

KAPITEL 18

Meningsmålinger under valgkampen

Af Jens Olav Dahlgaard, Frederik Hjorth, Martin Vinæs Larsen & Asmus Leth Olsen

Valget i 2015 blev den foreløbige kulmination på en tendens mod stadig flere meningsmålinger under folketingsvalgkampe. Denne udvikling sker samtidig med fornyet interesse for, hvordan medierne formidler målinger, og hvorvidt målinger påvirker vælgerne. Udviklingen understøtter vigtigheden af, at der udvikles robuste metoder til løbende i valgkampen at give et samlet indtryk af målingernes tendens og usikkerhed. Vi præsenterer og evaluerer målingernes samlede præcision med afsæt i en bayesiansk model og diskuterer, hvordan målingernes systematiske undervurdering af Dansk Folkeparti kan undgås ved fremtidige folketingsvalg.

Flere og flere meningsmålinger – men hvad viser de?

Da Viggo Kampmann førte Socialdemokratiet til sejr ved folketingsvalget den 15. november 1960, skete det efter en valgkamp, der handlede om alt andet end meningsmålinger. Det havde den simple forklaring, at der under hele valgkampen kun blev bragt én eneste måling. Den blev foretaget af Gallup og offentliggjort den 28. oktober, som var den faste dag for Gallups månedlige politiske måling. Valgkampens resterende 18 dage måtte derfor gå med meget andet end at diskutere den seneste måling.

Anderledes forholdt det sig ved folketingsvalget i 2015. I perioden fra valgets udskrivelse den 27. maj til valgdagen den 18. juni er der registreret 7.492 artikler og indslag i Infomedia, som indeholder ordet *meningsmå-*

ling. Medtager vi ordet *måling*, som oftest anvendes synonymt, så ender den samlede omtale på 11.556 artikler.

Der var også mange målinger at rapportere om. De fleste institutter kørte rullende målinger, som betød, at partiernes tilslutning dagligt kunne opdateres med 300-400 nye respondents svar fra hvert institut. Tæller vi de daglige opdateringer som individuelle målinger, bød valgkampen på intet mindre end 118 meningsmålinger. Dertil kommer målinger foretaget af partier og interesseorganisationer til intern brug, og som derfor ikke blev offentliggjort.

Valget i 2015 er dermed den foreløbige kulmination på stigningen i antallet af meningsmålinger – ikke mindst i løbet af selve valgkampen. Denne udvikling har særlig taget fart siden 2001. En opgørelse fra Politiken har vist, at der i perioden 1957 til 2000 blev lavet lige så mange målinger, som fra 2001 til 2015 (Kulager & Jensen 2015). På samme måde viser tidligere opgørelser også en kraftig stigning i omtalen af meningsmålinger i medierne i det seneste årti, og at målinger udgør en stadig større del af den politiske dækning (Hansen 2014e). Tabel 18.1 viser en opgørelse af antallet af meningsmålinger, der blev offentliggjort under de seneste årtiers folketingsvalg. Her fremgår det tydeligt, hvordan 2001-valget indvarsler en ny æra med langt flere målinger for hver valgkamp.

Tabel 18.1: Antallet af meningsmålinger i valgkampen 1984-2015

| Valg | 1984 | 1987 | 1988 | 1990 | 1994 | 1998 | 2001 | 2005 | 2007 | 2011 | 2015 |
|-------|------|------|------|------|------|------|------|------|------|------|------|
| Antal | 5 | 8 | 7 | 13 | 17 | 18 | 42 | 61 | 81 | 126 | 118 |

Kilde: Meningsmålinger i medierne. Tallene er fra Politiken (2015), Madsen & Birkemose (2007), Søren Risbjerg Thomsen og egen optælling.

Antallet af målinger og omfanget af omtalen var dog ikke de eneste udviklingstendenser for meningsmålinger ved valget i 2015. Gennem de seneste år har der været en forskningsmæssig interesse for, hvordan medierne formidler meningsmålinger (Bhatti & Pedersen 2016) og en fornyet interesse for, hvordan meningsmålinger påvirker vælgernes partipræferencer (Dahlgaard m.fl. 2015; jf. også Simon 1954; Elklit 1989). Begge spørgsmål får fornyet relevans af et stort antal meningsmålinger. Samtidig er offentlighedens interesse for disse spørgsmål blevet øget i kølvandet på danske og internationale eksempler på, at meningsmålinger har ramt meget skævt eller er blevet formidlet fejlagtigt i medierne (Thomsen

2013). I dette kapitel beskriver vi folketingsvalget med afsæt i meningsmålingerne under valgkampen. Dette sker på baggrund af vores egen daglige opgørelse af det samlede billede af alle meningsmåling, som vi, kapitlets forfattere, løbende publicerede i valgkampen.

Kapitlet har følgende struktur: For det første diskuterer vi forskellen mellem at udtrykke målingernes samlede usikkerhed/troværdighed med afsæt i frekventistisk og bayesiansk statistik. Derpå præsenterer vi den underliggende model for dagligt opdaterede grafiske præsentationer af målinger, der forsøger at vise alle de publicerede målinger og angive deres fælles statistiske usikkerhed. Herefter viser vi en række grafiske eksempler fra modellen, og vi evaluerer meningsmålingernes samlede præcision. Afslutningsvis diskuterer vi, hvorfor Dansk Folkeparti blev systematisk undervurderet i målingerne, og hvordan vi kan undgå lignende bias i målingerne ved fremtidige valg.

Meningsmålingers statistiske usikkerhed

Når medier omtaler meningsmålinger, sørger de typisk for at påhæfte en vending om, at man skal tage forbehold for den statistiske usikkerhed. Man får ofte indtryk af, at der med dét udtryk henvises til alle mulige typer usikkerheder, der kan omsætte sig i en forskel mellem det målingen indikerer og virkeligheden. Men udtrykket henviser reelt kun til én type usikkerhed, nemlig den statistiske usikkerhed. Som det vil fremgå længere fremme i kapitlet, illustrerede folketingsvalget 2015, at der er andre usikkerheder forbundet med meningsmålinger end den rent statistiske.

Det særlige ved den statistiske usikkerhed er imidlertid, at vi (med en given sikkerhed) kan angive præcist, hvor stor den er, hvis bestemte antagelser er overholdt. Så længe en stikprøve er foretaget ved simpel tilfældig udvælgelse, ved vi nøjagtigt, hvor meget variation i partiernes opbakning selve stikprøvetagningen vil medføre. Groft sagt lader det sig gøre på to forskellige måder: en frekventistisk og en bayesiansk. Ofte vil de to metoder give identiske resultater, men der er dog enkelte forskelle mellem dem, som er relevante for både vores model, og den måde medier formidler målinger på. Derfor skal de kort forklares her.

Af forskellige årsager har den frekventistiske metode været den mest anvendte og er det stadig. Den frekventistiske metode antager, at der er en sand, men uobserverbar, opbakning til hvert af partierne ude i virke-

ligheden. Hvis vi foretager gentagne stikprøver (meningsmålinger), får vi en række gæt på den sande opbakning til partierne. Hvis den sande opbakning til Socialdemokratiet fx er 25 pct., vil cirka 95 pct. af vores stikprøver ifølge den underliggende statistiske teori give Socialdemokratiet mellem 22,3 og 27,7 pct. af stemmerne, forudsat at hver måling er baseret på et tilfældigt udsnit af vælgerne.

Med den frekventistiske metode kan man danne et konfidensinterval om punkttestimatet. Konfidensintervaller er dannet på en måde, sådan at hvis man foretager virkelig mange målinger og for hver af dem konstruerer et konfidensinterval, vil 95 pct. af intervallerne indeholde den sande opbakning til partiet. I almindelighed angiver meningsmålingsinstitutter og medier, at konfidensintervallet med 95 pct. sandsynlighed indeholder den faktiske opbakning. Det er strengt taget en forkert fortolkning af frekventistiske konfidensintervaller, som ikke har belæg i den statistiske litteratur, omend den er udbredt selv i akademiske kredse (Hoekstra m.fl. 2014). I stedet er et konfidensinterval et interval skabt af en procedure, der, hvis vi hypotetisk gentog den et stort antal gange, i 95 pct. af tilfældene ville give et interval indeholdende det sande parameter. Konfidensintervallets 95 pct. knytter sig altså til proceduren og ikke til det enkelte interval (Hoekstra m.fl. 2014). Den udbredte fortolkning er dog bekvem og let forståelig, og i mange tilfælde er den tillige en god tilnærmelse, og derfor vil vi ikke forfølge denne unøjagtighed. Vi vil dog påpege, at den udbredte fortolkning er dækkende for bayesianske konfidensintervaller også kaldet *troværdighedsintervaller*.

Ved den bayesianske metode begynder man med at gøre sig klart, hvad man tror fx Socialdemokratiets opbakning er, inden man efterfølgende opdaterer den overbevisning med data fra en måling. Når man har indsamlet sin stikprøve, spørger man således, hvad sandsynlige niveauer for Socialdemokratiets opbakning er. På den måde adskiller tilgangen sig fra den frekventistiske ved ikke at se målingerne, men overbevisningen om partiernes opbakning som det varierende. Den store fordel ved den bayesianske metode er dermed, at den nemt og fleksibelt lader én bruge forudgående viden. Hvis man er meget sikker i sin overbevisning, lader man denne få stor vægt, og der skal meget data til at flytte den. Er man på usikker grund, lader man sig lede af målingen. Ved at kombinere sin forudgående viden med den nye viden, man får adgang til, kan man således skabe en opdateret overbevisning. Det kan man bruge til at skabe et baye-

siansk troværdighedsinterval, som man er 95 pct. sikker på, at den sande værdi befinder sig inden for.

Når vi bliver stillet over for en måling af partiernes opbakning, har vi på forhånd stor viden om denne. Sådanne målinger er derfor velegnede til brug af bayesianske teknikker. Man kan illustrere forskellen mellem den frekventistiske og bayesianske tilgang med et eksempel. Lad os sige, at vi foretager en måling med cirka 1000 interviews, der viser, at Socialdemokratiet står til 30 pct. af stemmerne. Frekventisten konstruerer på denne baggrund et konfidensinterval, der går fra 27 pct. til 33 pct. Bayesianeren begynder med at se på, hvad nylige målinger har sat Socialdemokratiet til. Lad os sige, at de målinger i gennemsnit giver partiet 25 pct. af stemmerne med en usikkerhed på plus/minus 1,5 pct. Når bayesianeren vægter målingerne sammen (jf. nedenfor), vil partiet ifølge hendes beregninger få 26 pct. med en usikkerhed på 1,3 pct. Havde hun været mindre sikker på de forudgående målinger, var skuddet havnet tættere på 30 pct. med større samlet usikkerhed.

Fordele ved den bayesianske metode

Der er flere fordele ved at anvende den bayesianske metode frem for den traditionelle, frekventistiske. For det første generer metoden troværdighedsintervaller med den egenskab, at de under den valgte sammenvægtning af forudgående og ny information indeholder de centrale 95 pct. af fordelingen af sandsynlige værdier for opbakningen til partierne. En egenskab der ofte fejlagtigt tillægges frekventistiske konfidensintervaller.

Mens den frekventistiske metode ser fuldstændig bort fra forudgående viden og lader som om, at vi ingenting ved om partiernes nuværende opbakning inden den nyeste måling, danner den bayesianske metode et kompromis mellem, hvad vi ved i forvejen, og hvad vi erfarer med den seneste måling. Udover at der er noget intuitivt appellerende i denne måde at anskue verden på, kan fremgangsmåden også bidrage til at reducere antallet af vilde sensationshistorier baseret på enkelte, afvigende målinger. Tidligere målinger vil fungere som et anker, der holder snor i tilfældige vildskud. Det klart mest iøjnefaldende eksempel er DR's ualmindeligt skæve *exit-prognose* fra kommunalvalget i 2013 (Thomsen 2013). Havde man dengang brugt eksisterende målinger fra dagene inden kommunalvalget til at kvalificere prognosens resultater, var man nok ikke kommet helt så galt af sted.

For feinschmeckerne er en anden fordel ved den bayesianske metode, at den underliggende estimationsteknik ikke giver symmetriske troværdighedsintervaller i modsætning til mere klassiske metoder. Særligt for små partier tæt på spærregrænsen gør det en forskel. Lad os sige, at et parti ligger til 3 pct. i en måling med 1000 personer. Den måde, som meningsmålingsinstitutter rutinemæssigt beregner konfidensintervaller på, giver et interval fra 1,9 til 4,1. Ret nervepirrende for partiet på grænsen. Man foranlediges måske til at tro, at det er omtrent lige sandsynligt at ramme over som under denne grænse. Det er det bare slet ikke. Faktisk afslører den bayesianske teknik, at der er omkring 4 pct. sandsynlighed for, at partiet står til mere end 4,1 pct. og kun 1 pct. til, at det står til mindre end 1,9 pct. Teknikken giver et troværdighedsinterval fra 2,1 pct. til 4,2 pct. med den fine egenskab, at der er 2,5 pct. sandsynlighed for, at partiet ligger over såvel som 2,5 pct. sandsynlighed for, at det ligger under intervallet. Denne krølle er sådan set ikke et slagsmål mellem bayesianske og frekventistiske teknikker. Det er nok et levn fra de dårlige gamle dage, hvor computerkraft var dyr og svær at komme i nærheden af. I sådanne tilfælde er standardmetoden til beregning af det frekventistiske interval væsentligt hurtigere. Det hører heldigvis fortiden til, og på en moderne computer tog det knap 12 sekunder for os at beregne usikkerheder omkring ni partier med en præcision til anden decimal med en relativt tung simuleringsmetode.

I relation til meningsmålinger skal man ikke overvurdere fordelene ved den bayesianske metode sammenlignet med brugen af vægtede gennemsnit. Således vil gennemsnittet af flere målinger lande tæt på det bayesianske estimat. Man kan dog som ovenfor argumentere for, at den er bedre til at begrænse indflydelsen af sensationalistiske målinger. Det vil sige målinger, der stikker ud i forhold til den generelle tendens. Ofte bliver disse af medier og eksperter fortolket som reelle vandringer, selv om de oftere bare vil være tilfældige udsving i. Den bayesianske metode udtrykker eksplicit, at vi skal tage anden information i betragtning og vil derfor nedvægte denne information. Igen vil man få samme effekt ved vægtede målinger, men det er måske mere eksplicit udtrykt ved den bayesianske model. Endelig kan man argumentere for, at vægtede målinger i bund og grund er bayesianske i deres natur. Fordelene ved bayesianske metoder er samtidig udtalte i andre og lidt mere komplicerede sammenhænge. Det gælder fx, hvor man ønsker at vægte kvalitativ og kvantitativ information sammen (Humphreys & Jacobs 2015).

En model til aggregering af valgkampens målinger

Uanset hvordan den statistiske usikkerhed angives, indikerer mange meningsmålinger bevægelser i partiernes opbakning, der er betydeligt mindre end denne. Dermed er deres informationsværdi om reelle ændringer i partiernes styrkeforhold ganske lille. På trods af det viser forskning i dansk mediedækning af meningsmålinger, at journalister meget ofte tolker på baggrund af bevægelser inden for den statistiske usikkerhed. Bhatti og Pedersen (2016) finder i en stikprøve af 672 danske avisartikler om meningsmålinger, at hele 67 pct. tolker på bevægelser, der er mindre end den statistiske usikkerhed. Journalister er ganske vist gode til at komme med retoriske forbehold, når de skal tolke på meningsmålinger – men i realiteten er de ret uforsigtige.

Op til valget blev der oven i købet offentliggjort adskillige daglige målinger. Det kunne bidrage yderligere til forvirringen, hvis formiddagens måling gav et lille forspring til rød blok, mens eftermiddagsmålingen havde blå blok i front. Især hvis begge historier blev krydret med potente overskrifter og uforsigtige konklusioner. Derfor udgav dette kapitels forfattere et dagligt plot, der forsøgte at samle alle de publicerede målinger og angive deres fælles statistiske usikkerhed. Det gjorde vi på baggrund af en statistisk model, der kombinerede alle nye målinger med information fra tidligere målinger. En liste med alle de målinger, der indgår i vores plots, er vist i tabel 18.2.

Tabel 18.2: Målinger i vores opgørelser over valgkampen

| Institut | Offentliggjort i | Antal målinger |
|----------|------------------|----------------|
| Epinion | DR | 23 |
| Gallup | Berlingske | 22 |
| Greens | Børsen | 21 |
| Megafon | TV 2/Politiken | 8 |
| Norstat | Norstat | 1 |
| Voxmeter | Ritzau | 22 |
| Wilke | Jyllands-Posten | 18 |
| YouGov | Metroxpress | 3 |

Note: Perioden 27. maj til 17. juni. Bemærk at antal målinger angiver antallet af offentliggjorte partifordelinger. Da de fleste institutter benytter rullende målinger, så er der et overlap af respondenter mellem målinger.

Konkret tog vi hver dag alle offentliggjorte målinger på internettet og vægtede dem sammen ud fra, hvor mange interviews målingen var baseret på. I realiteten svarer det til, at vi lagde målingerne sammen. Man kan forestille sig, at tre målinger med henholdsvis 3.000, 2.000 og 1.000 interviews bliver offentliggjort. Vi samlede blot data, så det svarede til én måling med 6.000 respondenter. Utroligt nok er det ikke altid, at antallet af interviews bliver rapporteret i medierne. Man skulle ellers mene, at læserne har et berettiget krav på at vide, om 100 eller 2.000 er blevet interviewet. I de tilfælde satte vi målingens størrelse til 1.000 interviews, hvilket omtrent svarer til den gængse størrelse på en meningsmåling.

Det næste skridt er at kombinere den nye viden med det, vi vidste i forvejen. Når man nu har adgang til en række tidligere målinger, er det oplagt at bruge et eller andet gennemsnit af dem. Spørgsmålet er hvilket. Den første begrænsning i valgkampen var, at mange institutter brugte rullende målinger over tre dage. Det vil sige, at de hver dag interviewede omkring 1000 personer, som de lagde oven i de omkring 2000 personer, de havde interviewet de foregående to dage. Det er for så vidt en fornuftig strategi, der bidrager til at reducere tilfældige svingninger fra dag til dag, som kunne blæses op til noget, de ikke er. Men vi ville tælle to tredjedele af vores data to gange, hvis vi brugte målingen fra dagen forinden til vores samlede målinger. For at undgå dette problem brugte vi den seneste måling, der var mindst tre dage gammel som forudgående information. Vi så bort fra information, som var mere end tre dage gammel.

Dernæst skulle vi afgøre, hvor meget den gamle information skulle fylde i forhold til den nye. I en valgkamp er det en rimelig antagelse, at partiernes opbakning ændrer sig hurtigere end ellers. Vælgerne udsættes i højere grad end ellers for indtryk, og de må samtidig forventes at bruge mere tid på at overveje, hvem de vil stemme på i stedet for bare at vælge det, de plejer. Samtidig kom der i folketingsvalgkampen konstant megen ny information, hvilket reducerede behovet for tidligere information. Vi lod derfor den gamle måling indgå med en fjerdedel af dens oprindelige størrelse. For den første dag i valgkampen brugte vi det vægtede gennemsnit fra Politiko.dk vægtet som 750 interviews. På dage, hvor et institut ikke offentliggjorde en ny måling, brugte vi deres seneste måling med en vægt på tre fjerdedele, hvis den var en dag gammel, en halv, hvis den var to dage gammel, og en fjerdedel, hvis den var tre dage gammel. I de tilfælde betød det også, at alle målinger før den seneste røg ud, da vi som

beskrevet ovenfor lagde et tre dages hul ind mellem hver måling og ikke så på målinger, der var mere end tre dage gamle.

Denne afskrivning svarer til, at man på tre dage nogenlunde fordobler den statistiske usikkerhed. For at vende tilbage til det løbende eksempel med Socialdemokratiet kan det være, at man til at starte med tror, at deres opbakning er mellem 22,5 og 27,5 pct. Tre dage senere tror man, i fravær af ny information, at den er mellem 20 og 30. Så hurtigt går afskrivningen ikke imellem valgkampe, men som nogle af figurerne nedenfor viser, kan der ske en ganske markant udvikling, når der er valgkamp. Andre vægtede gennemsnit, som fx Søren Risbjerg Thomsens for Altinget kvalitetsvægter målingerne, ud fra tidligere præcision. Vi valgte ikke at kvalitetsvægte. I en fremtidig model vil det være en oplagt udvidelse.

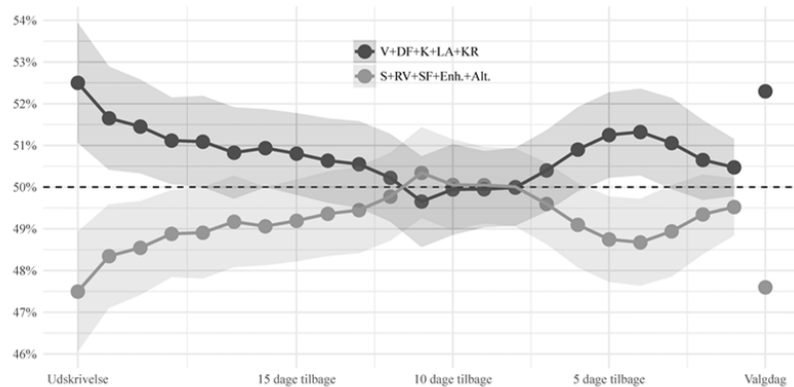
Når vi kombinerer, hvad vi ved i forvejen med den nye information, får vi et bedre bud på partiernes og blokkenes opbakning og en angivelse af den samlede usikkerhed i målingerne. Fordelen ved metoden er altså, at den for det første samler målinger, der svinger naturligt, til et fælles bud. For det andet nøjes den ikke med at danne et interval om en enkelt værdi, den kombinerer også informationen fra målingerne til at give et bud på den samlede usikkerhed.

Hvor præcise var meningsmålingerne?

I denne generelle diskussion om målingers præcision (eller mangel på samme) har meningsmålinger frem mod et folketingsvalg en særlig status, idet de kan sammenlignes direkte med valgresultatet, som i sagens natur kendes med fuldstændig sikkerhed. Hvordan klarede målingerne sig så ved folketingsvalget 2015?

I figur 18.1 har vi angivet vores gennemsnit af blokkenes opbakning i meningsmålingerne for hver af valgkampens dage. Rød blok er defineret som Socialdemokratiet, Radikale Venstre, Socialistisk Folkeparti, Enhedslisten og Alternativet. Blå blok er partierne Venstre, Dansk Folkeparti, Det Konservative Folkeparti, Liberal Alliance og Kristendemokraterne. Som figuren viser, er meningsmålingernes præcision i at forudsige valgresultatet i bedste fald blandet. Frem mod valgdagen pegede målingerne på en valgsejr til blå blok, svarende til resultatet. Men hvor målingerne viste et ret tæt løb, endte blå blok med en ganske solid valgsejr på i alt 52,3 pct. af stemmerne. Som det også fremgår af figur 18.1, lå det faktiske valgresultat

Figur 18.1: Gennemsnit af blokkenes opbakning i meningsmålingerne for hver af valgkampens dage



Note: Prikken på valgdagen angiver valgresultatet. De grå bånd angiver 95 pct. troværdighedsinterval, omtrent svarende til den statistiske usikkerhed.

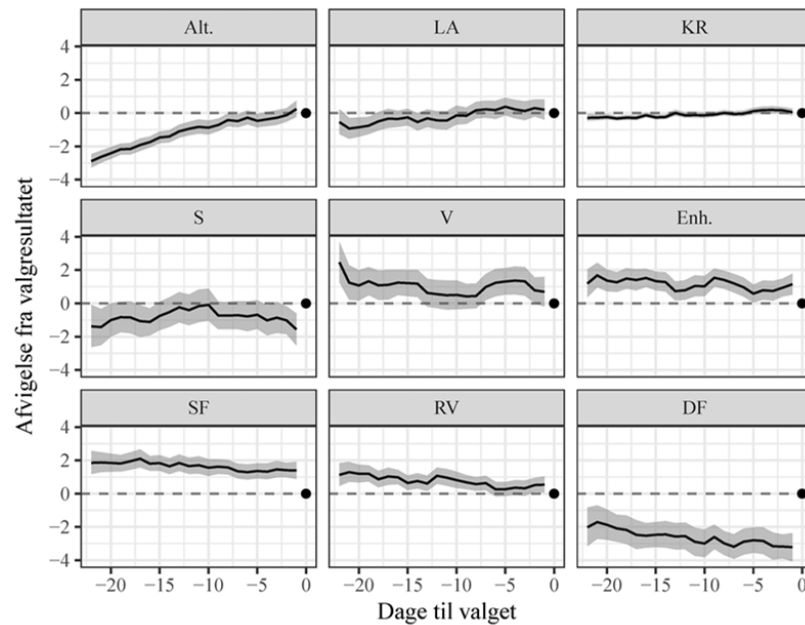
tat langt uden for den statistiske usikkerhed forbundet med målingerne. Afvigelsen skyldes altså ikke bare tilfældig stikprøvevariation, men en systematisk fejl i meningsmålingerne.

Vi kan komme fejlkilden lidt nærmere ved at se på meningsmålingernes præcision parti for parti. I figur 18.2 har vi opgjort vores aggregering af målingerne for hvert enkelt parti og den opbakning, partiet endte med at få på valgdagen. Som figuren viser, klarede målingerne sig for mange partiers vedkommende udmærket: Målingerne fangede for eksempel Venstres tilbagegang under valgkampen og Alternativets store fremgang. Men for især ét parti ramte målingerne helt skævt: Målingerne undervurderede Dansk Folkepartis opbakning med omkring 3 procentpoint – en forskel klart uden for den statistiske usikkerhed.

Fejlskønnet af Dansk Folkepartis opbakning har stor betydning for målingernes samlede præcision. Vi kan opgøre betydningen ved at se på målingernes samlede *root mean square error* eller RMSE, som angiver målingernes afvigelse fra valgresultatet for det gennemsnitlige parti. RMSE udregnes med formlen:

$$\sqrt{\frac{\sum_{i=1}^n (\pi_i - \pi_{15i})^2}{n}}$$

Figur 18.2: Aggregering af målingerne for hvert enkelt parti

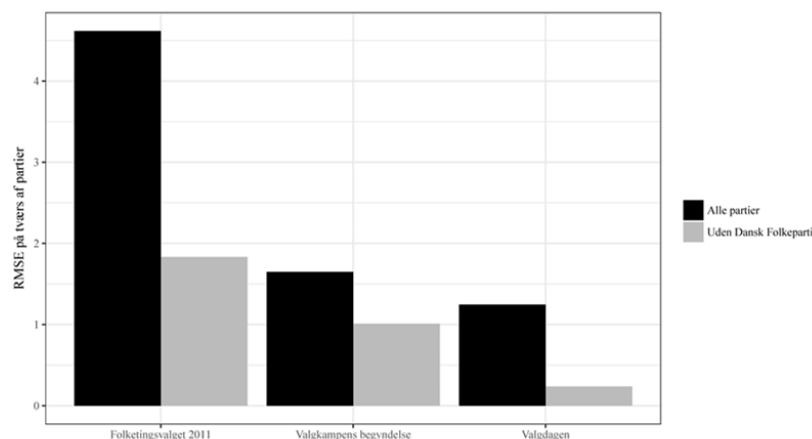


Note: Prikken på valgdagen angiver valgresultatet. Y-aksen angiver partiets målingers afvigelse fra valgresultatet. De grå bånd angiver 95 pct. troværdighedsinterval, omtrent svarende til den statistiske usikkerhed.

I formelen angiver π_i parti i 's opbakning i en måling, og π_{15i} angiver partiets valgresultat. RMSE udtrykker altså for en given prognose, hvor meget den samlede set afviger fra det faktiske valgresultat.

En måde at vurdere meningsmålingernes præcision er ved at sammenligne RMSE for forskellige typer prognoser. I figur 18.3 præsenterer vi RMSE på tværs af partier for tre typer prognoser: én kun baseret på valgresultatet i 2011 (svarende til ikke at bruge nogen information fra målinger overhovedet), én baseret på målingerne ved valgkampens begyndelse og én baseret på målingerne på valgdagen. For hver af de tre måder viser figur 18.3 også, hvor høj RMSE ville være, hvis man så bort fra målingerne af Dansk Folkepartis opbakning.

Figur 18.3: Fejlprocenter på tværs af partier opgjort for tre typer prognoser



Note: Root mean square error (RMSE) baseret kun på valgresultatet i 2011, baseret på målinger ved valgkampens begyndelse og baseret på målinger på valgdagen. De sorte søjler angiver RMSE for alle partier. De grå søjler angiver RMSE, hvis man ser bort fra målingerne vedr. Dansk Folkepartis opbakning.

Som vist i figur 18.3 er RMSE for gennemsnittet af meningsmålinger på valgdagen 1,25, som altså angiver, at målingerne ramte partiernes opbakning forkert med 1,25 procentpoint for det typiske parti. Heraf tegner fejlskønnet af Dansk Folkepartis opbakning sig alene for et helt procentpoint, svarende til 80 pct. af den samlede afvigelse. Målingerne klarede sig altså samlet set mindre godt, men man kan med god ret sige, at de med undtagelse af Dansk Folkeparti faktisk forudså valgresultatet med stor præcision. Endvidere giver de et meget bedre billede, end hvis vi havde brugt seneste valgresultat eller målinger fra valgkampens start.

Hermed ligner det overordnede billede fra valget 2015, hvad der er fundet i systematiske gennemgange af målinger ved tidligere folketingsvalg (Thomsen 1995; 1998; Thomsen 2002). I tabel 18.3 kan vi se fejlprocenter fra gennemsnit af meningsmålinger og endeligt resultat i perioden 1990 til 2011 (Thomsen 2013). Fejlprocenten beregnes ved at tage halvdelen af afvigelserne for alle partier mellem måling og landsresultatet. I 2015 var fejlprocenten 4,5 pct., hvilket er en smule over gennemsnittet af fejlprocenter siden 1990.

Tabel 18.3: Fejlprocenter for gennemsnit af målinger ved folketingsvalg 1990-2015

| Valg | 1990 | 1994 | 1998 | 2001 | 2005 | 2007 | 2011 | 2015 |
|-------------|------|------|------|------|------|------|------|------|
| Fejlprocent | 3,3 | 2,1 | 6,1 | 3,3 | 2,1 | 1,1 | 3,2 | 4,5 |

Note: Fejlprocenter fra 1990 til 2011 er her med udgangspunkt i Søren Risbjerg Thomsens kvalitetsvægtede gennemsnit af målinger (Risbjerg 2013). Fejlprocenten fra 2015 er fra egen model.

Simple justeringer er ikke nok til at undgå fejl i fremtiden

Indtrykket af målingernes præcision i 2015 aftvinger selvfølgelig spørgsmålet: Hvorfor ramte målingerne Dansk Folkepartis opbakning så skævt? Selvom vi ikke i skrivende stund kan forklare fejlskønnet med stor sikkerhed, illustrerer det nogle generelle tendenser, der vil have stor betydning for meningsmålingers præcision i fremtiden.

Som beskrevet ovenfor er et af de mest markante kendetegn ved 2015-målingerne, at de systematisk undervurderede Dansk Folkeparti under hele valgkampen. Det samme gjorde alle exit polls på valgdagen. Undervurderingen af Dansk Folkeparti ved valget i 2015 kan skyldes to ting. Enten at meningsmålingsinstitutterne ikke fik fat i Dansk Folkepartis vælgere, såkaldt *selektionsbias*, eller at meningsmålingsinstitutterne ikke formåede at få Dansk Folkepartis vælgere til at indrømme, hvad de ville stemme, *social desirability bias*.

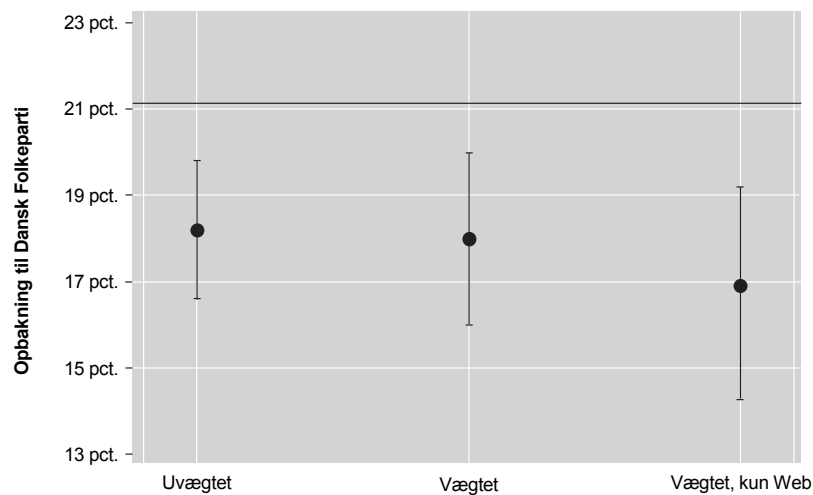
I diskussionen om hvorfor Dansk Folkeparti blev undervurderet, har sidstnævnte fået størst opmærksomhed. Som argument er det blevet anført, at vælgerne skulle være pinligt berørte over at stemme på et parti, der jævnligt bliver beskyldt for at være racistisk, populistisk eller, som tidligere statsminister Poul Nyrup Rasmussen formulerede det, ikke stuerent.

Et umiddelbart argument imod, at vælgerne skulle finde det pinligt at stemme Dansk Folkeparti, er, at meningsmålingerne i 2011 faktisk forudsagde partiets opbakning nogenlunde præcist. Noget tyder altså på, at vælgerne ikke fandt det pinligt at sige de stemte Dansk Folkeparti i 2011, så hvorfor skulle de finde det pinligt i 2015? Man kan dog forestille sig at de mange nye vælgere, der aldrig havde stemt Dansk Folkeparti før, fandt det pinligt at indrømme, at de nu stemte på partiet. I sidste ende er det imidlertid et empirisk spørgsmål, om det var social desirability bias eller selektionsbias der var på spil.

Det er imidlertid ikke så let at finde ud af, hvilken af disse to former for bias, der rent faktisk var på spil ved valget i 2015, og det er heller ikke let at justere sig ud af disse to fejlkilder, når man først har lavet en meningsmåling. Udfordringerne ved at skelne mellem og justere sig ud af de to former for bias kan illustreres ved at undersøge besvarelserne i Valgundersøgelsen, som i sig selv er en meningsmåling.

I figur 18.4 ser vi på forskellige bud på opbakningen til Dansk Folkeparti, som kan estimeres på baggrund af Valgundersøgelsen. Længst til venstre ser vi på andelen i stikprøven, der peger på Dansk Folkeparti. Vi kan her se, at respondenterne i Valgundersøgelsen er mindre tilbøjelige til at sige, de stemmer Dansk Folkeparti, end den danske befolkning som helhed var til at sætte deres kryds ved partiet. Præcis ligesom vi så i meningsmålingerne før valget. Det skal dog nævnes, at stikprøveudvælgelsen er meget anderledes i Valgundersøgelsen end de kommercielle meningsmålinger, så der er flere faktorer, som påvirker fordelingen i Valgundersøgelsen, også for Dansk Folkeparti (se kapitel 1).

Figur 18.4: Estimeret opbakning til Dansk Folkeparti på baggrund af Valgundersøgelsen



Note: Det uvægtede estimat tager udgangspunkt i andelen i stikprøven, der sagde, at de stemte på Dansk Folkeparti. Det vægtede estimat vejer stikprøven på baggrund af køn, alder, region og uddannelse. Det sidste estimat er beregnet som det vægtede, bortset fra at respondenter, der ikke afgav besvarelser over internettet, er sorteret fra. Den horisontale mørke linje er Dansk Folkepartis faktiske opbakning.

Billedet bliver imidlertid ikke klarere, når vi forsøger at justere for eventuel selektionsbias eller social desirability bias. I midten af figur 18.4 viser vi således et estimat af, hvem der stemmer Dansk Folkeparti vægtet efter køn, alder, region og uddannelse. Dette burde delvist afhjælpe selektionsbias, da vi tager højde for, at bestemte typer af mennesker er mindre tilbøjelige til at deltage i Valgundersøgelsen. Det ændrer dog ikke billedet, og om noget bliver vores estimat af opbakningen dårligere. En af udfordringer ligger dog også i, at man vægter op til den stemmeberettigede befolkning frem for dem, der rent faktisk stemmer (se kapitel 6).

Endelig sorterer vi længst til højre i figur 18.4 respondenter fra, hvis de ikke har deltaget i Valgundersøgelsen via internettet. Det skyldes, at web-besvarelser er kendetegnet ved en særlig grad af anonymitet, da der ikke er nogen interviewer til stede, hvilket, alt andet lige, reducerer problemer med social desirability bias. Dette gør dog vores estimat af Dansk Folkepartis opbakning betydeligt dårligere.

Det ser altså ikke ud til, at det er enkelt at fjerne de potentielle kilder til bias, som gjorde, at meningsmålingsinstitutterne undervurderede Dansk Folkeparti. Det er heller ikke så mærkeligt, da justering af ens meningsmålings resultater ofte ville kunne gøre målingen både bedre og dårligere. Det er for eksempel ikke tilfældigt, hvem der svarer over internettet og over telefon. Hvis man således kun fokuserer på web-respondenter, som vi gjorde ovenfor, så risikerer man at bytte den mindre social desirability bias ud med en forøget selektionsbias. I dette tilfælde vil kvaliteten af justeringen afhænge af *vekselkursen* mellem de forskellige former for bias – en vekselkurs, der ikke er kendt, før valgresultatet foreligger.

Konklusion – kan vi undgå at ramme forkert i fremtiden?

Det virker ikke, som om problemerne med at ramme forbi Dansk Folkeparti opbakning kan løses ved at lave simple justeringer af meningsmålingerne. I stedet kan man forsøge at lave mere gennemgribende ændringer i designet af meningsmålingerne. Vi vil afslutte kapitlet ved at se på, hvilke muligheder der er for at ændre den måde, vi laver meningsmålinger på.

En umiddelbar løsning er at gøre målingerne større ved at inkludere flere respondenter. Det har eksempelvis Jyllands-Posten valgt at gøre (Collignon 2015). Det vil højst sandsynligt betyde, at de enkelte målinger varierer mindre, men det er mere tvivlsomt, at det kan skabe mere præci-

sion på tværs af målingerne. I dagene op til valgdagen blev titusindvis af danskere bedt om at fortælle et meningsmålingsinstitut, hvem de ville stemme på. På trods af dette ramte vores model, der inkluderer al denne information, forkert. Noget tyder altså på, at større målinger ikke kan være hele løsningen.

I stedet anbefaler vi, at meningsmålingsinstitutter og andre interesserede ser nærmere på den voksende litteratur om, hvordan man måler socialt uacceptable holdninger i meningsmålinger (se f.eks. Blair m.fl. 2014; Blair & Imai 2012; Bullock m.fl. 2011; Holbrook & Krosnick 2010). Et eksempel fra denne litteratur er det såkaldte listeeksperiment, hvor man viser alle respondenter en liste af forskellige handlinger og beder dem angive, hvor mange – og ikke hvilke – handlinger på listen, de har tænkt sig at udføre. En tilfældig halvdel af respondenterne får imidlertid en liste med et ekstra punkt: den ting de muligvis ikke vil indrømme, at de gør – eksempelvis at stemme på Dansk Folkeparti. Ved at sammenligne, hvor mange ting fra listerne de to tilfældigt inddelte grupper nævner, at de har tænkt sig at gøre, kan man udregne, hvor stor en andel af befolkningen, der vil stemme på Dansk Folkeparti, uden at nogle enkelte respondenter direkte behøver at sige, at de rent faktisk har tænkt sig at gøre dette – de skal jo blot angive antallet af handlinger, de vil foretage sig. Ved at bruge metoder som denne, kan man højst sandsynligt undgå de fleste problemer med social desirability bias. For at adressere problemer med selektionsbias kan institutterne forsøge sig med mere statistisk avancerede vægtningsprocedurer.

Når man vægter, justerer man den stikprøve af vælgere, man har fået på baggrund af politiske og socio-demografiske variable. Har man eksempelvis fået en for lille andel respondenter fra Vestjylland, i forhold til hvad de udgør i befolkningen, så lægger man større vægt på de vestjyder, man har, når data skal analyseres. Et stort problem med vægtning er imidlertid, at man hurtigt kan komme til at vægte grupper med relativt få respondenter i – eksempelvis vestjyder, der stemte Enhedslisten – meget op, hvorved usikkerheden øges. I den seneste tid har forskere imidlertid udviklet metoder, der adresserer dette problem, så man kan foretage vægtning med et stort antal små grupper på en valid måde (se f.eks. Lax & Phillips 2009). Disse metoder, der ofte går under navnet MRP for *multi-level regression* med poststratifikation, har vist sig at kunne estimere valgresultater med stor præcision, selv hvis datagrundlaget lider af meget store problemer med selektionsbias. Wang m.fl. (2015) formåede således at

komme med et ganske præcist gæt på udfaldet af det amerikanske præsidentvalg i 2012 ved at bruge MRP på en meningsmåling, hvor deltagerne udelukkende var brugere af spillekonsollen Xbox.

Et vigtigt forbehold at tage i denne sammenhæng er, at både MRP og de førnævnte listeeksperimenter typisk vil kræve flere respondenter end én traditionel meningsmåling for at få usikkerhedsangivelser af en acceptabel størrelse. Som beskrevet ovenfor virker det dog ikke som om, at meningsmålingsinstitutternes har problemer med at skaffe store stikprøver – problemet er snarere at få omsat store stikprøver til meningsfulde estimater af partiernes opbakning. Her kan disse nye metoder hjælpe.

Vi er opmærksomme på, at de nye metoder kan virke skæve i forhold til institutternes gængse arbejdsgange. Vi tror imidlertid, at det er nødvendigt, at meningsmålingsinstitutterne går mere kreativt til værks i takt med, at svarprocenterne på meningsmålinger falder, og det dermed bliver sværere at tage kontakt til et tilfældigt udsnit af danskere.

En sidste og vigtig udfordring i forhold til at bruge meningsmålinger fremadrettet er åbenhed om, hvilke metoder institutterne bruger (Mandag Morgen 2016). Selvom de fleste afreporterer hvor mange, der er blevet spurgt i hver enkel måling, og i hvilken periode (Megafon var dog en undtagelse i den sammenhæng), så er det stadig uklart for dem, der bruger målingerne, kommentatorer, medier og forskere, hvordan de er skruet sammen. Det ville være lettere at vurdere de enkelte målingers værdi, hvis spørgsmålsformulering, indsamlingsmetode og vægtning blev beskrevet i et baggrundsnotat, når en meningsmåling blev offentliggjort. Uden disse informationer er det svært at vide, hvorfor institutterne er uenige, og som følge heraf, hvad der er det til enhver tid bedste bud på, hvem der vinder valget. Endnu vigtigere, så er det uden disse information umuligt at vide, om nogle meningsmålingsinstitutter bruger praksisser, der ikke har videnskabelig basis, såsom at vægte efter andre institutters målinger, såkaldt *herding*.

Som vi indledte kapitlet med at nævne, er meningsmålinger central i valgkampsdækningen, og forskning indikerer, at variation i målingerne kan flytte stemmer (Dahlgaard m.fl. 2015). Det gør meningsmåling til en vigtig disciplin, og vi vil derfor slutte af med at opfordre både de, der laver, videreformidler og kommenterer målinger, til at gøre sig umage med at sørge for, at målingerne leverer det mest valide og præcise resultat, og at der kommunikeres ærligt, hvad de kan og ikke kan.