elitepolarisering. Det skyldes blandt andet, at parlamentariske taler er styret af andre incitamenter og dynamikker end afstemninger. Derudover skal det også ses i lyset af det amerikanske studie, der har vist empiriske forskelle i udviklingen for et talebaseret polariseringsmål og et afstemningsbaseret mål

(Gentzkow et al., 2019). Men ligesom der er forskelle på udviklingen i polarisering i USA sammenlignet med Danmark, så er der formentlig også forskelle på den måde, som politiske taler anvendes af de folkevalgte. Det er derfor nødvendigt at overveje de konkrete dynamikker, som er på spil i en dansk kontekst, og hvordan de former forventningerne til forskellene mellem polarisering i henholdsvis sprogbrug og parlamentarisk afstemningsadfærd.

Der er særligt to forhold, som i en dansk sammenhæng taler for, at elitepolariseringen målt ved taler er forskellig fra elitepolarisering målt ved afstemninger. For det første er afstemninger i Folketinget præget af en stærk danske partidisciplin (Skjæveland, 2001). Det danske politiske system er bygget op omkring stærke partier, som har evnen til at påvirke medlemmers stemmeadfærd. Det skal blandt andet ses i relation til den indflydelse de enkelte partier har på medlemmers placering på valglister eller opstillingsmuligheder i bestemte kredse, hvor partiet står stærkt. Partierne har potentielt en ret betydelig indflydelse på enkelte medlemmers mulighed for at blive valgt, og det betyder alt andet lige, at partierne har mulighed for at straffe partimedlemmer, der ikke følger partiets linje og stemmer for de forslag, som er vigtige for partiet. Dette gælder i princippet også for parlamentariske taler, men jeg forventer, at tilskyndelsen til at disciplinere partimedlemmers tale er mindre end tilskyndelsen til at disciplinere deres afstemningsadfærd, da vedtagelsen af lovgivning indebærer relativt varige politiske ændringer, hvilket ikke i samme grad gælder for politiske taler.

For det andet er Danmark et konsensusland, hvor der ofte indgås politiske kompromiser i forbindelse med vedtagelsen af lovforslag (Lijphart, 1999). Som fremhævet ses det også ved, at størstedelen af lovforslagene i Folketinget vedtages af de fire gamle partier i enstemmighed (Kurrild-Klitgaard et al., 2006). Denne tendens skal ses i relation til den stærke partidisciplin, der gør det muligt for partierne at påvirke medlemmers stemmeadfærd med henblik på at sikre opbakning til politiske kompromisser. Netop de politiske kompromisser synes centrale, da de ofte indgås forud for vedtagelsen af lovgivning. En konsekvens heraf er, at den politiske konflikt i forbindelse med lovgivningsspørgsmål ikke afspejles i partiernes afstemningsadfærd, idet der før afstemningen er opnået enighed om indholdet af loven (Christiansen, 2019b). Samlet set taler de fremhævede dynamikker vedrørende partidisciplin og kompromisdannelse for, at polariseringen i afstemningsadfærd kan være mindre end den faktiske polarisering, der har været omkring en given

lovafstemning. Derfor forventer jeg, at *elitepolariseringen i sprogbrug er steget mere end elitepolariseringen af parlamentariske afstemninger igennem de seneste årtier* (forskelshypotesen).

Forventningen i forskelshypotesen også ses i sammenhæng med mit argument om, at elitepolarisering i sprogbrug udgør en særskilt dimension af elitepolarisering. Politiske taler reflekterer potentielt mål, værdier og udviklinger inden for politisk retorik, som ikke er fanget i politiske afstemninger. Derfor er det særligt interessant, at der over de seneste 25 år er sket en stigende professionalisering af danske partiers organisationer og kommunikation (Green-Pedersen og Kosiara-Pedersen, 2020). Det virker sandsynligt, at denne professionalisering indebærer et forøget fokus på strategisk kommunikation, som kan understøtte det enkelte partis profil og politiske standpunkter, herunder hvordan partiet adskiller sig fra andre. Samtidig giver parlamentariske taler politikerne mulighed for at fremhæve deres egne politiske synspunkter eller kritisere andres - både over for øvrige partier, men også befolkningen, da talerne bliver fulgt af medier, der viderebringer politikernes synspunkter. Parlamentariske taler giver politikerne mulighed for at fremhæve egne positioner i relation til øvrige partier, så det er tydeligt for vælgerne, hvad de enkelte partier står for. Det gør sig ikke i samme grad gældende for parlamentariske afstemninger, der i høj grad også spejler politisk kompromisdannelse, som formentlig ikke steget med den øgede professionalisering af partiorganisationer og politisk kommunikation. Det taler for, at polariseringen i sprogbrug er steget mere end polariseringen i afstemningsadfærd.

4 Metode og data

I dette afsnit beskriver jeg de metoder og den data, som jeg anvender til at teste mine hypoteser. Først introducerer jeg det undersøgelsesdesign, som jeg har valgt at anvende i besvarelsen af mit forskningsspørgsmål. Dernæst præsenterer jeg logikkerne bag tekst-som-data og superviseret maskinlæring, som udgør mit metodiske udgangspunkt for at måle elitepolarisering i sprogbrug. Herefter operationaliserer jeg de to centrale elitepolariseringsbegreber og præsenterer det framework, som jeg bruger til at validere målet for elitepolarisering i sprogbrug. Det særskilte fokus på valideringen af målet for elitepolarisering i sprogbrug skal ses i lyset af, at der tale om et nyt mål, som er baseret på relativt nye metodiske værktøjer. Endelig præsenterer jeg den analysestrategi, som anvendes til

at teste mine hypoteser.

4.1 Undersøgelsesdesign: Et deskriptivt casestudie

Det centrale formål med specialet er at undersøge udviklingen i elitepolariseringen i sprogbrug i Danmark over de seneste årtier. I problemformuleringen ligger hermed allerede en række implicitte beslutninger om hvilke undersøgelsesdesigns, der er relevante. For det første har specialet et deskriptivt fokus, da jeg er interesseret i at udføre en præcis og systematisk undersøgelse af i elitepolariseringen af sprogbrug over tid (Gerring, 2004). For det andet kan specialet karakteriseres som et casestudie af den sproglige polarisering mellem folketingspolitikere i perioden 1990-2020. Forskningsdesignet kan derfor betegnes som et deskriptivt longitudinelt casestudie, hvor jeg undersøger polariseringen i danske politikeres sprogbrug over tid. I de følgende afsnit uddyber jeg overvejelserne bag mit valg af forskningsdesign.

4.1.1 Et deskriptivt fokus

Når man undersøger en udvikling i et fænomen over tid, er et naturligt spørgsmål, hvorfor den pågældende udvikling har fundet sted. Et sådan spørgsmål har i modsætning til mit forskningsspørgsmål en kausal karakter. Det primære formål med dette speciale er ikke at forklare udviklingen i polariseringen af elitens sprogbrug. Fokus er derimod på at forstå, hvordan polariseringen i sprogbrug har udviklet sig. Det skal ses i lyset af, at der er relativt få beskrivelser af elitepolariseringen i Danmark, hvorfor der fortsat er et behov for at blive klogere på fænomenet i en dansk kontekst. På samme tid introducerer jeg også et nyt mål for elitepolarisering, der fokuserer på polariseringen i politisk sprogbrug, hvilket ikke tidligere er blevet undersøgt i en dansk kontekst. Jeg har på den baggrund vurderet, at det er hensigtsmæssigt at fastholde et deskriptivt fokus i specialet for at understøtte kvaliteten af den viden, der tilvejebringes om elitepolarisering i Danmark

Valget af et deskriptivt design bygger også på et argument om, at det kan være fordelagtigt at afkoble deskriptive studier fra kausale studier, da man undgår, at de beskrivende elementer er farvede af undersøgelsens kausale interesser (Gerring, 2012). Det er særligt relevant for fænomener, der besidder mange potentielle sande kausale forklaringer og effekter, hvilket netop gør sig gældende for elitepolarisering, men også på et mere generelt plan gælder de fleste fænomener i samfundsvidenskaben (Gerring,

2012). Det er dog vigtigt at understrege, at jeg ikke vil være ignorant over for de kausale implikationer af elitepolarisering i sprogbrug, men jeg vil fastholde, at der er en iboende værdi i, at jeg primært har et deskriptivt fokus. Desuden er den deskriptive undersøgelse et vigtigt skridt på vejen til at forstå elitepolarisering i sprogbrug, og det er afgørende for at kunne teste og udvikle forklaringer om fænomenet i fremtidig forskning. (King et al., 1994).

Selvom specialet overordnet set har et deskriptivt fokus, så kan det alligevel karakteriseres som et hypotetisk-deduktivt studie. Det skal ses i lyset af, at de opstillede hypoteser bygger på eksisterende litteratur om elitepolarisering og parlamentariske taler (Hansen og Stubager, 2020). Traditionelt set indebærer deskriptive studier et induktivt fokus, da der typisk ikke formuleres forventninger på forhånd (Hansen og Stubager, 2020). Imidlertid vil jeg fastholde, at det hypotetisk-deduktive fokus er foreneligt med den deskriptive tilgang, da studiet trækker på eksisterende teori og empiri, som anvendes til at danne forventninger om udviklingen i polarisering uden at indicere bestemte kausale sammenhænge. Det betyder, at forskningskriterier såsom repræsentative data og høj målingsvaliditet er særligt centrale, da disse er afgørende for at understøtte troværdigheden af mine hypotesetests. Jeg vil derfor løbende fokusere på disse forskningskriterier i præsentationen af mit data og min metodisk tilgang.

4.1.2 Danske folketingspolitikere som case

Specialet kan karakteriseres som et casestudie af elitepolariseringen blandt danske folketingspolitikere i perioden 1990 til 2020. Der er særligt to forhold, som har været afgørende for valget af den danske case. For det første er der tale om en relativt underbelyst case, hvor tidligere studier primært har fokuseret på elitepolarisering i lovfastemninger i Folketinget (Christiansen, 2019b; Kurrild-Klitgaard et al., 2006). For det andet har den metodiske udvikling i feltet om elitepolarisering og adgangen til digitaliserede data fra Folketinget givet mulighed for at undersøge netop den sproglige polarisering i eliten (Gentzkow et al., 2019; Goet, 2019). Den tidslige afgrænsning af casen er foretaget, da jeg ikke har haft de nødvendige ressourcer til at få adgang til data før 1990.

En begrænsning ved kun at undersøge danske folketingspolitikere er, at det potentielt svækker mulighederne for at generalisere resultaterne til andre lande. Det skal også ses i relation til, at polariseringsdynamikkerne i forskellige lande er påvirkede af den parla-

mentariske kontekst, herunder variationer i valgsystemer og graden af partidisciplin mv. (Goet, 2019). Generaliseringer af mine resultater forudsætter derfor grundige overvejelser om de forhold, der potentielt kunne medvirke til, at andre lande ikke har oplevet den samme udvikling i elitepolarisering i sprogbrug. Danmark er, som fremhævet, karakteriseret ved en stærk grad af konsensus (Lijphart, 1999). Jeg vil derfor argumentere for, at Danmark udgør en least-likely case for at finde en stigning i elitepolarisering. Viser der sig at være sket en stigning i elitepolariseringen af sprogbrug, vil det styrke potentialet for at generalisere resultaterne til lande, som er præget af en mindre grad af konsensus. Hvis der derimod ikke er sket en polarisering af eliten i Danmark, vil mulighederne for at generalisere resultaterne være begrænsede. Jeg vil dog understrege, at der er en iboende værdi i at undersøge den danske case, da det giver et billede af polariseringen uden for en amerikansk kontekst, hvor elitepolarisering som fænomen er underbelyst.

Valget af et longitudinelt casestudie giver mulighed for at gå i dybden med forskellige elementer i elitepolariseringen i Danmark, hvilket gør det velegnet til at besvare min problemformulering. Beslutningen om at undersøge danske folketingspolitikere i perioden 1990-2020 betyder, at den undersøgte case både har spatial (skiftende politikere) og temporal (længere årrække) variation, som bidrager til et dybdegående indblik i elitepolariseringen af sprogbrug. Dette ville være vanskeligt at opnå med andre forskningsdesign og casevalg. Et alternativt casevalg kunne være en undersøgelse af elitepolarisering i sprogbrug på Twitter. Over de senere år er Twitter i stigende grad blevet et medie, som bruges til politisk kampagneføring (Barberá og Rivero, 2015). Mediet giver en unik adgang til politisk sprogbrug, som kan undersøges over tid. Men hvor en sådan undersøgelse muligvis ville være gennemførlig i en amerikansk kontekst, så er mulighederne i en dansk kontekst begrænset. Det skyldes særligt, at det ikke er alle danske politikere, som bruger Twitter. Samtidigt har Twitter alene eksisteret siden 2006 – og den politiske kommunikation på mediet endnu kortere tid. Sammenlignet med folketingstaler vil valget af Twitter som case begrænse den spatiale og temporale variation.

Intuitivt fordrer valget af et casestudie en præcisering af de enheder og underheder, som indgår i den valgte case (Gerring, 2004). Idet jeg er interesseret i elitepolariseringen på et aggregeret niveau, har jeg valgt Folketingets partier som de primære enheder i undersøgelsen. Selvom det reelt er muligt at undersøge polariseringen mellem enkelte folketingsmedlemmer, så vurderes det uhensigtsmæssigt. Det skyldes, at det teoretiske

udgangspunkt for undersøgelsen er litteraturen om elitepolarisering, hvor de undersøgte enheder typisk er aggregeret til partier og politiske blokke (Dalton, 2008; Goet, 2019; Peterson og Spirling, 2018). Det skal også ses i lyset af, at jeg primært interesserer mig for den overordnede udvikling i elitepolarisering.

4.2 Data: Folketingstaler og -afstemninger

Den primære datakilde i specialet er folketingsstaler fra perioden 1990-2020. Derudover inddrager jeg et mål for elitepolarisering i afstemninger i testen af forskelshypotesen. Jeg benytter det mål, som Christiansen (2019b) bruger i sit studie af elitepolarisering i Danmark. Målet er stillet direkte til rådighed af Flemming Juul Christiansen i sin færdige form. Jeg har derfor ikke mulighed for at lave en dybdegående gennemgang af de beslutninger, som er truffet i relation til konstruktionen af datasættet. Derfor opsummerer jeg kort de beslutninger, som Christiansen (2019b) selv beskriver i sin artikel.

4.2.1 Folketingstaler fra 1990 til 2020

For at implementere de maskinlæringsmodeller, som jeg bruger til at måle elitepolarisering i sprogbrug, kombinerer jeg to forskellige datasæt. Det ene datasæt er stillet til rådighed af min vejleder Frederik Hjorth og dækker perioden fra 1990-2006 (Hjorth, 2019). Det andet datasæt er konstrueret med afsæt i offentligt tilgængelige folketingsstaler fra 2009 til 2020. Talerne er tilgængelige online som XML-filer. For at omdanne talerne til et brugbart format har jeg skrevet et Python-script, der for hver tale indhenter tidspunktet for talen, talerens navn, parti og titel (for eksempel minister eller formand)¹. Det har ikke været muligt at inkludere taler fra 2007 og 2008, hvorfor jeg ikke kan lave et polariseringsmål for disse år. Når jeg i de følgende afsnit omtaler perioden for datasættet, er det således med forbehold for, at jeg ikke har data for de pågældende år. Jeg vurderer ikke, at det udgør et større problem for undersøgelsen, da min problemformulering er fokuseret på den generelle tendens og udvikling i elitepolarisering i sprogbrug, som forventeligt er afspejlet i den øvrige data.

Det samlede datasæt består før den indledende databehandling af i alt 838.181 forskellige taler. Jeg betragter datasættet som populationsdata, da det dækker samtlige taler, som er blevet afholdt af politikere i den periode, som jeg har data for. Datasættet giver

¹Folketingstalerne er tilgængelige som XML-filer på <ftp://oda.ft.dk/ODAXML/Referat/samling/>

et unikt indblik i danske folketingspolitikeres sprogbrug i et ensartet format, hvilket er afgørende for specialets deskriptive fokus på udviklingen i polariseringen af elitens sprogbrug over tid. Talerne er opdelt efter folketingssessioner, som løber fra oktober til oktober. En konsekvens heraf er, at mange af figurerne i specialet vil vise værdier ud for året 1991 på x-aksen, selvom værdierne også er baseret på data fra 1990. Det betyder også, at mit databrud ser ud til omfatte 1990 og 2009 i figurerne, men data for disse år indgår under 1991 og 2010.

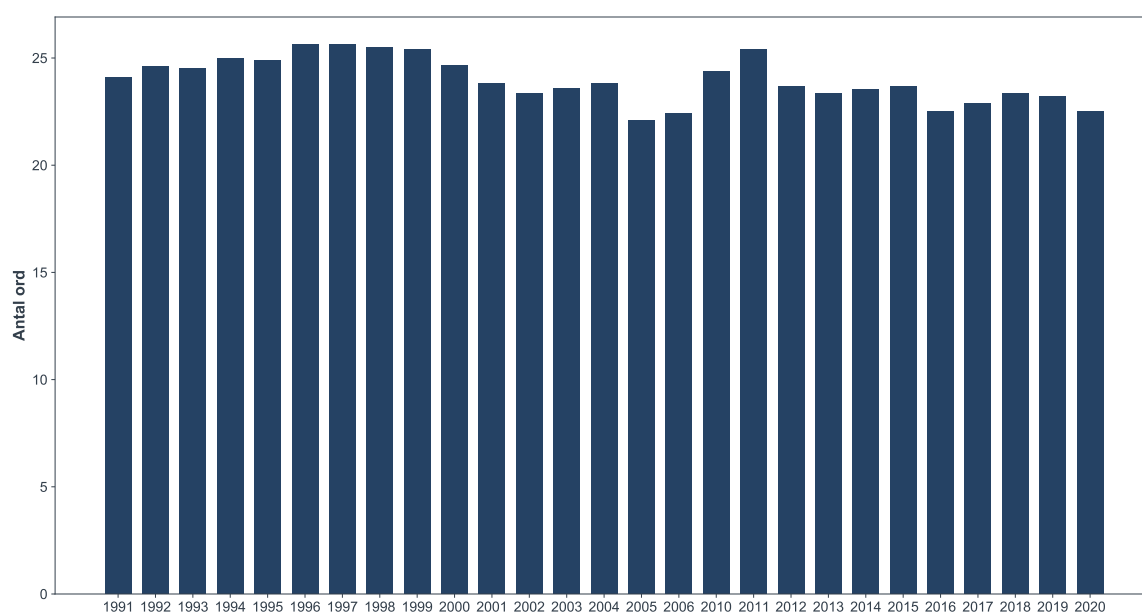
Der er foretaget en række modifikationer af datasættet for at sikre kvaliteten af data. For det første afgrænses data til større partier og samtidig ekskluderes løsgængere fra datasættet. Det betyder for eksempel, at Uffe Elbæk ikke er med i datasættet i perioden efter, at han trådte ud af Alternativet, og at partiet Frihed 2000 udgår, da partiet alene var i Folketinget fra 1999 til 2001. Derudover har jeg ikke taget grønlandske eller færøske partier med. Det kan være problematisk, hvis udeladte partier og løsgængere har en særlig betydning for polariseringen i sprogbrug. Jeg vurderer dog, at det ikke udgør et problem, fordi der er tale om meget få partier og politikere, som forventeligt ikke har en større betydning for den generelle polarisering af politikeres sprogbrug. Derudover er det netop polariseringen mellem Folketingets større og historiske partier, som er afgørende for at forstå de overordnede tendenser i elitepolarisering i sprogbrug. Tilpasningen af datasættet medfører et fald på ca. 10.000 folketingstaler.

For det andet har jeg fjernet en række observationer, som er dateret til urealistiske årstal såsom "1199". Der alene tale om ca. 800 taler, og der er ikke en systematik i afsenderen bag talerne. Jeg vurderer derfor, at det er uproblematisk at fjerne talerne. For det tredje har jeg fjernet taler, der alene indeholder processuelle vendinger ("tak for det" og "ja, det kan jeg bekræfte.") samt taler af Folketingets formand og næstformænd. Det sker ud fra en betragtning om, at sådanne vendinger og taler blot vil være unødigt støj i forhold til mit fokus på at undersøge polariseringen i sprogbrug (Fiva et al., 2021; Goet, 2019; Peterson og Spirling, 2018). Problematikken omkring proceduralt sprogbrug uddybes yderligere i afsnit 4.3.2. Da der et relativt stort antal formelle taler resulterer justeringen i, at antallet af taler falder til 543.326 taler.

Talerne varierer markant i deres længde og er dermed præget af stor heterogenitet. For eksempel er de korteste taler alene et ord lange ("ja" eller "nej"), mens de længste taler indeholder op imod 9.680 ord. Det gør det svært at betragte talerne som sammen-

lignelige enheder, da der er en klar forskel på længden af talerne (Hjorth, 2019). Der er således behov for at konvertere talerne til enheder, som i større grad er sammenlignelige, da det giver en bedre mulighed for at foretage manuelle valideringer af talerne i mit elitepolariseringsmål. En løsning på udfordringen er at bryde talerne op i mindre stykker på 3-5 sætninger. Alternativt kunne talerne være brudt op i enkelte sætninger. En risiko herved er dog, at meningen med sætningen ikke fastholdes, da den skal forstås i relation til sætningerne omkring den (Hjorth, 2019). Jeg vælger derfor at holde fast i løsningen med talestykker på 3-5 sætninger. I figur 4.1 fremgår det gennemsnitlige antal af ord per talestykke efter opdelingen. Det gennemsnitlige antal ord per talestykke konsistent befinder sig mellem ca. 20 til 25 ord per år, hvilket indikerer en succesfuld opdeling af talerne i sammenlignelige talestykker.

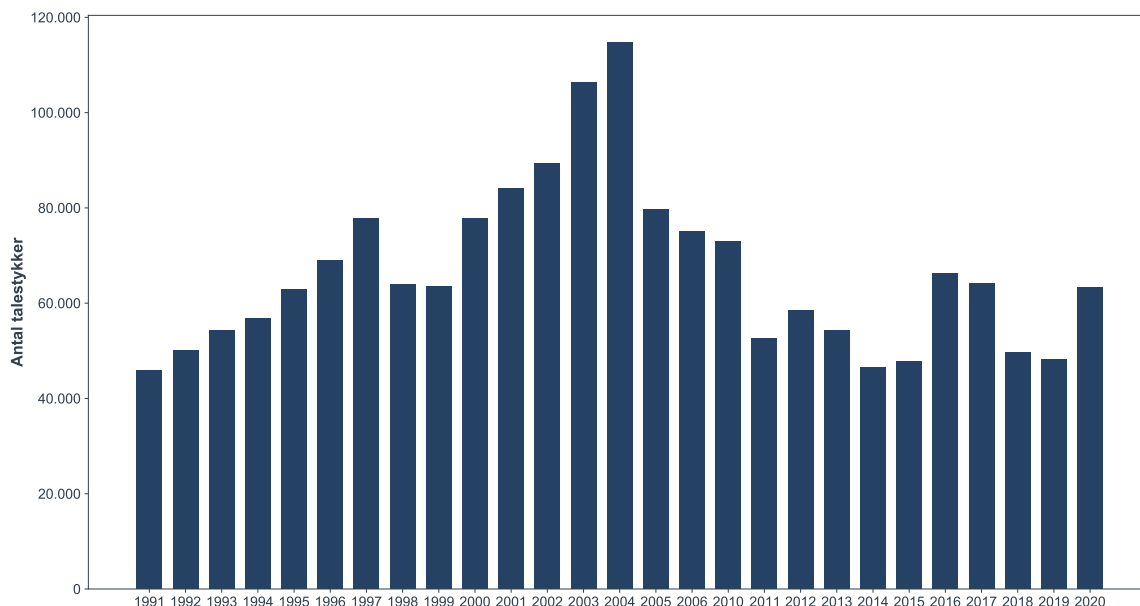
Figur 4.1: Gennemsnitligt antal ord per talestykke over tid



Det samlede datasæt med folketingstaler indeholder 1.819.747 talestykker fordelt mellem ca. 717 forskellige folketingspolitikere i perioden 1990 til 2020. Figur 4.2 viser antallet af taler per folketingsår for perioden. Antallet af talestykker svinger mellem ca. 45.500 og 114.500 årligt. Det fremgår også af figuren, at der ikke er nogen klar tendens til, at der kommer flere taler over tid i perioden. Der er således ikke umiddelbart grund til at forvente, at resultaterne af analysen påvirkes af specifikke tidslige mønstre i antallet af taler. Det bemærkes dog, at antallet af taler i perioden 2001 til 2004 er væsentligt højere end øvrige år, selvom det falder igen efter 2005. Jeg har dette for øje i forbindelse med

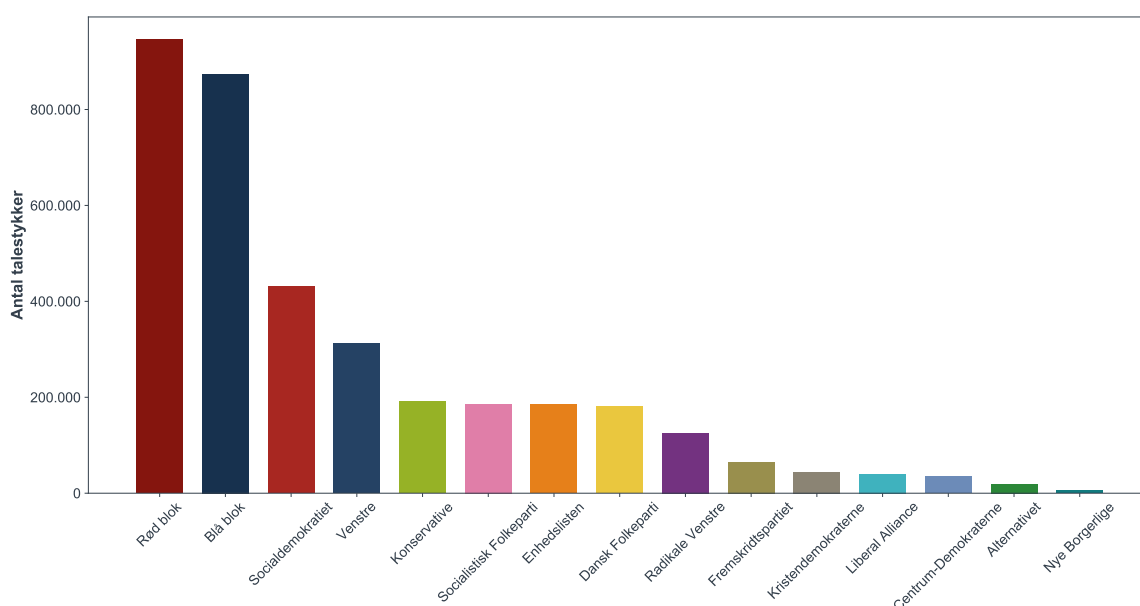
analysen af talestykkerne, men idet antallet af talestykker fordeler sig relativt konstant for de øvrige år, vurderes det ikke at være et større problem.

Figur 4.2: Antallet af talestykker over tid



Endelig har jeg som led i konstruktionen af datasættet tilføjet en række variable. To af de centrale variable i datasættet er politikeres parti- og bloktilhørsforhold. Disse variable er afgørende for at kunne måle polariseringen i sprogbrug, da enhederne i specialets analyser er politiske partier og blokke. Valget af enheder uddybes nærmere i afsnit 4.4 og 4.7. Figur 4.3 viser antallet af talestykker fordelt mellem de forskellige blokke og partier. Det ses, at fordelingen af talestykker mellem rød (945.638) og blå blok (874.109) er balanceret, selvom rød blok er afsenderen bag flest talestykker. Det fremgår også, at Venstre og Socialdemokratiet er de to partier med flest talestykker, mens Konservative, Socialistisk Folkeparti, Enhedslisten og Dansk Folkeparti alle befinder sig omkring ca. 200.000 talestykker. På trods af at data er balanceret på blokniveau, så er der en ubalance i data, når enkelte partier sammenlignes med resten af Folketinget. Det kan have indflydelse på anvendelsen af maskinlæring til at måle elitepolarisering i sprogbrug og behandles derfor særskilt i afsnit 4.4. Den samlede oversigt over variable fremgår af appendiks A.

Figur 4.3: Antallet af talestykker per parti



4.2.2 Lovafstemninger fra 2001 til 2014

Det andet datasæt er stillet til rådighed af Flemming Juul Christiansen, som tidligere har anvendt det til at undersøge polariseringen af afstemningsadfærd i Danmark (Christiansen, 2019b). Datasættet består af afstemninger fra Folketinget i perioden fra 2001 til 2014 og omfatter ca. 3.780 lovforslag. Ud af det samlede antal afstemninger blev 3.712 lovforslag vedtaget, og udviklingen i målet for polariseringen af afstemninger er således ikke drevet af private lovforslag, der er stemt ned (Christiansen, 2019a). Det skal understreges, at datasættet ikke omfatter samtlige afstemninger i Folketinget, men derimod kun lovforslag (Christiansen, 2019b). Det betyder, at datasættet ikke inkluderer beslutningsforslag eller forslag til vedtagelse. Beslutningen hviler på en betragtning om, at lovforslag, grundet deres bindende karakter, typisk er genstand for politisk uenighed såvel som kompromisser, hvilket er centralt for den polarisering i afstemningsadfærd (Christiansen, 2019a). Det vil blive mere klart i afsnit 4.5, hvor jeg operationaliserer målet for polarisering af afstemningsadfærd. Desuden betyder fokuset på lovforslag, at elitepolariseringsmålet bygger på afstemninger, som er sammenlignelige i deres indhold, og målet vil ikke være påvirket af, at der i nogle perioder er flere beslutningsforslag end andre.

Målet for elitepolariseringen af afstemningsadfærd bygger også på data, der er opdelt efter folketingssessioner (ligesom taledata), der løber fra oktober til oktober. Det afviger

fra udgangspunktet i Christiansen (2019b), men jeg har på forespørgsel fået tilsendt et årligt mål for elitepolariseringen i afstemninger. Det er centralt for min undersøgelse, da forskelshypotesen testes ved at sammenligne det talebaserede mål for elitepolarisering med det afstemningsbaserede mål. Derfor er det afgørende at have to mål, som bygger på data med sammenlignelige tidsenheder. Der er dog ikke et komplet tidsligt overlap for de to datasæt. Afstemningsdataen dækker en kortere periode (2001-2014) end taledataen (1990-2020). Desuden er der for den overlappende tidsperiode et databrud i taledatasættet (2007-2008). Det reducerer antallet af sammenlignelige datapunkter og dermed den samlede tidsperiode, for hvilken det er muligt at sammenholde de to forskellige mål. Jeg adresserer udfordringen i afsnit 4.7 om min analysestrategi. Endelig vil jeg igen understrege, at jeg har fået målet for polarisering i afstemninger tilsendt i sin endelige form, hvorfor det ikke er muligt at give en nærmere beskrivelse af databearbejdningsprocessen.

4.3 Tekst-som-data og superviseret maskinlæring

Det primære metodiske værktøj i mit speciale er superviserede maskinlæringsmodeller, som bruges til at måle elitepolariseringen i sprogbrug på baggrund af folketingstaler. Specialet skriver sig dermed ind i politologiens voksende tekst-som-data-litteratur. Det er ikke uden udfordringer, da anvendelsen af tekst-som-data forudsætter en grundig præprocessering for at sikre troværdige resultater (Grimmer og Stewart, 2013; Raschka, 2015). I de følgende afsnit vil jeg først introducere principperne bag tekst-som-data. Dernæst gennemgår jeg de præprocesseringstrin, som jeg har foretaget for at anvende talerne til superviseret maskinlæring. Endelig præsenterer jeg de maskinlæringsalgoritmer, som anvendes til at konstruere målet for elitepolarisering i sprogbrug. Inden jeg går videre er det dog nødvendigt med en kort beskrivelse af logikken bag det maskinlæringsbaserede mål for elitepolarisering i sprogbrug, da det motiverer nogle af mine beslutninger i præprocesseringen af data.

I simple termer er målet for elitepolarisering i sprogbrug konstrueret ved, at jeg træner en maskinlæringsmodel til forudsige partitilhørsforholdet bag en tale alene med udgangspunkt i de ord, som indgår i talen. Modellen lærer således at identificere bestemte ord og vendinger, som er associeret med bestemte partier og dermed kan bruges til at "gætte" partitilhørsforholdet for en tale (Goet, 2019). Den grundlæggende intuition er, at når modellen er "god" til at adskille partiernes taler, er der en høj polarisering i sprogbrug,

da partierne taler meget forskelligt. Det omvendte gælder, når modellen har ”svært” ved kende forskel på talerne. Dermed kan jeg anvende et løbende mål for maskinlæringsmodellens performance som et udtryk for, hvordan polariseringen i politikernes sprogbrug ændrer sig over tid. Den konkrete operationalisering og valget af performanceindikator uddybes i afsnit 4.4.

4.3.1 Overordnede antagelser bag arbejdet med tekst-som-data

Tekster, herunder folketalingstaler, kan betragtes som ustruktureret data, da de ikke på forhånd er struktureret med henblik på at blive brugt som data. Derimod er tekster og taler struktureret i overensstemmelse med sproglige og grammatiske regler. For at behandle tekst-som-data er det derfor nødvendigt med en vis grad af abstraktion og omstrukturering af teksters oprindelige form (Grimmer og Stewart, 2013). Det er særligt relevant i arbejdet med store mængder af tekst, hvor det ikke er muligt at læse samtlige tekster af ressourcemæssige årsager. Omstruktureringen af tekster til anvendelige kvantitative data indebærer en forsimpning af den kompleksitet, som tekster og sprog indeholder (Grimmer og Stewart, 2013). Et grundlæggende vilkår for automatiserede analyser af tekster er derfor, at analyserne aldrig kan give en præcis gengivelse af den datagenererende proces, som går forud for konstruktionen af en bestemt tekst. Alligevel kan tekst-som-dataanalyser være velegnede til at måle teoretisk relevante størrelser og udføre simple kategoriseringer af større tekstsamlinger, hvilket netop er hensigten i mit speciale (Grimmer og Stewart, 2013).

Det første skridt, som er nødvendigt for at anvende folketalingstaler til superviseret maskinlæring, er at omdanne talerne til kvantitative data. I den forbindelse anvender jeg i stil med store dele af tekst-som-data-litteraturen den såkaldte *bag-of-words-antagelse* (Grimmer og Stewart, 2013; Slapin og Proksch, 2008). Antagelsen indebærer, at rækkefølgen af ordene i en tale ikke er betydningsfuld for at analysere elitepolariseringen i politikeres sprogbrug. Denne antagelse kan virke urealistisk, da konteksten, som ord bliver sagt i, er relevant og betydningsfuld. Som fremhævet er det dog normal praksis i arbejdet med tekst-som-data, da det er nødvendigt for at transformere talerne til kvantitative data (Grimmer og Stewart, 2013). Det kan derudover tilføjes, at sprogbrug og skrivning kan betragtes som en stokastisk proces. For eksempel vil en specifik person sjældent anvende den præcis samme rækkefølge af ord, hvis vedkommende skulle producere et budskab to

gange (Egerod og Klemmensen, 2019). Derfor er opdelingen af tekster i enkelte ord (unigrams) eller ordpar (bigrams) ofte tilstrækkeligt til at indfange den overordnede mening i en tekst (Grimmer og Stewart, 2013). Selvom bag-of-words-antagelsen intuitivt kan virke urealistisk, vurderes den ikke at udgøre et problem for undersøgelsen. For at fastholde en større del af talernes mening har jeg dog valgt at inkludere både uni- og bigrams. Det sikrer blandt andet, at sammenhængende vendinger som ”illegale flygtninge” eller ”høj skat” ikke udelades af analysen.

Anvendelsen af *bag-of-word-antagelsen* er en af de mest omfattende skridt i min analyse af folketalingstaler. Selvom det forsimples arbejdet med talerne betragteligt, så er mit taledata fortsat præget af høj kompleksitet og dimensionalitet. Her refererer den høje dimensionalitet til den udfordring, at hvert enkelt ord i en samling af tekster udgør en variabel, som min superviserede maskinlæringsmodel skal forholde sig til. Udfordringen skal ikke forveksles med den generelle problematik om dimensionaliteten af det politiske rum, som præger studiet af polarisering. Dimensionaliteten i det politiske rum refererer derimod til det faktum, at politisk konflikt og polarisering kan ske langs mange forskellige dimensioner (for eksempel ideologi eller politiske emner), selvom dimensionerne ikke er af interesse for en given undersøgelse (Benoit og Laver, 2012; Goet, 2019). Problematikken omkring dimensionaliteten af det politiske rum uddybes nærmere i afsnit 4.4. I det følgende afsnit fokuserer jeg derfor i stedet på, hvordan jeg kan reducere kompleksiteten og dimensionaliteten af mit taledata.

4.3.2 Præprocessering af folketalingstaler

Overordnet set anvender jeg en Term Frequency Inverse Document Frequency (TF-IDF) til den kvantitative repræsentation af folketalingstaler. Det indebærer, at talerne repræsenteres ved en matrix, hvor hvert ord udgør en variabel. Værdien af ordet afhænger af, hvor ofte ordet indgår i en tale samt hvor mange taler, der indeholder ordet (Goet, 2019; Raschka, 2015). Helt konkret vil et ord, der indgår i en tale mange gange få en højere vægt end ord, der ikke indgår i talen særlig ofte. Imidlertid vil vægten blive reduceret, hvis ordet også indgår i mange andre taler (Raschka, 2015). Et ord er således særligt betydningsfuldt, hvis det indgår mange gange i én tale uden, at det samtidig indgår mange gange i andre taler.

Før transformationen af talerne til en TF-IDF-matrix er det nødvendigt at reducere

den høje dimensionalitet i talerne. Det er nødvendigt, da en høj dimensionalitet kan øge risikoen for ”overfitting”, hvor maskinlæringsalgoritmerne i specialet tilpasses talemønstre, som ikke er betydningsfulde for at forudsige politikeres partitilhørsforhold, hvilket kan skade validiteten af målet for elitepolarisering i sprogbrug (Raschka, 2015). En central opgave er derfor at finde en balance mellem at fastholde tekstuel information, som er relevant for at måle polarisering, og samtidig forsimple talerne tilstrækkeligt, så tale-data ikke bliver for kompleks og højdimensionel for de anvendte maskinlæringsmodeller (Grimmer og Stewart, 2013). For at opnå dette har jeg valgt (1) at ”stemme” ordene i folketingsalerne og fjerne såkaldte stopord, (2) fjerne punktummer og tal, (3) fjerne formelle vendinger samt (4) fjerne partinavne og navne på politikere. Nedenfor gennemgår jeg (1) og (3-4), da disse ændringer medvirker til at reducere dimensionaliteten af talerne og fokusere talerne på de ord, som er relevante for, at modellerne kan indfange forskelle i sprogbrug.

Formålet med at ”stemme” ord er at reducere ordene i de enkelte taler til deres mest grundlæggende betydning eller ordstamme (Raschka, 2015). For eksempel betyder det, at ord såsom ”familie, familier, familierne, familiernes” alle reduceres til ”familie” (Grimmer og Stewart, 2013). Det reducerer antallet af unikke ord i talerne og dermed dimensionaliteten af data, men fastholder dog essensen af de forskellige ord i det samlede korpus af tekster (Grimmer og Stewart, 2013). Jeg anvender Python-pakken NLTK, som har en algoritme, der understøtter det danske sprog. NLTK-pakken understøtter også fjernelsen af stopord – det vil sige ord, der ikke har en mening, men primært tjener grammatiske funktioner (for eksempel ”men”) (Grimmer og Stewart, 2013).² Da pakken ikke udelukkende er udviklet til dansk, har jeg også valgt at fjerne ord, der indgår i mere end 95 procent af alle taler for at fjerne hyppige ord, som ikke fanges af stopordene i NLTK-pakken. Det bemærkes, at 95 procent er relativt konservativt sat, da det er meget få ord, der indgår i så stor en andel af talerne. Beslutningen beror på et hensyn til ikke at fjerne ord, som potentielt er vigtige for forskelle i politikeres sprogbrug.

Et afgørende trin i arbejdet med politiske taler er at fjerne taler, som alene indeholder processuelle vendinger, der ikke er relevante for polariseringen i sprogbrug (Gentzkow et al., 2019; Peterson og Spirling, 2018). Formålet med at fjerne processuelle talestykker er at fjerne unødigt støj fra datasættet, som ikke er relevant for maskinlæringsmodellernes

²Dokumentation for pakken kan findes på <https://www.nltk.org/>

evne til at forudsige partitilhørsforhold. Som fremhævet i gennemgangen af datasættet, har jeg fjernet en stor del af de processuelle talestykker ved at fjerne taler af Folketingets formænd og næstformænd. I tillæg hertil har jeg lavet en liste med processuelle vendinger med udgangspunkt i gennemlæsningen af 500 tilfældige taler. De formelle vendinger fra listen er fjernet fra talerne. Denne metode er langt fra komplet, men jeg vurderer, at det er tilstrækkeligt til at fjerne de mest støjende tekststykker fra talerne. Det skal også ses i lyset af, at jeg som tidligere nævnt har valgt at fjerne ord, der er meget hyppige på tværs af taler. Dette vil højst sandsynligt også fange processuelle vendinger, som netop er karakteriseret ved at være frekvente.

Endelig har jeg også valgt at fjerne partinavne, bloknavne, partiforkortelser, politiker- og minister titler fra talerne. Det skyldes et specifikt hensyn til, at maskinlæringsmodellerne skal måle elitepolarisering. Det vil være u hensigtsmæssigt, hvis maskinlæringsmodellerne havde nemt ved at finde forskelle i folketingstaler alene fordi en tale indeholder referencer til bestemte partinavne eller navne på politikere (Fiva et al., 2021). Det er ikke et urealistisk scenarie, at politikere oftest refererer til politiske modstandere. En konsekvens heraf vil være, at en maskinlæringsmodel, der skal foretage en klassificering af en tales bloktilhørsforhold, kan få en høj performance ved at "gætte" på den modsatte blok af den, som nævnes i en given tale. Dermed vil modellens performancemål ikke afspejle sproglige forskelle, men derimod blot et mønster i den måde, som politikere refererer til andre politikere på. Hvis performancemålet skal være et validt mål for sproglig polarisering, er det afgørende, at det er forskelle i ordvalg, der styrer modellernes evne til at forudsige politiske tilhørsforhold.

4.3.3 Superviseret maskinlæring: Algoritmevalg og krydsvalidering

Formålet med maskinlæringsmodellerne, som anvendes i specialet, er at forudsige partitilhørsforholdet for politikerens bag en given tale. Idet partier og blokke kan betragtes som kategoriske variable, er der tale om et klassifikationsproblem (Hastie et al., 2009). Mere præcist er der tale om et binært klassifikationsproblem, da modellerne skal forudsige om en given tale tilhører (1) rød eller blå blok eller (2) et bestemt parti eller resten af Folketinget (Hastie et al., 2009). Denne opdeling er motiveret nærmere i afsnit 4.4, hvor jeg operationaliserer målet for elitepolarisering i sprogbrug.

Den grundlæggende idé med superviseret maskinlæring er, at den anvendte algorit-

me selv identificerer de mønstre i data, der er relevante for at forudsige et givent udfald (Grimmer og Stewart, 2013; Hastie et al., 2009). I mit tilfælde vil det sige, at den anvendte algoritme skal forudsige det politiske tilhørsforhold for en tale på baggrund af TF-IDF-matricen med de vægtede ord fra folketalingstalerne. Hvordan den valgte algoritme konkret estimerer sammenhængene, er ikke fokus for analysen. Fokus er derimod på modellens evne til at lave korrekte forudsigelser (Bach et al., 2019). Det betyder, at der i modsætning til undersøgelser af kausale sammenhænge ikke skal tages højde for problematikker såsom udeladt variabel bias og multikollinearitet. Det betyder også, at den konkrete formel, hvorved algoritmen forudsiger talers partitilhørsforhold ikke er af interesse for undersøgelsen, da fokus alene er på modellens evne til at lave forudsigelser.

I dette speciale afviger jeg delvist fra formålet om at sikre de bedst mulige forudsigelser. Det skyldes, at jeg anvender modellernes evne til at udføre korrekte forudsigelser som et mål for forskelle i politisk sprogbrug. Hvis jeg alene var interesseret i at prædiktere politiske tilhørsforhold, kunne jeg for eksempel have inddraget øvrige variable såsom profession eller køn, hvilket forventeligt vil øge modellernes præcision. Det er dog ikke formålet med specialet. Fokus er derimod at måle elitepolarisering i sprogbrug, hvorfor jeg alene anvender de konkrete ord i talerne som input i modellerne (Peterson og Spirling, 2018). Med denne præcisering på plads, vil jeg nu introducere de algoritmer, som anvendes i specialet og de forskellige trin, der indgår i træningen af modellen.

Der findes mange forskellige superviserede maskinlæringsalgoritmer, som kan løse klassifikationsproblemer (Hastie et al., 2009). Der er dog ikke en algoritme, der kan siges at være entydigt bedst på tværs af klassifikationsproblemer, da forskellige algoritmer har forskellige styrker og svagheder (Hastie et al., 2009). Jeg har valgt at anvende to forskellige algoritmer til analyserne i specialet – Multinomial Naïve Bayes (MNB) og Random Forest Classifier (RFC). Begge algoritmer kan anvendes til at løse klassifikationsproblemer og er ligeledes anvendt i den øvrige litteratur, der undersøger elitepolarisering med afsæt i parlamentariske taler (Goet, 2019; Peterson og Spirling, 2018). Valget af algoritmerne bygger på en samlet afvejning af ønsket om at generere gode forudsigelser samtidig med, at der også tages hensyn til simplicitet, reproducerbarhed, overfitting og computationelle ressourcer (Hastie et al., 2009). Algoritmerne omfavner disse hensyn i forskelligt omfang. Anvendelsen af begge algoritmer bidrager desuden til at validere resultaterne fra analysen, da udviklingen i elitepolarisering gerne skal være ens på tværs af algoritmerne - og

altså ikke algoritmeafhængige.

MNB er bredt anvendt i arbejdet med tekst-som-data (Hastie et al., 2009; Rennie et al., 2003). Det skyldes særligt, at algoritmen er hurtig, simpel og effektiv, hvilket gør den velegnet til store datamængder (Rennie et al., 2003). Algoritmen er givet ved følgende formel:

$$P(c|d) \propto P(c) \prod_{1 \leq k \leq n_d} P(t_k|c)$$

Hvis vi har et sæt af taler med $\in 1, 2, \dots, N$, kan den a posteriori sandsynlighed for, at en tale d_i tilhører et parti eller blok c_i betragtes som proportional med den a priori sandsynlighed for, at talen tilhører klasse $c(P(c))$ multipliceret med produktet af antallet af ord n_d i den tale, der bruges til klassifikationen og den viden, som et ord t giver i forhold til om talen tilhører parti eller blok ($P(t_k|c)$). En central udfordring ved MNB-algoritmen er, at den ikke er velegnet til at håndtere ubalancerede klasser. Algoritmen har en tendens til at reducere vægtene for underrepræsenterede klasser i træningsdata, hvilket skaber høj bias, da modellen underfitter og ikke ”lærer” af mønstrene i træningsdata (Rennie et al., 2003). Det er særligt problematisk for mit speciale, da jeg skal klassificere enkelte partier over for resten af Folketingets partier, hvilket medfører store ubalancer mellem klasserne. Derfor implementerer jeg en variation af MNB kaldet Complement Naïve Bayes, der anlægger en anden tilgang til udregningen af vægtene i algoritmen, som gør modellen bedre til at håndtere ubalanceret data (Rennie et al., 2003).

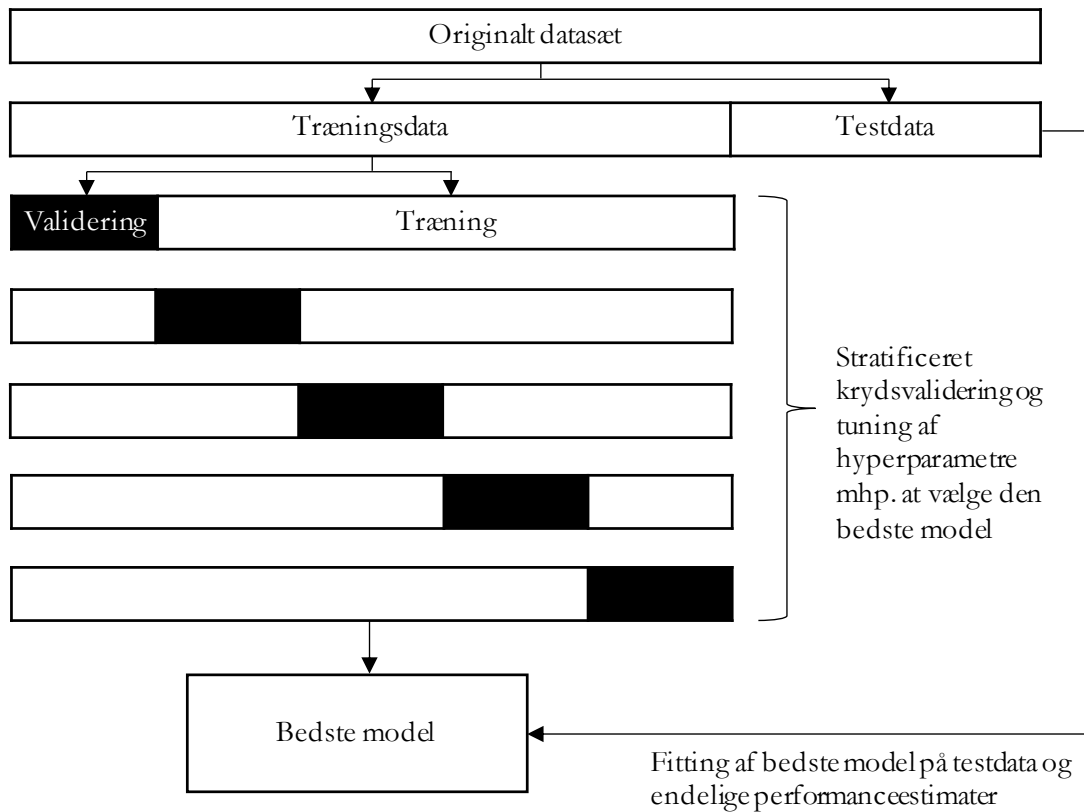
RFC-algoritmen er en ensemblemetode, der bygger på mange forskellige beslutnings-træer for at afgøre, om en tale tilhører en bestemt klasse (Hastie et al., 2009). Algoritmen virker ved at dele data tilfældigt op i forskellige mindre datasæt, hvorefter den samlede ”skov” af beslutningstræer ”stemmer” på det parti eller den blok, som talen tilhører. Fordelen ved at bruge flere beslutningstræer frem for ét beslutningstræ er, at RFC-algoritmen reducerer tendensen for beslutningstræer til at overfitte til træningsdata. Det reducerer modellens varians og bidrager til at øge modellens performance på testdata (Hastie et al., 2009). En ulempe ved RFC-algoritmen er dog, at den ofte er meget kompleks, hvilket kan gøre de estimerede modeller svære at fortolke. Samtidig betyder det også, at modellerne kræver betragtelige computationelle ressourcer (Hastie et al., 2009).

Jeg har valgt at rapportere resultaterne fra de superviserede maskinlæringsmodeller

fra MNB-algoritmen. Beslutningen beror på en betragtning om, at MNB-algoritmen er simpel, kræver mindre computerkraft og har nemt ved at håndtere højdimensionelt data, hvilket netop gælder taledataen i dette speciale. Samtidig er MNB-algoritmen bredt anvendt til at løse tekstklassifikationsproblemer (Rennie et al., 2003). Resultaterne af RFC-modellerne er fortsat afgørende for troværdigheden af analysens resultater, da de som tidligere nævnt tjener et validerende formål. Jeg har vedlagt resultaterne af RFC-modellerne i appendiks B.3. Resultaterne viser, at der er en ensartet relativ udvikling i performancemålene (elitepolarisering i sprogbrug) på tværs af algoritmerne. Det indikerer, at udviklingen i det maskinlæringsbaserede mål for elitepolarisering ikke blot skyldes algoritmiske forskelligheder, hvilket styrker målets troværdighed. Hensynet til computationel effektivitet er ikke uvæsentligt. For eksempel tager modellen, der måler forskelle i sprogbrug mellem politiske blokke, over to døgn at gennemføre. Dette er til trods for, at jeg anvender en computer med 16 CPU'er på Google Cloud Platform til at køre mine modeller.

En udfordring i arbejdet med maskinlæring er risikoen for overfitting, hvilket blandt andet viser sig ved, at modellen tilpasses arbitrære mønstre i træningsdata, som ikke er informative for modellernes evne til at forudsige mønstre uden for træningsdata (out-of-sample) (Hastie et al., 2009). Ovenfor har jeg kort adresseret, hvordan de valgte algoritmer er disponeret over for overfitting. Udfordringen kan dog også adresseres som led i træningen af modellerne ved at optimere algoritmernes hyperparametre og evaluere modellen på data, der ikke tidligere er anvendt til at træne modellen (**probst'tunability'2018**). I dette speciale anvender jeg hold-out validering til at evaluere de optimerede modellers out-of-sample performance. Det gør jeg ved at dele mit datasæt op i træningsdata (75 procent) og testdata (25 procent). Modellerne fittes på træningsdata og evalueres derefter på testdata for at vurdere modellernes evne til at klassificere ukendte taler. Som led i træningen af modellerne foretager jeg krydsvalidering med 5 folder, hvor jeg anvender en grid search til at optimere kombinationen af hyperparametre. Grid searchen fungerer ved, at jeg angiver et mulighedsrum for kombinationer af forskellige hyperparametre, der afprøves og evalueres enkeltvis i krydsvalideringen (Bach et al., 2019; Raschka, 2015). Det anvendte mulighedsrum for hyperparametrene kan findes i appendiks B.1. I figur 4.4 nedenfor har jeg opsummeret processen.

Figur 4.4: Proces for træning og test af superviserede maskinlæringsmodeller



Endelig har jeg anvendt en bootstrappingprocedure til at konstruere konfidensintervaller, som kan give et billede af usikkerheden bag forudsigelserne i mine maskinlæringsmodeller. Det indebærer konkret, at jeg gensampler mit datasæt 50 gange, og derefter træner 50 modeller for hvert år i analysen. Det resulterer i en fordeling af forudsigelser og performancemål, som jeg bruger til at estimere konfidensintervaller (Kuhn og Johnson, 2013). Det gør mig i stand til at afgøre, om udviklingen i maskinlæringsmodellernes performance bliver signifikant bedre eller dårligere over tid, og dermed også om en given udvikling i elitepolariseringen i sprogbrug er signifikant. Af hensyn til computationelle ressourcer er bootstrappingproceduren alene gennemført for modellerne, der bruges til at måle sproglige forskelle mellem de politiske blokke.

4.4 Operationalisering af elitepolarisering i sprogbrug

I teoriafsnittet definerede jeg elitepolarisering i sprogbrug som en udvikling, hvor forskellen mellem politiske partiers sprogbrug bliver større over tid. I litteraturen eksisterer der

ikke et klassisk mål, som anvendes til at tage højde for sådanne forskelle. Det skyldes blandt andet, at det først i de senere år er blevet muligt at automatisere ekstraheringen af store mængder information fra tekstdata (Grimmer og Stewart, 2013). Udviklingen har også påvirket studiet af elitepolarisering, hvor maskinlæring er blevet anvendt til at måle elitepolarisering i politiske taler (Fiva et al., 2021; Gentzkow et al., 2019; Goet, 2019; Peterson og Spirling, 2018). Der er dog ikke klar konsensus omkring, hvordan det måles bedst, hvilket er karakteristisk for arbejdet med big data i samfundsvidenskaben (Hjorth og Loftis, 2020). I operationaliseringen af målet for elitepolarisering i sprogbrug har jeg valgt at lægge mig op af den logik, om anvendes i Goet (2019) og Peterson og Spirling (2018). I begge studier bruges performancemålet *accuracy* for en superviseret maskinlæringsmodel til at måle polariseringen i taler.

Accuracy-målet er en af flere performanceindikatorer for en superviseret maskinlæringsmodels evne til at foretage korrekte klassifikationer. I mit tilfælde vil accuracy afspejle, hvor stor en andel af de samlede taler, som modellen er i stand til at klassificere korrekt og er givet ved:

$$Accuracy = \frac{sande\ positive + sande\ negative}{sande\ positive + sande\ negative + falske\ positive + falske\ negative}$$

Det afgørende element er, at de superviserede modeller alene forudsiger partitilhørsforholdet for en tale ud fra det sproglige indhold i talen. Ændringer i accuracy afspejler dermed, at forskelle og ligheder i partiernes ordvalg gør det henholdsvis ”lettere” og ”sværere” for modellen at lave korrekte forudsigelser. Som fremhævet tidligere er logikken, at en model med en høj præcision har nemt ved at kende forskel på partiernes taler, og det reflekterer en høj grad af polarisering i partiernes sprogbrug (Goet, 2019). Omvendt er polariseringen lav, når modellen klarer sig dårligt, da det indikerer, at de politiske partier taler ensartet omkring forskellige politiske emner. Det er dermed muligt at måle udviklingen i sproglig polarisering over tid ved at observere den relative udvikling i accuracy (Goet, 2019). Ved at estimere en model for hvert folketingsår er det muligt at give et billede af den årlige udvikling i elitepolarisering af sprogbrug.

En stor ulempe ved accuracy-målet er, at det er følsomt overfor ubalanceret data (He og Garcia, 2009). Det kompromitterer anvendeligheden af accuracy som et indirekte mål

for elitepolarisering, da min partivise sammenligning mellem enkelte partier og resten af Folketinget indebærer ubalanceret data. En løsning på problemet er at gruppere mine data i politiske blokke, der som udgangspunkt er relativt balancerede. Imidlertid mister jeg med denne tilgang en del af de nuancer, som potentielt gemmer sig i polariseringsniveauet for enkelte partier. I modsætning til Goet (2019) og Peterson og Spirling (2018) har jeg derfor også valgt at måle polariseringen for de enkelte partier ved at træne en model, der skal forudsige om en tale tilhører et givent parti eller resten af Folketinget. Det giver et billede af, om det enkelte partis sprogbrug er blevet mere polariseret fra resten af Folketinget over tid. Beslutningen udgør en tilpasning af elitepolariseringsmålet til den danske kontekst, da Danmark har et flerpartisystem, hvor opdelingen i politiske blokke indebærer en forsimpning af forskellene mellem partierne. Tilsammen giver den blokvis og partivise opdeling et bredt billede af polariseringen mellem danske partier. Opdelingerne betyder dog, at accuracy ikke længere er velegnet som indikator for polariseringen, da den partivise opdeling er stærkt ubalanceret.

For at løse problematikken omkring ubalanceret data har jeg valgt at anvende performancemålet AUC (area under the curve) til at måle polarisering i sprogbrug. AUC-scoren er ikke biased i retning af majoritets- eller minoritetsklassen, og kan derfor både anvendes til at måle performance for modeller med balancerede og ubalancerede klasser (He og Garcia, 2009). AUC-scoren kan udledes med afsæt i den såkaldte ROC-kurve (receiver-operator characteristics), som viser sammenhængen mellem sand-positiv-raten og falsk-positiv-raten. I den blokvis opdeling udtrykker sand-positiv-raten antallet af talestykker, der korrekt klassificeres som tilhørende blå blok, mens falsk-positiv-raten udtrykker antallet af talestykker, der fejlagtigt klassificeres som tilhørende blå blok (He og Garcia, 2009). ROC-kurven er en sandsynlighedskurve, hvor AUC-scoren er et mål for graden af separation (for et eksempel på en ROC-kurve se afsnit B.2). AUC-scoren udtrykker, hvor god en maskinlæringsmodel er til at skelne mellem politiske blokke eller de enkelte partier og resten af Folketinget. AUC-scoren kan antage værdier fra 0 til 1, hvor en værdi på 1 afspejler en model, der skelner perfekt mellem talestykker af de forskellige blokke eller enkelte partier og resten af Folketinget. En score på 0,5 udtrykker en model, der er ligeså god som et tilfældigt gæt på afsenderen bag et talestykke. Når AUC-scoren er høj betyder det eksempelvis, at den blokvis model er god til afgøre om en tale er sagt af rød eller blå blok. I lighed med accuracy udtrykker AUC-scoren om en model kan finde

meningsfulde forskelle i den måde, som partierne taler på. Anvendelsen af AUC-scoren som et mål for elitepolarisering i sprogbrug bevarer den underliggende logik, som ligger til grund for anvendelsen af accuracy i øvrige studier, og løser samtidig problematikken omkring ubalanceret data (Goet, 2019; Peterson og Spirling, 2018).

Den centrale intuition bag det maskinlæringsbaserede mål er, at de trænedede modeller ”lærer”, hvordan folketingsmedlemmer fra forskellige partier typisk taler (Goet, 2019). På baggrund af denne viden kan modellerne estimere sandsynligheden for, at en politiker bag en specifik tale tilhører parti A eller B. Folketinget er dermed polariseret, hvis det er nemt at skelne mellem partiernes sprogbrug. Idéen er, at retoriske tilgange, valg af underemner og øvrige semantiske forskelle bruges strategisk af politikere, og graden hvormed disse forskelle kan bidrage til at forudsige partitilhørsforhold afspejler polarisering i sprogbrug (Goet, 2019). Samtidig løser målet også problematikken omkring den høje dimensionalitet i det politiske rum, da forskellene mellem politikere reduceres til én dimension - politikernes sprogbrug (Benoit og Laver, 2012; Goet, 2019). Selvom målet umiddelbart løser en række udfordringer, som præger målingen af polarisering, så er der fortsat tale om et relativt nyt mål, der bygger på nye metodiske værktøjer. Jeg vurderer derfor, at der fortsat er et behov for en grundig validering af målet for at afdække, om målet på meningsfuld vis indfanger den konceptuelle udlægning af elitepolarisering i sprogbrug. Jeg behandler valideringen særskilt i afsnit 4.6.

4.5 Operationalisering af elitepolarisering i parlamentariske afstemninger

I teoriafsnittet definerede jeg elitepolarisering i parlamentariske afstemninger som en udvikling, hvor uenigheden mellem partierne i parlamentariske afstemninger bliver større over tid. For at måle udviklingen i polariseringen af parlamentariske afstemninger er det derfor nødvendigt med et mål, som kan indfange splittelsen mellem partier i lovaftemninger. Til dette formål anvender jeg Christiansen (2019b) indeks for polariseringen af afstemningsadfærd. Indekset er komplekst og kombinerer et mål for dissens i lovgivningsafstemninger med et indeks for polariseringen af partisystemer (Christiansen, 2019b; Dalton, 2008). Målet er givet ved:

$$PPI = \sqrt{\left\{ \sum dissens_{ijk} \times (stemmeandel_i) \times ([partiideologi_i - partisystemideologi]/5)^2 \right\}}$$

hvor i repræsenterer individuelle partier, j angiver lovforslag og k angiver tidsperioden. I tabel 4.1 nedenfor har jeg kort beskrevet de enkelte elementer, som indgår i beregningen af målet.

Tabel 4.1: Variable i målet for polarisering af parlamentariske afstemninger

Variabel	Beskrivelse
Dissens	Angiver afstanden til den vindende koalition. For et lovforslag, som bliver vedtaget, kan dissens antage værdien 0 for partier, som har stemt for lovforslaget, 1/2 for partier som har stemt blankt, og 1 for partier, som har stemt mod lovforslaget. Det modsatte gælder forslag, der bliver stemt ned.
Stemmeandel	Andelen af stemmer, som et parti har modtaget ved seneste folketingsvalg. Bruges til at vægte målet efter partiernes størrelse.
Partiideologi	Det enkelte partis ideologiske placering. Den ideologiske placering bygger på Chapel Hill Expert Survey (CHES), hvor en række eksperter har placeret danske partier på en skala fra 0 til 10 på forskellige politikområder. En undersøgelse af CHES-målet peger på, at det har en tilfredsstillende validitet og reliabilitet (Hooghe et al., 2010). Vurderingen bygger på en sammenligning med blandt andet manifest-data (CMP) og andre eksperterundersøgelser.
Partisystemideologi	Den gennemsnitlige ideologiske placering for partierne i et partisystem. Dette er også beregnet med udgangspunkt i CHES.

Målet for polariseringen af parlamentariske afstemninger er komplekst, da det både indeholder information om afstanden mellem partier i afstemninger, partistørrelsen og ideologiske forskelle (Christiansen, 2019b). Målet kan antage værdier fra 0 til 10, hvor en polarisering på 10 er udtryk for en situation, hvor alle placerer sig ved de ideologiske yderpoler, lovforslaget har opnået 50 procent opbakning og partierne konsekvent stemmer imod partier ved modsatte pol. Det betyder, at polariseringen for eksempel stiger, hvis partier flytter sig væk fra centrum af partisystemet, eller hvis yderpartier bliver

større på bekostning af partier tæt på centrum (Christiansen, 2019b). Samtidig vil polariseringen også stige, hvis partier omkring centrum ikke imødekommer øvrige partier, eller hvis et fløjparti aktivt vælger ikke at samarbejde om lovgivning.

En kritik af målet er, at det indfanger elementer, som ikke er relevante for definitionen af parlamentariske afstemninger. Det vil betyde, at målet har en svækket målingsvaliditet, da det substantielt set ikke indfanger min konceptualisering af polariseringen i parlamentariske afstemninger som uenighed i afstemninger. Jeg vil dog argumentere for, at det også er centralt at forholde sig til ideologiske forskelle og partiernes størrelse for at måle uenigheden i afstemninger. For eksempel vil jeg mene, at det samlede niveau af polariseringen i afstemninger er kvalitativt større, hvis modstanden kommer fra Folketingets fløjpartier, som har vokset sig større, end hvis det kommer fra store moderate partier grundet blokpolitik. Det skal ses i lyset af, at uenigheden i afstemninger ved situationer med blokpolitik kommer fra partier uden større ideologiske forskelle. Dette er blot et eksempel på en situation, hvor inddragelsen af ideologiske forskelle og partistørrelse oplyser nuancerne af uenigheden i lovfølgninger. Samlet set vurderer jeg, at målet af Christiansen (2019b) har en høj målingsvaliditet og indfanger centrale elementer af uenigheden i afstemninger.

4.6 Validering af målet for elitepolarisering i sprogbrug

Operationaliseringen af begrebet for elitepolarisering i sprogbrug indebærer anvendelsen af et nyt og alternativt maskinlæringsbaseret mål. Jeg har derfor valgt at foretage en særskilt validering af målet med udgangspunkt i Adcock og Collier (2001). Helt konkret vil jeg dels foretage en indholdsvalidering, dels en diskriminerende validering. Indholdsvalideringen belyser, om et givent mål indfanger en tilstrækkelig grad af indholdet i det begreb, der ønskes målt (Adcock og Collier, 2001). Den diskriminerende validering har derimod til formål at afgøre, om målet for et givent begreb er forskelligt fra et mål, som reelt indfanger et andet begreb (Adcock og Collier, 2001). I tabel 4.2 nedenfor fremgår de konkrete valideringstests, mens resultaterne af testene fremgår af appendiks D.

Tabel 4.2: Valideringsframework

Type af validering	Nøglespørgsmål	Test
--------------------	----------------	------

Indhold	Er indholdet af talerne med størst sandsynlighed for at tilhøre henholdsvis rød eller blå blok faktisk klart forskelligt fra hinanden?	Manuel
Indhold	Viser fordelingen af sandsynligheder for individuelle taler en tilfredsstillende grad af separation mellem blokke for de forskellige partier?	Visuel
Diskriminerende	Korrelerer målet for elitepolarisering i sprogbrug med målene for ideologisk polarisering og polarisering af afstemninger?	Korrelation

Jeg udfører to forskellige tests som led i indholdsvalideringen af målet for elitepolarisering i sprogbrug. For det første foretager jeg en manuel inspektion af de taler, som har den højeste sandsynlighed for at tilhøre rød eller blå blok. Den grundlæggende intuition er, at de taler, som med størst sandsynlighed tilhører en given blok, skal indeholde ord og vendinger, der afspejler et sprogbrug, som er forskelligt fra hinanden. Testen er foretaget på tværs af fordelings- og værdipolitiske emner (opdelingen af taler i politiske emner uddybes i afsnit 4.7). Resultaterne af testen kan findes i appendiks D.1.1. Her fremgår det for eksempel, at talen med størst sandsynlighed for at være blå på det udlændingepolitiske område er afholdt af Martin Henriksen. Talen indeholder ord såsom *parallelsamfund*, *ghettoområder* og *radikaliserede miljøer*. Omvendt er den udlændingepolitiske tale med størst sandsynlighed for at være rød afholdt af Margrethe Vestager, der blandt andet nævner, ”(...), at 1 år er den maksimale opholdstid for en voksen, og derfor er 1/2 år nok det, der kan strækkes til at være acceptabelt for et barn.” Forskellene mellem de to taler afspejler i stor grad forskelle, som korresponderer med den teoretiske forståelse elitepolarisering i sprogbrug, og det indikerer, at det maskinlæringsbaserede mål faktisk måler indholdsmæssige forskelle i politikernes sprogbrug.

For det andet foretager jeg en visuel inspektion af fordelingen af prædikterede sandsynligheder for enkelte taler i forhold til, om talen tilhører blå eller rød blok. Idéen er, at sandsynligheden for, at en tale tilhører blå blok som udgangspunkt bør være mindre for taler afholdt af røde partier end taler afholdt af blå partier. Hvis det ikke er tilfældet, vil det indikere, at modellerne ikke kan finde forskelle i partiernes tale, som stemmer overens med eksisterende skillelinjer i dansk politik. For eksempel vil det være bekymrende, hvis

fordelingen af sandsynligheder for Enhedslistens taler var skæve i retning af blå blok. Appendiks D.1.2 viser resultaterne af denne test. Det fremgår, at der både på tværs af alle taler og taler inden for bestemte værdi- og fordelingspolitiske emner er skæve sandsynlighedsfordelinger i retning af rød blok for røde partier og blå blok for blå partier. Det indikerer, at det maskinlæringsbaserede mål for elitepolarisering faktisk indfanger indholdsmæssige forskelle i partiernes sprogbrug, som stemmer med den eksisterende forståelse af partiernes bloktilhørsforhold.

Den diskriminerende validering udføres ved at korrelere målet for elitepolarisering i sprogbrug med målet for elitepolarisering i afstemninger og målet for ideologisk elitepolarisering. Målet for ideologiske elitepolarisering bygger på CHES-data. Som fremhævet i afsnit 4.1 udgør målet for ideologiske polarisering en delmængde af målet for polariseringen af parlamentariske afstemninger. Resultaterne af korrelationerne kan findes i afsnit 5.3 og appendiks D.2. Begge korrelationer er statistisk insignifikante. Det indikerer, at elitepolarisering i sprogbrug ikke blot er konceptuel forskellig fra ideologisk polarisering og polariseringen af afstemninger, men også empirisk forskellig. Det styrker mit argument om, at elitepolarisering i sprogbrug udgør en særkilt dimension af elitepolarisering.

4.7 Analysestrategi

Hypoteserne i specialet afspejler en række forventninger til udviklingen i elitepolarisering, og det er derfor centralt at fastsætte, hvordan disse forventninger konkret testes i analysen. Et første trin er i den sammenhæng at beslutte hvilke partier, der indgår i analysen. Jeg har valgt afgrænse analysen til de partier, som er en del af Folketinget i dag. Det betyder, at jeg ikke inkluderer selvstændige polariseringsmål for Centrum-Demokraterne, Fremskridtspartiet og Kristendemokraterne (som dog har ét mandat), selvom de på et tidspunkt har været en del af Folketinget i den undersøgte periode. Beslutningen hviler på en betragtning om, at det er mest interessant at se på de partier, som har en vælgerbase og er aktuelle i dansk politik. Jeg har desuden valgt at udelade Nye Borgerlige og Alternativet, da partierne ikke har været en del af Folketinget over en længere periode. Det skal ses i lyset af, at det anvendte polariseringsmål er baseret på den relative udvikling i AUC-scoren, som gør det nødvendigt at observere målet over en længere periode for at forstå udviklingen i elitepolarisering. Det bør understreges, at talestykkerne for de udeladte partier fortsat indgår i analysen af de øvrige partier, og der er således ikke tale

om en restriktion af datasættet, der problematiserer repræsentativiteten.

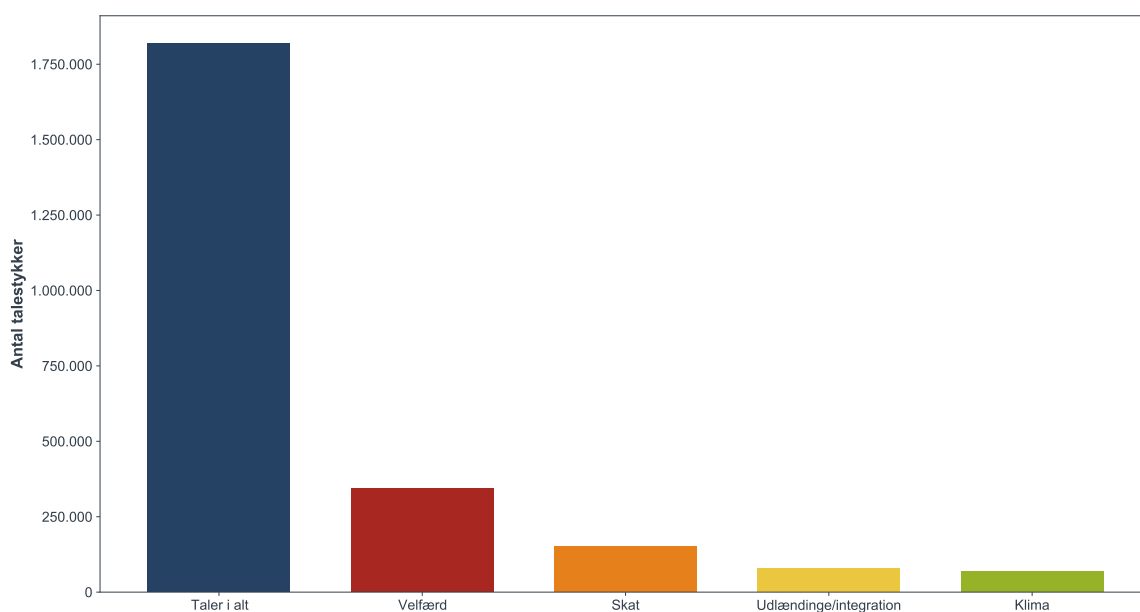
Med den generelle hypotese forventer jeg, at der er sket en stigning i elitepolarisering af sprogbrug over tid. Som udgangspunkt forventer jeg både, at udviklingen er sket mellem rød og blå blok samt de enkelte partier og resten af Folketinget. I den eksisterende litteratur, der anvender et tilsvarende mål for elitepolarisering, er der primært konstateret ændringer i elitepolarisering ved en visuel inspektion af udviklingen i accuracy (Goet, 2019; Peterson og Spirling, 2018). Jeg anvender i specialet den samme tilgang. Det betyder, at udviklingen i AUC-scoren alene understøtter forventningen i den generelle hypotese, hvis der for hele den undersøgte periode konstateres en stigende tendens i AUC-scoren. Hvis niveauet for AUC-scoren derimod ikke er højere i dag sammenlignet med tidligere, kan der ikke konstateres en klar stigning i polariseringen. Det er også tilfældet selvom polariseringen for eksempel er steget mellem 2010 og i dag.

Den værdi- og fordelingspolitiske hypotese bygger på en forventning om, at der er en større grad af polarisering inden for værdipolitiske emner end fordelingspolitiske emner. Det første skridt er i den sammenhæng at bryde det samlede taledatasæt op i værdi- og fordelingspolitiske emner. Der eksisterer ikke en prædefineret variabel, som angiver om en given tale er afholdt som led i en debat med et bestemt politisk fokus. Jeg har derfor selv opdelt talerne i forskellige værdi- og fordelingspolitiske områder. Af hensyn til at sikre et klart fokus har jeg valgt to konkrete politikområder inden for hver af de to dimensioner. For den fordelingspolitiske dimension har jeg valgt at fokusere på velfærds- og skattepolitik, mens jeg for den værdipolitiske dimension har valgt udlændinge- og klimapolitik. De valgte politikområder udgør typiske eksempler på områder, der falder under de to dimensioner og er dermed velegnede til at undersøge hypotesen (Andersen og Borre, 2003).

For hver af de fire politikområder har jeg udviklet en ordbog, som indeholder ord, der ofte anvendes i relation til debatter om de valgte politikområder. Ordbøgerne bygger på en manuel gennemlæsning af 500 tilfældigt udtrukne taler. I tillæg hertil har jeg selv tilføjet en række ord, som åbenlyst indikerer, at en tale vedrører et bestemt politikområde. Det kunne for eksempel være ”topskat”. Herefter har jeg konstrueret et datasæt for hvert af politikområderne, som består af talestykker, der indeholder minimum ét af ordene fra ordbøgerne. Ordbøgerne kan findes i appendiks C. Selvom der er en risiko for, at jeg ikke fanger samtlige taler, der falder under de forskellige politikområder, så er det centrale,

at de udvalgte taler vedrører de politikområder, som jeg er interesseret i. Desuden er der ingen grund til at forvente, at de taler, som jeg har valgt, er systematisk mere eller mindre polariserede end de taler, som jeg ikke indfanger med ordbøgerne. Derfor burde den anvendte tilgang ikke lede til forkerte estimater for udviklingen i elitepolarisering i sprogbrug. Figur 4.5 nedenfor viser fordelingen af taler på de forskellige politikområder.

Figur 4.5: Fordelingen af talestykker på værdi- og fordelingspolitiske emner



Værdi- og fordelingshypotesen undersøges også ved at observere den relative udvikling i AUC-scoren over tid. Der er trænet særskilte maskinlæringsmodeller for de ordbogsspecifikke datasæt og der laves et årligt elitepolariseringsmål for hvert af politikområderne. På den baggrund er det muligt at sammenholde polariseringstendenserne for de forskellige politikområder med henblik på at undersøge, om udviklingen i polarisering i sprogbrug er steget mere på de værdipolitiske områder. Jeg vil argumentere for, at førnævnte er tilfældet, hvis der kan konstateres en stigende tendens for AUC-scoren på de værdipolitiske områder samtidig med, at AUC-scoren enten er svagt stigende, stabil eller faldende for de fordelingspolitiske områder. Jeg anlægger en konservativ tilgang, hvorfor der er behov for at observere en tydelig forskel i den samlede udviklingstendens, før der kan konstateres en forskel.

Undersøgelsen af de politikspecifikke udviklinger i polariseringen tjener også et metodisk formål. Det skal ses i relation til undersøgelsen af den generelle hypotese, hvor

jeg undersøger polariseringen på tværs af alle emner. Det betyder, at jeg ikke kan være sikker på, at forskellene i politikernes sprogbrug skyldes, at politikerne taler forskelligt om de samme emner eller blot om forskellige emner. Ved at "fastfryse" politikområderne er det muligt at få et billede af, om udviklingen i den generelle polarisering også afspejler en polarisering af politikernes sprogbrug, når debatterne vedrører det samme emne. Det bidrager til at sandsynliggøre, at den eventuelle udvikling i AUC-scoren ikke er drevet af, at partierne taler om forskellige emner, men derimod taler forskelligt om de samme emner.

Den sidste hypotese er forskelshypotesen, der fokuserer på forskellen mellem elitepolarisering af parlamentariske afstemninger og elitepolarisering i sprogbrug. Grundlæggende forventer jeg, at udviklingen i elitepolarisering i sprogbrug har været stærkere end udviklingen i elitepolarisering af afstemninger. I lighed med værdi- og fordelingshypotesen undersøges det ved at sammenholde udviklingen i elitepolarisering af afstemninger med elitepolarisering i sprogbrug. Elitepolariseringen i sprogbrug er steget mere end den afstemningsbaserede elitepolarisering, hvis tendensen er stigende for polarisering i sprogbrug, mens tendensen for polarisering af afstemninger er svagt stigende, stabil eller faldende. Derudover foretager jeg en korrelation mellem de to mål. Idéen er, at der skal eksistere en svag eller ingen sammenhæng mellem målene, hvis de to mål skal afspejle forskellige dimensioner af elitepolarisering. Som fremhævet i præsentationen af mine data, så er der kun data for elitepolariseringen af afstemninger i perioden 2001-2014, hvorfor testen af forskelshypotesen fokuserer på denne periode. I denne periode er der et databrud for folketingstaler. Derfor vil korrelationen mellem de to mål alene bygge på datapunkter fra de år, hvor der er observationer for begge mål. Det er problematisk, da det lave antal af datapunkter i sig selv øger sandsynligheden for insignifikante resultater. Dermed skal testen primært ses som supplerende til den generelle sammenligning af udviklingstendenserne. Forskelshypotesen testes alene på tværs af alle taler og dermed ikke emnespecifikt, da jeg kun har haft adgang til et generelt mål for polariseringen af afstemninger.

For alle hypoteser sammenholdes udviklingen i AUC-scoren fra de primære modeller med udviklingen i AUC-scoren for modeller, som bygger på hypotetisk data, hvor jeg har randomiseret talernes partitilhørsforhold. Formålet med sammenligningen er, at konstruere et placeboscenarie, der viser elitepolariseringen i en situation, hvor afsenderen bag en tale er tilfældigt fordelt med udgangspunkt i partiernes relative andel af talerne.

Hvis det viser sig, at AUC-scoren for de primære modeller ikke er signifikant forskellig fra placeboscenarierne, så betyder det sandsynligvis, at elitepolariseringen i sprogbrug er et resultat af en underliggende spuriøs sammenhæng (Gentzkow et al., 2019; Peterson og Spirling, 2018). Tabel 4.3 opsummerer analysestrategien.

Tabel 4.3: Opsummering af analysestrategi

Hypotese	Empirisk udvikling
Den generelle hypotese	Udviklingen i AUC-scoren har en stigende tendens over hele tidsperioden, og målet er på et højere niveau i dag, end det har været tidligere i den undersøgte periode
Den værdi- og fordelingspolitiske hypotese	Udviklingen i AUC-scoren har en stigende tendens for værdipolitiske taler, mens AUC-scoren er svagt stigende, stabil eller faldende for fordelingspolitiske taler
Forskelshypotesen	Udviklingen i AUC-scoren har en stigende tendens, mens udviklingen i målet for elitepolarisering af afstemninger er svagt stigende, stabilt eller faldende. Svag eller insignifikant sammenhæng mellem elitepolarisering af afstemninger og elitepolariseringen i sprogbrug.

5 Analyse

I dette afsnit præsenterer jeg resultaterne af min analyse og besvarer specialets problemformulering. I afsnit 5.1 testes den generelle hypotese for at afdække den overordnede udvikling i elitepolarisering af sprogbrug på tværs af politiske emner. Derefter går jeg et spadestik dybere ved at analysere udviklingen i polariseringen inden for udvalgte værdi- og fordelingspolitiske emner. Endelig sætter jeg i afsnit 5.3 fokus på, om elitepolarisering i sprogbrug adskiller sig fra elitepolariseringen af afstemninger. Samlet set giver analysen en dybdegående indsigt i den sproglige polarisering mellem danske politikere.

5.1 Den generelle hypotese

Den første del af analysen fokuserer på den generelle udvikling i elitepolarisering i sprogbrug over tid. Udgangspunktet for denne del af analysen er den generelle hypotese:

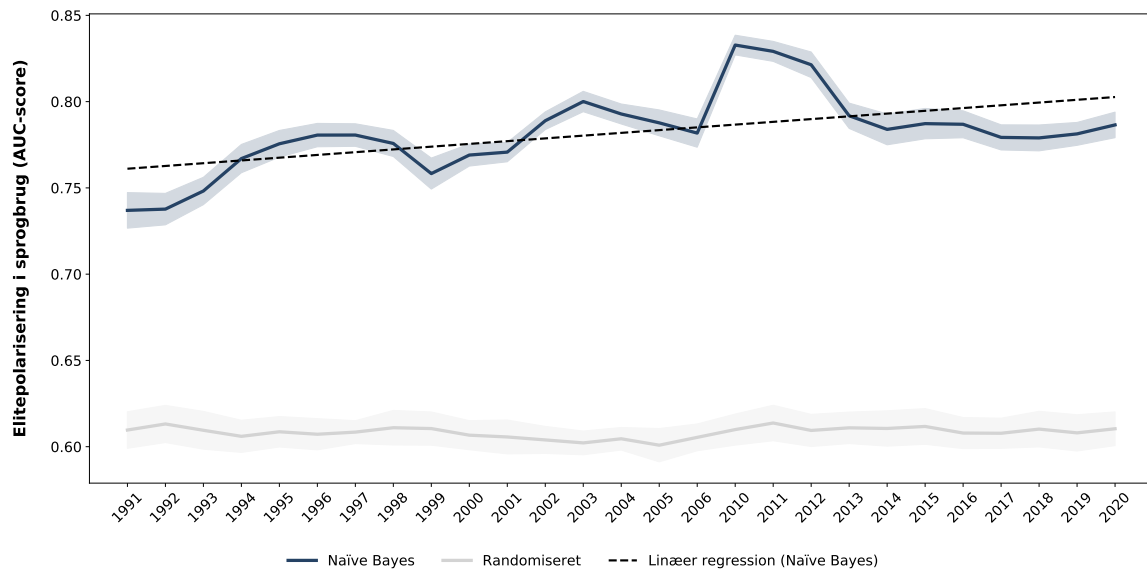
elitepolariseringen i sprogbrug er steget igennem de seneste årtier

Jeg tester hypotesen todelt ved først at analysere den blokvisse polarisering i sprogbrug og derefter den partivise polarisering. Som fremhævet tidligere betragter jeg analysen af den blokvisse polarisering som mine hovedresultater, mens den partivise polarisering bruges til at uddybe og forstå, om der er nogle partier, der er særligt polariserede sammenlignet med resten af Folketinget.

5.1.1 Den blokvisse udvikling

Figur 5.1 viser udviklingen i elitepolarisering i sprogbrug mellem rød og blå blok for perioden 1991-2020. Den blå linje viser den gennemsnitlige udvikling i de årlige maskinlæringsestimater for elitepolarisering i sprogbrug. For at give et billede af den samlede tendens over tid har jeg tilføjet en sort regressionslinje, som viser sammenhængen mellem elitepolarisering i sprogbrug og tid. Endelig har jeg tilføjet en grå linje, der viser placeboudviklingen for elitepolarisering i sprogbrug. Som tidligere nævnt afspejler placeboscenariet et elitepolariseringsmål baseret på et hypotetisk datasæt, hvor de enkelte talestykkers partitilhørsforhold er randomiseret med udgangspunkt i deres andel af de samlede taler. De gennemsnitlige polariseringsmål og tilhørende konfidensintervaller er estimeret med udgangspunkt i en bootstrappingprocedure, hvor jeg har konstrueret 50 forskellige datasæt ved at gensample observationer fra det originale datasæt (Goet, 2019; Hastie et al., 2009). Det fremgår af figuren, at den observerede udvikling i elitepolarisering er signifikant og substantiel forskellig fra placeboscenariet. Det taler for, at der eksisterer en grad af elitepolarisering i sprogbrug mellem blokkene, som ikke blot skyldes tilfældige talemønstre i data.

Figur 5.1: Udviklingen i elitepolarisering i sprogbrug mellem blokke



Note: De blå og grå bånd viser et 95 procent konfidensinterval. Intervallet er baseret på bootstrapping, hvor der for hvert år er simuleret 50 datasæt. Tilgangen er beskrevet nærmere i afsnit 4.7.

I starten af den undersøgte periode er polariseringen i sprogbrug omkring 0,74 med en stigende tendens frem til 2010-2012, hvor polariseringen toppe med et niveau på ca. 0,83. Her er det vigtigt at understrege databruddet fra 2007-2008, som potentielt er medvirkende til, at springet i polarisering fra 2006-niveauet til 2010-niveauet ser mere markant ud, end hvad der faktisk er tilfældet. Formentlig er stigningen sket gradvist, ligesom det er tilfældet for perioden, der leder op til databruddet. Fra 2012 og frem falder den sproglige polarisering til et niveau omkring 0,78 i 2013, og polariseringen er relativt stabilt omkring dette niveau frem mod 2020. En væsentlig observation er, at polariseringsniveauerne i 2013-2020 reelt er sammenlignelige med polariseringen fra midten til slutningen af 1990'erne.

Numerisk set er polariseringen i sprogbrug i 1991 5 procentpoint lavere end polarisering i 2021. Til sammenligning er forskellen mellem polariseringen i 1991 og 2011 omkring 10 procentpoint. Faldet i elitepolariseringen i sprogbrug fra 2013 og frem svarer således til ca. halvdelen af den samlede stigning i polarisering fra 1991 til 2012. Relativt set er faldet i polariseringen i sprogbrug efter 2012 altså ret markant, når man tager polariseringsudviklingen for den samlede periode i betragtning. Polariseringen i sprogbrug er ikke konstant stigende over hele perioden på trods af, at den sorte regressionslinje indikerer en generel stigning i polarisering. Derimod er polariseringen i større grad karakteriseret

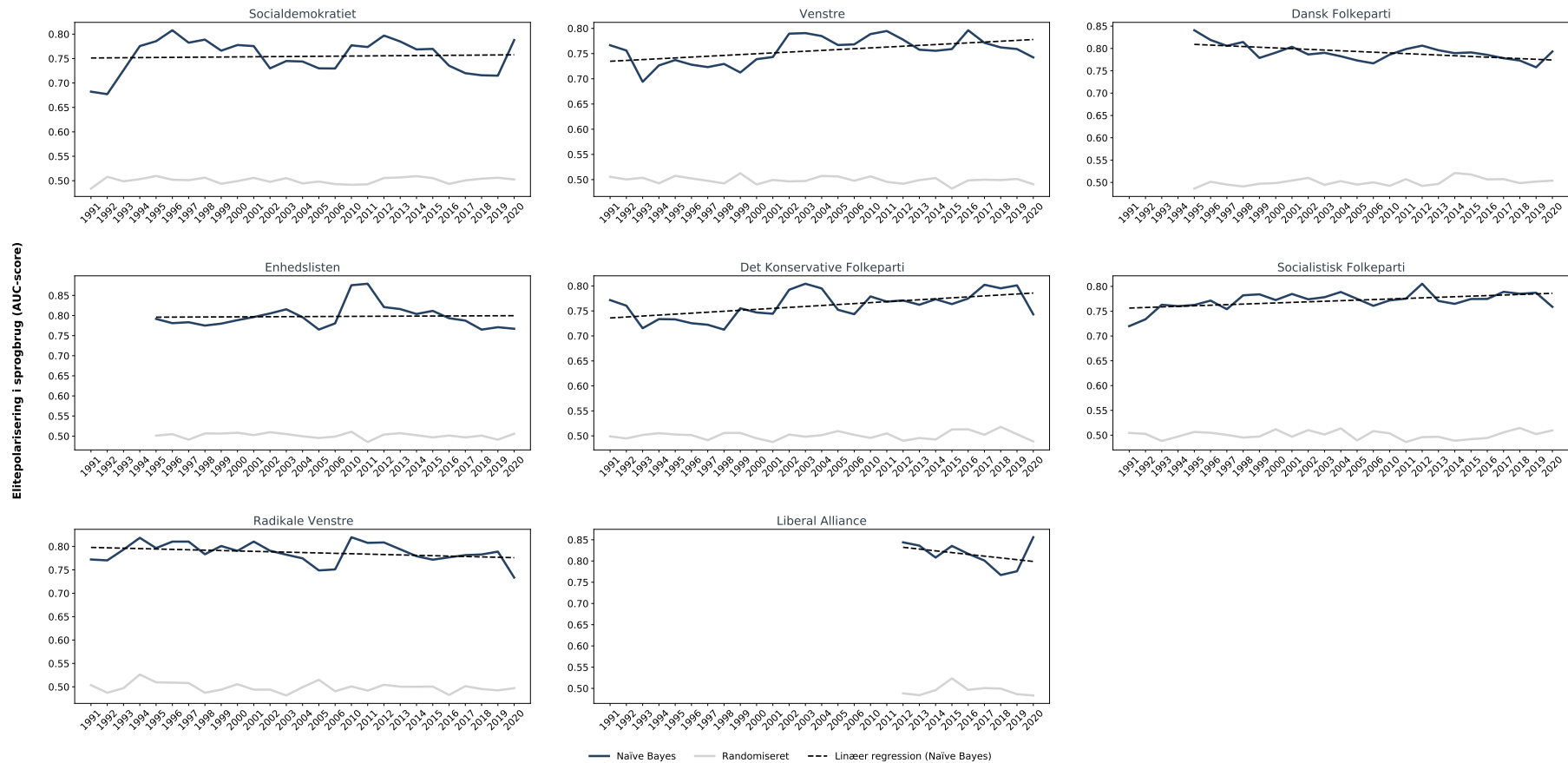
ved to tendenser - dels en stigning fra 1991 til 2012, dels et fald med en efterfølgende stabilisering fra 2013 til 2020.

I substantielle termer indikerer figur 5.1, at forskellen mellem røde og blå politikeres sprogbrug er størst i perioden 2010-2012, og måden hvorpå røde og blå politikere kommunikerer deres politik er mest forskellige fra hinanden i denne periode. Samtidig vokser forskellene mellem røde og blå politikere stødt fra 1991 til 2012, hvorefter forskelle i sprogbrug og ordvalg bliver mindre og mere stabile. Forskellen mellem røde og blå politikeres sprogbrug efter 2013 er ikke mere markant end forskellene mellem røde og blå politikere i midten af 1990'erne. Overordnet set er det derfor tvetydigt, om elitepolariseringen i sprogbrug mellem rød og blå blok er steget gennem de seneste årtier. Fra 1990'erne og frem til 2012 finder jeg støtte til den generelle hypotese om, at polariseringen i sprogbrug er steget gennem de seneste årtier, men fra 2013 til 2020 er polariseringen i sprogbrug stabil, og den er ikke forskellig fra polariseringen i midten af 1990'erne. For at få en bedre forståelse af udviklingen ser jeg nu nærmere på den partivise udvikling.

5.1.2 Den partivise udvikling

Figur 5.2 viser udviklingen i elitepolarisering i sprogbrug mellem enkelte partier og resten af Folketinget. Figuren indeholder en blå linje, som indikerer udviklingen i elitepolarisering, en sort regressionslinje og en grå linje, der viser et randomiseret placeboscenarie. Som fremhævet tidligere har jeg ikke udarbejdet bootstrap-konfidensintervaller for den partivise undersøgelse af polarisering, da der er markante computationelle omkostninger forbundet hermed og denne del af undersøgelsen samtidig indtager en sekundær/supplerende funktion i hypotesetesten (paragraffen fortsætter efter figuren).

Figur 5.2: Udviklingen i elitepolarisering i sprogbrug mellem enkelte partier og resten af Folketinget



Note: En række af partierne i figuren er først dannet i løbet af perioden 1990-2020. Derfor er der kun mål for disse partier i den periode, som de har eksisteret.

Det er derfor ikke muligt at afgøre, om polariseringstendenserne for partierne er signifikante - hverken over tid eller i forhold til placeboscenariet. Jeg vil dog fremhæve, at der er en substantiel forskel på elitepolariseringsniveauerne i placeboscenariet sammenlignet med den reelle elitepolarisering. Det taler umiddelbart for, at resultaterne ikke er en konsekvens af et underliggende mønster i data. Alligevel vil jeg primært kommentere på de overordnede tendenser med henblik på at uddybe forståelsen for resultaterne af den blokvisе sammenligning.

Den blokvisе udvikling i elitepolarisering viste en stigende grad af polarisering fra 1991-2012, mens polariseringen herefter falder. Det er derfor interessant at se, hvilke partier der har en lignende tendens i disse perioder. I den sammenhæng er en af de første ting, som bemærkes, at den sorte regressionslinje for mange røde partier ikke indikerer en stigende tendens i elitepolarisering. Både Socialdemokratiet og Enhedslisten har en meget flad udviklingstendens, mens udviklingen i polarisering er henholdsvis svagt faldende og stigende for Radikale Venstre og Socialistisk Folkeparti. En væsentligt pointe er dog, at polariseringen omkring 2010-2012 i større og mindre omfang viser en stigning i polarisering for alle de røde partier, hvilket også var tilfældet for polariseringen i den blokvisе sammenligning. Imidlertid er polariseringsniveauet for Socialdemokratiet, Radikale Venstre og Socialistisk Folkeparti ikke markant højere end polariseringen i tidligere dele af den undersøgte periode. Det står i kontrast til den blokvisе sammenligning. Det er alene polariseringsniveauet for Enhedslisten, der tydeligt er på sit højeste omkring 2010-2012. For alle de røde partier, med undtagelse af Socialistisk Folkeparti, sker der umiddelbart et fald i polariseringen efter 2012, hvilket er i overensstemmelse med den blokvisе sammenligning.

For de blå partier er fortællingen en anden. Både Venstre og Konservative har en klart stigende tendens i polariseringen af sprogbrug, mens Dansk Folkeparti og Liberale Alliance oplever en faldende polarisering. En interessant observation er, at elitepolariseringen for Venstre er særligt høj omkring 2001-2012, hvor partiet sad ved regeringsmagten i store dele af perioden. Først med Anders Fogh Rasmussen i spidsen fra 2001-2009 og derefter Lars Løkke Rasmussen fra 2009-2011. Denne periode var blandt andet karakteriseret ved den såkaldte "kontraktpolitik" og en generel professionalisering af den politiske kommunikation (Green-Pedersen og Kosiara-Pedersen, [2020](#); Kurrild-Klitgaard, [2011](#)). Både kontraktpolitik og professionaliseringen af politisk kommunikation er i et amerikansk stu-

die blevet fremhævet som to fænomener, der har et tidsligt sammenfald med stigninger i elitepolariseringen i sprogbrug i USA (Gentzkow et al., 2019).

En anden interessant tendens er den faldende elitepolarisering mellem Dansk Folkeparti og de øvrige partier i Folketinget. Faldet indikerer, at Dansk Folkepartis sprogbrug ligner de øvrige partier i Folketingets sprogbrug mere og mere over tid. Ændringen er samlet set knap 9 procentpoint fra 1995 til 2019, og faldet fra 2012 til 2019 er omkring 4-5 procentpoint. Sammenlignet med ændringerne i den blokwise sammenligning, er der tale om et relativt stort fald i forskellene mellem Dansk Folkeparti og resten af Folketingets sprogbrug. En overvejelse er i den sammenhæng, om udviklingen i et vist omfang afspejler en såkaldt *mainstreaming* af Dansk Folkepartis politiske positioner og sprogbrug. Her skal *mainstreaming* forstås som en proces, hvor et partis diskurser og holdninger skifter fra marginale positioner på det politiske spektrum til mere centrale positioner og dermed ændrer, hvad der udgør legitime politiske synspunkter (Brown et al., 2021). Særligt på det udlændingepolitiske område er der blevet argumenteret for, at Dansk Folkepartis positioner er blevet mere *mainstream*, og partiet er på mange punkter blevet mere ens med et parti som Socialdemokratiet (Christiansen, 2016, 2017; Kosiara-Pedersen, 2020). Derfor er det også interessant, at netop Socialdemokratiet i den samme periode også oplever en faldende polarisering i forhold til resten af Folketinget. *Mainstreamingen* af Dansk Folkeparti er således potentielt en mulig forklaring på det generelle fald i den blokwise udvikling i elitepolarisering.

Det er vigtigt at understrege, at de ovenstående overvejelser om hvad der påvirker udviklingen i elitepolarisering alene er nogle af de mange mulige forklaringer på det billede, som tegner sig på baggrund af den blok- og partivise udvikling i polarisering. Jeg vil samtidig understrege, at formålet med specialet ikke er at forklare udviklingen i polariseringen i sprogbrug, da et sådan formål vil have forudsat et andet undersøgelsesdesign. Derfor skal ovenstående overvejelser i større grad ses som mulige retninger, som fremtidig forskning med fordel kan fokuseres på for at forstå, hvad der påvirker elitepolarisering i sprogbrug. Jeg kommer nærmere ind på overvejelserne om, hvad der driver polariseringen i sprogbrug i min diskussion af de bredere implikationer af mine resultater. Først og fremmest er det imidlertid nødvendigt at håndtere en af de mest nærliggende forklaringer på de observerede polariseringstendenser. En indlysende udfordring er nemlig, at forskellene i sprogbrug blot skyldes, at politikere i forskellige partier taler om forskellige emner frem

for at tale forskelligt om de samme emner. Heldigvis kan dette håndteres ved at undersøge den emnespecifikke udvikling i elitepolarisering, hvilket sker i næste afsnit om værdi- og fordelingshypotesen.

5.2 Værdi- og fordelingshypotesen

Den andel del af analysen fokuserer på udviklingen i elitepolarisering i sprogbrug inden for værdi- og fordelingspolitiske emner. Formålet er at undersøge, om polariseringen over tid er forskellig på tværs af politiske emner. Mere specifikt er min forventning:

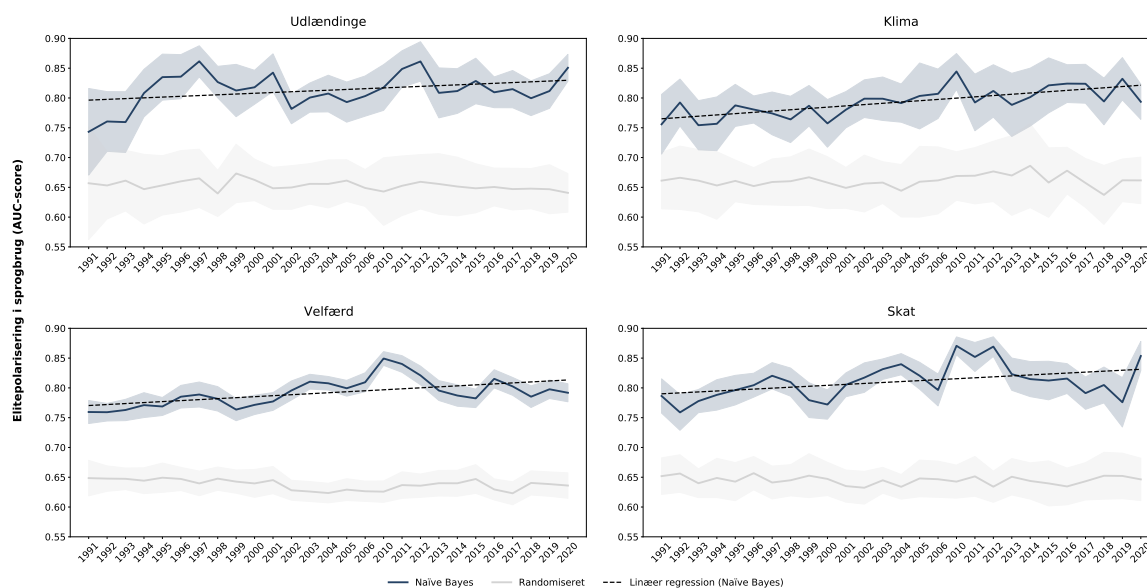
elitepolariseringen i sprogbrug er steget mere inden for værdipolitiske emner end fordelingspolitiske emner igennem de seneste årtier

Ligesom med den generelle hypotese tester jeg værdi- og fordelingshypotesen todelte. Først gennemgår jeg den blokvis udvikling i polarisering (hovedresultaterne). Dernæst gennemgår den partivise udvikling for at nuancere udviklingen.

5.2.1 Den blokvis udvikling

Figur 5.3 viser den blokvis udvikling i elitepolarisering i sprogbrug for de udvalgte værdi- og fordelingspolitiske emner. I lighed med figuren, der viser den blokvis elitepolarisering på tværs af alle taler, indikerer den blå linje udviklingen i elitepolarisering, mens den grå linje afspejler placeboscenariet. Den sorte regressionslinje er igen tilføjet for at give et billede af den overordnede polariseringstendens. Det fremgår af figuren, at der for hovedparten af den undersøgte periode ikke er et overlap mellem konfidensintervallerne for den faktiske udvikling i polarisering og polariseringen i placeboscenariet. Overordnet set er der således ikke grund til at tro, at polariseringen på de værdi- og fordelingspolitiske områder skyldes tilfældige mønstre i data. Der er dog enkelte år, hvor polariseringen på det udlændinge- og klimapolitiske område ikke er signifikant forskellig fra placeboscenariet. Det betyder, at jeg ikke kan afvise, at polariseringen i disse år blot skyldes et tilfældigt mønster i folketalingstalerne. Det gør sig dog alene gældende for en meget lille del af perioden og udgør ikke et større problem for validiteten af de samlede resultater.

Figur 5.3: Udviklingen i elitepolarisering i sprogbrug mellem blokke på tværs af værdi- og fordelingspolitiske emner



Note: De blå og grå bånd viser et 95 procent konfidensinterval. Intervallet er baseret på bootstrapping, hvor der for hvert år er simuleret 50 datasæt. Tilgangen er beskrevet nærmere i afsnit 4.7.

Overordnet set indikerer den sorte regressionslinje, at der er en stigende tendens i elitepolariseringen af sprogbrug i perioden 1991-2020 på tværs af værdi- og fordelingspolitiske emner. Et væsentligt spørgsmål er dog, om de årlige polariseringsniveauer er signifikant forskellige fra hinanden over tid. På de udlændinge- og klimapolitiske områder er polariseringen i 2020 henholdsvis ca. 11 og 4 procentpoint højere end polariseringen i 1991. Det skal dog bemærkes, at polariseringen på det klimapolitiske område i 2020 ikke er signifikant forskellig fra polariseringen i 1991, og på trods af den generelt stigende polariseringstendens er ændringen ikke signifikant. Det samme gør sig overordnet set gældende for det udlændingepolitiske område, hvor det alene er polariseringsniveauerne i 1997, 2012 og 2020, der er signifikant forskellige fra polariseringen i starten af 1990'erne og 2000'erne. Polariseringsniveauerne i de pågældende år er ikke signifikant forskellige fra hinanden, og der er altså ikke klare indikationer på en generel stigning i elitepolariseringen af sprogbrug over tid på det udlændingepolitiske område. En interessant observation er dog, at vi ligesom med den generelle polariseringsudvikling ser en tendens til, at polariseringen på det udlændingepolitiske område stiger fra 2001, toppe omkring 2012 og derefter falder. Det tyder igen på, at denne periode af dansk politik har haft et særligt polariseret sprogbrug. Polariseringen i 2012 er dog ikke signifikant forskellig fra niveauerne i 1997 og 2020.

Det fremgår endvidere af figuren, at polariseringen på de velfærds- og skattepolitiske områder er henholdsvis ca. 3 og 6 procentpoint højere i 2020 sammenlignet med 1991. Polariseringen på det velfærdspolitiske område i 2020 er ikke signifikant forskellig fra polariseringen i 1991. I forlængelse heraf ses det, at niveauerne for elitepolarisering i 1990'erne ikke er signifikant forskellig fra elitepolariseringen fra 2013-2020 for både det velfærds- og skattepolitiske område. Det indikerer tildeles en stagnerende tendens for polariseringen på de fordelingspolitiske områder. Det er dog væsentligt at bemærke, at mønsteret ikke gælder hele perioden. Ligesom med analysen af den generelle udvikling i elitepolarisering ses der en tendens til, at polariseringen stiger i perioden fra 2001 til 2012-13. Polariseringsniveauerne for denne periode er signifikant forskellige fra perioderne før og efter, og det gør sig både gældende for de velfærds- og skattepolitiske taler. De største forskelle i denne periode er omkring 9-11 procentpoint, og ændringerne i polarisering er sammenlignelige med ændringerne for den generelle udvikling i polarisering på tværs af alle politikområder. Udviklingen viser igen et tvetydigt resultat, hvor forskellene i rød og blå bloks sprogbrug varierer på tværs af perioden fra 1991-2020 og topper omkring 2010-2012.

Samlet set findes der ikke støtte til værdi- og fordelingshypotesen. Polariseringen på de værdipolitiske områder har en stigende tendens, men udviklingen i elitepolariseringen af sprogbrug er i store træk insignifikant med undtagelse af enkelte observationer på det udlændingepolitiske område. Samtidig er elitepolariseringen på de fordelingspolitiske områder også svagt stigende over perioden, men det er alene for perioden 2010-2012, at der observeres større stigninger, som også er substantielt forskellige for de øvrige dele af perioden. Modsat forventningen fra værdi- og fordelingshypotesen ser det ud til, at elitepolariseringen i sprogbrug ikke er steget mest på det værdipolitiske område. Faktisk set det ud til, at særligt de høje polariseringsniveauer omkring 2010-2012 går igen på tværs af politikområderne. Det indikerer, at elitepolariseringen i sprogbrug ikke kun er bestemt af de politiske emner, der diskuteres, men derimod hænger sammen med faktorer, der går på tværs af emner. På samme tid hjælper resultaterne også med at validere resultaterne fra testen af den generelle hypotese. Det skal ses i lyset af, at analysen også viser signifikante stigninger i polariseringen i sprogbrug, når de debatterede emner holdes fast.

En ting er dog, om der er polarisering i sprogbrug, og om politikerne taler forskelligt. Noget andet er, hvordan de reelt taler, og hvilke ord de bruger inden for de forskellige

emner. For at belyse dette nærmere, har jeg i tabel 5.1 lavet en oversigt, der viser de 15 mest frekvente ord i de taler, som min model prædikerer som henholdsvis røde og blå på tværs af perioden fra 1991 til 2020. Det fremgår af tabellen, at der er flere ord inden for de forskellige værdi- og fordelingspolitiske emner, der er i overensstemmelse med nogle af de politiske positioner og kommunikative strategier, som forbindes med blå og røde partier. På det skattepolitiske område ses det for eksempel, at blå partier ofte bruger ord som ”virksomhed” og ”skattestop”, mens ordet ”offentlig” ligger højt på listen for røde partier. Lignende forskelle viser sig på det udlændingepolitiske område, hvor de blå partier taler om ”grænsekontrol” og ”indfødsret”, mens røde partier oftere bruger ord som ”børn” og ”integration”. Det er værd at bemærke, at mange af de nævnte ord alene befinder sig i top-15 for den ene blok, men ikke for den anden blok. Det understreger forskellene i ordvalg mellem partierne i de respektive blokke. Det faktum, at forskellene i ordvalg i et vist omfang spejler kendte forskelle i politiske positioner styrker samtidig det maskinlæringsbaserede måls validitet.

Tabel 5.1: De mest frekvente ”røde” og ”blå” ord

Rang	Velfærd		Skat		Udlændinge		Klima	
	<i>Rød</i>	<i>Blå</i>	<i>Rød</i>	<i>Blå</i>	<i>Rød</i>	<i>Blå</i>	<i>Rød</i>	<i>Blå</i>
1	Arbejde	Kommune	Skat	Skat	Dansk	Dansk	Grøn	Miljø
2	Kommune	Give	Afgift	Skattelettelse	Opholds- tilladelse	Danmark	Energi	Energi
3	Arbejdsløs	Arbejde	Offentlig	Penge	Danmark	Udlænding	Vedvarende	Afgift
4	Menneske	Offentlig	Betale	Betale	Person	Statsborger- skab	Udvikling	Vindmøller
5	Børn	Børn	Kommune	Virksomhed	Udlændinge	Land	Problem	Kommune
6	Penge	Ældre	Regler	Penge	Udlændingelov	Opholds- tilladelse	Miljø	Virksomheder
7	Arbejds- marked	Forældre	Ændring	Skattestop	Land	Give	Omstilling	Udvikling
8	Ret	Person	Arbejde	Arbejde	Asylansøger	Person	Bæredygtig	Grøn
9	Folk	Regler	Beskatning	Ændring	Flygtning	Grænsekontrol	Klima	Energiaftale
10	Problem	Enkelte	Give	Høj	Menneske	Indvandrere	Penge	Mål
11	Uddannelse	Arbejds- marked	Fradrag	Erhvervsliv	Børn	Udlændinge- politik	Økonomisk	Regler
12	Social	Ret	Indkomst	Gæld	Asyl	Udlændingelov	Fremtid	Naturgas
13	Dagpenge	Uddannelse	Gæld	Regler	Arbejde	Flygtning	Arbejdsplads	Land
14	Offentlig	Gæld	Arbejdsplads	Bruger- betaling	Integration	Indfødsret	Landbrug	Begrund
15	Økonomisk	Virksomhed	Selskab	Skattetryk	Politi	Grænse	Investering	Høj

5.2.2 Den partivise udvikling

Den partivise analyse af udviklingen i elitepolarisering af sprogbrug for værdi- og fordelingspolitiske emner har igen til formål at nuancere billedet fra den blokvisе sammenligning. Som tidligere nævnt er der ikke konstrueret bootstrap-konfidensintervaller for den partivise udvikling. Det betyder, at jeg ikke med sikkerhed kan vide, at udviklingen i polarisering er signifikant forskellig fra placeboscenariet. Derfor vil jeg alene se på de overordnede tendenser. Resultaterne for den partivise udvikling fremgår af figur 5.4 nedenfor.

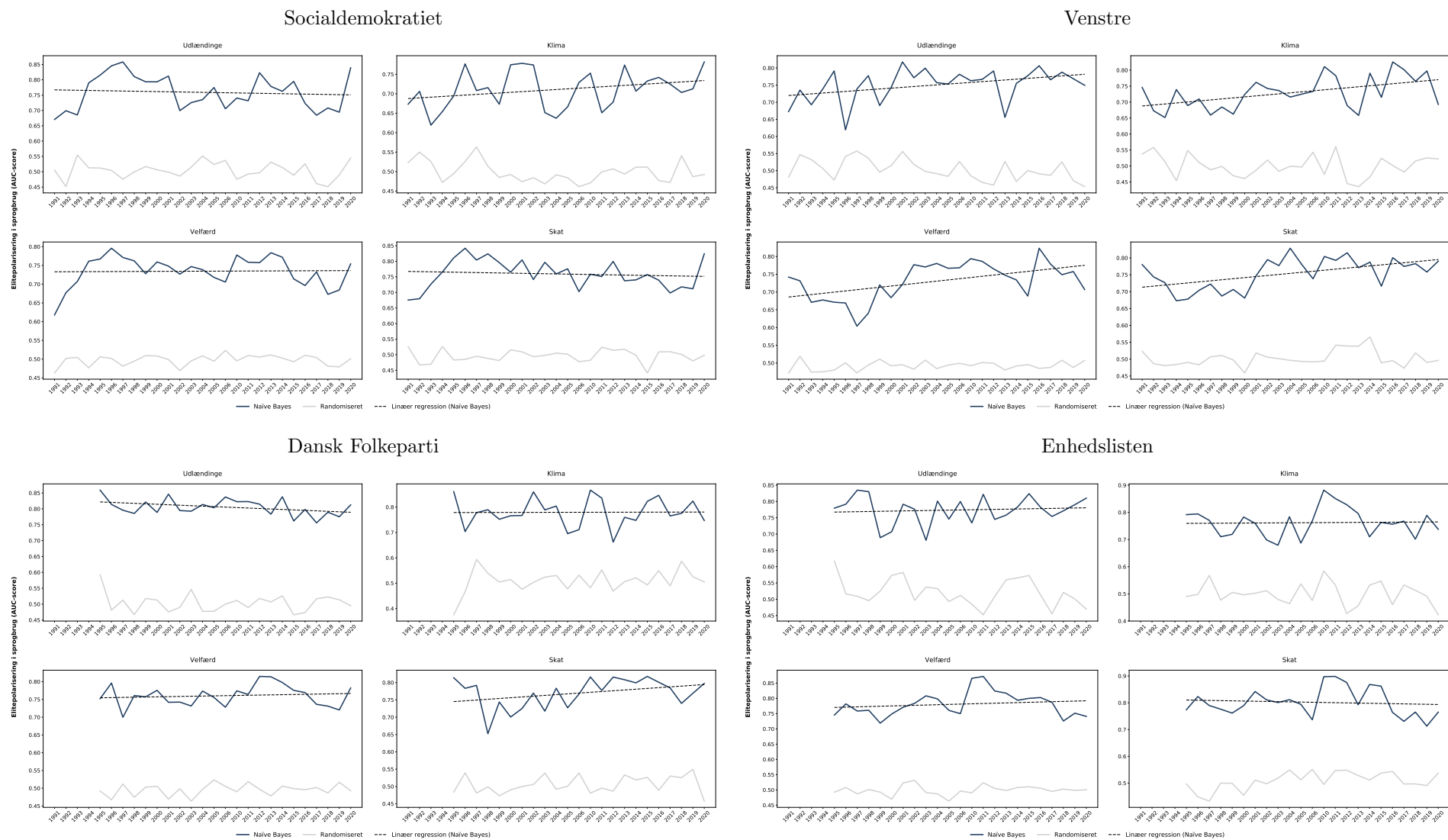
Det overordnede billede for den partivise udvikling er, at der for størstedelen af partierne ikke ses en generel stigning i polariseringen over tid. Hverken på det værdi- eller fordelingspolitiske område. Der kan i stedet observeres en forholdsvis stor variation i elitepolariseringen over tid, og der synes ikke at være et klart mønster i udviklingen på tværs af partierne. Samtidig observeres der heller ikke et klart mønster, som peger på, at polariseringen i sprogbrug er steget mere på det værdipolitiske område sammenlignet med det fordelingspolitiske område. Den partivise sammenligning giver således heller ikke støtte til værdi- og fordelingshypotesen.

Det eneste parti, som har haft en klart stigende polarisering i sprogbrug over for resten af Folketinget, er Venstre. Stigningen i polarisering gælder på tværs af alle politikområder, og der ser altså ikke ud til at være en klar forskel i den overordnede polariseringstendens, når polariseringen isoleres til bestemte værdi- eller fordelingspolitiske emner. Ligesom med den generelle hypotese ser det ud til, at polariseringen i sprogbrug for Venstre særligt stiger omkring 2001-2011, hvor de havde regeringsmagten. Det tyder på, at stigningen i elitepolarisering i denne periode for Venstre ikke blot skyldes, at partiet taler om andre emner end de øvrige folketingspartier. Derimod tyder det faktisk på, at der er sket en drejning i Venstres sprogbrug, der styrker maskinlæringsmodellens evne til at klassificere Venstres taler. Det peger således fortsat på, at fremtidig forskning med fordel kan se nærmere på betydningen af kontraktpolitik og professionaliseringen af kommunikation for at forklare elitepolariseringen i sprogbrug.

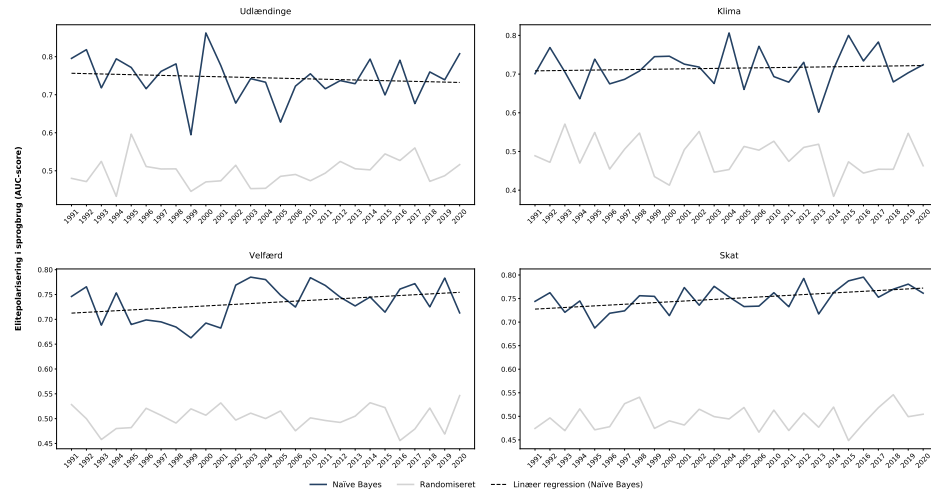
En udvikling, som jeg også kommenterede på i testen af den generelle hypotese, er den faldende elitepolarisering i sprogbrug for Dansk Folkeparti. Interessant nok tyder resultaterne fra den værdi- og fordelingspolitiske analyse på, at det observerede fald i elitepolariseringen mellem Dansk Folkeparti og de øvrige partier ikke gælder for alle politikområder.

Reelt set viser figur 5.4 kun en faldende tendens for de værdipolitiske områder, herunder særligt det udlændingepolitiske område. Det skal dog bemærkes, at ændringen fra 1991 til 2020 kun er knap 5 procentpoint på det udlændingepolitiske område. Dermed er faldet ikke særlig stort, når man sammenholder det med forskellene, der blev observeret på blokniveau. Derudover har jeg som nævnt heller ikke lavet bootstrap-konfidensintervaller for den partivise udvikling, og jeg kan derfor ikke afgøre om udviklingen er signifikant. Alligevel kan de modsatrettede polariseringstendenser på tværs af politikområder indikere, at den faldende elitepolarisering, som jeg observerede i testen af den generelle hypotese, er delvist drevet af, at Dansk Folkeparti taler om bestemte emner frem for andre. Her tænker jeg særligt på det udlændingepolitiske område. Det peger desuden på, at idéen om, at der er sket en mainstreaming af Dansk Folkepartis politiske positioner, forsat ser ud til at gælde Dansk Folkepartis sprogbrug, som er blevet mindre forskelligt fra resten af Folketinget over tid.

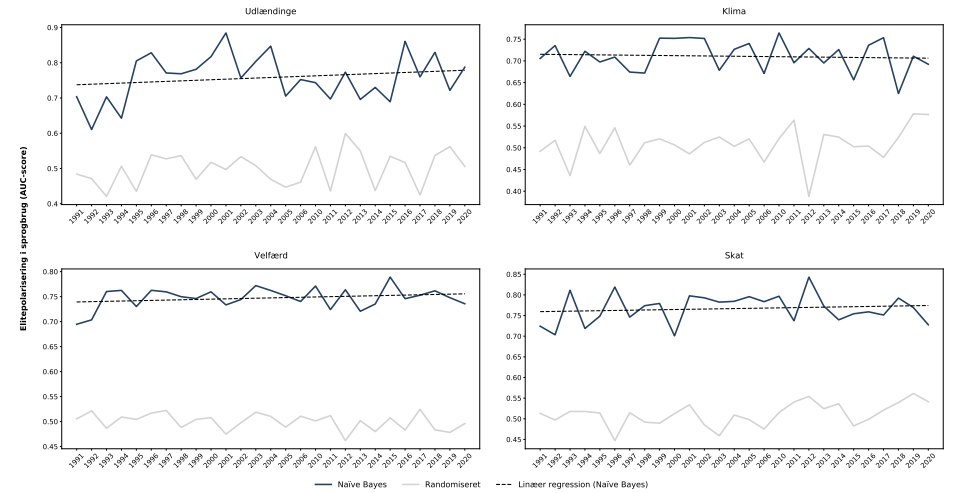
Figur 5.4: Udviklingen i elitepolarisering i sprogbrug mellem enkelte partier og resten af Folketinget på tværs af værdi- og fordelingspolitiske emner



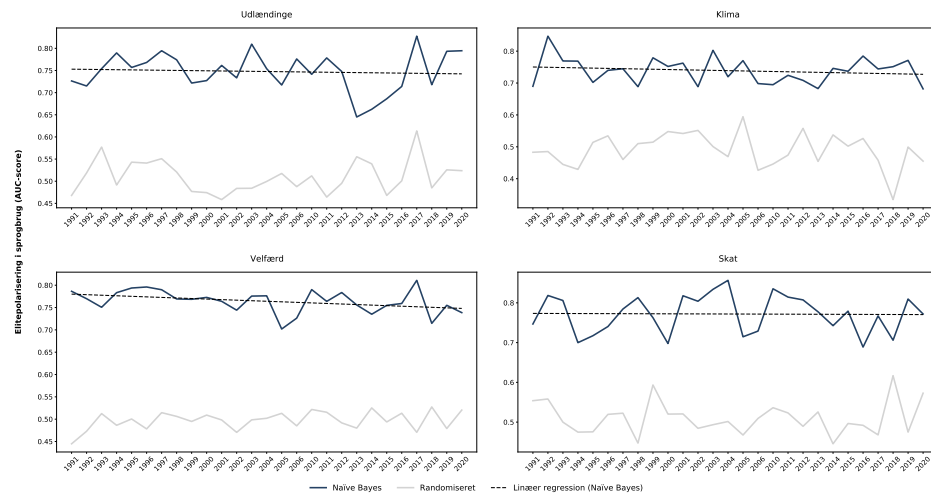
Det Konservative Folkeparti



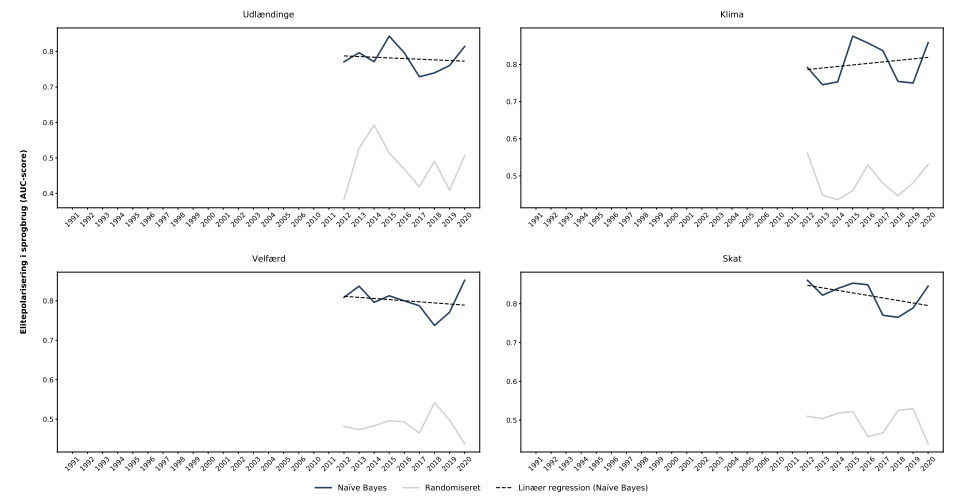
Socialistisk Folkeparti



Radikale Venstre



Liberal Alliance



5.3 Forskelshypotesen

Den sidste del af analysen fokuserer på forskellene mellem elitepolarisering i sprogbrug og elitepolariseringen af parlamentariske afstemninger. Formålet er at undersøge, om der udover en konceptuel forskel også er en empirisk forskel mellem de to dimensioner af elitepolarisering. Min forventning er, at:

elitepolariseringen i sprogbrug er steget mere end elitepolariseringen af parlamentariske afstemninger igennem de seneste årtier

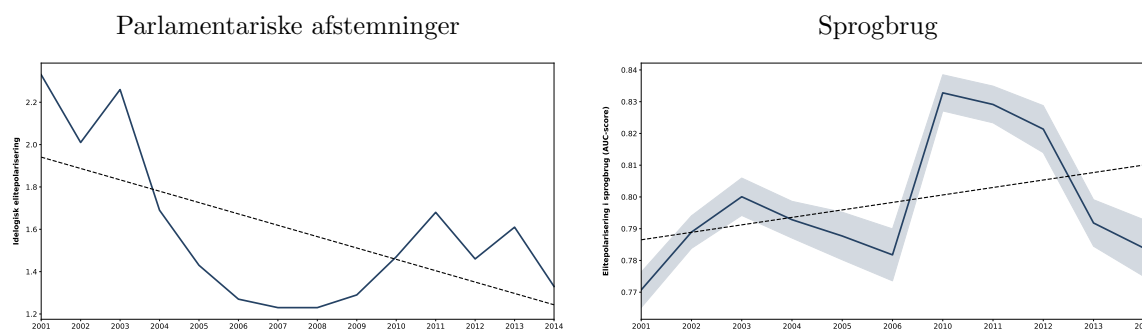
I modsætning til de andre hypoteser tester jeg kun forskelshypotesen på blokniveau. Jeg sammenholder polariseringen i sprogbrug med polariseringen af afstemninger. Det gør jeg ved (1) en visuel sammenligning af de to udviklingstendenser og (2) en simpel korrelation mellem de to mål.

Figur 5.5 viser udviklingen i elitepolarisering i sprogbrug og parlamentariske afstemninger for perioden 2001 til 2014. Som for de øvrige figurer viser den blå linje udviklingen i de anvendte mål, mens den sorte linje angiver en regression mellem elitepolariseringsmålet og tid. Overordnet set viser figuren, at elitepolariseringen af parlamentariske afstemninger har været faldende over perioden, mens elitepolariseringen i sprogbrug har en stigende tendens. Dette billede er i overensstemmelse med forventningerne i forskelshypotesen. Det tyder på, at forskellene mellem danske partiers sprogbrug samlet set er blevet større i perioden 2001-2014, mens uenigheden i lovfølgelse er blevet mindre. Her er det vigtigt at understrege, at der er tale om et overordnet billede, som tager udgangspunkt i tendensen for perioden fra 2001 til 2014.

En særligt interessant udvikling i figur 5.5 er, at der observeres en stigning i polariseringen af både sprogbrug og afstemninger fra 2006 til 2010-2012. Sammenlignet med polariseringsniveauer for den øvrige del af perioden er stigningen i polariseringen af afstemninger dog mindre end stigningen i polariseringen af sprogbrug. For eksempel er ændringen i polariseringen af afstemninger fra 2006 til 2010 på 0,2 indekspoint på en skala fra 1-10, mens den tilsvarende ændring i sprogbrug er på ca. 7 procentpoint. Det observeres også, at polariseringen af sprogbrug toppe i 2010, mens polarisering af afstemninger allerede toppe i 2001, hvor polariseringen i sprogbrug er på sit laveste. Samlet set giver billedet fra figur 5.5 støtte til forskelshypotesen. Der ser dog ud til at være visse

sammenfald for, hvornår polariseringsniveauerne falder og stiger, hvorfor jeg i næste afsnit ser nærmere på korrelationen mellem de to mål.

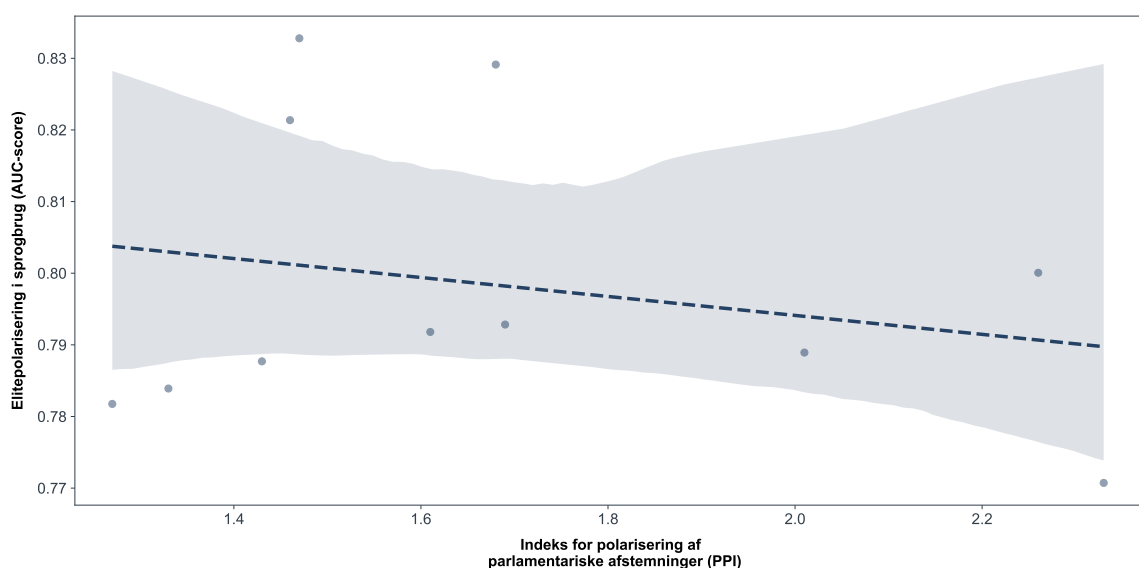
Figur 5.5: Udviklingen i elitepolarisering af sprogbrug og parlamentariske afstemninger



Note: Det blå bånd viser et 95 procent konfidensinterval. Der er et databrud for folketingstaler i 2007 og 2008. Jeg er dog interesseret i den overordnede udvikling, og det vurderes derfor ikke at udgøre et problem for at sammenholde tendenserne i figuren.

Figur 5.6 viser sammenhængen mellem elitepolarisering i sprogbrug og elitepolarisering af parlamentariske afstemninger. Det fremgår af figuren, at der ikke er en signifikant sammenhæng mellem de to dimensioner af elitepolarisering ($r = -0.23, p < 0.49$). Det indikerer, at de to dimensioner af elitepolarisering er forskellige fra hinanden, hvilket styrker mit argument om, at elitepolarisering i sprogbrug skal betragtes som en særskilt dimension af elitepolarisering. I substantielle termer betyder det, at der ikke er en klar relation mellem udviklingen i forskellene mellem de politiske blokkes sprogbrug og uenigheden mellem partierne i lovførelser. Med andre ord kan partierne i de politiske blokke begynde at tale mere forskelligt uden, at det er kædet sammen med større politisk uenighed i afstemninger. Dette er en grundlæggende forudsætning for forskelshypotesen, der netop indebærer, at der har været forskellige udviklinger inden for de to dimensioner af elitepolarisering. Som nævnt er korrelationen baseret på få observationer, hvilket medvirker til de store konfidensintervaller, og dermed skaber usikkerhed om resultaterne. Jeg vil dog argumentere, for at korrelationen i kombinationen med de observerede udviklinger i figur 5.5 giver tilstrækkelig støtte til hypotesen. Det skal ses i lyset af de modsatrettede tendenser, hvor polariseringen af afstemninger falder, mens polariseringen i sprogbrug stiger.

Figur 5.6: Korrelation mellem elitepolarisering i sprogbrug og parlamentariske afstemninger



Note: $N = 11$. Det er alene punkterne, hvor der er data for det samme år, som er korreleret for at sikre sammenligneligheden af datapunkterne.

6 Konklusion

Polarisering er blevet et ord, som oftere og oftere bruges til at karakterisere det politiske landskab i Danmark. Alligevel er det af flere omgange også blevet fremhævet, at dansk politik er konsensuspræget sammenlignet med lande som USA. Det rejser spørgsmål om, hvad der faktisk er tilfældet. Er dansk politik blevet mere eller mindre polariseret? Og varierer polariseringen på tværs af forskellige dimensioner af elitepolarisering? Det er nogle af de spørgsmål, som jeg har forsøgt at besvare med dette speciale, hvor jeg har undersøgt udviklingen i polariseringen af danske politikeres sprogbrug og parlamentariske afstemningsadfærd. Det primære formål med specialet har været at kaste mere lys på udviklingen i elitepolarisering i Danmark, som kun er blevet undersøgt i et begrænset omfang.

For at undersøge udviklingen i elitepolariseringen af sprogbrug, har jeg udført et deskriptivt casestudie af folketingstaler for perioden 1990 til 2020. Min undersøgelse er baseret på et nyt mål for elitepolarisering, som udnytter superviseret maskinlæring til at finde forskelle i politikeres sprogbrug. Elitepolariseringen i sprogbrug er både undersøgt på blok- og partiniveau samt på tværs af værdi- og fordelingspolitiske emner. Derudover har jeg argumenteret for, at der er en konceptuel forskel mellem elitepolarisering i sprogbrug

og elitepolariseringen af parlamentariske afstemninger. Derfor har jeg også sammenstillet udviklingen i elitepolariseringen af sprogbrug og afstemninger med henblik på at afgøre, om der også findes empirisk støtte til dette argument.

Mine resultater viser, at der for den samlede periode fra 1990 til 2020 er en tendens til stigende polarisering i danske politikeres sprogbrug. Stigningen er dog ikke konsistent, da der er væsentlige variationer i polariseringen over tid. Desuden er polariseringen i sprogbrug i dag ikke højere end polariseringen i midten af 1990'erne, men dog på et højere niveau end starten af 1990'erne. Til gengæld viser mine resultater entydigt, at polariseringen i sprogbrug steg stødt under 2000'ernes VK-regeringer med den højeste polarisering omkring 2011-2012. Når der sættes fokus på folketalingstaler og polariseringen af politikernes sprogbrug, er det altså ikke klart, at Danmark er et konsensusland, hvor uenigheden er lille og ensartet over tid.

Jeg finder derudover mod forventning, at der ikke er store forskelle mellem elitepolariseringen i sprogbrug på tværs af værdi- og fordelingspolitiske emner. Det står i kontrast til resultaterne af et andet dansk studie, som fremhæver, at elitepolariseringen af afstemningsadfærd er steget på det værdipolitiske område, mens udviklingen på det fordelingspolitiske område er stabilt/svagt faldende (Christiansen, 2019b). Det tyder på, at forskellene mellem partiernes sprogbrug ikke nødvendigvis følger de samme dynamikker, som gør sig gældende for uenigheden i lovaftemninger. Det indikerer samtidig, at der potentielt er faktorer, som har en betydning for polariseringen af sprogbrug på tværs af politiske emner.

Netop de empiriske forskelle mellem elitepolariseringen af sprogbrug og elitepolariseringen af afstemningsadfærd udgør et af de centrale fund i specialet. Mine resultater viser, at elitepolariseringen i sprogbrug stiger i perioden fra 2001-2014, mens polariseringen af afstemninger falder i samme periode. Samtidig viser min validering af målet for elitepolarisering i sprogbrug også, at målet er forskelligt fra ideologisk elitepolarisering. Det giver empirisk støtte til mit argument om, at sprogbrug bedst forstås som en særskilt dimension af elitepolarisering, der muligvis både påvirker og er påvirket af andre faktorer end polariseringen i afstemninger og ideologi. En implikation heraf er, at fremtidig forskning med fordel kan skelne mellem de forskellige typer af elitepolarisering for at få en bedre forståelse af elitepolariseringens betydning.

Samlet set har jeg med mit speciale demonstreret, at billedet af Danmark som et

konsensusland uden elitepolarisering i stor grad afhænger af hvilken dimension af elitepolarisering, der sættes fokus på. Udviklingen i elitepolariseringen af sprogbrug varierer over tid, og det er således ikke helt uforståeligt, at medierne til tider taler om en stigende polarisering mellem danske politikere. Selvom dette i sig selv kan være interessant, så kan der på baggrund af resultaterne også rejses en række andre spørgsmål. For hvor kommer den sproglige polarisering egentlig fra? Og er det et problem, at der observeres forskelle i politikernes sprogbrug, som ikke nødvendigvis omsættes til forskelle i politikernes afstemningsadfærd? Besvarelsen af disse spørgsmål går ud over formålet med min undersøgelse, men jeg vil alligevel runde specialet af med en diskussion heraf, da det kan hjælpe med at sætte en retning for fremtidig forskning.

7 Diskussion

Resultaterne af analysen har rejst en række spørgsmål, som jeg diskuterer i dette afsnit. Som det første diskuterer jeg resultaternes eksterne validitet og muligheden for at generalisere dem. Dernæst vil jeg diskutere resultaternes teoretiske og metodiske implikationer, herunder den konceptuelle distinktion mellem elitepolarisering i sprogbrug og polariseringen af afstemninger. Endelig diskuterer jeg de empiriske implikationer af mine resultater med fokus på, om der er nogle problematiske konsekvenser af mine resultater. Jeg reflekterer løbende over mulige fokuspunkter for fremtidig forskning.

7.1 Elitepolarisering i sprogbrug uden for en dansk kontekst

Det danske demokrati er præget af en stor grad af konsensus, og jeg har derfor valgt at betragte undersøgelsen af elitepolariseringen i Folketinget som en least-likely case eller en kritisk case (Flyvbjerg, 2010; Lijphart, 1999). Forudsætningen for at generalisere den observerede udvikling til andre cases er således, at der er sket en stigning i elitepolariseringen over tid, idet der forventeligt også vil være sket en stigning i elitepolariseringen i lande med en mindre grad af konsensus. Mine resultater peger dog ikke entydigt på, at elitepolariseringen i Danmark er steget. Derimod viser mine resultater, at elitepolariseringen i sprogbrug har været varierende over tid, og der er perioder, hvor polariseringen er steget uden, at polariseringen i dag er på det højeste niveau i et historisk perspektiv. Eftersom resultaterne ikke klart indikerer en konsistent stigning i elitepolariseringen i

sprogbrug, vil jeg argumentere for, at det svækker muligheden for at anvende least-likely logikken til at sandsynliggøre en stigning i elitepolarisering af sprogbrug over tid uden for Danmark.

Variationen i elitepolariseringen af sprogbrug over tid, hvor der både observeres perioder med relativ høj og lav elitepolarisering, gør det svært at generalisere resultaterne for udviklingen i elitepolarisering over tid. En mulig løsning er dog at sammenholde den konkrete udvikling i Danmark med resultaterne fra øvrige lande, hvor man har målt elitepolariseringen i parlamentarisk taler. For eksempel viser et britisk studie, at elitepolariseringen i Storbritannien generelt set har været faldende fra 2001 og frem (Petersen og Roepstorff, 2020). Det er en periode, hvor mine resultater indikerer, at elitepolariseringen i sprogbrug både når at toppe (omkring 2010-2013) for derefter at falde til et niveau, som er sammenlignelig med polariseringen i 1990'erne. I kontrast hertil viser et studie af elitepolariseringen i taler fra USA en klart stigende tendens i 2010'erne, hvor polariseringen når de højeste niveauer i et historisk perspektiv (Gentzkow et al., 2019). Det samme gælder for et norsk studie, hvor stigningerne i polariseringen mellem de politiske blokke dog relativt set er væsentlig mere moderate end i USA (Fiva et al., 2021). Der er altså ikke en entydig tendens for udviklingen i elitepolarisering af sprogbrug på tværs af lande, hvilket yderligere svækker muligheden for at generalisere resultaterne af min analyse til andre lande.

Det er særligt interessant, at udviklingen i elitepolarisering i sprogbrug i Norge viser en relativt moderat, men stigende grad af sproglig polarisering over tid. Det skal ses i lyset af, at Norge og Danmark umiddelbart er sammenlignelige lande med sammenlignelige parlamentariske strukturer, statslige institutioner og kulturer. Dette kunne indikere, at elitepolariseringen i sprogbrug kan være påvirket af relativt partikulære forhold, som kan variere på tværs af ellers sammenlignelige lande. Forskellen kan dog også skyldes, at mit mål for elitepolarisering ikke én til én afspejler det mål, som undersøgelsen af elitepolarisering i parlamentariske taler i Norge bygger på (se Fiva et al. (2021)). Derfor kan de forskellige udviklinger i polariseringen skyldes metodiske forskelle. Uanset hvad der er tilfældet, peger det på relevansen af, at der laves flere studier af elitepolarisering i sprogbrug med henblik på at forstå hvilke faktorer, der kan forklare variationen i sproglig polarisering og samtidig finde frem til den mest retvisende metodiske målingstilgang.

7.2 Teoretiske og metodiske implikationer

Resultaterne af specialet og den anvendte metode giver anledning til en række overvejelser omkring undersøgelsens teoretiske og metodiske implikationer. I det følgende afsnit vil jeg først diskutere de teoretiske implikationer med fokus på den konceptuelle distinktion mellem elitepolarisering i sprogbrug og øvrige dimensioner af elitepolarisering samt behovet for at danne en bedre forståelse for de faktorer, som påvirker elitepolarisering i sprogbrug. Dernæst diskuterer jeg de metodiske implikationer, herunder særligt overvejelser om målingen af elitepolarisering i flerpartisystemer.

7.2.1 Elitepolariseringens dimensioner: Teoretiske og empiriske forskelle

Resultaterne fra forskelshypotesen viser, at der er behov for en konceptuel og teoretisk skelnen mellem elitepolarisering i sprogbrug og elitepolariseringen af afstemninger. Som fremhævet i afsnit 3.1, er der en række argumenter for, at polariseringen af afstemninger og polariseringen i sprogbrug bedst forstås som to forskellige dimensioner af elitepolarisering. Det blev yderligere sandsynliggjort i analysen, hvor en sammenligning af den sproglige og afstemningsbaserede polarisering viser, at de to dimensioner er undergået forskellige udviklinger. En mulig implikation heraf er, at de to dimensioner af elitepolarisering er drevet af forskellige faktorer. Udover den konceptuelle skelnen kan det derfor være meningsfuldt med en teoretisk skelnen mellem polarisering af afstemninger og sproglig polarisering, da det kan bidrage til at udvikle teoretiske forklaringer på, hvilke forskellige faktorer, der påvirker hver af de to dimensioner af elitepolarisering.

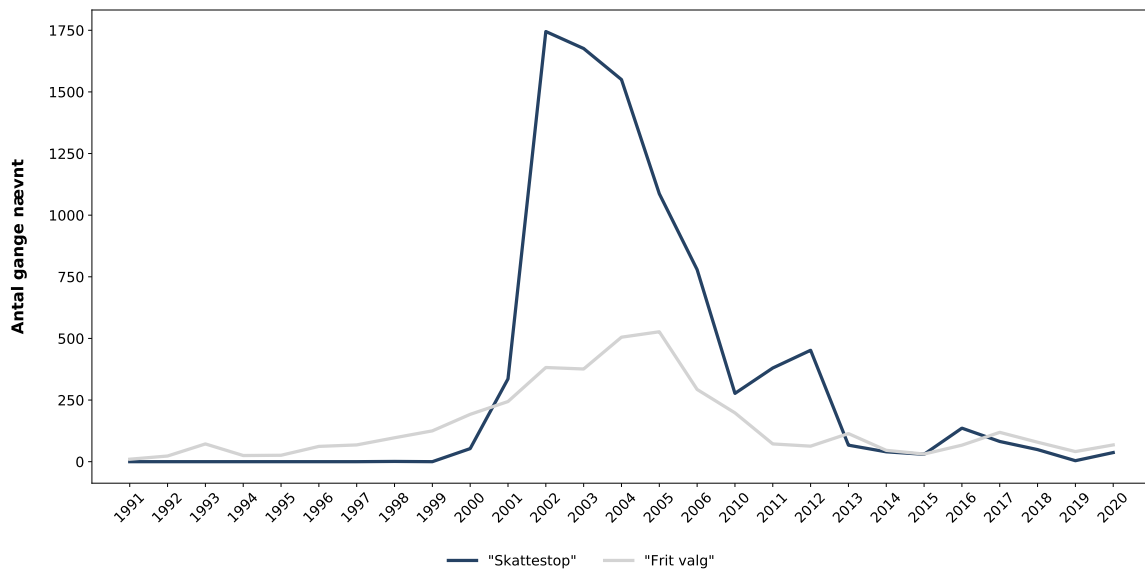
I to lignende studier af elitepolarisering i parlamentariske taler fra Storbritannien argumenterer forfatterne for, at deres talebaserede mål for elitepolarisering måler ideologiske forskelle (Goet, 2019; Peterson og Spirling, 2018). Deres argument bygger på en antagelse om, at politikeres sprogbrug og tale primært bygger på politiske og ideologiske overvejelser (Goet, 2019). Forskellene mellem et afstemningsbaseret mål for ideologisk polarisering og et talebaserede mål for elitepolarisering skyldes derfor ikke, at man måler to forskellige ting. Det skyldes derimod, at afstemninger kan være påvirkede af institutionelle forhold såsom stærk partidisciplin, og disse forhold påvirker i mindre grad parlamentariske taler, hvorfor der observeres forskellige udviklinger (Peterson og Spirling, 2018). Jeg vil dog argumentere for, at denne forståelse af talebaserede mål for elitepolarisering indebærer en meget bred forståelse af, hvad der konstituerer ideologi som begreb. Forståelsen ko-

ger ideologi ned til at være defineret ved forskelle i politikeres sprogbrug. Disse forskelle kan dog i stor grad afspejle ikke-ideologiske faktorer såsom strategiske, kommunikative og retoriske overvejelser. Derfor mener jeg på trods af tilgangene i øvrige studier, at et talebaseret mål for elitepolarisering bedst forstås som et mål for elitepolarisering i sprogbrug og ikke ideologisk elitepolarisering. Hvis man ikke tager højde for, at forskellene i politikeres taler også er styret af andet end ideologiske hensyn, så risikerer man at få et skævt billede af, hvordan ideologiske forskelle mellem politikere i virkeligheden udvikler sig. Det understøttes desuden af den diskriminerende validering i appendiks D.2, som viser, at korrelationen mellem et mål for ideologisk polarisering og målet for elitepolarisering i sprogbrug er insignifikant. Hvis polariseringen af sprogbrug reelt skulle måle det samme underliggende fænomen - ideologisk elitepolarisering - ville den klare forventning være, at der var en positiv sammenhæng mellem målene. Det er ikke tilfældet.

7.2.2 Mulige forklaring på elitepolarisering i sprogbrug

Mit studie giver også anledning til at reflektere over den teoretiske forståelse af de faktorer, som påvirker udviklingen i elitepolarisering i sprogbrug. Selvom jeg har haft et deskriptivt fokus, hvor jeg ikke har forsøgt at forklare udviklingen i elitepolarisering, er der alligevel et par mønstre, som er gået igen på tværs af mine hypoteser. Her tænker jeg på stigningen i polarisering omkring 2001 til 2012-2013, som sker i periode, hvor Danmark primært ledes af Anders Fogh Rasmussens VK-regeringer (se bilag E.1). Som tidligere fremhævet er en af de interessante udviklinger i denne periode Foghs kontrakt-politiske tilgang til politik (Kurrild-Klitgaard, 2011). Kontraktpolitikken er koblet op på en professionalisering og fokusering af politisk sprogbrug. To eksempler på, hvordan kontraktpolitikken er koblet til ændringer i sprogbrug, ses blandt andet i det skattepolitiske fokus på "skattestop" og velfærdspolitiske fokus på "frit valg" (Kurrild-Klitgaard, 2011). Det er derfor interessant at se, hvordan disse to ord ser ud til at finde deres vej til dansk politik i samme periode, som vi ser en stigning i polariseringen af sprogbrug, jævnfør figur 7.1. Kontraktpolitikken og dens indvirkning på politisk sprogbrug er blot en af mange forklaringer på, hvorfor der sker en stigning i polariseringen af politikeres sprogbrug.

Figur 7.1: Udviklingen i brugen af ”skattestop” og ”frit valg”



En nærliggende alternativ forklaring på stigningen i sproglig elitepolarisering er, at den samme periode er blevet omtalt som en periode med en høj grad af blokpolitik. Netop manglen på samarbejde mellem blokkene kan også tænkes at hænge sammen med en øget sproglig polarisering, da det tydeliggør politiske forskelle, som politikerne kan italesætte. Imidlertid viser afstemningsresultater fra Folketinget, at VK-regeringen ikke stemte mere med sin egen blok end man har gjort under tidligere regeringer (Kurrild-Klitgaard et al., 2006). Det er dog ikke utænkeligt, at det politiske fokus på fortællingen om en regering, der kun fører blokpolitik kan kædes sammen med en stigende grad af polarisering i sprogbrug. Blok- og kontraktpolitikken er to blandt mange forklaringer, som potentielt hjælper med at forstå udviklingen i sproglig polarisering. Det er ikke utænkeligt, at begge faktorer har bidraget til stigningen i elitepolariseringen af sprogbruget i perioden fra 2001 til 2012-2013. En mulighed er også, at sammenfaldet mellem blok- og kontraktpolitikken samt stigningen i elitepolarisering i sprogbrug skyldes, at både blok- og kontraktpolitik indebærer retoriske overvejelser, som bidrager til at øge forskellene i politikeres sprogbrug. Uanset hvad kalder den manglende forståelse af, hvad der påvirker elitepolarisering i sprogbrug på yderligere forskning, der kan belyse hvilke faktorer, der hænger sammen med fænomenet. Og her vil jeg med afsæt i udviklingen fra figur 7.1 argumentere for, at særligt kontraktpolitikken ser ud til at være et nyttigt sted at starte.

7.2.3 Måling af elitepolarisering af sprogbrug i flerpartisystemer

Den metodiske tilgang i mit speciale giver anledning til refleksioner om, hvordan man måler elitepolarisering i en flerpartiskontekst. Jeg har tidligere fremhævet, at den nuværende litteratur om elitepolarisering primært er amerikansk. En konsekvens heraf er, at det teoretiske og metodiske apparat omkring polarisering oftest er tilpasset et topartisystem, hvor det er forholdsvist simpelt at undersøge polariseringen mellem de to poler. Det er imidlertid mere kompliceret i et flerpartisystem, da der er mange partier og mange mulige poler. Det rejser en række spørgsmål. Er det for eksempel tilstrækkeligt, at der er polarisering mellem to partier? Eller skal der være polarisering mellem flere partier på én gang? Det er langt fra åbenlyst, hvornår den parlamentariske situation er karakteriseret ved polarisering, og det er samtidig også udfordrende at måle polariseringen.

I dette speciale har jeg primært løst problematikken omkring den danske flerpartikonstekst ved at forsimple det politiske system til en rød og en blå blok. Denne tilgang er også anvendt i den øvrige europæiske litteratur om elitepolarisering (Kurrild-Klitgaard et al., 2006; Peterson og Spirling, 2018). Fordelen ved denne tilgang er, at det løser nogle af de metodiske udfordringer, som der eksisterer, når man forsøger at sammenligne mange partier. Ulempen er derimod, at det er tale om en forsimpning af den empiriske virkelighed, der reelt gør sig gældende i flerpartisystemer. Derfor har jeg i specialet forsøgt at nuancere et blokvist mål for elitepolariseringen i sprogbrug ved at inkorporere en sammenligning mellem enkelte partier og resten af Folketinget. Det er i særdeleshed muligt, da den metodiske tilgang bag målet bygger på en klassifikationsmodel, der skal indplacere taler i to forskellige grupper. Med andre ord er det muligt at konstruere et polariseringsmål, der afspejler mange forskellige opdelinger og grupperinger af Folketinget.

Polariseringsmålet, som bygger på en sammenligning af enkelte partier med Folketinget, er ikke uden videre uproblematisk. En central udfordring er blandt andet, at det øger den analytiske kompleksitet markant, da der skal foretages ligeså mange sammenligninger, som der er partier. Samtidig efterlader det igen et teoretisk spørgsmål om, hvornår det samlede partisystem faktisk er polariseret. Er det tilstrækkeligt, at et parti er polariseret i forhold til resten af Folketinget, eller kræver det flere partier? Der er således stadig et potentiale for at forbedre det metodiske såvel som det teoretiske grundlag for at analysere polarisering i flerpartisystemer. Et muligt fokuspunkt for fremtidig forskning er at udforske denne problematik yderligere. Det skal også ses i lyset, at resultaterne i

specialet viser, at inddragelsen af flere sammenligninger, som fokuserer på polariseringen for enkelte partier, tilføjer nuancer, der ikke ses i en blokvis opdeling.

7.3 Empiriske implikationer: Konsekvenser af elitepolarisering

I analysen observeres der perioder, hvor polariseringen er tydeligt højere end andre, og forskellene i politikernes sprogbrug ændrer sig løbende i perioden fra 1991 til 2020. Spørgsmålet er dog, om vi bør være bekymrede for denne udvikling. Er det for eksempel bedre, at vi i dag observerer et polariseringsniveau, som er lavere end polariseringen omkring 2010-2012? En udfordring ved at diskutere dette spørgsmål er, at der er meget få studier, der har undersøgt effekterne af polarisering i sprogbrug. Det gør det svært at komme med præcise svar på, hvad der er de empiriske implikationer af mine resultater. Alligevel vil jeg i de følgende diskutere, om polariseringen i sprogbrug overhovedet udgør et demokratisk problem, og hvordan fremtidig forskning kan hjælpe med at belyse dette bedre.

Det er langt fra entydigt, at høj polarisering i sprogbrug skal betragtes som et demokratisk problem. Alligevel nævnes polarisering sjældent som et fænomen, der har en positiv indvirkning på samfundet. En af argumenterne bag denne betragtning er, at høj elitepolarisering kan skabe stilstand, da politikere ikke ønsker at samarbejde om politikforslag (Jones og McDermott, 2004). Historisk set er det dog ikke et stort problem i dansk politik, hvor der selv under VK-regeringen var samarbejde på tværs af politiske blokke (Kurrild-Klitgaard et al., 2006). Der kan endda argumenteres for, at elitepolariseringen i sprogbrug reelt styrker demokratiet, da det hjælper vælgerne med at forstå, hvad forskellige partier står for. Denne betragtning er koblet til det, der i litteraturen kaldes *the responsible party model*, og polarisering er derfor positiv, da det styrker vælgernes mulighed for at holde politikere ansvarlige for politiske løfter og beslutninger, og samtidig vælge de partier, som bedst spejler deres egne holdninger (Jones og McDermott, 2004; Poole og Rosenthal, 1985). En vis grad af polarisering kan være positivt i det omfang, at det hjælper vælgere med at finde det "rigtige" parti.

En forudsætning for, at polarisering i sprogbrug har en positiv indvirkning på demokratiet er, at polariseringen i sprogbrug faktisk også spejler reelle forskelle mellem partier. Det er derfor bekymrende, at polariseringen i sprogbrug ikke følger polariseringen i afstemninger. Hvis polariseringen i sprogbrug blot udtrykker strategiske beslutninger og

overvejelser, som har til formål at vinde vælgere, så er en mulig konsekvens, at vælgerne stemmer på et parti uden, at partiets politiske beslutninger stemmer overens med dets sprogbrug. Afkoblingen mellem politikernes sprogbrug og afstemningsadfærd kan gøre det svært for vælgerne at navigere i politikernes reelle standpunkter. En væsentlig indvending imod dette er dog, at vælgere med al sandsynlighed vil blive opmærksomme på, at politikere ikke leverer på deres løfter og derfor ændrer deres stemmer ved næste valg. Jeg vil derfor argumentere for, at der ikke er grund til at forvente, at førnævnte problematik har varige negative effekter på vælgerne mulighed for at vælge det ”rigtige” parti. Overordnet set vil jeg derfor fastholde, at en vis grad af polarisering i sprogbrug kan være positivt for demokratiet i det omfang, at det hjælper vælgere med at stemme på politikere og partier, som bedst repræsenterer deres holdninger.

Endelig er et vigtigt aspekt i diskussionen om, hvorvidt perioder med høj elitepolarisering er problematisk, hvordan det påvirker polariseringen mellem vælgerne. Her er effekten af polarisering i sprogbrug på affektiv polarisering særlig interessant, da affektiv polarisering er blevet koblet til forfordeling af ligesindede og forskelsbehandling - ikke blot i den politisk sfære, men også udenfor (Druckman et al., 2021; Iyengar et al., 2019). Det skal også ses i relation til, at flere nye studier netop viser, at affektiv polarisering også findes i en dansk kontekst (Brink og Hansen, 2019; Hjorth et al., 2019). En faktor, der kan føre til affektiv polarisering, er ideologisk elitepolarisering (Skytte, 2020). Imidlertid viser tidligere studier, at der ikke har været en særlig stor grad af ideologisk elitepolarisering i Danmark over de senere år (Christiansen, 2019b). Derfor kan det være nødvendigt at kigge andre steder hen for at forstå, hvorfor der er fundet affektiv polarisering i Danmark. Her er polariseringen i sprogbrug en mulig forklaring. Det er ikke utænkeligt, at polariseringen i sprogbrug mellem politikere kan føre til affektiv polarisering. For eksempel har et tidligere studie vist, at en hård tone mellem eliten kan føre til affektiv polarisering (Skytte, 2020). Det viser, at politikeres sprogbrug ikke er uvæsentligt for affektiv polarisering. Selvom polarisering i sprogbrug ikke nødvendigvis indebærer en hårdere tone, kan det være et interessant udgangspunkt for fremtidig forskning i affektiv polarisering. Det vil samtidig bidrage med endnu et perspektiv til diskussionen af, om elitepolarisering i sprogbrug kan have en negativ betydning for demokratiet.

Litteratur

- Abramowitz, A. I. & Saunders, K. L. (2008). Is Polarization a Myth? *The Journal of Politics*, 70(2), 542–555.
- Adcock, R. & Collier, D. (2001). Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *The American Political Science Review*, 95(3), 529–546.
- Aldrich, J. H. & Rohde, D. W. (1997). The Transition to Republican Rule in the House: Implications for Theories of Congressional Politics. *Political Science Quarterly*, 112(4), 541–567.
- Andersen, J. G. & Borre, O. (2003). *Politisk forandring. Værdipolitik og nye skillelinjer ved folketingsvalget 2001*. Systime.
- Bach, A., Svejgaard, J. & Hjorth, F. (2019). Maskinlæring som politologisk værktøj. *Politica*, 51(2), 168–186.
- Barber, M. & McCarty, N. (2015). Causes and consequences of polarization. *Political negotiation: A handbook*. Brookings Institution Press.
- Barberá, P. & Rivero, G. (2015). Understanding the Political Representativeness of Twitter Users. *Social Science Computer Review*, 33(6), 712–729.
- Benoit, K. & Laver, M. (2012). The Dimensionality of Political Space: Epistemological and Methodological Considerations. *European Union Politics*, 13, 194–218.
- Bille, L. (2006). *Det nye flertal. Dansk partipolitik 2001-2005*. Djøf Forlag.
- Bille, L. (2014). *Blå eller rød...? Dansk partipolitik 2005-2011 i perspektiv*. Djøf Forlag.
- Bond, J. R. & Fleisher, R. (Red.). (2000). *Polarized Politics: Congress and the President in a Partisan Era*. CQ Press.
- Brewer, M. (2005). The Rise of Partisanship and the Expansion of Partisan Conflict Within the American Electorate. *Political Research Quarterly*, 58(2), 219–229.
- Brink, L. & Hansen, T. (2019). Eksisterer der et partipolitisk ”os og dem”? Et surveyeksperimentelt studie af affektiv polarisering i Danmark. *Økonomi og Politik*, 92(3), 10–30.
- Brown, K., Mondon, A. & Winter, A. (2021). The far right, the mainstream and mainstreaming: towards a heuristic framework. *Journal of Political Ideologies*, 1–18.
- Bräuninger, T., Müller, J. & Stecker, C. (2016). Modeling Preferences Using Roll Call Votes in Parliamentary Systems. *Political Analysis*, 24(2), 189–210.

- Christiansen, F. (2019a). »Jeg vil ikke stemple dem«: Ny minister dropper at bruge ordet ghetto. *Politiken*.
- Christiansen, F. J. (2016). The Danish People's Party: Combining cooperation and radical positions. I T. Akkerman, S. L. de Lange & M. Rooduijn (Red.), *Radical Right-Wing Populist Parties in Western Europe: Into the Mainstream?* (s. 94–112). Routledge.
- Christiansen, F. J. (2017). Conflict and co-operation among the Danish Mainstream as a condition for adaption to the populist radical right. I P. Odmalm & E. Hepburn (Red.), *The European Mainstream and the Populist Radical Right* (s. 49–70). Routledge.
- Christiansen, F. J. (2019b). Polarisering af afstemninger om lovgivning i Folketinget. *Økonomi og Politik*, 92(3), 76–90.
- Cirone, A., Cox, G. W. & Fiva, J. H. (2021). Seniority-based nominations and political careers. *American Political Science Review*, 115(1), 234–251.
- Dalton, R. J. (2008). The Quantity and the Quality of Party Systems: Party System Polarization, Its Measurement, and Its Consequences. *Comparative Political Studies*, 41(7), 899–920.
- Damgaard, E. & Svensson, P. (1989). Who governs? Parties and policies in Denmark. *European Journal of Political Research*, 17(6), 731–745.
- Dinesen, P. T. (2019). Politisk polarisering i Danmark. *Økonomi og Politik*, 92(3), 6–9.
- Downs, A. (1957). An Economic Theory of Political Action in a Democracy. *Journal of Political Economy*, 65(2), 135–150.
- Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M. & Ryan, J. B. (2021). Affective polarization, local contexts and public opinion in America. *Nature Human Behaviour*, 5(1), 28–38.
- Egerod, B. C. & Klemmensen, R. (2019). Scaling Political Positions from Text. I L. Curi- ni & R. Franzese (Red.), *The SAGE Handbook of Research Methods in Political Science and International Relations* (s. 498–521). Sage Publications.
- Fiva, J. H., Nedregard, O. & Oien, H. (2021). Polarization in Parliamentary Speech. *Working paper*, 61.
- Flyvbjerg, B. (2010). Fem misforståelser om casestudiet. I S. Brinkmann & L. Tanggaard (Red.), *Kvalitative metoder* (s. 463–487). Hans Reitzels Forlag.

- Gabel, M. J. & Huber, J. D. (2000). Putting Parties in Their Place: Inferring Party Left-Right Ideological Positions from Party Manifestos Data. *American Journal of Political Science*, 44(1), 94–103.
- Gentzkow, M., Shapiro, J. M. & Taddy, M. (2019). Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech. *Econometrica*, 87(4), 1307–1340.
- Gerring, J. (2004). What Is a Case Study and What Is It Good for? *American Political Science Review*, 98(2), 341–354.
- Gerring, J. (2012). Mere Description. *British Journal of Political Science*, 42(4), 721–746.
- Goet, N. D. (2019). Measuring Polarization with Text Analysis: Evidence from the UK House of Commons, 1811–2015. *Political Analysis*, 27(4), 518–539.
- Green, J., Edgerton, J., Naftel, D., Shoub, K. & Cranmer, S. J. (2020). Elusive consensus: Polarization in elite communication on the COVID-19 pandemic. *Science Advances*, 6(28), 1–5.
- Green-Pedersen, C. & Kosiara-Pedersen, K. (2020). Det danske partisystem mellem kontinuitet og forandring. *Politica*, 52(3), 215–233.
- Grimmer, J. & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3), 267–297.
- Hansen, K. M. & Stubager, R. (2020). At lave undersøgelser inden for statskundskab. I S. W. Hansen, K. M. Hansen & L. B. Andersen (Red.), *Metoder i statskundskab* (3.). Hans Reitzels Forlag.
- Hansen, M. E. (2008). Back to the Archives? A Critique of the Danish Part of the Manifesto Dataset. *Scandinavian Political Studies*, 31(2), 201–216.
- Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2. udg.). Springer-Verlag.
- He, H. & Garcia, E. (2009). Learning from Imbalanced Data. *Knowledge and Data Engineering, IEEE Transactions on*, 21(19), 1263–1284.
- Hetherington, M. J. (2009). Review Article: Putting Polarization in Perspective. *British Journal of Political Science*, 39(2), 413–448.

- Hjorth, F. (2019). Establishment Responses to Populist Challenges: Evidence from Legislative Speech. *Prepared for presentation at the 2019 Annual Meeting of the European Political Science Association*, 1–31.
- Hjorth, F. & Loftis, M. W. (2020). "Big data": alternative datakilder og analysemuligheder. I K. M. Hansen, L. B. Andersen & S. W. Hansen (Red.), *Metoder i Statskundskab* (3. udg.). Hans Reitzels Forlag.
- Hjorth, F., Sønderskov, K. & Dinesen, P. (2019). Affektiv Polarisering i Danmark: Et Listeeksperiment om Social Distance til Politiske Modstandere. *Økonomi og Politik*, 92(3), 31–40.
- Hooghe, L., Bakker, R., Brigeovich, A., Vries, C. D., Edwards, E., Marks, G., Rovny, J., Steenbergen, M. & Vachudova, M. (2010). Reliability and validity of the 2002 and 2006 Chapel Hill expert surveys on party positioning. *European Journal of Political Research*, 49(5), 687–703.
- Huber, J. & Inglehart, R. (1995). Expert Interpretations of Party Space and Party Locations in 42 Societies. *Party Politics*, 1(1), 73–111.
- Inglehart, R. (2008). Changing Values Among Western Publics from 1970 to 2006. *West European Politics*, 31, 130–146.
- Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N. & Westwood, S. J. (2019). The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science*, 22(1), 129–146.
- Iyengar, S., Sood, G. & Lelkes, Y. (2012). Affect, Not Ideology. *Public Opinion Quarterly*, 76(3), 405–431.
- Iyengar, S. & Westwood, S. J. (2015). Fear and Loathing across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science*, 59(3), 690–707.
- Jones, D. R. & McDermott, M. L. (2004). The Responsible Party Government Model in House and Senate Elections. *American Journal of Political Science*, 48(1), 1–12.
- Kam, C. J. (2009). *Party Discipline and Parliamentary Politics*. Cambridge University Press.
- King, G., Keohane, R. O. & Verba, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press.
- Kosiara-Pedersen, K. (2020). Tillykke til Dansk Folkeparti! *Politica*, 52(3), 269–289.

- Kuhn, M. & Johnson, K. (2013). *Applied Predictive Modeling* (1. udg.). Springer.
- Kurrild-Klitgaard, P. (2011). Kontraktpolitik, kulturkamp og ideologi 2001-2011. *Økonomi og Politik*, 83(3), 47–62.
- Kurrild-Klitgaard, P., Klemmensen, R. & Hansen, M. E. (2006). Blokpolitik og det ”samarbejdende folkestyres” fire gamle partier, 1953-2005. *Økonomi og Politik*, 79(1), 79–85.
- Layman, G. C. & Carsey, T. M. (2002). Party Polarization and Party Structuring of Policy Attitudes: A Comparison of Three NES Panel Studies. *Political Behavior*, 24(3), 199–236.
- Levendusky, M. S. (2010). Clearer cues, more consistent voters: A benefit of elite polarization. *Political Behavior*, 32(1), 111–131.
- Levendusky, M. (2009). *The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans*. University of Chicago Press.
- Lijphart, A. (1999). *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries*. New Haven: Yale University Press.
- McCarty, N. (2019). *Polarization: What Everyone Needs to Know*. Oxford University Press.
- McCarty, N., Poole, K. & Rosenthal, H. (2006). *Polarized America: The Dance of Ideology and Unequal Riches*. The MIT press.
- Merkley, E., Bridgman, A., Loewen, P. J., Owen, T., Ruths, D. & Zhilin, O. (2020). A Rare Moment of Cross-Partisan Consensus: Elite and Public Response to the COVID-19 Pandemic in Canada. *Canadian Journal of Political Science/Revue canadienne de science politique*, 53(2), 311–318.
- Müller, W. C. & Strøm, K. (Red.). (1999). *Policy, Office, or Votes?: How Political Parties in Western Europe Make Hard Decisions*. Cambridge University Press.
- Petersen, M. B. & Roepstorff, A. (2020). *The Danes’ behaviour during the Corona Crisis* (tekn. rap.).
- Peterson, A. & Spirling, A. (2018). Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems. *Political Analysis*, 26(1), 120–128.
- Poole, K. T. & Rosenthal, H. (1985). A spatial model for legislative roll call analysis. *American Journal of Political Science*, 357–384.

- Proksch, S.-O. & Slapin, J. B. (2012). Institutional Foundations of Legislative Speech. *American Journal of Political Science*, 56(3), 520–537.
- Raschka, S. (2015). *Python Machine Learning*. Packt Publishing Ltd.
- Rennie, J., Shih, L., Teevan, J. & Karger, D. (2003). Tackling the Poor Assumptions of Naive Bayes Text Classifiers. *Proceedings of the Twentieth International Conference on Machine Learning*, 41.
- Ritzau. (2019). Ny boligminister dropper ordet ghetto. *Avisen.dk*.
- Robison, J. & Mullinix, K. J. (2016). Elite Polarization and Public Opinion: How Polarization Is Communicated and Its Effects. *Political Communication*, 33(2), 261–282.
- Sartori, G. (1976). *Parties and Party Systems: A Framework for Analysis*. ECPR Press.
- Schlesinger, J. A. (1985). The New American Political Party. *The American Political Science Review*, 79(4), 1152–1169.
- Skjæveland, A. (2001). Party Cohesion in the Danish Parliament. *The Journal of Legislative Studies*, 7(2), 35–56.
- Skytte, R. (2020). Dimensions of Elite Partisan Polarization: Disentangling the Effects of Incivility and Issue Polarization. *British Journal of Political Science*, 1–19.
- Slapin, J. B. & Proksch, S.-O. (2008). A Scaling Model for Estimating Time-Series Party Positions from Texts. *American Journal of Political Science*, 52(3), 705–722.
- Theriault, S. M. (2008). *Party Polarization in Congress*. Cambridge University Press.
- Thomsen, J. P. F. (2017). Indvandring som politisk skillelinje blandt ”os”. *Oprør fra udkanten : Folketingsvalget 2015* (s. 265–279). Djøf Forlag.

Appendiks

Indhold

A	Oversigt over variable	2
B	Superviseret maskinlæring	4
B.1	Mulighedsrum for hyperparametre	4
B.2	Receiver-operator characteristics (ROC) kurver	5
B.3	Random Forest modeller	8
C	Værdi- og fordelingspolitiske ordbøger	13
D	Validering af mål for elitepolarisering i sprogbrug	15
D.1	Indholdsvalidering	15
D.2	Diskriminerende validering	24
E	Regeringer i dansk politik	25

A Oversigt over variable

Tabellen nedenfor viser en oversigt over de forskellige variable, som indgår i datasættet med folketingstaler. Derudover har jeg tilføjet en kort beskrivelse af, (1) hvad variablen indeholder, og (2) hvilken type variablen er.

Tabel A.1: Oversigt over variable i datasættet

Variable	Beskrivelse	Type
Talestykke	Indeholder alle ord i en given tale efter, at der fjernet tegn, tal og stopord mv. fra talen. Talestykker udgør 3-5 sætninger og er dermed en mindre del af den samlede tale, som talestykket oprindeligt indgik i.	Tekststreng
Samlede tale	Indeholder den fulde tale som et bestemt talestykke oprindeligt har været en del af i sin originale form.	Tekststreng
Rolle	Angiver den rolle som en given politiker har i Folketinget. Rollerne omfatter ministertitel eller formandstitler (for eksempel formand, aldersformand mv.)	Kategorisk
Parti	Angiver det parti som er afsenderen bag en given tale. Variablen er one-hot-encoded således, at der er en variabel for hver kategorisk enhed. Hver af disse variable kan antage værdien 1 (hvis et talestykke tilhører partiet) eller 0 (hvis et talestykke tilhører en af de andre partier)	Kategorisk (binær)
Blok	Angiver den blok som er afsenderen bag en given tale. Variablen er binær, hvor 0 angiver, at talestykket tilhører rød blok, mens 1 angiver, at talestykket tilhører blå blok. (Rød blok = S, RV, SF, EL og ALT. Blå blok = K, V, CD, FP, KD, DF, LA og NB)	Kategorisk (binær)
Politikområde	En række variable som angiver om et givent talestykke falder under et af følgende politikområder: velfærd, skat, udlændinge eller klima. Hver af disse variable kan antage værdien 1 (hvis et talestykke tilhører politikområdet) eller 0 (hvis talestykket ikke gør). Et talestykke kan godt falde under flere politikområder på samme tid.	Kategorisk (binær)
Antal ord	Antallet af ord som indgår i et givent talestykke.	Kontinuer
Dato	Indeholder tidspunktet for talen. Tidspunktet er angivet som: dd-mm-åååå.	Kontinuer

TF-IDF	En matrix bestående af de forskellige ord som indgår i det samlede korpus af tekster for et givent år. Hvert ord udgør sin egen variabel og har en værdi/vægt, som afhænger af, (1) hvor meget ordet fylder i den specifikke tale, og (2) hvor meget ordet fylder i det samlede korpus. En høj vægt opnås hvis ordet har en høj frekvens i en given tale, men en lav frekvens i det samlede korpus af tekster.	Kontinuer
--------	--	-----------

B Superviseret maskinlæring

Denne del af appendiks giver et detaljeret indblik i nogle af de væsentlige elementer fra mine superviserede maskinlæringsmodeller. Først giver jeg et overblik over mulighedsrummet for hyperparametrene, som jeg har optimeret i mine grid searches. Dernæst præsenterer jeg ROC-kurverne fra Naïve Bayes modellerne. Endelig viser jeg resultaterne af minde Random Forest modeller, som bidrager til at validere Naïve Bayes modellerne.

B.1 Mulighedsrum for hyperparametre

Tabellen nedenfor viser de hyperparametre, som jeg har optimeret i forbindelse med træningen af mine superviserede maskinlæringsmodeller. Tabellen indeholder en kort beskrivelse af de enkelte parametre, herunder hvilken model parameteret tilhører.

Tabel B.1: Hyperparametre i grid search

Parameter	Beskrivelse	Mulighedsrum	Algoritme
Alpha	En udfordring ved Naïve Bayes algoritmen er risikoen for nul sandsynlighed. Dette problem kan for eksempel opstå, hvis der i test-/valideringsdata er et nyt ord, som modellen ikke før har set, og dermed ikke er trænet på. Dette kan håndteres med Laplace-smoothing, som opjusterer den aposteriori sandsynlighed for at observere et ord. Et højere alpha vil skubbe sandsynligheden for at observere et givent ord mod 0,5, og regulerer dermed algoritmen.	10, 1, 0.1, 0.01, 0.001, 0.0001	Naïve Bayes
No. of estimators	Antallet af beslutningstræer, som indgår i en model. Et højere antal af beslutningstræer styrker ikke nødvendigvis modellen. Det leder ikke til overfitting, men det øger kompleksiteten af modellen.	16, 32, 64, 128	Random Forest Classifier

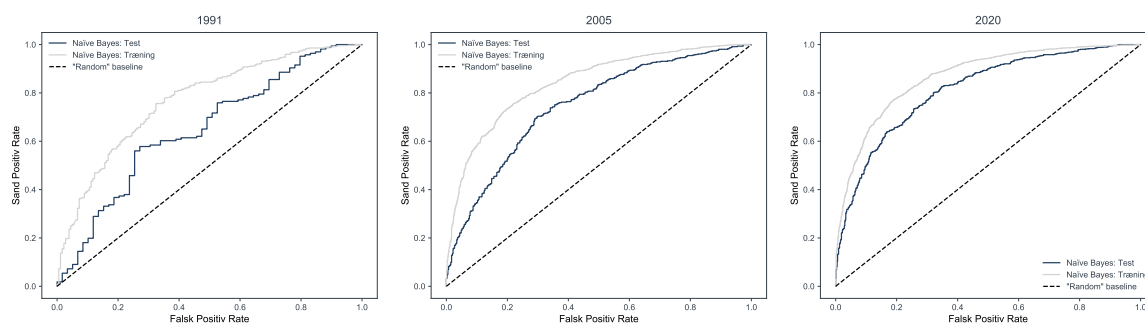
Max depth	Antal niveauer i hvert beslutningstræ. Parametret styrer højden af træet. Højere træer kan ofte styrke en models accuracy. Det kan dog lede til overfitting og lavere accuracy, hvis dybden bliver for høj.	3, 5, 7, 9	Random Forest Classifier
Min samples split	Mindste antal af datapunkter tilladt i hvert interne "blad" (parameter) før, at det kan splitte til et nyt "blad". Hvis værdien er for lav, vil et givent træ blive ved med at vokse sig større, hvilket vil lede til overfitting. Flere samples i hvert interne blad kan reducere overfitting.	2, 4, 6	Random Forest Classifier
Min samples leaf	Mindste antal af datapunkter tilladt i hvert "blad" efter, at det er blevet splittet. Hvis værdien er for lav, kan det føre til overfitting, da et træ vokser sig større og inkluderer flere "blade".	1, 2, 4	Random Forest Classifier

B.2 Receiver-operator characteristics (ROC) kurver

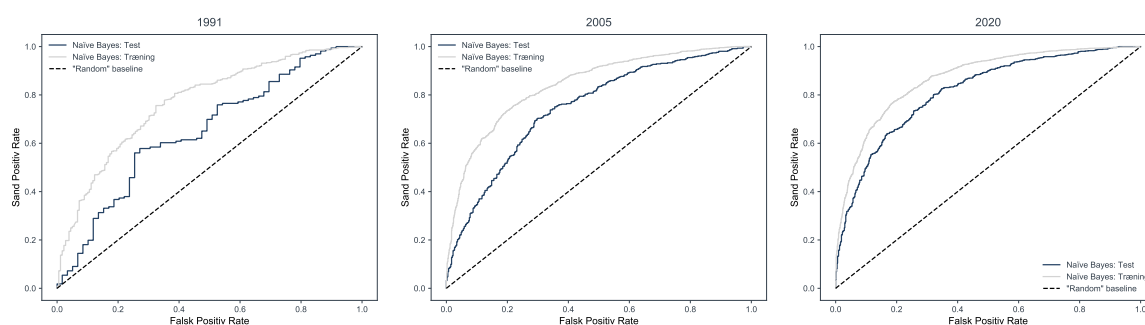
De følgende figurer viser ROC-kurverne for mine Naïve Bayes modeller. Af hensyn til fremstillingen har jeg alene inkluderet ROC-kurverne for den blokvisе sammenligning på tværs af alle politikområder. Samtidig viser jeg kun ROC-kurverne for 1991, 2005 og 2020. Denne afgrænsning skal ses i lyset af, at der alternativt skulle være vist over 100 ROC-kurver, hvilket er uhensigtsmæssigt. De resterende ROC-kurver kan fremsendes ved forespørgsel. Som forventeligt er modellernes performance på træningsdata bedre end performance på testdata. Det overordnede billede peger dog på, at der ikke er store tab i performance, når modellerne skal forudsige partitilhørsforhold på ukendt data. Med andre ord har mine modeller en relativt god out-of-sample performance.

Figur B.1: ROC-kurver

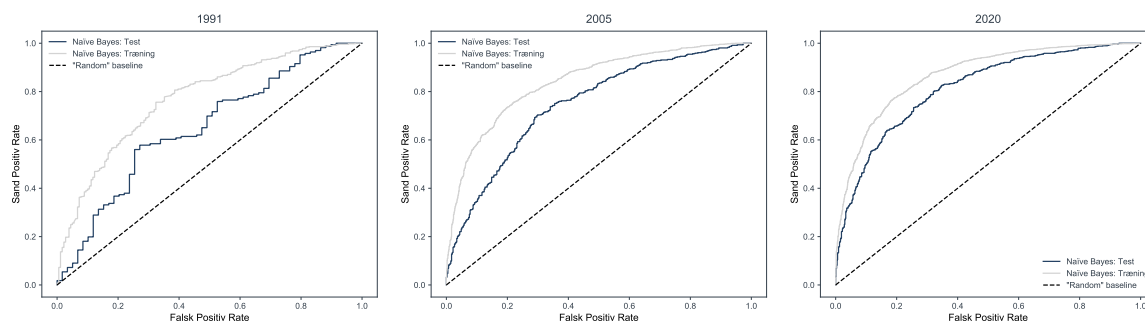
Blokke



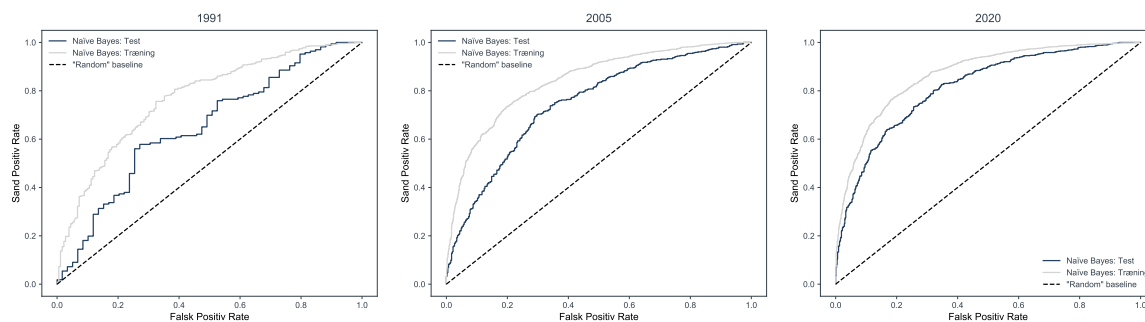
Socialdemokratiet



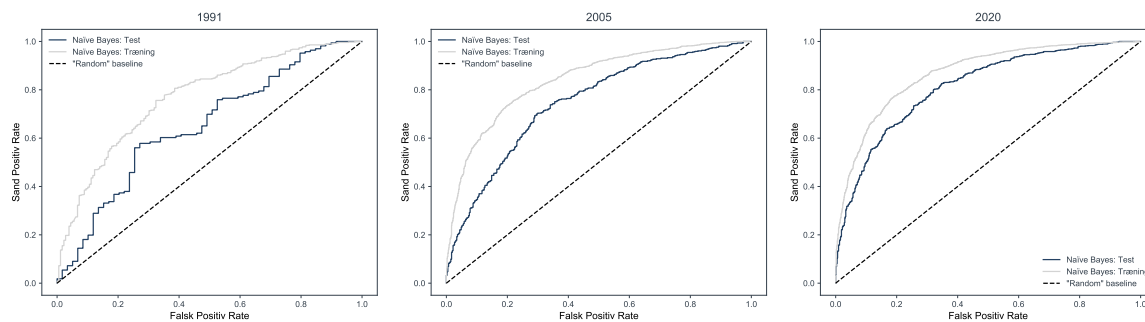
Venstre



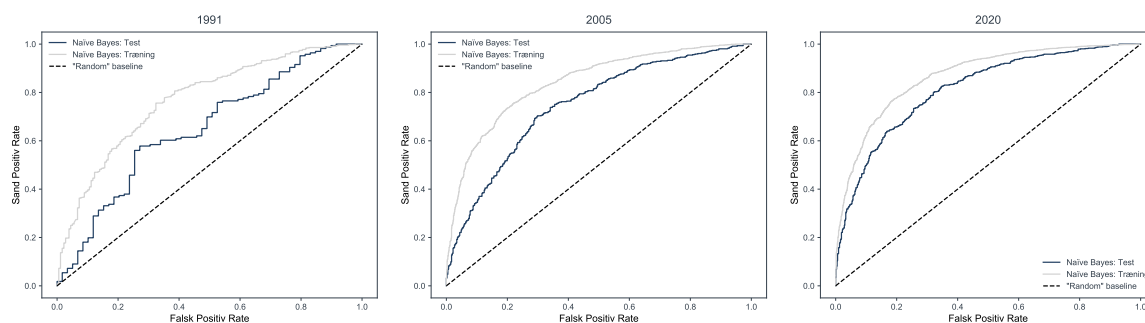
Dansk Folkeparti



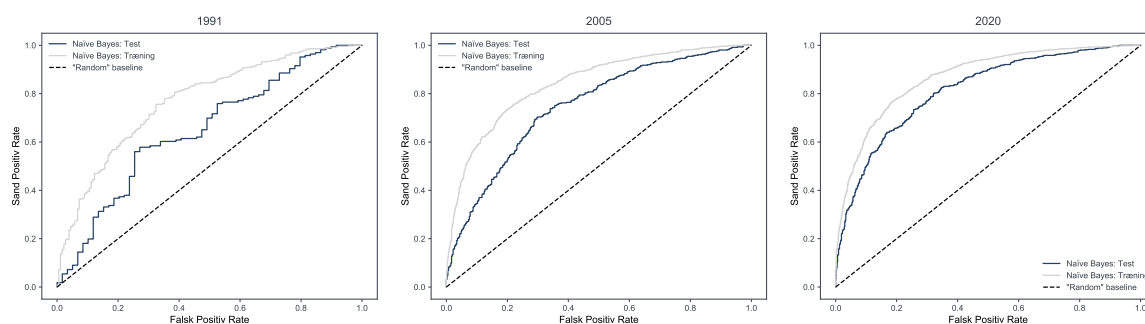
Enhedslisten



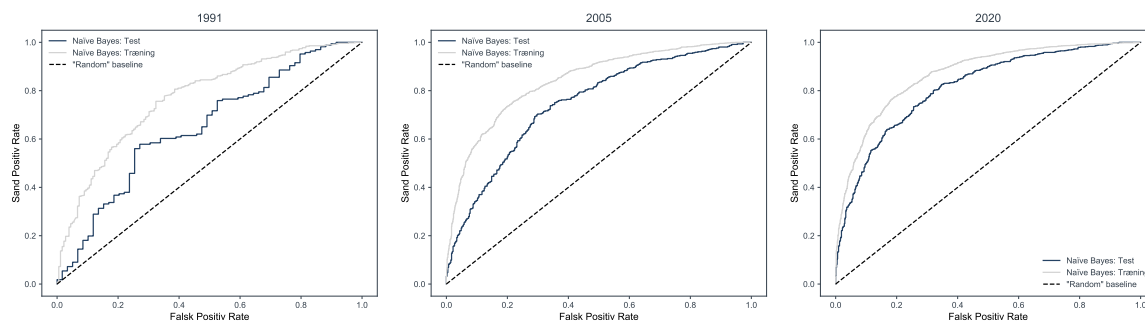
Konservative



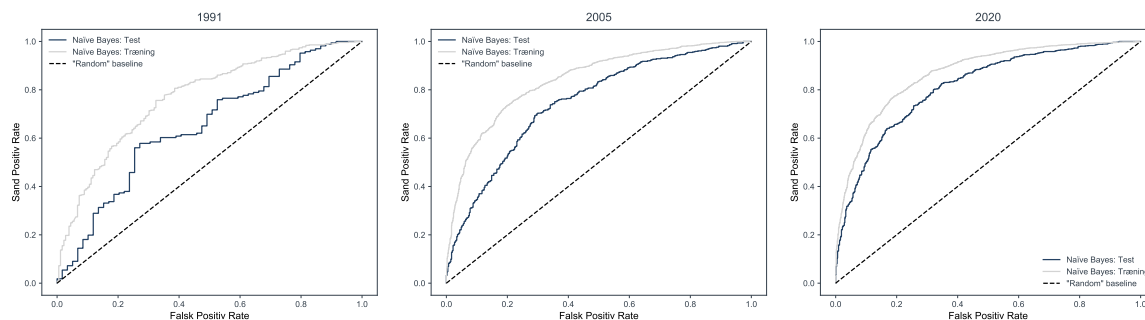
Socialistisk Folkeparti



Liberal Alliance



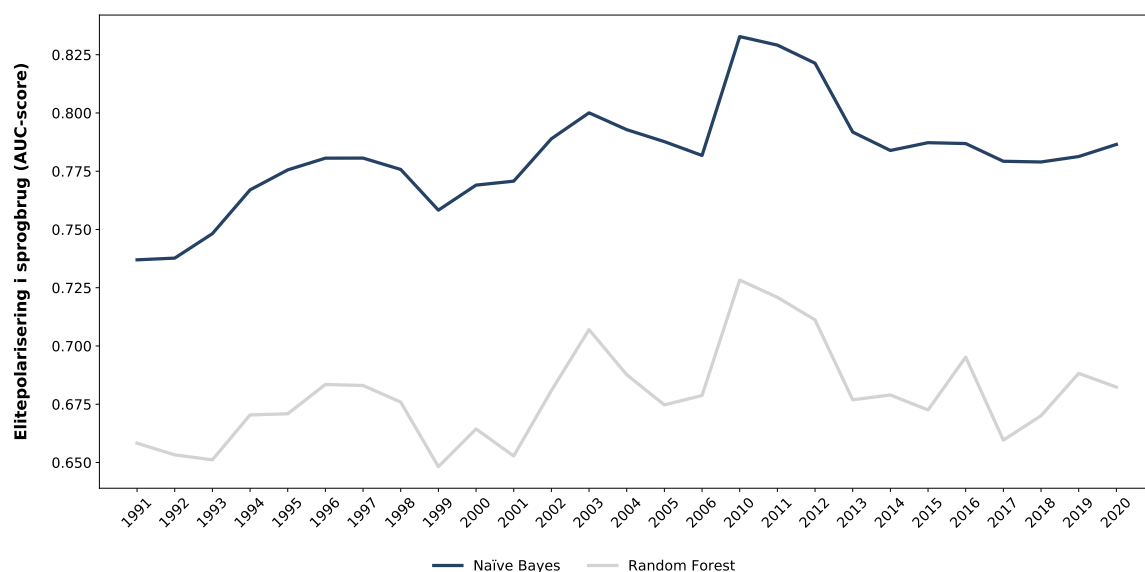
Radikale Venstre



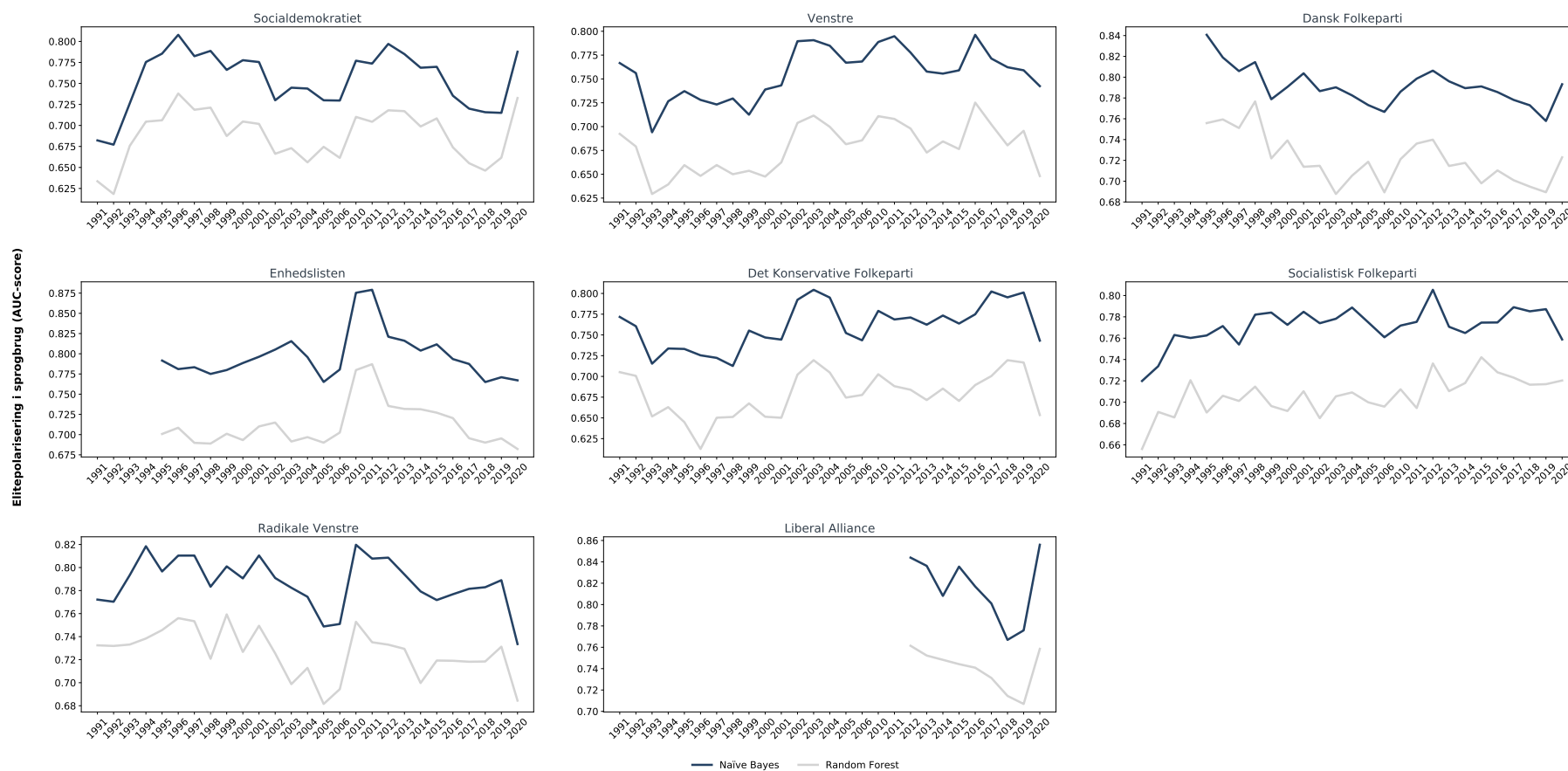
B.3 Random Forest modeller

Figurerne nedenfor viser udviklingen i AUC-scoren for de maskinlæringsmodeller, der er trænet med udgangspunkt i Random Forest algoritmen. Det fremgår, at der overordnet set er en ensartet udviklingstendens, når Random Forest sammenlignes med Naïve Bayes. Det fremgår dog også, at Naïve Bayes algoritmen performer bedre end Random Forest algoritmen. Dette mønster gør sig gældende på tværs af de blok- og partivise sammenligning og på tværs af opdelingen i værdi- og fordelingspolitiske emner. Det tyder på, at udviklingen i elitepolarisering af sprogbrug ikke er algoritme specifik, hvilket styrker troværdigheden af mine resultater.

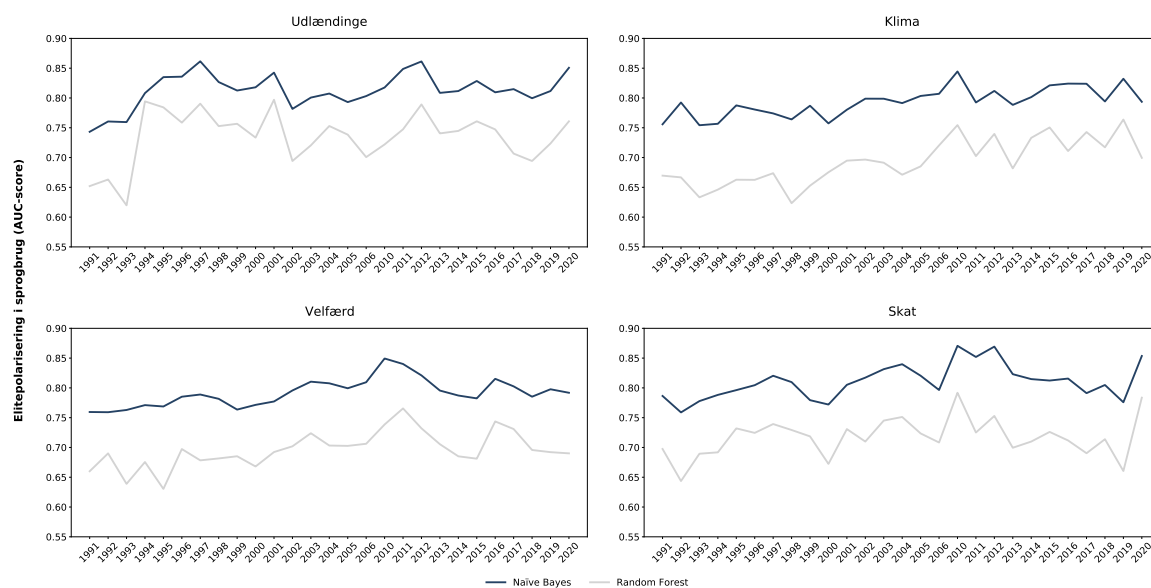
Figur B.2: Udviklingen i elitepolarisering i sprogbrug mellem blokke (RFC)



Figur B.3: Udviklingen i elitepolarisering i sprogbrug mellem enkelte partier og resten af Folketinget (RFC)

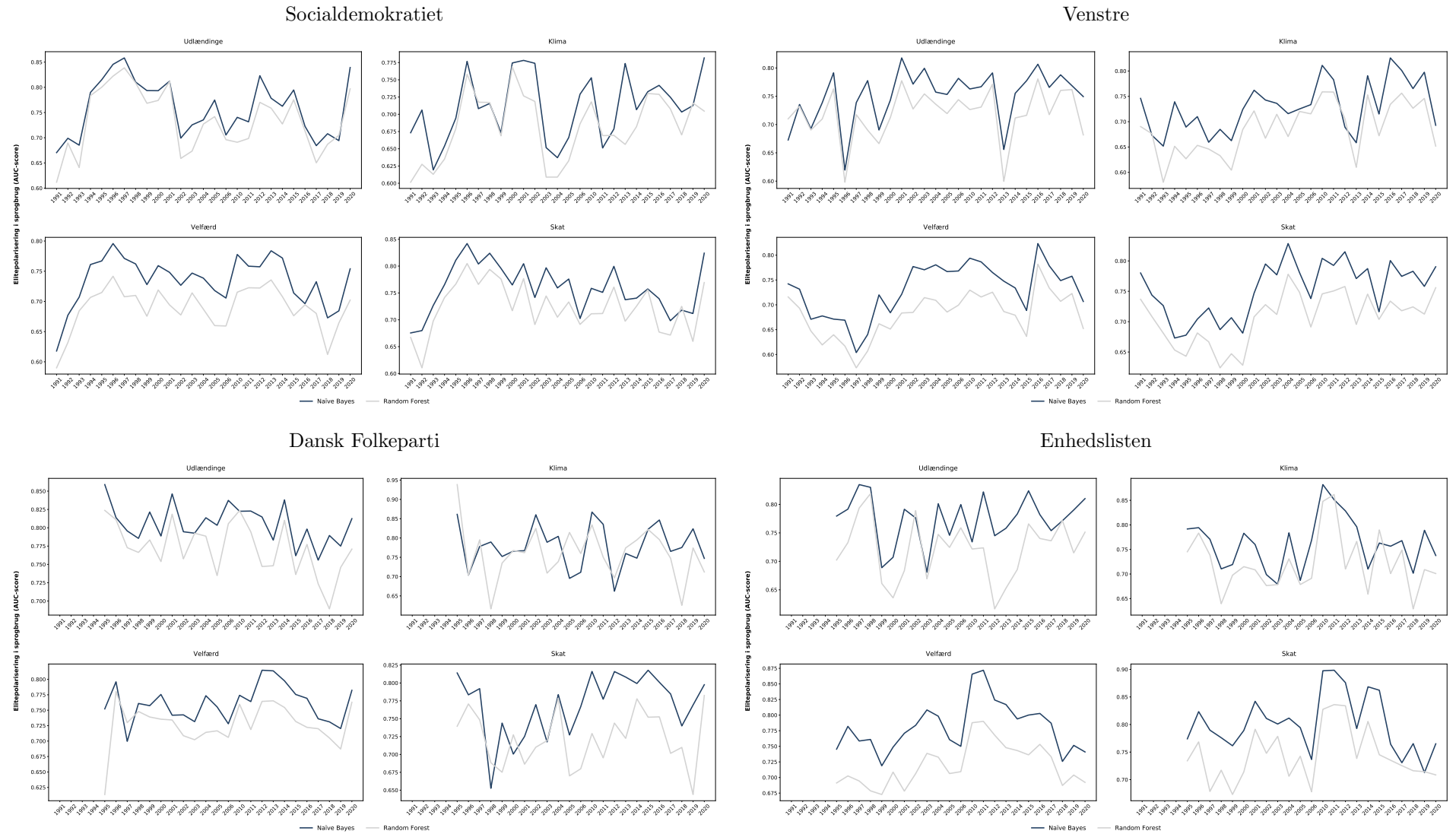


Figur B.4: Udviklingen i elitepolarisering i sprogbrug mellem blokke på tværs af værdi- og fordelingspolitiske emner (RFC)

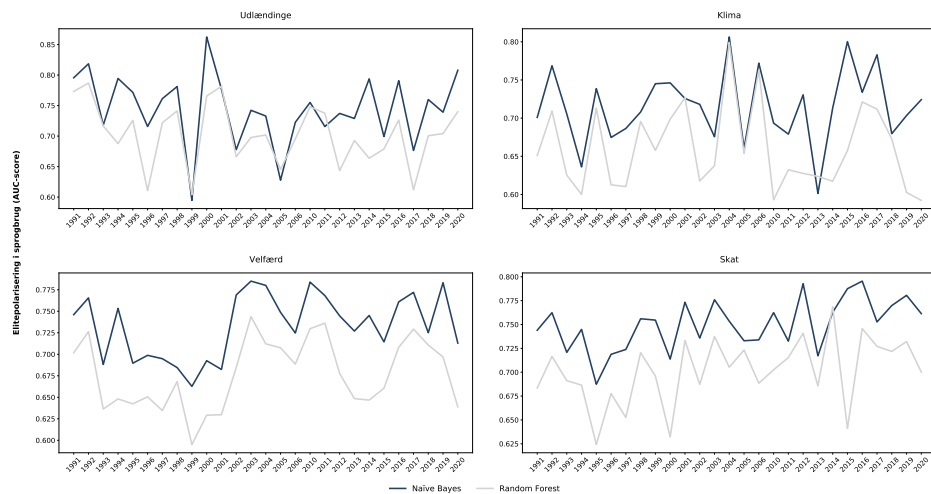


Figur B.5: Udviklingen i elitepolarisering i sprogbrug mellem enkelte partier og resten af Folketinget på tværs af værdi- og fordelingspolitiske emner (RFC)

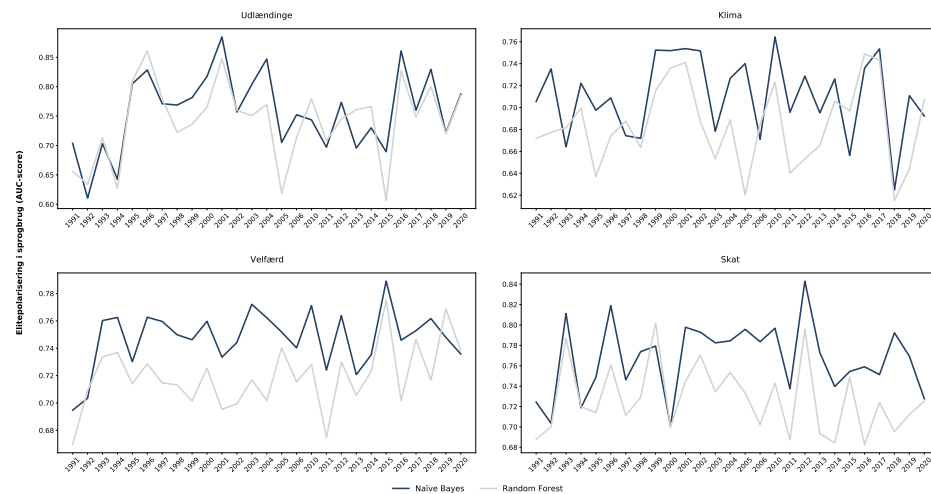
11



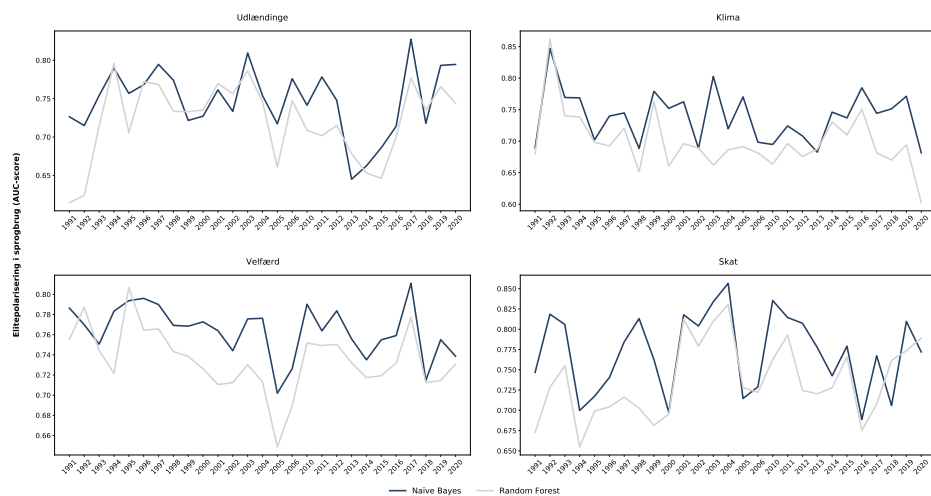
Det Konservative Folkeparti



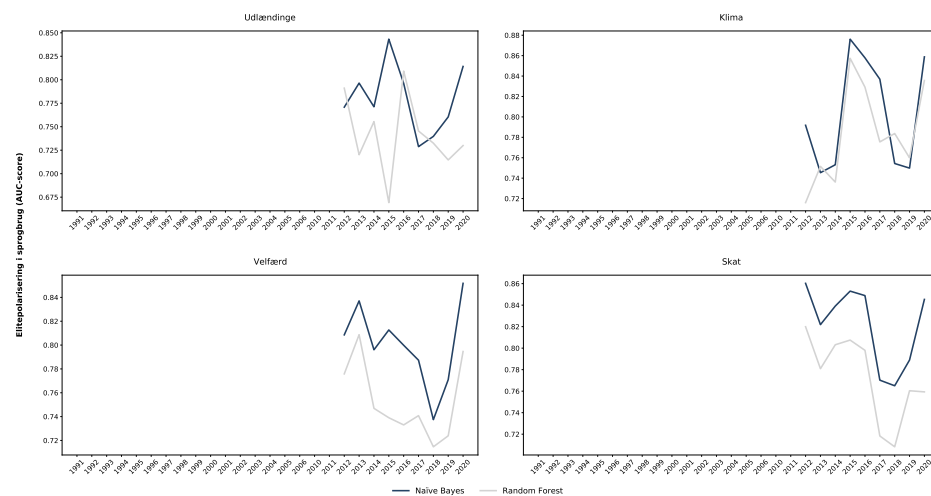
Socialistisk Folkeparti



Radikale Venstre



Liberal Alliance



C Værdi- og fordelingspolitiske ordbøger

Tabellen nedenfor viser de værdi- og fordelingspolitiske ordbøger, som er brugt til at opdele folketalingstalerne i forskellige politiske emner. En given tale indgår under et politiske emne, hvis talen indeholder en af ordene for et givent emne. Det bemærkes, at talerne godt kan indgå i flere kategorier.

Tabel C.1: Fordelings- og værdipolitiske ordbøger

Politik- område	Ordbog
Velfærd	'velfærdsreformer', 'efterløn', 'dagpasning', 'dagpenge', 'mindsteløn', 'ulighed', 'arbejdsløs', 'aktivering', 'pension', 'offentlige ydelser', 'førtidspension', 'kontanthjælp', 'kontanthjælpsloft', 'kontanthjælpsniveau', 'velfærd', 'plejehjem', 'børnehaver', 'børnehave', 'folkeskoler', 'folkeskolen', 'velfærdssamfund', 'overenskomst', 'dagpengesystem', 'velfærdssystem', 'hjemmehjælp', 'forsøgelse', 'forsørgelsesgrundlag', 'folkepension', 'arbejdsmarkedet', 'ældre', 'beskæftigelse', 'arbejdsmarkedskøen', 'fattigdom', 'beskæftigelse', 'flexicurity', 'velfærdsydelser', 'børnecheck', 'ældrechecken', 'kernevelføerden', 'kernevelføerd', 'serviceovens', 'kerneydeler', 'velfærdspolitik', 'velfærdspolitisk', 'bistandshjælp', 'dagpengeniveau', 'dagpengeperiode'
Skat	'skattestop', 'skattehuller', 'brugerbetaling', 'skattesystemet', 'skatte', 'skat', 'skattetryk', 'skattetryk', 'beskatning', 'skattevæsenet', 'moms', 'skattereform', 'skatterabat', 'topskatten', 'fradrag', 'skatteyder', 'topskatteyder', 'jobfradrag', 'skattenedsættelse', 'skattekontrolloven', 'SKAT', 'skatteudvalget', 'kommuneskat', 'skatteprocent', 'skattekort', 'trækprocent', 'beskatningsgrundlaget', 'skatter adet', 'skattebelagt', 'skatteministeren', 'ubeskattet', 'afgifter', 'gebyrer', 'udligningskatten', 'skattelettelse', 'skatteudspil', 'skattefri', 'skatteforslag', 'skatteaftalen', 'topskattelettelse', 'sambeskatningsordningen', 'skattestoppet', 'momsregistrerede', 'momsregnskab', 'skatteværdi', 'skattepligtige', 'skattepolitik', 'skattepolitisk', 'skatteløsninger', 'skattely'

Udlændinge	'statsborgerskab', 'opholdstilladelse', 'integration', 'flygtninge', 'indvandrere', 'tosprogede', 'udlændinge', 'islam', 'imamer', 'parallelsamfund', 'asylpolitik', 'integrationspolitik', 'udlændingepolitik', 'flygtningepolitik', 'asylpolitisk', 'integrationspolitisk', 'udlændingepolitisk', 'flygtningepolitisk', 'flygtningenævnet', 'ghetto', 'ghettoriserings-effekten', 'ghettosering', 'udlændingestyrelsen', 'flygtningekonvention', 'muslimer', 'racehygiejne', 'starthjælp', 'sprogbonus', 'indfødsret', 'indfødsretsprøve', 'udlændingepakken', 'danskundervisning', 'integrationsprogram', 'integrationsindsats', 'integrationsbarometer', 'udlændingepolitikken', 'migration', 'muslimske', 'asylpladser', 'asylsøgere', 'asylbørn', 'integrationsplaner', 'ikkevestlige', 'grænsekontrol', 'maskeringsforbuddet', 'maskeringsforbud', 'udlændingedeбат', 'udlændingekområdet', 'flygtningestrømme', 'integration', 'udlændinge', 'etnisk', 'anden etnisk', 'sharia', 'masseindvandring'
Klima	'forurening', 'klima', 'klimaforandringer', 'forurening', 'miljøeffekt', 'giftstoffer', 'co2', 'co2-reduktion', 'grøn omstilling', 'grønne omstilling', 'biodiversitet', 'PTX', 'IPCEI', 'vindmøller', 'vindmølle', 'grøn energi', 'klimaplan', 'kulbrinte', 'kulbrintebeskatning', 'vandmiljø', 'vandmiljøhandlingsplan', 'gødning', 'miljøskadelig', 'klimaindsatsen', 'vindmølle', 'energi', 'klimaændringer', 'luftforandringer', 'vandressourcer', 'skovressourcer', 'C02-miljøvenlig', 'benzinforurening', 'miljøoptimering', 'miljøøkonomi', 'fossilt brændstof', 'fossilt brændsel', 'fossile brændstoffer', 'fossile brændsler', 'solceller', 'miljøpolitik', 'klimapolitik', 'miljøbelastning', 'miljøvenlig', 'klimatiltag', 'miljøpolitisk', 'klimapolitisk', 'klimamål', 'klimakommission', 'atomkraft', 'fossil energi', 'klimamål', 'klimaregulering', 'miljøafgift'

D Validering af mål for elitepolarisering i sprogbrug

Nedenfor følger resultaterne af valideringen af mit maskinlæringsbaserede mål for elitepolarisering i sprogbrug. Først præsenteres resultaterne af indholdsvalideringen. Dernæst følger resultaterne af den diskriminerende validering.

D.1 Indholdsvalidering

Indholdsvalideringen består af to dele. To tabeller, som viser de mest røde og blå taler, samt en række figurer, der viser fordelingen af sandsynligheder for partiernes taler i forhold til, om de tilhører rød eller blå blok.

D.1.1 De mest røde og blå taler

Tabellerne nedenfor viser de taler som har den største sandsynlighed for at tilhøre rød og blå blok for de forskellige værdi- og fordelingspolitiske emner. Sandsynlighederne er estimeret med udgangspunkt i den blokvisе model for elitepolarisering.

Tabel D.1: De mest ”røde” taler

Politik- område	Tale	Taler/Sand- synlighed
Velfærd	Jeg vil gerne have Dansk Folkepartis ordfører til at bekræfte, at Dansk Folkeparti altså med sin tilslutning til denne her aftale, som også Hr. Mikkel Dencker har sagt, støtter den meget lave offentlige vækst, som er forudsætningen for hele den pakke, man har indgået, og at det allerede betyder, at den lavere vækst ude i kommunerne fra 2004 til 2005 skal bidrage med 4 mia. kr. til underholdning 4 mia. kr. i mindre vækst.	Jacob Buksti (0,7 pct)

Skat	Dansk Folkeparti har gennem årene været med til at gennemføre en lang stribe skattelettelser, og de rigeste har været topscorere hver eneste gang. Nu mangler pengene, og så skal regningen betales. Den bliver så sendt til børnefamilierne, til de arbejdsløse, til pensionisterne og til de lavtlønnede. og så ved jeg godt, man siger: nå ja, men de rige skal også betale.	Maja Panduro (0,3 pct)
Udlændinge	Nej, jeg har ikke den statistik, men det undrer mig, at Hr. Karsten Lauritzen ikke stiller spørgsmål om det, der er debattens indhold, nemlig forholdene for de børn, der lever i de omtalte familier. Som Hr. Karsten Lauritzen sikkert ved, fordi han er ordfører og trænet på området, er 1½ år meget, meget lang tid for et menneske at opholde sig i et asylcenter. Røde Kors siger, at 1 år er den maksimale opholdstid for en voksen, og derfor er ½ år nok det, det kan strækkes til at være acceptabelt for et barn. derfor må jeg bare konstatere, at Hr. Karsten Lauritzen synes, det er vigtigere, at man følger udlændingelovens § et eller andet, litra c, end at diskutere, hvordan man sørger for, at børn forskånes for at få yderligere psykiske lidelser end dem, de bringer med sig fra deres hjemland.	Margrethe Vestager (1,1 pct)
Klima	Så det var to gode ting. Så er det tredje et spørgsmål, jeg vil stille ministeren. Hvis vi nu er enige om, at vi trods alt nok har en målsætning om, at vi vil nedbringe danmarks afhængighed af det fossile brændstof, ville det så ikke være en god idé at have en sigtelinje, der handlede om, at i 2025 skulle vores olieforbrug være reduceret til halvdelen og vores c02-udslip også skulle være reduceret til halvdelen?	Per Clausen (3,5 pct)

Tabel D.2: De mest ”blå” taler

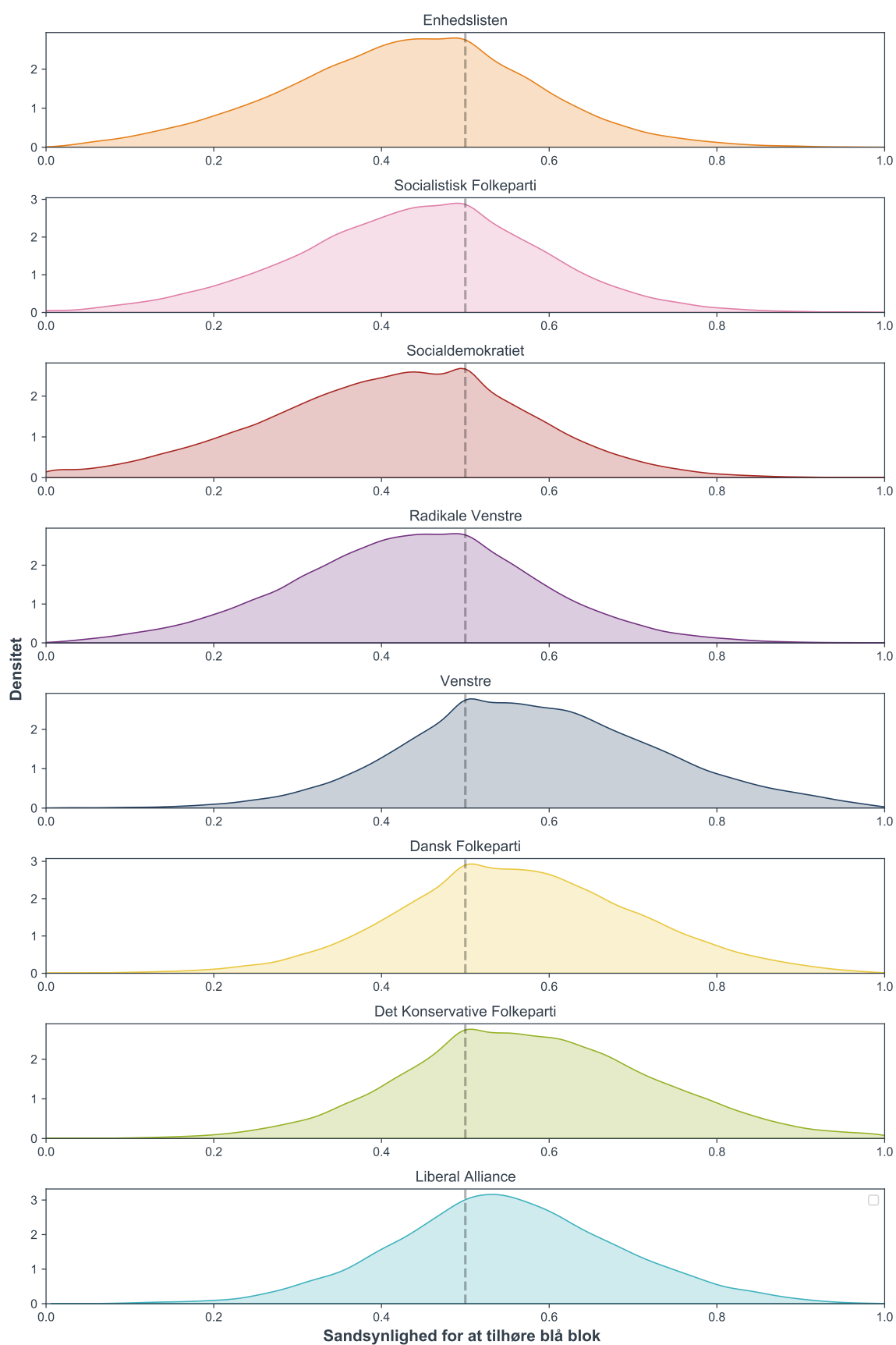
Politik- område	Tale	Taler/Sand- synlighed
Velfærd	Som Hr. Eigil Andersen helt korrekt refererer, påtænker jeg at gå nogle af reglerne på mit område efter i sømmene. Målet for mig er, at vi mindsker bureaukratiet og frigør ressourcer, så alle gode kræfter kan bruges på at hjælpe mennesker i job igen. Jeg kan oplyse Hr. Eigil Andersen om, at initiativerne er lige på trapperne, så Hr. Eigil Andersen må altså udvise en lille smule tålmodighed med hensyn til det nærmere indhold. Afslutningsvis vil jeg gerne sige, at det selvfølgelig er en præmis, at vi fastholder det overordnede mål med beskæftigelsespolitikken, nemlig at den skal bidrage til at få mennesker hurtigst muligt tilbage i job.	Inger Støjberg (96,4 pct)
Skat	Jeg var selv finansminister – at nu skal vi altså også lige have et øje for, hvad det er for en løndannelse, der er i det danske samfund. så jeg tager absolut ikke let på det. Det ved jeg også at Fru. Margrethe Vestager ikke gør, og derfor tror jeg da også, at Fru. Margrethe Vestager og jeg jo i virkeligheden er enige om, at det ville være meget ærgerligt, hvis de tanker, man gør sig i den anden del af oppositionen – om at lave ændringer af vores skattesystem, der trækker skævt i forhold til det her ved at øge byrderne på erhvervslivet og ved at øge topskatten i det her samfund – blev til virkelighed.	Lars Løkke Rasmussen (99,1 pct)

Udlændinge	Det er indlysende. Det er også indlysende, at der, hvor vi ser nogle af de her meget radikaliserede miljøer, jo netop er i nogle af parallelsamfundene. Det er jo også indlysende. Hvis det så er, at man har en tilstrømning til danmark, som alt andet lige gør, at der på sigt vil ske det, at de nuværende ghettoområder bliver større eller der vil komme nye ghettoområder, så skabes der jo også grobund for, at der kommer nye parallelsamfund og kommer nye områder, hvor der vil kunne opstå radikaliserede miljøer.	Martin Henriksen (99,9 pct)
Klima	Vindmøller, hvor danmark har en verdensmarkedsandel på 35-40 pct., har et stort vækstpotentiale. Der er forventning om en mangedobling af den globale vindkraftkapacitet frem mod 2030. Også markedet for miljøeffektiv teknologi til vandområdet vokser.	Eva Kjer Hansen (99,2 pct)

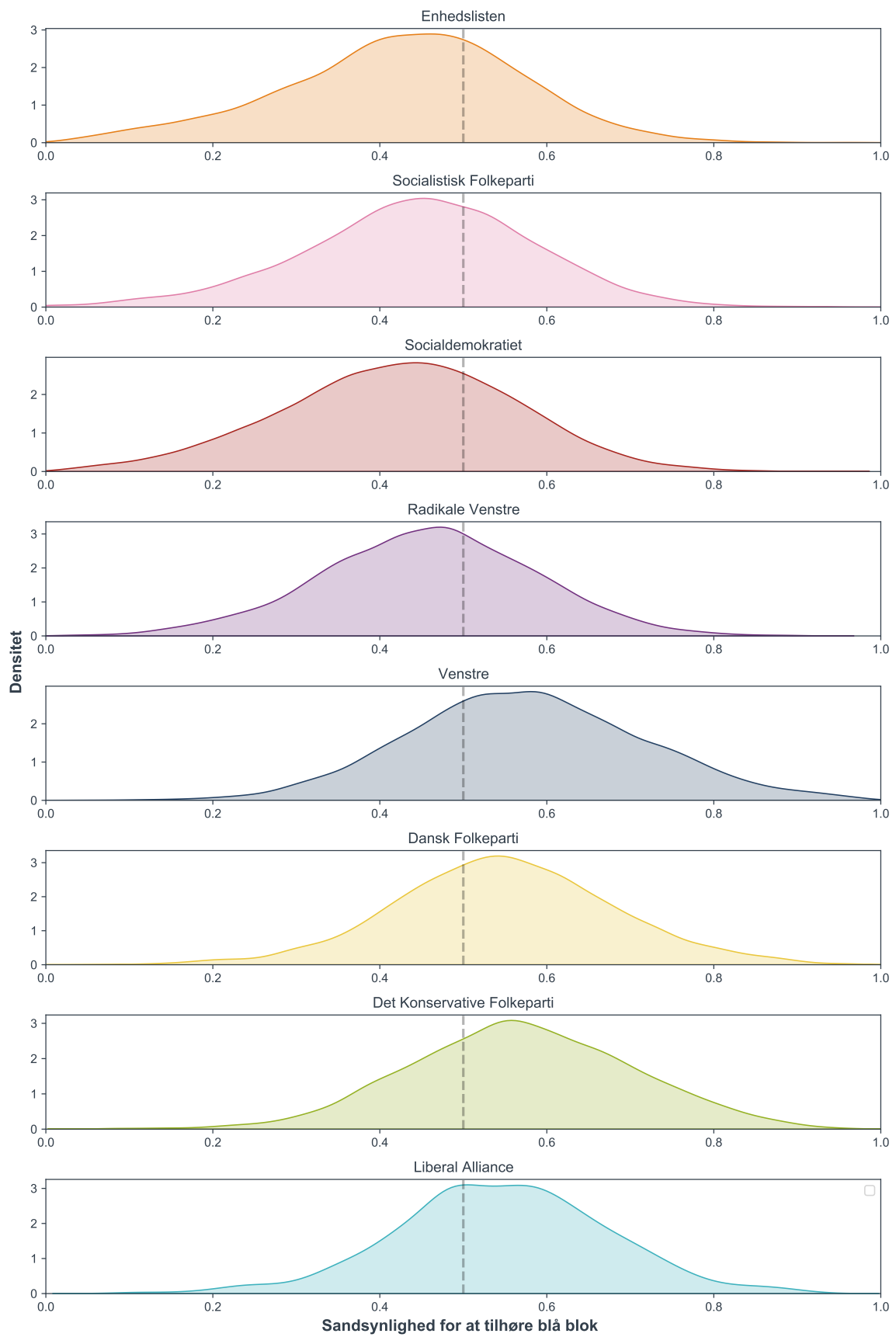
D.1.2 Fordelingen af sandsynligheder: Rød eller blå blok?

Nedenfor følger en række figurer, som viser fordelingen af sandsynligheder for de forskellige partiers taler i forhold til, om de tilhører blå blok. Lave sandsynligheder indebærer, at en tale tilhører rød blok og omvendt med de høje sandsynligheder. Der er både en figur for modellen, som måler elitepolarisering på tværs af alle politiske emner, og figurer for hvert af de værdi- og fordelingspolitiske emner.

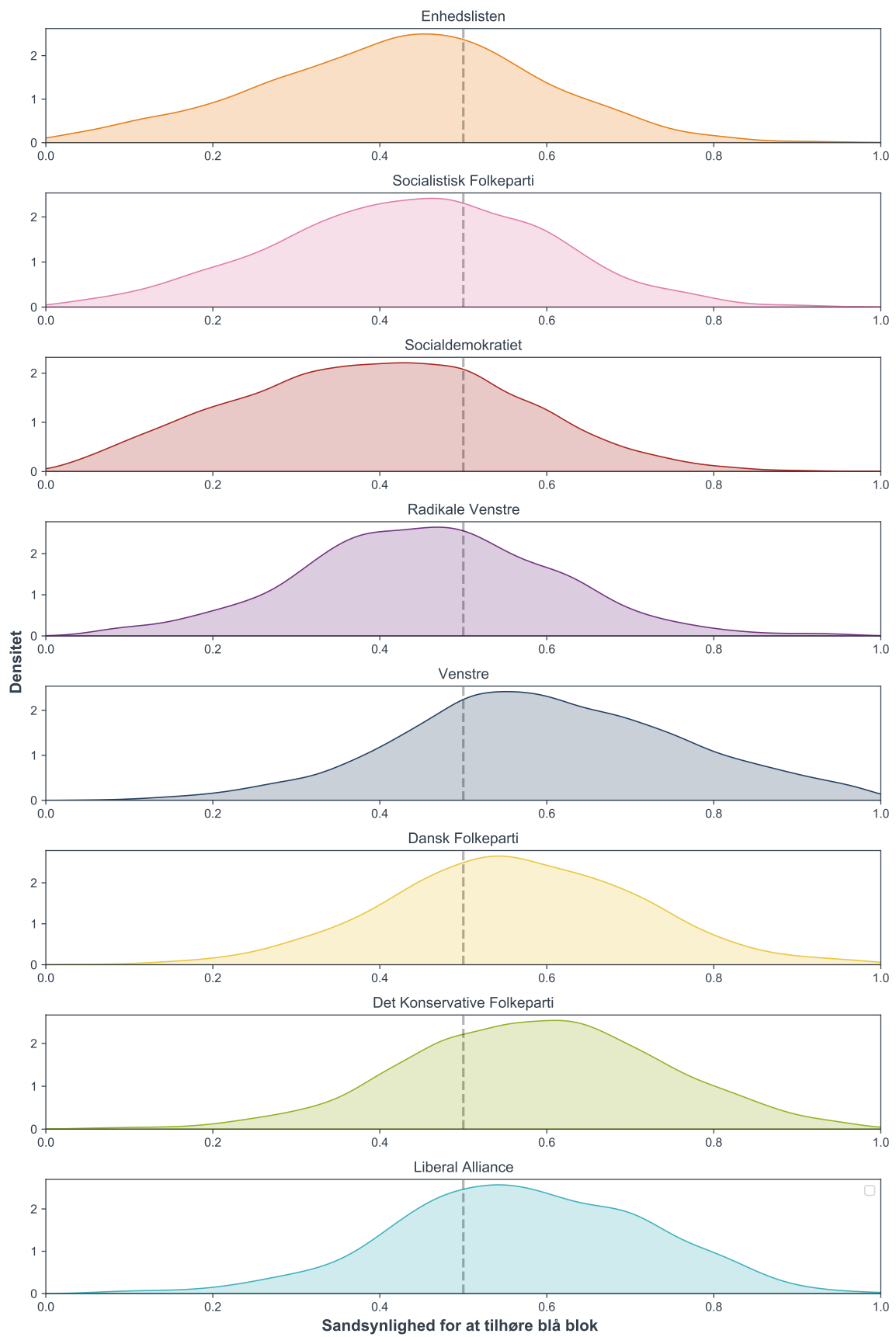
Figur D.1: Prædikterede sandsynligheder for alle taler



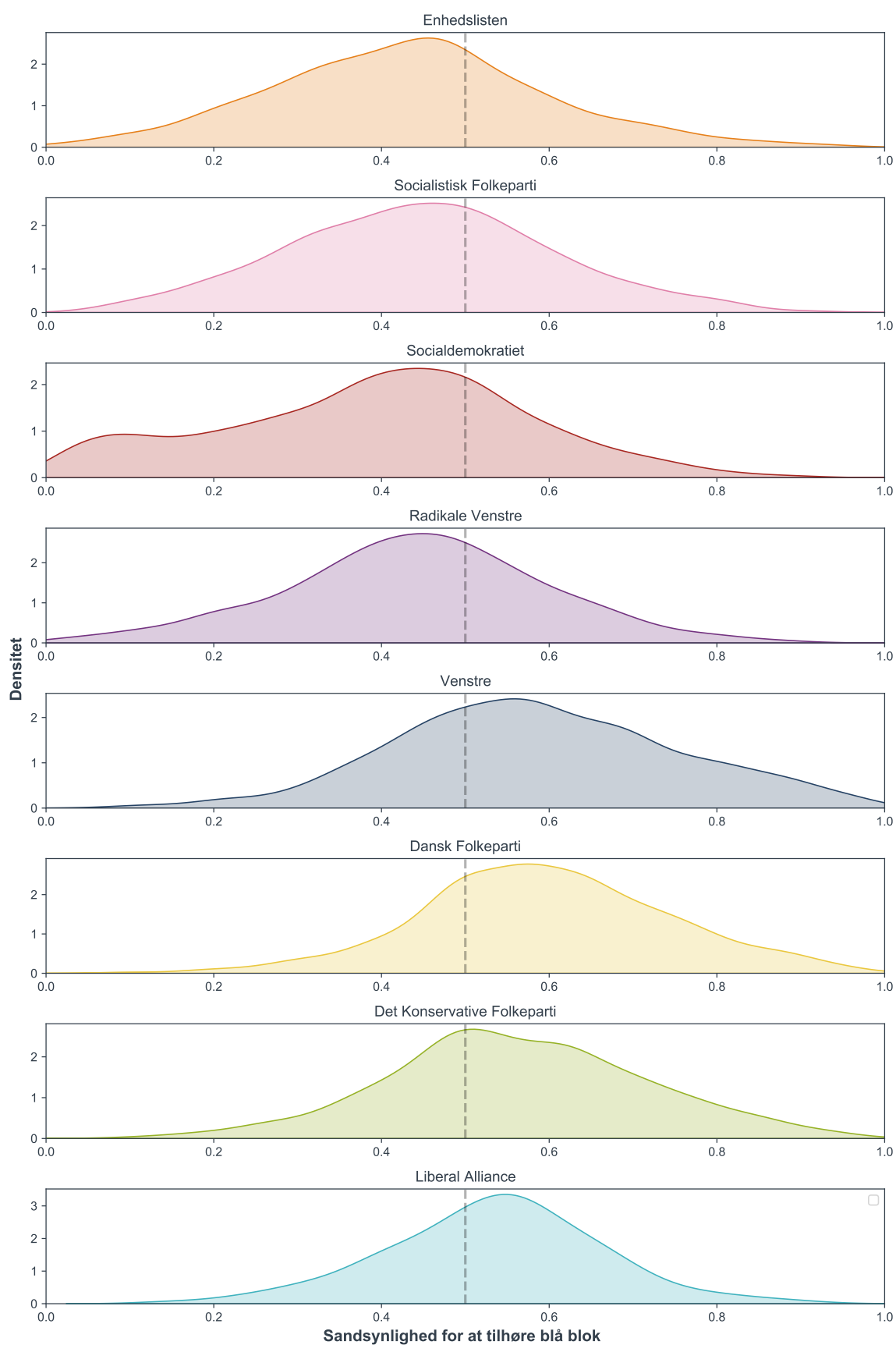
Figur D.2: Prædikterede sandsynligheder for velfærdspolitiske taler



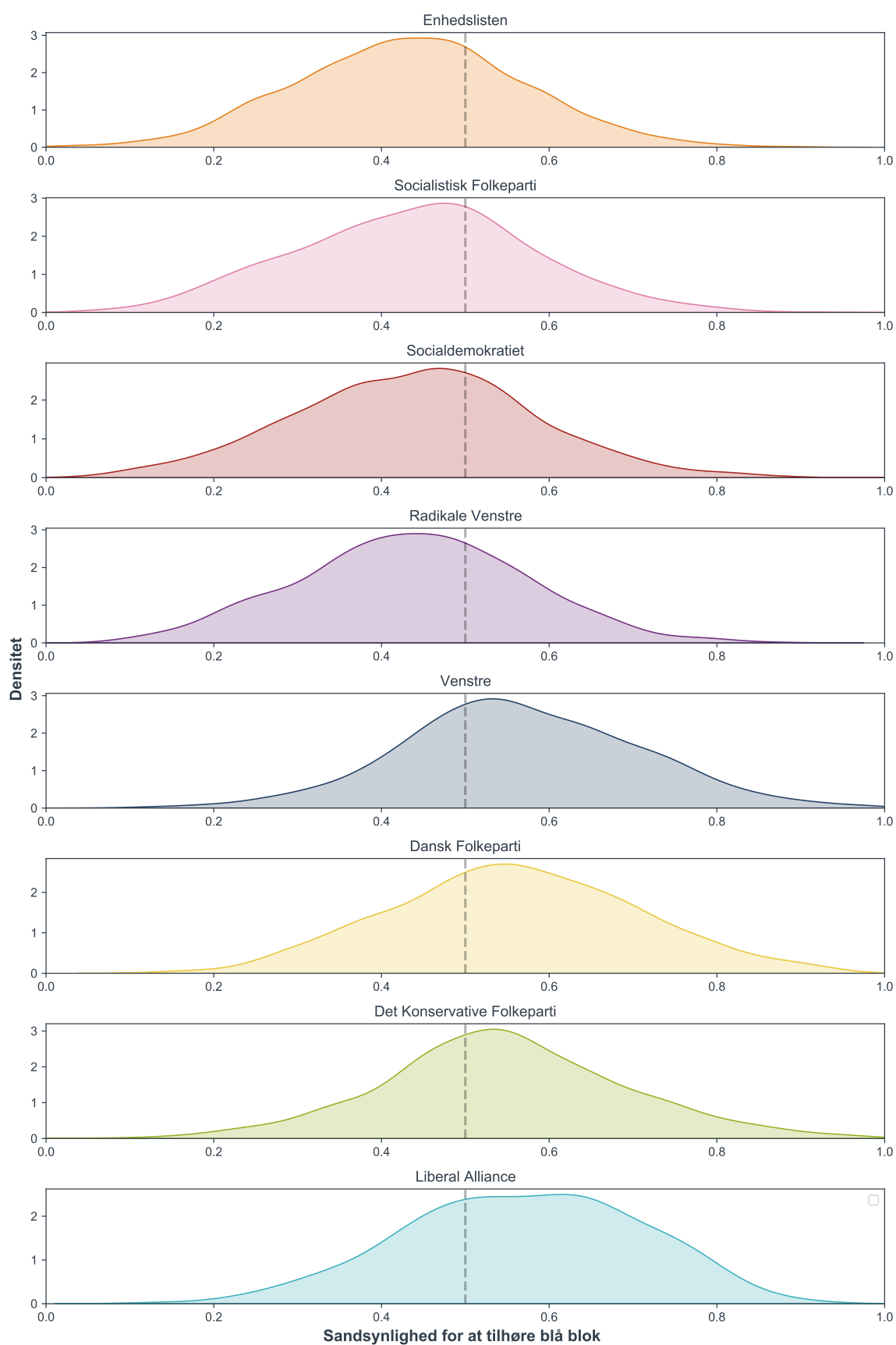
Figur D.3: Prædikterede sandsynligheder for skattepolitiske taler



Figur D.4: Prædikterede sandsynligheder for udlændingepolitiske taler



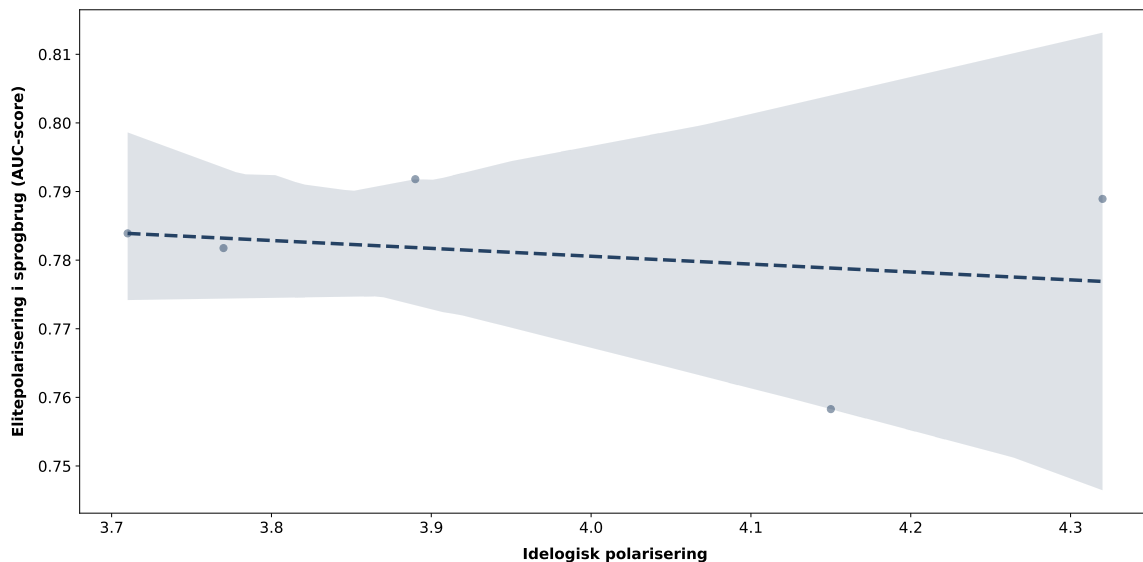
Figur D.5: Prædikterede sandsynligheder for klimapolitiske taler



D.2 Diskriminerende validering

Figuren nedenfor viser en korrelation mellem målet for elitepolarisering i sprogbrug mellem de politiske blokke på tværs af alle politiske emner og målet for ideologisk polarisering, som er baseret på CHES data. Sammenhængen er insignifikant, hvilket indikerer, at ideologisk polarisering og elitepolarisering i sprogbrug udgør to forskellige dimensioner af elitepolarisering ($r = -0.22, p < 0.71$). Det skal dog bemærkes, at korrelationen alene bygger på 5 observationer, da der alene er CHES-data for fem tidspunkter (1999, 2002, 2006, 2010 og 2014). Det kan derfor ikke udelukkes, at de insignifikante resultater skyldes de store konfidensintervaller som følge af det lave antal observationer. Imidlertid er udviklingen i ideologisk elitepolarisering nogenlunde stabilt med tendens til et margi- nalt fald (se Christiansen (2019b)). Resultaterne fra min analyse viser, at polariseringen i sprogbrug steg i førnævnte periode. Der er med andre ord tale om en modsatrettet udviklingstendens, hvilket understøtter, at der er tale om to forskellige dimensioner af elitepolarisering i sprogbrug.

Figur D.6: Korrelation mellem mål for elitepolarisering i sprogbrug og ideologisk elitepolarisering



Note: N = 5.

E Regeringer i dansk politik

Tabellen nedenfor giver et overblik over de regeringer, som har været ved magten i den periode, som min undersøgelse fokuserer på.

Tabel E.1: Oversigt over regeringer fra 1990 til i dag

Regering	År	Partier	Parlamentarisk situation
Poul Schlüter IV	1990-1993	V, K	Mindretalsregering
Poul Nyrup Rasmussen I	1993-1994	S, CD, RV, KF	Flertalsregering
Poul Nyrup Rasmussen II	1994-1996	S, RV, CD	Mindretalsregering
Poul Nyrup Rasmussen III	1996-1998	S, RV	Mindretalsregering
Poul Nyrup Rasmussen IV	1998-2001	S, RV	Mindretalsregering
Anders Fogh Rasmussen I	2001-2005	V, K	Mindretalsregering
Anders Fogh Rasmussen II	2005-2007	V, K	Mindretalsregering
Anders Fogh Rasmussen III	2007-2009	V, K	Mindretalsregering
Lars Løkke Rasmussen I	2009-2011	V, K	Mindretalsregering
Helle Thorning-Schmidt I	2011-2014	S, RV, SF	Mindretalsregering
Helle Thorning-Schmidt II	2014-2015	S, RV	Mindretalsregering
Lars Løkke Rasmussen II	2015-2016	V	Mindretalsregering
Lars Løkke Rasmussen III	2016-2019	V, LA, K	Mindretalsregering
Mette Frederiksen	2019-	S	Mindretalsregering