



UPPSALA
UNIVERSITET

Om det vore val idag...?

En studie om att predicera valresultat i svenska riksdagsval med strukturella modeller och bayesiansk regression

Av: Ida Frölander & Leonora Uddhammar

Statistiska institutionen
Uppsala universitet

Handledare: Måns Magnusson

HT 2020

Sammanfattning

Vi har i denna uppsats undersökt strukturella faktorer som kan användas för att predicera valresultat i svenska riksdagsval minst tre månader före valdagen. Uppsatsens syfte motiveras utifrån en empirisk kunskapslucka i forskning kring faktorer som kan tänkas förutspå svenska riksdagsval. Vi har utifrån bayesiansk regression och "Leave-One-Out Cross-Validation" identifierat två slutgiltiga modeller. Den första modellen består av Statistiska centralbyråns partisynpatiuundersökning från maj under valåret och en offset som är definierad utifrån valresultat i föregående riksdagsval. Den andra modellen består av faktorerna arbetslöshet och partiernas omnämnande i media samt en offset som också definierats utifrån partiets valresultat i tidigare riksdagsval. Modellerna har utvärderats genom att predicera valresultatet i riksdagsvalet 2018 där de fick ett genomsnittligt prediktionsfel (RMSE) på 1,2 respektive 2,0 och där samtliga 90% prediktionsintervall innefattade de sanna valresultaten. Modellerna har jämförts med en modell från tidigare forskning av Sundell och Lewis-Beck (2014). Resultaten visar att modellerna som har identifierats i denna uppsats har ett lägre prediktionsfel och snävare prediktionsintervall jämfört med modellen från studien av Sundell och Lewis-Beck. Våra resultat indikerar även att normalfördelningen - som varit normen i tidigare studier av strukturella modeller - inte är optimalt att använda i länder med flerpartisystem. Ett normalfördelningsantagande tycks bidra till att den prediktiva fördelningen och prediktionsintervallen blir för snäva.

Nyckelord: valresultat, riksdagsval, flerpartisystem, strukturella modeller, prediktioner, prediktionsintervall, bayesiansk regression

Förord

Vi skulle vilja rikta ett stort tack till Miriam Hurtado Bodell som varit vänlig att hjälpa oss med uppsatsen genom att ta fram data över riksdagspartiernas omnämnande i media. Vi skulle också vilja rikta ett stort tack till vår handledare Måns Magnusson för hjälpfulla kommenterar och intressanta diskussioner under uppsatsskrivandet.

Innehållsförteckning

1. Introduktion	5
<i>Syfte och frågeställning.....</i>	<i>6</i>
2. Tidigare forskning.....	7
<i>Strukturella modeller.....</i>	<i>7</i>
<i>Strukturella modeller i flerpartisystem</i>	<i>8</i>
<i>Strukturella modeller i Sverige.....</i>	<i>11</i>
Partiledarpopularitet	13
Näst bästa parti.....	13
Omnämnande i media.....	14
<i>Avgränsningar.....</i>	<i>15</i>
3. Datamaterial.....	16
<i>Beroende variabel</i>	<i>16</i>
<i>Förklarande variabler</i>	<i>16</i>
4. Metod.....	19
<i>Bayesiansk statistik och inferens</i>	<i>19</i>
<i>Bayesiansk multipel linjär regression.....</i>	<i>20</i>
<i>Den prediktiva fördelningen för framtida observationer och prediktionsintervall</i>	<i>21</i>
<i>Tolkning av parameterestimater.....</i>	<i>22</i>
<i>Modellutvärdering.....</i>	<i>22</i>
<i>RMSE.....</i>	<i>23</i>
<i>Prediktionsintervall</i>	<i>24</i>
<i>Kolmogorov-Smirnov Goodness-of-Fit Test.....</i>	<i>24</i>
<i>En alternativ metod: frekventistisk linjär regression (OLS).....</i>	<i>26</i>
5. Resultat.....	28
<i>Utvärdering av modell 1–8</i>	<i>30</i>
<i>Utvärdering av riksdagsvalet 2018.....</i>	<i>36</i>
6. Diskussion och slutsatser	39
Källförteckning.....	42
Appendix A: Q-Q plots för modell 1–8.....	45
Appendix B: Variabler och källor	50

1. Introduktion

Att förutspå valresultat har blivit en framträdande del av samhällsdebatten i många demokratiska länder. En stor del av debatten och forskningen har framförallt fokuserat på de amerikanska presidentvalen där det finns en lång tradition av opinionsmätningar. Intresset för valprognoser nådde dock nya höjder när Nate Silver, grundaren av nyhetssidan FiveThirtyEight, inför 2008 års presidentval korrekt lyckades förutspå valresultaten i alla delstater förutom Indiana. Inför presidentvalet 2016 lyckades dock Silver (och de flesta andra prognosmakarna) sämre då Hillary Clinton förlorade valet till Donald Trump trots att FiveThirtyEight gav henne en sannolikhet på cirka 70% att vinna. Förvisso innebär detta även att det fanns en sannolikhet på 30% att Hillary Clinton skulle förlora – vilket kan anses vara relativt högt – men att korrekt förutspå valresultat tycks uppenbarligen vara en svår uppgift trots att tillgängligheten av data och opinionsmätningar aldrig har varit större.

Lewis-Beck och Stegmaier (2014) har delat in tillvägagångssätten för att förutspå valresultat i fyra huvudsakliga grupper: strukturella modeller (“structuralists”), opinionsmätningar (“aggregators”), en kombination av strukturella modeller och opinionsmätningar (“synthesizers”) och expertprognoser (“judges”). Strukturalisterna använder regression för att skatta den sittande regeringens stöd baserat på olika grundläggande ekonomiska och politiska faktorer. Den andra gruppen (aggregators) använder istället opinionsundersökningar där ofta flera olika mätningar vägs samman för att få mer korrekta och tillförlitliga prognoser. Den tredje gruppen (synthesizers), som Nate Silver och hans prognoser på FiveThirtyEight tillhör, kombinerar strukturella modeller med opinionsmätningar. Den sista gruppen (judges) gör en kvalitativ bedömning av olika informationskällor. Varje grupp är förknippad med sina egna för- och nackdelar och den främsta samt oftast diskuterade skillnaden är mellan hur långt i förväg man vill göra prognosen och hur precis man vill att den ska vara (Vestin och Öhrvall 2020). Strukturella modeller har blivit populära eftersom de har visat sig relativt träffsäkra en längre tid före val medan opinionsundersökningars precision ökar ju närmare valdagen de mäts. Det har därför blivit vanligt att kombinera dessa två metoder och ge dem olika vikt baserat på hur lång tid det är kvar till valet (Ibid.).

Studier som fokuserar på att förutspå riksdagsvalen i Sverige är ett relativt outforskat ämne. Förvisso har det länge genomförts opinionsundersökningar som mäter hur svenska väljare avser att rösta men specifika valmodeller lyser nästan helt med sin frånvaro. Inför riksdagsvalet 2014

gjordes dock ett par försök där två av dessa var sajten “Botten Ada”, utvecklad av Måns Magnusson och Jens Finnäs, samt sajten trefyrario, utvecklad av Daniel Walther (Öhrvall 2014). Båda sajterna använde sig av bayesiansk statistik och opinionsundersökningar för att predicera utgången i riksdagsvalet. Ytterligare ett försök gjordes av Sundell och Lewis-Beck (2014) som istället för att använda opinionsundersökningar konstruerade en strukturell modell bestående av två ekonomiska variabler (genomsnittlig inflation och förändring i arbetslöshet under mandatperioden). Modellen predicerade att den sittande regeringen skulle få 49,7% av rösterna i valet i september 2014, vilket visade sig vara en överskattning med 10,3 procentenheter. Mot denna bakgrund är det tydligt att det finns utrymme för vidare utforskning av teoretiskt motiverade variabler såväl som en förbättring av precisionen i prognoserna av svenska riksdagsval.

Syfte och frågeställning

Denna uppsats kommer att försöka bidra till dessa kunskapsluckor genom att fokusera på att identifiera strukturella faktorer som kan användas för att predicera valresultat i svenska riksdagsval. De övergripande forskningsfrågorna som uppsatsen avser att besvara är:

- Vilka faktorer kan bäst förutspå valresultatet i svenska riksdagsval minst tre månader före valdagen?
- I vilken utsträckning kan vår modell generera väl kalibrerade prediktionsintervall för valresultatet i svenska riksdagsval minst tre månader före valdagen?

För att avgöra vilka strukturella faktorer som bör studeras närmare kommer vi först att presentera en litteraturöversikt av tidigare forskning kring strukturella modeller som använts för att prognostisera valresultat både internationellt och i Sverige.

2. Tidigare forskning

Strukturella modeller

Strukturella modeller har länge varit den vanligaste metoden för att förutspå valutgången i nationella val (Lewis-Beck och Dassonnville 2015a: 275–6). Dessa modeller bygger på idén att det utifrån ekonomiska och politiska faktorer går att förklara hur folk kommer att rösta. En fördel jämfört med opinionsundersökningar är därmed att de strukturella modellerna kan hjälpa oss förstå orsakerna till ett visst valresultat och vilka faktorer som spelar roll för hur människor röstar (Ibid.). De vanligaste ekonomiska faktorerna, ofta kallade ”de stora två”, är någon form av mått på ekonomisk tillväxt (mätt i BNP) och arbetslöshet (Lewis-Beck och Paldam 2000). Även något mått som mäter inflation är en vanlig faktor som ofta inkluderas. Strukturella modeller har sin teoretiska grund i litteraturen om ekonomisk röstning, vilket kort kan sammanfattas som idén att människor röstar utifrån det samhällsekonomiska läget. Detta fält utgår bland annat från ”ansvarshypotesen” (the responsibility hypothesis) som antar att väljare ser den sittande regeringen som den huvudsakliga ansvariga aktören för en positiv såväl som negativ samhällsekonomisk utveckling – om ekonomin går bra kommer väljarna att belöna regeringen och om den går dåligt kommer regering att straffas (Ibid.: 114).

En strukturell modell som fått stor genomslagskraft är den så kallade ”Time-for-Change”-modellen som konstruerades av Abramowitz 1988 för att förutspå valutgången i amerikanska presidentval (Abramowitz 1988, 2008). Modellen bygger på idén att presidentvalen huvudsakligen är en omröstning om huruvida den sittande presidenten har presterat väl under sin mandatperiod, oavsett om hen ställer upp i det aktuella valet eller inte (Ibid.). Modellen består av tre förklarande variabler: ekonomins tillväxthastighet under valårets andra kvartal, den sittande presidentens popularitet runt valårets andra kvartal och hur länge den sittande presidentens parti innehaft presidentämbetet. Inför 2008 års presidentval prognostiserade Abramowitz med hjälp av denna modell att Obama skulle få 54,3% av rösterna. Det slutgiltiga valresultatet visade att Obama fick 52,9% av rösterna.

Varianter av de två första förklarande variablerna i Abramowitz modell utgör grunden för de flesta strukturella modeller, som enligt Lewis-Beck (2005: 150) sammanfattas på följande vis:

$$\text{Röster sittande regering} = \text{samhällsekonomiska läget}_t + \text{regeringens popularitet}_t + \varepsilon$$

Strukturella modeller i flerpartisystem

Strukturella modeller har huvudsakligen utvecklats utifrån en amerikansk kontext – ett presidentiellt tvåpartisystem. Många länder, däribland Sverige, är dock ett flerpartisystem där fler än två partier tävlar om makten och där koalitioner ofta är nödvändiga för att inta regeringsmakten. En potentiell utmaning med dessa länder är att definiera den beroende variabeln eftersom valresultatet inte längre är ett tydligt nollsummespel mellan endast två partier/kandidater (Nadeau & Lewis-Beck 2020; Walther 2015). De flesta studier som undersökt länder med flerpartisystem har dock valt att behålla definitionen av den beroende variabeln som andel röster på den sittande regeringen, även när detta är en koalition av flera partier (se till exempel Norpoth & Gschwend 2003, 2010; Lewis-Beck & Dassonneville 2015a¹, 2015b; Foucault & Nadeau 2012, Magalhães et al. 2012, Sundell & Lewis-Beck 2014, Nadeau et al. 2012). Vissa varianter förekommer dock, till exempel har två studier av en skandinavisk kontext (Norge respektive Danmark) istället använt andel röster på vänsterblocket som beroende variabel (Arnesen 2012; Nadeau & Lewis-Beck 2020). En annan variant användes för en studie i Belgien där författarna förvisso använde andel röster på den sittande regeringen som beroende variabel, men till skillnad från andra studier skattade de först varje parti individuellt (Dassonneville & Hooghe 2012). Författarna kunde göra detta genom att inkludera andelen aktiva partimedlemmar för respektive parti som en förklarande variabel i modellen med argumentet att detta bör fånga partiernas relativa popularitet.

När det kommer till valet av förklarande variabler finns det en något större variation i litteraturen, även om det fortfarande nästan uteslutande rör sig om olika varianter av det samhällsekonomiska läget och regeringens popularitet. Exakt vilka varianter och kombinationer av dessa variabler som har använts beror på valkontexten i det land som modellen utvecklats för (Dassonneville & Hooghe 2012; Lewis-Beck & Tien 2011). I den översikt vi har gjort är BNP-tillväxt och arbetslöshet de två vanligaste variablerna, vilka som tidigare nämnts också utgör ”de stora två”. Även ett mått på den sittande regeringens popularitet, antingen från en opinionsundersökning eller en liknande survey, används nästan alltid. Det kan verka märkligt att en strukturell modell inkluderar en opinionsundersökning som förklarande variabel eftersom en opinionsmätning inte rimligtvis kan anses vara strukturell – den bör snarare ses som ett utfall av potentiella strukturella faktorer. Utifrån vad vi har sett är detta dock inget som har problematiserats eller diskuterats i tidigare forskning.

¹ Lewis-Beck och Dassonneville (2015a, 2015b) använder egentligen en kombination av metoder (synthesizers), där de i första steget definierar och estimerar en strukturell modell och i nästa steg använder sig av opinionsundersökningar.

Vi kan vidare konstatera att dessa relativt få förklarande variabler generellt har lyckats producera träffsäkra prediktioner av nationella valresultat (lågt prediktionsfel). I en studie på tyska val under åren 1980–2009 av Lewis-Beck och Dassonneville (2015a) kunde exempelvis 82,2% av variationen i andel röster på den sittande regeringen förklaras av de två variablerna BNP-tillväxt och andel stöd för regeringen, båda med en ledtid på sex månader. Modellen hade vidare ett prediktionsfel på 0,5 procentenheter för valet till Förbundsdagen år 2011. Just sex månaders ledtid har i flera studier visat sig ge bäst prediktioner (se till exempel Magalhães et al. 2012, Lewis-Beck & Dassonneville 2015a). Även studier som ämnar att prognostisera presidentvalet i USA har visat att denna ledtid har givit bäst effekt, med förklaringen att detta är så pass nära valet man kan komma före dess att den prediktiva kraften i regeringens popularitet påverkas i för stor utsträckning av partiernas valkampanjande (Lewis & Rice 1992: 123).

En utmaning i flerpartisystem som har lyfts fram i litteraturen är vem eller vilket parti som ska hållas ansvarig för den ekonomiska politiken när en koalitionsregering styr landet. Det är inte självklart att väljare straffar respektive belönar alla regeringspartier på samma sätt, vilket gör att effekten av ”ansvarshypotesen” kan ses som svagare i flerpartisystem jämfört med tvåpartisystem (Lewis-Beck & Paldam 2000: 119; Anderson 2000). Walther (2015), som i en studie prognostiserat valresultat i Sverige och Tyskland med hjälp av opinionsundersökningar, menar att svårigheterna med ekonomiskt och politiskt ansvarsutkrävande bidrar till att strukturella modeller inte är lämpliga att applicera på flerpartisystem. Vissa forskare anser dock att en del av denna komplexitet kan fångas upp genom att inkludera en interaktionsterm i den strukturella modellen (se till exempel Dassonneville & Hooghe 2012; Magalhães et al. 2012; Arnesen 2012). Magalhães et al. (2012) beskriver logiken med en interaktionsterm som ett försök att fånga en sorts ”policyorienterad” röstning istället för en ”incumbent-oriented” (vilket är den typ av röstning som bland annat Abramowitz ”Time-for-Change model” bygger på). Hypoteser kring policyorienterad röstning grundar sig i studier som visat att en ökad arbetslöshet kan leda till högre, eller åtminstone inte försämrat, stöd för vänsterorienterade partier medan stödet för högerpartier tvärt emot förväntas sjunka om arbetslösheten ökar (Ibid.). Effekten av arbetslöshet kan således skilja sig beroende på om landet styrs av en vänsterregering eller högerregering. Magalhães et al. (2012) menar att detta kan förklaras utifrån det faktum att partier till vänster traditionellt sett har prioriterat arbetslöshetsfrågan högt och satsat mer pengar på olika sociala och ekonomiska skyddsnät. Författarna kom vidare fram

till att den bästa modellen i en spansk kontext inkluderade en interaktionsterm mellan arbetslöshetsnivån och en dummy som antog värdet 1 om det största högerpartiet var i regering och värdet 0 om det inte var det, samt inflationstakt och andelen medborgare som tyckte att den politiska situationen var bra eller mycket bra. Samtliga variabler hade en ledtid på sex månader och regressionskoefficienten tillhörande interaktionstermen fick som förväntat ett negativt värde. Modellen hade en förklaringskraft justerad R^2 -värde på 93% och predicerade att den sittande regeringen skulle få 34,5% av rösterna i valet 2011, vilket var en överskattning med 5,8 procentenheter.

Avslutningsvis bör något också sägas om den metod som huvudsakligen har använts för att skatta strukturella modeller. Tidigare forskning har nästan uteslutande använt sig av multipel linjär regression som skattats genom OLS och där modellerna appliceras på så många tidigare val som möjligt (Lewis-Beck & Dassonneville 2015). Den ”bästa” modellen har sedan använts för att göra prediktioner för ett kommande val. Stickprovsstorleken i studierna är vanligtvis liten och varierar oftast mellan 10 till 15 observationer (Linzer 2013). Detta beror framförallt på att det helt enkelt inte finns så många val att studera och att det ibland saknas data för strukturella faktorer långt bak i tiden. Det förekommer dock vissa studier där det utifrån det specifika valsystemet är möjligt att kombinera val på flera nivåer (exempelvis lokala eller regionala) med de nationella och på så vis erhålla större stickprov (se exempelvis Foucault & Nadeau 2012; Arnesen 2012). Det är vidare vanligt förekommande att modellerna utvärderas utifrån det justerade R^2 -värdet, där ett högt värde anses tyda på en bra modell. Man kan dock fråga sig hur relevant detta utvärderingsmått faktiskt är. Vi har ovan beskrivit hur studier som använt 2–3 förklarande variabler har uppnått justerade R^2 -värden på över 80% (ibland även över 90%), vilket kan anses högt för strukturella modeller (Lewis-Beck 2005). Magalhães et al. (2012) hade dock en modell där justerad $R^2 = 93\%$ men där modellen ändå hade ett tämligen högt prediktionsfel på 5,8 procentenheter. Om det är relativt ”lätt” för strukturella modeller att uppnå ”höga” R^2 -värden och det inte är en garanti för bra prediktioner så kan man ifrågasätta dels vad som ska klassas som ett ”högt” R^2 -värde och dels om det överhuvudtaget är ett relevant utvärderingsmått för strukturella modeller i nationella val.

Vi har vidare inte hittat någon tidigare studie som använder bayesiansk linjär regression för att skatta en strukturell modell för nationella val i flerpartisystem. Det är däremot vanligt bland ”synthesizers” att först skatta en strukturell modell genom ”klassisk” OLS och sedan applicera en bayesiansk metodologi där den strukturella modellen används som en a-priori-fördelning

tillsammans med opinionsundersökningar (data) för att förutspå valutgången (se exempelvis Linzer 2013). Mer specifikt används ofta vad som kallas en Dynamic Linear Model (DLM) vilket blivit en ledande metod bland prognosmakare och som ofta visat sig generera träffsäkra prediktioner (Walther 2015). Om man vill kombinera en strukturell modell med en DLM är det viktigt att få en uppskattning av osäkerheten i prediktionerna, vilket vi bäst kan få genom bayesianska prediktionsintervall.² Vi kommer därför i analysen nedan att använda oss av bayesiansk statistik för att undersöka våra strukturella modeller.

Strukturella modeller i Sverige

Studier som försöker förutspå valresultaten i Sverige är ett sällsynt fenomen. Vad vi har kunnat hitta är studien av Sundell och Lewis-Beck (2014) en av få, och den enda där en strukturell modell har använts i en svensk kontext. Som beroende variabel valde Sundell och Lewis-Beck att använda andelen röster på partier som ingår i regeringsunderlaget, alltså inte endast de partier som formellt ingår i regeringen. Författarna studerade valen mellan 1973 och 2010 för att sedan göra en prediktion gällande utgången i riksdagsvalet 2014. Författarna började med att identifiera och testa tre förklarande variabler: inflation, BNP-tillväxt och arbetslöshet. Samtliga variabler mättes vid två tillfällen: som ett genomsnitt av mandatperioden mellan riksdagsval och första kvartalet under valåret.³ Variablerna som var ett genomsnitt av mandatperioden visade högst korrelation med utfallsvariabeln och författarna valde därför att använda dessa i fyra olika modeller genom multipel linjär regression (OLS). Resultaten visade dock att BNP-tillväxt hade en negativ effekt på den beroende variabeln vilket författarna inte kunde motivera teoretiskt och de valde därför att inte inkludera denna variabel i den slutgiltiga modellen som presenteras nedan:

*Government and support parties vote share = 50.6 - 1.1*Unemployment term change - 0.4*Inflation term average.*

En intressant fråga utifrån deras modell är om det vore mer relevant att använda sig av en offset istället för ett intercept vid denna typ av modell. En offset ersätter ett intercept och estimeras inte på samma sätt utan specificeras istället på förhand. Givet att vi inte skulle ha någon information om valutgången är det rimligt att anta att de flesta väljare skulle rösta likadant som

² För vidare resonemang kring detta se s. 26–27 i metodavsnittet nedan.

³ Genomsnittet av mandatperioden är mätt som första kvartalet året efter ett riksdagsval och första kvartalet under det aktuella riksdagsvalet.

de gjort tidigare år (se även figur 1, s. 30). Resultatet i det föregående riksdagsvalet skulle således kunna vara en rätt bra indikation på hur kommande resultat skulle bli. Självklart kommer vissa personer byta parti och det går därför att argumentera att det är just röstningsbeteendet bland denna grupp som vi främst bör fokusera på att försöka förklara. Detta resonemang finner även stöd i forskning som har visat andelen rörliga väljare (de som byter parti) förvisso har ökat från 16% 1973 till 32,6% 2010 men att en klar majoritet ändå röstar som de gjorde i det föregående valet (Oscarsson & Holmberg 2013: 15). Använder vi en modell med en offset som är specificerad utifrån det tidigare valresultatet förklarar vi det kommande valresultatet för just de väljare som har bytt parti sedan det senaste riksdagsvalet, vilket vi anser är ett bättre sätt att specificera modellen än med ett ospecificerat intercept. Ytterligare en fördel med att försöka förklara valresultat bland de som byter parti är att denna grupp potentiellt är mer benägna att påverkas av strukturella faktorer än de som inte byter parti. Trogna väljare som i praktiken alltid röstar på samma parti gör det troligtvis av mer ideologiska skäl än på grund av strukturella faktorer, såsom ekonomin, arbetslöshet och partiledarpopularitet, då dessa fluktuerar över tid. En ökning av andelen rörliga väljare över tid skulle då indikera att färre väljare röstar utifrån ideologiska värderingar och alltmer på basis av strukturella faktorer.

Sundell och Lewis-Beck testar i deras studie även att estimerar en modell utifrån en opinionsundersökning (Sifos maj-mätning 2014). Denna modell predicerar att den sittande regeringen (och stödpartierna) kommer att få 43,4% av rösterna i riksdagsvalet 2014 – vilket är en stor skillnad från deras första prediktion på 49,7%. Författarna testar också att kombinera deras ursprungliga modell med en opinionsundersökning men konstaterar att detta inte resulterar i en mer kraftfull modell med motiveringen att tillskottet av opinionsundersökningen bara marginellt ökar R^2 -värdet och minskar det justerade R^2 -värdet. Som beskrivits tidigare kan dock denna utvärderingsmetod problematiseras. Slutligen kan vi också konstatera att modellen av Sundell och Lewis-Beck skiljer sig från övriga studier som använt strukturella modeller då deras modell inte innehåller någon förklarande variabel som mäter regeringens popularitet.

Modellen som presenterats av Sundell och Lewis-Beck resulterade således inte i en särskilt träffsäker prediktion av riksdagsvalet 2014 då det faktiska valresultatet för regeringen (och stödpartierna) blev 39,4%. Författarna skriver även själva att det inte finns några variabler kring vilka det råder konsensus om i vilken utsträckning de påverkar en regerings chanser att bli omvalda i Sverige. Det finns dock mycket forskning, framförallt från det statsvetenskapliga

fältet, om svenskars röstningsbeteende. Det är således möjligt att vi utifrån den litteraturen kan hitta potentiella faktorer som kan vara relevanta när vi försöker predicera valresultat. Dessa redogörs för nedan.

Partiledarpopularitet

Forskning har visat att det har skett en personifiering i politiken där väljare i allt större utsträckning röstar på en person istället för ett parti (Manin 1997). En omdiskuterad faktor kopplat till detta är hur viktig partiledaren är för hur många röster ett parti får i riksdagsvalen. Partiledaren kan ses som den yttersta representanten för sitt parti och en central aktör för hur partiet presterar i opinionsundersökningar och val (Hagevi et al. 2015). Om partiledaren uppfattas positivt av väljarkåren kan hen dra till sig röster (flöte) men om partiledaren istället är impopulär kan hen ha en avstötande effekt (sänke) (Ibid.). Det är dock inte entydigt vilken eller hur stor betydelse partiledaren har för partiets valresultat och forskningen pekar även på att partiledareffekter är svagare i proportionella valsystem (Holmberg & Oscarsson 2011). Hagevi (2015: 108) sammanfattar de svenska forskningsresultaten på området på följande vis: “partiledareffekterna är begränsade, men inte obefintliga, och dess effekter är störst bland borgerliga väljare.” Valforskningsprogrammet vid Göteborgs universitet har även visat att det finns en svagt ökande trend sedan 1979 som pekar på att partiledarnas popularitet har en effekt på hur folk väljer att rösta (Holmberg & Oscarsson 2013). Sammantaget anser vi att det finns stöd för att partiledarpopularitet kan utgöra en potentiell förklarande faktor när det kommer till att försöka förutspå valresultat i svenska riksdagsval.

Näst bästa parti

En annan potentiell faktor är inspirerad från litteraturen kring partiidentifikation och rörliga väljare. Undersökningar har visat att allt färre väljare anser sig vara starka partianhängare, från cirka 50% av väljarna vid valet 1964 till cirka 15–17% vid valet 2010 (Oscarsson & Holmberg 2013). Detta samvarierar i sin tur med att allt fler väljare också mer frekvent byter parti mellan olika riksdagsval. Tittar vi på väljare som byter mellan blockgränserna är det dock fortfarande ovanligt då denna grupp utgör cirka 10% av väljarna under valen sedan millennieskiftet (Ibid.) Det är således alltmer vanligt med rörliga väljare men fortfarande ovanligt att väljare byter block. Det är därmed möjligt att tänka sig att väljare kan byta parti relativt lättvindigt så länge det finns ett alternativ som ligger nära ideologiskt inom samma block. Oscarsson (2020: 1) har undersökt just detta och konstaterar att “de som byter parti i huvudsak byter till ideologiskt närstående grannpartier i partisystemet.” En variabel som skulle kunna fånga detta

röstningsbeteende är hur många väljare som har ett visst parti som sitt näst bästa alternativ. Om något händer som gör att förstahandsvalet inte längre är lika attraktivt så är ett rimligt antagande att väljaren istället går till andrahandsvalet.

Omnämnande i media

På ett övergripande plan kan opinionsbildning sägas handla om att framföra sin åsikt och övertyga andra om att partiet är det bästa regeringsalternativet. Ett viktigt medel att sprida denna politiska information är genom media. Journalister pekas ofta ut som en maktfaktor i valrörelser då de kan styra vilka frågor som lyfts fram och vilka perspektiv som ska speglas. Med hjälp av media kan partier också själva sprida sitt budskap, både genom debattinlägg och genom att journalister rapporterar från presskonferenser eller dylikt. Ett kvantitativt mått på ett partis omnämnande i media kan således ses som ett utfall och/eller en proxy av det opinionsbildningsarbete som ett parti utför. Tidigare forskning från en nederländsk respektive dansk kontext har även visat att politikernas och partiernas omnämnande i media (visibility) har en positiv effekt på deras valresultat (Oegema & Kleinnijenhuis 2000; Hopmann et al. 2010). Det finns således skäl att tro att partiernas omnämnande i media kan förutspå valresultat även i en svensk kontext.

Sammanfattningsvis går det att konstatera att det finns ett behov av att identifiera och testa faktorer med både en hög prediktiv styrka och träffsäkerhet för en strukturell modell i en svensk kontext. För att fylla detta behov kommer vi i denna uppsats att använda bayesiansk linjär regression för att undersöka olika strukturella modeller. Uppsatsen kommer att fortgå med följande disposition: på nästkommande sida ges en kort redogörelse av de avgränsningar vi gjort i uppsatsen. Därefter presenteras den data som resultaten bygger på samt en beskrivning av de variabler som ingår i de olika modellerna. Sedan följer ett metodavsnitt där uppsatsens metod presenteras och motiveras. Efter detta kommer uppsatsens resultat att presenteras. Uppsatsen avslutas sedan med ett diskussionsavsnitt där resultaten och slutsatser diskuteras.

Avgränsningar

Vi har i denna uppsats gjort ett antal avgränsningar. Vi har exempelvis valt att inte inkludera riksdagsvalen före 1973 i vår analys. Det finns två huvudsakliga skäl till denna avgränsning. Det första grundar sig i att även Sundell och Lewis-Beck valde att exkludera riksdagsvalen före 1973 i deras studie och genom att använda samma avgränsningsperiod blir det lättare att jämföra våra modeller. Det är vidare svårt att hitta data på många av de förklarande variabler som vi vill använda i uppsatsens modeller längre tillbaka än 1970-talet och ett utökat antal studerade riksdagsval hade således lett till att vi hade behövt exkludera ett antal potentiella förklarande variabler ur vår analys.

Ytterligare en avgränsning i uppsatsen är att vi inte kommer att predicera valresultat för Sverigedemokraterna (SD). Partiet har suttit i riksdagen sedan 2010 och var för första gången med i Statistiska centralbyråns (SCB:s) partisympatiundersökning (PSU) inför riksdagsvalet 2014. Om vi vill använda oss av PSU:n som förklarande variabel har vi därför för få datapunkter för att kunna predicera valresultatet för SD i riksdagsvalet 2018. För en diskussion kring en eventuell lösning på detta problem, se avsnitt 6.

3. Datamaterial

Vi kommer i detta avsnitt presentera de variabler som vi utifrån tidigare forskning har identifierat som potentiella förklarande variabler för valresultat i svenska riksdagsval. Vissa av de utvalda variablerna har tidigare testats i strukturella modeller och visat sig ha en god prediktiv förmåga. Vissa variabler har inte tidigare använts i strukturella modeller men har i tidigare forskning identifierats som relevanta för hur väljare i Sverige röstar. Nedan följer en specifikation av de variabler som kommer ingå i resultatdelen av uppsatsen.

Beroende variabel

Valresultat: Denna variabel beskriver respektive parti eller sittande regerings valresultat⁴ (GOV) vid varje riksdagsval. Variabeln avser andel röster (av giltiga röster) totalt för riket. Datan är inhämtad från SCB.

Förklarande variabler

Valresultat i föregående riksdagsval (Res_tid_val): Denna variabel beskriver respektive parti eller sittande regerings valresultat vid föregående riksdagsval. Variabeln avser andel röster (av giltiga röster) totalt för riket. Datan är inhämtat från SCB.

PSU_majdiff: Denna variabel beskriver skillnaden mellan SCB:s PSU i maj under valåret och resultatet i föregående riksdagsval.⁵ Resultatet från undersökningen visar "valresultatet om det varit val idag". År 1982 genomfördes inte undersökningen och vi använder därför istället resultatet från maj 1981.

Arbetslöshet: Denna variabel beskriver skillnaden i andelen arbetslösa mellan mars under valåret och oktober föregående valår. Om skillnaden är positiv har arbetslösheten ökat under mandatperioden och om den är negativ har arbetslösheten minskat. Valet av mars månad är baserat på tidigare forskning som visat att en ledtid på sex månader är mest effektiv. Valet av oktober månad är baserat på att det vanligtvis är den månad då den nya regeringen tillträder. Datan är inhämtad från SCB:s arbetskraftsundersökning (AKU) och avser andel arbetslösa i åldrarna 15–74 år efter arbetskraftstillhörighet. För perioden före januari 2001 var populationen definierad som 16–64 år.

⁴ Variabeln sittande regering (GOV) definieras genomgående i denna uppsats som valresultatet för regeringspartier+stödpartier.

⁵ Variabeln är framräknad genom att ta PSU i maj under valåret minus resultatet föregående riksdagsval.

Sittande_regering (Inc): Denna variabel beskriver om respektive parti är en del av regeringsunderlaget eller inte på valdagen. Variabeln är en dummy som antar värdet 1 om partiet sitter i regering eller agerar stödparti till regeringen och värde 0 om partiet varken är i regering eller agerar som stödparti.

*Arbetslöshet*sittande_regering:* Denna variabel är en interaktionsterm mellan variablerna arbetslöshet och sittande regering. Partier som inte är i regering bör rimligtvis inte påverkas lika negativt av en ökad arbetslöshet som de partier som sitter i regering då de inte kan anses vara ansvariga för utfallet av den politik som förs.

Inflation: Denna variabel mäter förändringen i inflationsnivå under mandatperioden. Variabeln är framräknad genom att ta skillnaden i Konsumentprisindex (KPI) mellan mars under valåret och oktober föregående valår. Datan är inhämtad från OECD:s databas och mäter procentuell förändring i KPI jämfört med samma period föregående år.

BNP_tillväxt: Denna variabel beskriver förändringen i BNP-tillväxt under mandatperioden. Variabeln är framräknad genom att skillnaden mellan BNP-nivå året innan ett riksdagsval och BNP-nivå året efter föregående riksdagsval. Datan är inhämtad från SCB och avser årlig volymförändring av BNP i procent. BNP avser BNP till marknadspris.

Partiledarpopularitet (Partiled_pop): Denna variabel avser att mäta hur populär respektive partiledare är proportionellt till de andra partiledarna. När den beroende variabeln är definierad som GOV mäts endast populariteten för den partiledare som innehar statsministerposten. Det har inte varit möjligt att hitta data för denna variabel från samma källa för samtliga år. Vi har heller ingen data före år 1979. Datan för åren 1979–2006 kommer från Valforskningsprogrammet vid Göteborgs universitet och den nationella SOM-undersökningen. I undersökningarna har följande fråga/påstående ställts: “Den här frågan gäller hur mycket man allmänt sett gillar respektive ogillar de olika partiledarna.” Varje partiledare får ett medelpoäng framräknat från svaren som ges på en skala från -5 (ogillar) till +5 (gillar). För åren 1979–2006 har datan insamlats under hösten året innan riksdagsval. För valen 2010–2018 har data från Dagens Nyheter, SVT/SIFO och Ipsos använts. Datan från dessa undersökningar anger andelen av väljarkåren som uppgett att de har stort eller mycket stort förtroende för respektive partiledare. Mätpunkterna för dessa val är mars 2010, april 2014 och maj 2018. Det faktum att

datan är insamlad vid olika tidpunkter och från olika källor utgör en brist i datamaterialet och bör tas i beaktning när modellerna analyseras och den prediktiva förmågan i variabeln undersöks.

Näst_bästa_parti (Nb_parti): Denna variabel mäter vilket parti väljarna anser vara näst bäst. Datan är inhämtad från en rapport av Valforskningsprogrammet vid Göteborgs universitet som vid varje valår genomför en valundersökning av svenska väljare. Det framgår inte exakt när datan är insamlad men rapporten nämner att insamlingsmetoden utgörs av en rullande panel där hälften av paneldeltagarna i stickprovet intervjuats i samband med föregående val och hälften vid nästkommande val (Oskarsson & Holmberg 2018: 1). Det är således inte möjligt att säkerställa exakt hur långt innan valdagen datan är insamlad vilket utgör en klar brist i datamaterialet. Vi anser dock ändå att det från en explorativ synvinkel är av intresse att undersöka i vilken utsträckning denna variabel har någon prediktiv styrka. Om variabeln har en sådan effekt bör det vara möjligt för framtida forskning att få fram data för en specifik tidpunkt.

Mediaomnämnannde (Media): Denna variabel mäter andelen gånger respektive partinamn omnämnts i svenska nyhetstidningar proportionellt till de andra riksdagspartierna året innan riksdagsval. Variabeln är framräknad genom att ta totala antalet gånger respektive partinamn omnämnts under hela året före ett valår delat med totala antalet gånger samtliga partier har omnämnts under samma tidsperiod. Även ändelser på partinamnen ingår, exempelvis: Kristdemokraterna, Kristdemokraternas och kristdemokratiska. De tidningar som datan har insamlats ifrån är: Dagens Nyheter, Expressen, Aftonbladet och Svenska Dagbladet.

4. Metod

För att besvara de frågeställningar som uppsatsen avser att undersöka kommer vi att använda bayesiansk multipel linjär regression. Innan denna metod beskrivs närmare kommer vi först ge en kort introduktion till bayesiansk statistik.

Bayesiansk statistik och inferens

Den statistiska disciplinen brukar ofta delas in i två olika inriktningar: frekventistisk och bayesiansk. Inom frekventistisk (även kallad "klassisk") statistik ses parametrarna i en modell som bestämda konstanter som går att skatta genom att dra stickprov från en population. Inom den bayesianska statistiken ses parametrarna (θ) istället som slumpvariabler med tillhörande sannolikhetsfördelningar. Centralt för den bayesianska inriktningen är att man använder data för att uppdatera kunskapen om parametrarna samt att inferens gällande parametrarna uttrycks i termer av sannolikheter. Bayesianska modeller utgår från Bayes teorem, vilket kan sägas utgöra grundbulten i den bayesianska statistiken och kan skrivas enligt följande:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)},$$

där θ är en uppsättning av parametrar. Den första termen kallas för a-posteriori-fördelningen och är sannolikhetsfördelningen för θ givet datan. Den andra termen, $p(\theta|y)$, är likelihooden och beskriver fördelningen av möjliga utfall i stickprovet givet θ . Den tredje termen, $p(\theta)$, är a-priori-fördelningen och reflekterar vår uppfattning om θ innan någon data har observerats. Slutligen har vi även $p(y)$, vilket är en normaliseringskonstant som ser till att a-posteriori-fördelningen alltid summerar till 1. Denna term beror dock inte på θ och kan därför betraktas som en konstant (k) som kan elimineras. Av detta följer att a-posteriori-fördelningen kan skrivas som proportionell till likelihooden multiplicerad med a-priori-fördelningen enligt följande:

$$p(\theta|y) = kp(y|\theta)p(\theta) \propto p(y|\theta)p(\theta).$$

Bayesiansk multipel linjär regression

Multipel linjär regression är en av de vanligaste metoderna för att studera linjära samband mellan en utfallsvariabel (Y) och två eller flera förklarande variabler (X). Metoden är också användbar om vi vill säga någonting om framtiden, det vill säga att göra prediktioner baserat på ny data. Som tidigare nämnts har tidigare forskning nästan uteslutande använt sig av "klassisk" linjär regression (OLS) för att förutspå valresultat med hjälp av strukturella modeller. Vi kommer därför i denna uppsats också att använda oss av linjär regression men utifrån ett bayesianskt perspektiv. Vid bayesiansk regression estimeras en modell men till skillnad från "klassisk" regression så får vi fram en fördelning över de värden som parametrarna i modellen kan anta. Detta gör det möjligt att beskriva utfallsvariabeln (och parametrarna i modellen) som en fördelning av möjliga värden istället för enbart punktskattningar. Regressionerna i denna uppsats genomförs i programvaran R genom att använda rstanarm-paketet. Vi kan genom att använda vektormatrisnotation skriva ner en linjär regressionsmodell enligt följande:

$$Y = X\beta + \varepsilon; \quad \varepsilon \sim N(0, \sigma^2 I),$$

där Y är en vektor för alla möjliga värden för n observationer. X är en $n \times k$ matris för de oberoende variablerna, β är en vektor med de skattade regressionskoefficienterna. Vi antar även att feltermerna är normalfördelade med medelvärde 0 och konstant varians. Av detta följer att utfallsvariabeln $y = (y_1, \dots, y_n)$ har följande fördelning (likelihood):

$$y \sim N_n(X\beta, \sigma^2 I_n).$$

För att estimeras en bayesiansk modell behöver vi även tillskriva en a-priori-fördelning till alla okända parametrar i modellen (β_k och σ^2). A-priori-fördelningen återspeglar vår uppfattning om parametrarna före dess att någon data har observerats. Det är vanligt att a-priori-fördelningen bestäms med hjälp av information från tidigare studier eller annan expertkunskap (Carlin och Louis, 2000: 22). Eftersom vi i denna uppsats inte har någon förutbestämd kunskap om parametrarna i modellen kommer vi inte att specificera a-priori-fördelningarna på förhand utan istället använda oss av "defaultvärden" i rstanarm-paketet. Dessa a-priori-fördelningar är svagt informativa med liten effekt på parametrarna och utformade så att de bidrar med en avvägd regularisering samt en mer stabil process vid skattningarna (Gelman et al. 2020: 124). I strikt mening är fördelningarna inte heller bayesianska eftersom de är baserade på datan efter den har

observerats men Gelman et al. (2020) menar att de ändå bör betraktas som approximativt bayesianska. Det slutgiltiga steget i att estimerar en bayesiansk modell är sedan att identifiera parametrarnas a-posteriori-fördelning med hjälp av Bayes teorem. Det kan dock vara svårt att analytiskt identifiera a-posteriori-fördelningarna och det görs därför ofta genom att simulera dragningar med hjälp av Markov Chain Monte Carlo (MCMC). Vi får då fram en simulerad fördelning som vi kan använda för att göra inferens och prediktioner.

Den prediktiva fördelningen för framtida observationer och prediktionsintervall

Centralt för denna uppsats är att göra prediktioner - det vill säga att med hjälp av våra modeller säga något om framtiden. Genom att simulera dragningar från parametrarnas a-posteriori-fördelningar med MCMC (som beskrivits ovan) med nya värden för X kan vi få fram den prediktiva fördelningen för framtida observationer. Denna fördelning kan skrivas enligt följande:

$$\begin{aligned} p(\tilde{y}|y) &= \int p(\tilde{y}|y, \theta)p(\theta|y)d\theta \\ &= \int p(\tilde{y}|\theta)p(\theta|y)d\theta. \end{aligned}$$

Denna prediktiva fördelning används när vi vill göra prediktioner på ny data som inte har använts för att estimerar modellen. I vårt fall motsvarar den nya datan värdena på de förklarande variablerna i modellen för det riksdagsval som vi avser att predicera. Vi får då fram en fördelning över alla möjliga värden givet den nya datan - med andra ord sannolikhetsfördelningen för prediktionerna. Fördelningens medelvärde kan liknas vid en punktskattning och kan sägas representera vår "bästa" prediktion av valresultatet för det specifika riksdagsvalet. Vi kommer fortsättningsvis beteckna detta medelvärde som \tilde{y} i uppsatsen.

Vi kan utifrån den prediktiva fördelningen även beräkna prediktionsintervall för de prediktioner som görs. Ett prediktionsintervall anger med $100(1 - \alpha)\%$ sannolikhet att en prediktion av en framtida observation innefattas av det angivna intervallet. Vi kommer i denna uppsats att använda oss av 90% prediktionsintervall vilket betyder att vi med 90% sannolikhet kan säga att vår prediktion av ett framtida riksdagsval kommer att befinna sig inom det angivna intervallets gränser.

Tolkning av parameterestimat

Parameterestimat i bayesiansk regression är uttryckt i termer av sannolikheter. Vi får alltså inte fram en enskild punktskattning utan en fördelning över möjliga värden. Vi kan då ange hur sannolikt det är att en parameter antar ett visst värde och därmed även hur osäker skattningen är.

Om inget annat specificeras består vanligtvis en linjär regressionsmodell av ett intercept (β_0). Interceptet kan definieras som väntevärdet för utfallsvariabeln när alla förklarande variabler har värde 0. Vi kommer i denna uppsats (förutom i modell 1 och 2a) inte att estimerar ett intercept utan istället använda en offset som bestämts på förhand. Mer specifikt kommer vi att använda variabeln “Valresultat i föregående riksdagsval” som offset. Detta medför att när vi försöker predicera utfallsvariabeln (resultat i riksdagsvalet) gör vi detta givet resultatet i det föregående riksdagsvalet.⁶

Modellutvärdering

Modellerna som presenteras i uppsatsen kommer att utvärderas i två steg. I det första steget kommer vi att estimerar och jämföra de olika modellerna baserat på data för riksdagsvalen 1973–2014. Vi kommer i detta steg helt att utelämna 2018 års riksdagsval. När vi har identifierat den modell som vi anser vara bäst kommer vi i det andra steget att utvärdera modellen genom att predicera 2018 års riksdagsval.

En väsentlig aspekt i denna uppsats, och vid prediktiv modellering generellt, är på vilka grunder en modell bör utvärderas. Med andra ord, vad är det som skiljer en bra modell från en dålig modell? Det är svårt att besvara denna fråga genom en exakt definition och vi kommer i denna uppsats därför att utgå från tre huvudsakliga utvärderingsmått för att göra denna bedömning: Root Mean Square Error (RMSE), 90% prediktionsintervall och Kolmogorov-Smirnov-test (K-S-test). Även om vi använder oss av dessa mått finns det en överhängande risk att våra strukturella modeller överanpassas. Med detta menas att en modell anpassas utifrån det tillgängliga datamaterialet och därmed genererar bra prediktioner på befintlig data men att modellen fungerar sämre när den används på ny data. Risken med överanpassning ökar även när vi har ett litet stickprov. För att undvika detta kommer vi att använda oss av “Leave-One-Out Cross-Validation” (LOOCV) när vi estimerar och jämför de olika modellerna för

⁶ För ett utförligare resonemang kring detta, se s. 11–12.

riksdagsvalen 1973–2014. Vi kommer även att testa hur modellen fungerar på ny data (riksdagsvalet 2018). LOOCV är en metod som är vanlig inom prediktiv modellering och bygger på idén att man delar upp datan i ett träningsset och ett valideringsset (bestående av en enda observation), där det förstnämnda används för att estimeras modellen och det senare används för att utvärdera den. Vi estimerar således modellen på $n - 1$ observationer i träningssetet och vi kan sedan utvärdera modellen genom att predicera den observation som finns i valideringssetet (\tilde{y}_i). Ett alternativ till LOOCV som också är vanligt förekommande är “Leave-Future-Out Cross-Validation” (LFOCV). Istället för att estimeras modeller på $n - 1$ observationer i träningssetet använder man då endast tidigare datapunkter än den som finns i valideringssetet. Man utelämnar alltså alla framtida observationer. Som framgår ovan har vi i denna uppsats valt att använda oss av LOOCV. Eftersom vi i del två av resultatdelen kommer att predicera utgången i riksdagsvalet 2018 och då endast använder tidigare datapunkter kan detta i viss bemärkelse sägas likna LFOCV. Vi kommer nu att mer ingående att beskriva de tre utvärderingsmått som vi kommer använda för att utvärdera de olika modellerna i resultatdelen.

RMSE

Ett vanligt sätt att utvärdera en modell är att undersöka hur pass bra prediktioner modellen ger. Detta kan göras genom att beräkna modellens prediktionsfel: skillnaden mellan det sanna värdet på utfallsvariabeln (y_i) och det av modellen predicerade värdet (\tilde{y}_i). Vi kommer i denna uppsats att utvärdera modellerna genom att använda utvärderingsmättet RMSE som definieras enligt följande:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\tilde{y}_i - y_i)^2}{n}},$$

där n är antalet prediktioner och \tilde{y}_i representerar det predicerade valresultatet i varje valideringsset. Vi kommer för varje modell som estimeras i resultatavsnittets första del med LOOCV att beräkna en genomsnittlig RMSE för samtliga valideringsset och parti, vilket kan sammanfattas på följande sätt:

$$CV_{n,k} = \frac{1}{n} \sum_{i=1}^n RMSE_{i,k},$$

där n är antalet prediktioner och k representerar respektive parti (inklusive GOV). $CV_{n,k}$ representerar alltså en genomsnittlig RMSE och utgör en skattning av det prediktionsfel vi förväntar oss av ett visst parti om en specifik modell skulle användas på ny data. Vi vill i denna uppsats dock huvudsakligen jämföra modeller och vi kommer därför främst att utgå från den totala genomsnittliga RMSE:n för en hel modell när vi jämför och utvärderar modellerna, vilket kan skrivas på följande sätt:

$$\overline{CV}_{n,k} = \frac{1}{k} \sum CV_{n,k}.$$

Ett lågt RMSE-värde är att föredra jämfört med ett högt RMSE-värde, eftersom det betyder att modellens prediktioner då är närmare det faktiska valresultatet. En bra modell karaktäriseras således av ett så lågt RMSE-värde som möjligt.

Prediktionsintervall

Ett annat sätt att utvärdera modellen är att undersöka om det faktiska valresultatet innefattas av ett prediktionsintervall med en specificerad konfidensgrad. Som beskrivits innan så anger ett prediktionsintervall med $100(1 - \alpha)\%$ sannolikhet att en prediktion av en framtida observation innefattas av det angivna intervallet. Vi kommer i denna uppsats att använda 90% prediktionsintervall. För en modell som predicerar riksdagsvalet väl kommer prediktionsintervallet således innefatta det faktiska riksdagsvalet. Vidare är ett snävare intervall att föredra framför ett bredare eftersom det visar på mindre osäkerhet i prediktionerna. En modell som är dåligt anpassad och därmed innehåller mycket osäkerhet kommer generera breda prediktionsintervall eller helt enkelt inte innefatta det sanna valresultatet alls. Ett prediktionsintervall som exempelvis ger oss en sannolikhet på 90% att Liberalerna kommer att få mellan 0–20% av rösterna vid nästa riksdagsval är inte särskilt svårt att gissa för gemene man.

Kolmogorov-Smirnov Goodness-of-Fit Test

Ytterligare ett sätt att utvärdera om vi har en bra modell är att undersöka om den prediktiva fördelningen kan anses vara korrekt. Vi kan göra detta genom att först beräkna den prediktiva fördelningen (sannolikhetsfördelningen) för en framtida observation. Vi kan sedan utifrån den prediktiva fördelningen undersöka i vilken percentil det sanna valresultat (y_i) hamnar och sedan repetera denna procedur för varje valideringsset med LOOCV så att vi totalt får 13 (eller 7

beroende på antal datapunkter) percentiler. Om den prediktiva fördelningen är korrekt bör percentilerna fördela sig enligt en likformig fördelning och därmed ha samma sannolikhet att inträffa. Om percentilerna inte är likformigt fördelade finns antydning om att modellen inte är korrekt.

Vi kan även undersöka hur väl kalibrerade prediktionsintervallen är genom att titta på percentilerna. Om sannolikheten hos modellen är korrekt bör 90% av valresultaten hamna mellan den 5:e och 95:e percentilen. Om percentilerna inte är likformigt fördelade utan istället får en fördelning som liknar en u-kurva betyder det att vi har många valresultat som hamnat i väldigt höga och låga percentiler och vi har då underskattat osäkerheten i valresultatet. Vi kommer då att få för snäva prediktionsintervall då vår prediktiva fördelning förmodligen är för tajt. Om vi istället har för breda prediktionsintervall, och alltså överskattar osäkerheten i valresultatet, kommer fördelningen istället att likna ett uppochneru. Vi har då istället en för försiktig prediktiv fördelning som är bredare än vad den bör vara.

Vi kommer att undersöka om den prediktiva fördelningen kan anses korrekt med hjälp av Kolmogorov-Smirnov-test (K-S-test). Detta är ett test som ofta används för att undersöka om empirisk data (percentilerna) överensstämmer med en känd teoretisk fördelning (likformig). Testet bygger på att den empiriska kumulativa fördelningen av den empiriska datan jämförs med den kumulativa fördelningsfunktionen av en teoretisk sannolikhetsfördelning (Laguna & Marklund 2018). Vi kommer i denna uppsats att testa följande hypoteser:

H_0 : Data överensstämmer med en likformig fördelning

H_1 : Data överensstämmer inte med en likformig fördelning

Teststatistikan definieras enligt följande:

$$D = \max_{1 \leq i \leq N} \left(F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i) \right),$$

där F är den teoretiska kumulativa fördelningen av den fördelning som testas (likformig). Vi förkastar nollhypotesen om $D > D_{krit.} = D_\alpha$ eller om p-värdet understiger signifikansnivån

(α).⁷ Vi kommer i denna uppsats att använda oss av $\alpha = 0,1$. En fördel med K-S-testet är att det inte kräver ett stort stickprov för att fungera väl (Ibid.).

Det är även möjligt att visuellt undersöka om den empiriska datan överensstämmer med en specificerad teoretisk fördelning genom att använda Q-Q plots. Om så är fallet ska observationerna (percentilerna) ligga längs en rak diagonal linje. Det kan dock vara svårt att avgöra detta visuellt när man har ett litet stickprov och ett hypotestest är därför enklare att tolka. Q-Q plots för alla partier och modeller (1–8) finns dock att hitta i appendix A.

Riksdagsvalet 2018 kommer att utvärderas genom att undersöka vilken total genomsnittlig RMSE ($total \overline{RMSE}$) modellen har, om prediktionsintervallen innefattar det sanna valresultatet samt storleken på intervallen.

Avslutningsvis vill vi också nämna något om antaganden vid prediktiv modellering. Det finns ett antal antaganden som vid linjär regression vanligtvis brukar kontrolleras och utvärderas innan man utför inferens eller gör prediktioner. Vi gör även i denna uppsats ett antal antaganden men vi kommer inte explicit att undersöka huruvida dessa kan anses vara uppfyllda eller inte.⁸ Detta kan motiveras utifrån att vi huvudsakligen är intresserade att undersöka hur väl vår modell fungerar på ny data. Genom att använda LOOCV och ett "testset" (riksdagsvalet 2018) kommer vi fånga upp de eventuella fel som finns och hur väl modellen går att generalisera. Det är således inte nödvändigt att utvärdera antagandena eftersom vi ändå kommer att fånga upp de fel som finns och därmed implicit testat antagandena.

En alternativ metod: frekventistisk linjär regression (OLS)

Som nämnts innan har tidigare forskning vanligtvis använt "klassisk" eller frekventistisk multipel linjär regression (OLS) för att estimerat strukturella modeller. Detta hade även varit en möjlig metod för denna studie. Faktum är att resultaten från de bayesianska regressionsmodeller vi använder i denna uppsats är mycket lika de resultat vi skulle fått från klassisk regression eftersom vi inte har specificerat några informativa a-priori-fördelningar och använder en likelihood från normalfördelningen (gaussian). En skillnad mellan metoderna, och fördelen

⁷ Programvaran R kan inte utföra K-S-testet om det förekommer "ties". När två värden är identiska, till exempel om det förekommer två värden som antar värdet 1 så har vi därför ändrat det ena värdet till 0,999.

⁸ De antaganden som görs är följande: Modellen är linjär (i parametrarna), feltermerna och variablerna är inte korrelerade ($Cov(X_i, u_i) = 0$), väntevärdet av feltermerna är 0 ($E(u_i|X_i) = 0$), homoskedasticitet (feltermerna har konstant varians) ($Var(u_i|X_i) = \sigma^2 I$) och ingen autokorrelation i feltermerna ($Cov(u_i, u_j|X_i, X_j) = 0$).

med att använda bayesiansk regression, är dock att det bayesianska tillvägagångssätt tar mer osäkerhet i beaktning samt gör det möjligt att uttrycka osäkerheten i termer av sannolikheter. De bayesianska prediktionsintervallen kommer därför förmodligen vara bredare jämfört med "klassiska" men förhoppningsvis också mer korrekta eftersom de tar större hänsyn till den osäkerhet som finns. Eftersom vi i denna uppsats är intresserade av att göra så bra prediktionsintervall som möjligt anser vi att den bayesianska metoden är att föredra.

5. Resultat

Vi kommer i detta avsnitt att presentera uppsatsens resultat, uppdelat i två delar. Som tidigare nämnts kommer vi i den första delen att presentera och utvärdera ett antal olika modeller som avser att predicera resultatet i svenska riksdagsval under åren 1973–2014. Samtliga av dessa modeller estimeras genom att använda LOOCV. Resultaten för alla modeller i den första delen finns sammanfattade i tabell 1 (se sida 29 nedan). När vi i denna del hänvisar till RMSE för specifika partier avser detta $CV_{n,k}$ och när vi hänvisar till en total RMSE för en hel modell avser detta $\overline{CV}_{n,k}$ (genomsnittet för alla partier i en specifik modell). I den andra delen kommer vi att testa den modell som vi finner bäst genom att undersöka hur väl modellen predicerar riksdagsvalet 2018 (ny data).

Tabell 1: Utvärdering av resultat för modell 1–8

		Modell 1	Modell 2a	Modell 2b	Modell 3	Modell 4	Modell 5	Modell 6 ⁹	Modell 7	Modell 8
Partier	Antal obs. ¹⁰	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
V	13	1.91	2.13	1.71	1.96	1.91	2.07	1.90	2.17	2.76
S	13	4.56	2.52	2.49	4.60	4.56	2.76	2.41	2.54	5.71
MP	7	1.59	1.30	1.02	3.86	1.66	1.19	1.84	1.27	1.32
C	13	3.12	2.18	2.13	2.77	2.12	2.01	1.70	1.76	3.24
L	13	3.56	2.98	3.74	4.31	4.73	4.82	4.28	4.13	5.26
M	13	4.53	3.58	3.35	4.46	5.97	3.84	3.51	3.44	6.19
KD	7	3.44	5.75	2.76	4.25	3.03	5.40	4.20	4.08	4.65
GOV	13	3.71	2.27	2.06	3.35	3.32	2.43	-	2.24	3.39
$\overline{CV}_{n,k}$	-	3.30	2.84	2.41	3.70	3.41	3.07	2.83	2.70	4.07
Tot. ant. obs utanför 90% PI	-	11	12	12	14	12	10	10	13	12
Andel (%) obs. utanför 90% PI	-	12	13	13	15	13	13	13	14	13
Tot. antal p-värden<0.10	-	2	0	1	0	0	0	1	0	0

⁹ Modell 6 har totalt 79 observationer då GOV inte prediceras.

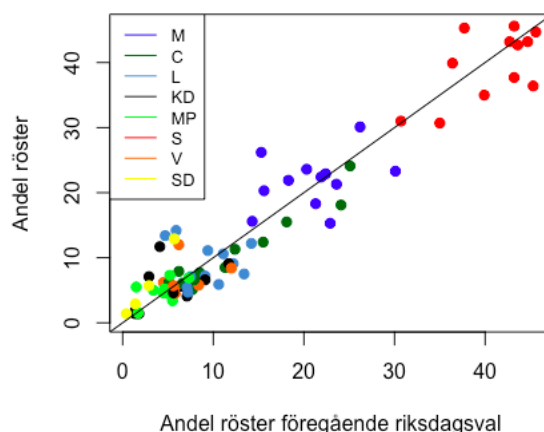
¹⁰ Modell 5 har totalt 79 observationer (för MP finns 6 observationer och för KD finns 7 observationer. För övriga partier finns 11 observationer).

Utvärdering av modell 1–8

Modellerna i denna del kommer att presenteras successivt där modeller byggs på med variabler som identifieras utifrån tidigare forskning. Vi börjar med den absolut enklaste modellen (1) där vi predicerar valresultatet utifrån tidigare års valresultat. Allt annat lika kan vi anta att människor kommer att agera (och således även rösta) på ett liknande sätt som tidigare om de inte utsätts för ny information. Modell 1 specificeras enligt följande:

$$\text{Modell 1: } \text{Valresultat}_k = \beta_0 + \beta_1 \text{Res_tid_val}_k + \varepsilon$$

Vi kan från tabell 1 utläsa att Socialdemokraterna (S) och Moderaterna (M) är de partier med högst RMSE medan Miljöpartiet (MP) är det parti som har lägst RMSE på 1,59. Modell 1 har en total RMSE på 3,30 och totalt 11 observationer, eller 12% av samtliga observationer, som inte innefattas av prediktionsintervallen. Vi förkastar även nollhypotesen för Vänsterpartiet (V) och MP - vilket antyder på 10% signifikansnivå att den empiriska datan inte är likformigt fördelad för dessa två partier och att percentilerna i prediktionsintervallen således inte är korrekta. I figur 1 nedan kan vi även se att det finns ett tydligt linjärt samband mellan tidigare valresultat och nästkommande valresultat. Stöddlinjen i figuren, som har ett intercept på 0 och en lutning på 1, visar att det är rimligt att anta ett intercept på 0, det vill säga att tidigare valresultat predicerar framtida valresultat väl. Figuren kan tolkas som att de partier som ligger över stöddlinjen har ökat i andel röster jämfört med föregående riksdagsval medan de observationer som ligger under istället har fått färre röster jämfört med föregående val.



Figur 1: Resultat i föregående riksdagsval visas på horisontella axeln för respektive parti och resultatet av nästkommande riksdagsval visas på den vertikala axeln (år 1973–2014)

Vi går sedan vidare med att bygga på modell 1 genom att lägga till variabeln "PSU_majdiff" (modell 2a). Mot bakgrund av figur 1 (och tidigare resonemang kring att använda en offset, se s. 11-12) kommer vi även att testa modell 2b där vi istället för ett intercept använder variabeln "Valresultat i föregående riksdagsval" som en offset. Vi lägger även i denna modell till variabeln "PSU_majdiff". Modell 2a och modell 2b specificeras enligt följande:

$$\text{Modell 2a: } \text{Valresultat}_k = \beta_0 + \beta_1 \text{Res_tid_val}_k + \beta_2 \text{PSU_majdiff}_k + \varepsilon$$

$$\text{Modell 2b: } \text{Valresultat}_k = \text{Res_tid_val}_k + \beta_1 \text{PSU_majdiff}_k + \varepsilon$$

Vi kan från tabell 1 utläsa att modell 2b har en lägre total RMSE (2,41) jämfört med modell 2a (2,84). Båda modellerna har 12 observationer utanför prediktionsintervallen vilket är något fler än de 10% (9–10 observationer) som vi hade kunnat förvänta oss. Detta är således lite fler fel än förväntat om prediktionsintervallen skulle vara optimalt kalibrerade. Vi kan genom att använda en binomialfördelning undersöka hur sannolikt det är att få ett utfall på 12 observationer eller fler om vi totalt har 92 observationer och sätter $p=0,1$.¹¹ Vi får då fram att sannolikheten att få 12 observationer eller fler är 21%, vilket inte är jättehögt men inte heller helt osannolikt. Ett 90% prediktionsintervall verkar således fungera någorlunda bra men tycks vara ett gränsfall och det är möjligt att både modell 2a och 2b har för snäva prediktionsintervall. Vi kan vidare konstatera att samtliga p-värden från K-S-testen för modell 2a överstiger signifikansnivån på 10%. För modell 2b förkastar vi nollhypotesen för K-S-testen för S vilket antyder, på 10% signifikansnivå, att den empiriska datan inte är likformigt fördelad och att den prediktiva fördelningen inte är väl kalibrerad. Det kan dock poängteras att p-värdet för S var 0,09 och därmed väldigt nära signifikansnivån. Efter att ha utvärderat och jämfört de två modellerna kan vi konstatera att det inte är särskilt stora skillnader mellan dem. Utifrån ovanstående resultat och tidigare resonemanget om att använda en offset, samt att det blir färre parametrar att skatta i modellen, anser vi att det är motiverat att fortsätta använda en offset i uppsatsens övriga modeller.

Vi kan studera modell 2b ytterligare genom att undersöka vilka värden parameterestimaterna ($\hat{\beta}_1$) tillhörande "PSU_majdiff" antar för respektive parti i modell 2b. Ett $\hat{\beta}_1$ -värde runt 1 antyder att "PSU_majdiff" har en större prediktiv styrka jämfört med modellens offset ("Valresultat i föregående riksdagsval"). Ett $\hat{\beta}_1$ -värde närmare 0 skulle istället betyda att modellens offset har

¹¹ p är sannolikheten att en observation hamnar utanför prediktionsintervallet.

en högre prediktiv styrka jämfört med variabeln ”PSU_majdiff”. Samtliga $\hat{\beta}_1$ -värde för modell 2b och parti finns sammanfattade i tabell 2 nedan. Utifrån tabell 2 kan vi utläsa att de flesta partier har ett $\hat{\beta}_1$ -värde mellan 0,7–1,3. För S ligger dock nästan alla $\hat{\beta}_1$ -värden på 0,6, förutom valet 1998 då det ligger på 0,5. För MP är $\hat{\beta}_1$ -värdena ännu lägre på antingen 0,3 eller 0,4. Detta antyder att vår offset istället bidrar med mer information jämfört med ”PSU_majdiff” när vi predicerar valresultatet för S och MP. Vid riksdagsvalet 1998 visade ”PSU_majdiff” att S skulle tappa 5,4 procentenheter jämfört med riksdagsvalet 1994 men det riktiga valresultatet visade att S tappade hela 8,9 procentenheter. Det var också riksdagsvalet 1998 som hamnade utanför prediktionsintervallet för S (i percentil 0,003), vilket indikerar att vi har underskattat osäkerheten i valresultatet och att prediktionsintervallet därmed är för snävt.

Tabell 2: $\hat{\beta}_1$ -värde för modell 2b (samtliga partier)

	V ($\hat{\beta}_1$)	S ($\hat{\beta}_1$)	MP ($\hat{\beta}_1$)	C ($\hat{\beta}_1$)	L ($\hat{\beta}_1$)	M ($\hat{\beta}_1$)	KD ($\hat{\beta}_1$)	GOV($\hat{\beta}_1$)
1973	1.2	0.6	-	0.8	1.1	0.8	-	0.7
1976	1.2	0.6	-	0.8	1.3	0.8	-	0.7
1979	1.2	0.6	-	0.7	1.4	0.8	-	0.7
1982	1.3	0.6	-	0.9	1.3	0.8	-	0.7
1985	1.3	0.6	-	0.8	1.2	0.9	-	0.7
1988	1.3	0.6	-	0.9	1.3	0.8	-	0.7
1991	1.2	0.6	0.3	0.8	1.2	0.9	1	0.7
1994	1.3	0.6	0.4	0.8	1.3	0.8	0.9	0.6
1998	0.8	0.5	0.4	0.8	1.2	0.9	0.9	0.7
2002	1.1	0.6	0.4	0.8	1.2	0.8	0.9	0.7
2006	1.4	0.6	0.4	0.8	1.1	0.7	1	0.6
2010	1.2	0.6	0.4	0.8	1.3	0.8	1	0.7
2014	1.5	0.6	0.3	0.8	1.3	0.8	1	0.6

Vi går sedan vidare med att testa ytterligare två modeller, modell 3 och modell 4, som specificeras enligt följande:

$$\text{Modell 3: } Valresultat_k = Res_tid_val_k + \beta_1 PSU_majdiff_t + \beta_2 Inflation_t + \beta_3 Arbetslöshet_t + \beta_4 Inc_k + \beta_5 Arbetslöshet_t * Inc_k + \varepsilon$$

$$\text{Modell 4: } Valresultat_k = Res_tid_val_k + \beta_1 PSU_majdiff_t + \beta_2 BNP_tillväxt_t + \beta_3 Arbetslöshet_t + \beta_4 Inc_k + \beta_5 Arbetslöshet_t * Inc_k + \varepsilon$$

Samtliga av dessa nya variabler är teoretiskt motiverade utifrån tidigare forskning. Vi kan utifrån tabell 1 utläsa att både modell 3 och 4 har en högre total RMSE jämfört med modell 2b och att modell 3 även har fler antal observationer utanför prediktionsintervallen. P-värdena från samtliga K-S-test (både för modell 3 och 4) överstiger 10% vilket antyder att de prediktiva fördelningarna är väl kalibrerade. Utifrån utvärderingsmåttan kan vi sammantaget konstatera att variablerna “arbetslöshet”, “BNP-tillväxt” och “inflation” inte tydligt förbättrar den prediktiva kraften i modellerna trots att dessa variabler har visat sig effektiva i tidigare forskning (för en diskussion om detta se avsnitt 6).

Vid en jämförelse av samtliga modeller (1–4) där vi även tar i beaktning att modell 2b är en relativt simpel modell, vilket är att föredra, gör vi bedömningen att modell 2b är den “bästa” modellen av de fem vi testat hitintills. Utifrån resultaten presenterade i tabell 1 kan vi dock konstatera att det finns utrymme att förbättra modell 2b. För att göra detta kommer vi därför att titta närmare på de riksdagsval som modell 2b inte lyckades predicera särskilt väl - de val som hamnade utanför prediktionsintervallen. Dessa riksdagsval finns sammanfattade i tabell 3 nedan. Förhoppningen är att vi på så vis kan urskilja anledningen till varför vissa riksdagsval predicerades sämre och därmed identifiera vilka förklarande faktorer som kan tänkas förbättra modellen.

Tabell 3: Riksdagsval som inte innefattas av prediktionsintervallen

	V	S	MP	C	L	M	KD	GOV
Riks- dagsval	1998 2014	1998	1994	1979	1985 2002	1985 2002	1998	2006 2010

Från tabell 3 kan vi utläsa att det finns likheter mellan vissa av partierna. Till exempel så prediceras riksdagsvalet 1998 dåligt för både V och S samt Kristdemokraterna (KD). Både Liberalerna (L) och M prediceras dåligt 1985 och 2002. År 2002 ökade L med 8,7 procentenheter samtidigt som M minskade med 7,6 procentenheter. Man kan tänka sig att dessa partier ligger nära varandra ideologiskt och därför har väljare som relativt lätt rör sig mellan partierna - om det ena partiet ökar i andel röster minskar stödet för det andra som en konsekvens. Detta tyder på att näst bästa parti kan vara en potentiell förklarande variabel för dessa val. Vid valet 1985 ökade dock L mer än vad M minskade. Alla partier backade faktiskt i detta riksdagsval förutom L som då hade Bengt Westerberg som partiledare. Detta val har kommit att kallas "Westerbergeffekten" på grund av den uppskattade positiva effekten som Westerberg hade på partiets valresultat (Oscarsson & Holmberg 2008: 234). När det gäller KD kan vi tänka oss att partiledaren hade en liknande effekt på valresultatet i riksdagsvalet 1998. Detta val predicerades dåligt av modell 2b samtidigt som Alf Svenssons popularitet var väldigt hög. I valet 1998 ökade KD till 11,7% från 4,1% i valet 1994, vilket antyder att Svenssons popularitet hade en stor effekt på partiets valresultat. En annan potentiell, och mindre utforskad, förklarande variabel är medias effekt på resultaten i riksdagsvalen. Som nämnts innan skriver tidningar mycket om de olika partierna och den politiska utvecklingen i landet, vilket kan tänkas påverka hur människor röstar. Studier i andra länder har hittat att politikens och partiers omnämnande i media korrelerar med en ökning av andel röster på partiet (se sida 14).

Utifrån ovanstående iakttagelser väljer vi att i modell 5 lägga till variabeln "partiledarpopularitet", i modell 6 lägga till variabeln "näst bästa parti" och i modell 7 lägga till variabeln "mediaomnämnannde". Dessa modeller specificeras enligt följande:

$$\text{Modell 5: } Valresultat_k = Res_tid_val_k + \beta_1 PSU_majdiff_k + \beta_2 Partiled_pop_k + \varepsilon$$

$$\text{Modell 6: } Valresultat_k = Res_tid_val_k + \beta_1 PSU_majdiff_k + \beta_2 Nb_parti_k + \varepsilon$$

$$\text{Modell 7: } Valresultat_k = Res_tid_val_k + \beta_1 PSU_majdiff_k + \beta_2 Media_k + \varepsilon$$

Vi kan utifrån tabell 1 utläsa att modell 5 och modell 6 har färre observationer utanför prediktionsintervallen jämfört med modell 2b. Modell 7 har däremot fått ytterligare en observation utanför prediktionsintervallen. Vi predicerar dock inte "GOV" för modell 6 då variabeln "näst bästa parti" inte kan anses teoretiskt motiverad när utfallsvariabeln är definierad som andel röster på den sittande regeringen, vilket kan förklara varför antalet observationer utanför prediktionsintervallen har blivit färre. Om vi därför istället tittar på total andel

observationer utanför prediktionsintervallen ser vi att det endast är modell 5 som har en lägre andel än modell 2b. För modell 5 hamnar riksdagsvalet 1985 för M inte längre utanför prediktionsintervallet (gäller även modell 7) samt 1994 för MP och 2006 för GOV. Däremot hamnar istället riksdagsvalet 2010 för MP och 1991 för KD utanför intervallen för modell 5. För modell 6 prediceras inte riksdagsvalet 1994 för MP längre dåligt men däremot så har riksdagsvalet 1991 för KD tillkommit (hamnat utanför utanför), vilket också gäller modell 7. Sammantaget kan vi konstatera att ingen av modellerna tycks ha förbättrat modell 2b - några riksdagsval som tidigare predicerades dåligt har försvunnit men vissa har också tillkommit och den totala RMSE:n har ökat. Vi anser att dessa resultat pekar på att varken “partiledarpopularitet”, “näst bästa parti” eller “mediaomnämmande” tydligt har lyckats förbättra modell 2b. Mot bakgrund av detta anser vi därför att modell 2b fortfarande är att föredra framför samtliga modeller vi hitintills presenterat.

En potentiell kritik mot modell 2b är att den förklarande variabeln utgörs av en opinionsundersökning (PSU:n). Förvisso brukar någon form av opinionsundersökning ingå i strukturella modeller men då kombinerat med andra förklarande faktorer såsom exempelvis BNP-tillväxt eller arbetslöshetsnivå (se till exempel Lewis-Beck & Dassonneville 2015a; Norpoth & Gschwend). Endast en opinionsundersökning hjälper oss inte heller att förstå de underliggande faktorerna till varför väljare röstar som de gör. Dessutom använder Sundell och Lewis-Beck (2014) inte någon opinionsundersökning i deras studie. Eftersom deras studie är den enda vi hittat som använder en strukturell modell för att förutspå riksdagsvalen i Sverige - och vi vill kunna jämföra våra resultat - så kommer vi även att estimerar en modell där vi exkluderar PSU:n och endast använder andra strukturella faktorer. Vi väljer dock fortfarande att behålla en offset baserad på tidigare års valresultat eftersom vi anser att det är ett bättre sätt att tillhandahålla information jämfört med ett ospecificerat intercept. Vi använder samma process som tidigare för att jämföra olika modeller - det vill säga vi använder LOOCV för att estimerar modellerna och utvärderar dem sedan utifrån de tre utvärderingsmått. Vi testat ett antal modeller med olika förklarande variabler och utifrån utvärderingsmått finner vi att följande modell är den bästa strukturella modellen:

$$\text{Modell 8: } \text{Valresultat}_k = \text{Res_tid_val}_k + \beta_1 \text{Arbetslöshet}_k + \beta_2 \text{Media}_k + \varepsilon$$

Variabeln “arbetslöshet” ingår ofta i strukturella modeller och räknas till en av “de stora två”. Den ingick även i den modell som konstruerades av Sundell och Lewis-Beck, även om den i

deras studie var definierad på ett något annorlunda sätt. Variabeln "mediaomnämmande" är kanske ett desto mer förvånande resultat även om det förvisso finns visst stöd i tidigare forskning för denna variabel (Oegema & Kleinnijenhuis 2000; Hopmann et al. 2010). Om vi jämför resultaten med modell 2b kan vi se att modell 8 har en högre total RMSE på 4,07. Till skillnad från modell 2b så indikerar resultaten från K-S-testen för modell 8 att alla prediktiva fördelningar är korrekta (samtliga p-värden överstiger 0,1). Vi kommer nu gå vidare med att testa hur väl modell 2b och modell 8 lyckas predicera resultatet i ett riksdagsval utifrån ny data - 2018 års riksdagsval.

Utvärdering av riksdagsvalet 2018

Ett slutgiltigt test för modell 2b och modell 8 blir alltså hur väl de lyckas predicera utfallet i riksdagsvalet 2018. Modell 8 kommer även att jämföras med Sundells och Lewis-Becks modell som framöver kallas modell Sundell och Lewis-Beck (2014).¹² Resultaten för samtliga modeller finns sammanfattade i tabell 4 (se s. 38). Om vi börjar titta på resultaten i tabell 4 kan vi konstatera att modell 2b överlag ger starka prediktioner med en total RMSE på 1,2 och att samtliga prediktionsintervall innefattar det sanna valresultatet för riksdagsvalet 2018. KD och M är de två partier som prediceras sämst med en RMSE på 3,2 respektive 2,6 medan Centern (C) är det parti med lägst RMSE på 0,3. Allra lägst RMSE (0,2) får vi dock när vi predicerar valresultatet för hela den sittande regeringen (V+S+MP). Prediktionsintervallen varierar i storlek från 5,5 procentenheter för V till 11,8 procentenheter för M.

Vid en jämförelse av modell 8 och modell Sundell och Lewis-Beck (2014) kan vi konstatera att modell 8 har en lägre total RMSE på 2,0 jämfört med 3,2 för modell Sundell och Lewis-Beck (2014). Vi kan också se att modell 8 tycks ha en mer jämn fördelning av prediktionsfelen bland samtliga partier medan modell Sundell och Lewis-Beck (2014) har vissa mycket låga prediktionsfel men även några mycket stora (se till exempel S och GOV). Vi kan vidare konstatera att samtliga prediktionsintervall för modell 8 innefattar de sanna riksdagsvalsresultaten medan modell Sundell och Lewis-Beck (2014) missar valresultatet när hela den sittande regeringen (GOV) prediceras. Storleken på intervallen kan anses stora för båda modellerna, men där modell 8 har snäppet lägre total intervallstorlek än modell Sundell

¹² Vi har utifrån Sundells och Lewis-Becks studie själva tagit fram data och variabler. Vi har efter bästa förmåga försökt att replikera variablerna så att de överensstämmer med den ursprungliga studien. Vi har sedan estimerat modellen och predicerat valresultaten för riksdagsvalet 2018. Eftersom vi inte har tillgång till originaldata kan vi dock inte garantera att den data vi använt är densamma som Sundell och Lewis-Beck använde i deras studie.

och Lewis-Beck (2014). Utifrån ovanstående resultat anser vi att modell 8 är att föredra framför modell Sundell och Lewis-Beck (2014) men att ingen av modellerna är särskilt bra.

Sammantaget kan vi konstatera att modell 2b har en lägre total RMSE och mer väl kalibrerade prediktionsintervall jämfört med modell 8 och modell Sundell och Lewis-Beck (2014). Modell 2b är således den bästa modellen även när vi predicerar riksdagsvalet 2018.

Tabell 4: Utvärdering av resultat för modell 2b, 8 och modell Sundell och Lewis-Beck (2014) för riksdagsvalet 2018

Partier	Modell 2b				Modell 8			Modell Sundell och Lewis-Beck (2014)		
	Valresultat (%)	Prediktion (%)	RMSE	90% PI	Prediktion (%)	RMSE	90% PI	Prediktion (%)	RMSE	90% PI
V	8.0	7.8	0.2	[5.1, 10.6]	5.0	3.0	[0.5, 9.1]	7.8	0.2	[4.2, 11.3]
S	28.3	29.1	0.8	[24.5, 33.7]	29.6	1.3	[21.6, 37.6]	36.0	7.7	[27.2, 44.9]
MP	4.4	5.9	1.5	[3.8, 8.1]	6.5	2.1	[3.9, 9.2]	5.2	0.8	[2.4, 8.0]
C	8.6	8.3	0.3	[5.4, 11.2]	5.2	3.4	[0, 11.4]	7.9	0.7	[0, 18.5]
L	5.5	4.8	0.7	[0, 11.3]	6.3	0.8	[0, 14.3]	10.1	4.6	[4.6, 15.6]
M	19.8	22.4	2.6	[16.4, 28.2]	22.1	2.2	[12.4, 31.4]	19.8	0	[11.7, 27.7]
KD	6.3	3.1	3.2	[0, 8.9]	5.3	1.0	[0, 13.4]	9.0	2.7	[3.5, 14.7]
GOV	40.7	40.9	0.2	[37.2, 44.6]	42.6	1.9	[36.3, 48.9]	49.3	8.6	[42.1, 56.7]
Total \overline{RMSE}			1.2			2.0			3.2	

6. Diskussion och slutsatser

Vi har i denna uppsats undersökt vilka strukturella faktorer som kan användas för att bäst predicera valresultat i svenska riksdagsval minst tre månader före valdagen. Uppsatsens syfte motiveras utifrån en empirisk kunskapslucka i forskning kring faktorer som kan tänkas predicera svenska riksdagsval. Vi har därmed identifierat och testat ett flertal förklarande faktorer i olika modeller. Utifrån resultaten har vi identifierat två slutgiltiga modeller - en som innehåller SCB:s PSU från maj under valåret och en som består av faktorerna arbetslöshet och partiernas omnämnande i media. Båda modellerna har en offset som är definierad utifrån variabeln "valresultat i föregående riksdagsval". Modellerna har en ledtid på 3–4 respektive 4–5 månader (beroende på när data blir tillgänglig) och har jämförts med en strukturell modell från tidigare forskning av Sundell och Lewis-Beck (2014). Våra resultat tyder på att de två slutgiltiga modellerna som presenteras i denna uppsats genererar mer träffsäkra prediktioner och mer väl kalibrerade prediktionsintervall av resultaten i riksdagsvalet 2018 än modellen framtagen av Sundell och Lewis-Beck (2014).

En första slutsats vi kan dra av resultaten i denna uppsats är att om vi endast ser till vilka faktorer som bäst kan predicera valresultat i svenska riksdagsval - den prediktiva förmågan - så är PSU:n från SCB i särklass bäst. Modell 2b, som endast innehöll en offset och variabeln "PSU_majdiff", hade lägst total RMSE och samtliga prediktionsintervall innehöll det sanna valresultatet. Storleken på prediktionsintervallen var också mindre jämfört med de andra modellerna. Denna slutsats skiljer sig från tidigare forskning på området av Sundell och Lewis-Beck (2014) som istället konstaterade att valet att inkludera en opinionsundersökning inte förbättrade deras strukturella modell. En möjlig förklaring till dessa olika slutsatser kan vara att Sundell och Lewis-Beck framförallt utvärderade modellens prediktiva förmåga utifrån det justerade R^2 -värdet. Som nämnts tidigare (s. 10) anser vi att prediktionsmodeller inte främst bör utvärderas utifrån justerat R^2 eftersom det har visat sig relativt enkelt att få höga justerade R^2 -värden. Ett högt justerat R^2 -värde bör således inte betraktas som synonymt med en bra prediktionsmodell.

En andra slutsats som vi kan dra utifrån resultatdelen är att "traditionella" strukturella faktorer såsom arbetslöshet, BNP-tillväxt och inflation inte tycks fungera särskilt väl som förklarande variabler när det kommer till att förutspå resultaten i svenska riksdagsval, eller åtminstone att

dess faktorerna inte bidrar med ny relevant information utöver den som vi kan få från PSU:n. En potentiell anledning till detta resultat är att Sverige är ett flerpartisystem med för närvarande åtta riksdagspartier och det finns därmed många alternativ för väljarna när de går till valurnan. Har man istället endast två stora partier, som i exempelvis USA, kan man tänka sig att en väljare som är missnöjd med den politik som förs eller det samhällsekonomiska läget helt enkelt vänder sig till det andra alternativet. I Sverige är det dock inte lika lätt att förutspå vilket parti en väljare kommer att gå till om hen är missnöjd med den rådande situationen. En intressant skillnad med vår studie är att dessa mer traditionella strukturella faktorer har använts framgångsrikt i andra studier gjorda på flerpartisystem med skillnaden att den förklarande variabeln då har varit definierad som andelen röster på den sittande regeringen eller ett specifikt block. Just skillnaden i hur man definierar den beroende variabeln skulle kunna vara en potentiell anledning till att dessa variabler fungerat bättre i andra länder eftersom endast förutspå stödet för den sittande regeringen efterliknar situationen i ett tvåpartisystem - antingen är man för eller emot den sittande regeringen. Svensk politik har länge haft en uppdelning mellan höger- och vänsterpartier och har sedan början av 2000-talet även haft två tydliga block - de Rödgröna och Alliansen. Efter riksdagsvalet 2018 upplöstes dock denna tydliga blockstruktur när två av allianspartierna (C och L) valde att släppa fram Stefan Löfven som statsminister och förhandla med de rödgröna partierna kring budgeten. Då det inte längre finns några tydliga block inom svensk politik går det att argumentera att det är än mer viktigt att identifiera modeller som med god precision kan predicera andelen röster för varje enskilt parti. Sverigedemokraternas intåg i riksdagen har ytterligare komplicerat framtida regeringsbildningar, vilket också bidrar till vikten av att kunna predicera valresultaten för enskilda partier.

En annan potentiell förklaring till varför strukturella faktorer är relativt sämre på att predicera svenska riksdagsval jämfört med andra länders parlamentsval skulle kunna vara att svenska väljare drivs mer av ideologiska värden än av strukturella faktorer. Denna förklaring motiveras av tidigare studier på hur svenskar placerar sig på en höger-vänsterskala. I en rapport från SCB beskrivs svenskars valarbeteende som "internationellt sett ovanligt endimensionell och strukturerad av den klassiska vänster-högerkonflikten" (Oscarsson & Holmberg 2013: 18). Även om det finns skäl att tro att denna dimension har blivit mindre viktig över tid då antalet rörliga väljare ökar för varje val är det rimligt att anta att ideologi fortsatt spelar en stor roll för svenska väljare. Det finns därmed skäl att vidare utforska hur stor effekt ideologi har i jämförelse med de strukturella faktorer som vi har testat i denna uppsats.

Fokus för denna uppsats har inte endast varit att göra så bra punktskattningar som möjligt utan även undersöka i vilken utsträckning vi kan uppnå väl kalibrerade prediktionsintervall. En slutsats som vi kan dra från uppsatsens resultat är att det normalfördelningsantagande som vi har gjort för modellerna (och därmed också för den prediktiva fördelningen) tycks göra att den prediktiva fördelningen blir för snäv. Resultaten som finns sammanfattade i tabell 1 visar att något för många av valresultaten (över 10%) för modell 2b och modell 8 inte innefattas av prediktionsintervallen. Valresultaten hamnar istället strax utanför prediktionsintervallen, i svansarna av den prediktiva fördelningen. Exempel på sådana "uteliggare" är KD och S 1998 samt L och M 2002. Detta tyder på att normalfördelningskurvan går ner för snabbt och att prediktionsintervallen därför blir för snäva. Det finns därmed anledning att ifrågasätta om användandet av normalfördelningen - som varit normen i tidigare studier av strukturella modeller - är optimalt i länder med flerpartisystem där det finns fler alternativ för missnöjda väljare. Framtida studier skulle kunna undersöka detta vidare genom att använda sig av robust regression där man inte antar en normalfördelning utan istället använder en t-fördelning och skattar frihetsgraderna. Det är möjligt att en t-fördelning som är plattare och därmed har en "längre" svans skulle vara bättre anpassad till datan. Det finns således ett behov att undersöka vilka modeller man bör använda för att göra prediktioner i flerpartisystem.

En övergripande kritik mot strukturella modeller är vidare att de har svårt att predicera valresultat för nya partier. En brist med denna uppsats är att vi inte har predicerat valresultatet för SD. Partiet har suttit i riksdagen sedan 2010 och om vi vill återspegla svensk politik och framtida riksdagsval behöver vi därför en modell som även kan predicera valresultatet för SD. Detta har varit utanför omfånget för denna uppsats men det finns möjlighet för andra studier att utifrån modellernas parameterestimater komma fram till informativa a-priori-fördelningar som eventuellt kan användas för att predicera resultatet för SD i kommande val. På så vis är det möjligt att komma runt problemet med att strukturella modeller inte klarar av att hantera uppkomsten av nya partier. Denna möjlighet kan även framhävas som ytterligare en fördel med att använda bayesiansk regression istället för frekventistisk regression för att predicera framtida riksdagsval.

Källförteckning

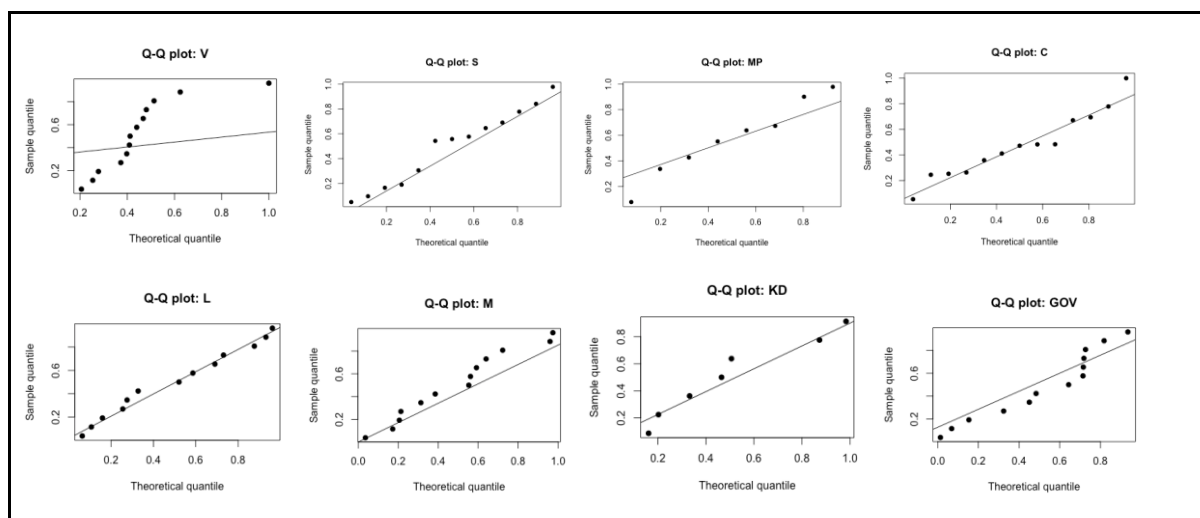
- Abramowitz, A. (2008) Forecasting the 2008 Presidential Election with Time-for-Change Model. *PS: Political Science & Politics*, 41(4), pp. 691-695.
- Abramowitz, A. (1988) An Improved Model for Predicting the Outcomes of Presidential Elections. *PS: Political Science and Politics*, 21, pp. 843-7.
- Anderson, C. J. (2000) Economic voting and political context: a comparative perspective, *Electoral Studies*, 19(2-3), pp. 151-170.
- Arnesen, S. (2012) Forecasting Norwegian elections: Out of work and out of office, *International Journal of Forecasting*, 28(4), pp. 789-796.
- Carlin, B. P., & Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Boca Raton: Chapman & Hall/CRC.
- Dassonneville, R. & Hooghe, M. (2010) Election forecasting under opaque conditions: A model for Francophone Belgium, 1981– 2010. *International Journal of Forecasting*, 28(4), pp. 777-788.
- Foucault, M. and Nadeau, R. (2012). Forecasting the 2012 French Presidential Election. *PS: Political Science and Politics*, 45(2), pp. 218-222.
- Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge University Press.
- Hagevi, M. (2015). *Den svenska väljaren 2014*. Gleerups Utbildning AB.
- Holmberg, S., & Oscarsson, H. (2013). Party Leader Effects on the Vote. In K. Aarts, A. Blais, & H. Schmitt (Eds.), *Political leaders and democratic elections*. OUP Oxford.
- Hopmann, D. N., Vliegenthart, R., De Vreese, C., & Albæk, E. (2010). Effects of election news coverage: How visibility and tone influence party choice. *Political communication*, 27(4), 389-405.
- Lewis-Beck, M.S. (2005) Election Forecasting: Principles and Practice. *The British Journal of Politics and International Relations*, 7(2), pp. 145-164.
- Lewis-Beck, M.S. & Dassonneville, R. (2015a) Comparative election forecasting: Further insights from synthetic models. *Electoral Studies*, 39, pp. 275-283.
- Lewis-Beck, M. & Dassonneville, R (2015b). Forecasting elections in europe: Synthetic models. *Research & Politics*, 2(1), pp. 1-11.
- Lewis-Beck, M.S. & Stegmaier, M. (2014) US Presidential Election Forecasting. *PS: Political Science & Politics*, 47(2), pp. 284-288.

- Lewis-Beck, M. S. & Paldam, M. (2000) Economic voting: an introduction. *Electoral Studies*, 19(2), pp. 113-121.
- Lewis-Beck, M. S. & Tien, C. C. (2011) Election forecasting. In Michael P. Clements, & David F. Hendry (Eds.), *The Oxford handbook of economic forecasting* (pp. 655–671). New York: Oxford University Press.
- Lewis-Beck, M. S. & Rice, T. W. (1992) *Forecasting elections*. Washington, DC: Cq Pr.
- Linzer, D. (2013) Dynamic Bayesian Forecasting of Presidential Elections in the States. *Journal of the American Statistical Association*, 108:501, pp. 124-134.
- Magalhães, P.C., Aguiar-Conraria, L. & Lewis-Beck, M.S. (2012) Forecasting Spanish elections. *International Journal of Forecasting*, 28(4), pp. 769-776.
- Marklund, J. (2018). *Business Process Modeling, Simulation and Design*. CRC Press.
- Nadeau, R., Bélanger, É. & Lewis-Beck, M. S. (2012) Choosing the dependent variable: Sarkozy's forecasting lesson. *French Politics*, 10(4), pp. 364–368.
- Nadeau, R. & Lewis-Beck, M. S. (2020) Election forecasts: Cracking the Danish case. *International Journal of Forecasting*, 36 (3), pp. 892-898.
- Norpoth, H. & Gschwend, T. (2003) Against all odds? The red–green victory. *German Political Studies*, 21, pp. 15–34.
- Norpoth, N. & Gschwend, T. (2010) The chancellor model: Forecasting German elections, *International Journal of Forecasting*, 26 (1), pp. 42-53.
- Oegema, D., & Kleinnijenhuis, J. (2000) Personalization in political television news: A 13-wave survey study to assess effects of text and footage. *Communications*, 25(1), 43-60.
- Oscarsson, Henrik (2020) Flytande väljare 2018. Valforskningsprogrammets rapportserie 2020:10. Göteborgs universitet: Statsvetenskapliga institutionen.
- Oscarsson, Henrik. & Holmberg, S. (2013) Allt mer lätttrörliga väljare - Vad är nytt och vad är gammalt? SCB, Demokratistatistik rapport 16.
- Oscarsson, Henrik. & Holmberg, S. (2008) Alliansseger: Redogörelse för 2006 års valundersökning i samarbete mellan Statsvetenskapliga institutionen vid Göteborgs universitet och Statistiska centralbyrån. Allmänna valen 2006, Del 4.
- Sundell, A. & Lewis-Beck, M.S. (2014) Forecasting the 2014 Parliamentary Election in Sweden. Available at: SSRN 2450229.
- Vestin, E., & Öhrvall, R. (2020). Att förutsäga amerikanska presidentval. *Statsvetenskaplig Tidskrift*, 122(5), 127-139.

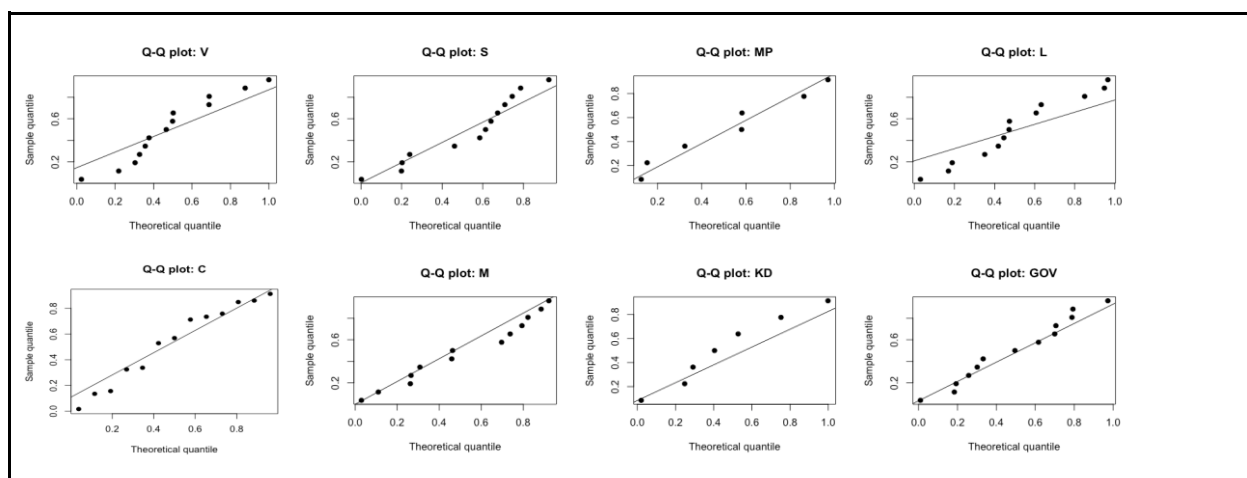
Walther, D. (2015) Picking the winner(s): Forecasting elections in multiparty systems. *Electoral Studies*, 40 (C), pp. 1–13.

Öhrvall, R. (2014) Trehundratolv: Att förutsäga valresultat. *Svenska Dagbladet*.
<http://blog.svd.se/trehundratolv/2014/08/25/att-forutsaga-valresultat/> (Hämtad 2020-12-28).

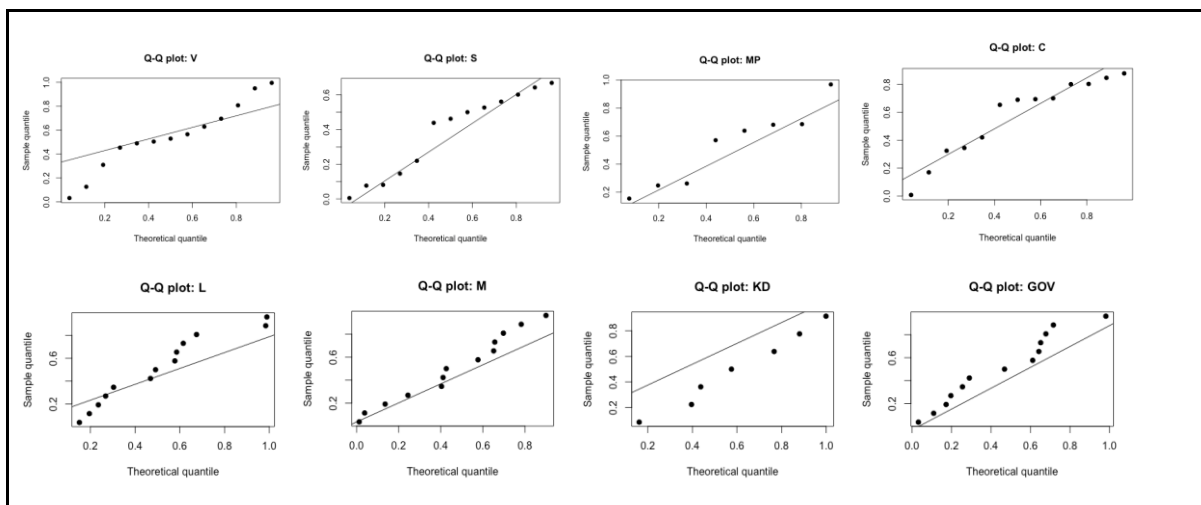
Appendix A: Q-Q plots för modell 1–8



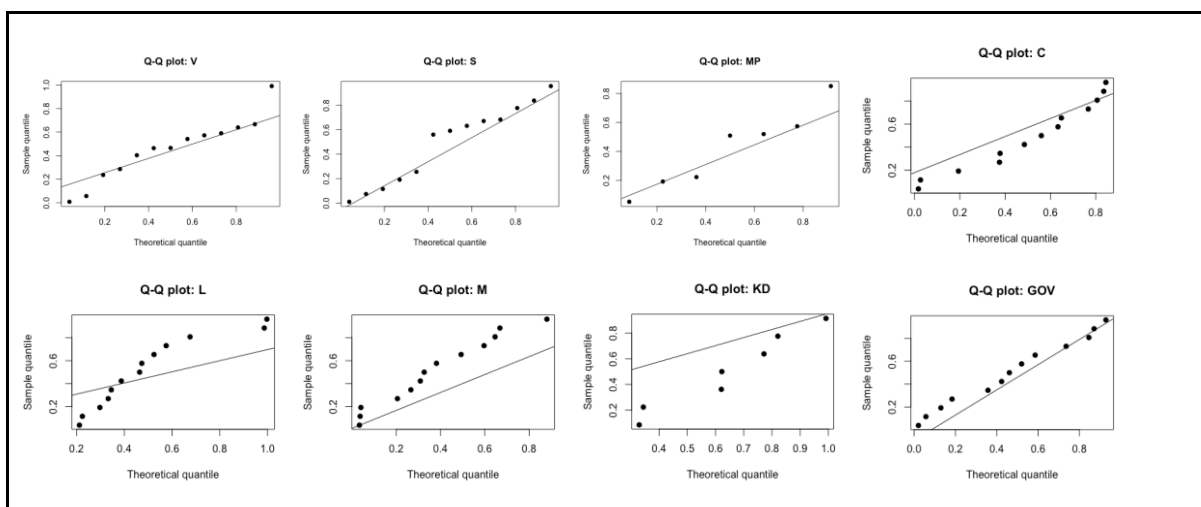
Figur 2: Q-Q plot för respektive parti som jämför en teoretisk likformig fördelning på den horisontella axeln med empirisk data (percentiler) för modell 1 på den vertikala axeln



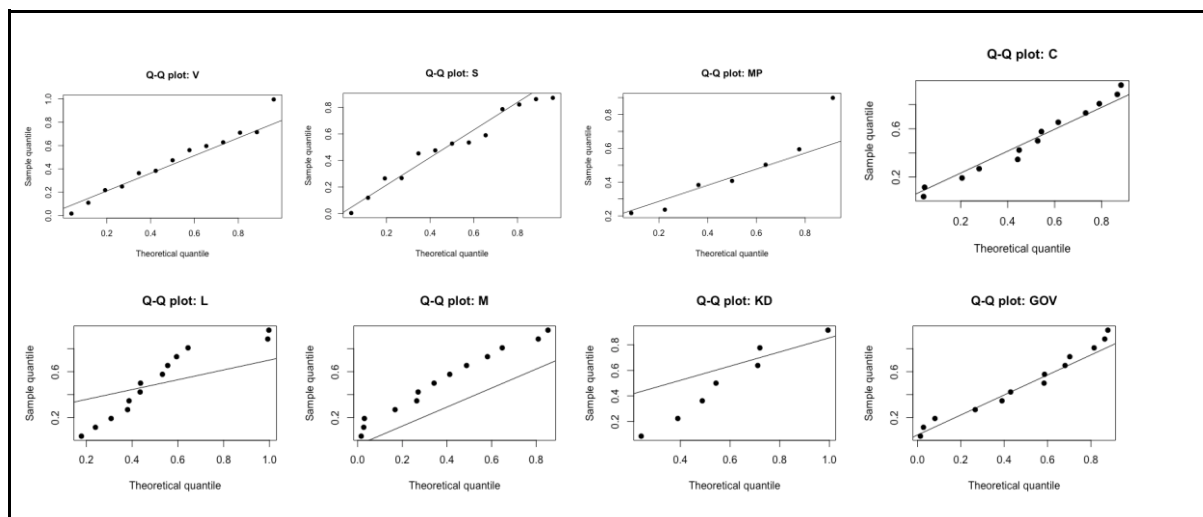
Figur 3: Q-Q plot för respektive parti som jämför en teoretisk likformig fördelning på den horisontella axeln med empirisk data (percentiler) för modell 2a på den vertikala axeln



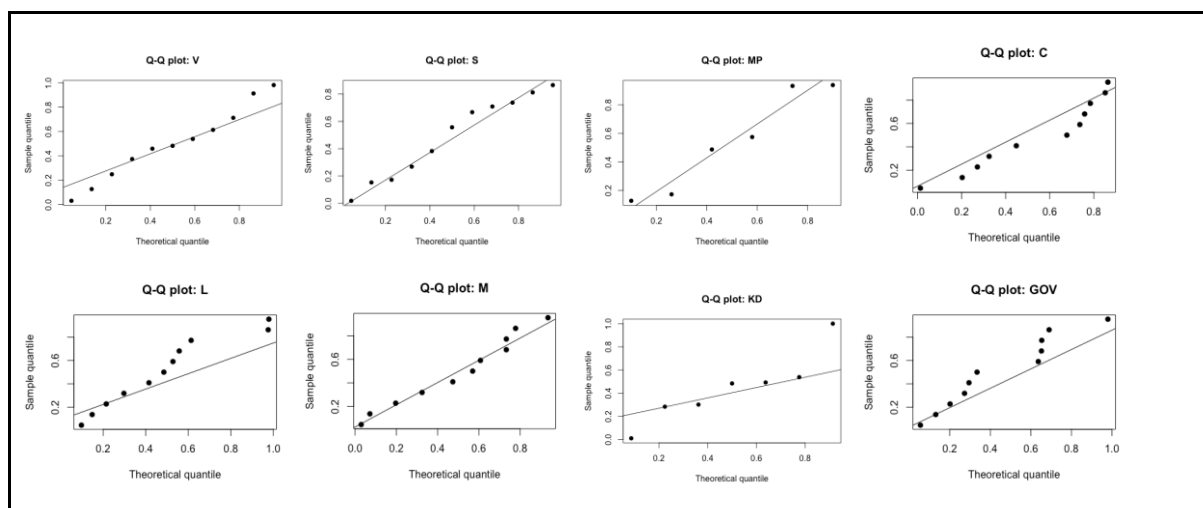
Figur 4: Q-Q plot för respektive parti som jämför en teoretisk likformig fördelning på den horisontella axeln med empirisk data (percentiler) för modell 2b på den vertikala axeln



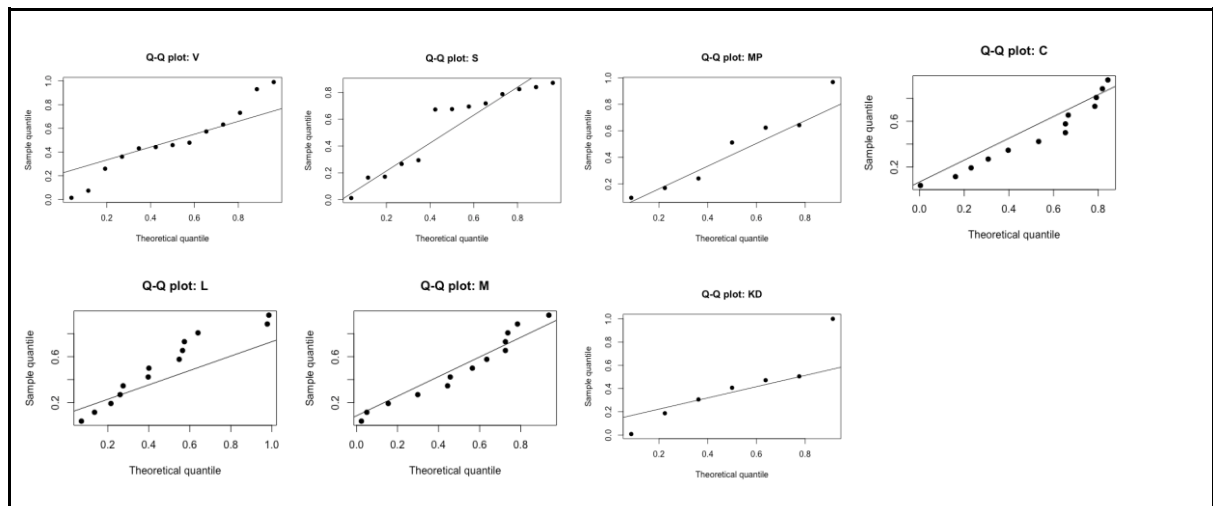
Figur 5: Q-Q plot för respektive parti som jämför en teoretisk likformig fördelning på den horisontella axeln med empirisk data (percentiler) för modell 3 på den vertikala axeln



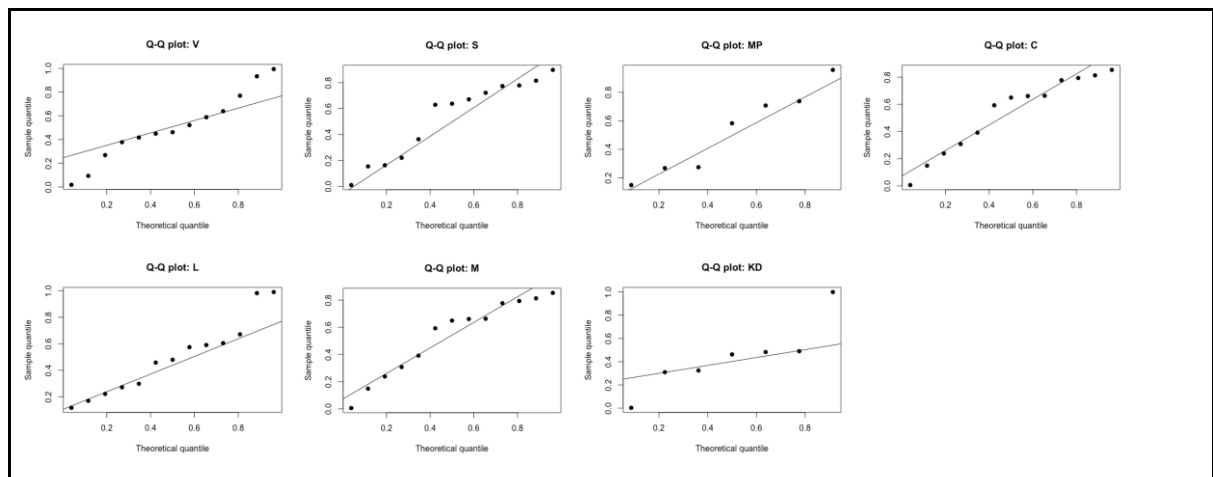
Figur 6: Q-Q plot för respektive parti som jämför en teoretisk likformig fördelning på den horisontella axeln med empirisk data (percentiler) för modell 4 på den vertikala axeln



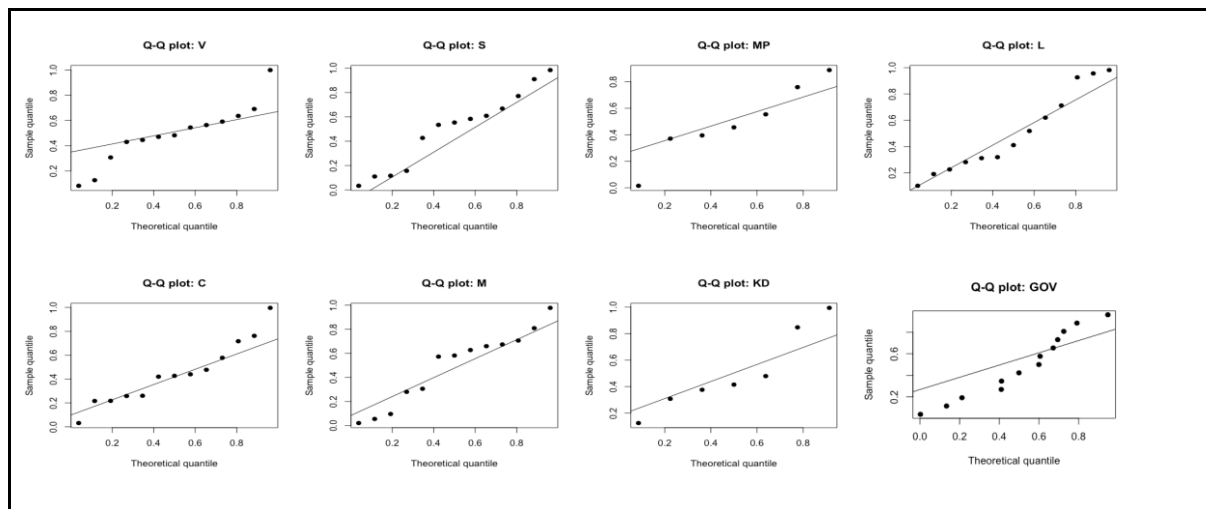
Figur 7: Q-Q plot för respektive parti som jämför en teoretisk likformig fördelning på den horisontella axeln med empirisk data (percentiler) förmodell 5 på den vertikala axeln



Figur 8: Q-Q plot för respektive parti som jämför en teoretisk likformig fördelning på den horisontella axeln med empirisk data (percentiler) för modell 6 på den vertikala axeln



Figur 9: Q-Q plot för respektive parti som jämför en teoretisk likformig fördelning på den horisontella axeln med empirisk data (percentiler) för modell 7 på den vertikala axeln



Figur 10: Q-Q plot för respektive parti som jämför en teoretisk likformig fördelning på den horisontella axeln med empirisk data (percentiler) för modell 8 på den vertikala axeln

Appendix B: Variabler och källor

Variabel: Valresultat

SCB: "Riksdagsval - erhållna mandat efter region och parti. Valår 1973 - 2018"

https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START_ME_ME0104_ME0104C/Riksdagsmandat/

Variabel: Valresultat i föregående riksdagsval (VFR)

SCB: "Riksdagsval - erhållna mandat efter region och parti. Valår 1973 - 2018"

https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START_ME_ME0104_ME0104C/Riksdagsmandat/

SCB: "Historisk statistik över valåren 1910–2014. Procentuell fördelning av giltiga valsedlar efter parti och typ av val" (för valresultat före 1973)

<https://www.scb.se/hitta-statistik/statistik-efter-amne/demokrati/allmanna-val/allmanna-val-valresultat/pong/tabell-och-diagram/historisk-valstatistik/historisk-statistik-over-valaren-19102014.-procentuell-fordelning-av-giltiga-valsedlar-efter-parti-och-typ-av-val/>

Variabel: PSU_majdiff

SCB: "Valresultat om det varit val idag (PSU)"

https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START_ME_ME0201_ME0201A/Vid10/#

Variabel: Arbetslöshet

SCB: "Befolkningen 15-74 år (AKU) efter kön, ålder och arbetskraftstillhörighet. Kvartal 1970K1 - 2020K3"

https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START_AM_AM0401_AM0401A/NAKUBefolkning2K/#

Variabel: Inflation

OECD: "Inflation (CPI)"

https://data.oecd.org/price/inflation-cpi.htm?fbclid=IwAR0aHomn5gsWREcdOMX64I8_Lv9FlxMhtxjAMdF2nXPHtz11eVx9xhyVS3Y

Variabel: BNP_tillväxt

SCB: “BNP från användningssidan, försörjningsbalans efter användning, aggregerad. År 1950 - 2018”

https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START_NR_NR0103_NR0103E/NR0103T01BNAr/

Variabel: Partiledarpopularitet (partiled_pop)

Oscarsson, Henrik & Sören Holmberg (2020) Swedish Voting Behavior. Swedish National Election Studies, Working Paper Series. Report 2020:1. University of Gothenburg, Department of Political Science.

Holmberg, S. (2008). Mona Sahlin-The Comeback Woman. SOM-institutet.

https://ipd.gu.se/digitalAssets/1287/1287273_kapitel_07.pdf

Dagens Nyheter: “Förtroendet minskar för partiledarna”

<https://www.dn.se/nyheter/sverige/fortroendet-minskar-for-partiledarna/>

SVT/Sifo Partiledarförtroendet –Maj 2015

https://www.kantarsifo.se/sites/default/files/reports/documents/partiledarf_rtroende_2016_maj2_1.pdf

DN/IPSOS PARTILEDARFÖRTROENDE - maj 2018

https://www.ipsos.com/sites/default/files/ct/news/documents/2018-05/dn_ipsos_partiledarfortroende_20180528.pdf

Variabel: Näst_bästa_parti (NBP)

Oscarsson, Henrik & Sören Holmberg (2020) Swedish Voting Behavior. Swedish National Election Studies, Working Paper Series. Report 2020:1. University of Gothenburg, Department of Political Science.

Variabel: Mediaomnämnannde (media)

Kodord som har använts:

Vänsterpartiet

Vänsterpartiets
Vänsterpartistiska
Socialdemokraterna
Socialdemokraternas
Socialdemokratiska
Miljöpartiet de Gröna
Miljöpartiet
Miljöpartiets
Centerpartiet
Centerpartiets
Centerpartistiska
Folkpartiet liberalerna
Folkpartiet liberalernas
Folkpartiet
Folkpartiets
Folkpartistiska
Liberalerna
Liberalernas
Moderata samlingspartiet
Moderaterna
Moderaternas
Nya Moderaterna
Nya Moderaternas
Kristdemokraterna
Kristdemokraternas
Kristdemokratisk