

Justerad skattningsmetod i Partisymptatiundersökningen

**En beskrivning av de metodjusteringar som
implementerats i maj 2020**

2020

Justerad skattningsmetod i Partisymptiundersökningen

Producent SCB, Statistiska centralbyrån
 Avdelningen för befolknings- och
 välfärdsstatistik
 Box 24300, 104 51 Stockholm
 010-479 40 00

Förfrågningar Regina Vilkénas
 010 -479 4118
 psu@scb.se

Det är tillåtet att kopiera och på annat sätt mångfaldiga innehållet.
Om du citerar, var god uppge källan på följande sätt:
Källa: SCB, Justerad skattningsmetod i Partisymptiundersökningen

1 Innehåll

1	Innehåll	2
2	Sammanfattning	4
3	Inledning	5
4	Kort om Partisympatiundersökningen	6
4.1	Om Partisympatiundersökningen.....	6
4.1.1	Huvudsakliga variabler.....	6
4.1.2	Skattningsförfarande till och med november 2019, Val idag.	8
4.1.3	Skattningsförfarande till och med november 2019, partisympati samt eu- och eurosympati.....	9
5	Kvalitetsuppföljning och förändringsarbete	10
5.1	Bortfall.....	10
5.2	Bortfallets konsekvenser.....	11
5.3	Metodundersökning 2018.....	13
5.4	Övrigt kvalitetsarbete.....	14
5.4.1	Utvecklingsbehov.....	14
6	En ny och gemensam skattningsmetod	16
6.1	Datatunderlag i skattningen.....	16
6.2	Hjälpinformation.....	16
6.2.1	Utvärdering av hjälpvariabler.....	17
6.2.2	Partival vid senaste riksdagsval.....	19
6.3	Estimator.....	20
6.4	Imputering av partiellt bortfall och vet ej-svar.....	21
6.4.1	Imputeringsmetoder, partival i senaste riksdagsval.....	22
6.4.2	Imputeringsmetoder, val idag.....	22
6.5	Slutgiltigt förslag.....	24
6.6	En jämförelse med metodundersökningen 2018.....	25
7	Publicerade resultat och tidsserier	26
7.1	Nya tidsserier, november 2010 – november 2019.....	26
8	Referenser	28
	Bilaga 1. Indelning av hjälpinformation	29
	Bilaga 2. Teknisk beskrivning av estimationen	31

2 Sammanfattning

Metoden i Partisymptiundersökningen, PSU, har i grunden sett likadan ut sedan undersökningarna startade i början på 1970-talet. Jämfört med hur det såg ut då så har bortfallet i undersökning ökat och det politiska landskapet ser annorlunda ut med fler partier. Utifrån detta har vi sett ett behov av att uppdatera metoden, till en metod som är anpassad till dagens, och framtidens, förutsättningar.

PSU har två huvudmått, *Val idag* och *Partisympti*. *Val idag* avser resultatet vid ett hypotetiskt riksdagsval vid undersökningstillfället, medan *Partisympti* avser "bästa parti" utan att specificera politisk nivå.

I design- och utvärderingsarbetet har ett stort antal alternativa skattningsmetoder för *Val idag* och *Partisympti* testats och utvärderats, bland annat gentemot den metodundersökning som gjordes 2018.

Tidigare har måtten tagits fram med olika metoder, men från och med maj 2020 kommer en gemensam skattningsmetod att användas för båda måtten. Utöver det har vi utökat den hjälpinformation som används för att ta fram skattningarna samt justerat imputeringsmetoden. Metodjusteringen syftar till att ge bättre precision i skattningarna, förbättrad jämförbarhet mellan de olika måtten samt en möjlighet att på sikt utöka redovisningen av *Val idag* för fler delpopulationer.

I samband med publiceringen uppdateras SCBs Statistikdatabasen med omräknade skattningar från och med november 2010. Detta möjliggör jämförelser över tid utan att hänsyn behöver tas till metodjusteringen.

Val idag-skattningen påverkas på totalnivå i förhållandevis liten utsträckning, medan den justerade skattningsmetoden har något större inverkan på *Partisymptiskattningen*. I synnerhet för de partier som är under- eller överrepresenterade bland de svarande när det kommer till röstning vid senaste riksdagsvalet eftersom vi lagt till den uppgiften i hjälpinformationen för partisympti.

3 Inledning

I denna rapport redogör vi för den metodomläggning som gjorts för Partisympatiundersökningen och som implementeras i samband med undersökningen i maj 2020.

Omläggningen är ett resultat av ett långsiktigt kvalitetsarbete. Översynen av metoden har fokuserat på det ökade bortfallets implikationer på undersökningen i stort samt på imputeringsförfarandet i skattningen av valresultatet om det vore val idag (*Val idag-skattningen*). Vi har även beaktat användarbehovet och tagit hänsyn till önskemål från användare som främst handlar om möjligheten till utökad redovisning av *Val idag-skattningen* och möjligheten att titta på relationen mellan ett hypotetiskt valresultat och väljarkårens partisympatier vid ett givet tillfälle och över tid.

SCB genomför ett kontinuerligt kvalitetsarbete som i synnerhet baseras på bortfallsanalys och en återkommande validering av *Val idag-skattningen*. Som följd av bortfallsutvecklingen genomförde SCB en utökad metodundersökning vid 2018 års val. Den har legat till grund för analyser rörande de förändringar som nu implementeras.

Tidsserierna som finns i Statistikdatabasen har uppdaterats från 2010 med skattningar framtagna med den nya metoden för att möjliggöra jämförelser över tid. Skattningarna med tidigare metod finns under en övergångsperiod tillgängliga i ett exceldokument på SCBs webbplats.

Rapporten inleds med en beskrivning av undersökningen, dess huvudsakliga mått och hur dessa tidigare tagits fram. Vidare diskuteras bortfall och andra metodologiska utmaningar, samt hur kvaliteten i undersökningen kan studeras och utvärderas. I avsnitt sex ges en teoretisk beskrivning av de metodval som gjorts, innan rapporten avslutas med en beskrivning av de omräkningar av äldre resultat som tagits fram för ökad jämförbarhet över tid.

4 Kort om Partisymptatiundersökningen

4.1 Om Partisymptatiundersökningen

Partisymptatiundersökningen (PSU) genomförs i maj och november varje år. Undersökningen har genomförts sedan 1972. Syftet med undersökningarna är att beskriva det aktuella politiska opinionsläget (partival vid val idag, partisymptati, EU- och Eurosymptati) i såväl hela som delar av väljarkåren samt förändring över tid.

Urvalet består av cirka 9 000 personer, folkbokförda i Sverige, som hade varit röstberättigade om det varit ett riksdagsval andra söndagen i september aktuellt undersökningsår. Urvalet består av tre paneler som var och en är ett slumpmässigt urval från ramen som hämtas från Registret över totalbefolkningen (RTB). Varje panel är med i tre på varandra följande undersökningar, vilket innebär att en panel roterar ut varje omgång. Urvalet görs systematiskt från en ram sorterad efter län, kommun, distriktkod, fastighetsbeteckning och familjeidentitet. Urvalet betraktas som ett obundet slumpmässigt urval (OSU) i skattningen. Databasinsamlingen görs sedan 2016 genom kombinerad insamling, telefonintervjuer och webbenkät, där respondenterna själva kan välja svarsmetod.

Partisymptatiundersökningen ger en bild av opinionsläget vid den aktuella tidpunkten för insamlingen, det är inte en prognos över kommande valresultat. Efter varje undersökningsomgång publicerar SCB en skattning över hur väljarnas röster skulle fördela sig mellan riksdagspartierna vid ett hypotetiskt riksdagsval vid undersökningstillfället, samt väljarflöden. Väljarflöden beskriver hur väljare rör sig mellan de olika partierna jämfört med förra undersökningsomgången och med riksdagsvalet. Vi redovisar också hur partisymptati (bästa parti) fördelar sig över riksdagspartierna och övriga partier. Det stora urvalet gör det även möjligt att redovisa hur partisymptatierna fördelar sig inom ett stort antal delgrupper i väljarkåren. Eftersom undersökningen har pågått sedan 1972 finns goda möjligheter till studier över tid.

4.1.1 Huvudsakliga variabler

Val idag – partival vid ett hypotetiskt riksdagsval

Val idag-skattningen avser resultatet vid ett riksdagsval i maj respektive november. Här tillfrågas alltså respondenten om röstningsintention och den politiska nivån är definierad till riksdagsnivå. Eftersom vi här vill mäta ett hypotetiskt valresultat för den aktuella tidpunkten tas även

hänsyn till röstnings sannolikhet, se i avsnitt 4.1.2, i skattningsförfarandet.

Frågorna som ställs är:

- Om vi tänker oss att det vore riksdagsval någon av de närmaste dagarna: skulle du rösta, skulle du kanske rösta eller skulle du inte rösta?
- Vilket parti skulle du rösta på?

De som anger att de inte skulle rösta/inte vet får istället följdfrågan:

- Om du trots allt bestämde dig för att rösta, vilket parti skulle du då rösta på?

De som anger att de inte vet vilket parti de skulle rösta på eller anger att de skulle rösta blankt får frågan:

- Finns det ändå något parti du skulle kunna överväga att rösta på?

Val idag-skattningen redovisas för riket samt efter region och kön.

Partisynpati (bästa parti)

Partisynpati avser vilket parti man står närmast, eller har som bästa parti, oavsett politisk nivå eller tillfällig röstningsintention vid ett val. I skattningsförfarandet tas inte heller någon hänsyn till om man skulle rösta eller inte om det vore val. Även om många röstar på partiet de sympatiserar med är det en del som inte gör det.

Frågorna som ställs är:

- Är det något av de politiska partierna som står dig närmare än de andra? Vilket parti är det?

De som inte har ett parti de står närmare får följdfrågan:

- Men vilket parti har du störst sympati för?

Partisynpati redovisas efter demografiska registervariabler som kön, ålder, civilstånd, antal barn, län och födelse land. Därutöver används även vissa variabler som mäts i undersökningen som redovisningsvariabler, så som facktillhörighet, socioekonomisk tillhörighet, sektor för anställning samt boendeförhållanden.

EU-sympati

Frågan om EU-sympati har ställts i PSU på det sätt som den gör idag sedan 1996.

Frågan som ställs lyder:

- Är du i huvudsak för eller mot det svenska medlemskapet i EU eller har du ingen bestämd åsikt?

EU-sympati redovisas efter samma demografiska variabler som *Partisympati*.

Eurosympati

Eurosympati har redovisats på nuvarande sätt sedan 2004.

Frågan som ställs lyder:

- Om vi idag skulle ha folkomröstning om att ersätta kronan som valuta, skulle du då rösta ja eller nej till att införa euron som valuta i Sverige?

Eurosympati redovisas efter samma demografiska variabler som *Partisympati*.

4.1.2 Skattningsförfarande till och med november 2019, Val idag

Val idag-skattningen, det vill säga resultatet i ett hypotetiskt val vid undersökningstillfället, skattas med en poststratifieringsestimater. Som stratifieringsvariabler används röstningsalternativ vid senaste riksdagsval (2018 var det åtta riksdagspartier, övriga partier, "ej röstat" och "för unga") samt valkretstillhörighet (10 strata). Sammanlagt är det alltså 11×10 poststrata.

Med avsikten att skatta valresultatet vid ett hypotetiskt val tar vi hänsyn till valdeltagande. Varje person i urvalet tilldelas en "röstnings-sannolikhet" som bestäms av svaret på en fråga om intentionen att delta i ett hypotetiskt val samt uppgifter om valdeltagande från metodundersökningen, läs mer om den i avsnitt 5.2. Kortfattat ges de som uppger att de skulle rösta en större tyngd i undersökningen jämfört med dem som uppger att de inte skulle rösta.

Personer som inte uppgivit vilket parti de skulle rösta på i ett val vid undersökningstillfället (röstningssympati) antas komma att rösta på det parti de sagt sig ha störst sympati för (partisympati – "bästa parti"). Personer för vilka uppgift fortfarande saknas om röstningssympati och/eller röstningsalternativ vid senaste riksdagsval tilldelas ett parti utifrån samma fördelning som de svarande individerna har inom respektive region. Till och med maj 2019 hade de svarande möjlighet att

ange information om block-sympati om de inte kunde ange ett specifikt parti. Den informationen användes då vid fördelningen.

Utöver imputeringen som beskrivs ovan, som avser de svarande, imputeras även hela objektsbortfallet utifrån fördelningen bland svarande. För en mer detaljerad beskrivning av estimatorn, se SCB 2019.

4.1.3 Skattningsförfarande till och med november 2019, partisympati samt eu- och eurosympati

Fördelningen av sympatier beräknas med hjälp av den generaliserade regressionsestimatorn, GREG. Information från den aktuella populationen om kön*åldersklasser (12), region (10), utbildning (2) och födelseland (2) används som hjälpinformation, där siffrorna inom parenteser anger antal grupper eller klasser, se bilaga 1. För en mer detaljerad beskrivning av estimatorn, se SCB (2019).

5 Kvalitetsuppföljning och förändringsarbete

SCB fick 1968 uppdraget att ta fram en metod för att mäta parti-sympatier i befolkningen. Då inleddes ett gediget utredningsarbete som resulterade i ett flertal rapporter och testundersökningar för att ta fram en robust metod för att mäta opinionen. Den första undersökningen som sedan publicerades genomfördes 1972. Skattningsmetoderna har, med några undantag, i stort sett varit desamma under perioden 1972-2019. Metodfrågorna har dock sedan start varit viktiga för PSU och SCB har löpande kontrollerat metoderna genom de så kallade metodundersökningar som genomförs i anslutning till riksdagsvalen. De möjliggör en validering av metoden mot det sanna valresultatet.

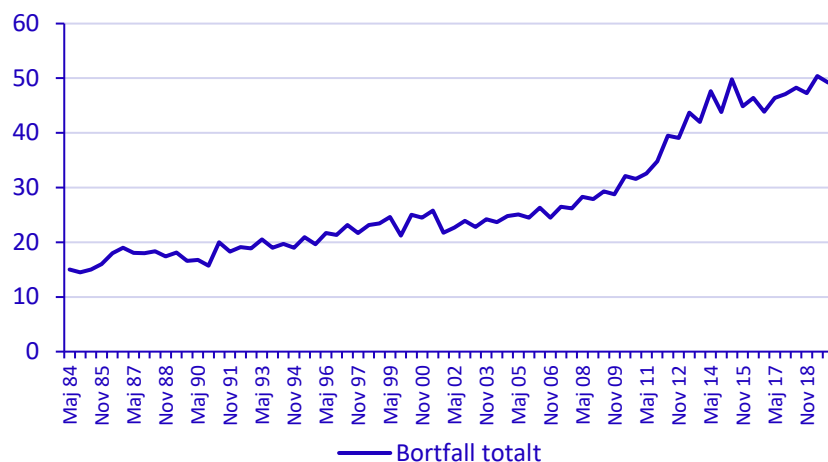
Under perioden har antalet partier i riksdagen ökat, och även bortfallet har ökat över tid. Båda dessa fenomen får konsekvenser för den skattningsmetod som designades i slutet av 60-talet. I detta kapitel adresserar vi dessa konsekvenser och beskriver bortfallets utveckling och påverkan. Vi beskriver även den metodundersökning som genomfördes vid senaste riksdagsvalet 2018 samt annat kvalitetsarbete som har genomförts och avslutar med en genomgång av det utvecklingsbehov som vi har sett vara nödvändigt för PSU.

5.1 Bortfall

Bortfallsproblematiken är numera ett välkänt problem i individundersökningar och är inget unikt för PSU eller för SCB. Det är dock något man måste förhålla sig till. Under de senaste tio, femton åren har bortfallet ökat markant och sedan 2010, då vi införde kalibrering av vikterna vid skattning av *Partisynpati*, har det ökat med nästan 20 procentenheter. Sedan SCB startade mätningarna för 50 år sedan har bortfallet mer än tredubblats. De senaste undersökningsomgångarna har bortfallet varit strax under 50 procent.

Diagram 5.1. Bortfallet i PSU, 1984-2019

Procent



Bortfall är i sig nödvändigtvis inget problem, utan det blir framförallt problematiskt om vi har en skev svarsmängd sett utifrån målvariabeln. Alltså om de svarande skiljer sig från urvalet på ett väsentligt sätt i fråga om det man vill mäta. Samma bortfall skulle alltså kunna påverka två olika undersökningar på olika sätt beroende på hur bortfallet korrelerar med målvariabeln. Ett ökat bortfall kan även få praktiska konsekvenser för såväl vilken skattningsmetod som bör användas som för vilka grupper resultat kan redovisas.

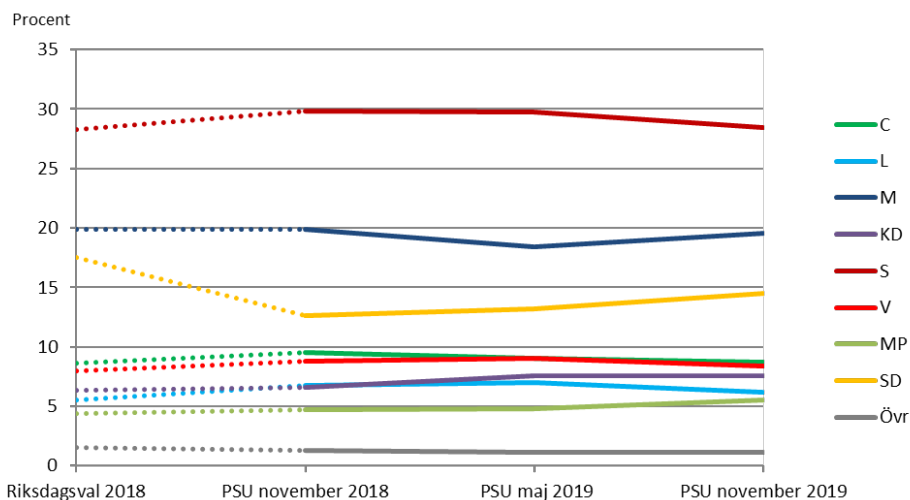
Vanligtvis har man skäl att misstänka att bortfallet är snedvridande och därför söks metoder att reducera den effekten. Ett sätt att kompensera för bortfall är imputering, ett annat är genom viktning. Vid imputering ersätts saknade värden med proxyvärden vilka kan tas fram på olika sätt. I avsnitt 6 beskrivs hur detta görs i PSU. Viktning görs ofta genom att dela upp de svarande i grupper och inom varje grupp bilda en korrigeringsvikt som används i skattningen. Målsättningen med indelningen är att föra samman objekt som har samma sannolikhet att svara. Mer om val av skattningsmetod finns i avsnitt 6 och i bilaga 2 ges en mer teknisk beskrivning av skattningsmetoden.

5.2 Bortfallets konsekvenser

Genom att titta på hur de svarande har svarat på frågan om partival vid senaste riksdagsval kan vi få en bild av hur väl respektive partis väljare representeras bland de svarande i jämförelse med fördelningen i riksdagsvalet. Har vi en större andel svarande som uppgett att de röstade på ett visst parti jämfört med det faktiska valresultatet kan vi prata om en viss överrepresentation av det partiets väljare och vice versa om andelen är mindre jämfört med valresultatet.

I diagrammet nedan ser vi till exempel att det i vår svarsmängd vid undersökningen i november 2018 är en mindre andel som uppger att de röstade på Sverigedemokraterna i förhållande till partiets faktiska valresultat 2018.

Diagram 5.2. Valresultatet 2018 i förhållande till vad respondenterna i PSU november 2018 – november 2019 uppgett att de röstat på vid senaste riksdagsval (2018).



Valresultat i riksdagsvalet, längst till vänster, i jämförelse med hur respondenterna i PSU angett att de röstade vid senaste riksdagsvalet.

För *Val idag-skattningen* innebär det ökade bortfallet problem med skattningsmetoden som bygger på ett stort antal poststrata. Till följd av minskade svarsmängder har antalet svar i vissa poststrata minskat kraftigt. Utöver bortfallstrenden har även antal poststrata i estimatorn successivt ökat till följd av att antalet partier i riksdagen har ökat. Till en början delades hjälpinformationen upp i 10 regioner och 5 partistrata + valskolkare + nya unga (7*10=70). Sedan 2010 har antalet strata ökat till 9 partistrata + valskolkare + unga kombinerat med 10 regioner (11*10=110). Att antal partier ökar får alltså även det en negativ effekt i detta sammanhang. Det beror på att vi då får allt fler strata med ett väldigt litet antal observationer vilket leder till ökad osäkerhet i skattningarna.

Tabell 5.2. Hjälpinformationen partival i senaste riksdagsval fördelat över 10 regioner. Antal svar i maj 2019 i respektive poststrata.

Partival senaste RV	Regioner									
	1	2	3	4	5	6	7	8	9	10
C	20	25	46	18	42	43	49	63	38	32
L	14	17	14	21	52	39	54	53	18	9
M	59	47	63	42	127	83	123	125	64	33
KD	18	28	46	11	53	22	40	57	19	21
S	71	78	103	55	181	102	139	217	112	178
V	30	18	28	39	43	52	37	59	32	38
MP	22	12	9	13	26	37	27	32	7	14
SD	44	67	57	18	88	28	62	101	50	34
ÖVR	1	5	4	3	7	4	4	9	3	7
Röstade inte/röstade blankt	351	406	420	272	699	397	579	807	434	457
Unga	8	10	8	9	13	7	19	22	9	10

Regioner, 1– Stormalmö, 2– Övriga södra Götaland, 3- Småland med öarna, 4- Göteborg, 5– Övriga Västsverige, 6- Stockholm, 7- Övriga Stockholms län, 8- Östra Mellansverige, 9- Norra Mellansverige, 10-Mellersta och övre Norrland.

5.3 Metodundersökning 2018

Som tidigare nämnts genomför SCB återkommande metodundersökningar veckorna innan riksdagsvalen för att validera undersökningens metod. Insamlingen i samband med 2018 års riksdagsval genomfördes från den 22 augusti till och med den 7 september 2018. Själva riksdagsvalet genomfördes den 9 september.

Urvalet bestod av 6 025 personer, två tredjedelar av respektive panel i den ordinarie novemberomgången 2018. En tredjedel av urvalet deltog således för första, en för andra och en för tredje gången i SCBs parti-sympatiundersökning. Svarsfrekvensen uppgick till drygt 50 procent.

Urvalets storlek var, trots att det var mindre än vid en ordinarie undersökning, utökat jämfört med en normal metodundersökning. Tidigare metodundersökningar har bestått av 3 000 personer, men med anledning av den större översynen av metoden utökades det 2018.

Även om vi med metodundersökningen försöker efterlikna en ordinarie undersökningsomgång i så stor utsträckning som möjligt så finns det en del skillnader som är värda att nämna.

- Kortare insamlingsperiod, drygt två veckor istället för fyra veckor.
- Vi frågar om en kommande partiröst i ett faktiskt val, ett prognosförsök. Vid en ordinarie undersökning frågar vi om partiröst vid ett hypotetiskt riksdagsval de närmsta dagarna.
- Mindre urval.

Resultaten från metodundersökningen visar att skattningen av *Val idag* har varit relativt nära valresultatet, se tabell 5.1. I tabellen finns valresultatet, resultatet från metodundersökningen 2018 samt differenser mellan metodundersökningarna, med då aktuell metod, 2010, 2014 och 2018 och valresultatet respektive år.

Tabell 5.1. Metodundersökningen i jämförelse med valresultat, 2010, 2014 och 2018

Partinamn	Valet 2018	Metod und. 2018	Diff. 2018	Diff. 2014*	Diff. 2010*	
C	8,6	8,8	0,2	0,1	-1,3	
L	5,5	6,7	1,2	0,6	0,4	
M	19,8	18,4	-1,4	-2,9	-1,5	
KD	6,3	4,6	-1,7	-0,3	1,1	
S	28,3	27,9	-0,4	0,4	-0,1	
V	8,0	10	2,0	1,7	-0,3	
MP	4,4	5,5	1,1	1,1	1,8	
SD	17,5	16,2	-1,3	-1,0	-0,2	
ÖVR	1,5	1,9	0,4	0,2	0	
Genomsnitts- avvikelse				1,1	0,9	0,7

*2010 och 2014 har en estimator med tre regioner använts för att anpassa den till det mindre urvalet som var 3 000. Vid undersökningen 2018 användes ordinarie estimator, med 10 regioner, då urvalet var större.

5.4 Övrigt kvalitetsarbete

Under perioden 2009-2017 har Partisymptiundersökningen genomlysts ur en rad olika aspekter, som metodfrågor, redovisning och produktionsprocessen. Genomlysningarna har bland annat resulterat i ny tabellplan, införandet av GREG-estimatoren för skattning av partisympti, EU-sympati och eurosympatier, införande av kombinerad insamling samt omläggning av produktionssystemet.

Åren 2015-2018 har studier fokuserats till undersökningens mest centrala skattning, valresultat vid ett val idag. Analyser har genomförts avseende kvaliteten i den hjälpinformation (valt parti i senaste riksdagsval) som används i estimationen, samt den imputering som görs. Utöver det har vi framförallt fokuserat på hur bortfallet påverkar *Val idag-skattningen* och *Partisymptatiskattningen*. Vi redogör nedan för de huvudsakliga utvecklingsbehov och metodutmaningar som identifierats.

5.4.1 Utvecklingsbehov

Utökad redovisning av *Val idag-skattningen*

Under många år har det förekommit frågor från media och allmänheten om vad som skiljer de olika skattningarna åt och hur SCB har tänkt att de ska användas. Man har framfört att det är synd att SCB har en annorlunda och mindre detaljerad redovisning av *Val idag* jämfört med *Partisymptati*. Detta eftersom det är den tidigare som många inför ett val anser är mest intressant att följa. Genom förfrågningar och kontakt med användare får vi löpande önskemål om att utöka redovisningen för *Val idag*, eller möjliggöra att på uppdrag kunna redovisa skattningen efter olika bakgrundsvariabler.

Anledningen till den begränsade möjligheten till utökad redovisning av *Val idag* är en otillfredsställande poststratifiering som fungerar bra på aggregerade totaler men inte i en mer finfördelad redovisning, eftersom endast regioner finns med i hjälpinformationen.

Jämförbarhet

Skattningen av *Partisymptati* och *Val idag* har hittills tagits fram med olika skattningsmetoder, beskrivna i avsnitt 4.1. I samband med att bortfallet ökat har vi kunnat se att skattningarna påverkats olika och vi har uppmuntrat att jämförelser mellan måtten görs med försiktighet. Möjligheten att jämföra *Partisymptati* och *Val idag* för att se hur måtten förhåller sig till varandra och om det förändras över tid skulle bidra till en ökning av undersökningens användbarhet.

Robusthet över tid

Som beskrevs i avsnitt 5.3 så har det ökade bortfallet i kombination med att antal strata ökat i poststratifieringen av *Val idag* lett till en förekomst av strata med få svarande. Poststratifieringsestimatorn är även känslig för andra förändringar som kan påverka antal observationer i strata, som ytterligare ett parti i riksdagen eller ett väldigt lågt regionalt stöd för ett eller flera partier. Därför är det önskvärt med en skattningsmetod som inte är lika känslig för ökat bortfall och inte heller för förändringar i det politiska landskapet.

Behov av ny imputeringsinformation på grund av ett förändrat politiskt läge

I PSU har vi tidigare följt upp svarande som är osäkra på vilket parti de sympatiserar med och/eller vilket parti de skulle rösta på med hjälp av en fråga om vilket block de föredrar. Informationen från dessa blockfrågor har sedan använts i imputeringen.

I och med överenskommelsen mellan S, MP, C och L i samband med regeringsbildningen efter 2018 års riksdagsval har det politiska landskapet förändrats och det blev svårt att ställa tydliga frågor om blocktillhörighet. I och med det har det varit nödvändigt att se över de rådande imputeringsmetoderna. Översynen har utgått från all imputering som gjorts och inte enbart fokuserat på de uteblivna blockfrågorna.

6 En ny och gemensam skattningsmetod

Målsättningen har varit att ta fram en gemensam skattningsmetod som fungerar för skattningar av *val idag* och *partisympati* med syfte att öka jämförbarhet mellan måtten, vilket ger ytterligare värde till undersökningen. För att nå den behövs en skattningsmetod som är mer flexibel med avseende på hjälpvariabler och robust över tid med avseende på antal partier i riksdagen och bortfallets utveckling framöver.

Samtidigt har det varit viktigt att behålla de två måttens särdrag. I design- och utvärderingsarbetet har ett stort antal alternativa skattningsmetoder för *Val idag* och *Partisympati* testats och utvärderats utifrån följande:

- Medelvärde av absoluta avvikelser för respektive parti mellan valresultatet och skattning i metodundersökningen
- Huruvida valresultatet faller inom konfidensintervallet för skattningen för respektive parti eller inte
- I metodundersökningarna 2014 och 2018 – undersöka om det finns systematik i träffsäkerheten
- Teoretiska avvägningar

6.1 Dataunderlag i skattningen

Tidigare har olika datamängder använts vid skattning av *Val idag* och *Partisympati*. I den nya skattningsmetoden används istället samma datamängd för båda måtten.

För att få en gemensam datamängd har mängden svarande definierats av de som har svar i åtminstone någon av frågorna om val idag, partisympati, partival vid senaste riksdagsval, EU-sympati eller Euro-sympati. Detta innebär att partiellt bortfall förekommer i svarsmängden för både *Val idag* och övriga målvariabler.

6.2 Hjälpinformation

Vid val av hjälpvariabler är det tre kriterier som ska beaktas, se Särndal och Lundström (2005)

- (i) Det första kriteriet är att variabeln samvarierar väl med svarsbenägenheten (-sannolikheten). Det är det viktigaste kriteriet eftersom det leder till en minskning av bortfallsrisken för alla skattningar.
- (ii) Det andra kriteriet är att variabeln samvarierar väl med (viktiga) målvariabler. Om så är fallet minskar bortfallsbiasen för de

skattningar som byggs upp av dessa målvariabler. Även variansen minskar för dessa skattningar.

- (iii) Det tredje kriteriet är att variabeln avgränsar (viktiga) redovisningsgrupper. Det leder framförallt till minskad varians i skattningar för dessa redovisningsgrupper.

För att se om hjälpvariablerna uppfyller kriterium (i), kan man studera sambandet mellan den dikotoma variabeln svarande/bortfall och hjälpvariablerna. Detta görs genom att beräkna skattad andel svarande i olika grupper, bestämda av hjälpvariablerna. Vid skattningen används designvikten som i PSU är populationsstorlek delat med urvalsstorlek. Vid stora skillnader mellan svarsandelarna utgör variabeln som definierar grupperna en stark kandidat till hjälpvariabel.

6.2.1 Utvärdering av hjälpvariabler

Skattad andel svarande i olika grupper redovisas vid publicering av *Partisympti* bl.a. för de variabler som ingår i hjälpinformationen för partisympti. Från dessa resultat kan man se att svarsandelarna skiljer sig mycket åt för t.ex. variablerna födelseland (utrikes född/inrikes född), utbildning och för olika åldersgrupper, medan svarsandelarna i olika regioner är mer lika.

Vissa hjälpvariabler samvarierar, vilket gör att vi behöver ta reda på hur de fungerar tillsammans. I Särndal och Lundström (2009) presenteras en indikator, H_3 -indikatorn, som kan användas för att välja hjälpinformation och för att se den sammanlagda effekten av olika hjälpvariabler. Indikatorn kan användas på följande sätt; beräkna värdet på indikatorn för olika variabler, lägg sedan stegvis till de variabler i hjälpvektorn som har störst effekt på indikatorn och beräkna värdet på indikatorn på nytt. Ett högt värde på H_3 -indikatorn indikerar hjälpvariablernas möjlighet att reducera eventuell bortfallsrisk, se kriterium (i).

I tabellerna 6.1 och 6.2 nedan redovisas värden på indikatorn för ett urval av möjliga hjälpvariabler och för kombinationer av hjälpvariabler. De variabler som redovisas här är de som tidigare använts i skattningen av partisympti, se avsnitt 4.1.3, samt valdeltagande. Variabeln partival vid senaste riksdagsval som tidigare använts vid skattning av val idag, se avsnitt 4.1.2, går inte att utvärdera med H_3 -indikatorn eftersom värdet på variabeln behövs på individnivå för hela urvalet och vi har bara tillgång till variabeln för dem som svarat i undersökningen. Variabeln valdeltagande har vi för hela urvalet och den variabeln ingår i partival vid senaste riksdagsval, se bilaga 1.

I tabell 6.1 och 6.3 visas H_3 -indikatorns värde för omgången maj 2019 för enskilda variabler och i tabell 6.2 visas hur H_3 -indikatorn ökar då hjälpvariabler läggs till i hjälpvektorn.

Tabell 6.1. H_3 -indikatorn, maj 2019

Variabel	$H_3 \times 10^5$
Valdeltagande	297,3
Utbildningsnivå	183,3
Kön kombinerat med åldersklasser	157,6
Födelseland	98,0
Regioner	36,2

Tabell 6.2. H_3 -indikatorn, kombinationer av variabler maj 2019

Variabler	$H_3 \times 10^5$
Valdeltagande och utbildningsnivå	332,5
Valdeltagande, utbildningsnivå och kön kombinerat med åldersklasser	363,8
Valdeltagande, utbildningsnivå, kön kombinerat med åldersklasser och födelseland	367,1
Valdeltagande, utbildningsnivå, kön kombinerat med åldersklasser, födelseland och regioner	368,3

I tabellerna 6.1 och 6.2 redovisas endast resultat från maj 2019. Resultaten från tidigare omgångar skiljer sig något vad gäller värdet på H_3 -indikatorn men i studerade omgångar ger valdeltagande¹ högst H_3 -värde följt av utbildningsnivå, kön kombinerat med åldersklasser, födelseland och regioner. Av resultaten i tabell 6.2 ser vi att H_3 -värdet för valdeltagande och utbildningsnivå är större än H_3 -värdet för valdeltagande i tabell 6.1. Läger man till kön kombinerat med åldersklass ökar H_3 -värdet ytterligare men effekten av att lägga till födelseland och speciellt regioner ger inte en lika stor ökning, vilket innebär att t.ex. regioner inte är lika viktigt för kriterium (i) ovan. Eftersom regioner är viktiga redovisningsgrupper och partifördelningen i olika regioner ser olika ut enligt valresultat bör regioner ändå ingå i hjälpvektorn enligt

¹ Valdeltagande finns endast för omgångarna från och med riksdagsvalet 2018.

kriterium (iii) om redovisningsgrupper och för att minska variansen, se kriterium (ii).

Gällande kriterium (iii) så är variablerna, kön, ålder, utbildning, födelseland och region viktiga redovisningsgrupper i redovisningen av *Partisymptati*. I och med att dessa variabler även finns med som hjälpinformation vid skattningen av *Val idag* kommer det möjliggöra en utökad redovisning av *Val idag*.

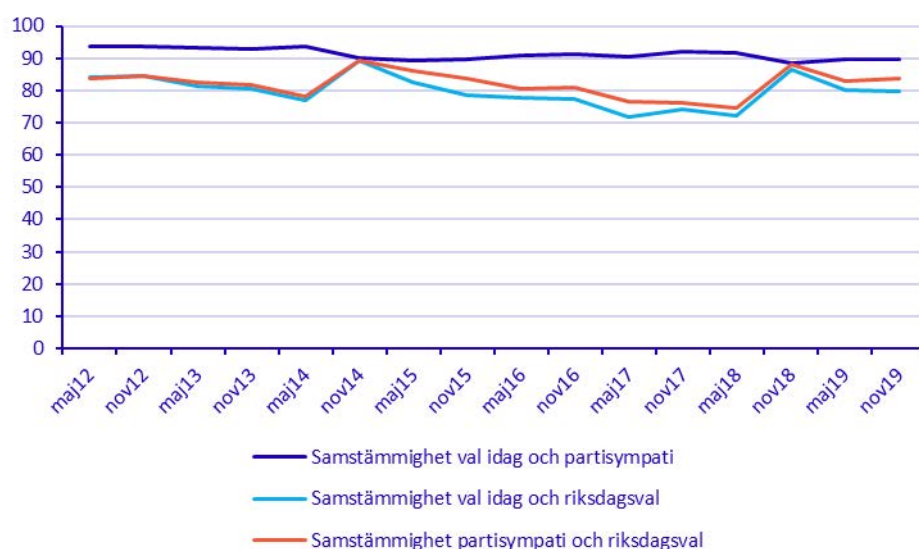
6.2.2 Partival vid senaste riksdagsval

Som nämnts i avsnitt 6.2.1 går det inte att inkludera partival vid senaste riksdagsval i beräkningen av H_3 -indikatorn för val av hjälpinformation. Den variabel som gav högst värde på indikatorn var valdeltagande, se tabell 6.1. Variabeln partival vid senaste riksdagsval har delats in i 11 grupper varav en grupp är ogiltiga och ej röststat, se bilaga 1, och valdeltagande ingår i den gruppen. Dvs. om vi använder partival vid senaste riksdagsval som hjälpvariabel så ingår variabeln valdeltagande indirekt.

Analys av resultaten i metodundersökningen visar också att partival vid senaste riksdagsval är viktig hjälpinformation vid skattningen av *Val idag*.

Val idag och *Partisymptati* mäter två skilda fenomen, som dock båda i mycket stor utsträckning samvarierar med en persons rösthandling vid riksdagsval. Under en mandatperiod kan vi följa hur partival för partisymptati och val idag förhåller sig till partival vid senaste riksdagsvalet.

Diagram 6.1: Samstämmighet mellan partival för *Val-idag* och *Partisymptati* samt konsistens mellan respektive mått och partival i senaste riksdagsvalet (2018). Maj 2012 - november 2019



I analysen ovan ingår svarande som har uppgett ett partisvar för val idag, partisymptati och partival vid senaste riksdagsval.

Av diagrammet framgår att runt 80 procent anger att de sympatiserar med eller skulle rösta på samma parti som de röstade på i senaste riksdagsvalet. Andelen varierar dock lite över tid och samstämmigheten är genomgående lite högre mellan partival i riksdagsvalet och partisympati jämfört med val idag. Båda variablerna samvarierar alltså med *partival vid senaste riksdagsval*, även om det i sig är två skilda mått som inte ska sammanblandas. Vi kan därför utgå ifrån att en skattningsmetod som passar väl för partival vid ett val även bör fungera för att skatta partisympati och att metodundersökningen kan ge oss en idé om kvaliteten även i *Partisimpatiskattningen*.

Med tanke på att båda variablerna samvarierar gentemot hjälpinformationen *Partival vid senaste riksdagsval* har vi kommit fram till att den hjälpinformationen bör inkluderas i skattningsmetoden för båda variablerna. De variabler som används i hjälpinformationen beskrivs närmare i avsnitt 6.5 nedan.

6.3 Estimator

I avsnitt 4.1.2 beskrevs den skattningsmetod som använts för skattning av partifördelning för *Val idag* till och med november 2019. Poststratifiering har använts med poststrata bildade med hjälp av partival i senaste riksdagsval och regioner (bildade av valkretsar).

Att utöka poststratifieringen med flera variabler, t.ex. kön och åldersklasser, är inte möjligt med den skattningsmetoden eftersom valresultatet inte är känt uppdelat på t.ex. kön och åldersklasser. Vill man lägga till flera variabler måste man alltså ändra på estimatorn.

Det finns flera skäl till att utöka antalet hjälpvariabler i estimatorn för *Val idag*. Förutom att förbättra precisionen i skattningen på totalnivå, även förbättra redovisningen i delpopulationer. Att redovisa resultaten på en mer detaljerad nivå har varit ett önskemål från användare. För att möjliggöra detta införs en generaliserad regressionsestimator (GREG-estimatorn) i vilken det är möjligt att använda fler hjälpvariabler. GREG-estimatorn är egentligen en ”familj” av estimatorer där poststratifiering ingår som en medlem, se t.ex. Särndal, Swensson och Wretman (1992).

GREG-estimatorn har använts för att skatta partisympati sedan 2010/2011. Vid en jämförelse av poststratifiering och GREG med samma hjälpinformation var det ingen större skillnad på resultaten för *Val idag* med de två metoderna. I GREG-estimatorn användes hjälpinformationen partival vid senaste riksdagsval och regioner, men inte genom att kombinera de två variablerna som vid poststratifiering. GREG-estimatorn som tidigare endast används för *partisimpatiskattningen* fungerar alltså väl även för *Val idag-skattningen* och är den estimator som passar bäst utifrån möjligheten att justera hjälpinformationen.

En mer teknisk beskrivning av GREG-estimatorn ges i bilaga 2.

6.4 Imputering av partiellt bortfall och vet ej-svar

I den datamängd som används i skattningen finns en del partiellt bortfall i målvariablerna *Val idag* och *Partisymptati* och även i variabeln *partival* i senaste riksdagsval som används som hjälpvariabel i skattningen.

För variabeln *Partisymptati* görs ingen imputering, för *Val idag* görs viss imputering och för *partival* i senaste riksdagsval imputeras allt partiellt bortfall och en del vet ej-svar i svarsmängden. Att vi imputerar för *val idag* beror på att vi har en stor grupp som anger att de inte vet att vilket parti de skulle rösta på samtidigt som vi vet att den här gruppen faktiskt bestämmer sig för ett parti och röstar vid ett faktiskt val. Det här är något vi vill ta hänsyn till när vi skattar ett hypotetiskt valresultat.

Vid metodundersökningen inför riksdagsvalet 2018 angav nästan 20 procent att de inte visste vilket parti de skulle rösta på. Valdeltagandet i den här osäkra gruppen var dock högt, cirka 92 procent, vilket är högre än bland de som inte svarar på undersökningen (bortfallet), där valdeltagandet var cirka 82 procent. Dessutom har vi kunnat se att partipreferens skiljer sig åt i de här båda grupperna. De osäkra skiljer sig även från de som redan visste vilket parti de skulle rösta på. Det har vi sett genom att titta på hur de båda grupperna röstade i riksdagsvalet 2018. Vi har jämfört partipreferens bland de som hade ett partisvar, de som var osäkra samt de som inte svarade på metodundersökningen 2018 och som sedan svarade i nästkommande undersökningar. I tabellen nedan ser vi en partifördelning för respektive grupp. Skillnaderna kvarstår under mandatperioden och är ofta tydligare än de var i metodundersökningen.

Tabell 6.3. Partival i riksdagsvalet 2018, endast svarande i metodundersökningen och/eller undersökningarna i november 18 eller i maj 19.

Parti	Partival i riksdagsvalet 2018 bland de som uppgav ett partisvar i metodundersökningen	Partival i riksdagsvalet 2018 bland "osäkra" i metodundersökning	Partival i riksdagsvalet 2018 bland bortfall i metodundersökningen
C	9,1	13,0	8,2
L	6,6	7,7	6,9
KD	6,4	7,2	6,1
M	19,4	22,3	19,0
S	28,7	32,8	34,3
V	9,0	4,7	9,3
MP	5,2	3,7	3,6
SD	14,2	7,4	12,1
Antal svar	2168	430	562

6.4.1 Imputeringsmetoder, partival i senaste riksdagsval

Partival vid senaste riksdagsval används endast som hjälpvariabel i estimatorn och imputeras upp till svarmängden, vilket innebär att vi imputerar det partiella bortfallet i variabeln. I och med att partiellt bortfall och vet ej/blankt-svar imputeras har alla i svarmängden ett värde.

Imputeringen genomförs genom att först hämtas det första partisvaret som har lämnats i någon av undersökningsomgångarna för dem som är med för andra eller tredje (och sista) gången. Detta görs för alla i svarmängden utifrån antagandet att man minns sitt partival bättre ju närmre i tid riksdagsvalet var. De vet ej-svar och det partiella bortfall som kvarstår imputeras med hjälp av en partifördelning. Vilken partifördelning som används beror på om personen har ett partisvar i val idag eller inte. Om det inte finns ett partisvar i val idag beräknas partifördelningen per region (tre regioner), utbildningsnivå och kön. Om partisvar finns i val idag tas även hänsyn till parti i val idag i fördelningen.

6.4.2 Imputeringsmetoder, val idag

Imputering av val idag görs endast för dem som svarat ”vet ej” eller ”skulle rösta blankt” och som anger ett partisvar på frågan om de har ett parti de överväger att rösta på² (”lutar-fråga”), i frågan om partisympati eller har svarat på frågan om partival i senaste riksdagsval. I först hand väljs partisvar i lutar-frågan, i andra hand partisvar om partisympati och i tredje hand väljs ett parti slumpmässigt från en partifördelning, dvs.,

- 1) svar i lutar-frågan om partisvar finns,
- 2) svar om partisympati om ej partisvar i lutar-frågan men det finns partisvar om partisympati och
- 3) ett parti väljs slumpmässigt från en partifördelning

Partifördelningen i 3) beräknas med hjälp av svarande och de som imputerats enligt 1) eller 2). Beräkningen av partifördelning görs per region, utbildningsnivå, kön och partival i senaste riksdagsval, dvs. partifördelningen för val idag bland svarande ges av

$$p_l^{(ijkm)} = \frac{x_{ijklm}}{x_{ijk.m}} \quad (6.1)$$

där x_{ijklm} är antalet svarande i region, i , med utbildning, j , kön, k , partival nu, l , och partival vid senaste riksdagsval, m , och där $x_{ijk.m}$ är antalet svarande i region, i , med utbildning, j , kön, k , och partival vid senaste riksdagsval, m . Imputering görs genom att slumpa ett värde för

² En fråga som infördes i PSU i maj 2019 samtidigt som frågan om partiblock utgick.

partival nu från fördelningen i (6.1). Information om region, utbildningsnivå, kön och partival vid senaste riksdagsval från individen som ska imputeras används.

I tabell 6.3 redovisas antal imputerade värden för partival nu och partival i senaste riksdagsval. För partival nu redovisas antalen per imputeringsmetod samt totalt. För partival vid senaste riksdagsval redovisas endast antalet imputerade med hjälp av partifördelning.

Tabell 6.3. Antal imputerade partival nu och partival i senaste riksdagsval per omgång, november 10-november 2019

Omgång	Antal imputerade per metod (partival nu)			Totalt antal imputerade partival nu	Antal vet ej/blanktsvar	Antal imputerade: partival vid senaste riksdagsval	Antal svarande totalt	Urvalsstorlek
	Lutarfrågan	Partisympati	Partifördelning					
Nov10		240	264	504	552	213	6192	9054
Maj11		335	501	836	897	193	6147	9123
Nov11		426	601	1027	1107	194	5907	9060
Maj12		389	528	917	981	177	5473	9039
Nov12		390	509	899	961	167	5479	8999
Maj13		374	488	862	934	168	5098	9059
Nov13		413	543	956	1034	188	5267	9081
Maj14		423	429	852	927	191	4757	9085
Nov14		264	196	460	507	167	5072	9031
Maj15		465	365	830	887	160	6067	12091
Nov15		418	388	806	851	130	4972	9021
Maj16		405	360	765	808	127	4838	9033
Nov16		460	397	857	917	157	5021	8952
Maj17		383	370	753	810	145	4808	8973
Nov17		351	358	709	771	157	4715	8919
Maj18		385	329	714	780	151	4632	8951
Sep18 ³		307	213	520	569	136	3019	6014
Nov18		224	140	364	397	138	4721	8966
Maj19	197	105	142	444	467	85	4506	9092
Nov19	268	141	170	579	628	125	4645	9149

I omgångarna november 2010 till november 2018 fanns inte "lutarfrågan" och imputering har för dessa omgångar gjorts med hjälp av partisvar i frågan om partisympati eller med partifördelning. I imputering med partifördelning har svar i frågan om partiblock använts för den perioden, se vidare avsnitt 7.1.

³ Metodundersökningen innan valet 2018.

6.5 Slutgiltigt förslag

I den nya gemensamma GREG-estimatorn har vi en gemensam data-mängd och använder samma hjälpinformation för alla skattningar, nämligen

- Kön*ålder (6 åldersklasser*kön, se bilaga 1)
- Region (10 regioner, se bilaga 1)
- Utbildning (upp till gymnasial respektive eftergymnasial)
- Födelseland (inrikes eller utrikes född)
- Partival i senaste riksdagsval (partier C, L, KD, M, S, V, MP, SD, övriga partier, ogiltiga eller valskolkare och unga)

Partival i senaste riksdagsval är inte känd på förhand på individnivå utan samlas in i undersökningen. I den nya estimatorn kombineras alltså de hjälpvariabler som tidigare använts i skattningen av *Partisynpati* respektive *Val idag*.

I tabellen nedan beskrivs kortfattat det nya skattningsförfarandet i kolumnen Ny gemensam estimator i jämförelse med hur estimationen gjordes för *Val idag-skattningen* och *Partisynpatiskattningen* tidigare.

Tabell 6.4. Översikt över huvudsakliga skillnader mellan tidigare skattningsmetoder och den nya gemensamma metoden.

	Tidigare estimator för <i>Val idag-skattningen</i>	Tidigare estimator för <i>Partisynpati</i>	Ny gemensam estimator
Dataunderlag i skattningen	Urvalet exklusive övertäckning	Svarande	Svarande
Skattning	Poststratifiering: Partival vid senaste riksdagsval (11)*regioner (10)	GREG-estimator: Kön*ålder (12) Region (10) Utbildning (2) Utrikes/inrikes född(2)	GREG-estimator: Partival vid senaste riksdagsval (11) Kön*ålder (12) Region (10) Utbildning (2) Utrikes/inrikes född(2)
Imputering	Ja *målvariabel *hjälpinformation (partival vid senaste riksdagsval) Vet ej/blankt/skulle inte rösta, partiellt bortfall och objektsbortfall imputeras	Nej	* Hjälpvariabel (partival vid senaste riksdagsval) Ja *Målvariabel Ja för val idag

6.6 En jämförelse med metodundersökningen 2018

Genomgående i arbetet metodjusteringarna har vi stämt av skattningar i metodundersökningen mot valresultatet 2018. Sammanfattningsvis visar dessa jämförelser att för *Val idag-skattningen* skiljer sig den modifierade skattningsmetoden relativt lite jämfört med den tidigare metoden. Vi har fortsatt en relativt låg genomsnittsavvikelse när vi summerar skillnad mellan skattat resultat och det faktiska valresultatet.

Tabell 6.5 Jämförelse av resultatet i metodundersökningen med tidigare och ny skattningsmetod samt valresultatet, 2018

Parti	Valresultat 2018	Tidigare estimator	Ny estimator
C	8,6	8,8	9,1
L	5,5	6,7	6,4
M	19,8	18,4	19,2
KD	6,3	4,6	4,3
S	28,3	27,9	27,2
V	8,0	10	10,0
MP	4,4	5,5	5,0
SD	17,5	16,2	16,6
ÖVR	1,5	1,9	2,2
Genomsnittsavvikelse		1,1	1,0

Vid jämförelser som ovan bör man dock tänka på att mätningen görs veckorna innan valet och inte på valdagen, många är osäkra på sitt partival vid den tidpunkten och dessutom kan en faktisk valhandling inte helt likställas med att svara på en undersökning om ett kommande val. Metodundersökningen skiljer sig dessutom lite från den ordinarie undersökningen, se avsnitt 5.2. Även om metodundersökningen inte bör ses som ett absolut facit så ger den en bra indikation på kvaliteten i den nya skattningsmetoden.

7 Publicerade resultat och tidsserier

Den justerade metoden implementeras i och med majundersökningen 2020. I samband med det uppdaterar vi tabellerna i [Statistikdatabasen](#) från november 2010 och framåt med skattningar framtagna med den nya estimatorn.

Tidigare skattningar och nya skattningar till och med november 2019 kommer under en övergångsperiod att finnas tillgängliga i ett Excel-dokument på SCBs webbplats för jämförelser.

Sammanfattningsvis påverkas skattningarna, som förväntat, lite olika av de justeringar som vi har gjort. Fördelningen för *Val-idagskattningen*, totalt, påverkas mycket lite av de justeringar som vi har gjort, och alla förändringar är inom felmarginalen. Det beror på att vi redan tidigare hade partival vid senaste riksdagsval i hjälpinformationen samt att skillnaderna i imputeringsmetod inte är så stora i realiteten. Att lägga till registerinformation i estimatorn påverkar dock redovisningen efter kön och det gäller i synnerhet för partier där stödet skiljer sig i betydande grad mellan kvinnor och män. Skattningarna för olika regioner påverkas mindre eftersom vi även tidigare hade uppgift om region i hjälpinformationen.

Den justerade skattningsmetoden har något större inverkan på *Partisimpatiskattningen*. I synnerhet för de partier som är under- eller överrepresenterade bland de svarande när det kommer till röstning vid senaste riksdagsvalet eftersom vi lagt till den uppgiften i hjälpinformationen för partisynpati. Effekten blir tydligare i mindre redovisningsgrupper. I de mindre redovisningsgrupperna är samtidigt osäkerhetsintervallen bredare, så även om en förändring i skattningen är förhållandevis stor behöver det inte innebära att det är en statistiskt säkerställd skillnad mellan skattningarna.

7.1 Nya tidsserier, november 2010 – november 2019

Den gemensamma GREG-estimatorn har använts för att ta fram alla skattningar från november 2010 till november 2019. All hjälpinformation har funnits tillgänglig och har kunnat användas på samma sätt som är avsett även bakåt i tiden.

Däremot har imputeringen i de nya tidsserierna justerats något utifrån tidigare förutsättningar. Lutar-frågan infördes i partisynpatiuundersökningen i maj 2019, samtidigt som frågan om partiblock utgick. Imputering med hjälp av svar i lutar-frågan kan således bara göras

fr.o.m. maj 2019. För perioden november 2010 till november 2018 imputeras val idag enligt

- 1) svar om partisympati om partisvar finns
- 2) ett parti väljs slumpmässigt från en partifördelning

Imputering enligt 2) görs på liknande sätt som beskrivs i avsnitt 6.4.2, men om svar i frågan om partiblock (vid val idag) finns så används partifördelningen per parti i partiblocket istället för fördelning över alla partier. Dvs. om en individ har angett ett partiblock så väljs ett parti slumpmässigt från partiblocket enligt fördelningen inom blocket. Imputeringsgrupperna definierade av region, utbildningsnivå, kön och partival i senaste riksdagsval används även då svar om block finns.

På samma sätt som för imputering av val idag för perioden november 2010 till november 2018 används partiblock (i senaste riksdagsval) vid imputering av partival vid senaste riksdagsval om ett blocksvar finns.

Då bortfallet i undersökningarna innan 2010 var klart lägre jämfört med hur det sett ut under de senare åren bedöms det fortfarande möjligt att göra jämförelser från 2010 och bakåt trots att de skattningarna inte är uppdaterade.

8 Referenser

SCB (2019). Statistikens framställning Partisynpatiuundersökningen (PSU).

https://www.scb.se/contentassets/ed452064a5a54c189df3da7e576cdf4c/me0201_staf_2019_nov_rv_191203.pdf.

Särndal, C-E. och Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. Wiley & Sons.

Särndal, C-E. och Lundström, S. (2009). *Design for estimation: Identifying auxiliary vectors to reduce nonresponse bias*. Research and Development – Methodology reports from Statistics Sweden, 2009:1

Särndal, C.-E., Swensson, B. och Wretman J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag

Bilaga 1. Indelning av hjälpinformation

B1.1. Hjälpinformation i skattning av partisympati till och med november 2019.

Kön kombinerat med åldersgrupper där åldersgrupperna är definierade enligt tabell B1.

Tabell B1. Åldersklasser i hjälpinformationen

Åldersklasser	Ålder (år)
1	18-29
2	30-39
3	40-49
4	50-64
5	65-74
6	75-

Födelseland är indelat i två grupper, inrikes född och utrikes född. Utbildningsnivå är indelat i två grupper, upp till gymnasial nivå och eftergymnasial utbildning. Regioner är uppdelade i 10 grupper enligt tabell B2.

Tabell B2. Indelning av regioner i hjälpinformationen

Regioner	
1	Stor-Malmö
2	Övriga södra Götaland (Blekinge län och Skåne län exklusive Stor-Malmö)
3	Småland med öarna (Jönköpings län, Kronobergs län, Kalmar län och Gotlands län)
4	Göteborgs kommun
5	Övriga Västsverige (Hallands län och Västra Götalands län)
6	Stockholms kommun
7	Övriga Stockholms län (Stockholms län exklusive Stockholms kommun)
8	Östra Mellansverige (Uppsala län, Södermanlands län, Östergötlands län, Örebro län och Västmanlands län)
9	Norra Mellansverige (Värmlands län, Gävleborgs län och Dalarnas län)
10	Mellersta och övre Norrland (Västernorrlands län, Jämtlands län, Västerbottens län och Norrbottens län)

B1.2. Hjälpinformation i skattning av val idag till och med november 2019

För att bilda poststrata används regioner och partival vid senaste riksdagsval. Regionerna är desamma som i tabell B2 i avsnitt B1.1 och partival vid senaste riksdagsval är indelat 11 kategorier enligt tabell B3.

Tabell B3. Indelning av partival i senaste riksdagsval

Grupp	
1	C
2	L
3	KD
4	M
5	S
6	V
7	MP
8	SD
9	Övriga partier
10	Ogiltiga röster och ej röstat
11	Unga

För novemberomgången det år det är riksdagsval används endast de första 10, dvs. unga finns inte med.

Bilaga 2. Teknisk beskrivning av estimationen

Vi har en population U bestående av N personer. De parametrar vi är intresserade av är funktioner av två totaler $t_y = \sum_{k \in U} y_k$ och $t_z = \sum_{k \in U} z_k$, där y_k är värdet på variabeln y för person k och z_k värdet på en annan variabel för samma person. Om y (och även z) är en dikotom variabel, dvs. om

$$y_k = \begin{cases} 1 & \text{om person } k \text{ har den studerade egenskapen} \\ 0 & \text{annars} \end{cases} \quad (\text{B2.1})$$

så är $t_y = \sum_{k \in U} y_k$ antal personer i populationen med den studerade egenskapen.

Vi är också intresserade av parametrar för redovisningsgrupper dvs. delmängder av populationen. Låt oss benämna dessa U_1, U_2, \dots, U_D . Totalen för variabeln y i redovisningsgrupp d kan skrivas

$$t_{yd} = \sum_{k \in U} y_{dk} \quad (\text{B2.2})$$

där,

$$y_{dk} = \begin{cases} y_k & \text{om person } k \text{ tillhör redovisningsgrupp } d \\ 0 & \text{annars} \end{cases} \quad (\text{B2.3})$$

Totalen för variabeln z , t_{zd} , bildas på liknande sätt.

En generell parameter för redovisningsgrupp d (där d också kan avse hela populationen) kan skrivas $\theta_d = C \cdot \frac{t_{yd}}{t_{zd}}$, där C är en konstant. Om vi är intresserade av en procentuell andel, t.ex. ett partis andel av sympatier, sätts $C = 100$ och $z_k = 1$ för alla k , och y_k är definierad enligt (B2.1) dvs. $y_k = 1$ om person k sympatiserar med partiet. Om vi låter N_d vara antalet personer i redovisningsgrupp d , då kan parametern skrivas

$$P_d = 100 \cdot \frac{\sum_{k \in U} y_{dk}}{N_d} = 100 \cdot \frac{t_{yd}}{N_d} \quad (\text{B2.4})$$

En skattning av andelen i (B2.4) fås genom att skatta totalerna i täljare och nämnare i högerledet i (B2.4), dvs. en skattning ges av

$$\hat{P}_d = 100 \cdot \frac{\hat{t}_{yd}}{\hat{N}_d} \quad (\text{B2.5})$$

Om vi betraktar svarmängden, r , som ett obundet slumpmässigt urval från populationen av storlek N , kan vi skatta t_{yd} på följande sätt

$$\hat{t}_{yd} = \sum_{k \in r} \frac{N}{m} y_{dk}, \quad (\text{B2.6})$$

där m är antal svarande. Totalen N_d skattas på motsvarande sätt. I skattningen i (B2.6) används ingen hjälpinformation och den resulterande skattningen av andelen är helt enkelt andelen i svarmängden.

I syfte att erhålla en estimator med mindre urvalsfel och bortfallsskevhet än estimatorn (B2.6) används hjälpinformation i skattningen av totalerna i (B2.5). En hjälpvektor, \mathbf{x}_k , bildas som anger till vilka kategorier av hjälpvariablerna som person k tillhör. I PSU består hjälpvektorn av de indelningar av variabler som beskrivs i bilaga 1, dvs.

Åldersklasser*kön (12) + födelseland (2) + utbildningsnivå (2) + regioner (10) + partival vid senaste riksdagsval (11)

Hjälptotaler, totala antalet personer, i de olika indelningarna hämtas från Registret över totalbefolkningen (RTB) och från valresultat i senaste riksdagsvalet (variabeln partival vid senaste riksdagsval).

Den generaliserade regressionsestimatoern, GREG, för totalen t_{yd} ges av

$$\hat{t}_{yd} = \sum_{k \in r} (d_k / \hat{\theta}_k) g_k y_{dk} \quad (\text{B2.7})$$

där d_k är designvikten för person k , $\hat{\theta}_k$ är skattad svarssannolikhet för person k och där g_k är en justeringsfaktor som baseras på hjälpinformationen och ges av

$$g_k = 1 + \left(\sum_{k \in U} \mathbf{x}_k - \sum_{k \in r} d_k^* \mathbf{x}_k \right)' \left(\sum_{k \in r} d_k^* \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \mathbf{x}_k \quad (\text{B2.8})$$

där $\sum_{k \in U} \mathbf{x}_k$ är de kända populationstotalerna för hjälpvariablerna, $d_k^* = d_k / \hat{\theta}_k$ och r är svarmängden. För en mer utförlig beskrivning av GREG-estimatoern hänvisas till Särndal, Swensson och Wretman (1992).

I PSU är $d_k / \hat{\theta}_k = \left(\frac{N}{n} \right) / \left(\frac{m}{n} \right) = N/m$ dvs. populationsstorlek delat med antal svarande och den totala vikten för person k ges av $w_k = \frac{N}{m} * g_k$.

Ett partis andel av sympatier (Partisympti) i redovisningsgrupp d , skattas enligt formel (B2.5), där

$$\hat{t}_{yd} = \sum_{k \in r} \frac{N}{m} g_k y_{dk}$$

där $y_{dk} = 1$ om person k tillhör redovisningsgrupp d och sympatiserar med parti p dvs. om $y_k = 1$ och person k tillhör redovisningsgrupp d och $y_{dk} = 0$ annars. Skattningen av antalet personer i redovisningsgrupp d ges av

$$\hat{N}_d = \sum_{k \in r} \frac{N}{m} g_k z_{dk}$$

där $z_{dk} = 1$ om person k tillhör redovisningsgrupp d och $z_{dk} = 0$ annars.

Skattning av ett partis andel av väljarkåren i ett hypotetiskt val vid undersökningstillfället (*Val idag*) görs på liknande sätt som för partisympti med skillnaden att i *Val idag-skattningen* används röstningssannolikheter, se avsnitt 4.1.2. Variabeln y_k , se (B2.1), ges av

$$y_k = \begin{cases} q_k & \text{om person } k \text{ skulle rösta på parti } p \\ 0 & \text{annars} \end{cases}$$

där q_k är röstningssannolikheten för person k och i skattningen \hat{N}_d ges z_{dk} av

$$z_{dk} = \begin{cases} q_k & \text{om person } k \text{ tillhör redovisningsgrupp } d \\ 0 & \text{annars} \end{cases}$$

SCB beskriver Sverige

Statistikmyndigheten SCB förser samhället med statistik för beslutsfattande, debatt och forskning. Vi gör det på uppdrag av regeringen, myndigheter, forskare och näringsliv. Vår statistik bidrar till en faktabaserad samhällsdebatt och väl underbyggda beslut.