



Kunstig intelligens i makroøkonomisk prognosearbeid

*En empirisk studie av hvor godt maskinlæring evner å predikere norsk
økonomisk vekst*

Christian Alex Bankson og Andreas Madsen Holm

Veileder: Torfinn Harding

Masterutredning i økonomi og administrasjon

Hovedprofil: Finansiell økonomi

NORGES HANDELSHØYSKOLE

Dette selvstendige arbeidet er gjennomført som ledd i masterstudiet i økonomi- og administrasjon ved Norges Handelshøyskole og godkjent som sådan. Godkjenningen innebærer ikke at Høyskolen eller sensorer inntår for de metoder som er anvendt, resultater som er fremkommet eller konklusjoner som er trukket i arbeidet.

Forord

Denne masterutredningen er skrevet som det avsluttende, selvstendige arbeidet i masterstudiet i økonomi og administrasjon ved Norges Handelshøyskole (NHH). Utredningen utgjør 30 studiepoeng innenfor hovedprofilen Finansiell Økonomi, og tar for seg emner innenfor makroøkonomisk prognosearbeid og kunstig intelligens.

Utredningens formål er å undersøke i hvilken grad økonomisk vekst kan anslås ved hjelp av maskinlæring, en sentral gren innen kunstig intelligens. I løpet av vårt studieløp på NHH har vi fulgt kurs som omfatter samspillet mellom finansielle markeder og makroøkonomiske forhold. Her har vi blant annet opparbeidet god innsikt i de komplekse forholdene bak konjunktursvingninger. Med maskinlæringens framgang innenfor flere fagområder, synes vi det både er interessant og motiverende å utforske maskinlæringens potensielle bidrag til det makroøkonomiske fagmiljøet.

Vi vil benytte anledningen til å rette en stor takk til vår veileder, Torfinn Harding, for inspirerende og konstruktiv veiledning denne høsten. Videre ønsker vi å takke Manasa Patnam og det øvrige teamet i IMF for faglige innspill og viktig innsikt i algoritmisk tilnærming til makroøkonomiske anslag. Avslutningsvis takker vi venner og familie som har bidratt med støtte og motiverende ord i løpet av høsten.

Norges Handelshøyskole

Bergen, desember 2019

Sammendrag

Denne masterutredningen introduserer anvendelsen av maskinlæring i utarbeidelsen av makroøkonomiske anslag. Vårt studie analyserer hvor godt maskinlæring predikerer økonomisk vekst i Fastlands-Norge. Basert på et kvartalsvis datasett trenes tre ulike algoritmer til å produsere kvartalsvise og årlige prediksjoner. For å vurdere prediksjonene sammenlignes de med tidligere avgitte prediksjoner fra Norges Bank, SSB, Finansdepartementet og DNB. For å sikre et rettferdig informasjonsgrunnlag for alle prediksjoner tilpasses fremgangsmåten for maskinlæringen slik at prediksjonstidspunktet for de ulike aktørene ivaretas. Maskinlæringen vil likevel ha en fordel ettersom prediksjonene er basert på senere reviderte data, til forskjell fra de historiske prediksjonene fra de resterende aktørene. For de kvartalsvise prediksjonene gjennomføres også en robusthetstest ved å utvide prediksjonshorizonten til å inkludere finanskrisen i 2008 og 2009.

Våre resultater underbygger klart at maskinlæring produserer minst like gode prediksjoner som de etablerte aktørene. For både kvartalsvise og årlige prediksjoner leverer den beste maskinlæringsmodellen, Super Learner, de mest presise prediksjonene for Fastlands-Norge. For den kvartalsvise sammenligningen med Norges Bank viser analysen at årene 2010 og 2011 er avgjørende. Maskinlæringen evner å fange opp tidspunktet for den kraftige veksten norsk økonomi opplevde på en bedre måte enn Norges Bank. Det samme mønsteret går igjen for de årlige prediksjonene. Spesielt for 2011 evner Super Learner å predikere mer presist enn Norges Bank, SSB, DNB og Finansdepartementet. Prediksjonene til de fire aktørene forbedrer seg utover perioden, mens maskinlæringen viderefører sin presise prediksjon. Det kan virke som at maskinlæringen evner å fange opp de strukturelle endringene etter finanskrisen raskere enn de andre aktørene, både ved kvartalsvis og årlig prediksjon. I robusthetstesten på de kvartalsvise prediksjonene sliter maskinlæringen med å predikere omfanget av finanskrisen. Likevel leverer både Super Learner og Elastic Net modellene absolutte avvik som i snitt er lavere enn 1 %-poeng. Selv om maskinlæringen besitter et informasjonsfortrinn gjennom reviderte input data, mener vi en algoritrisk tilnærming til prediksjon av norsk økonomisk vekst vil være verdifullt for det makroøkonomiske fagmiljøet.

Nøkkelord – Maskinlæring, Prediksjon, Norsk økonomisk vekst

Innhold

1	Innledning	1
2	Bakgrunn	4
2.1	Litteraturgjennomgang	6
3	Data	8
3.1	Data til prediksjoner	8
3.2	Data til evaluering av prediksjoner	11
3.3	Overfitting	13
3.4	Utfordringer med makroøkonomiske data	15
4	Metode	16
4.1	Prediksjon med maskinlæring	16
4.1.1	Overvåket og uovervåket læring	16
4.1.2	Trenings- og testdata	16
4.1.2.1	Fremgangsmåte for å unngå look-ahead bias	17
4.1.3	Kryssvalidering	19
4.1.4	Rolling Forecasting Samples	20
4.1.4.1	Caret	21
4.2	Elastic Net	22
4.2.1	Ridge regresjon	23
4.2.2	Lasso regresjon	24
4.2.3	Elastic Net	24
4.3	Random Forest	26
4.3.1	Bootstrap aggregating	27
4.3.2	Random Forest	27
4.4	Super Learner	28
4.5	Måltall for vurdering av presisjon	30
5	Resultater	32
5.1	Vurderingsgrunnlag og resultater	32
5.1.1	Kvartalsvise prediksjoner	32
5.1.1.1	Robusthet ved prediksjon av kriseperioder	36
5.1.2	Årlige prediksjoner	38
6	Diskusjon	42
6.1	Verdien av ny informasjonen	42
6.2	Evaluering av kvartalsvise prediksjoner	43
6.2.1	Variasjonen mellom maskinlæringsmodellene	43
6.2.2	Norges Bank og Super Learner	45
6.2.2.1	Nærmere om avvikene i perioden 2010-2012	48
6.2.3	Robusthet ved prediksjon av kriseperioder	50
6.3	Evaluering av årlige prediksjoner	52
6.4	Variable Importance	55
6.4.1	Sentrale variabler i kvartalsvise prediksjoner for 2007-2018	56
6.5	Potensielle mangler ved maskinlæring	58

7 Konklusjon	61
Referanser	63
Appendiks	66
A1 Data	66
A1.1 Prediksjoner innhentet fra andre aktører	66
A2 Figurer	68
A2.1 Data og metode	68
A2.2 Utarbeidelsen av kvartalsvise prediksjoner	69
A2.3 Utarbeidelsen av årlige prediksjoner	70
A2.4 Øvrige figurer	71
A3 Variabler	72

Figurliste

3.1	Overfitting - resultatet av økt kompleksitet	14
4.1	Fremgangsmåte for utarbeidelse av kvartalsvise prediksjoner	17
4.2	Informasjonsgrunnlag for prediksjon av kvartal T+1	18
4.3	Årlige prediksjonstidspunkt for ulike aktører	19
4.4	Rolling Forecasting Origin - Inndeling av tidsseriedata i Caret	22
4.5	Eksempel på et beslutningstre med regresjoner	26
5.1	Indeksert absolutt avvik - kvartalsvise prediksjoner for 2010-2018	35
5.2	Indeksert absolutt avvik - kvartalsvise prediksjoner for 2007-2018	37
5.3	Avvik fra faktisk vekst - årlige prediksjoner for 2010-2017	40
5.4	Indeksert absolutt avvik - årlige prediksjoner for 2010-2017	41
6.1	Kvartalsvise prediksjoner av firekvartalsvekst - Norges Bank, Super Learner og faktisk vekst	46
6.2	Kvartalvise prediksjoner av firekvartalsvekst - SAM, Super Learner og faktisk vekst	47
6.3	Kvartalvise prediksjoner av firekvartalsveks - Super Learner, Elastic Net og faktisk vekst	52
6.4	De 10 «viktigste» variablene bak kvartalsvis prediksjon	56
A2.1	Bias vs. Variance - Overfitting	68
A2.2	Kryssvalidering Elastic Net	69
A2.3	Kryssvalidering Elastic Net - Plot av kombinasjoner gitt RMSE	69
A2.4	Kryss-validering Random Forest	70
A2.5	Kryss-validering Super Learner - vekting av modeller	70
A2.6	Prognoser av Super Learner med og uten informasjonsfortrinn	71
A2.7	Fordeling av absolutte avvik i prediksjoner fra 2007 til 2018	71
A3.1	Korrelasjonsmatrise	72

Tabelliste

4.1	Utvalgte algoritmer for Super Learner	29
5.1	Resultat av kvartalsvise prediksjoner for 2010-2018	32
5.2	Konfidensintervaller for avvik - kvartalsvise prediksjoner	33
5.3	Deskriptiv sammenligning av Super Learner og Norges Bank - kvartalsvise prediksjoner for 2010-2018	34
5.4	Deskriptiv statistikk - kvartalsvise prediksjoner for 2007-2018	36
5.5	Resultat av årlige prediksjoner for 2010-2017	39
5.6	Deskriptiv statistikk - årlige prediksjoner for 2010-2017	39
6.1	Look-ahead bias i kvartalsvise prediksjoner for 2010-2018	42
6.2	Deskriptiv statistikk - kvartalsvise prediksjoner av kriseperioder	50
A1.1	Kvartalsvise prediksjoner Norges Bank	66
A1.2	Årlige prediksjoner fra ulike aktører	67
A1.3	Faktisk BNP vekst Fastlands-Norge 2018-2010	68
A2.1	Prediksjoner for årlig vekst, 2018-2020	70
A3.1	Norske forklaringsvariabler i datasettet	73
A3.2	Norske forklaringsvariabler i datasettet	74
A3.3	Internasjonale forklaringsvariabler i datasettet	75
A3.4	Deskriptiv statistikk av utvalgte variabler fra 1996 til 2018	76

1 Innledning

Makroøkonomiske prognoser er svært viktige for å forstå hvilken situasjon et lands økonomi befinner seg i. Brutto nasjonalprodukt (BNP) setter et målbart tall på verdiskapingen i landet, og anses som den viktigste variabelen for beskrivelsen av den økonomiske tilstanden (SSB, 2019). Prognoser som forsøker å anslå videre vekst vil derfor være avgjørende for en rekke beslutningstakere, både i politikken og næringslivet. Sentralbanker og regjeringer ønsker jevn og stabil vekst når styringsrente og statsbudsjett avgjøres, mens beslutningstakere i næringslivet må vite den makroøkonomiske risikoen ved store investeringsbeslutninger. I denne prosessen kan prediksjoner gi et viktig innblikk i hvordan fremtiden vil se ut, og hvilke tiltak som burde innføres basert på dette. Presisjonen i prediksjonene vil være svært avgjørende for å treffe de rette beslutningene.

Viktigheten av nøyaktighet i prediksjonene gjør at en rekke aktører har spesialisert seg mot å tilby de mest presise prediksjonene i markedet. Formålet med prediksjonene varierer fra aktør til aktør. Offentlige aktører som Norges Bank og Finansdepartementet er avhengig av presise prediksjoner for å kunne utøve den riktige økonomiske politikken. Private aktører som banker, meglerhus og prognosehus tilbyr ekspertise overfor sine kunder, som i mange sammenhenger er viktige beslutningstakere i næringslivet. I tillegg leverer frittstående internasjonale organisasjoner som IMF prediksjoner for å informere politiske aktører og næringsliv om norsk økonomi. Makroøkonomiske prognoser publiseres gjerne to til fire ganger i året, med tilhørende dybdeanalyser av ulike sektorer i økonomien.

Tradisjonelt har statistiske modeller hatt en viktig rolle i utarbeidelsen av prediksjoner. Norges Bank har for eksempel sin egenutviklede SAM (System of Averaging Models) modell, mens DNB benytter seg av enklere regresjonsmodeller. Norges Bank har investert tungt i utarbeidelsen av SAM og prediksjonene representerer et viktig utgangspunkt for deres offisielle prediksjoner som publiseres i pengepolitiske rapporter (PPR)(Aastveit et al., 2011). DNB bruker i større grad skjønnsmessige vurderinger i utarbeidelsen av sine prediksjoner, der modeller i mindre grad blir vektlagt. Maskinlæring har fått økende grad av oppmerksomhet de siste årene, men innenfor prognoseutarbeidelse har det vært lav interesse i forskningsmiljøet(Jung et al., 2018).

Innenfor en rekke sektorer i næringslivet er maskinlæring raskt blitt tatt i bruk for å effektivisere tidligere standardiserte arbeidsoppgaver. I makroøkonomisk prediksjonsarbeid er derimot bruken av maskinlæring fortsatt på et tidlig stadie. Dette til tross for at mange er enige om at maskinlæring vil spille en viktig rolle i fremtidens prediksjonsarbeid. Eksempelvis bruker Norges Bank sin SAM modell som et utgangspunkt for den skjønnsmessige vurderingen tilknyttet hver enkelt prediksjon. Dersom en modellbasert prediksjon kan gi en presis og troverdig output vil det være et ideelt utgangspunkt for videre diskusjon og vurdering.

Vi ønsker å gi en indikasjon på hvilken verdi maskinlæring faktisk kan tilføre dagens utarbeidelse av makroøkonomiske anslag. Vår problemstilling for denne oppgaven er derfor:

Hvor godt evner maskinlæring å predikere norsk økonomisk vekst?

For å belyse problemstillingen vil vi benytte maskinlæringsmodellene Random Forest (RF), Elastic Net og Super Learner til å predikere BNP-vekst for Fastlands-Norge i perioden 2010-2018. Vi vil gjennomføre kvartalsvise og årlige prediksjoner. I tillegg til å vurdere modellenes prestasjoner opp mot hverandre, vil vi vurdere prestasjonen til den beste modellen opp mot etablerte aktørers prediksjoner for den respektive perioden. På denne måten ønsker vi å belyse hvor godt maskinlæring fungerer til å predikere norsk økonomisk vekst.

I besvarelsen av problemstillingen har vi foretatt flere begrensninger som må presiseres. Modellene som benyttes tilhører ulike grupper maskinlæringsmodeller. Random Forest er en type *beslutningstre*, Elastic Net kombinerer to *lineære modeller* og Super Learner er en *samlemodell*. De nevnte modellene representerer generelle tilnærminger for sine respektive grupper. Innenfor hver gruppe finnes det nemlig flere varianter av hver modell. I tillegg finnes det andre typer modellgrupper som ikke er inkludert. Eksempelvis er Nevrale nettverk en mye anvendt gruppe som vi ikke dekker i denne oppgaven. Valget om å fokusere på de tre nevnte modellene impliserer at vi ikke kan utelukke at det finnes andre varianter eller grupper som kan levere mer presise prediksjoner for norsk økonomisk vekst. I tillegg vil vårt valg av generelle representanter fra noen grupper gi et mindre fokus på andre aspekter i analysen. Mange maskinlæringsmodeller produserer presise prediksjoner, men gir begrenset innsyn i bakgrunnen for prediksjonene. I vår analyse har vi fokusert på modellenes evne til å predikere, men vektlegger ikke hvorvidt de samme modellene er

praktisk anlagt for analyse av bakgrunnen for prediksjonene. For videre forskning vil det derfor være interessant med en dypere analyse av flere typer modeller, enten med fokus på presisjon eller analyse av bakgrunnen for prediksjonen.

Våre resultater viser at standard maskinlæring produserer meget gode prediksjoner for norsk økonomisk vekst. Vi finner likevel relative forskjeller i prestasjon mellom de ulike modellene. Super Learner kommer desidert best ut både for kvartalsvise og årlige prediksjoner. Sammenlignet med Norges Bank leverer Super Learner kvartalsvise prediksjoner som i snitt har 19%¹ lavere avvik i perioden 2010-2018. Dette innebærer at Super Learner sine prediksjoner i snitt ligger 0.12 %-poeng nærmere faktisk vekst, relativt til Norges Bank. Resultatet kommer først og fremst som en følge av at Norges Bank har flere store avvik, spesielt i årene 2010 og 2011. Våre funn viser videre at Super Learner også leverer mer presise årlige prediksjoner relativt til flere etablerte aktører innen fagmiljøet. For årene 2010-2017 leverer Super Learner i snitt 38.6% mer presise prediksjoner enn Norges Bank, DNB, Finansdepartementet og Statistisk Sentralbyrå (SSB). Eksempelvis er maskinlæringens prediksjoner i snitt 0.28 %-poeng nærmere faktisk vekst enn SSB sine prediksjoner. Spesielt oppsiktsvekkende er det at maskinlæringen produserer mer presise prediksjoner sammenlignet med DNB, som har et informasjonsfortrinn ved sine prediksjoner. Til tross for at DNB sine prediksjoner inneholder to måneder ekstra informasjon, leverer Super Learner 38% mer presise prediksjoner.

I oppgavens videre struktur vil vi først presentere datagrunnlaget for prediksjonene, både med hensyn til kilder, valg av variabler og utfordringer som knytter seg til makroøkonomiske data. Videre tar vi for oss teorien bak hver maskinlæringsmodell. Her vil vi i detalj gjennomgå det metodiske bak maskinlæringens prediksjoner. De viktigste resultatene vil følgelig presenteres, både for kvartalsvise og årlige prediksjoner. Til slutt vil vi drøfte mulige forklaringer bak resultatene vi har presentert. Dette vil blant annet inkludere interne forskjeller mellom maskinlæringsmodellene, i tillegg til en diskusjon rundt maskinlæringens prestasjon relativt til de etablerte aktørene.

¹Relativ sammenligning av RMSE. Beskrives nærmere i seksjon 4.5

2 Bakgrunn

Våren 2019 var det stor diskusjon i mediene tilknyttet prediksjoner for den økonomiske veksten i eurosonen. Sjefsstrateg i Danske Bank, Christian Lie, var sterkt uenig med DNB sin makroøkonom Ole Kristian Kjennerud. Lie mente at den økonomiske veksten i Europa skulle bremse opp, mens Kjennerud mente veksten skulle tilta (Christensen, 2019). Vi ble svært fascinert av hvordan to så dyktige makroøkonome kunne være så uenige om retningen for den økonomiske veksten. Vi har i flere år fulgt med på makroøkonomiske oppdateringer fra begge banker, men sjelden har vi lagt merke til at man er uenige om retningen på veksten. Vårt inntrykk er heller at uenighetene ofte består i hvor mye veksten kommer til å stige eller synke. Det ble tydelig at de to økonomene vektla ulike makroøkonomiske indikatorer i sine analyser, men hvordan kunne man vite hvem som hadde rett?

Sommeren 2019 arbeidet en av forfatterne hos DNB Markets i Oslo. Forfatteren benyttet muligheten til å spørre DNB sin sjefsøkonom, Kjersti Haugland, om hvordan makroøkonomene i DNB arbeider med makroøkonomiske prediksjoner. Gjennom en svært interessant og uformell samtale, kom det frem at prediksjonene blir til gjennom en skjønnsmessig faglig vurdering. At prediksjonene i liten grad hviler på modeller, er spesielt interessant. Ifølge Haugland er deres boligprismodell den mest avanserte modellen som anvendes i forbindelse med prediksjonen.

Maskinlæring og kunstig intelligens har de siste årene vært mye omtalt i medier og diverse rapporter fra næringslivet. Det virker tydelig at maskinlæring er kommet for å bli, og trolig vil revolusjonere mange ulike fagområder. Aktører i finansmarkedet ligger langt fremme i bruken av maskinlæring. Hedgefond som Bridgewater Associates og Renaissance Technologies var tidlig ute med å implementere maskinlæring og deep-learning i deres trading systemer. Maskinlæringsalgoritmene i disse hedgefondene har blitt rapportert til å utkonkurrere kvantefond, tradisjonelle hedgefond og den brede indeksen S&P500 over lengre perioder (Aegasis AI, 2019). Renaissance Technologies har utmerket seg markant med en årlig avkastning på 71.8% fra 1994 til midten av 2014 (Collins og Rubin, 2015).

Makroøkonomisk prediksjon er et av områdene der maskinlæring potensielt vil spille en viktig rolle. Gitt vår interesse for makroøkonomi er det derfor interessant å undersøke hvor gode prediksjoner vi kan lage ved hjelp av maskinlæring. Med lite forkunnskaper om maskinlæring så vi det som en gylden mulighet til å tilegne oss verdifull erfaring. Kunnskapen vi tilegner oss kan bli nyttig i flere situasjoner i næringslivet. Det virker å være en bred enighet om at maskinlæring vil bli viktig innen makroøkonomisk prediksjon. Likevel er det interessant at få aktører anvender maskinlæring i dagens utarbeidelse av prediksjoner.

Maskinlæring benytter historisk informasjon til å forstå sammenhenger og predikere økonomiens videre utvikling. På mange måter ligner dette på prosessen i prediksjonsarbeidet til DNB. Forskjellen er at det benyttes en maskin som arbeider raskt, systematisk og evner å analysere komplekse sammenhenger. Spørsmålet er hvorvidt maskinlæring i praksis kan benyttes som et verktøy for prediksjoner av økonomisk vekst. I fagmiljøet virker det å være en bred oppfatning om at implementering av maskinlæring foreløpig ikke vil tilføre verdi. Oppfatningen varierer derimot fra aktør til aktør. Enkelte foretrekker skjønnsmessige vurderinger, mens andre har investert tungt i bruken av modeller. Blant annet bruker SSB komplekse modeller i sitt arbeid og Norges Bank satser på utvikling av deres SAM modell. SAM er en samlingsmodell som representerer den nærmeste tilnærmingen til maskinlæring innenfor det norske fagmiljøet.

Basert på de nevnte forholdene ønsker vi å se nærmere på hvor godt maskinlæring predikerer veksten i norsk fastlandsøkonomi. Valget av norsk fastlandsøkonomi er sammensatt. I seksjon 2.1 presenteres flere pågående studier der maskinlæring anvendes for å predikere økonomisk vekst internasjonalt, men få har gjort dette for norsk økonomi. I tillegg ønsker vi å være mest mulig relevant for næringsliv og fagmiljø i Norge. Formålet med oppgaven er ikke nødvendigvis å anvende den mest avanserte maskinlæringen som finnes, men heller å belyse hvor gode prediksjoner standard maskinlæring leverer relativt til etablerte aktørers prediksjoner. Vi har derfor bevisst anvendt maskinlæringsmodeller som er forholdsvis enkle å anvende og er intuitive. På denne måten ønsker vi å være like relevant for de med lite, som de med mye, erfaring med maskinlæring.

Vår ambisjon er å bidra til økt oppmerksomhet rundt temaet i det makroøkonomiske fagmiljøet, slik at vi engasjerer til mer debatt og studier innenfor området.

2.1 Litteraturgjennomgang

Vår utredning er et bidrag til en liten, voksende litteratur innenfor anvendelsen av maskinlæring til økonomisk prediksjon. Likevel foreligger det en konsensus i den eksisterende litteraturen om at maskinlæring vil utgjøre et fordelaktig verktøy i fremtidig økonomisk analyse. Tkacz og Hu (1999) var tidlig ute med å presentere bruken av Artificial Neural Network (ANN) til å predikere BNP vekst. I dette kanadiske studiet ble det gjennomført prediksjoner for både ett og fire kvartaler frem i tid. Metodikken ga ingen signifikante utslag på kort sikt, men resulterte i 15-19%² mer presise estimater fire kvartaler frem i tid relativt til etablerte lineære benchmarkmodeller. ANN modellen viste en evne til å fange opp ikke-lineære sammenhenger mellom utviklingen i BNP relativt til de inkluderte finansielle indikatorene. Store datamengder er en forutsetning for en god ANN modell, hvilket kan anses å være en svakhet. Lignende resultater ble replikert av Chuku et al. (2017) som benyttet ANN på makroøkonomiske tidsserier fra en rekke afrikanske land og oppnådde høyere presisjon enn tradisjonelle økonometri- og ARIMA-modeller.

Random Forest blir anvendt av Biau og D'Elia (2010) til å predikere BNP utviklingen i eurosonen for perioden 2004-2009. Studiet fokuserer på å predikere veksten i inneværende kvartal (nowcasting) ved hjelp av data hentet fra månedlige rapporter og undersøkelser. Av resultatene fremstår modellen som et langt mer presist prediksjonsverktøy enn en standard Autoregressiv (AR) modell. Frem til 2008 er Random Forest også i stand til å overgå estimatene utarbeidet av den Europeiske Kommisjon. Biau og D'Elia opprettholder presisjonen etter 2008 ved å kombinere Random Forest med en lineær modell (Bridge) som har gode forutsetninger for kortsiktig prediksjon. Tiffin (2016) lyktes med å utarbeide gode prediksjoner for den økonomiske utviklingen i en fremvoksende markedsøkonomi. Han tok i bruk Elastic Net og Random Forest til å nowcaste BNP-utviklingen i Libanon; et land der offisielle veksttall blir publisert med to års etterslep. Chakraborty og Joseph (2017) sammenligner nøyaktigheten i prediksjonene til en rekke maskinlæringsmodeller opp mot tradisjonelle metoder. Her vurderes maskinlæring til å være det beste verktøyet i prognosearbeidet.

²Relativ sammenligning av RMSE. Beskrives nærmere i seksjon 4.5

Estimeringsfeil som følge av støy, forventningsskjevhet og varians kan minimeres ved å slå ulike modeller sammen (Watson og Stock (2016), Hendry og Clements (2004), Timmermann (2007)). Valland (2019) nowcaster Norsk BNP utvikling med en modelltilnærming som ligner SAM modellen til Norges Bank. Vallands modell sammenlignes opp mot maskinlæringsmodellene Random Forest, Support Vector Machine og ANN. Resultatet viser at maskinlæring ikke nødvendigvis overgår tradisjonelle metoder, men at presisjonen kan forbedres hvis maskinlæringsmodellene anvendes i kombinasjon med andre modeller fra litteraturen. Sett bort fra forskjeller ved valg av algoritmer, vil Valland sin tilnærming skille seg fra vår ettersom Valland fokuserer på nowcasting av BNP-veksten.

Fremgangsmåten i vår utredning tar utgangspunkt i et IMF Working Paper utarbeidet av Jung et al. (2018). Her anvendes datadrevne maskinlæringsmodeller til å predikere reell BNP-vekst, uten å legge restriksjoner for det underliggende forholdet mellom input og output variabler. Studiet er basert på den økonomiske veksten i følgende land: USA, Storbritannia, Tyskland, Spania, Mexico, Filipinene og Vietnam. Modellene som anvendes er Elastic Net, Recurrent Neurale Network (RNN) og Super Learner, en algoritme som evner å sette sammen flere ulike modeller tilpasset prediksjonsformålet.

Outputen fra modellene vurderes opp mot prediksjoner gjort av IMF World Economic Outlook (WEO) for både kvartaler (å/å) og år. På et overordnet nivå fremgår det av studiet at prediksjonene fra maskinlæringen er mest presise for alle land, uavhengig av prediksjonshorizont. Prediksjonsforbedringen på kvartalsvise data svinger mellom 49-83%³, med et gjennomsnitt på 61% på tvers av alle land. Både Elastic Net, og spesielt Super Learner, kommer ut som spesielt gode tilnærminger til kvartalsvise prediksjoner. På årlige data varierer forbedringen mellom 4-38%, men for Spania og Mexico evner ikke maskinlæringsmodellene å overgå estimatene til WEO. På årlig basis fremstår RNN som den beste maskinlæringsmodellen, mens Super Learner underpresterer relativt til de øvrige tilnærmingene i studiet.

³Relativ sammenligning av RMSE. Beskrives nærmere i seksjon 4.5

3 Data

Studier av utviklingen i makroøkonomiske forhold innebærer håndtering av to mye omtalte dimensjoner innen økonometrien; tverrsnitts- og tidsdimensjonen. Wooldridge (2012) presiserer at tidsseriedata skiller seg fra tverrsnittsdata gjennom en tidsbestemt rangering av observasjonene. I makroøkonomisk sammenheng representerer tidsserier gjerne en sekvens tverrsnittsdata indeksert på tid, også kalt paneldata. Ettersom vårt endelige datasett består av flere variabler observert over tid, er datasettet av typen paneldata.

3.1 Data til prediksjoner

For å produsere de beste prediksjonene for norsk økonomisk vekst er det avgjørende med et datasett som dekker alle aspekter ved økonomien, og inkluderer et tilstrekkelig antall observasjoner. I utgangspunktet ønsket vi å samle data tilbake til 1980, men vi fant tidlig ut at det var store variasjoner i historisk tilgjengelighet for de aktuelle variablene. Derfor er variabler med kortere tidsserier også inkludert i datagrunnlaget. Ettersom IMF sitt arbeidsnotat av Jung et al. (2018) har vært en viktig inspirasjonskilde, ble IMF sin World Economic Outlook database et naturlig startpunkt for innsamlingsprosessen.

IMF databasen inneholder mange relevante variabler for vårt prediksjonsformål, slik at vi raskt fikk bygget opp et relativt stort datasett. Datasettet består av variabler som er oppgitt indeksert, på vekst- eller absolutt form. Variablene på vekstform er av typen firekvartalsvekst (å/å), slik at endringene reflekterer utviklingen relativt til samme kvartal forrige år. Vår avhengige variabel, BNP-vekst for Fastlands-Norge, er også på formen firekvartalsvekst. En viktig implikasjon er derfor at prediksjonene som produseres må avleses som firekvartalsvekst. I dette valget hadde vi også en baktanke rundt vår tilnærming til prediksjon av årlig vekst, der vi ønsker å bruke firekvartalsveksten i 4.kvartal. Ved å velge firekvartalsvekst kunne vi holde oss til ett datasett for produksjonen av både årlige og kvartalsvise prediksjoner, fremfor å benytte to ulike datasett. Variabler som oppgis indeksert eller på absolutt form ble hentet ned som de er. Eksempler på dette er aksjeindekser som OSEBX, tall på oljeinvesteringer og indekser for industriproduksjon. Videre er flere makroøkonomiske variabler utsatt for sesongvariasjoner. De aktuelle variablene er innhentet ferdig sesongjustert.

IMF sin database publiseres to ganger i året. Ved hver publisering blir nye datapunkter lagt til, i tillegg til at eksisterende datapunkter oppdateres. Revideringer er svært vanlig for makroøkonomiske variabler, ettersom målingene foregår kontinuerlig. Vi har valgt å anvende den nyeste databasen, fra 3.kvartal 2019. Variablene som ikke var tilgjengelige hos IMF har vi hentet fra Bloomberg og Norges Bank. Bloomberg er en søkemotor som samler inn store mengder data fra ulike nasjonale og internasjonale aktører. De nasjonale statistiske byråene i de respektive landene er vanlige primærkilder, som for Norge tilsvarer SSB. Fra Norges Bank har vi først og fremst hentet data om ulike rente- og valutavariabler, som den norske styringsrenten.

I tradisjonelle statistiske regresjonsmodeller er stasjonaritet et viktig aspekt for sammensettingen av datasett. Stasjonaritet innebærer at variabler har konstant gjennomsnitt, varians og kovarians over tid (Wooldridge, 2012). Problemet med ikke-stasjonære variabler i slike modeller er at det kan oppstå spuriøse⁴ sammenhenger som medfører feilaktige konklusjoner. Dette kommer av at regresjoner med spuriøse sammenhenger ofte genererer høy forklaringskraft og høyt signifikante koeffisienter, som i realiteten er verdiløse (Brooks, 2002). Etablerte fremgangsmåter for å løse problemstillingen er differensiering eller omregning av verdier til vekstform. For de fleste maskinlæringsmodeller er stasjonaritet ikke en forutsetning på lik linje som i tradisjonelle regresjonsmodeller. Modellene som anvendes i denne oppgaven besitter egenskaper som omgår forutsetningen om stasjonaritet ved å minimere eller eliminere variabler som ikke bidrar i prediksjonen. Et eksempel er variabelreduksjon i Elastic Net, som både motvirker multikolaritet og problemer Ordinary Least Squares (OLS) ville hatt ved manglende stasjonaritet. Multikolaritet er problematisk ettersom det medfører skjevhet i variablene bak prediksjonene. Ettersom majoriteten av variablene i vårt datasett er indeksert eller på vekstform, er det mindre sannsynlig at det ville oppstått problemstillinger knyttet til stasjonaritet ved bruk av tradisjonelle modeller.

En viktig problemstilling for prediksjonsformålet var valget av hvilke deler av datagrunnlaget som skulle anvendes i modelleringen. Ettersom våre modeller kun godtar fullstendige tidsserier, innebærer det at variabler uten tilstrekkelig antall observasjoner holdes utenfor. Dimensjonene til datasettet for modelleringen ble funnet gjennom en vektning

⁴En «falsk» korrelasjon mellom variabler som skyldes en bakenforliggende variabel som påvirker begge variablene (Brooks, 2002)

av antall observasjoner opp mot antall variabler. For eksempel ville et datasett med start i 1999 gitt mange variabler, men begrenset antall observasjoner i modelleringen. Alternativt ville et datasett med start i 1985 gitt betraktelig flere observasjoner, men redusert antall variabler. Ulike sammensetninger ble testet og vurdert basert på prediksjonene i treningssettene. Valget falt på perioden fra 1.kvartal 1996 til og med 3.kvartal 2018, hvor presisjonen var høyest.

Det endelige datasettet for våre prediksjoner består av kvartalsvise observasjoner av 54 unike variabler tilbake til 1.kvartal 1996. For de kvartalsvise prediksjonene har det vært formålstjenlig å inkludere ledende variabler, «leads», av de 28 variablene med daglig frekvens⁵. Det kvartalsvise datasettet består derfor av 82 variabler totalt. Høyfrekvente data, som månedlige eller daglige noteringer, er regnet om til et kvartalsvis vektet snitt. 34 av variablene inneholder informasjon om norsk økonomi og de resterende 20 variablene inneholder informasjon om internasjonal økonomi. Hvilke variabler som er representert for hvert land varierer, men BNP-vekst er inkludert for alle. Internasjonale variabler er inkludert ettersom Norge er en liten åpen økonomi og er svært avhengig av handel med andre land. Den økonomiske situasjonen hos Norges viktigste handelspartnere vil derfor være viktig for veksten i norsk økonomi.

Appendiks tabellene A3.1, A3.2 og A3.3 viser en fullstendig oversikt over variablene i det endelige datasettet. Tabellene presenterer variablenes opprinnelse, måleenhet, kilde og frekvens i publiseringer. I tillegg har vi gruppert variablene etter karakteristikker brukt av Aastveit et al. (2011). Blant de norske variablene finner vi seks ulike grupperinger. Gruppen med flest variabler kalles *finanssiell*. I denne inngår 11 forskjellige indekser notert på Oslo børs. Hovedindeksen består av de største og viktigste selskapene i norsk økonomi, mens de resterende er mer sektorspesifikke. Gjennom aksjemarkedets evne til å være fremoverskuende vil inkluderingen av indeksene gi verdifull informasjon om tilstanden i de ulike sektorene i norsk fastlandsøkonomi.

Variabler i gruppen *nasjonal* er hentet fra nasjonalregnskapet til SSB. Nasjonalregnskapet gir viktig informasjon om verdiskapingen i norsk økonomi gjennom ulike beregninger. Vi har plukket ut de vi mener er spesielt representative for aktiviteten i økonomien. *Renter*, *valuta*, *industri*, *eiendom* og *survey* utgjør de resterende gruppene i datasettet. Renter er

⁵Inkluderingen av leads beskrives nærmere i seksjon 4.1.2

avgjørende for vektingen mellom sparing og investering i økonomien, mens valuta er spesielt viktig for norsk eksport. Både renter og valuta påvirker aktiviteten i industrien, som er en mer direkte temperaturmåler på aktiviteten i norsk fastlandsøkonomi. Boligprisene kan være viktige for veksten på ulike måter. Vekst i boligpriser kan være positivt ved å reflektere økte lønninger og høyere etterspørsel etter boliger. Alternativt kan det være et tegn på kunstig høy etterspørsel basert på høye låneopptak. I det første tilfellet bidrar boligmarkedet til økt vekst, mens det i andre tilfellet indikerer bobletendenser i markedet. De internasjonale variablene er gruppert på samme måte som de norske, der fokuset har vært å hente de mest sentrale variablene fra hvert land.

3.2 Data til evaluering av prediksjoner

Som et utgangspunkt for å evaluere maskinlæringens prediksjonsevne har vi samlet inn prediksjoner utarbeidet av Norges Bank, Finansdepartementet, SSB og DNB. Prediksjonene er hentet fra offentlige rapporter som publiseres en eller flere ganger årlig. Hos Norges Bank finner vi både kvartalsvise og årlige prediksjoner, mens de øvrige aktørene kun systematisk publiserer årlige prediksjoner. Gjennomgående har vi fokusert på at de årlige prediksjonene skal være avgitt på samme tidspunkt, slik at informasjonsgrunnlaget til prediksjonene er identisk. Ettersom Finansdepartementet sin publisering av Nasjonalbudsjettet kun forekommer i 3.kvartal er også prediksjonene fra Norges Bank og SSB hentet fra 3.kvartal. DNB sine prediksjoner er hentet fra desember, ettersom vi ønsker å undersøke hvor mye to måneder med ekstra informasjon påvirker presisjonen i prediksjonene.

For en rettferdig sammenligning av de årlige prediksjonene har vi hentet historiske data for den årlige veksten i norsk fastlandsøkonomi. Sucarrat og Gharsallah (2019) har nylig publisert en studie som sammenligner årlige prediksjoner fra Finansdepartementet, SSB og Norges Bank opp mot prediksjoner fra tre ulike statistiske modeller. Vi ønsker å anvende nylig reviderte data for faktisk BNP vekst. Samtidig ønsker vi å bruke et datasett som i størst mulig grad er representativt for prediksjonene som er samlet inn. Det ideelle er derfor å anvende BNP-data som senest ble revidert i 3.kvartal 2019, ettersom Finansdepartementet, SSB og Norges Bank sine prediksjoner er avgitt i dette kvartalet. Ettersom Sucarrat og Gharsallah (2019) har publisert sine data for faktisk BNP-vekst i

Fastlands-Norge, har vi valgt å anvende de samme dataene. Disse er lastet ned 3.september 2019 og har SSB som primærkilde.

På kvartalsbasis sammenligner vi prediksjonene fra maskinlæringen med Norges Bank sine prediksjoner. Sentralbanken publiserer prediksjoner for det påfølgende kvartalet i pengepolitisk rapport, som normalt publiseres siste måned i hvert kvartal. Vi har hentet PPR prediksjoner for firekvartalsvekst fra 1.kvartal 2010 til 3.kvartal 2018. Etersom kun kvartal til kvartal (k/k)⁶ vekst ble publisert for 4.kvartal 2016 og 1.kvartal 2017, mangler vi PPR prediksjoner i disse kvartalene. Prediksjonene til Norges Bank hviler tungt på deres SAM-modell. Resultatene fra SAM er ofte gjort tilgjengelig i den pengepolitiske rapporten. Vi har derfor også hentet SAM sine prediksjoner fra 1.kvartal 2010 til og med 3.kvartal 2016. SAM er en kombinasjon av tre ulike modellklasser: vektor autoregresjon, ledende indikatorer og faktormodeller (Aastveit et al., 2011). For å dra full nytte av de ulike samlingsmodellene, utarbeides prediksjonene gjennom to steg. I første steg genereres et komprimert prediksjonsintervall basert på prediksjonene til de individuelle modellkomponentene i hver modellklasse. I andre steg blir resultatene fra de tre modellklassene tildelt ulike vektorer og slått sammen til ett resultat. SAM kan vektlegge modeller med nyere og mer relevant data i løpet av et kvartal, hvilket gjør den spesielt godt egnet for nowcasting.

For å illustrere det potensielle omfanget av datarevideringer er en eldre og urevidert versjon av faktiske BNP data for Fastlands-Norge hentet inn. Intensjonen er å undersøke om prediksjonene presterer bedre eller dårligere med dette som beregningsgrunnlag. Ureviderte BNP data er hentet fra samfunnsøkonomenes årlige prognosepris (Sucarrat, 2019). Samfunnsøkonomenes prognosepris er en årlig konkurranse der flere etablerte aktører avgir sin prediksjon for BNP-veksten i Fastlands-Norge for det påfølgende året. Når den faktiske veksten er kjent blir prediksjonene vurdert, og en vinner kåret basert på det absolutte avviket. Genarro Succarat har hatt ansvar for prisen siden 2010. Selv om BNP-dataene ikke representerer en fullstendig urevidert versjon, mener vi de er tilstrekkelige til vårt analyseformål når vi skal illustrere effekten av senere revideringer.

En fullstendig oversikt over innhentede prediksjoner, i tillegg til data for revidert og ureviderte BNP-tall, er presentert i tabell A1.1, A1.2 og A1.3 i appendiks.

⁶Eksempelvis veksten fra 1.kvartal til 2.kvartal

3.3 Overfitting

Maskinl ring appellerer til sluttbrukere med en evne til   oppdage komplekse strukturer uten forh ndprogrammerte innstillinger; en egenskap som kommer godt med i en tid med mer tilgjengelig informasjon enn noen gang. I hvilken grad den store tilgangen til informasjon er nyttig, avhenger av formålet ved prediksjonen.

En rekke hedgefond anvender maskinl ring hovedsakelig til trendklassifisering. Her kjennetegnes modellene av h y kompleksitet, i form av sv rt mange inkluderte variabler (Aggarwal, 2017). Renaissance Technologies vil eksempelvis ikke bry seg nevneverdig om sammenhengen mellom S&P500 og iskremprisen i New Dehli er kausal, s  lenge korrelasjonen er tilstede og er rimelig konsistent over en viss tid. For prediksjonsform l er  kt kompleksitet kun ideelt frem til et gitt punkt. Studier viser nemlig at modellenes evne til   skille signal⁷ fra st y⁸ svekkes n r kompleksiteten i datasettet blir for h y (Science, 2019). Fortman-Roe (2012) illustrerer problemstillingen mellom kompleksitet og prediksjonsavvik ved spredningsplottet i figur 3.1. Eksemplet tar for seg tre ulike modelleringer av lykke opp mot velstand, hvor kompleksiteten  ker gjennom inkluderingen av flere h yere-ordens variabler for velstand.

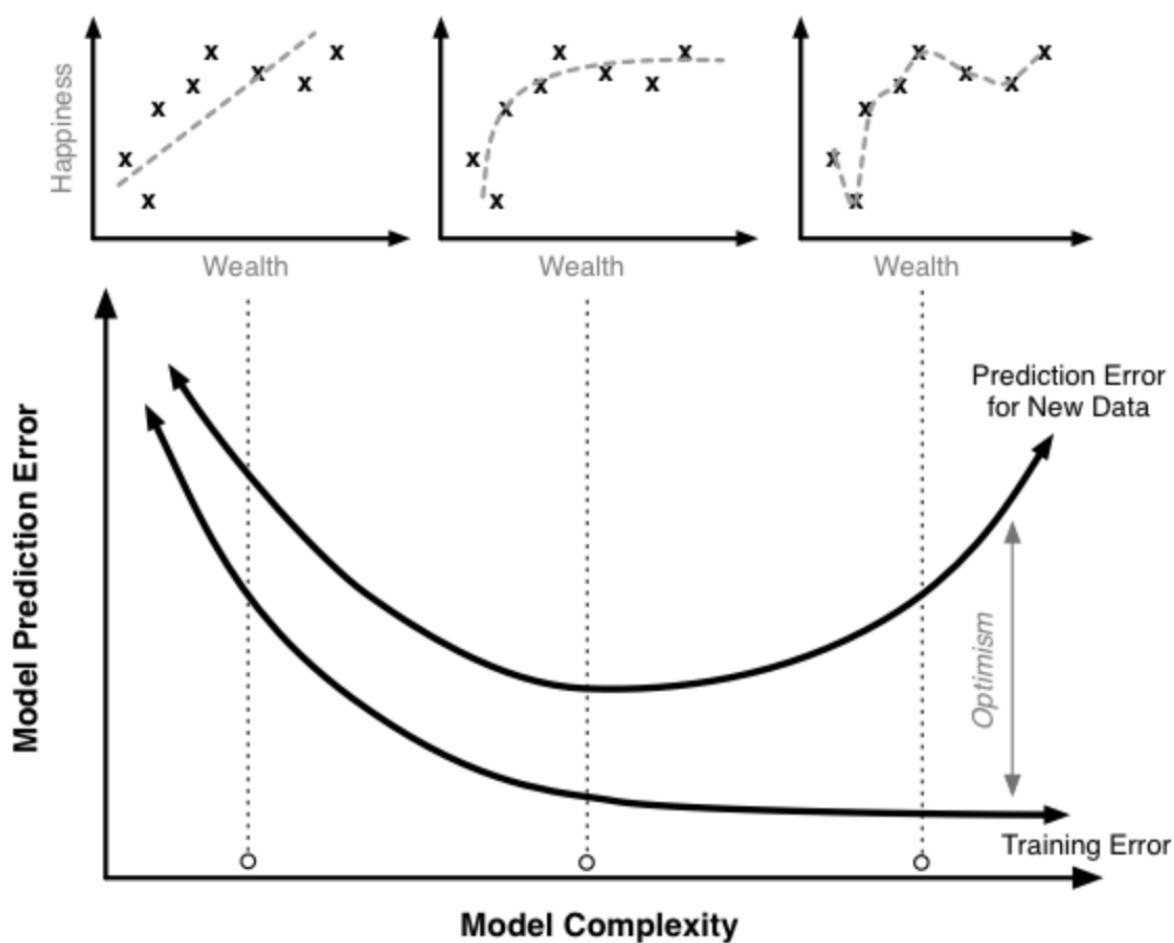
Spredningsplottet viser hvordan vi beveger oss mellom store og sm  prediksjonsavvik i trening- og testsett. Store prediksjonsavvik i b de trening- og testsettet (fra venstre) g r over til lave avvik i trening og store avvik i test (til h yre). Variasjonen skyldes kompleksiteten i variablene, der venstre side har lav kompleksitet og h yre side har h y kompleksitet. Den optimale tilpasningen for enhver modell er punktet der  kt variabelkompleksitet ikke kan redusere prediksjonsavviket i treningen uten at det g r p  bekostning av testsettet⁹ (Fortman-Roe, 2012). En av de st rste utfordringene ved anvendt maskinl ring, overfitting, omfatter ethvert tilfelle med h yere kompleksitet enn dette punktet. Overfitting beskriver en modelltilpasning som i for stor grad er basert p  treningssettene og dermed underpresterer i prediksjonen av testsettene (Mullainathan og Spiess, 2017). Modellen har med andre ord en manglende evne til   generalisere mellom treningsdata og testdata.

⁷Refererer til den underliggende trenden i datasettet

⁸Refererer til irrelevant informasjon eller tilfeldigheter i datasettet

⁹Figur A2.1 gir en alternativ presentasjon av dette

Figure 3.1: Overfitting - resultatet av økt kompleksitet



* Inkluderingen av høyere ordens variable av *velstand* gir bedre tilpasning i treningen, men generaliserer dårlig i prediksjonen av testsettet

For å besitte presise og robuste maskinlæringsmodeller er det avgjørende å unngå overfitting. I seksjon 4.1.3 går vi nærmere inn på et sentralt verktøy for å redusere sannsynligheten for overfitting. Uavhengig av hvor god modell man besitter, er det også nødvendig med en systematisk tilnærming til vektingen av antallet variabler opp mot antallet observasjoner. Fornuftig bruk av økonomisk teori og intuisjon i denne prosessen kan være avgjørende for hvorvidt algoritmene identifiserer strukturelle signaler eller støy først (Mullainathan og Spiess, 2017). En tilnærming der så mange variabler som mulig inkluderes er dermed ikke bærekraftig. I praksis eksisterer det ingen god analytisk måte å finne balansen mellom observasjoner og variabler, men det er mulig å eksperimentere med variabelkompleksitet. I seksjon 3.1 benytter vi kunnskap illustrert i figur A2.1 til å finne en fornuftig balanse.

3.4 Utfordringer med makroøkonomiske data

Over tid er det sannsynlig at mange av de inkluderte variablene i vårt datasett er blitt revidert. Dette er normal praksis for makroøkonomiske variabler og skaper problemer når vi skal sammenligne våre prediksjoner opp mot aktørene omtalt i 3.2. I utarbeidelsen av prediksjoner er det avgjørende at informasjonsgrunnlaget er så presist som mulig. Vårt datasett vil kunne være mer «korrekt» som følge av nyere revideringer, sammenlignet med informasjonen de resterende aktørene hadde tilgjengelig. Dette vil potensielt utgjøre en fordel for våre prediksjoner. Vi har undersøkt mulighetene for å lage et datasett som tar utgangspunkt i første publisering av observasjonene. Derimot har vi ikke hatt tilgang på tilstrekkelige publiseringer til å danne et komplett datasett. Aastveit et al. (2011) løser noe av problemet gjennom sitt arbeid for Norges Bank, med et mer detaljert datagrunnlag enn vi har tilgjengelig. Problemstillingen må derfor anses som en potensiell svakhet ved maskinlæringens prediksjoner.

En annen problemstilling ved makroøkonomiske tidsseriedata omhandler forskjellen i publiseringsdato for hver enkelt variabel. Blant annet publiseres mange «survey» data, som for eksempel PMI og forbrukertillit, i løpet av måneden etter den gjeldende perioden. Data for BNP publiseres to måneder inn i kvartal $T+1$ for kvartal T . Ved prediksjoner av historiske tidspunkter er det en utfordring å ivareta det faktiske publiseringstidspunktet til alle input variabler, noe som kan forårsake *Look-ahead bias*. Look-ahead bias innebærer bruk av informasjon som ikke er tilgjengelig på prediksjonstidspunktet (Walimbe, 2017). For eksempel ved prediksjon av 1.kvartal 2017 er det avgjørende at det ikke benyttes informasjon om industriproduksjon og arbeidsledighet fra desember 2016. Dette fordi informasjonen først publiseres i løpet av januar 2017. Uten eventuelle tilpasninger vil våre modeller inkludere informasjon fra desember 2016 som i praksis ikke er tilgjengelig. Prediksjoner med look-ahead bias vil kunne gi misvisende resultater og urealistisk gode estimater. Vi har derfor tilpasset fremgangsmåten slik at vi unngår look-ahead bias.

I vår analyse innebærer look-ahead bias også et likt informasjonsgrunnlag for alle aktører. Vi har derfor vært konsekvente i innhenting av prediksjoner fra de ulike aktørene. Som beskrevet i seksjon `refsubsec:InnhentingEksternePrediksjoner` har vi systematisk hentet prediksjoner som er avgitt på tilnærmet samme tidspunkt. Tilpasningene vi har gjort for å unngå look-ahead bias er beskrevet i seksjon 4.1.2.1.

4 Metode

4.1 Prediksjon med maskinlæring

Innenfor klassisk statistisk metode hviler det tunge forutsetninger om datagrunnlagets distribusjon og forholdet mellom avhengig og uavhengig variabel. På et overordnet nivå eksisterer to «fagkulturer» (Breiman, 2001b). Den første antar en spesifikk stokastisk datamodell med et randomisert utvalg av uavhengige variabler. Den andre forsøker å lage en funksjon som best forklarer den avhengige variabelen basert på utvalgte uavhengige variabler. Maskinlæring tilhører den sistnevnte kulturen, innenfor en kategori som kalles algoritrisk modellering. For prediksjonsformål er det mindre fokus på forholdet mellom den avhengige variabelen og de uavhengige variablene, ettersom målet er å skape en funksjon som best predikerer den avhengige variabelen.

Maskinlæring består hovedsakelig av to elementer: En *læringsprosess* som bestemmer den beste tilpasningen for de uavhengige variablene, og en *algoritme* som basert på læringen modellerer sammenhengen mellom avhengig og uavhengige variabler (Jung et al., 2018).

4.1.1 Overvåket og uovervåket læring

Læringsprosessen kan deles inn i to ulike former; overvåket og uovervåket læring (Hastie et al., 2004). Overvåket læring handler om problemstillinger der det er liten tvil om sammenhengen mellom variablene og hva slags type output som forventes fra modellen. Utfordringen er å produsere den korrekte kvantifiseringen av den avhengige variabelen, og effekten fra de ulike uavhengige variablene. Uovervåket læring handler i større grad om en udefinert problemstilling, der man samler et stort sett av data for å avdekke sammenhenger. På denne måten kan de uavhengige variablene kategoriseres basert på sammenhengene algoritmen finner.

4.1.2 Trenings- og testdata

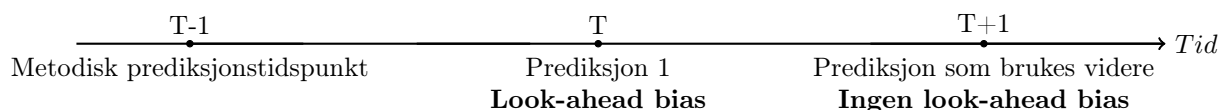
Læring av modeller til prediksjonsformål fordrer noen helt nødvendige forberedelser for å sikre et troverdig utfall. I første omgang innebærer dette et tydelig skille mellom data som benyttes til trening og data som brukes til å evaluere prediksjonsevnen. Dersom modellen

testes p  data som allerede har blitt brukt til   tilpasse modellen, vil det forhindre et troverdig m l p  prediksjonsevnen.

Et treningssett beskriver delen av det opprinnelige datasettet som legger grunnlaget for treningen av modellens parametere. Treningssettet utgj r normalt en betydelig andel av datasettet, men det nøyaktige omfanget vil avhenge av de modellspesifikke egenskapene. Testsettet utgj r den gjenv rende delen av det opprinnelige datasettet. Denne delen er utgangspunktet for en troverdig evaluering av modellens evne til   predikere p  nye data. Testsettet kan variere i st rrelse, men b r v re minst like stor som den lengste prediksjonshorizonten (Hyndman og Koehler, 2006). I litteraturen refereres trenings- og testsett ogs  som henholdsvis «in-sample» og «out-of-sample». Inndelingen av datasettet kan gj res p  flere m ter, men det er avgj rende   skille fremgangsm ten for tverrsnitts- og tidsseriedata. Randomisert inndeling av trenings- og testsett som anvendes for tverrsnittsdata egner seg ikke tidsseriedata. V r fremgangsm te blir gjennomg tt i seksjon 4.1.4.

4.1.2.1 Fremgangsm te for   unng  look-ahead bias

For   l se problemstillingen tilknyttet look-ahead bias omtalt i seksjon 3.4, m  vi gj re flere tilpasninger i v r fremgangsm te. For *kvartalsvis* vekst konstruerer vi konsekvent prediksjoner for kvartal T og kvartal T+1 samtidig. Prediksjonen for kvartal T inneholder look-ahead bias, gjennom bruk av informasjon om kvartal T-1 som f rst blir tilgjengelig i l pet av kvartal T. Prediksjonen for kvartal T+1 unng r look-ahead bias ettersom informasjonen fra kvartal T ikke anvendes. Vi trekker konsekvent ut prediksjonen for kvartal T+1 i v r analyse, frem til vi har prediksjoner for alle kvartaler i perioden. Fremgangsm ten er illustrert i figur 4.1

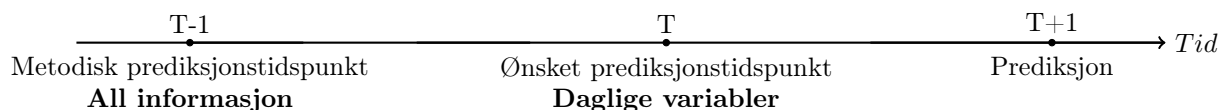


Figur 4.1: Fremgangsm te for utarbeidelse av kvartalsvise prediksjoner

I praksis vil tiln rmingen inneb re at v re prediksjoner utelater all informasjon fra kvartal T i prediksjonen av kvartal T+1. En viktig implikasjon fra dette vil v re at data fra de to f rste m nedene i kvartal T ikke inkluderes. Denne informasjonen er tilgjengelig p 

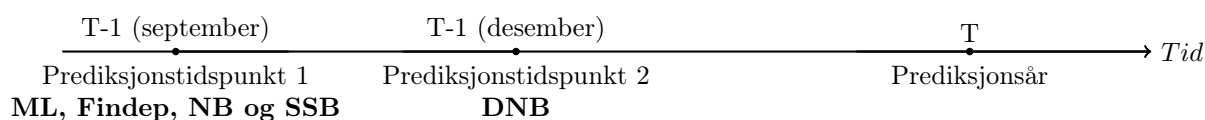
prediksjonstidspunktet til Norges Bank, som publiserer sine prediksjoner i siste m ned av kvartalet. V r forsiktige tiln rming inneb rer at Norges Bank sine prediksjoner konsekvent besitter mer informasjon fra kvartal T for alle m nedlige variabler. I praksis vil dette eksempelvis medf re at v r prediksjon av veksten i 1.kvartal 2017 inkluderer industriproduksjonen til og med 3.kvartal 2016, mens Norges Banks prediksjoner i tillegg bruker informasjon fra to av tre m neder i 4.kvartal 2016.

For   best replikere de reelle omstendighetene ved prediksjon ett kvartal frem i tid, gjennomf rer vi nye tilpasninger i metoden. Informasjonsgrunnlaget for prediksjonene fra maskinl ringen er illustrert i figur 4.2. All informasjon om variabler med daglig frekvens er inkludert for prediksjon av kvartal T+1. Informasjonen inkluderes ved hjelp av leads for alle daglige variabler i datasettet, som nevnt i seksjon 3.1. Leads gir informasjon om variablenes verdi det neste kvartalet. I praksis vil dette inneb re at de daglige dataene fra kvartal T inkluderes i informasjonsgrunnlaget n r prediksjonene for kvartal T+1 lages. Basert p  Norges Banks prediksjonstidspunkt inneb rer dette at h yfrekvent data fra den siste m neden, som ikke var tilgjengelig for deres prediksjoner, blir brukt i v re prediksjoner. Likevel mener vi ikke det medf rer ufordelaktige forhold for den relative sammenligningen, ettersom Norges Bank har store fordeler gjennom de m nedlige variablene.



Figur 4.2: Informasjonsgrunnlag for prediksjon av kvartal T+1

En lignende tankegang benyttes for   unng  look-ahead bias i de  rlige prediksjonene. Vi anvender firekvartalsveksten for 4.kvartal i  r T som en tiln rming for den  rlige veksten i  r T. I praksis reflekterer prediksjonen utviklingen fra 4.kvartal  r T-1 til og med 4.kvartal  r T. Prediksjonstidspunktet settes til siste dag i 3.kvartal  r T-1. Prediksjonshorizonten endres til fem kvartaler, slik at vi unng r look-ahead-bias. Gitt v r tiln rming er det kun prediksjonene av 4.kvartal vi trekker ut og anvender som  rlige prediksjoner. Informasjonsgrunnlaget for prediksjon av  r T er derfor all informasjon til og med 3.kvartal  r T-1. De ledende variablene er ikke inkludert, ettersom akt rene vi  nsker   sammenligne oss med har publisert sine prediksjoner i 3.kvartal  r T-1. Inkludering av leads ville derfor bidratt til look-ahead bias, som vi  nsker   unng .



Figur 4.3:  rlige prediksjonstidspunkt for ulike akt rer

Tiln rmingen fungerer godt til   skape et rettferdig sammenligningsgrunnlag opp mot prediksjonene fra SSB, Finansdepartementet og Norges Bank. Den eneste forskjellen i informasjonsgrunnlaget kommer fra publiseringstidspunktet til den enkelte akt r i 3.kvartal. Siden v re prediksjoner inneholder all informasjon fra 3.kvartal i  r T-1, vil de inneholde daglig informasjon som de andre akt rene ikke har hatt tilgjengelig. Normalt publiserer derimot akt rene sine prediksjoner mot slutten av kvartalet. Vi vurderer derfor informasjonsasymmetrien til   v re liten. Etter v rt skj nn vil dette v re en prosess som replikerer prediksjonssituasjonen til de andre akt rene p  en tilfredsstillende m te.

4.1.3 Kryssvalidering

Alle maskinl ringsmodeller gjennomf rer omfattende «tuning» for   tilpasse modellen, slik at presisjonen i out-of-sample prediksjonene optimaliseres. Fremgangsm ten kalles kryssvalidering og inneb rer at ulike modellparametere justeres optimalt. P  denne m ten motarbeides overfitting, et problem som medf rer redusert presisjon i prediksjonene. Ved overfitting vil man produsere d rlige out-of-sample prediksjoner, grunnet for stor vektlegging av forhold i in-sample. V rt datasett er strukket over flere ti r for   oppn  et akseptabelt antall observasjoner. For   lage presise prediksjoner er derfor kryssvalidering avgj rende for at modellene skal fange opp relevante karakteristikker fra ett ti r som ogs  er viktig i det neste.

Kryssvalideringen gjennomf res ved   sette en startverdi for tuning-parametere i hver enkelt modell. For eksempel vil det gjelde hvor mange tr er som skal danne det ferdige beslutningstreet i Random Forest, eller vektningen av λ og α i restleddet til Elastic Net. I prosessen deles treningssettet i K forskjellige grupper. En av gruppene legges vekk og betegnes som et valideringssett. De resterende gruppene utgj r treningssettene. Innenfor treningssettene tunes og valideres modellen gjennom testing av ulike sammensetninger av parameterne. Deretter velges den beste tilpasningen til   lage prediksjoner av testsettet. Eksempelvis vil ulike verdier av λ og α bli anvendt i kryssvalideringen av Elastic Net. Hver

kombinasjon produserer egne in-sample prediksjoner som vurderes opp mot hverandre, der fokuset ligger i   minimere prediksjonsavvik. Resultatet av kryssvalideringen inneholder kombinasjonen av λ og α med best forutsetninger til   predikere i out-of-sample.

4.1.4 Rolling Forecasting Samples

Tidsdimensjonen i v rt datasett gir rekkef lgen til observasjonene en avgj rende betydning for den underliggende inndelingen av datasettet. En potensiell fremgangsm te, som ivaretar dette hensynet, ville v rt   bevilge de eldste 80% av observasjonene til trening og de gjenv rende 20% til test. Hyndman og Koehler (2006) presenterer *Rolling Forecasting Origin* som en mer sofistikert teknikk for kryssvalidering og inndeling av tidsseriedata. Datasettet deles inn i tidsserier best ende av individuelle testsett og tilh rende treningssett som kun inneholder observasjoner som har forekommet f r observasjonene i de respektive testsettene. Med denne prosedyren blir eldre test- og treningssett inkludert i nyere treningssett. Eksempelvis vil alle observasjoner til og med 2010 brukes i treningen for prediksjonen av  ret 2011. Ved neste iterasjon vil treningssettet v re utvidet til 2011, for prediksjonen av  ret 2012. Helt overordnet inneb rer dette en forflytning av trening- og testsett over tid.

Bakgrunnen for at vi benytter rullerende trening og validering henger sammen med viktigheten av tidsdimensjonen i v rt prediksjonsform l. Ved prediksjoner av  konomisk vekst er det n rliggende at den seneste informasjonen om tilstanden i  konomien p virker modellens presisjonsniv . Tidsdimensjonen gj r at observasjoner i periode $T-1$, T og $T+1$ vil v re tett korrelert med hverandre. Dette skiller seg fra for eksempel prediksjon av tverrsnittsdata som ikke inneholder en tidsdimensjon. Her vil en ekstra observasjon i mindre grad kunne utgj re en forskjell. Viktigheten av nyere informasjon er analysert n rmere i seksjon 6.1. Et rullerende tidsvindu gjenspeiler de reelle omstendighetene med ny informasjon som blir tilgjengelig fra en periode til den neste.

4.1.4.1 Caret

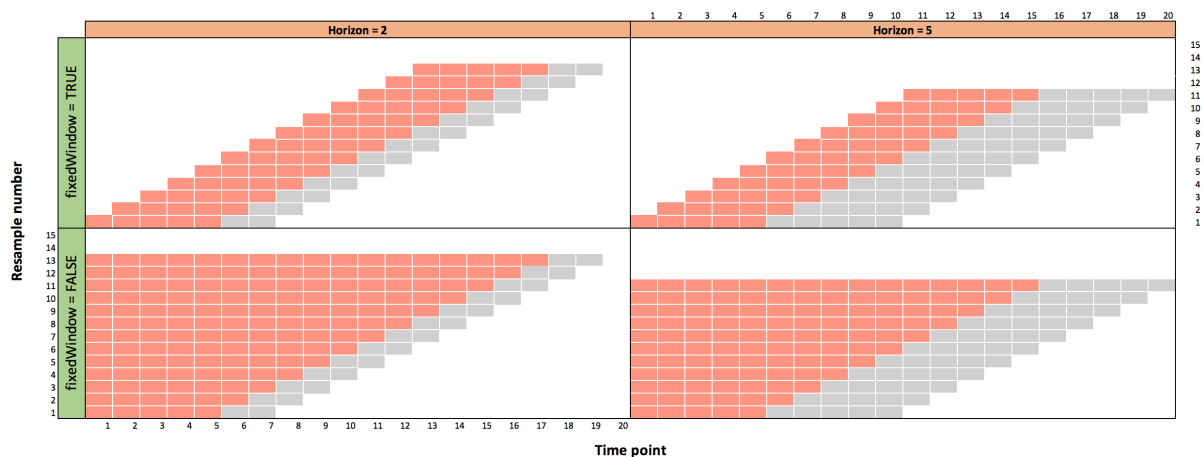
For den tekniske gjennomf ringen av kryssvalidering, inndeling av datasett og prediksjon i tr d med Hyndman og Koehler (2006) sin fremgangsm te, anvender vi *Caret* i det statistiske programmeringsspr ket R. *Caret*¹⁰ er en pakke best ende av en rekke funksjoner som effektiviserer prosessen med   bygge prediksjonsmodeller (Kuhn, 2008). Her finnes det funksjoner for datasplitting, feature selection og modell tuning. Dette gjelder b de klassifisering- og prediksjonsform l i tversnitts- eller tidsseriedata. Totalt inneholder *Caret* 238 ulike modeller, noe som gir oss fleksibilitet i valget av hvilke modeller vi  nsker   benytte. Pakken tillater oss   anvende modeller med ulik kompleksitet p  en forholdsvis enkel og intuitiv m te, der vi slipper   gj re den mest avanserte kodingen. Vi vil n  forklare de mest sentrale funksjonene i v r fremgangsm te.

I funksjonen «*Traincontrol*», gjennom metoden «*Timeslice*», oppn r vi  nsket inndeling av trening- og testsett p  en enkel og intuitiv m te. *Timeslice* inneholder tre parametere som krever spesifisering i forkant av datainndelingen: *initial window*, *horizon* og *fixed window*. I *Initial Window* defineres hvor mange observasjoner som skal inng  i det initielle treningssettet. For eksempel vil dette tilsvare 56 kvartaler i analysen der f rste prediksjon er 1.kvartal 2010. *Horizon* definerer prediksjonshorisonten, det antallet prediksjoner som gjennomf res i hvert testsett. Som vi har v rt inne p  i seksjon 4.1.2.1 bruker vi en horisont p  to kvartaler ved kvartalsvise prediksjoner og en horisont p  fem kvartaler for  rlige prediksjoner. Avslutningsvis spesifiseres *Fixed Window*, som avgj r hvorvidt st rrelsen p  alle treningssett holdes konstant¹¹ eller utvides over tid.

Figur 4.4 gir en god oversikt over mulighetsområdet ved ulike spesifikasjoner av de nevnte parameterne i *timeslice*.

¹⁰Classification And REgression Training

¹¹Gitt initial window

Figur 4.4: Rolling Forecasting Origin - Inndeling av tidsseriedata i Caret

* Oversikt over ulike innstillinger for parametere i timeslice. For kvartalsvise prediksjoner er innstillingene representert i nedre venstre halvdel. For årlige prediksjoner er innstillingene i nedre høyre halvdel. Røde ruter = treningsett, Grå ruter = testsett

I analysen settes fixed window til false, hvilket innebærer at treningssettene inkluderer samtlige historiske observasjoner frem til det respektive prediksjonstidspunktet. Dette kan bidra til mer støy i treningen, men vi vurderer det økte antallet observasjoner til å være mer avgjørende for våre modeller. For kvartalsvise prediksjoner setter vi initial window slik at første prediksjon forekommer i 1.kvartal 2010 i den generelle analysen og 1.kvartal 2007 i analysen av robusthet. For de årlige prediksjonene settes initial window slik at første prediksjonsår er 2010. Horisonten endres mellom årlige og kvartalsvise prediksjoner. Nedre venstre halvdel av figur 4.4 viser den kvartalsbaserte tilnærmingen, mens nedre høyre halvdel viser den årlige.

4.2 Elastic Net

Elastic Net er en maskinlæringsmodell som opprinnelig ble introdusert i 2005 (Hastie og Zou, 2005). Modellen er en kombinasjon av Ridge og Lasso regresjon. Disse bygger på den tradisjonelle OLS regresjonen, der målet er å estimere koeffisienten $\hat{\beta}$ som best representerer den sanne parameteren β . OLS gjør dette gjennom å finne koeffisientene som minimerer summen av de kvadrerte residualene (Wooldridge, 2012). Et residual er et mål på avviket som oppstår mellom prediksjon (\hat{y}_n) og observert verdi (y_n). De mest vanlige feilkildene ved OLS regresjoner omhandler for stor varians eller skjevhet i de estimerte koeffisientene. For stor varians kan forekomme dersom regresjonen inneholder for mange forklaringsvariabler eller i tilfeller der forklaringsvariablene er sterkt korrelerte. Elastic

Net forsøker å løse problemet gjennom å kombinere aspekter ved Ridge og Lasso, i form av dimensjonsreduksjon og selektering av variabler. I praksis betyr dette at det inkluderes et restledd bestående av to ulike straffetermer som vektet forskjellig. For å forstå de ulike komponentene av Elastic Net er det avgjørende å forstå Ridge og Lasso regresjoner hver for seg.

4.2.1 Ridge regresjon

Ridge regresjon forsøker å redusere variansen ved å introdusere en straffeterm som minimeres når de estimerte koeffisientene i regresjonen nærmer seg null. Regresjonen inneholder fortsatt det tradisjonelle OLS optimeringsproblemet tilknyttet minimering av summen til de kvadrerte residualene. Ridge løser optimeringsproblemet ved å redusere koeffisientene til variablene som er mest korrelert. Dette medfører at koeffisientene som forklarer den samme variansen får en lavere koeffisient enn de ellers ville hatt. På denne måten evner Ridge regresjonen å redusere variansen samtidig som den beholder alle variabler. Optimeringsproblemet ser slik ut:

$$\hat{\beta} = \arg \min_{\hat{\beta}_j} \left[\underbrace{\left(\sum_{i=1}^n (Y - X\hat{\beta})^2 \right)}_{\text{RSS}} + \lambda \underbrace{\sum_{j=1}^p (\hat{\beta}_j)^2}_{\text{Ridge penalty}} \right] \quad (4.1)$$

Y representerer den faktiske verdien til den avhengige variabelen, mens X er verdien til de ulike uavhengige variablene. n er antall observasjoner og p er antall forklaringsvariabler. Det første leddet representerer det tradisjonelle OLS-problemet, mens straffetermen er representert i andre ledd. Parameteren λ bestemmer hvor stor vekt straffetermen får, og fastsettes gjennom kryssvalidering. Høyere λ vil medføre en større betydning for straffetermen, mens en lavere λ vil bidra til et resultat som ligner en tradisjonell OLS regresjon. I praksis vil en høyere λ derfor innebære at de mest korrelerte variablene får redusert sin koeffisient i den endelige modellen, slik at variansen reduseres. Et godt eksempel kan være koeffisientene til styringsrenten og inflasjonsnivået. Etersom de fleste sentralbanker fører en form for inflasjonsstyring vil korrelasjonen mellom variablene være høy. Ridge straffetermen reduserer derfor koeffisientene til de to variablene, for å forhindre økt varians.

4.2.2 Lasso regresjon

Lasso regresjon ble først introdusert av Tibshirani i 1996. Metoden fungerer på samme måte som Ridge regresjon gjennom å redusere koeffisientene som er mest korrelerte, i tillegg til å beholde OLS funksjonen. Lasso skiller seg fra Ridge gjennom muligheten til å fullstendig utelukke forklaringsvariabler. Straffetermen baserer seg på summen av de absolutte koeffisientene, ikke de kvadrerte koeffisientene som i Ridge. I praksis betyr dette at en høy λ medfører at enkelte forklaringsvariabler kan bli helt utelukket fra regresjonen, fremfor å bli redusert som i Ridge. Optimeringsproblemet ser slik ut:

$$\hat{\beta} = \underset{\hat{\beta}_j}{\operatorname{arg\,min}} \left[\underbrace{\left(\sum_{i=1}^n (Y - X\hat{\beta})^2 \right)}_{\text{RSS}} + \lambda \underbrace{\sum_{j=1}^p |\hat{\beta}_j|}_{\text{LASSO penalty}} \right] \quad (4.2)$$

Første del av ligningen inneholder OLS problemet, mens siste del er Lasso straffetermen. λ spiller igjen en viktig rolle for hvor avgjørende straffetermen blir, der en høyere λ åpner opp for fullstendig utelukkelse av variablene som er sterkest korrelert. Såkalt variabelseleksjon blir mer gjeldende. En mye omtalt kritikk mot Lasso er at modellen ikke tar hensyn til hvilken av de mest korrelerte variablene som blir kastet ut, noe som kan medføre at viktige forklaringsfaktorer blir utelatt¹². Selv om to makroøkonomiske variabler korrelerer, er det fullt mulig at de har egne direkte effekter på den avhengige variabelen. Eksempelvis kan den norske valutakursen mot Euro bli kastet ut som følge av sterk korrelasjon med den norske styringsrenten. Et slikt scenario kan påvirke prediksjonene av norsk BNP-vekst. Valutakursen har en selvstendig effekt som er frikoblet fra den norske styringsrenten. Dette vil ikke reflekteres i modellen hvis en av variablene utelates.

4.2.3 Elastic Net

Elastic Net er en sammensetning av en OLS regresjon og straffetermene fra Ridge og Lasso. Dermed kan modellen anvende dimensjonsreduksjon gjennom Ridge, i tillegg til variabel seleksjon gjennom Lasso. På denne måten blir Elastic Net en fleksibel metode som reduserer betydningen av sterkt korrelerte variabler, og i ytterste konsekvens fjerner de fra den endelige modellen. Optimeringsproblemet blir som følger:

¹²Omitted variable bias (Wooldridge, 2012)

$$\hat{\beta} = \arg \min_{\hat{\beta}_j} \left[\underbrace{\left(\sum_{i=1}^n (Y - X\hat{\beta})^2 \right)}_{\text{RSS}} + \lambda \sum_{j=1}^p \left[\underbrace{(1 - \alpha)(\hat{\beta}_j)^2}_{\text{Ridge}} + \underbrace{\alpha|\hat{\beta}_j|}_{\text{Lasso}} \right] \right] \quad (4.3)$$

Igjen er n antall observasjoner og p antall forklaringsvariabler. Ridge og Lasso straffetermene utgjør et felles restledd. λ definerer størrelsen på restleddet, slik at man får vektet betydningen til de to straffetermene. Lavere λ gjør modellen mer lik OLS. Parameteren α introduseres som en relativ vektning mellom de to straffetermene. Dersom α nærmer seg én vil Lasso straffetermen vektlegges mest. Hvis α er lav, nærmere null, vil Ridge straffetermen være mest avgjørende. α vil med andre ord avgjøre om det er dimensjonsreduksjon eller variabel seleksjon som bidrar til å redusere variansen i modellen. Avgjørelsen av α og λ bestemmes gjennom kryssvalidering.

Hastie og Zou (2005) ser på restleddet i Elastic Net som ideell i tilfeller der antallet forklaringsvariabler overgår antallet observasjoner. I slike tilfeller vil en ren Lasso modell aldri velge et antall forklaringsvariabler i den endelige modellen som overgår antallet observasjoner. Dette innebærer at variabler som potensielt kunne påvirket prediksjonene blir utelatt, kun på bakgrunn av få observasjoner. Ved å inkludere Ridge straffetermen har man også muligheten til å inkludere alle variabler, og heller gi de som korrelerer sterkt en lavere koeffisient. Videre er Elastic Net en ideell metode i tilfeller der enkelte forklaringsvariabler viser parvis sterk korrelasjon. Dette er et aktuelt aspekt for vårt datasett, der mange av variablene kan tenkes å påvirke hverandre indirekte. I en slik situasjon vil en tradisjonell Lasso utelate en av de korrelerte variablene. Hvilken variabel som utelates er derimot svært tilfeldig, og gjør prosessen lite effektiv med tanke på å lage de beste prediksjonene. Elastic Net unngår problemstillingen ved å inkludere Ridge i restleddet, som tillater at begge variabler beholdes.

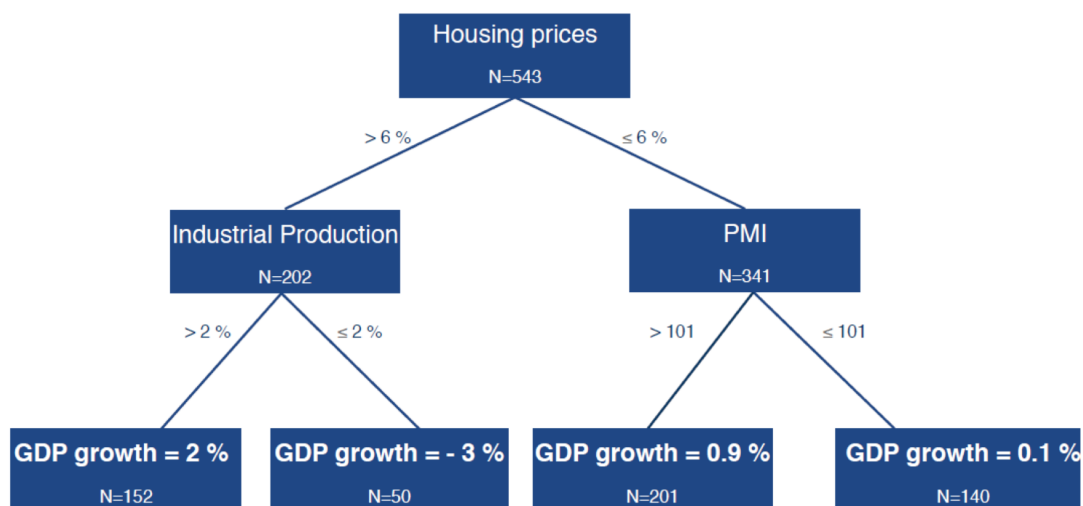
På et overordnet nivå er den største styrken til Elastic Net muligheten til å inkludere både variabel seleksjon og dimensjonsreduksjon i modellen. Variabler med likhetstrekk kan inkluderes i modellen uten problemer. Dette gjør Elastic Net motstandsdyktig mot potensielle problemer tilknyttet multikolinearitet mellom forklaringsvariablene. Dette er en viktig egenskap for prediksjon av økonomisk vekst, der det er avgjørende å kunne inkludere alle sentrale aspekter i økonomien.

4.3 Random Forest

Random Forest er en anerkjent maskinlæringsmodell som ble introdusert i 2001 (Breiman, 2001a). Metoden er effektiv og mye anvendt i litteraturen, ettersom den like godt håndterer problemstillinger tilknyttet klassifisering som regresjoner. Metoden deler datasettet inn i flere sett og gjennomfører regresjoner på hver av disse. En kombinasjon av prediksjoner fra de ulike regresjonene vil være bedre enn en enkeltstående prediksjon basert på det originale datasettet.

Et beslutningstre vokser basert på k regresjoner som analyserer sammenhengen mellom ulike variabler fra mindre utvalg av datasettet (Efron og Tibshirani, 1993). Sentralt i hvert beslutningstre står *bagging*, også kalt *bootstrap aggregating* (Hyndman og Athanasopoulos, 2018). I et standard beslutningstre begrenses antallet grener dersom kompleksiteten i modellen blir for stor. Bootstrap aggregating løser kompleksitetsproblemet ved å utvide beslutningstreeet med et stort antall regresjoner. Sammen danner de et stort beslutningstre der hver gren kalles en node. Innenfor hver node gjennomføres de ulike regresjonene på hvert datasett. På denne måten senker Random Forest variansen i den endelige modellen. Figur 4.5 viser et eksempel på hvordan et beslutningstre med regresjoner kan se ut (Woloszko, 2018).

Figur 4.5: Eksempel på et beslutningstre med regresjoner



* Med bakgrunn i det utvalgte datasettet blir observasjonene splittet basert på om boligprisene i perioden øker med mer enn 6% eller er mindre eller lik 6%. Blant observasjonene med boligprisvekst over 6% velger modellen industriproduksjon blant alle variablene til å videre splitte observasjonene. Bakgrunnen er at denne variabelen gir den beste splitten. For observasjonene hvor boligprisene var 6% eller lavere, velges PMI for videre splitt. De endelige nodene er homogene med tanke på BNP-vekst.

I figur 4.5 kan vi se hvordan prediksjon av BNP-vekst påvirkes av ulike variabler i et beslutningstre. Dersom boligprisene stiger mer enn 6%, samtidig som industriproduksjonen synker med mer enn 2%, ender prediksjonen fra noden på -3%. Den lave veksten i industriproduksjon avdekker at de økte boligprisene kan skyldes en boligboble, som kan føre til negativ vekst. Den endelige prediksjonen lages ved et vektet snitt av prediksjonene fra alle nodene.

4.3.1 Bootstrap aggregating

Bagging, eller bootstrap aggregating, er en metode som står sentralt i Random Forest. Intuisjonen bak konseptet er relativt enkel; ved å gjøre randomiserte utvalg fra datasettet, for deretter å kjøre regresjoner på hvert utvalg, kan en kombinere resultatene til å lage en representativ prediksjon (Breiman, 1994). Utvalgene trekkes fra treningssettet basert på randomisering, slik at det er helt tilfeldig hvilke variabler som blir valgt. De ulike utvalgene danner grenene i beslutningstreet, der regresjoner blir gjennomført. Den mest egnede variabelen bestemmes for å avgjøre splitten i det videre beslutningstreet. Når analysene er gjennomført blir utvalget ført tilbake til det originale datasettet, før nye randomiserte utvalg trekkes. I løpet av prosessen vil man ofte få mange hundre ulike noder. Basert på prediksjonene fra de ulike nodene blir det regnet ut et vektet gjennomsnitt som utgjør den endelige prediksjonen fra modellen.

En negativ side ved bootstrap aggregating er at det ikke foreligger føringer for hvilke karakteristikkene som skal være gjeldende for de ulike utvalgene. Dermed risikerer man at viktige variabler forekommer i et flertall av nodene. For eksempel vil rentenivå og valutakurs være korrelerte, samtidig som de har individuell forklaringskraft på BNP-veksten. Dette kan ofte resultere i at nodene i beslutningstreet blir høyt korrelerte; dersom variablene med stor forklaringskraft korrelerer. En ren bootstrap aggregating metode vil derfor medføre overfitting.

4.3.2 Random Forest

På samme måte som ved bootstrap aggregating danner Random Forest ulike noder ved hjelp av et randomisert utvalg med tilbakelegging fra det originale datasettet. Random Forest skiller seg fra bootstrap aggregating ved å legge til et randomisert valg internt i

hver node i valget av variabelen som skal bestemme splitten videre i beslutningstreet. Variabler som er sterkt korrelert og potensielt forklarer noe av den samme variasjonen, vil derfor kun forekomme i et fåtall noder. Random Forest oppnår dermed et endelig beslutningstre der hver node i større grad representerer ulike effekter på den økonomiske veksten. Fremgangsmåten reduserer korrelasjonen mellom nodene i den endelige modellen. Det vektete snittet av prediksjonene fra hver node representerer en samlet prediksjon som inneholder flere av karakteristikkene ved datasettet. Random Forest tilpasser seg til dataene på en god måte, der metoden tar hensyn til ikke-lineære effekter, i tillegg til å unngå overfitting (Ishwaran, 2007). Dette senker variansen i modellen og øker presisjonen i de endelige prognosene.

Til tross for at Random Forest både er intuitiv og anerkjent, er det viktig å poengtere at flere prosesser i modellen er vanskelige å analysere i detalj. Skapelsen av de ulike nodene og splittelsen videre i nye noder er matematisk lite dokumentert innenfor fagmiljøet (Breiman (2001a), Lin og Jeon (2006), Biau et al. (2008a), Biau et al. (2008a), Biau et al. (2008b)). Dette medfører mindre innsikt i hva algoritmen faktisk gjør, hvilket gjør det vanskelig å underbygge hvorfor Random Forest fungerer godt eller dårlig for et prediksjonsformål.

Random forest er likevel en anvendelig metode som egner seg godt til flere ulike prediksjonsformål, der modellens evne til å håndtere ikke-lineære sammenhenger er spesielt gunstig. Samtidig er det positivt at metoden inneholder mekanismer som forhindrer overfitting og stor varians i den endelige modellen. Dette er relevant for vårt prediksjonsformål ettersom våre forklaringsvariabler sannsynligvis korrelerer med hverandre.

4.4 Super Learner

Den siste modellen i analysen er Super Learner, introdusert av Van der Laan et al. (2007). Denne er tilgjengelig i R-pakken SuperLearner, og er den eneste modellen vi benytter utenfor Caret. Rent teknisk er kodingen mer komplisert i R, men den metodiske fremgangsmåten for prediksjonene er lik for alle modeller¹³. Super Learner er en samlingsmodell som benytter en algoritme til å vekte egenskapene til en rekke maskinlæringsmodeller. Hvilke modeller den endelige modellen består av avgjøres gjennom

¹³Seksjon 4.1.2.1

kryssvalidering. Basert på det aktuelle prediksjonsproblemet får hver modell sin optimale vektning (Dietterich, 2000). Bakgrunnen for inkluderingen av Super Learner ligger i svært lovende resultater i lignende studier. Jung et al. (2018) presenterer Super Learner som den beste maskinlæringsmodellen i prediksjonen av økonomisk vekst i fem svært forskjellige økonomier.

Hvilke modeller som brukes som input i Super Learner spesifiseres manuelt. Vi har valgt ut 10 ulike modeller, listet opp i tabell 4.1. Jung et al. (2018) argumenterer for at dette utvalget av modeller egner seg best for prediksjon av økonomisk vekst. Dermed får vi testet et bredt spekter av modeller, fra nevralt nettverk til vektormaskiner.

Tabell 4.1: Utvalgte algoritmer for Super Learner

Algoritme	Beskrivelse
glmnet	Elastic Net
gam	Generalized additive models
nnet	Neural Network
svm	Support vector machine
bayesglm	Bayesian generalized linear model
polymars	Polynomial spline regression
rpart	Recursive partitioning
mean	Arithmetic mean
randomForest	Random Forest

Vektingen av algoritmene i den endelige modellen gjøres ved hjelp av kryssvalidering. Kryssvalideringen minimerer residualene, ved testing av utallige kombinasjoner av algoritmene. Den endelige modellen inneholder en vektning av hver algoritme som gir in-sample prediksjoner med det laveste avviket.

Super Learner har en tilnærming som løser kritikken mot svært mange modeller; at de ikke er tilstrekkelig tilpasset det aktuelle prediksjonsproblemet. Med Super Learner anvendes flere algoritmer, optimalt tilpasset den aktuelle problemstillingen. Kryssvalideringen reduserer sannsynligheten for overfitting, slik at inkluderingen av flere algoritmer blir uproblematisk. Van der Laan og Polley (2010) finner modellen som spesielt robust i de tilfellene der datasettet er lite, noe som gjør den svært aktuell for vårt prediksjonsformål.

4.5 Måltall for vurdering av presisjon

For å kunne vurdere presisjonen til prediksjonene fra maskinlæringen, Norges Bank, Finansdepartementet, SSB og DNB er vi avhengig av pålitelige beregninger som egner seg for evaluering av prediksjoner. Vi fokuserer hovedsakelig på to ulike beregningsmetoder for å analysere prestasjonen til maskinlæringen og de ulike aktørene.

Root-Mean-Squared-Error (RMSE) representerer standardavviket til residualer. Det gir et mål på spredningen av residualene og indikerer dermed hvor konsentrert prediksjonene er rundt den faktiske utviklingen (Holmes, 2000). Metodisk anvender vi et utvidende vindu i datasettet, slik at et nytt residual utarbeides for hver nye prediksjon. Clark og McCracken (2001) bruker en lignende fremgangsmåte og argumenterer for at dette skaper en mer representativ RMSE for prediksjoner av tidsseriedata opp mot benchmarkmodeller. RMSE uttrykkes som kvadratroten til den gjennomsnittlige kvadrerte verdien til residualene for hele out-of-sample perioden (Formel 4.4).

$$RMSE = \sqrt{\left(\sum_{i=1}^n (\hat{y}_n - y_n)^2\right)/n} \quad (4.4)$$

Dette medfører at RMSE kan avleses med samme skala og enhet som vår avhengige variabel. I praksis betyr det at en RMSE på 0.7 gir prediksjoner som i snitt avviker 0.7 %-poeng fra faktisk verdi. Som følge av at residualene kvadreres vil store avvik straffes hardere sammenlignet med små avvik. Dette gjør en direkte tolkning av RMSE mer komplisert, men samtidig er den harde straffen for store avvik ideell for evaluering av prediksjoner. Makroøkonomiske prediksjoner er viktige verktøy i en rekke sentrale beslutninger i samfunnet, og store avvik kan derfor få konsekvenser.

En alternativ tilnærming til målingen av prediksjonsevne er mean absolute error (MAE), eller snitt absoluttfeil. MAE representerer gjennomsnittlig absolutt størrelse på residualene. Med samme notasjoner som tidligere blir beregningen som følger:

$$MAE = \left(\sum_{i=1}^n |\hat{y}_n - y_n|\right)/n \quad (4.5)$$

MAE kan avleses på samme skala og enhet som vår avhengig variabel og RMSE. En MAE på 0.4 innebærer at prediksjonene i snitt ligger 0.4 %-poeng fra faktisk økonomisk vekst. Forskjellen fra RMSE er at residualene ikke kvadreres. Det betyr at MAE ikke skiller mellom store og små avvik på lik linje med RMSE, og er et mindre ideelt enkeltstående mål for evaluering av makroøkonomiske prediksjoner. Likevel har den en mer intuitiv tolkning, noe som vil være nyttig i analysen av prediksjonene. Spesielt interessant er det å evaluere RMSE og MAE for prediksjonene opp mot hverandre. To aktører kan nemlig ha relativt like MAE, men forskjellige RMSE. En slik situasjon vil indikere at prediksjonene til aktøren med høyere RMSE har flere store avvik relativt til den andre.

Til slutt vil vi ved analyse av enkeltstående prediksjoner beregne absolutte avvik og faktisk avvik. Absolutt avvik brukes blant annet til indeksering av aktørenes prestasjoner, mens faktisk avvik først og fremst er gunstig for å si noe om enkeltstående prediksjoner er forventningsrette. Snittet av faktisk avvik anvendes i liten grad, ettersom et slikt snitt vil kunne være misvisende for vårt prediksjonsformål. I prediksjon av økonomisk vekst er prestasjonen like dårlig uavhengig av om avviket er positivt eller negativt. En snittberegning av faktisk avvik gjenspeiler ikke et slikt aspekt. I beregningen vil negative og positive avvik kunne nulle hverandre ut, slik at prediksjonene ser mer presise ut enn de er.

5 Resultater

5.1 Vurderingsgrunnlag og resultater

Med utgangspunkt i tidligere beskrevet datasett og metodisk fremgangsmåte sammenlignes resultatene fra våre tre maskinlæringsmodeller med utvalgte benchmarks. En sentral problemstilling for evalueringen av modellenes resultater omhandler ulike publiseringer av faktisk BNP utvikling. Stark og Croushore (2002) presenterer tre alternative tilnærminger: (1) den siste oppdateringen, (2) siste versjonen før en strukturell revisjon og (3) data publisert en bestemt periode etter første publisering. For våre modeller benytter vi konsekvent den siste oppdateringen av BNP-vekst i Fastlands-Norge.

5.1.1 Kvartalsvise prediksjoner

Første del av analysen omfatter en sammenligning av våre maskinlæringsmodeller opp mot en benchmark for prediksjoner for ett kvartal frem i tid. Som benchmark tar vi utgangspunkt i prediksjoner laget av Norges Bank, gjort tilgjengelig i hver pengepolitiske rapport fra 2010. I en slik kvartalsrapport publiserer Norges Bank normalt prediksjoner for firekvartalsveksten i inneværende kvartal og det påfølgende kvartalet, hvor vi fokuserer på sistnevnte.

Tabell 5.1 viser prestasjonen til Norges Bank opp mot våre tre maskinlæringsmodeller, fra 1.kvartal 2010 til og med 3.kvartal 2018.

Tabell 5.1: Resultat av kvartalsvise prediksjoner for 2010-2018³

obs: 91	Norges Bank	Elastic Net	Random Forest	Super Learner	Økt presisjon ML ¹
RMSE ²	0.63	0.80	1.00	0.51	19%

¹ Relativ sammenligning av beste maskinlæringsmodell og Norges Bank

² Absolutt avvik %-poeng

³ Fra 1.kvartal 2010 til 3.kvartal 2018

Resultatene gir innsikt i flere interessante aspekter. Hverken Elastic Net eller Random Forest er i stand til å levere mer konsistente prediksjoner enn Norges Bank. Med en RMSE på 0.63 er Norges Bank markant bedre enn de to nevnte maskinlæringsmodellene. Super Learner er derfor den eneste av våre modeller som leverer mer presise prediksjoner

enn Norges Bank. I snitt leverer Super Learner prediksjoner med 0.12 %-poeng lavere RMSE enn Norges Bank; en forbedring på 19%. Det er likevel viktig å presisere at prestasjonen til maskinlæringsmodellene preges av at det er norsk økonomi som predikeres. Til sammenligning presenterer Tiffin (2016) og Jung et al. (2018) resultater der også Elastic Net konsistent overgår prediksjoner gjort av IMF World Economic Outlook for både avanserte og fremvoksende økonomier.

Tabell 5.2: Konfidensintervaller for avvik - kvartalsvise prediksjoner

Konfidensintervall (95%)	Elastic Net	Random Forest	Super Learner
Øvre grense ¹	0.78	0.99	0.53
Nedre grense ¹	0.41	0.55	0.33

¹ Absolutt avvik i %-poeng

En videre sammenligning av resultatene kan gjøres ved å beregne konfidensintervaller for de absolutte avvikene til de ulike maskinlæringsmodellene. Et 95% konfidensintervall impliserer at man ved gjentatte forsøk vil i 95% av tilfellene ha en gjennomsnittsverdi som ligger innenfor det oppgitte intervallet (Wooldridge, 2012). Et smalere konfidensintervall vil derfor indikere mer konsistente prediksjoner. Konfidensintervallene til de tre maskinlæringsmodellene er illustrert i tabell 5.2. Super Learner har den laveste nedre og øvre grensen, med et intervall på 0.33 til 0.53. I praksis betyr dette at ved gjentatte forsøk vil gjennomsnittsverdien av Super Learner sine absolutte avvik i 95% av tilfellene ligge mellom 0.33 og 0.53 %-poeng. Etersom avviket er absolutt, kan avvikene både være negative og positive. Intervallet gir støtte til resultatene i tabell 5.1, som beskriver Super Learner sine prediksjoner som de mest presise. Konfidensintervallet til Super Learner er også det smaleste av de tre modellene. Dette indikerer at Super Learner gir mer stabile prediksjoner enn de to andre modellene. I den videre vurderingen opp mot Norges Bank vil Super Learner derfor representere maskinlæringsmodellene.

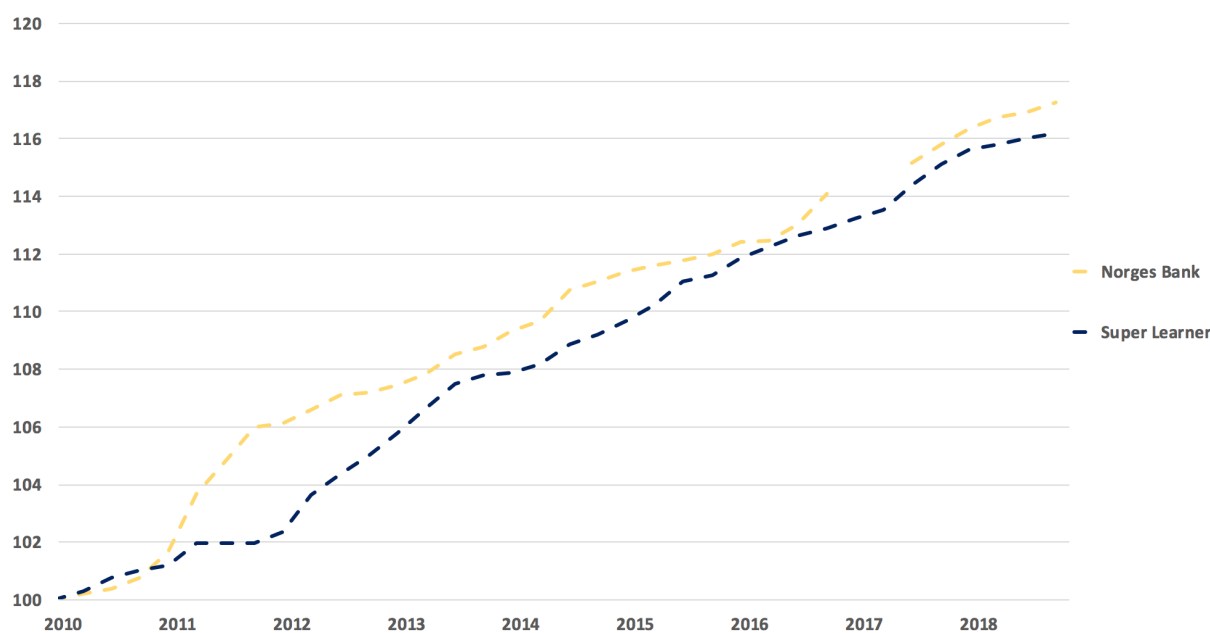
Tabell 5.3: Deskriptiv sammenligning av Super Learner og Norges Bank - kvartalsvise prediksjoner for 2010-2018

	Super Learner	Norges Bank
Kvartaler vunnet (%) ¹	31	69
Minste absolutt avvik ²	0.00	0.01
Største absolutt avvik ²	1.23	1.95
Median absolutt avvik ²	0.36	0.35
Snitt absoluttfeil (MAE) ²	0.43	0.48

¹ Andel kvartaler der prediksjonen ligger nærmest faktisk vekst

² Absolutt avvik %-poeng

Tabell 5.3 gir en deskriptiv sammenligning av prediksjonene til Super Learner opp Norges Bank. En direkte sammenligning av samtlige prediksjoner viser at Norges Bank predikerer bedre enn Super Learner i 69% av kvartalene. Dette til tross for at Super Learner har mer presise prediksjoner basert på både RMSE og MAE. Både Super Learner og Norges Bank har minste avvik som er ekstremt lave, men Norges Bank har en betydelig høyere verdi for sitt største avvik. Hele 0.72 %-poeng større er det største avvik sammenlignet med Super Learner. I tillegg er det interessant at avstanden mellom de to MAE verdiene er lavere enn avstanden mellom de to RMSE verdiene. Dersom Norges Bank har flere store avvik enn maskinlæringen vil RMSE straffe dette hardere enn MAE. Det kan derfor virke som at Norges Bank ofte predikerer bedre enn Super Learner, men når de store avvikene oppstår klarer Super Learner seg betraktelig bedre enn Norges Bank. Dette kan forklare hvorfor den totale presisjonen er best for maskinlæringen.

Figur 5.1: Indeksert absolutt avvik - kvartalsvise prediksjoner for 2010-2018

* Absolutte avvik for prediksjon av firekvartalsvekst i Fastlands-Norge. 4.kvartal 2009 som utgangspunkt og 1.kvartal 2010 som første vekstkvartal. Absolutt avvik i hvert kvartal legges til i indekseringen. Mangler prediksjoner fra Norges Bank i 4.kvartal 2016 og 1.kvartal 2017.

Basert på resultatene er det interessant å identifisere hvilke perioder prediksjonene til Super Learner og Norges Bank er mest/minst presise. Figur 5.1 er en behjelpelig grafisk illustrasjon i en slik sammenheng. Figuren viser en indeksering av de absolutte avvikene til Super Learner og Norges Bank, med 4.kvartal 2009 som utgangspunkt. Avstanden mellom de stiplede linjene og den horisontale akse representerer det akkumulerte absolutte avviket til den respektive aktøren. Grafen bekrefter inntrykket av to jevne prestasjoner, som følger hverandre tett. Derimot legger vi merke til at Super Learner i større grad følger den faktiske utviklingen i starten av perioden sammenlignet med Norges Bank. Avstanden Super Learner her opparbeider seg i form av lavere akkumulert absolutt avvik ser ut til å bli avgjørende. Avviket holder seg lavere i hele perioden, til tross for at avstanden til Norges Bank i perioder synker. Dette indikerer at prediksjonene til Norges Bank er mer presise i resten av perioden, noe som samsvarer med at Norges Bank har mest presise prediksjoner i 69% av kvartalene. Den avgjørende perioden ser ut til å være slutten av 2010 og starten av 2011, noe vi vil analysere nærmere i 6.2.2.

5.1.1.1 Robusthet ved prediksjon av kriseperioder

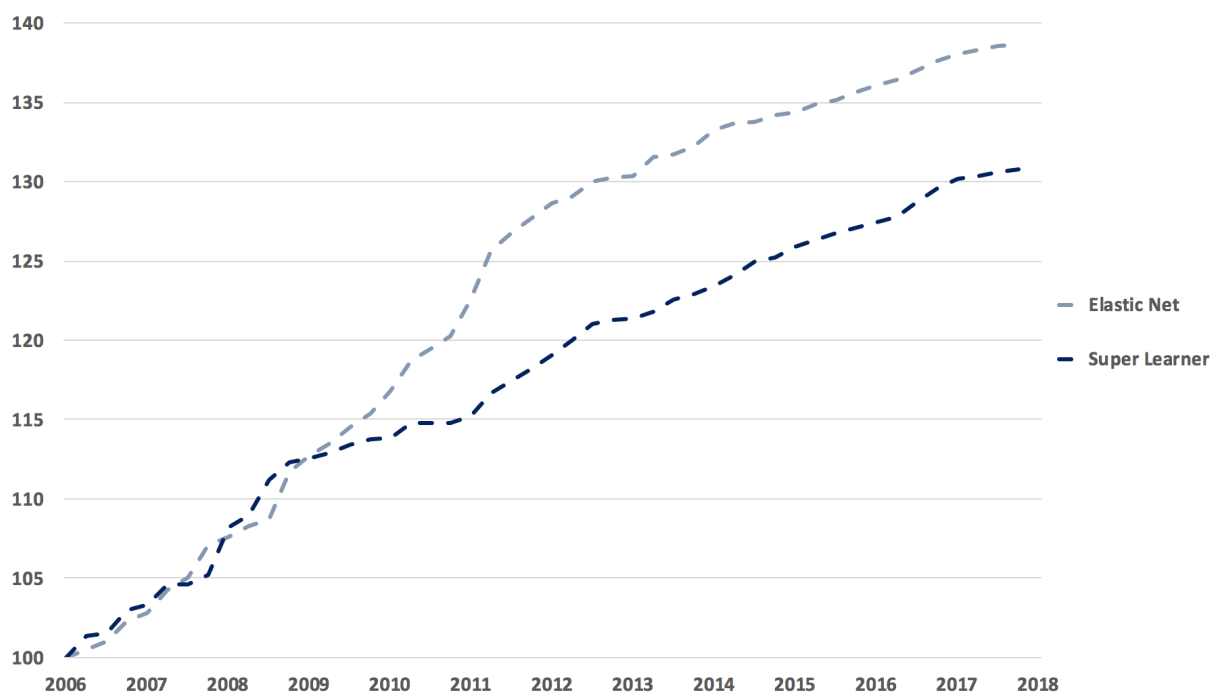
I den påfølgende delen av analysen undersøker vi maskinlæringsmodellenes anvendelighet i den virkelige verden. Et sentralt aspekt ved dette er å teste modellenes robusthet ved prediksjon av kriseperioder. I seksjon 5.1.1 er den første prediksjonen satt til å være 1.kvartal 2010. Dette innebærer at en av verdenshistoriens mest omfattende finansielle kriser foreløpig er inkludert i treningssettet. Ved å utvide tidshorisonten for prediksjonene til 1.kvartal 2007 utarbeides en out-of-sample RMSE som tar hensyn til en ekstraordinær omstendighet. En naturlig hypotese, uavhengig av metode, vil være en svekkelse av presisjonen gjennom en høyere RMSE. I tabell 5.4 oppsummeres prestasjonen for kvartalsvise prediksjoner i den utvidede perioden for de respektive maskinlæringsmodellene.

Tabell 5.4: Deskriptiv statistikk - kvartalsvise prediksjoner for 2007-2018

obs: 91	Elastic Net	Random Forest	Super Learner
RMSE ¹	0.92	1.45	0.79
Minste absolutt avvik ¹	0.06	0.02	0.00
Største absolutt avvik ¹	2.71	4.13	2.88
Median absolutt avvik ¹	0.49	0.77	0.44
Snitt absoluttfeil (MAE) ¹	0.70	2.57	0.57

¹ Absolutt avvik %-poeng

Som forventet observerer vi en høyere RMSE for samtlige modeller sammenlignet med tidligere prediksjonshorisont. I snitt er det snakk om 27% høyere RMSE for de tre modellene når finanskrisen inkluderes i prediksjonene. De minste absolutte avvikene holder seg fortsatt lave, og innebærer at presisjonen i de beste prediksjonene fortsatt er høy. Den vesentlige forskjellen kan spores til de største absolutte avvikene. Gitt at modellene nå forsøker å predikere en finansiell krise, er det ikke overraskende med økte avvik. Random Forest sitt største avvik er på hele 4.13 %-poeng, mens Elastic Net og Super Learner ligger relativt nærme hverandre med 2.71 og 2.88 %-poeng. For Super Learner innebærer dette en økning på 1.65 %-poeng relativt til det største avviket tabell 5.3. At dette avviket oppstår under finanskrisen er sannsynlig. Videre er det interessant at median og MAE holder seg relativt lave for Super Learner. Dette kan indikere at presisjonen fortsatt er stabil.

Figur 5.2: Indeksert absolutt avvik - kvartalsvise prediksjoner for 2007-2018

* Absolutte avvik for prediksjon av firekvartalsvekst i Fastlands-Norge. 4.kvartal 2006 som utgangspunkt og 1.kvartal 2007 som første vekstkvartal. Absolutt avvik i hvert kvartal legges til i indekseringen.

For å identifisere hvilke perioder som er avgjørende for forskjellene i Super Learner og Elastic Net sin prestasjoner, indekseres avvikene slik vi gjorde i seksjon 5.1.1 (Figur 5.2). Med utgangspunkt i 4.kvartal 2006 er de to maskinlæringsmodellene indeksert, der den absolutte feilen til modellene akkumuleres med størrelsen på feilen hvert kvartal. Av figuren legger vi merke til at begge modeller ser ut til å prestere jevnt under finanskrisen og oljekrisen. Det kan se ut til at modellene i større grad beveger seg vekk fra faktisk utvikling under finanskrisen sammenlignet med oljekrisen, noe som indikerer at de har størst problemer med å predikere finanskrisen. Gapet mellom Super Learner og Elastic Net ser ut til å oppstå i perioden mellom de to krisene, fra 2010 til 2014.

Vi vurderer presisjonen til å være akseptabel på et overordnet nivå, ettersom perioden nå inneholder to kriseperioder. Jung et al. (2018) gjennomfører en identisk robusthetstest for fem ulike økonomier. Her har de i tillegg inkludert mer tradisjonelle metoder som Autoregressiv Modell (AR), Bayesian Averaging Model (BAM) og Vektor Autoregressiv Modell (VAR). For samtlige av økonomiene produserer Elastic Net og Super Learner prediksjoner med en signifikant lavere RMSE enn både tradisjonelle metoder og benchmarkmodellen til World Economic Outlook. Den beste maskinlæringsmodellen

i hvert land leverte i snitt 41% mer presise prediksjoner relativt til den beste av de øvrige modellene. I likhet med våre funn er likevel avvikene større sammenlignet med prediksjonene gjort fra 2010; RMSE for beste maskinlæringsmodell er 0.02-0.3 %-poeng høyere i robusthetstesten for de ulike økonomiene.

5.1.2 Årlige prediksjoner

Mange aktører utarbeider årlige prediksjoner for norsk økonomi. Vi ønsker derfor å undersøke hvor gode årlige prediksjoner maskinlæring kan produsere, basert på den kvartalsvise modellen. Etersom kvartalsprediksjonene er på formen år-over-år, er det interessant å anvende firekvartalsveksten i 4.kvartal som et mål på den årlige veksten i BNP for Fastlands-Norge. Tankegangen er enkel; år-over-år veksten for 4.kvartal er endringen fra 4.kvartal år T-1 til 4.kvartal år T. I praksis vil dette illustrere utviklingen i første, andre, tredje og fjerde kvartal, slik at vi får en representativ tilnærming til den årlige veksten.

For de årlige prediksjonene presterer Super Learner klart best av de tre maskinlæringsmodellene. Med en RMSE på 0.40 utkonkurrerer den Elastic Net og Random Forest, som henholdsvis har en RMSE på 1.1 og 1.06. Vi bruker derfor Super Learner som representant for maskinlæring i den videre analysen. For å vurdere prestasjonen til prediksjonene fra Super Learner har vi samlet historiske prediksjoner fra Norges Bank, DNB, Finansdepartementet og SSB. Vår fremgangsmåte i seksjon 3.2 er inspirert av Sucarrat og Gharsallah (2019), som løser problemstillingen rundt prediksjonstidspunkt ved å konsekvent hente prediksjoner utgitt på samme tidspunkt hvert år. Både Norges Bank, Finansdepartementet og SSB slipper sine rapporter i september, men på ulike datoer. Det vil derfor være små forskjeller i informasjonsgrunnlagene av variabler med daglig frekvens. Vi mener derimot denne informasjonsasymmetrien er liten, og anser informasjonsgrunnlaget til aktørene som tilnærmet likt. DNB har vi gitt et bevisst informasjonsfortrinn, slik at vi får et referansepunkt på verdien av mer informasjon.

Tabell 5.5: Resultat av årlige prediksjoner for 2010-2017

	RMSE ¹	Økt presisjon med ML ²	Prediksjonstidspunkt	RMSE ³
ML - Super Learner	0.40	-	September	-
Norges Bank	0.64	37.5%	September	0.66
DNB	0.65	38%	Desember	0.49
Finansdepartementet	0.65	38%	September	0.53
SSB	0.68	41%	September	0.64

¹ Absolutt avvik %-poeng. Basert på revidert BNP Fastlands-Norge nedlastet 3. september 2019

² Relativ sammenligning av beste maskinlæringsmodell (Super Learner)

³ Absolutt avvik %-poeng. Basert på urevidert BNP Fastlands-Norge fra Samfunnsøkonomenes prognosepris

Resultatene for prediksjonene er vist i tabell 5.5. $RMSE^1$ er beregnet på de respektive prediksjonene fra de ulike aktørene og revidert BNP vekst for Fastlands-Norge fra september 2019. I tillegg er det inkludert en $RMSE^3$ beregning som baserer seg på urevidert BNP-vekst fra samfunnsøkonomenes prognosepris. Dette for å gi et inntrykk av hvilken effekt senere revideringer av veksten har på prestasjonen til prediksjonene.

Som vi ser av tabell 5.5 presterer maskinlæringen best av samtlige aktører, med en RMSE på 0.40. Sammenlignet med Norges Bank, Finansdepartementet, DNB og SSB leverer maskinlæringsmodellen betraktelig mer presise prediksjoner i perioden. Spesielt interessant er det at både Super Learner og Norges Bank presterer bedre enn DNB, som har et informasjonsfortrinn. For tre av de fire aktørene som har vært en del av prognoseprisen er det interessant at prediksjonene er mer presise hvis avvikene beregnes ut i fra den ureviderte veksten.

Tabell 5.6: Deskriptiv statistikk - årlige prediksjoner for 2010-2017

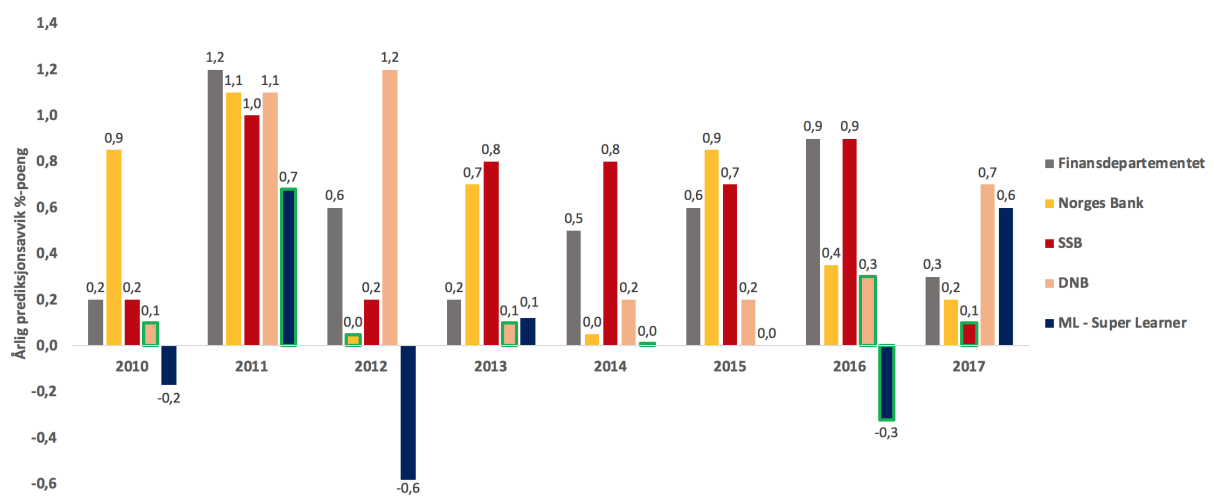
	Min ¹	Maks ¹	Median ¹	Snitt absoluttfeil (MAE) ¹
ML - Super Learner	0.00	0.68	0.25	0.31
DNB	0.10	1.20	0.25	0.49
Norges Bank	0.05	1.10	0.53	0.52
Finansdepartementet	0.20	1.20	0.55	0.56
SSB	0.10	1.00	0.75	0.59

¹ Absolutt avvik %-poeng

Den sterke prestasjonen til Super Learner bekreftes videre ved å studere deskriptiv statistikk for de absolutte avvikene, illustrert i 5.6. Super Learner har både den laveste

maks- og minverdien. For de resterende aktørene registrerer vi at Norges Bank har den laveste minverdien, mens SSB har den laveste maksverdien. Avstanden fra Super Learner til de resterende aktørene er gjennomgående større for RMSE enn for MAE. Dette indikerer at Super Learner har mer presise prediksjoner, med færre store avvik, relativt til de andre aktørene. Det er forventet at DNB leverer presise prediksjoner gitt deres informasjonsfordel, og derfor er det bemerkelsesverdig at Super Learner leverer bedre prediksjoner til tross for dette.

Figur 5.3: Avvik fra faktisk vekst - årlige prediksjoner for 2010-2017

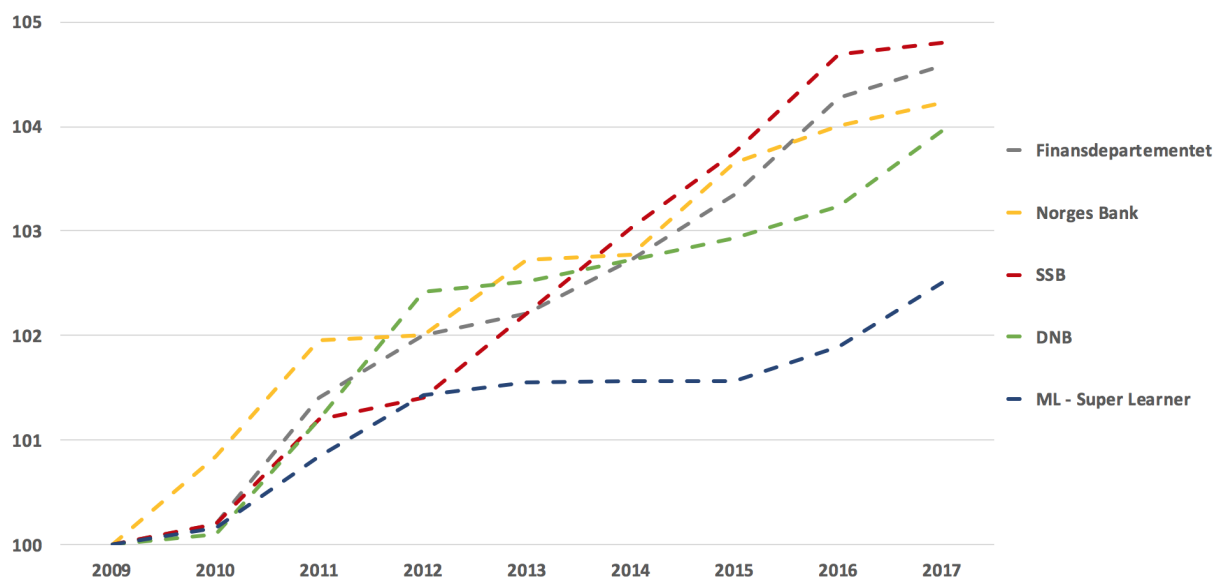


* Avvik beregnes med basis i prediksjon og faktisk BNP vekst for Fastlands-Norge. Laveste avvik hvert år er markert med grønn ytterkant.

I figur 5.3 er avviket fra faktisk vekst illustrert for alle prediksjonene til hver aktør. Faktisk avvik gir et godt bilde på om prediksjonene er forventningsrette og rasjonelle (Sucarrat og Gharsallah, 2019). For hvert år er det laveste avviket merket med en grønn kant, for å fremheve den beste prediksjonen. Maskinlæringen går seirende ut i fire av de åtte årene, henholdsvis 2011, 2014, 2015 og 2016. DNB den beste prediksjonen i tre av årene, mens Norges Bank er best i det siste. Det er interessant å merke seg at maskinlæringen presterer svært godt i perioden fra 2013 til og med 2016. Avvikene i perioden er imponerende lave, spesielt tatt i betraktning at norsk økonomi opplevde en krevende oljekrise i 2016. Oversikten bekrefter inntrykket av at maskinlæringen har gjennomgående lavere avvik, men det er interessant hvordan maskinlæringen predikerer at veksten vil ta en annen retning enn majoriteten av de andre aktørene predikerte i 2010, 2016 og 2017. Blant de resterende aktørene legger vi merke til at Norges Bank og DNB

sine prestasjoner i stor grad preges av store avvik i 2010, 2011 og 2012, men ellers leverer relativt presise prediksjoner.

Figur 5.4: Indeksert absolutt avvik - årlige prediksjoner for 2010-2017



* Absolutte avvik for prediksjon av firekvartalsvekst i 4.kvartal i Fastlands-Norge. 2009 som utgangspunkt og 2010 som første vekstår. Absolutt avvik i hvert år legges til i indekseringen.

I figur 5.4 er de absolutte prediksjonsavvikene indeksert mot faktisk BNP-vekst med 2009 som utgangspunkt. Første prediksjonsår er 2010 og markerer starten på prediksjonsperioden. De stiplede linjene viser utviklingen i akkumulert absolutt avvik for hver aktør. Figur 5.4 bekrefter at samtlige aktører utenom Norges Bank presterer relativt godt i 2010. I 2011 observerer vi en knekk oppover for samtlige aktører, noe som indikerer at prediksjonene ikke evner å fange opp den faktiske utviklingen. Fra 2012 til 2016 er det derimot tydelig at maskinlæringen presterer best. I denne perioden vokser det akkumulerte absolutte avviket hos de andre aktørene betraktelig mer. Det oppstår et tydelig skille mellom maskinlæringen og de øvrige aktørene, ettersom maskinlæringen leverer svært presise prediksjoner. Fra 2016 til 2018 jevner det seg mer ut, der det akkumulerte gapet mellom aktørene holder seg relativt stabilt. Helningen på linjene blir for samtlige aktører slakere, noe som indikerer at flere av prediksjonene er presise med lave avvik.

6 Diskusjon

6.1 Verdien av ny informasjonen

Datasettet vi benytter er bygget opp av variabler som oppdateres daglig, månedlig eller en gang i kvartalet. Som følge av ulike frekvenser foreligger det forsinkede publiseringsdatoer som begrenser vår mulighet til å anvende all informasjon knyttet til siste kvartal før en prediksjon. En evaluering av hvordan denne begrensningen påvirker presisjonen kan derfor være interessant å se nærmere på.

En mulig tilnærming til dette er å sammenligne RMSE for kvartalsvise prediksjoner med og uten hensyn til look-ahead bias. I seksjon 4.1.2.1 beskriver vi hvordan reelle prognoser et kvartal frem i tid replikeres ved hjelp av to-stegsprediksjoner av datapunkter som inneholder leads av høyfrekvente variabler. I en tilnærming som inneholder look-ahead bias tar vi noen forbehold som ikke gjør seg gjeldende i den virkelige verden. Prediksjonstidspunktet settes til siste dag i kvartal T-1 for prediksjonen av kvartal T. Som en forenkling forutsetter vi at all informasjon om kvartal T-1 er tilgjengelig for modellen. Dette innebærer at vi bryter med prinsippet om å unngå look-ahead bias, men vurderer det til å være akseptabelt når vi skal illustrere konsekvensene av informasjonsasymmetri ved prediksjoner.

Tabell 6.1: Look-ahead bias² i kvartalsvise prediksjoner for 2010-2018

obs: 91	Elastic Net	Random Forest	Super Learner
RMSE ¹ uten skjevhet	0.80	1.00	0.51
RMSE ¹ med skjevhet	0.75	0.91	0.37
Økt presisjon med skjevhet (%)	6.25	9.00	27.45

¹ Absolutt avvik %-poeng

² Informasjonsskjevhet

I tabell 6.1 presenteres resultatet av modellenes kvartalsvise prediksjoner for perioden 1.kvartal 2010 til og med 3.kvartal 2018. I første rad finner vi RMSE fra vår opprinnelige analyse, hvor vi ivaretar hensynet til look-ahead bias. Andre rad viser RMSE for prediksjoner med et urettmessig informasjonsfortrinn. Ikke overraskende kommer det frem at samtlige modeller leverer mer presise estimater med informasjonsfortrinn. Den

største presisjonsforbedringen er på 27.45% og tilhører Super Learner, som leverer en RMSE på lave 0.37. I appendiks figur A2.7 har vi lagt ved en grafisk fremstilling av prediksjonene opp mot faktisk utvikling. Basert på våre observasjoner virker verdien av nyere informasjon å være avgjørende for prediksjonsutfallet.

Lignende resultater er å finne i eksisterende litteratur. Aastveit et al. (2011) gjør en inngående analyse av effekten ny informasjon har på prediksjonsevnen til SAM. Deres fremgangsmåte bygger på et system der informasjon sorteres i blokker basert på publiseringsdato, slik at RMSE for én enkel prediksjonshorisont kan beregnes etter hver publisering av en blokk. Dermed er man i stand til å evaluere hvordan presisjonen endres etterhvert som mer data blir gjort tilgjengelig i løpet av et kvartal. Av resultatene fremgår det tre underliggende funn. For det første er at usikkerheten i prediksjonene er tiltakende ved økt horisont. Videre viser det seg at RMSE reduseres for kortere horisonter når mer informasjon blir gjort tilgjengelig. Dette omfatter prediksjoner av inneværende (nowcasting) og neste kvartal. Her har man funnet at publiseringen av BNP-veksten fra det foregående kvartalet er mest utslagsgivende for RMSE-forbedringen. Avslutningsvis er den positive effekten av nyere informasjon på kortere prediksjonshorisonter tilnærmet ikke-eksisterende for prediksjoner to og tre kvartaler frem i tid (Aastveit et al., 2011). Med andre ord virker ikke verdien av nyere informasjon å være av lik betydning for prediksjoner utover ett kvartal frem i tid, noe vi mener er et meget interessant funn. Vi bemerker oss at de kvartalsvise prediksjonene presentert i Norges Banks Staff Memo er for kvartalsvekst i fastlandsøkonomien, og dermed skiller seg fra firekvartalsvekst tilnærmingen i vår analyse.

6.2 Evaluering av kvartalsvise prediksjoner

6.2.1 Variasjonen mellom maskinlæringsmodellene

Av resultatene i seksjon 5.1.1 fremgår det relative prestasjonsforskjeller mellom de ulike maskinlæringsmodellene. For en bedre forståelse av resultatene kan det være interessant å undersøke hvorfor vi observerer de modellspekifikke forskjellene. Av modellene fra Caret presterer Elastic Net gjennomgående bedre enn Random Forest, uavhengig av horisont. I likhet med våre funn, identifiserer Tiffin (2016) tilsvarende forskjeller mellom modellene i sitt arbeid for Libanon. Begge metodene har i utgangspunktet gode forutsetninger til å

håndtere datasett med mange variabler relativt til antall observasjoner. Overordnet er Elastic Net spesielt gunstig ettersom den kombinerer dimensjonsreduksjon og variabel seleksjon i ett steg, i tillegg til å produsere resultater som er robuste overfor potensielle korrelasjoner mellom forklaringsvariabler. Random Forest evner på sin side å effektivt sortere et bredt spekter av forklaringsvariabler, slik at modellen kan fange opp ikke-lineære forhold og interaksjoner mellom variablene.

En undersøkelse av kryssvalideringen i Elastic Net kan gi et bedre innblikk i modellens sammensetning og prestasjoner. Elastic Net er en sammenfatning av en OLS-regresjon med et restledd som inkluderer straffeledene i Ridge og Lasso. Med Ridge kommer muligheten for dimensjonsreduksjon, mens Lasso åpner for variabel seleksjon. Kryssvalideringen velger α og λ som minimerer RMSE. En oversikt over ulike sammensetninger av parameterne er gjengitt i figurene A2.2 og A2.3 i appendiks. De endelige verdiene som benyttes er $\alpha = 0.1$ og $\lambda = 0.2836$. For restleddet medfører dette en prioritering av dimensjonsreduksjonen fra Ridge, som innebærer at regresjonskoeffisientene minimeres proporsjonalt; spesielt høyt korrelerte variabler får tildelt tilnærmet like koeffisienter uten at noen favoriseres. Lambdaverdien gir en lavere vektning av restleddet relativt til OLS-regresjonen. Kort oppsummert får vi generert en modifisert OLS-regresjon som også søker å minimere koeffisientenes størrelse. At Elastic Net gir bedre prediksjoner enn Random Forest tilsier at det underliggende forholdet mellom våre forklaringsvariabler og BNP-utviklingen i fastlandsøkonomien potensielt er lineært.

Som vi nevnte i seksjon 4.3.2 er Random Forest noe vanskeligere å analysere. Det kommer av at sentrale prosesser i modellens tilnærming er lite matematisk dokumentert i fagmiljøene. Til tross for en intuitiv oppbygning har man ikke vært i stand til å opparbeide en dybdeforståelse for hvorfor Random Forest genererer gode eller dårlige resultater ved prediksjonsformål (Breiman (2001a), Lin og Jeon (2006), Biau et al. (2008a), Biau et al. (2008a), Biau et al. (2008b)).

Super Learner står frem som den klart beste maskinlæringsmodellen. Til forskjell fra våre to øvrige maskinlæringsmodeller er dette en algoritme som anvender en sammensetning av flere modeller, spesifisert i tabell 4.1 under seksjon 4.4. Forklaringen på algoritmens prestasjoner ligger trolig i dens evne til å tilpasse seg det aktuelle prediksjonsproblemet. Kryssvalideringen resulterer i en individuell vektning av de mest egnede modellene for det

endelige utgangspunktet for prediksjonene. I appendiks, figur A2.5, presenteres utfallet av kryssvalideringen. Koeffisientene gjengir vektingen til den enkelte modell i det vektete gjennomsnittet; en *koeffisient* = 0 indikerer at den aktuelle modellen er utelatt. Vi legger spesielt merke til at Elastic Net får tildelt nest høyeste vekt, mens Random Forest utelates helt. Dette er i samsvar med våre funn, knyttet til modellenes individuelle prestasjoner. Videre ser vi at Support Vector Machine (SVM) får tildelt høyest vekt, på hele 0.515. SVM er en algoritme som deler input variabler inn i to hovedkategorier med et hyperplan¹⁴. Ved bruk av kernerfunksjoner er den i stand til å transformere data som ikke er lineært separerbare på et todimensjonalt plan, og løser det istedet på et tredimensjonalt plan. Det optimale hyperplanet er det som maksimerer avstanden mellom seg og nærmeste observasjon (Vapnik og Cortes, 2005). SVM fungerer forøvrig godt på både lineære og ikke-lineære problemer. Super Learner åpner for testingen av en rekke modeller uten overdreven bruk av datakraft.

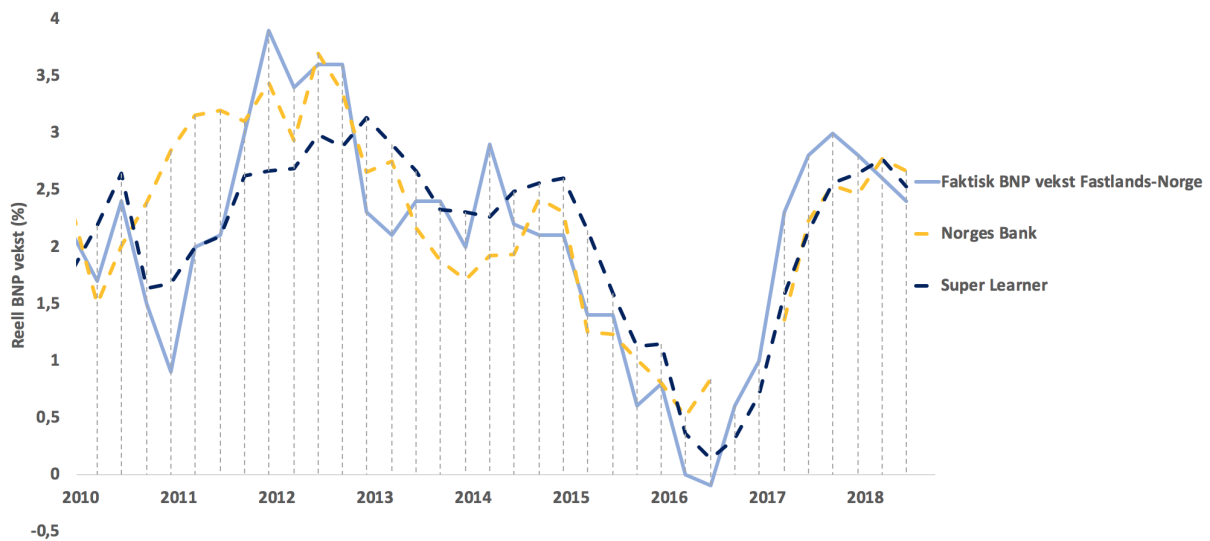
Det bør fremheves at Super Learner er spesielt robust på små datasett, noe vårt datasett kan kategoriseres som. Både Tiffin (2016) og Valland (2019) finner at prestasjonen til sammensatte modeller overgår selv de beste individuelle modellene. Dette kan forklare hvorfor både Super Learner og Elastic Net, som kombinerer flere modeller, presterer best.

6.2.2 Norges Bank og Super Learner

I sammenlikningen med Norges Bank i seksjon 5.1.1 produserte den beste maskinlæringsmodellen, Super Learner, i snitt 19% mer presise kvartalsvise prediksjoner for perioden 2010-2018. Dette til tross for at Norges Bank leverer mer presise prediksjoner i 69% av kvartalene, mot maskinlæringsens 31%. Dette gjør det interessant å se nærmere på *hvilke* kvartaler Norges Bank taper mot maskinlæringen, for å undersøke om det er spesifikke mønstre i prestasjonsforskjellene.

¹⁴En generalisering av et vanlig (todimensjonalt) plan til n dimensjoner norske leksikon (2018)

Figur 6.1: Kvartalsvise prediksjoner av firekvartalsvekst - Norges Bank, Super Learner og faktisk vekst



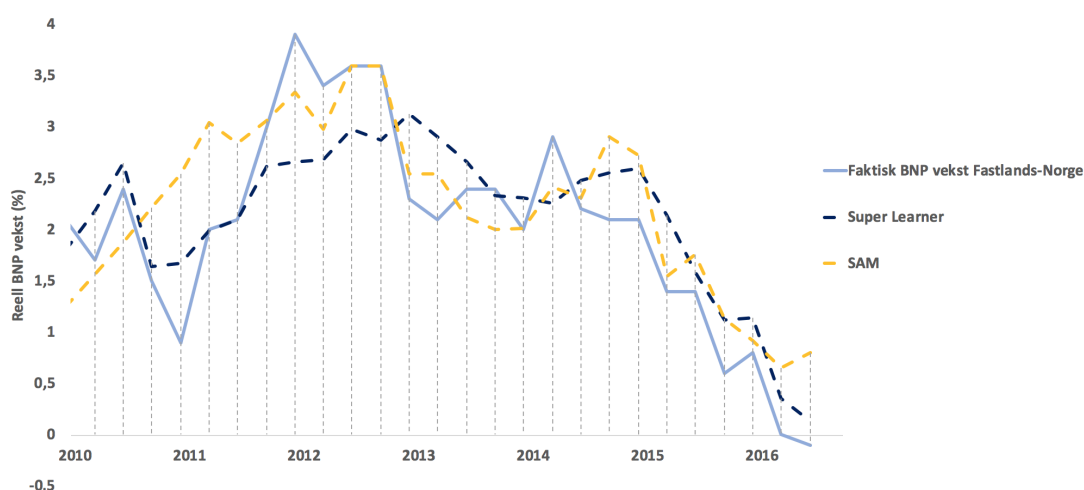
* Faktisk vekst er for BNP i Fastlands-Norge, fra 1.kvartal 2010 til 3.kvartal 2018. Mangler prediksjoner fra Norges Bank i 4.kvartal 2016 og 1.kvartal 2017.

I figur 6.1 har vi plottet prediksjonene til Super Learner opp mot prediksjonene til Norges Bank. I tillegg er faktisk firekvartalsvekst i fastlandsøkonomien inkludert. Norges Banks prediksjoner er hentet fra deres pengepolitiske rapporter. Der det har vært mulig har vi systematisk valgt anslag publisert i kvartalet før det predikerte kvartalet. Som nevnt i seksjon 3.2 mangler vi observasjoner for 4.kvartal 2016 og 1.kvartal 2017, som følge av at prediksjoner for firekvartalsveksten i disse kvartalene ikke ble publisert.

Vi observerer at Super Learner jevnt over presterer godt, med få store avvik. Likevel er det en tydelig tendens at maskinlæringen presterer bedre i andre halvdel av tidsserien, der avvikene er mindre sammenlignet med første halvdel. I første halvdel sliter modellen med å treffe omfanget av vekstendringene, selv om den stort sett evner å predikere riktig retning for veksten. I andre halvdel ligger prediksjonene betydelig nærmere den faktiske utviklingen. Spesielt merker vi oss at maskinlæringen predikerer bunnen av oljekrisen i 2016 bedre enn Norges Bank. At Norges Banks prediksjoner ligger nærmere faktisk BNP-vekst i 69% av kvartalene er likevel tydelig. Derimot skiller 4.kvartal 2010 og 1.kvartal 2011 seg ut. Norges Bank har her predikert en kontinuerlig oppgang i økonomien, men den faktiske utviklingen var avtakende. Det virker sannsynlig at disse avvikene er årsaken til at maskinlæringen ender opp med en lavere RMSE. Dette til tross for at Norges Bank har mer presise prediksjoner i et flertall av kvartalene.

Norges Bank sine prediksjoner hviler i stor grad på prediksjoner produsert av deres SAM modell. SAM prediksjonene danner utgangspunktet for prediksjonene, før en skjønnsmessig vurdering avgjør hva den endelige prediksjonen blir. I en pengepolitisk rapport publiseres både SAM prediksjonen og den endelige prediksjonen. Det er derfor interessant å sammenligne disse i 4.kvartal 2010 og 1.kvartal 2011, der de pengepolitiske prediksjonene treffer svært dårlig. Spørsmålet er om forklaringen til de store avvikene ligger i den skjønnsmessige vurderingen, eller om det er SAM som ga et feilaktig utgangspunkt. Hvis SAM er grunnen, er det interessant å analysere maskinlæringen sin prestasjon relativt til SAM. Prediksjonene fra maskinlæringen og SAM, er fremstilt sammen med faktisk firekvartalsvekst i fastlandsøkonomien i figur 6.2.

Figur 6.2: Kvartalsvise prediksjoner av firekvartalsvekst - SAM, Super Learner og faktisk vekst



* Faktisk vekst er for BNP i Fastlands-Norge, fra 1.kvartal 2010 til 3.kvartal 2016. Mangler prediksjoner fra SAM etter 3.kvartal 2016.

Det er tydelig at den dårlige prestasjonen til prediksjonene fra pengepolitisk rapport i 2010 og 2011 kan være en følge av SAM. Der den faktiske BNP-utviklingen avtar i 4.kvartal 2010 og 1.kvartal 2011, predikerer SAM stigende økonomisk vekst. Prediksjonene for 2010 og 2011 samsvarer derfor i liten grad med den faktiske utviklingen i norsk fastlandsøkonomi. Det er interessant at maskinlæringen evner å predikere retningen på veksten mellom 2010 og 2012, til tross for at den sliter med å identifisere omfanget av toppene og bunnene. Spesielt i 2012 sliter maskinlæringen med å fange opp den sterke veksten i norsk økonomi. I denne perioden presterer SAM betydelig bedre. Fra 2012 til og med 2016 er prestasjonen til SAM og maskinlæringen jevn, og modellene evner å predikere utviklingen godt.

6.2.2.1 Nærmere om avvikene i perioden 2010-2012

Det er tydelig at både Super Learner og SAM sliter med å predikere den økonomiske utviklingen i årene 2010, 2011 og 2012. SAM har store problemer med å fange opp både retning og omfang av utviklingen i 2011. Maskinlæringen fanger ikke opp omfanget av den avtagende veksten i 2011 eller den sterke veksten i 2012. En mulig forklaring til deres svake prestasjoner kan ligge i de strukturelle endringene som oppstod i kjølevannet av finanskrisen i 2009. Det virker ikke tilfeldig at modellene strever med å predikere veksten i etterkant av den største finansielle krisen siden den store depresjonen.

Nøkkelen til presise prediksjoner for alle prediksjonsmodeller ligger i treningen. Hvilke elementer som vektlegges i prediksjonene vil avhenge av modellspesifikke tolkninger av sammenhengen mellom den økonomiske veksten og forklaringsvariablene i treningsperioden. I seksjon 6.1 diskuterte vi viktigheten av den nyeste informasjonen i treningssettet. I den anledning er det viktig å poengtere at den siste informasjonen i treningssettet for prediksjonene i 2010 kommer fra 2009; året som ble sterkt preget av finanskrisen. Norsk økonomi klarte seg relativt bra gjennom krisen sammenlignet med resten av verden. Likevel må vi drøfte to viktige implikasjoner som gjør det utfordrende å lage prediksjoner for årene etter 2009.

For det første representerer finanskrisen en ekstraordinær hendelse som kan forårsake støy i treningssettet til prediksjonen av de første kvartalene. Veien videre etter en krise av et slikt omfang er vanskelig å spå, ettersom hver krise er unik og har individuelle særtrekk. At den siste informasjonen modellene trener på er fra finanskrisen, vil derfor kunne skape stor usikkerhet i prediksjonen av den videre utviklingen.

I tillegg er det viktig å reflektere rundt det strukturelle bruddet finanskrisen medførte for underliggende forhold i økonomien. Veien ut av krisen var preget av store redningspakker, og svært ekspansiv pengepolitikk preget den økonomiske politikken i store deler av verden (Grytten og Hunnes, 2016). Redningspakkene bestod av enorme lån fra statlige myndigheter og internasjonale organisasjoner. Lånene gikk til land som ikke hadde midlene til å stimulere aktiviteten i egen økonomi, samt betjene sin eksisterende gjeld. Slike løsninger bar preg av kortsiktighet. Ved gjeldsforfall var det få land som kunne tilbakebetale, og for å unngå konkurs, ble nye lånepakker utstedt. Problemene var spesielt store i Europa, der EU samarbeidet ble satt på store prøver. Plutselig ble det tydelig hvor utfordrende det var

for den europeiske sentralbanken å utforme en felles økonomisk politikk for en union der det var store forskjeller mellom medlemslandenes økonomiske situasjon. Finanskrisen hadde også store innvirkninger på forventningene til private bedrifter og husholdninger. Lavere vekstutsikter resulterte i en mer pessimistisk tilnærming med færre investeringer og høyere sparerate, hvilket forsterket oppbremsingen. Fortsatt er det svært høye gjeldsnivåer internasjonalt, og europeiske sentralbanker sliter med å øke den økonomiske aktiviteten. Lave renter for å stimulere økonomien er blitt den nye normen for verdens sentralbanker.

For maskinlæringen er det avgjørende hvor raskt den evner å oppfatte det strukturelle bruddet og tilpasse seg de nye økonomiske forholdene. Perioden etter 2009 er en utfordrende periode å predikere, ettersom tidligere strukturelle forhold ble mindre relevante og nye elementer vokste frem som drivere for den økonomiske veksten. Styringsrentene i en rekke land er et godt eksempel på dette. Rentene ble satt ned til rekordnivåer for å stimulere økonomien. Lavrentepolitikken har bidratt til å holde veksten nogenlunde oppe, men ikke bidratt til den samme stimulansen som en tradisjonelt ville forvente. Blant annet ble kvantitative lettelsener introdusert som et pengepolitisk virkemiddel for å senke de lange rentene, og gjøre investeringer mer attraktivt (Olsen, 2019). Den Europeiske sentralbanken har holdt styringsrenten nær null i lang tid, uten at veksten i økonomien har vist tegn til bedring. Spørsmålet er hvorvidt det kan ha forekommet en frikobling mellom rentesettingen og markedssaktørens atferd. I fagmiljøet er man tydelige på at Norge, som en liten åpen økonomi, blir sterkt påvirket av internasjonale vekstutsikter og renteforhold (Norman og Orvedal, 2012). Omveltninger i økonomien til våre nærmeste handelspartnere kan ha smitteeffekter på forhold i norsk økonomi. Isolert sett kan strukturelle endringer være noe av forklaringen bak den dårlige prestasjonen til prediksjonene til SAM og maskinlæringen i henholdsvis 2010, 2011 og 2012. I kjølevannet av finanskrisen måtte endringene i de strukturelle forholdene identifiseres, observeres og læres for å kunne produsere presise prediksjoner for den videre utviklingen. At både SAM og maskinlæringen leverer betraktelig mer presise prediksjoner fra 2013 til 2018 kan derfor ses på som en konsekvens av at modellene evner å plukke opp de strukturelle endringene og anvende de videre.

6.2.3 Robusthet ved prediksjon av kriseperioder

I seksjon 5.1.1.1 presenteres funn som impliserer redusert modellpresisjon når finanskrisen inkluderes i testsettene. Finansielle kriser representerer ekstraordinære omstendigheter som beskriver en sammenfallende krise i realøkonomien og i finansmarkedene (Grytten og Hunnes, 2016). Det er derfor av stor verdi for beslutningstakere å kunne forutse kriseoppbygning i økonomien. Fra et økonometrisk standpunkt er makroøkonomiske dataserier korte, og i slike serier er forekomstene av finansielle kriser sjeldne. Dette har medført at tradisjonelle økonometrimodeller stort sett har feilet med å utarbeide robuste metoder for å identifisere kriseoppbygning. Maskinlæringsmodeller er stort sett ikke-lineære og fleksible, hvilket kan gi bedre forutsetninger for å predikere slike uvanlige omstendigheter.

Tabell 6.2: Deskriptiv statistikk - kvartalsvise prediksjoner av kriseperioder

	Min ¹	Maks ¹	Median ¹	MAE ¹	RMSE
Q12007 - Q42009					
Super Learner	0.01	2.88	0.85	0.99	1.29
Elastic Net	0.41	2.71	0.70	1.0	1.22
Q12014 - Q42016					
Super Learner	0.07	0.74	0.34	0.40	0.43
Elastic Net	0.06	0.97	0.31	0.35	0.43

¹ Absolutt avvik %-poeng

Testsettet inneholder nå to spesielt urolige perioder for norsk fastlandsøkonomi; finanskrisen (2007-2009) og oljekrisen (2014-2016). Under finanskrisen opplevde norsk økonomi et fall i både nominelle og reelle boligpriser på henholdsvis 14% og 18%. Spesielt finansmarkedene ble hardt rammet gjennom et 64% fall på Oslo Børs. Rask tilførsel av likviditet i pengemarkedet, statsfinanser fra høye oljeinntekter, stans i konstruksjonen av nybygg og en lav rentepolitikk var likevel forhold som bidro til at Norge ble mindre rammet av finanskrisen enn andre land (Grytten, 2011). Oljekrisen var en realøkonomisk krise forårsaket av en oljepriskollaps i 2014. Prisfallet fikk konsekvenser for bunnlinjen i petroleumssektoren, hvilket medførte økt arbeidsledighet og lavere etterspørsel innenfor Norges største eksportindustri. Til tross for en sektorspesifikk krise, opplevde man en smitteeffekt på etterspørselen fra husholdningene og svekkede finansmarkeder grunnet en

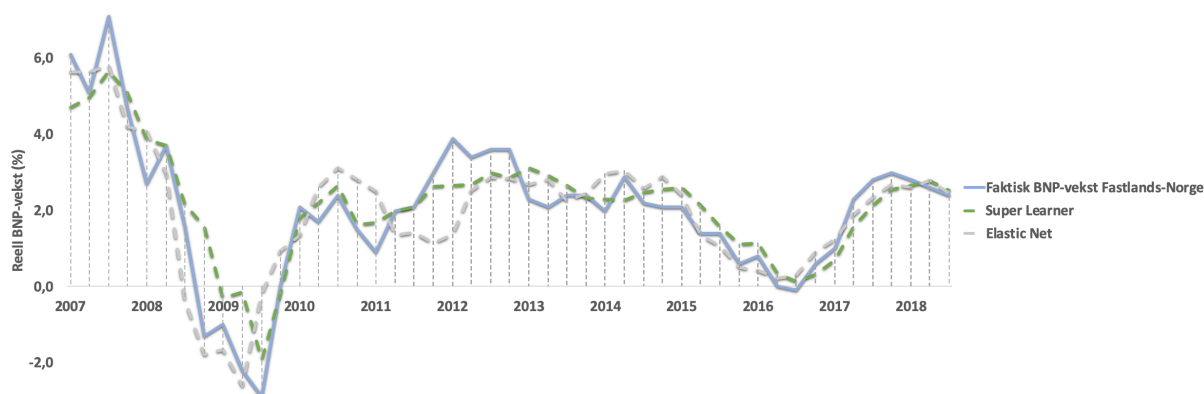
overvekt av tunge oljeaksjer på Oslo Børs (Juelsrud og Wold, 2019). Omfanget av krisen ble likevel begrenset av at den øvrige fastlandsøkonomien gikk godt i perioden.

Tabell 6.2 indikerer at finanskrisen var mer problematisk å predikere. Her varierer prediksjonsavvikene mellom 0.01 og 2.88 %-poeng. Av den deskriptive statistikken legger vi spesielt merke til at den mest presise prediksjonen til Elastic Net har et 0.40 %-poeng større avvik fra faktisk utvikling enn Super Learner. Likevel har Elastic Net lavere median og maksimalt avvik, hvilket indikerer en mer snever spredning av prediksjonsavvikene. Resultatet er dermed en lavere RMSE relativt til Super Learner. Dette gjør Elastic Net til den beste maskinlæringsmodellen i prediksjonen av finanskrisen. Sammenlignet med den overordnede statistikken for alle prediksjoner i tabell 5.4, er det tydelig at denne perioden inneholder mange av de største prediksjonsavvikene for begge modeller.

Prediksjonene for oljekrisen er markant bedre enn for finanskrisen. Det største avviket observert for maskinlæringsmodellene tilhører Elastic Net, og er på 0.97 %-poeng. Det gjennomsnittlige absolutte avviket for modellene holder seg under 0.40 %-poeng; mer enn en halvering sammenlignet med under finanskrisen. En lik RMSE på 0.43 reflekterer en helt jevn prestasjon mellom de to modellene i løpet av oljekrisen.

Figur 6.3 gir en grafisk fremstilling av modellenes prediksjoner opp mot faktisk BNP-utvikling fra 2007 til 2018. Overordnet klarer begge modellene å predikere riktig retning for veksten i begynnelsen av hver krise. Elastic Net klarer i større grad å fange opp dybden til krisene, men sliter relativt til Super Learner i perioden mellom krisene. Dette kan skyldes endringer i strukturelle forhold som vanskeligere lar seg forklare med en mer lineær modell. RMSE i perioden mellom krisene er henholdsvis 0.59 og 1.1 for Super Learner og Elastic Net. Det største avviket for begge modeller finner vi innenfor fire kvartaler midt i finanskrisen: Super Learner underestimerer nedgangen i 4.kvartal 2008 med 2.88 %-poeng, mens Elastic Net feilestimerer utviklingen i 3.kvartal 2009 med 2.71 %-poeng. Størrelsen på hvert av avvikene er signifikant, og kan alene sammenlignes med den forventede årlige vekstraten i norsk fastlandsøkonomi under en høykonjunktur.

Figur 6.3: Kvartalvise prediksjoner av firekvartalsveks - Super Learner, Elastic Net og faktisk vekst



* Faktisk vekst er for BNP i Fastlands-Norge, fra 1.kvartal 2007 til 3.kvartal 2018.

Som det fremgår av både figur og deskriptiv statistikk klarer maskinlæringsmodellene seg betraktelig bedre under oljekrisen enn finanskrisen. En forklaring kan ligge i omfanget av de to krisene. Globalisering bidro til at en amerikansk finanskrisen spredte seg til en rekke internasjonale finansmarkeder og realøkonomiske sektorer, og medførte en omfattende nedgangskonjunktur på tvers av landegrensene. Oljekrisen var derimot sektorspesifikk og geografisk begrenset til områder med høy petroleumsaktivitet. Et høyt kostnadsnivå i utvinningsprosessen relativt til internasjonale oljeprodusenter gjorde dette til en særnorsk krise (Juelsrud og Wold, 2019). Oljeprisnedgangen fikk store konsekvenser for petroleumssektoren, men hadde kun mindre innvirkninger på den øvrige fastlandsøkonomien utenfor Sørvest-Norge.

En annen forklaring kan være den utvidede treningen. I forkant av finanskrisen inneholder treningssettet kun én større nedgangskonjunktur, fra dotcom-krisen i 2000. At finanskrisen etterhvert inngår i de utvidende treningssettene kan ha gitt modellene bedre forutsetninger til å predikere utviklingen under oljekrisen. Etter vårt skjønn viser maskinlæringen en god evne til å predikere nedgangskonjunkturer i norsk fastlandsøkonomi.

6.3 Evaluering av årlige prediksjoner

Resultatene for de årlige prediksjonene understreker anvendeligheten til maskinlæring i prediksjonen av økonomisk vekst over ulike horisonter. Igjen kommer Super Learner sterkest ut av maskinlæringsmodellene. Med en RMSE på 0.4 er Super Learner mer enn 63% mer presis enn nest beste modell. Random Forest hever prestasjonen sin relativt til

Elastic Net, og produserer en lavere RMSE for årlige prediksjoner enn for kvartalsvise prediksjoner. Dette kan indikere at modellens ikke-lineære egenskaper kommer mer til sin rett. Prediksjonen av årene 2010 og 2012 er spesielt utslagsgivende for både Random Forest og Elastic Net. Det absolutte prediksjonsavviket over de to årene er i snitt 1.6 %-poeng for Random Forest og 2.0 %-poeng for Elastic Net. I vår diskusjon om kvartalsvise prediksjoner identifiserte vi denne perioden som problematisk, grunnet strukturelle endringer i kjølevannet av finanskrisen.

Det er interessant at Super Learner opprettholder høy presisjon, som ved kvartalsvise prediksjoner. Jung et al. (2018) argumenterer for at maskinlæringsmodeller på overordnet nivå underpresterer ved årlige prediksjoner grunnet korte og mer volatile tidsserier. Årsaken er at årlige prediksjoner tradisjonelt baseres på årlige observasjoner, hvilket gir langt færre observasjoner. Treningsgrunnet blir redusert, og bestående av aggregerte variabler som kan være mer volatile. Vår tilnærming til årlige prediksjoner, ved bruk av firekvartalsvekten for 4.kvartal, kan derfor være en mulig forklaring på Super Learners prestasjon. På den ene siden er argumentet at kvartalsvise observasjoner kan forårsake mer støy i prediksjonen av årlig vekst. På den annen side vil maskinlæringsmodellene besitte et større informasjonsgrunnlag om utviklingen for hvert år i treningen. Hvorvidt dette utgjør en fordel eller ulempe vil derfor avhenge av modellenes evne til å anvende den ekstra informasjonen som kommer av vår tilnærming.

Super Learner er videre blitt sammenliknet med fire etablerte aktører med ulike formål for sine prognoser: (1) Norges Bank søker et utgangspunkt for sin rentestyring, (2) Finansdepartementet identifiserer virkningene av nåværende statsbudsjett for utarbeidelsen av nye, (3) SSB oppfyller et av sine samfunnsoppdrag, mens (4) DNB tilbyr en tjeneste til egen kundegruppe og et perspektiv i utarbeidelsen av konsensus fra privat sektor. Resultatene i seksjon 5.1.2 viser at Super Learner kommer best ut i sammenlikningen av RMSE, MAE og akkumulert absolutt avvik. Finansdepartementet er likevel påpasselige med å poengtere at vurderingen av historisk treffsikkerhet bør baseres på BNP-tallet publisert i begynnelsen av året etter anslagsåret (Finansdepartementet, 2019). Argumentet er at senere revisjoner kan bygge på andre definisjoner og beregningsmetoder enn de som var i bruk da nasjonalbudsjettet for det enkelte år ble utarbeidet. Et slikt forbehold har vi forsøkt å ta til følge i tabell 5.5. Her er det inkludert en RMSE-beregning for aktørene

basert på ureviderte BNP-tall¹⁵. RMSE på ureviderte BNP-tall viser en mindre forbedring for noen aktører, men vår overordnede konklusjon forblir uendret. Maskinlæringsmodellen viser gjennomgående høyere presisjon enn de øvrige aktørene. Målt i RMSE, er de utvalgte aktørene rimelig jevne. Av figur 5.3 er det interessant å se at Norges Bank, SSB og Finansdepartementet i stor grad har prediksjonsavvik i samme retning. Over perioden viser de en tendens til å overestimere veksten i fastlandsøkonomien. Det kan derfor virke som at prognosene ikke er like uavhengige som en kan forvente.

En mulig forklaring kan ligge i aktørenes tilnærminger til prediksjonene. Vi vet med sikkerhet at de tre overnevnte aktørene presenterer modellbaserte prognoser. Som vi gjennomgikk i seksjon 3.2 benytter Norges Bank SAM som utgangspunkt for sin endelige prognose. Videre er det verdt å merke seg at både Finansdepartementet og SSB tar utgangspunkt i KVARTS (Von Brasch (2019), Finansdepartementet (2019)). KVARTS¹⁶ er en makroøkonomisk kvartalsmodell for norsk økonomi, utviklet av Statistisk Sentralbyrå (Haakonsen, 2015). Modellen er basert på økonomisk teori og historiske data. Finansdepartementet får oversendt oppdatert modell og prognoser etter hver publisering av konjunkturtendensene fra SSB. Finansdepartementet utvikler så sine egne prognoser for norsk økonomi, der de også inkluderer det økonomiske opplegget for kommunene Von Brasch (2019). Med KVARTS som fellesnevner, i tillegg til den institusjonelle relasjonen som eksisterer mellom Norges Bank og Finansdepartementet, er det ikke unaturlig at aktørene fremstår rimelig samstemte i sine prognoser. Etter vår beste kunnskap er DNB sine prognoser mindre modellbaserte, der faglige skjønnsmessige vurderinger står mer sentralt. De faglige diskusjonene er fundert på dybdekunnskap om norsk økonomi og databanker med historisk informasjon. DNB har ingen klar trend i prediksjonsavvikene.

På samme måte som med kvartalsvise prediksjoner, er årene i etterkant av finanskrisen spesielt vanskelige å predikere. Unntaket finner vi i året ut av krisen, 2010, hvor Norges Bank er eneste aktør med et betydelig avvik. I 2011 overestimeres veksten av samtlige, mens fire av fem underestimerer veksten i 2012. Tatt i betraktning at veksten økte med 1.8 %-poeng fra 2011 til 2012, kan det virke som at aktørene fanger opp at veksten vil tilta, men feilestimerer tidspunktet. Videre ser vi at Super Learner utkonkurrer de øvrige aktørene fra 2013 til 2017. Som ved kvartalsvise prediksjoner, kan det virke som

¹⁵Kolonne fire fra venstre

¹⁶En matematisk modellformulering av norsk økonomi

maskinlæringsalgoritmen fanger opp de endrede strukturelle forholdende raskt. I årene etter 2012 har fastlandsøkonomien opplevd avtakende vekst. At både Norges Bank, Finansdepartementet og SSB stort sett overestimerer veksten i årene etter 2012 kan indikere at deres modellbaserte tilnærminger har problemer med å fange opp endringene i de strukturelle forholdende. I samme periode har DNB en tendens til å underestimere veksten, men produserer gode estimater relativt til de andre aktørene.

Overordnet ser vi store variasjoner i presisjonen til prognosene fra den enkelte aktør. Makroøkonomiske anslag er forbundet med mye usikkerhet, hvilket gjør det urealistisk å forvente nøyaktige prognoser fra år til år. Ser vi bort ifra 2011, har samtlige aktører levert enkelte gode estimater i løpet av perioden. Likevel impliserer MAE en dominerende prestasjon av Super Learner. Presisjonen til maskinlæringsmodellen er også best når vi vektlegger størrelsene til avvikene, gjennom RMSE. Super Learner virker med andre ord å ha best forutsetninger for årlig prediksjon av norsk økonomisk vekst etter 2010. Det er viktig å poengtere at evalueringen av prediksjonsevne på årlig vekst baseres på langt færre prediksjoner enn ved prediksjonen av firekvartalsvekst. Vi mener derfor at resulterende prediksjoner og konklusjoner bør behandles med varsomhet.

6.4 Variable Importance

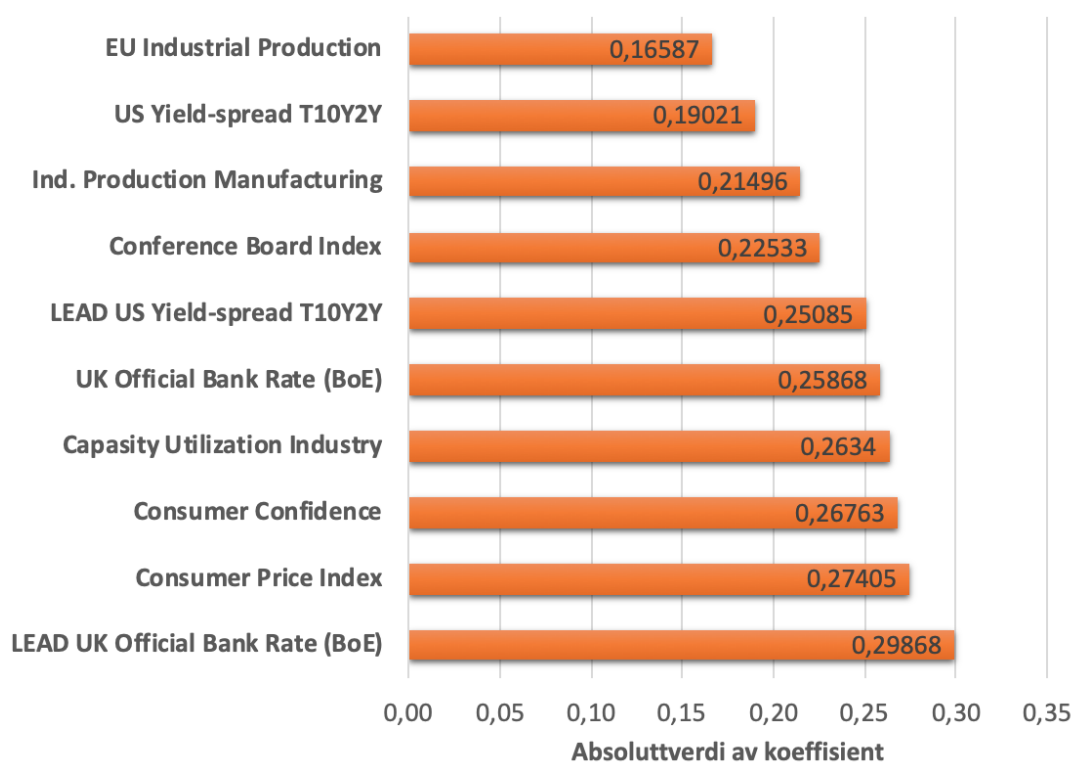
Makroøkonomiske sammenhenger er i stadig endring. Dermed er det nærliggende at maskinlæringsmodellene, gjennom kontinuerlig kryssvalidering, vektlegger forklaringsvariabler forskjellig for de ulike prediksjonene over tid. Forståelsen av sammenhengene står sentralt for økonomisk intuisjon, og det ville vært interessant å analysere hvilke variabler som har vært avgjørende for hver enkelt prediksjon. Funksjonene anvendt i vår analyse gir dessverre ikke en slik innsikt. I seksjon 6.5 går vi nærmere inn på hvilke problemstillinger en slik mangel medfører for anvendbarhet og økonomisk tolkning. I Caret pakken finnes det likevel en funksjon som gjør det mulig å analysere hvilke variabler som har vært viktigst på tvers av alle prediksjoner. Anvendbarheten til en slik funksjon er begrenset, og reduserer muligheten for å gjennomføre dybdeanalyser. Likevel ønsker vi å illustrere hvilken output funksjonen gir og hvilke overordnede linjer det er mulig å trekke basert på dette. Oversikten gir et generelt inntrykk av hvordan modellen vektlegger de forskjellige variablene, basert på alle prediksjonene.

Funksjonen som Caret introduserer kalles *VarImp*, som står for Variable Importance. VarImp åpner for en modellbasert estimering av bidraget til hver enkel variabel på tvers av alle prediksjoner. I praksis vil en oversikt fra VarImp for våre kvartalsvise prediksjoner fra 2007 til 2018 vise en aggregert oversikt av de viktigste variablene for alle de kvartalsvise prediksjonene i perioden. Fordelen med en modellbasert tilnærming er at man får en estimering som er tett knyttet opp mot modellens prestasjoner og potensielt får inkludert korrelasjonsstrukturer mellom variablene i utregningen (Kuhn, 2008). I den videre analysen tar vi utgangspunkt i Elastic Net, vår beste maskinlæringsmodell fra Caret, i estimeringen av variablenes viktighet. VarImp for Elastic Net rangerer variablenes viktighet basert på absoluttverdien til koeffisientene i den kryssvaliderte modellen.

6.4.1 Sentrale variabler i kvartalsvise prediksjoner for 2007-2018

Figur 6.4 viser en oversikt over de 10 viktigste variablene for prediksjonene gjort med Elastic Net i perioden fra 1.kvartal 2007 til 3.kvartal 2018.

Figur 6.4: De 10 «viktigste» variablene bak kvartalsvis prediksjon



* VarImp for Elastic Net vurderer viktighet ut i fra absoluttverdien til koeffisientene. Merk at Elastic Net minimerer koeffisientene til høyt korrelerte variabler. Prediksjoner fra 1.kvartal 2007 til 3.kvartal 2018.

Koeffisientene til de 10 viktigste variablene gir et sammensatt inntrykk av hva som driver prediksjonene fra Elastic Net i perioden. De åtte viktigste variablene har alle koeffisient over 0.20, der ingen virkelig skiller seg ut. Den viktigste variabelen er styringsrenten i Storbritannia, etterfulgt av det norske inflasjonsnivået og den norske forbrukertilliten. Blant de 10 viktigste variablene er fire norske, mens de resterende seks er utenlandske. De dominerende utenlandske aktørene er Storbritannia, EU og USA; alle sentrale aktører for handelsnasjonen Norge og den globale økonomien.

Til tross for at Norge er en liten åpen økonomi som avhenger av sine handelspartnere, er også nettoeffekter innenlands avgjørende for konjunktursvingningene i økonomien. Konsumprisindeks (KPI), forbrukertillit og industriell kapasitetsutnyttelse utgjør de tre viktigste norske variablene, basert på absoluttverdien til koeffisientene. At KPI får så stor vekt er interessant. Normalt vurderes inflasjon som en etterslepene indikator på økonomisk aktivitet, men samtidig representerer den et viktig mål for Norges Banks rentepolitikk. Husebø og Wilhelmsen (2005) finner at KPI faktisk har ledende informasjonsverdi om utviklingen i fastlandsøkonomien. Sammenhengen er motsyklisk og variabelen skal ut i fra deres funn lede økonomisk vekst med 5 kvartaler. Kapitalutnyttelse beskriver i hvilken grad aktører innenfor norsk industri sektor evner å nyttiggjøre seg av tilgjengelig kapasitet i produksjonsprosessen. Under oppgangstider vil etterspørsel drive behovet for økt utnyttelse av tilgjengelig kapasitet, og skape et press i økonomien. Husebø og Wilhelmsen (2005) bekrefter dens prosykliske egenskaper og kommer frem til at kapasitetsutnyttelse leder økonomisk vekst med ett kvartal. Forbrukertilliten gjenspeiler husholdningenes forventninger til egen og landets økonomi, og vil sammen med industriens kapasitetsutnyttelse kunne gi et godt temperaturmål for aktivitetsnivået i fastlandsøkonomien.

Nasjonale variabler er ikke nødvendigvis tilstrekkelige når de lange konjunkturtrendene i norsk fastlandsøkonomi skal fanges opp. Som en liten åpen økonomi har den norske økonomien alltid vært avhengig av sine handelspartnere, med stor eksport av blant annet olje og fisk. At veksten i norsk økonomi derfor avhenger av utviklingen hos de viktigste handelspartnerne er naturlig.

Antallet utenlandske variabler i figur 6.4 gir et innblikk i hvor avgjørende maskinlæringen vurderer utviklingen i internasjonal økonomi til å være for norsk økonomisk vekst. Den

nomielle styringsrenten til Bank of England (BoE) er representert ved den største og femte største absoluttverdien, gjennom henholdsvis lead og sammenfallende verdi. Rentesettingen til BoE påvirker handlingsmønsteret til britiske bedrifter og husholdninger, noe som kan være viktig for etterspørselen etter norske varer. Endringer i den britiske styringsrenten vil derfor kunne medføre skift i den norske eksportsektoren.

Videre er Conference Board Leading Index og den amerikanske yield-kurven¹⁷ viktige for prediksjonene. USA er en viktig handelspartner, men også den ledende økonomiske makten i verdensøkonomien. Historisk har de to variablene vært ansett som ledende indikatorer for den økonomiske utviklingen i USA. Stadig flere økonomer har undersøkt og bekreftet yield-kurvens egnethet til å predikere fremtidig vekst, men det understrekes at finansielle indikatorer kan gi falske signaler i perioder der finansielle kriser ikke forplanter seg til realøkonomien (Estrella og Mishkin, 1997). Industriproduksjonen for eurosonen er den siste internasjonale variabelen som er representert i figur 6.4. Flere av Norges viktigste handelspartnere, deriblant Sverige, Tyskland og Storbritannia, er medlemmer av EU. At industriproduksjonen i EU derfor er en viktig variabel for prediksjonene virker derfor naturlig.

Avslutningsvis er det viktig å presisere at det foreligger svakheter ved VarImp. I seksjon 6.2.1 diskuterer vi hvordan kryssvalideringen av Elastic Net vektlegger dimensjonsreduksjon. Dermed reduseres koeffisientene til høyt korrelerte variabler, fremfor at noen blir forkastet. Variabler kan gi avgjørende informasjon for prediksjonene, men likevel få tildelt lave koeffisienter grunnet relativ sterk korrelasjon med én eller flere variabler. Koeffisientenes størrelse er dermed ikke ensbetydende med variablenes viktighet. Vi kan derfor ikke trekke slutninger om variablenes faktiske innvirkning på norsk økonomi basert på VarImp resultatene.

6.5 Potensielle mangler ved maskinlæring

Anvendelsen av maskinlæring til prediksjonsformål åpner en rekke muligheter for ulike bransjer, men det er viktig å presisere at utviklingen bringer med seg nye utfordringer. Som vi har vært inne på tilhører maskinlæring kategorien algoritmisk modellering, der fokuset hovedsakelig ligger i å produsere et gitt output basert på læring fra et sett med input. I

¹⁷Rentedifferansen mellom 2 årige og 10 årige amerikanske statsobligasjoner

likhet med mer tradisjonelle modelleringsmetoder ble de første maskinlæringsmodellene bygget opp ved hjelp av instruksjoner mennesker kunne forklare. Innenfor en rekke ønskede bruksområder foreligger det oppgaver som er for store og kompliserte til at det kan skrives enkle tilhørende instruksjoner. Læringsprosessen har dermed blitt automatisert. Over tid har kompleksiteten økt, slik at modellene vanskeligere lar seg forklare. Arbeidsprosessene innen maskinlæring er nemlig ikke bundet til å passe med menneskelig forståelse (Woloszko, 2017). Terminologien «Black box-algoritme» blir stadig mer vanlig, og beskriver modeller med et kjent output og input, men med en ukjent virkemåte. I kjølevannet av denne utviklingen foreligger det en debatt rundt balansen mellom «accuracy» (presisjon) og «interpretability» (transparens). Mer komplekse algoritmer har vist evnen til å produsere langt mer presise prediksjoner som følge av mer avansert læring. Spørsmålet er hvilken verdi presise prediksjoner har innenfor det enkelte bruksområdet hvis den underliggende resonneringen ikke kan forklares.

Selv om våre resultater indikerer at maskinlæring kan fungere som et nyttig verktøy for det makroøkonomiske fagmiljøet, vil hensynet til «interpretability» være et viktig aspekt som foreløpig trekker verdien til maskinlæring ned. I makroøkonomiske miljøer som lever av å gi råd til både private og institusjonelle kunder, er det viktig å kunne underbygge bakgrunnen for prediksjonene. I dette ligger det en opparbeidelse av tillit, som står helt sentralt i ethvert rådgivende kundeforhold. For en investor som bruker makroavdelingen til DNB som sin rådgiver, vil prediksjoner for veksten i norsk økonomi være en viktig forutsetning for hvilke investeringer som velges, og når de gjennomføres. Når rådene skal leveres til en slik kunde, er det avgjørende at makroøkonomene kan komme med en detaljert og velbegrunnet bakgrunn for prediksjonene. I tillegg må makroøkonomen selv kunne stole på at prediksjonen er troverdig, en tillit det er vanskelig å skape uten informasjon om hva prediksjonene har vektlagt.

Transparens i modellens prediksjoner er også avgjørende for Norges Bank. Ved fastsettelsen av både styringsrente og rentebane er hovedstyret avhengige av en forklaring av de underliggende drivkreftene for den predikerte vekstutviklingen i norsk fastlandsøkonomi. En rekke sentralbanker har derfor anvendt DSGE-modeller¹⁸ til pengepolitisk analyse. Til tross for svakheter ved prediksjon, særlig i kriseperioder, tilbyr slike modeller et rammeverk som kan identifisere kilder til fluktusjon og strukturelle brudd, samt predikere effekten

¹⁸Dynamisk, Stokastisk, Generell likevektsmodell

av endringer i pengepolitikken (Tovar, 2009). Norges Banks makroøkonomiske modell, NEMO (Norwegian Economic Model), er et eksempel på en DSGE-modell. Det er den skjønnsmessige vurderingen av SAM prediksjonene som setter de betingede forutsetningene for utarbeidelsen av NEMO (Aastveit et al., 2011).

Løsninger som kan bidra til å håndtere utfordringene tilknyttet black box-algoritmer er under utvikling. Vi har tidligere presentert funksjonen VarImp, som gir en oversikt over de viktigste variablene på tvers av prediksjonene. Den aggregerte tilnærmingen gjør det vanskelig å gjennomføre en dybdeanalyse. Pakkene Lime og Shap er to eksempler på nylig utviklede pakker som forsøker å gjøre det mulig å gå i dybden på mekanismene bak prediksjonene. Begge fungerer som verktøy for å forklare hva som skjer bak utarbeidelsen av den enkelte prediksjon i en black box-algoritme. Lime ble introdusert i 2016 (Ribeiro et al., 2016) og står for «local interpretable model-agnostic explanations». Algoritmen kan forklare prediksjoner i tilknytning til både klassifisering og regresjon. Shapley Additive Explanations (SHAP) utgjør et helhetlig rammeverk av metoder som tilegner uavhengige variabler en verdi for deres individuelle forklaringskraft ved den enkelte prediksjonen (Lundberg og Lee, 2017). I praksis vil pakker som Lime og Shap kunne gjøre det mulig å få en oversikt over bakgrunnen for prediksjonen av et spesifikt kvartal eller år.

Utarbeidelsen av løsninger som LIME og SHAP viser at utviklingen går i retning av å gjøre maskinlæring mer anvendelig til praktiske situasjoner i arbeidslivet. Slike løsninger vil være avgjørende for at maskinlæring skal bli inkludert i utarbeidelsen av makroøkonomiske prediksjoner, siden det skaper tillit og kunnskap rundt metoden. Likevel mener vi resultatene fra denne oppgaven viser hvor langt maskinlæringen er kommet med tanke på presisjon og kvalitet, og at det allerede nå kan tilføre verdi i utarbeidelsen av prediksjoner. Frem til man har en skreddersydd modell tilpasset makroøkonomisk prediksjon, kan modellene brukes som et utgangspunkt for diskusjon og videre vurdering når prediksjoner utarbeides.

7 Konklusjon

I denne oppgaven har vi presentert resultater som viser potensialet til maskinlæring som et nyttig verktøy i utarbeidelsen av prediksjoner for norsk økonomisk vekst. Med utgangspunkt i kvartalsvise observasjoner har Super Learner vist seg å være den beste maskinlæringsmodellen blant de tre vi har anvendt, for både kvartalsvise og årlige prediksjoner. I sammenligningen av Super Learner og Norges Bank på kvartalsvise prediksjoner leverer Super Learner i snitt 19% mer presise prediksjoner, målt ved RSME. For årlige prediksjoner er prestasjonen enda mer imponerende, der Super Learner i snitt har 39% mer presise prediksjoner sammenlignet med Norges Bank, DNB, Finansdepartementet og SSB. Spesielt interessant er det at maskinlæringen utkonkurrerer DNB, som i sine prediksjoner har hatt et informasjonsfortrinn ved å avgi prediksjoner to måneder senere.

En nærmere analyse av de kvartalsvise resultatene har avdekket at Norges Bank produserer de beste prediksjonene i et flertall av kvartalene, men at den totale presisjonen svekkes av store avvik i slutten av 2010 og starten av 2011. Årene markerer kjernen i kjølevannet av finanskrisen, med stor økonomisk usikkerhet og strukturelle endringer. Etter en krise følger som regel en sterk oppgangsperiode, noe som er et tydelig standpunkt ved Norges Bank sine prediksjoner i perioden. Prediksjonene i de avgjørende kvartalene i 2010 og 2011 er alle i overkant optimistiske. Super Learner evner i større grad å predikere tidspunktet for oppgangen, som virkelig slo til mot slutten av 2011. Den samme tendensen ser vi i de årlige prediksjonene, der Norges Bank, Finansdepartementet, SSB og DNB har store avvik i 2011. Alle aktørene leverte for optimistiske prediksjoner. Maskinlæringen leverer jevnt over stabile prediksjoner, som i store deler av perioden ligger nærmere faktisk vekst enn de nevnte aktørers prediksjoner. Spesielt interessant er det at maskinlæringen presterer svært godt i perioden fra 2013 til og med 2016, tatt i betraktning at norsk økonomi opplevde en krevende oljekrise i 2016.

Etter de lovende resultatene er det viktige spørsmålet hva dette kan signalisere om den videre implementeringen av maskinlæring til prognoseformål for norsk økonomi. Et viktig aspekt ved dette er å påpeke eventuelle svakheter ved resultatene vi har presentert. På et overordnet nivå er det svært viktig å poengtere at datagrunnlaget som er brukt som input for maskinlæringsmodellene er reviderte data. Datagrunnlaget representerer på

mange måter mer korrekt informasjon sammenlignet med dataene de ulike aktørene hadde tilgjengelig da deres prediksjoner ble laget. Problemet er vanskelig å finne løsninger på, og må derfor vurderes som et mulig informasjonstrinn for våre prediksjoner. Videre vil det kunne komme potensielle informasjonfordeler fra de ledende variablene vi har inkludert i de kvartalsvise prediksjonene. Prediksjonene inkluderer daglig informasjon fra den siste måneden i kvartal T-1 som ikke var tilgjengelig for Norges Bank i deres prediksjon av kvartal T. Vi mener likevel fordelene ikke påvirker resultatene nevneverdig, ettersom Norges Bank også har et informasjonsfortrinn. Der vår tilnærming utelukker samtlige månedlige variabler fra kvartal T-1 i prediksjon av kvartal T, har Norges Bank informasjon om to av tre måneder i kvartal T-1.

Vi mener våre resultater viser verdien maskinlæring kan tilføre aktører som utarbeider makroøkonomiske prediksjoner. Likevel er vi tydelige på at maskinlæring på nåværende tidspunkt ikke bør erstatte eksisterende modeller og skjønnsmessige vurderinger. Maskinlæring burde heller anses som et supplerende verktøy, som gir et verdifullt og objektivt utgangspunkt i produksjon av prediksjoner for norsk økonomi. Vi håper vår utredning kan bidra til økt oppmerksomhet rundt bruken av maskinlæring i makroøkonomiske sammenhenger. I tillegg ønsker vi å motivere til videre forskning som kan fremme maskinlæringens anvendelighet for makroøkonomiske prognoser.

På sikt er det opp til hver enkelt aktør å utvikle fremgangsmetoder som er best mulig tilpasset prediksjonsformålet. Spesielt viktig vil det være å utvikle metodikker som utnytter det tilgjengelige informasjonsgrunnet, og gjør en fornuftig vektning av observasjoner opp mot variabler. Blant annet vil det være avgjørende at maskinlæring ikke bare egner seg for historisk prediksjon, men leverer presise prediksjoner frem i tid. Når slike problemstillinger er tilstrekkelig håndtert, vil maskinlæring kunne bidra til svært presise prediksjoner for økonomisk vekst. Det blir derfor spennende å se hvor godt Super Learner sin prediksjon for 2020, på 2.66% ¹⁹, sammenfaller med BNP-veksten i Fastlands-Norge.

¹⁹Årlige prediksjoner for 2018, 2019 og 2020 inkludert i appendiks tabell A2.1 for Elastic Net, Random Forest og Super Learner. Prediksjonene baserer seg på nylig publiserte data, og er derfor kun lagt ved som et supplerende element i denne oppgaven. Vårt fokus har vært på prediksjon av kvartalene/årene der faktisk vekst er kjent (for å kunne evaluere prediksjonene) og der data har vært tilgjengelig fra starten av masterarbeidet

Referanser

- Aastveit, K. A., Gerdrup, K. R., og Jore, A. S. (2011). Short-term forecasting of gdp and inflation in real-time: Norges bank's system for averaging models.
- Aegasis AI, L. (2019). *Machine Learning for Portfolio Optimization Trading*. Aegasis AI.
- Aggarwal, S. (2017). Comparative analysis of hedge funds in financial markets using machine learning models. *International Journal of Computer Applications*, 163(3).
- Biau, D., Kernéis, S., og Porcher, R. (2008a). Statistics in brief: The importance of sample size in the planning and interpretation of medical research. *Clinical orthopaedics and related research*, 466:2282–8.
- Biau, G., Devroye, L., og Lugosi, G. (2008b). Consistency of random forests and other averaging classifiers. *Journal of Machine Learning Research*, 9:2015–2033.
- Biau, O. og D'Elia, A. (2010). Euro area gdp forecasting using large survey datasets: A random forest approach.
- Breiman, L. (1994). Bagging predictors.
- Breiman, L. (2001a). Random forests. *Machine Learning*, 45(1):5–32.
- Breiman, L. (2001b). Statistical modeling: The two cultures.
- Brooks, C. (2002). *Introductory econometrics for finance*, volume 2. Cambridge University press.
- Chakraborty, C. og Joseph, A. (2017). Machine learning at central banks.
- Christensen, J. (2019). Dnb markets kontrer: Fem grunner til at du som eier aksjer kan sove som en baby. *Dagens Næringsliv*.
- Chuku, C., Oduor, J., og Simpasa, A. (2017). Intelligent forecasting of economic growth for african economies: Artificial neural networks versus time series and structural econometric models.
- Clark, T. E. og McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105(1):85–110.
- Collins, M. og Rubin, R. (2015). How an exclusive hedge fund turbocharged its retirement plan. *Bloomberg*.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. *International Workshop on Multiple Classifier Systems*, sider 1–15.
- Estrella, A. og Mishkin, F. S. (1997). The predictive power of the term structure of interest rates in europe and the united states: Implications for the european central bank.
- Finansdepartementet (2019). Meld. st. 1 (2019–2020) - nasjonalbudsjettet 2020.
- Fortman-Roe, S. (2012). *Accurately Measuring Model Prediction Error*. Elite Data Science.
- Grytten, O. H. (2011). Financial crises and monetary expansion.

- Grytten, O. H. og Hunnes, A. (2016). *Krakk og kriser i historisk perspektiv*. Cappelen Dam.
- Haakonsen, L. (2015). Kvartar i praksis v.
- Hastie, T., Tibshirani, R., og Friedman, J. (2004). The elements of statistical learning, data mining, inference, and prediction.
- Hastie, T. og Zou, H. (2005). Regularization and variable selection via the elastic net. *Journal of the American Statistical Association*, 67(2):301–320.
- Hendry, D. F. og Clements, M. P. (2004). Pooling of forecasts. *The Econometrics Journal*, 7(1):1–31.
- Holmes, S. (2000). *RMSE*. Stanford University, Stanford, USA.
- Husebø, T. A. og Wilhelmssen, B. T. (2005). Norwegian business cycles 1982-2003. *Staff Memo*, (2).
- Hyndman, R. J. og Athanasopoulos, G. (2018). *Forecasting: principles and practice*, volume 2. OTexts: Melbourne, Australia.
- Hyndman, R. J. og Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688.
- Ishwaran, H. (2007). Variable importance in binary regression trees and forests. *Electronic Journal of Statistics*, 1(1):519–537.
- Juelsrud, R. E. og Wold, E. G. (2019). Effekten av redusert jobbsikkerhet på sparing og sysselsetting – hva kan vi lære av oljekrisen? *Bankplassen blogg*.
- Jung, J.-K., Patnam, M., og Ter-Martirosyan, A. (2018). An algorithmic crystal ball: Forecasts-based on machine learning.
- Kuhn, M. (2008). Building predictive models in r using the caret package. *Journal of Statistical Software*, 28(5).
- Lin, Y. og Jeon, Y. (2006). Random forests and adaptive nearest neighbors. *Journal of the American Statistical Association*, 101:578–590.
- Lundberg, S. M. og Lee, S.-I. (2017). A unified approach to interpreting model predictions.
- Mullainathan, S. og Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106.
- Norman, V. D. og Orvedal, L. (2012). *En liten, åpen økonomi*, volume 3. Gyldendal.
- norske leksikon, S. (2018). Hyperplan.
- Olsen, (2019). Pengepolitikkenes verktøykasse. *Bankplassen blogg*.
- Ribeiro, M. T., Singh, S., og Guestrin, C. (2016). “why should i trust you?” explaining the predictions of any classifier.
- Science, E. D. (2019). *Overfitting in Machine Learning: What It Is and How to Prevent It*. Elite Data Science.
- SSB (2019). Hva er egentlig bnp?

- Stark, T. og Croushore, D. (2002). Forecasting with a real-time dataset for macroeconomists. *Journal of Macroeconomics*, 24(4):507–531.
- Sucarrat, G. (2019). *Samfunnsøkonomenes prognosepris*.
- Sucarrat, G. og Gharsallah, S. (2019). How precise are the forecasts of the norwegian national budget?
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288.
- Tiffin, A. (2016). Seeing in the dark: A machine-learning approach to nowcasting in lebanon.
- Timmermann, A. (2007). An evaluation of the world economic outlook forecasts. 54(1):1–33.
- Tkacz, G. og Hu, S. (1999). Forecasting gdp growth using artificial neural networks. (99).
- Tovar, C. E. (2009). Dsge-models and central banks. *Economics E-journal*, 3.
- Valland, R. A. (2019). Nowcasting av norsk bnp med maskinlæringsalgoritmer. Master's thesis, Universitetet i Bergen.
- Van der Laan, M. J. og Polley, E. C. (2010). Super learner in prediction. (266).
- Van der Laan, M. J., Polley, E. C., og Hubbard, A. E. (2007). Super learner.
- Vapnik, V. og Cortes, C. (2005). Support vector networks. *Kluwer Academic Publishers*, 20(3):273–297.
- Von Brasch, T. (2019). Hvorfor lager ssb prognoser for utvikling i norsk økonomi.
- Walimbe, R. (2017). *Avoiding Look-Ahead Bias in Time Series Modelling*. Data Science Central.
- Watson, M. W. og Stock, J. H. (2016). Factor models and structural vector autoregressions in macroeconomics.
- Woloszko, N. (2017). Making better economic forecasts with machine learning.
- Woloszko, N. (2018). Adaptive trees: A novel approach to macroeconomic forecasting.
- Wooldridge, J. M. (2012). *Introductory Econometrics: A Modern Approach*, volume 5. Cengage Learning.

Appendiks

A1 Data

A1.1 Prediksjoner innhentet fra andre aktører

Tabell A1.1: Kvartalsvise prediksjoner Norges Bank

Prediksjonskvartal	Prediksjon	Prediksjonstidspunkt	Kilde
3.kvartal 2018	2.67	20/06/18	Pengepolitisk rapport
2.kvartal 2018	2.77	21/03/18	Pengepolitisk rapport
1.kvartal 2018	2.46	13/12/17	Pengepolitisk rapport
4.kvartal 2017	2.54	20/09/17	Pengepolitisk rapport
3.kvartal 2017	2.23	21/06/17	Pengepolitisk rapport
2.kvartal 2017	1.36	15/03/17	Pengepolitisk rapport
1.kvartal 2017	Manglende	-	Pengepolitisk rapport
4.kvartal 2016	Manglende	-	Pengepolitisk rapport
3.kvartal 2016	0.84	23/06/16	Pengepolitisk rapport
2.kvartal 2016	0.51	17/03/16	Pengepolitisk rapport
1.kvartal 2016	0.81	17/12/15	Pengepolitisk rapport
4.kvartal 2015	1.01	24/09/15	Pengepolitisk rapport
3.kvartal 2015	1.23	18/06/15	Pengepolitisk rapport
2.kvartal 2015	1.25	19/03/15	Pengepolitisk rapport
1.kvartal 2015	2.30	11/12/14	Pengepolitisk rapport
4.kvartal 2014	2.42	18/09/14	Pengepolitisk rapport
3.kvartal 2014	1.93	19/06/14	Pengepolitisk rapport
2.kvartal 2014	1.92	27/03/14	Pengepolitisk rapport
1.kvartal 2014	1.71	05/12/13	Pengepolitisk rapport
4.kvartal 2013	1.88	19/09/13	Pengepolitisk rapport
3.kvartal 2013	2.17	20/06/13	Pengepolitisk rapport
2.kvartal 2013	2.75	14/03/13	Pengepolitisk rapport
1.kvartal 2013	2.65	14/03/13	Pengepolitisk rapport
4.kvartal 2012	3.36	31/10/12	Pengepolitisk rapport
3.kvartal 2012	3.70	22/06/12	Pengepolitisk rapport
2.kvartal 2012	2.93	14/03/12	Pengepolitisk rapport
1.kvartal 2012	3.44	19/10/11	Pengepolitisk rapport
4.kvartal 2011	3.10	19/10/11	Pengepolitisk rapport
3.kvartal 2011	3.20	22/06/11	Pengepolitisk rapport
2.kvartal 2011	3.15	16/03/11	Pengepolitisk rapport
1.kvartal 2011	2.85	27/10/10	Pengepolitisk rapport
4.kvartal 2010	2.39	27/10/10	Pengepolitisk rapport
3.kvartal 2010	2.01	23/06/10	Pengepolitisk rapport
2.kvartal 2010	1.50	23/06/10	Pengepolitisk rapport
1.kvartal 2010	2.29	28/10/09	Pengepolitisk rapport

Tabell A1.2: Årlige prediksjoner fra ulike aktører

Aktør	År	Prediksjon ¹	Dato	Kilde
Norges Bank	2018	2.00	15/09/17	PPR nr.3 s.59
Norges Bank	2017	1.80	16/09/16	PPR nr.3 s.63 og 64
Norges Bank	2016	1.25	18/09/15	PPR nr.3 s.58 og 59
Norges Bank	2015	2.25	11/09/14	PPR nr.3 s.46 og 47
Norges Bank	2014	2.25	12/09/13	PPR nr.3 s.35 og 51
Norges Bank	2013	3.00	25/10/12	PPR nr.3 s.25 og 39
Norges Bank	2012	3.75	13/10/11	PPR nr.3 s.25 og 47
Norges Bank	2011	3.00	21/10/10	PPR nr.3 s.28 og 39
Norges Bank	2010	2.75	22/10/09	PPR nr.3 s.25 og 47
Finansdepartementet	2018	2.50	29/09/17	Nasjonalbudsjettet 2018 s.18
Finansdepartementet	2017	1.70	23/09/16	Nasjonalbudsjettet 2017 s.17
Finansdepartementet	2016	1.80	25/09/15	Nasjonalbudsjettet 2016 s.17
Finansdepartementet	2015	2.00	26/09/14	Nasjonalbudsjettet 2015 s.26
Finansdepartementet	2014	2.70	04/10/13	Nasjonalbudsjettet 2014 s.29
Finansdepartementet	2013	2.50	28/09/12	Nasjonalbudsjettet 2013 s.26
Finansdepartementet	2012	3.10	30/09/11	Nasjonalbudsjettet 2012 s.26
Finansdepartementet	2011	3.10	01/10/10	Nasjonalbudsjettet 2011 s.30 og 31
Finansdepartementet	2010	2.10	09/10/09	Nasjonalbudsjettet 2010 s.31 og 32
SSB	2018	2.10	05/09/17	Økonomiske analyser 2017 nr.3 s.1 og 30
SSB	2017	2.10	13/09/16	Økonomiske analyser 2016 nr.4 s.19
SSB	2016	1.80	01/09/15	Økonomiske analyser 2015 nr.3 s.17
SSB	2015	2.10	02/09/14	Økonomiske analyser 2014 nr.4 s.15
SSB	2014	3.00	24/09/13	Økonomiske analyser 2013 nr.4 s.16
SSB	2013	3.10	04/09/12	Økonomiske analyser 2012 nr.4 s.46
SSB	2012	3.50	06/09/11	Økonomiske analyser 2011 nr.4 s.42
SSB	2011	2.90	31/08/10	Økonomiske analyser 2010 nr.4 s.52
SSB	2010	2.10	01/09/09	Økonomiske analyser 2009 nr.4 s.52
DNB	2017	1.30	06/12/16	Økonomiske utsikter 2017 s.28
DNB	2016	1.20	09/12/15	Økonomiske utsikter 2016 s.35 og 37
DNB	2015	1.20	16/12/14	Økonomiske utsikter 2015-2018 s.37 og 39
DNB	2014	2.00	16/12/13	Makroutsikter 2014 s.25 og 32
DNB	2013	2.40	12/12/12	Prognoseprisen 2013, innhentet via epost
DNB	2012	2.50	19/12/11	2012 Forecasts s. 21-22
DNB	2011	3.00	16/12/10	Hva skjer i 2011? s. 22
DNB	2010	2.00	22/12/09	Hva skjer i 2010? s. 20

¹ Årlig vekst i %

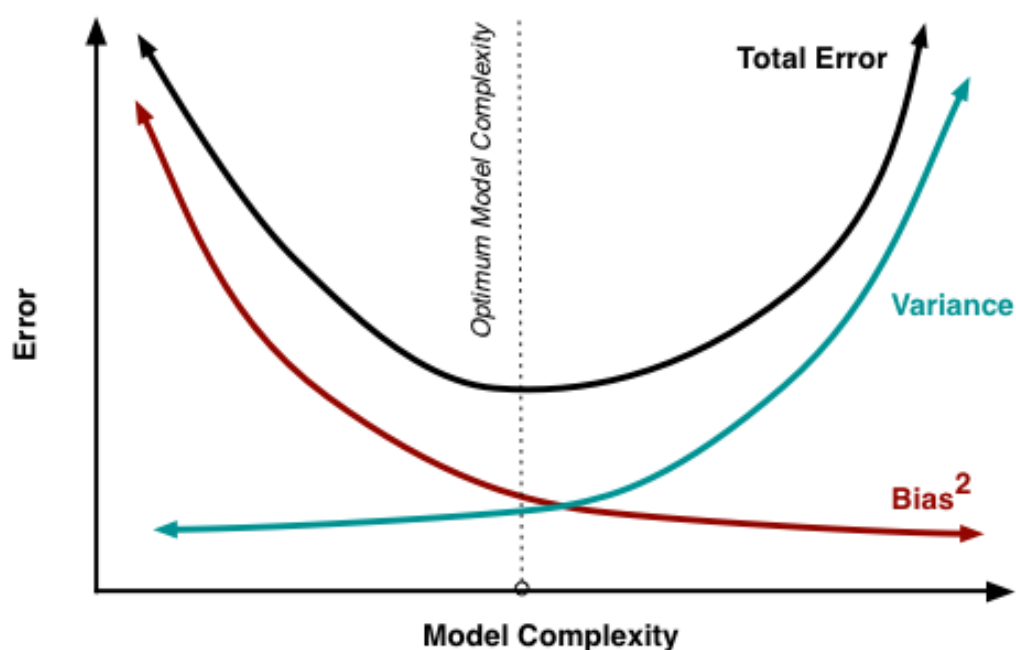
Tabell A1.3: Faktisk BNP vekst Fastlands-Norge 2018-2010

Aktør	År	BNP-vekst %	Dato	Kilde
SSB	2018	2.20	03/09/19	SSB statistikkbank Tabell 09189
SSB	2017	2.00	03/09/19	SSB statistikkbank Tabell 09189
SSB	2016	0.90	03/09/19	SSB statistikkbank Tabell 09189
SSB	2015	1.40	03/09/19	SSB statistikkbank Tabell 09189
SSB	2014	2.20	03/09/19	SSB statistikkbank Tabell 09189
SSB	2013	2.30	03/09/19	SSB statistikkbank Tabell 09189
SSB	2012	3.70	03/09/19	SSB statistikkbank Tabell 09189
SSB	2011	1.90	03/09/19	SSB statistikkbank Tabell 09189
SSB	2010	1.90	03/09/19	SSB statistikkbank Tabell 09189
Prognoseprisen	2017	2.00	28/11/18	SSB statistikkbank
Prognoseprisen	2016	1.10	06/12/17	SSB statistikkbank
Prognoseprisen	2015	1.10	22/12/16	SSB statistikkbank
Prognoseprisen	2014	2.40	09/12/15	SSB stahtistikkbank
Prognoseprisen	2013	2.00	16/12/14	SSB statistikkbank
Prognoseprisen	2012	3.40	06/12/13	SSB, Økonomiske Analyser 2013 nr.5 s. 13
Prognoseprisen	2011	2.50	06/12/12	SSB, Økonomiske Analyser 2012 nr.6 s. 12
Prognoseprisen	2010	1.90	08/12/11	SSB, Økonomiske Analyser 2011 nr.6 s. 11

A2 Figurer

A2.1 Data og metode

Figur A2.1: Bias vs. Variance - Overfitting



A2.2 Utarbeidelsen av kvartalsvise prediksjoner

Figur A2.2: Kryssvalidering Elastic Net

```

glmnet
91 samples
81 predictors

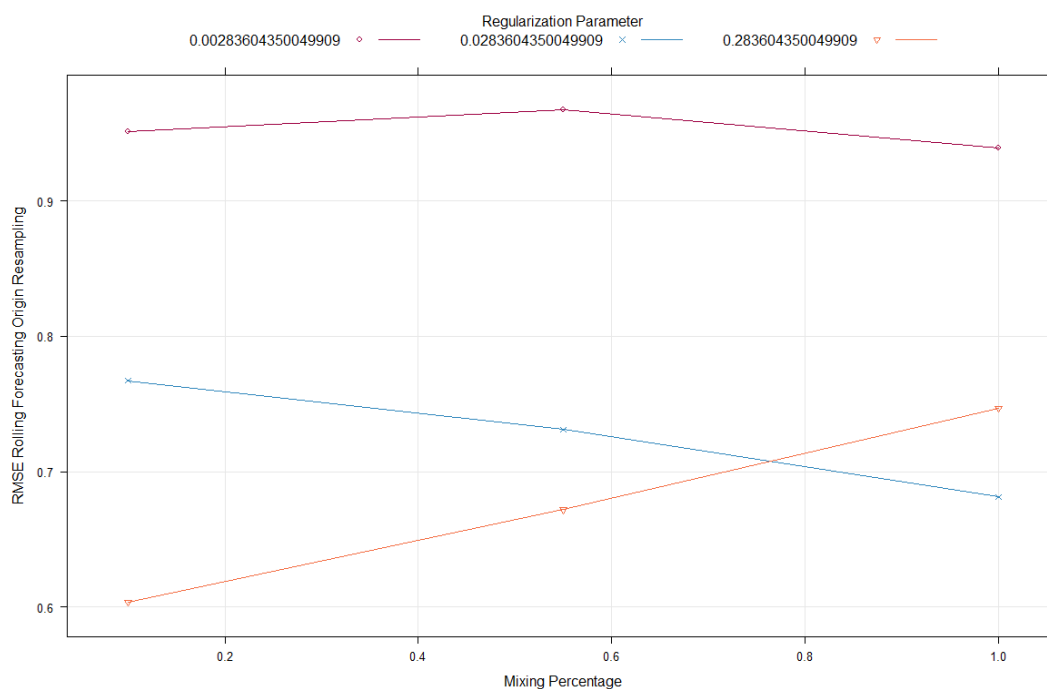
Pre-processing: centered (80), scaled (80)
Resampling: Rolling Forecasting Origin Resampling (2 held-out with no fixed window)
Summary of sample sizes: 55, 56, 57, 58, 59, 60, ...
Resampling results across tuning parameters:

alpha  lambda      RMSE      Rsquared  MAE
0.10   0.002836044  0.9511389  1         0.8823961
0.10   0.028360435  0.7672261  1         0.7049859
0.10   0.283604350  0.6036221  1         0.5608342
0.55   0.002836044  0.9674125  1         0.8872330
0.55   0.028360435  0.7310933  1         0.6623722
0.55   0.283604350  0.6720216  1         0.6167072
1.00   0.002836044  0.9389376  1         0.8561129
1.00   0.028360435  0.6812837  1         0.6223817
1.00   0.283604350  0.7467856  1         0.6793809

RMSE was used to select the optimal model using the smallest value.
The final values used for the model were alpha = 0.1 and lambda = 0.2836044.

```

Figur A2.3: Kryssvalidering Elastic Net - Plot av kombinasjoner gitt RMSE



Figur A2.4: Kryss-validering Random Forest

Random Forest

91 samples
81 predictors

Pre-processing: centered (80), scaled (80)

Resampling: Rolling Forecasting Origin Resampling (2 held-out with no fixed window)

Summary of sample sizes: 55, 56, 57, 58, 59, 60, ...

Resampling results across tuning parameters:

mtry	RMSE	Rsquared	MAE
2	0.8183744	1	0.7664976
41	0.7406537	1	0.6947043
80	0.8041472	1	0.7513176

RMSE was used to select the optimal model using the smallest value.

The final value used for the model was mtry = 41.

Figur A2.5: Kryss-validering Super Learner - vekting av modeller

Call: SuperLearner(Y = y.train, X = x.train, family = gaussian(), SL.library = SL.library, verbose = FALSE)

	Risk	Coef
SL.glmnet_A11	1.105654	0.35538727
SL.gam_A11	NA	0.00000000
SL.glm_A11	484.746455	0.00000000
SL.nnet_A11	3.897761	0.00000000
SL.svm_A11	1.127400	0.51476366
SL.bayesglm_A11	13.702741	0.01001457
SL.polymars_A11	1.723287	0.11983450
SL.rpart_A11	NA	0.00000000
SL.mean_A11	3.945947	0.00000000
SL.randomForest_A11	NA	0.00000000

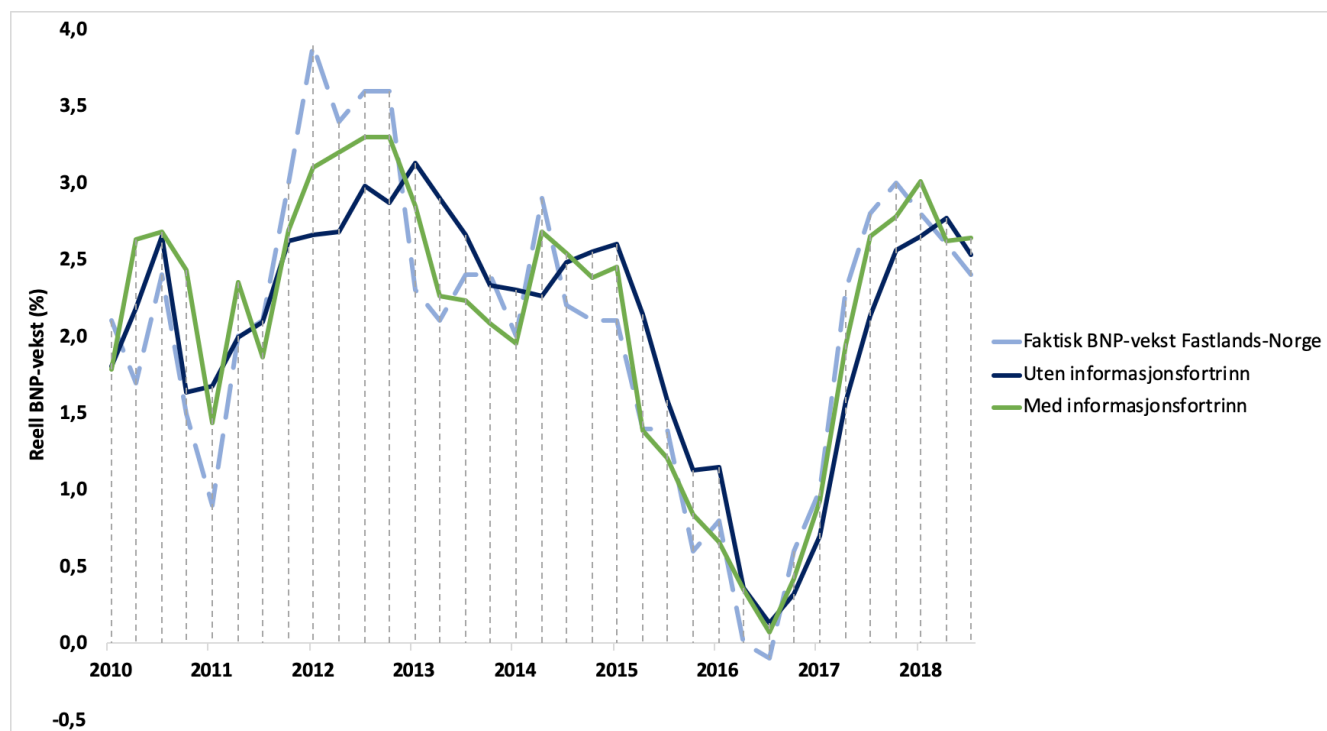
A2.3 Utarbeidelsen av årlige prediksjoner

Tabell A2.1: Prediksjoner for årlig vekst, 2018-2020

År	Super Learner	Elastic Net	Random Forest
2018	1.96	2.52	2.11
2019	2.63	2.29	2.17
2020	2.66	2.41	2.57

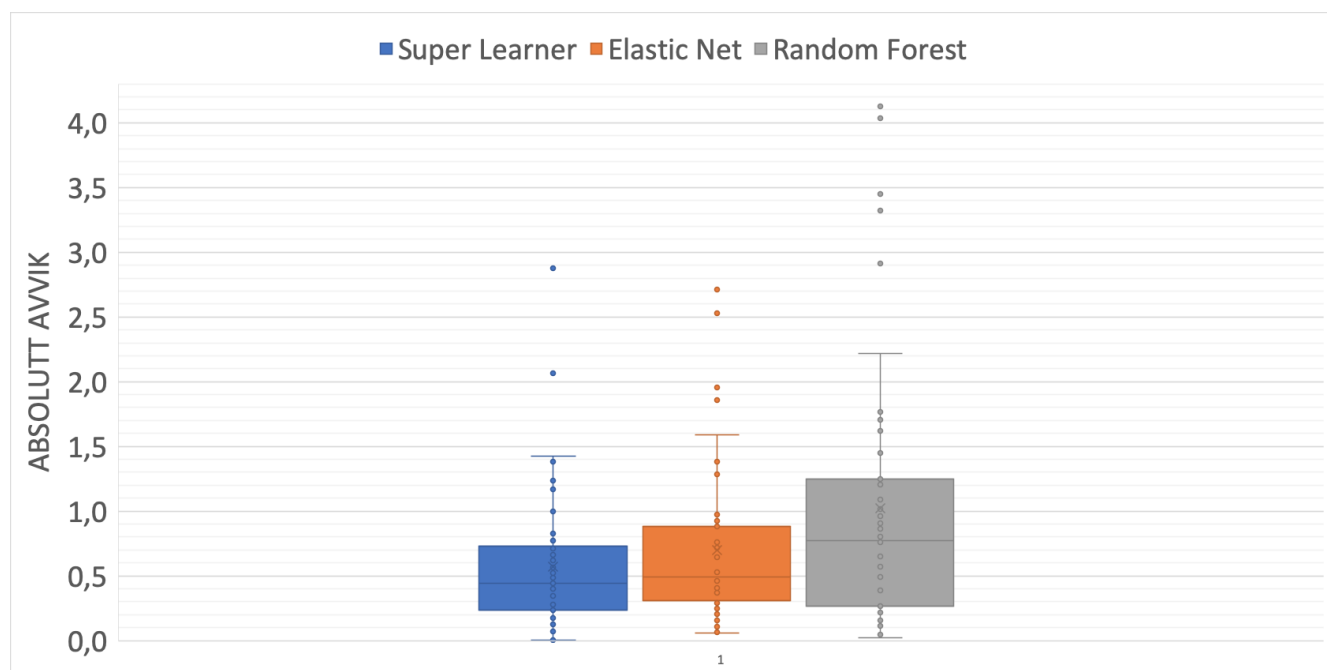
A2.4 Øvrige figurer

Figur A2.6: Prognoser av Super Learner med og uten informasjonsfortrinn



* Kvartalsvise prediksjoner for periode T. Informasjonsfortrinn innebærer prediksjoner fra kvartal T-1. Prediksjoner fra kvartal T-2 omgår look-ahead-bias

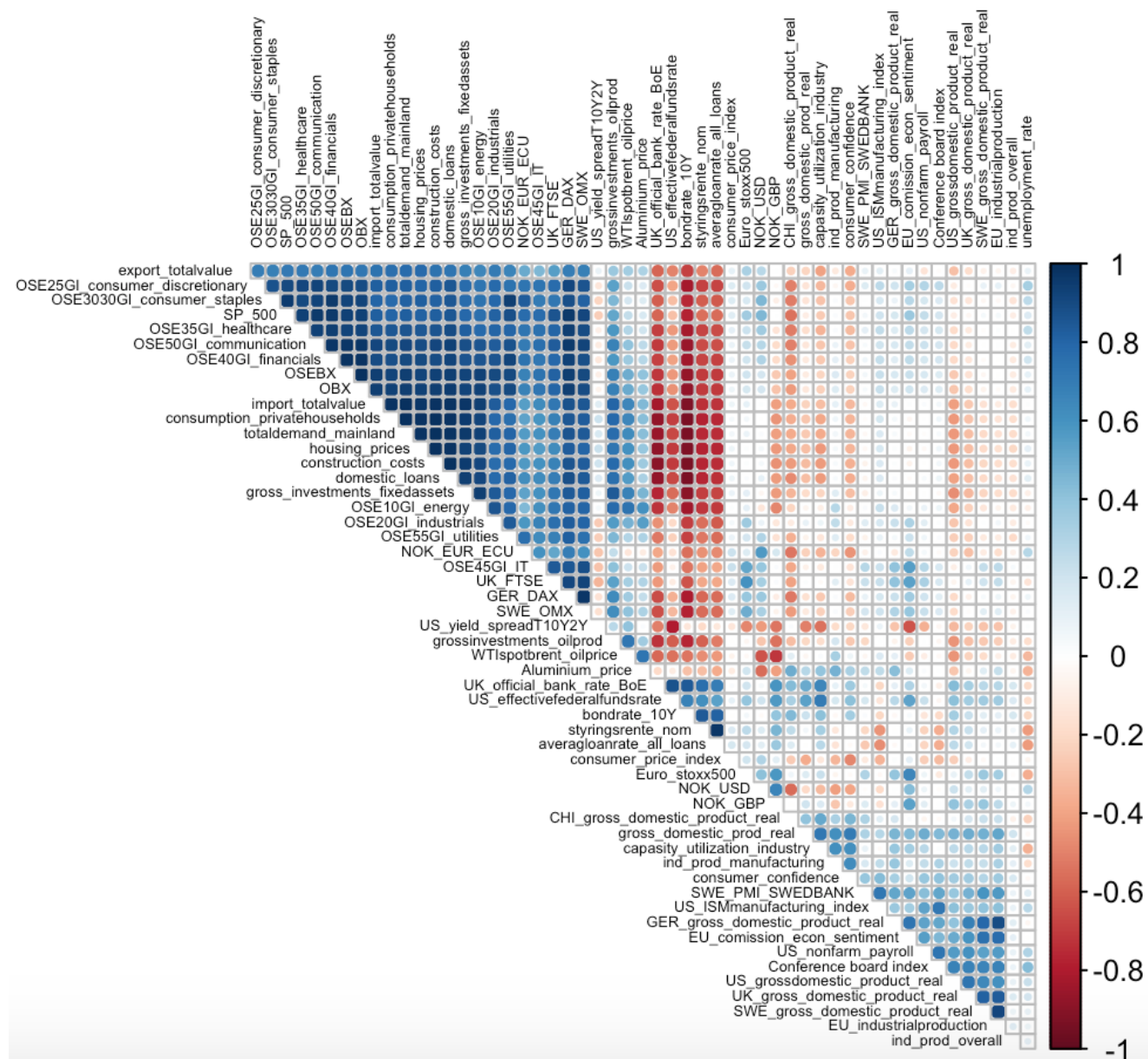
Figur A2.7: Fordeling av absolutte avvik i prediksjoner fra 2007 til 2018



* Absolutte avvik for kvartalsvise prediksjoner.

A3 Variabler

Figur A3.1: Korrelasjonsmatrise



* Korrelasjonsmatrise for datasettets 54 variabler. Fargekoder varierer fra rød (perfekt negativ korrelasjon) til blå (perfekt positiv korrelasjon)

Tabell A3.1: Norske forklaringsvariabler i datasettet

Område	Variabel	Type	Enhet	Kilde	Frekvens
Norge	Produksjonsindeks, totalt	Industri	% , å/å	Bloomberg	m
Norge	Produksjonsindeks, industri	Industri	% , å/å	Bloomberg	m
Norge	Kapasitetsutnyttelse	Industri	%	SSB	k
Norge	Vekslingskurs, NOK/EUR	Valuta	Kurs	IMF	d
Norge	Vekslingskurs, NOK/USD	Valuta	Kurs	IMF	d
Norge	Vekslingskurs, NOK/GBP	Valuta	Kurs	IMF	d
Norge	Arbeidsledighet	KPI	%	IMF	m
Norge	Konsumprisindeks	KPI	% , å/å	IMF	m
Norge	Boligpriser	KPI	Indeks	Bloomberg	k
Norge	Nominell styringsrente	Renter	%	Norges Bank	d
Norge	Statsobligasjon, 10 år	Renter	%	Norges Bank	d
Norge	Gj.snitt lånerente, alle lån	Renter	%	SSB	d
Norge	Konsum, husholdninger	Nasjonalt	Indeks	SSB	m
Norge	Brutto oljeinvesteringer	Nasjonalt	Indeks	SSB	m
Norge	Bruttoinvesteringer i realkapital	Nasjonalt	Indeks	SSB	m
Norge	Total etterspørsel, fastland	Nasjonalt	Indeks	SSB	m
Norge	Eksport, totalt	Nasjonalt	Indeks	SSB	m
Norge	Import, totalt	Nasjonalt	Indeks	SSB	m

* Frekvenser inndelt i daglig (d), månedlig (m) og kvartalsvis (k)

Tabell A3.2: Norske forklaringsvariabler i datasettet

Område	Variabel	Type	Enhet	Kilde	Frekvens
Norge	Innenlandsk lånegjeld i NOK	Nasjonal	Absolutt	SSB	m
Norge	Konstruksjonskostnad	Eiendom	Indeks	Bloomberg	m
Norge	Forbrukertillit	Survey	Indeks	Finans Norge	k
Norge	OSEBX, hovedindeks	Finansiell	Indeks	Oslo Børs	d
Norge	OBX, 25 mest likvide aksjer	Finansiell	Indeks	Bloomberg	d
Norge	OSE25GI, konsumvarer	Finansiell	Indeks	Bloomberg	d
Norge	OSE3030GI, forbruksvarer	Finansiell	Indeks	Bloomberg	d
Norge	OSE10GI, energi	Finansiell	Indeks	Bloomberg	d
Norge	OSE40GI, finans	Finansiell	Indeks	Bloomberg	d
Norge	OSE35GI, helsevern	Finansiell	Indeks	Bloomberg	d
Norge	OSE20GI, industri	Finansiell	Indeks	Bloomberg	d
Norge	OSE45GI, IT	Finansiell	Indeks	Bloomberg	d
Norge	OSE50GI, kommunikasjon	Finansiell	Indeks	Bloomberg	d
Norge	OSE55GI, forsyning	Finansiell	Indeks	Bloomberg	d

* Frekvenser inndelt i daglig (d), månedlig (m) og kvartalsvis (k)

Tabell A3.3: Internasjonale forklaringsvariabler i datasettet

Område	Variabel	Type	Enhet	Kilde	Frekvens
Tyskland	Reell BNP	Nasjonal	%, å/å	Bloomberg	k
Tyskland	DAX, hovedindeks	Finansiell	Indeks	Bloomberg	d
Sverige	Reell BNP	Nasjonal	%, å/å	IMF	k
Sverige	PMI Swedbank	Survey	Indeks	Bloomberg	m
Sverige	OMX, hovedindeks	Finansiell	Indeks	Bloomberg	d
UK	Reell BNP	Nasjonal	%, å/å	IMF	k
UK	BoE base rate	Renter	%	Bank of England	d
UK	FTSE, hovedindeks	Finansiell	Indeks	Bloomberg	d
USA	Reell BNP	Nasjonal	%, å/å	IMF	k
USA	Effective Federal Funds Rate	Renter	%	Federal Reserves	d
USA	Yield-spread (10Y-2Y)	Renter	%	Federal Reserves	d
USA	ISM Manufacturing Index	Survey	Indeks	Bloomberg	m
USA	Conference Board Leading Index	Survey	%, å/å	Bloomberg	m
USA	Standard & Poor 500	Finansiell	Indeks	Bloomberg	d
EU	Produksjonsindeks, Industri	Industri	%, å/å	Bloomberg	m
EU	Economic Sentiment	Survey	Indeks	Bloomberg	m
EU	Euro Stoxx 500	Finansiell	Indeks	Bloomberg	d
Kina	Reell BNP	Nasjonal	%, å/å	Bloomberg	k
Internasjonal	WTI Spot	Råvare	USD/fat	Bloomberg	d
Internasjonal	Aluminium	Råvare	USD/MT	Bloomberg	d

* Frekvenser inndelt i daglig (d), månedlig (m) og kvartalsvis (k)

Tabell A3.4: Deskriptiv statistikk av utvalgte variabler fra 1996 til 2018

Variabel	Min	Maks	Median	Mean	Std
BNP					
Norge	-2.90	7.10	2.40	2.71	1.96
Sverige	-5.30	8.10	3.00	2.58	2.36
Tyskland	-7.03	5.93	1.75	1.47	2.14
UK	-5.93	4.79	2.46	2.19	1.85
USA	-3.92	6.85	2.6	2.66	1.91
Kina	6.4	15.0	8.8	9.09	2.06
Renter					
Norges Banks styringsrente	0.50	8.00	2.25	3.20	2.19
Norsk gj.snittlig lånerente, alle lån	3.36	9.72	4.79	5.60	1.84
ECB Official Interest rate	-0.40	3.75	1.00	1.07	1.28
BoE Base Rate	0.25	7.50	4.00	3.18	2.48
FED Effective funds rate	0.07	6.52	1.45	2.38	2.25
Survey					
Norsk forbrukertillit	-15.50	34.91	18.99	17.51	12.26
Svensk PMI Swedbank	34.0	64.50	55.10	54.60	5.94
ISM Manufacturing	34.50	60.60	53.50	52.74	4.74
Conference Board Index (LEI)	-19.83	10.80	2.80	1.57	5.85
EU Economic Sentiment	70.07	117.10	101.37	100.89	9.26
Annet					
Norsk arbeidsledighet	2.20	5.0	3.57	3.65	0.67
Norsk konsumprisindeks	-1.43	4.70	2.15	2.08	1.01
Norsk kapasitetsutnyttelse	75.90	85.20	79.60	79.90	2.27
WTI Spot oljepris	12.93	123,78	49.14	54.52	29.35