

# **Nowcasting av norsk BNP med maskinlæringsalgoritmer**

Rene Alexander Valland

**Masteroppgave**

Masteroppgaven er levert for å fullføre graden

**Master i samfunnsøkonomi**

Universitetet i Bergen, Institutt for økonomi

September 2019



UNIVERSITETET I BERGEN

## **Forord**

Hva som driver økonomisk vekst, og arbeidet som går inn i å predikere dette er noe som har fascinert meg lenge. Samtidig er det kommet nye datadrevne metoder å utføre arbeidet på. I denne oppgaven benytter jeg det jeg har lært de siste årene til å predikere økonomisk vekst med maskinlæring. Det har vært en reise jeg har lært mye av. Stor takknemlighet rettes til veileder Arild Aakvik for støtte og verdifull feedback.

## Sammendrag

Hvordan det går med økonomien er noe som aktører i finans, bedrifter og det offentlige tjener på å vite så fort som mulig. Blant de viktigste makrotallene er veksten i bruttonasjonalprodukt, som viser verdien av alt som produseres i et land over en periode, minus varer og tjenester som blir brukt under denne produksjonen. Det er en indikator for et lands samlede verdiskaping, og gir uttrykk for opptjent bruttoinntekt for innenlands produksjon. En endring i BNP-vekst har stor betydning for befolkningens optimisme rundt deres fremtidige økonomiske situasjon, arbeidsledighet, finanspolitikk og pengepolitikk. Det er nødvendig å kjenne økonomiens tilstand og utvikling for å tilrettelegge stabiliserende tiltak til rett tid.

I denne oppgaven tester vi om, og hvordan maskinlæring kan bidra til å forbedre prediksjoner på inneværende kvartals BNP-vekst over ulike tidspunkt i perioden. Vi vurderer treffsikkerheten til tre maskinlæringsmodeller opp mot tre tradisjonelle på et datasett med over 120 tidsserier, ved å predikere BNP-vekst for fastlands-Norge. Funnene i denne oppgaven er tredelt, først finner vi at maskinlæringsmetoder ikke nødvendigvis er bedre enn tradisjonelle metoder, men at prediksjon kan forbedres ved å inkorporere maskinlæringsmodeller med andre mye brukte modeller innen litteraturen, i en samlemodell. For det andre finner vi at å slå sammen modellene med vekting basert på historisk treffsikkerhet ikke nødvendigvis gir bedre prediksjonsevne enn et enkelt gjennomsnitt. Når vi evaluerer samlemodellenes prediksjonsevne på ulike tidspunkt, oppdager vi også at den optimale kombinasjonen av modeller og vekting varierer med hvilke makrovariabler som er nylig publisert, og prediksjonshorisont.

## Innhold

1. Introduksjon .....	5
1.1 Bakgrunn .....	5
1.2 Litteratur.....	6
2. Teori.....	7
2.1 Faktoranalyse .....	7
2.1.1 Prinsipalkomponentanalyse .....	8
2.2 Random Forest .....	9
2.2.1 Bootstrap sampling og aggregering.....	9
2.2.3 Beslutningstrær .....	9
2.3 Support Vector Machine .....	11
2.4 Nevral Nettverk.....	12
2.5 Modellkombinering .....	13
3.0 Data .....	15
3.1 Stasjonarisering .....	17
3.2 Metode .....	17
4.0 Resultater .....	17
4.1 Første måned.....	18
4.2 Andre måned .....	25
4.3 Tredje måned .....	32
4.4 Oppsummering.....	38
5.0 Konklusjon .....	40
Referanseliste.....	42
Appendix.....	45
A.1 Variabelviktighet .....	45
A.2 Tidsserietransformasjon.....	46
A.3 Prosent av total varians forklart ved komponenter i faktormodell .....	48
A.4.1 Variabelviktighet i random forest, første måned.....	49
A.4.2 Variabelviktighet i random forest, andre måned.....	50
A.4.3 Variabelviktighet i random forest, tredje måned .....	51

## **1. Introduksjon**

I oppgaven undersøker jeg om tre utvalgte maskinlæringsalgoritmer kan bidra til å gjøre korttidsprognoser for veksten i bruttonasjonalprodukt mer treffsikker. Tradisjonelle prognosemetoder bygger på en a priori antagelse om forholdet mellom variablene, det innebærer da sterke begrensninger når en ønsker å avdekke den egentlige prosessen fra økonomisk aktivitet til BNP. Jeg viser hvordan en datadreven metode med maskinlæring kan benyttes til å lage prognoser for veksten i Norge, uten å måtte på forhånd spesifisere forholdet mellom variablene. Jeg tar ML-modellene random forest, nevralt nettverk og support vector machine, og anvender disse på et stort datasett med makroøkonomiske variabler til å predikere kvartalsvis BNP-vekst. Jeg sammenligner så resultater med en annen mye brukt metode – faktormodell. Resultater fra modellene slås sammen i en modell for korttidsprognose, etter Norges Banks System for Averaging Models (SAM), og samlet skal disse gi et best mulig vekstestimat.

### **1.1 Bakgrunn**

Hvordan det går med økonomien er noe som aktører i finans, bedrifter og det offentlige tjener på å vite så fort som mulig. Blant de viktigste makrotallene er bruttonasjonalprodukt, som viser verdien av alt som produseres i et land over en periode, minus varer og tjenester som blir brukt under denne produksjonen. Det er en indikator for et lands samlede verdiskaping, og gir uttrykk for opptjent bruttoinntekt for innenlands produksjon. En endring i BNP-vekst har stor betydning for befolkningens optimisme rundt deres fremtidige økonomiske situasjon, arbeidsledighet, finanspolitikk og pengepolitikk. Det er nødvendig å kjenne økonomiens tilstand og utvikling for å tilrettelegge stabiliserende tiltak til rett tid. Likevel publiseres ikke økonomisk vekst tidligere enn 40 dager etter at den rapporterte perioden er over. Prediksjon av bruttonasjonalprodukt er derfor av interesse, og i de siste årene har feltet opplevd utvikling

i teknologi og nye metoder. Maskinlæring og stordata er ikke nytt, men krever mye datakraft. I de senere årene har teknologien gjort ML-metodene praktiske også til bruk i makroøkonomi (Varian, 2014).

## 1.2 Litteratur

Der er mange måter å måle og predikere økonomisk vekst på, og det skrives stadig om nye metoder. Siden BNP publiseres kvartalsvis, ender vi opp med korte tiddserier. På den annen side har vi på svært mange serier for relevante finansvariabler, sektorverdier og makrotall med månedlige og daglige publiseringer. Det vi ender opp med er et kort, men bredt datasett, der prediksjonen blir mer treffsikker jo mer betimelig data en har tilgang på. Viktigheten ved å ha mest mulig oppdaterte tidsserier er størst jo nærmere nåtid en lager prognoser for. Det nærmeste en kommer er nowcasting, der en måler økonomisk tilstand for perioden en befinner seg i. Tidligere fremgangsmåter for å få med seg data så snart det er publisert har vært å transformere månedlige tidsserier til å bli kvartalsvis ved å bruke bridging (Baffigi og Golinelli 2004) og mixed data sampling (Ghysels mfl 2004) i nowcasting. Faktormodeller er populære fordi de greier å håndtere store mengder asynkrone data, og støter ikke på fellene som kan oppstå ved å regne ut kvartalsvis gjennomsnitt over flere perioder (Aastveit og Trovik, 2012).

Teräsvirta mfl. (2005) er tidlig ute med å anvende nevralt nettverk når de bygger prognosemodeller for makroøkonomiske tidsserier. Makridakis og Spiliotis (2018) sammenligner prognosenøyaktighet ved ulike populære maskinlæringsmetoder mot åtte ulike tradisjonelle benchmarks. De finner at out-of-sample nøyaktighet ved ML er lavere enn ved tradisjonelle metoder. Chakraborty og Joseph (2017) gjør en lignende analyse og finner at ML stort sett gjør det bedre enn tradisjonelle metoder. Random forest er mer kjent som klassifiseringsverktøy, og er mye brukt innen medisin. Den er også godt egnet til regresjon og estimering av økonomisk vekst, som brukt i Biau & D'elia (2012) til å lage prognoser for BNP i EU, og Nyman & Ormerod (2016) til bruk i Storbritannia og USA. Kim (2003) får lovende resultater når han bruker Support Vector Machine i prediksjon på aksjemarkedet. Videre gir SVM og nevralt nettverk gode resultater i prediksjon av BNP i ganske nylig (Richardson mfl, 2018) og tidligere litteratur (Teräsvirta mfl 2005, Ahmed mfl 2010).

Selv om noen modeller historisk gjør det bedre enn andre, er det likevel fare for estimeringsfeil som følge av støy, forventingsskjevhet og varians. Ved å slå ulike modeller sammen i et ensemble kan disse feilkildene minimeres (Stock og Watson, 2004, Hendry og Clements, 2004, og Timmermann, 2006). En fremgangsmåte for modellsamling vi tar utgangspunkt i er den Norges Bank benytter. NB publiserer predikert BNP kvartalsvis i Pengepolitisk rapport for inneværende kvartal (siste tre månedene fra publikasjon), og ett kvartal fremover. De bruker en rekke modeller når de predikerer brutto nasjonalprodukt. I et av Norges Banks Staff Memos forfattet av Aastveit mfl. (2011) forklares det at de mest brukte modellene er vektor autoregresjon, ledende indikatorer og faktormodeller. De viser så at ved å samle resultater fra flere ulike modeller gir det bedre resultater enn å plukke ut det som er antatt å være den beste ex-ante. De gjør dette ved å vekte resultatene fra de ulike modellene ulikt og slår disse sammen til ett resultat i System of Averaging Models (SAM). Modeller med nyere og mer relevant data vektet høyere i løpet av kvartalet. Vi kan spesielt trekke paralleler til metode og resultater i et working paper gjort av sentralbanken i New Zealand (Richardson mfl, 2018). Forskjellen er at mens jeg sammenligner maskinlæringsmodeller med tradisjonelle metoder som faktormodell og autoregresjon, bruker Richardson New Zealands sentralbanks estimerer som benchmark. En mulig konsekvens er at sentralbankens estimerer er bedre enn de jeg lager selv med benchmarkmodeller, og dermed vil mine maskinlæringsmodeller se forholdsvis bedre ut.

## **2. Teori**

### **2.1 Faktoranalyse**

Faktormodeller er mye brukt i sentralbanker verden over, og er spesielt nyttig i datasett med relativt mange variabler. Spesielt viktig er muligheten til å redusere et bredt datasett til et mindre, mer håndterlig et der en ivaretar frihetsgrader. I faktoranalyse søker en å beskrive samvariasjonen mellom et stort sett med variabler, med utgangspunkt i færrest mulig underliggende dimensjoner (faktorer). Vi antar da at et stort datasett kan representeres ved langt færre, uobserverte variabler. Dette bidrar til at en ikke lenger er begrenset i forhold til hvor mange serier en ønsker å bruke. Det blir da mulig å samle inn store mengder asynkrone data med høyere publiseringsfrekvens, og anvende disse umiddelbart i analysen, sånn at prognosen kan oppdateres kontinuerlig. Modellen har altså to funksjoner, å redusere kompleksitet, samt avdekke mønstre. Vi begynner med å gruppere variablene, slik at vi ender opp med en håndfull variabler kalt faktorer. Denne forenklingen sørger for parsimoni, og

krever mindre datakraft. La vektor  $X_t = (x_{1t}, x_{2t}, \dots, x_{nt})'$  beskrive  $n$  transformerte, stasjonære tidsserier for periode  $t = 1, \dots, T$ :

$$X_t = \Lambda F_t + \xi_t \quad (1)$$

der  $\Lambda = (\lambda_{0,ij})$  er en  $(n \times r)$  matrise med faktorladninger,  $r$  antall faktorer, faktorvektor er  $F_t = (f_{1t}, \dots, f_{rt})'$ , og  $\xi_t$  er idiosynkratisk komponent. Den idiosynkratiske komponenten  $\xi_t = (\xi_{1t}, \dots, \xi_{nt})'$  er uobserverbar og har forventning lik null og kovariansmatrise  $\Psi = E[\xi_t, \xi_t']$ . Modellen forenkles ved at antall faktorer  $r$ , typisk er mindre enn antall variabler  $n$ .

### 2.1.1 Prinsipalkomponentanalyse

For å redusere forklaringsvariablene til noen få felles faktorer, estimerer vi prinsipalkomponenter av datasettet ved lineær ortogonal transformering.

Prinsipalkomponentanalyse er en populær fremgangsmåte for faktoranalyse, da den er enkel å bruke, og gir gode resultater selv med store datasett. Vi ønsker å få så mye som mulig av essensen i datasettet inn i færrest mulig komponenter, samtidig som disse nye komponentene skal ha stor nok forklaringskraft til å kunne brukes i prediksjon. Transformeringen fra originale variabler, til prinsipalkomponenter foregår ved at den første komponenten skal ha høyest mulig varians. Neste komponent er den med nest høyest varians, betinget at den er ortogonalt uavhengig til den forrige. Og sånn fortsetter det videre, til alle variablene er transformert, de resulterende komponentene korrelerer da ikke med hverandre. Hvor mange faktorkomponenter en bør bruke er en del av modelltuningen, en ønsker å bruke færrest mulig faktorer for å ivarete frihetsgrader, samtidig som en vil at faktorene skal forklare så mye som mulig av variasjonen. I analysen ender vi opp med en modell basert på  $r$  komponenter, som til sammen

Prinsipalkomponentsanalysen bruker vi til å løse et MKM-problem i (3), der  $\Lambda$  og  $F_t$  er ukjente parametre vi estimerer ved:

$$\min_{\Lambda, F} V_r(\Lambda, F) \quad (2)$$

$$V_r(\Lambda, F) = \frac{1}{NT} \sum_{t=1}^T (X_t - \Lambda F_t)' (X_t - \Lambda F_t) \quad (3)$$



Løsningen på (3) er prinsipalkomponentestimator til faktorene,  $\hat{F}_t = N^{-1}\hat{\Lambda}'X_t$ , der  $\hat{\Lambda}$  er egenvektorene i utvalgets variansematrise til  $X_t$ , og inneholder de  $r$  største eigenverdiene estimert ved  $\hat{\Sigma}_x = T^{-1} \sum_{t=1}^T X_t X_t'$ . (Stock, 2016).

## 2.2 Random Forest

Francis Galton (1907) spurte innbyggerne i en landsby hvor mye en okse veier. Ingen av de gjettet riktig, men vet å ta gjennomsnittet av svarene oppnådde han et nesten perfekt estimat. Random Forest er basert på bootsrappede binære beslutningstrær, og ble først foreslått av Breiman (2001). Essensen i modellen er at vi reduserer variansen i estimatene, ved å bruke et regresjonstre mange ganger på flere bootstrap-utvalg vi har hentet fra datasettet. Det endelige estimatet finner vi enkelt ved gjennomsnittet, en prosess kalt bagging. Spesielt i prediksjon med mange variabler og mye støy, er fremgangsmåten nyttig. Random forest er en videreutvikling av bagging

### 2.2.1 Bootstrap sampling og aggregering

Anta en modell  $Z$  som tilpasses et treningssett av det fulle datasettet, der  $Z = (z_1, z_2, \dots, z_N)$ , og  $z_i = (x_i, y_i)$ . Vi trekker  $B$  nye datasett med tilbakelegg fra  $Z$ , der hvert nye utvalg har samme størrelse som  $Z$ . Noen observasjoner kan altså dukke opp flere ganger, mens andre ikke i det hele tatt. Modellen tilpasses samtlige  $B$  datasett, som vi kaller bootstrap-utvalg. Bootstrap-aggregering, eller bagging, kan forbedre estimatet ved å tilpasse modellen til hvert av bootstrap-utvalgene,  $b = 1, 2, \dots, B$ , som gir  $\hat{f}^{*b}(x)$ . Baggingen estimeres som gjennomsnittet for modellene over alle bootstraputvalg vi har:

$$\hat{f}_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^{*b}(x) \quad (4)$$

### 2.2.3 Beslutningstrær

Beslutningstrær lages for samtlige  $B$  bootstrap-utvalg. Fra variablene velges  $m$  tilfeldige variabler, og vi finner det beste splitt-punktet som minimerer estimeringsfeil. En node som splittes gir oss to nye noder som enten splittes videre eller er terminale.

Vi har et datasett med  $p$  forklaringsvariabler  $x_{ip}$ , og en responsvariabel  $y_i$ , med  $N$  antall observasjoner. Vi deler responsvariabelen inn i  $M$  regioner  $R_1, R_2, \dots, R_M$ , der responsvariabelen er en konstant  $c_m$ . Dette vises ved:

$$f(x) = \sum_{b=1}^B c_m I(x \in R_m) \quad (5)$$

der  $I$  er en indikatorfunksjon. I ordinær bagging vurderes alle variabler når den beste splitten identifiseres. I Random forest benyttes kun et tilfeldig utvalg av variablene. Antallet variabler som skal vurderes ved splitten i random forest er en hyperparameter vi velger på forhånd. Fremgangsmåten for valg av hyperparametere diskuteres videre i seksjonen om data og metode. Vi finner den verdien for  $c_m$  som minimerer summen av kvadratfeil, et mål på avviket mellom  $f(x)$  og  $y_i$ . Beslutningstreet bygges i flere steg. Ved første node, treroten, deles utvalget inn i to delutvalg ved hjelp av en grådig algoritme, der utvalget splittes ved en variabel  $j$  og et punkt  $s$ . Dette gir to halvplan:

$$R_1(j, s) = \{X | X_j \leq s\} \quad (6)$$

$$R_2(j, s) = \{X | X_j > s\} \quad (7)$$

Ved å minimere summen av kvadratfeil til å finne optimal  $s$  og  $j$  får vi følgende:

$$\min_{j,s} [\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2] \quad (8)$$

Den indre minimeringen løses for hver splitt ved:

$$\hat{c}_1 = gj. snitt(y_i | x_i \in R_1(j, s)) \quad (9)$$

$$\hat{c}_2 = gj. snitt(y_i | x_i \in R_2(j, s)) \quad (10)$$

Prosessen gjentas for hver splitt til et kriterie for terminalnode er oppnådd. Et relevant kriterie kan være at en splitt ikke oppfyller et forhåndbestemt krav om minimum antall observasjoner,

at maks antall nivåer på treet er nådd, eller at å splitte videre ikke gir nevneverdig reduksjon i avviksmålet. Når prosessen er fullført for samtlige beslutningstrær, filtreres støyen ut ved å slå dem sammen med bagging, der det endelige resultatet er gjennomsnittet av trærne:

$$\hat{f}_{rf}^B(x) = \frac{1}{B} \sum_{b=1}^B T_b(x, \Theta_b), \quad (11)$$

der  $\Theta_b$  karakteriserer et tre  $b$  ved dets splitt-variabler og –punkt, og verdier ved terminalnodene.

Et eksempel på et beslutningstre finner en i appendix A.1, generert ved prediksjon for Januar 2019. I hver node av treet kontrolleres det om en tidsseries verdi er høyere eller lavere enn den oppgitte verdien i noden. Deretter fortsetter en nedover til en kommer til predikert verdi i terminalnoden, eller «bladene». Den endelige random forest prediksjonen gjøres ved å ta gjennomsnittet av alle trærne i modellen. Ulempen med maskinlæringsmodeller er at det kan være vanskelig å se hva som driver estimatet. Det utvikles hele tiden nye pogrampakker som forsøker å bøte på dette. Et eksempel kan sees på plot for variabelviktighet for perioden 2019Q1 i appendix A.4.

### 2.3 Support Vector Machine

Support Vector Machine er en maskinlæringsalgoritme av Vapnik (1992) som kan brukes til både regresjon og klassifisering. Kort fortalt går modellen ut på å plote datasettet i et *feature space*, og dele opp datamengder i to kategorier med et hyperplan. Modellen bygger på kernel-funksjoner, i tilfeller der mønstre mellom variablene ikke kan løses lineært, lar vi kernelfunksjonen transformere data som ikke er lineært separerbart, ved å tillegge en ekstra dimensjon. En løser da problemet ved å se på det tredimensjonalt. Det beste hyperplanet er det som maksimerer avstanden mellom seg og nærmeste observasjon. De nærmeste datapunktene på begge sider av hyperplanet gjør opp det som kalles støttevektorer, som er de eneste modellen vektorer med noe annet enn null. Modellen var opprinnelig laget for binær klassifikasjon, men tilpasses for regresjon ved å tillate verdier som ligger innenfor en avstand  $\varepsilon$  fra støttevektorene (Smola & Schölkopf, 2004).

## 2.4 Nevralt Nettverk

Nevralt nettverk er en samlebetegnelse for metoder som er inspirert av og etterligner biologisk nervevev i hjernen. Metodegrunnlaget kan spores tilbake til Frank Rosenblatt (1957) sin forskning på perceptroner. Et perceptron består av ett enkelt skjult lag med vekter, og er forgjengeren til nevralt nettverk.

Et enkelt nevralt nettverk består av tre lag, et for inndata, et skjult lag, og et for utdata. Hvert lag i modellen består av ulike vektornoder en kaller nevroner. I inndata har vi et nevron for hver tidsserie. Utdata består av kun ett nevron, estimert BNP-vekst. Lagene fungerer som filtre, der inndata siles ut og transformeres til noe vi kan bruke i prediksjonen. I utgangspunktet er vektingen tilfeldig, hvis en har to uavhengige variabler  $v_1$  og  $v_2$ , og en avhengig variabel  $y$ , skal en vekte de to inndataene slik at summen deres blir så lik  $y$  som mulig. Formelt kan det skrives som

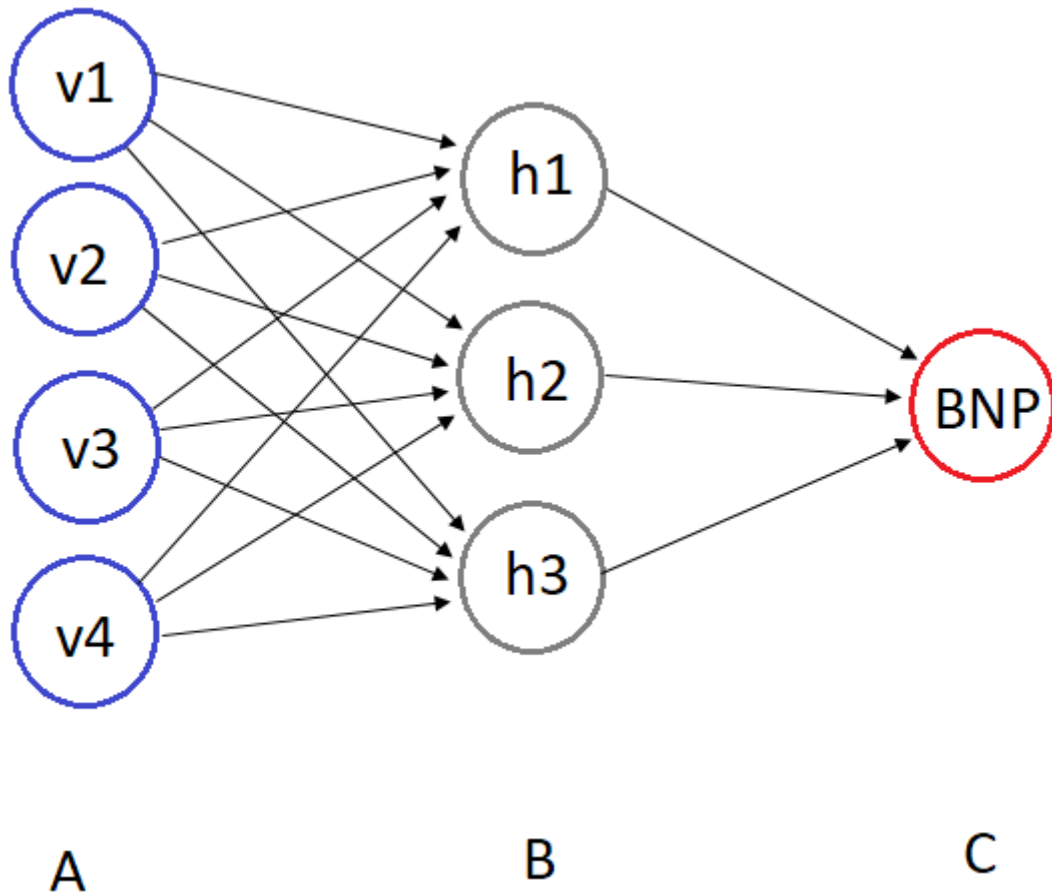
$$\hat{y} = x_1 w_1 + x_2 w_2, \quad (12)$$

Der  $\hat{y}$  er utdata og  $w_i$  er variabelvekt. I figuren under har vi inndata i lag A, som vektet og summeres på veien til lag B. Vektingen er representert ved pilene mellom nevronene.

I hvert nevron ligger sigmoid aktiveringsfunksjoner som vekter inndataene, og kobler vekting mot utdata. Aktiveringsfunksjonen i et skjult nevron er her definert ved

$$h(v_i) = (1 + e^{-v_i})^{-1}, \quad (13)$$

der  $h(v_i)$  er transformert data fra det  $i$ -te nevron i det skjulte laget, og  $v_i$  den vektete summen fra inndata til nevronet. Læring oppnås ved at modellen vekter inndata på nytt hvis estimatet er for langt unna observert verdi. Modellen oppdaterer vektene etter Levenberg-Marquardt optimering. Optimeringen lar en estimere en Jacobian-matrise, som inneholder in-sample estimeringsfeil. Den bruker så disse feilene til å gå tilbake og vekte de aktuelle nodene på nytt, til den har minimert tapsfunksjonen MSE (19).



Figur 1: Illustrering av et nevralt nettverk.

## 2.5 Modellkombinering

Selv om én enkelt modell gjør det veldig bra, kan prediksjonen forbedres ved å slå sammen resultater fra flere modeller inn i et ensemble. Vi vekter da resultatene og samler de i én enkelt prediksjon  $\hat{Y}_t$  for et gitt tidspunkt. Gitt at vi har  $N$  modeller  $j = 1, 2, 3, \dots, N$ , definerer vi den samlede prediksjonen i tidspunkt  $t$  som

$$\hat{Y}_t = \sum_{j=1}^N w_{j,t} \hat{Y}_{j,t} . \quad (14)$$

Der  $\hat{Y}_t$  er ensembles estimat i tidspunkt  $t$ ,  $w_j$  er vektcoeffisienten for modell  $j$  ved tidspunkt  $t$ , og  $\hat{Y}_{j,t}$  modell  $j$ 's prediksjon i tidspunkt  $t$ . En søker å oppnå en vekting som minimerer prediksjonens kvadrerte residual, der denne defineres ved:

$$MSE_t = (\widehat{Y}_t - Y_t)^2, \quad (15)$$

der  $Y_t$  er samme periodes observert verdi. Etterpå summerer vi modellens MSE over hele prediksjonsperioden, og bruker dens gjennomsnitt som mål på om den gjør det bra eller ikke, ved å sammenligne modellene.

Det finnes ulike fremgangsmåter for vektfordeling. Hvilken som gir best resultat er der ingen klare regler for, det kan variere med modellsammensetningen og dataene en bruker. Her bruker jeg to ulike metoder for vekting, og vurderer disse opp mot hverandre på resultatene. Den første, og enkleste metoden vi bruker for å angi vekt er et enkelt gjennomsnitt av prediksjonene fra hver enkelt modell,

$$w_j = \frac{1}{N}. \quad (16)$$

En modells vekt er da konstant over perioden, så lenge samme antall modeller er i bruk. Fremgangsmåten vektet altså alle modellene likt, og ser bort i fra historisk treffsikkerhet. Til tross for metoden kan virke for enkel, har den gitt gode resultater i tidligere litteratur, et fenomen kjent som «the forecast puzzle», og videre beskrevet og forklart av Claeskens mfl (2016).

Den andre vektemetoden vi vurderer er basert på arbeid av Bates og Granger (1969), som er mye brukt i relevant litteratur, blant annet Richardson mfl. (2018) og Chikamatsu mfl. (2018). En vektet enkeltmodellens prediksjon ved å ta invers MSE til modellen, og forholdet til IMSE hos de andre modellene i samlingen. Modeller med relativt god historisk treffsikkerhet vektet da høyere enn andre modellene over perioden.

$$\omega_{i,t} = \frac{\frac{1}{MSE_{i,t}}}{\sum_{j=1}^N \frac{1}{MSE_{j,t}}} \quad (17)$$

der  $i$  er modellen som skal vektet,  $t$  tidsperioden som estimeres for, og  $N$  antall modeller i ensemblet.

### 3.0 Data

Til analysen brukes et datasett med tidsserier for hovedsaklig Norge, samt et utvalg variabler for Norges nærmeste handelspartnere. Det inneholder i hovedsak det som kalles «hard data», tidsserier fra realøkonomi, i tillegg til finansvariabler. Aastveit og Trovik (2012) finner at finansdata har mye å si for prediksjon på norsk kvartalsvis økonomisk vekst, spesielt tidlig i kvartalet, når nyeste tall fra realøkonomien ennå ikke er tilgjengelig. Jeg har tatt paneldata brukt i Aastveit og Trovik som utgangspunkt. Dette strekker seg fra andre kvartal i 2000 til og med første kvartal i 2019. Jeg har hentet datasettet fra Macrobond, det består av 128 tidsserier, der 80 er for norsk makroøkonomi, 14 utenlandsk makroøkonomi, 33 fra finans, og 1 for sentiment. Variablene for makroøkonomi representerer arbeidsmarkedet, konsumpriser, eksport/import, boligmarkedet, industriproduksjon, detaljhandel, råvarepriser, renter, data fra finansmarkeder for ulike sektorer, og gjeld. Noen av tidsseriene brukt i av Aastveit og Trovik er ikke lenger tilgjengelige, derfor har jeg noen serier færre i forhold til deres 148.

Tidsserieverdier for en gitt periode endrer seg over tid, som følge av sesongjustering og revisjoner. Det beste ville dermed vært å ha et unikt datasett, et vintage, for hvert prediksjonstidspunkt, slik at det blir så virkelighetsnært som mulig. Sentralbanker bruker ofte denne fremgangsmåten i sine modelltester. Dette blir ikke gjort her, siden Macrobond kun tilbyr nyeste vintage av tidsseriene. Noen tidsserier som publiseres månedlig, er ofte ikke tilgjengelig før noen uker etter at måneden er over. Norges Bank publiserer BNP mellom 45-50 dager etter at kvartalet er over. Når jeg predikerer BNP-vekst for eksempel for årets første kvartal, vil tilgjengeligheten på betimelig data variere med prediksjonspunktet. Jo lenger ut i kvartalet en kommer, jo mer oppdatert vil datasettet være, og jo mer nøyaktig prediksjon kan en regne med få. Derfor er det nyttig med høyfrekvent data, da den er mest mulig oppdatert uansett prediksjonstidspunkt. Tidsserier fra finans, rente, råvarepriser er eksempler på daglig oppdatert data, mens makrotall for en måned kan ta alt fra en uke til seks uker etter måneden er over før seriene blir publisert. Selv en gitt tidsserier kan variere noe i når den publiseres, som følge av helg eller ferie. Vi vet altså ikke eksakt dato for når tidsseriene publiseres. Vi vet likevel hvilke tidsserier vi har tilgjengelig helt på slutten av måneden, og bruker dermed månedens siste dag som prediksjonstidspunkt.

Vi har simulert publiseringsforsinkelser i analysen, ved at serier fra industriproduksjon for en periode er tilgjengelig til bruk først i andre måned etter at perioden er over. Finansdata publiseres daglig, og er tilgjengelig for inneværende periode, mens resten av tidsseriene for en periode er tilgjengelig i slutten av måneden etter at perioden er over. Hvis vi predikerer

økonomisk vekst for perioden januar-mars, og vi befinner oss i slutten av februar, vet vi at vi har data for industriproduksjon og detaljhandel fra desember, månedlig finansdata for februar, og resten fra januar. En slik analyse er en tilnærming til pseudo sanntidsanalyser tidligere arbeid gjort av Rünstler mfl. (2009) og Aastveit og Trovik (2012). Aastveit og Trovik deler datasettet opp i 13 blokker etter publiseringsdato, og bruker faktormodell til å predikere for hver gang en blokk publiseres. Fremgangsmåten i denne oppgaven er forenklet, publiseringsdato er tatt hensyn til, men prediksjon forekommer kun én gang månedlig, på månedens siste dag. Ulempen er at vi ikke får sett variabelbetydningen i form av endringer i prediksjonsfeil over månedens løp, men en sånn analyse egner seg heller ikke når vi skal sammenligne ulike modeller, da disse benytter seg av data ulikt.

At noen av variablene kan regnes som utdaterte gjør at vi mangler relevante verdier for de tregeeste publiseringene på et prediksjonspunkt. Dette *jagged-edge* tilfellet løser vi ved å estimere de manglende verdiene med førsteordens autoregressiv modell, AR(1). Løsningen skiller seg fra den typiske fremgangsmåten for faktormodeller, der en bruker Kalmanfilter eller EM-algoritme på faktorer (Shumacher og Breitung, 2008), som ikke vil kunne anvendes i maskinlæringsmodellene uten å redusere datasettet til faktorer. Når vi reduserer datasettet til faktorer, mister vi muligheten til å se betydningen av enkelte tidsserier.

Noen av tidsseriene er preget av sesongeffekter, blant annet er konsum og varehandel høyere rundt juletider, og produksjon lavere i fellesferien. Slike gjentakende fenomen er vanlig å filtrere ut ved såkalt sesongjustering, for å gjøre dataene lettere å fortolke og analysere. Jeg har hentet ferdig sesongjusterte tidsserier der det er tilgjengelig, andre relevante tidsserier er sesongjustert med X-13ARIMA-SEATS i R. Programpakken er utviklet av United States Census Bureau, og er blant de mest brukte i verden til sesongjustering (Sax & Eddelbuettel, 2018).

Data som publiseres daglig er først slått sammen til månedlig gjennomsnitt. Siden BNP-vekst her er representert som kvartalsvis vekstrate, konverterer jeg månedvis variabler til aggregert tre-måneders gjennomsnitt, som beskrevet i Mariano og Murasawa, 2003:

$$x_{i,t}^Q = \frac{1}{3}(x_{i,t}^M + 2x_{i,t-1}^M + 3x_{i,t-2}^M + 2x_{i,t-3}^M + x_{i,t-4}^M). \quad (18)$$



### **3.1 Stasjonarisering**

Jeg undersøker stasjonaritet i settet med augmented Dickey-Fuller test på enhetsrot, drift og trend, og differensierer ikke-stasjonære variabler for å sørge for stasjonærhet. Stasjonære seriers gjennomsnitt og varians holder seg uendret over tid. Diverse sjokk på variablene vil da ikke ha en permanent effekt, som gjør at vi lettere kan skille sjokkeffekter fra tidseffekter. Stasjonaritet er derfor foretrukket i tidsserieranalyser. Variabler på prosentform er stort sett greie å ha med å gjøre, da de som regel flukterer rundt en fast verdi. Likevel har vi etter ADF-test endt opp med å ta førstedifferansen på de fleste av disse. Resten av settet på nivåform er først transformert til log-form, og tatt førstedifferansen av. Hvilke tidsserier som er inkludert, og hvordan de er transformert er beskrevet i appendix A.2. Før modellene tas i bruk, normaliseres forklaringsvariablene slik at de får gjennomsnitt lik null, og varians lik én. På den måten vil serier som normalt har store verdier ikke få større betydning enn små.

### **3.2 Metode**

Datasettet splittes opp i to deler ved prediksjonstidspunktet, et treningssett og et testsett. Modellene trenes først på treningssettet, og anvendes videre til å gjøre prediksjoner på testsettet. Aller først trenes modellen på data fra andre kvartal i 2000 til og med første kvartal 2007, og brukes til å predikere for andre kvartal i 2007. Estimater sammenlignes da med observert verdi for perioden, og kvadrert residual registreres. Deretter utvides treningssettet til å inkludere 2007Q2, modellen trenes på nytt, og predikerer for neste periode, 2007Q3. Prosessen gjentas for hele testsettet, som går fra 2007Q2 til 2019Q1. Ved første prediksjon vil modellene ha trent seg på 28 observasjoner, som tilsvarer 7 år. Resultatene fra hele perioden blir da på 48 observasjoner, som tilsvarer 12 år.

### **4.0 Resultater**

Vi begynner med å se på de ulike modellenes ytelse hver for seg over perioden, hvor godt de predikerer BNP-vekst i neste kvartal. Deretter samler vi modellene og vurderer hvilken kombinasjon av modeller og vektning som gir best prediksjon i en gitt måned. Analysen gjøres først i slutten av kvartalets første måned, deretter andre, så tredje. Resultatene for perioden oppsummeres i del 4.4.

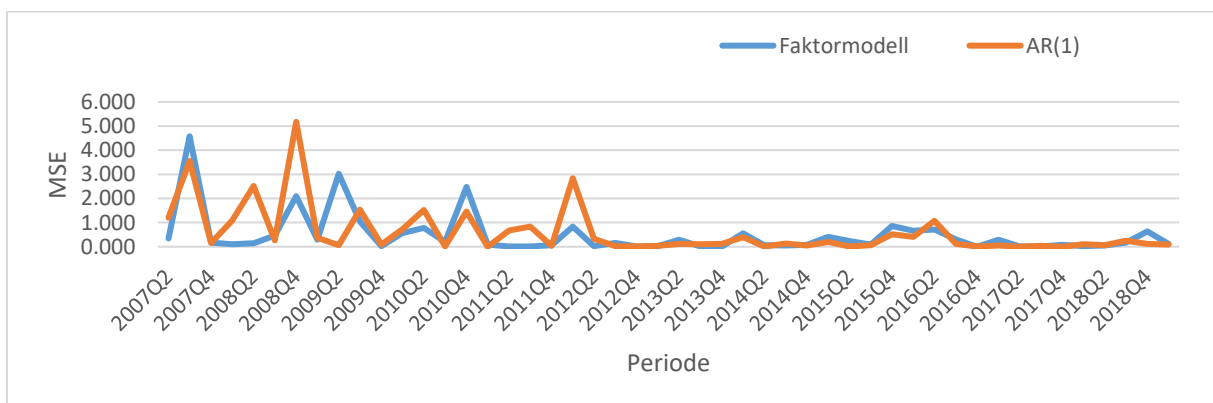
## 4.1 Første måned

I tabell 1 har vi gjennomsnittlig prediksjonsfeil for hver enkeltmodell over perioden 2007Q2-2019Q1. Random Walk gjorde det klart dårligst, og brukes som sammenligningsgrunnlag for de andre modellene i prosent. Faktormodeller, AR(1) og random forest har alle en ganske lav og lik gjennomsnittlig målefeil over perioden, og virker som gode kandidater for sammenslåing, gitt at de varierer ulikt.

Prediksjonsfeil		
Modell	1. måned	
	MSE	%
Random Forest	0.576	59.30 %
Nevralt Nettverk	0.982	101.06 %
Support Vector Regression	1.147	118.12 %
Faktormodell	0.595	61.27 %
AR(1)	0.542	55.80 %
Random Walk	1.187	100.00 %

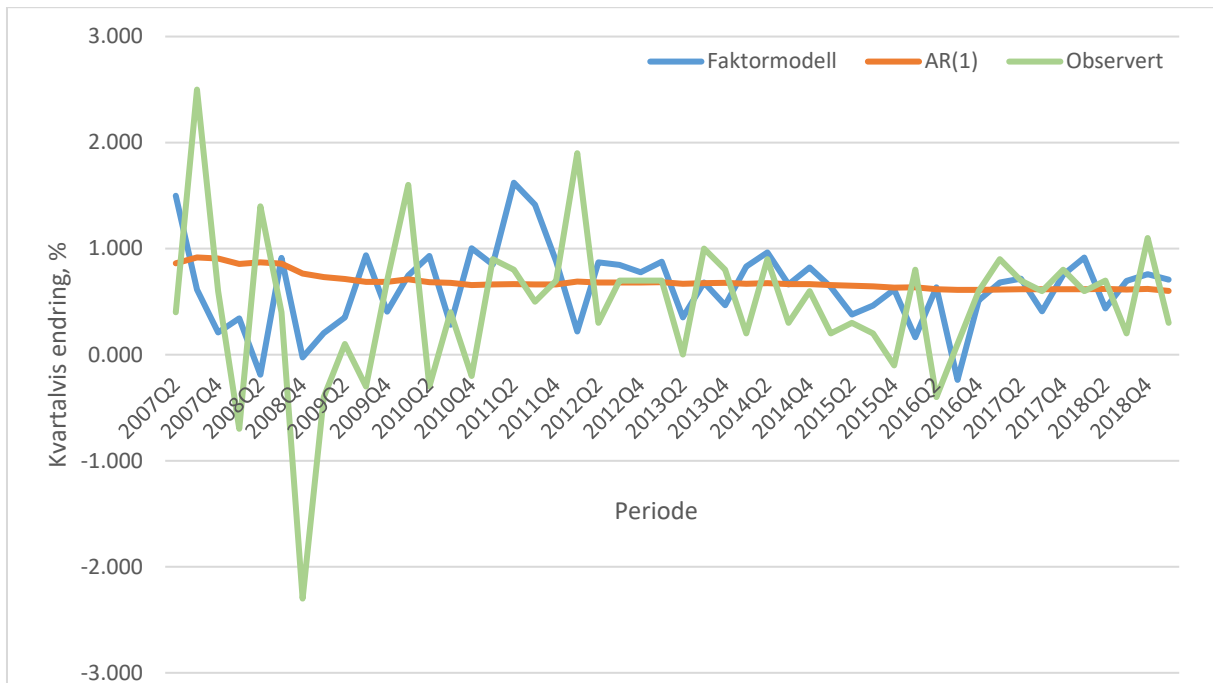
Tabell 1: Gjennomsnittlig feilledd for enkeltmodeller 2000-2018.

Det er også interessant å se hvordan feilleddene varierer over perioden, i figur 2 vises det for faktormodell og AR(1). Begge modellene bommer ganske grovt i begynnelsen av perioden, spesielt sterkt negativt BNP-endring greide ikke modellen å følge. Videre utover perioden flater den ut, noe som kan følge av at BNP-endring også har stabilisert seg.



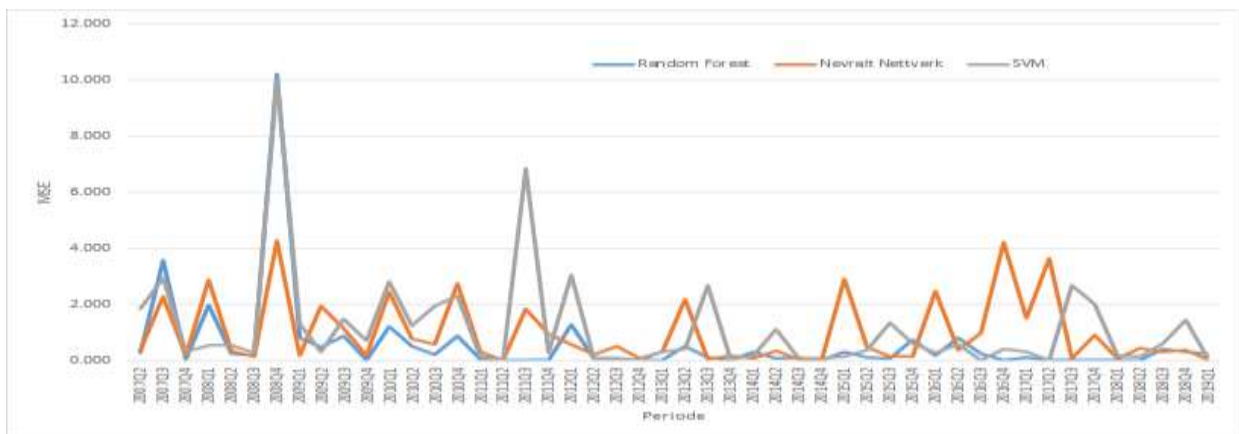
Figur 2: Feilledd over perioden for faktormodell og førsteordens autoregresjon.

I figur 3 ser vi at selv om modellen har ganske like feilledd, varierer de veldig i prediksjon. Mens førsteordens autoregresjon er flat over perioden, følger faktormodell mer med på svingningene. Det medfører og at faktormodellen straffes med høy MSE når den bommer på svingninger.



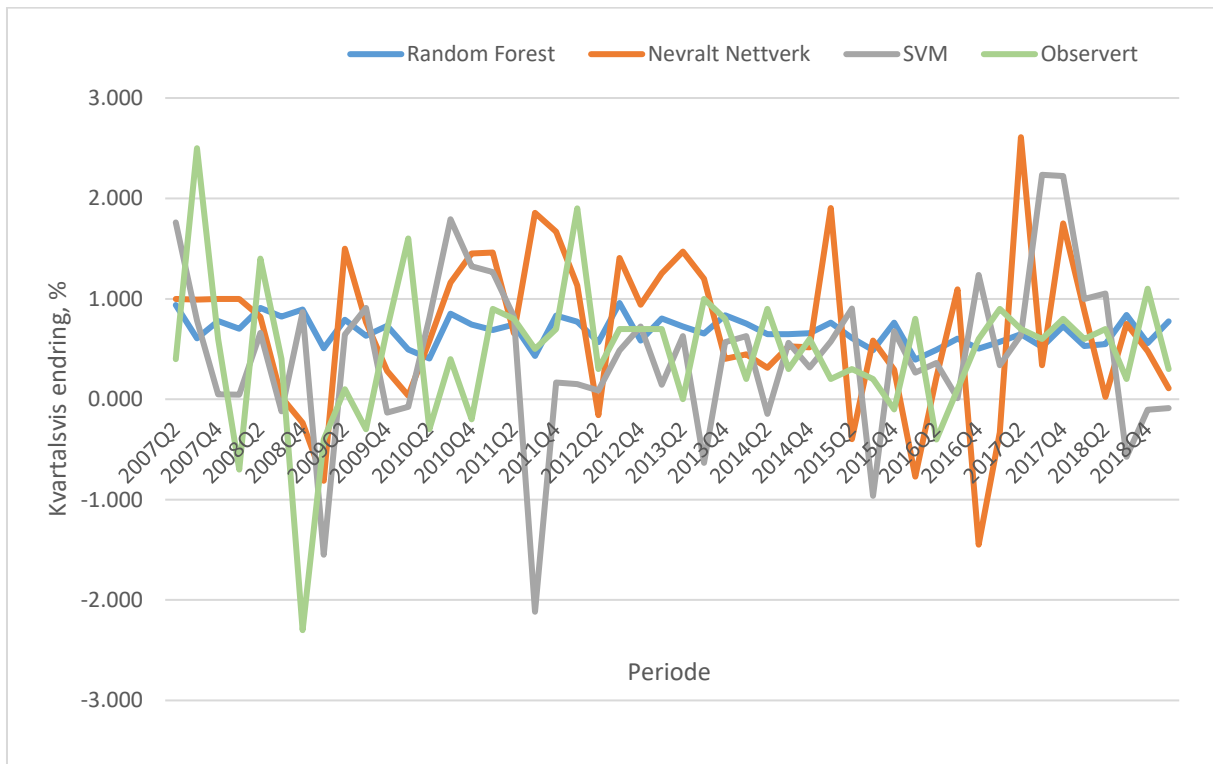
Figur 3: Predikert økonomisk vekst, mot observert. Faktormodell og førsteordens autoregresjon.

For maskinlæringsmodellene ser vi mye det samme, der de sliter med å predikere finanskrisen i slutten av 2018. Vi ser og at de fortsetter å bomme ganske mye over perioden, i motsetning til faktormodellen og AR(1).



Figur 4: Feilledd over perioden for Random forest, nevralt nettverk og SVM.

I figur fem ser en hvordan maskinlæringsmodellene varierer fra hverandre. Mens random forest predikerer riktig retning på vekstendring, ligger den fortsatt ganske flat. Modellen virker å mangle styrke til å følge vekstendringene fulgt ut. I motsatt fall er nevralt nettverk og SVM mer volatile, de svinger mye mer med observert verdi, men bommer også i perioder der BNP-vekst ligger flatt. Spesielt i slutten av perioden, der vekst har vært jevn, bommer SVM og nevralt nettverk grovere enn de andre modellene.



Figur 5: Predikert økonomisk vekst, mot observert. Random forest, Nevralt Nettverk, SVM.

I neste steg samler vi prediksjoner fra modellene i to grupper, en benchmarkmodell som består av faktormodell og AR(1), og en maskinlæringsgruppe kalt ML, som består av random forest, nevralt nettverk og support vector machine. Vi sammenligner resultatene for samlemodellene og prediksjonsfeil over perioden, og ser hvordan de enkelte modellene bidrar til prediksjonen i en samlet modell.

I tabell 2 vises avvik mellom predikert og observert verdi når vi samler modellene. Enkeltmodellene i de to ensemble-gruppene er vektet etter de to metodene diskutert i avsnitt 2.5. Vi fokuserer på resultatene fra invers MSE vektning (IMSE), men har med resultater for vektning ved enkelt gjennomsnitt (Avg) som sammenligningsgrunnlag når vi evaluerer vektemetodene. Vi observerer at Benchmark-samlemodellen gjør det bedre enn maskinlærings-

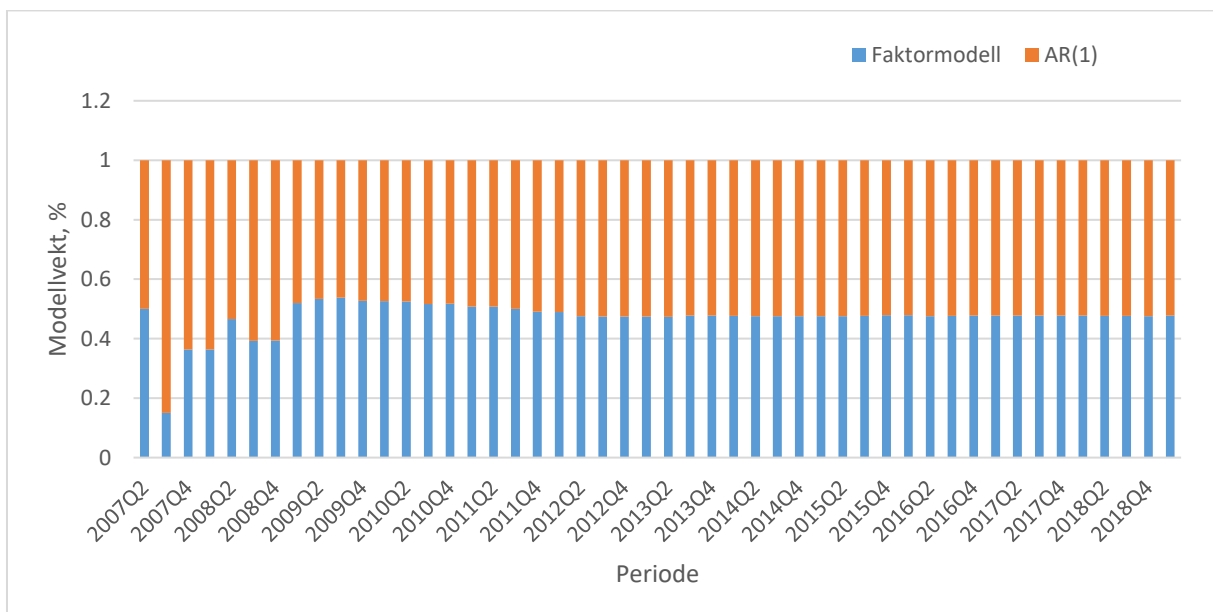
på tvers av de ulike vektemetodene, men bare så vidt. Fra tabell 1 så vi at modellene nevnt nettverk og SVM hadde svært høyt gjennomsnittlig feilledd. Likevel, når vi slår disse sammen til ML får vi en enda bedre gjennomsnittlig prediksjon enn selv den beste enkeltmodellen i samlingen. Også benchmarkensemblen gir bedre resultater enn enkeltmodellene den utgjøres av.

### Prediksjonsfeil

Modell	Vekt	1.måned	
		MSE	%
Benchmark	IMSE	0.525	54.07 %
	Avg	0.549	56.58 %
Maskinlæring	IMSE	0.546	56.23 %
	Avg	0.542	55.82 %

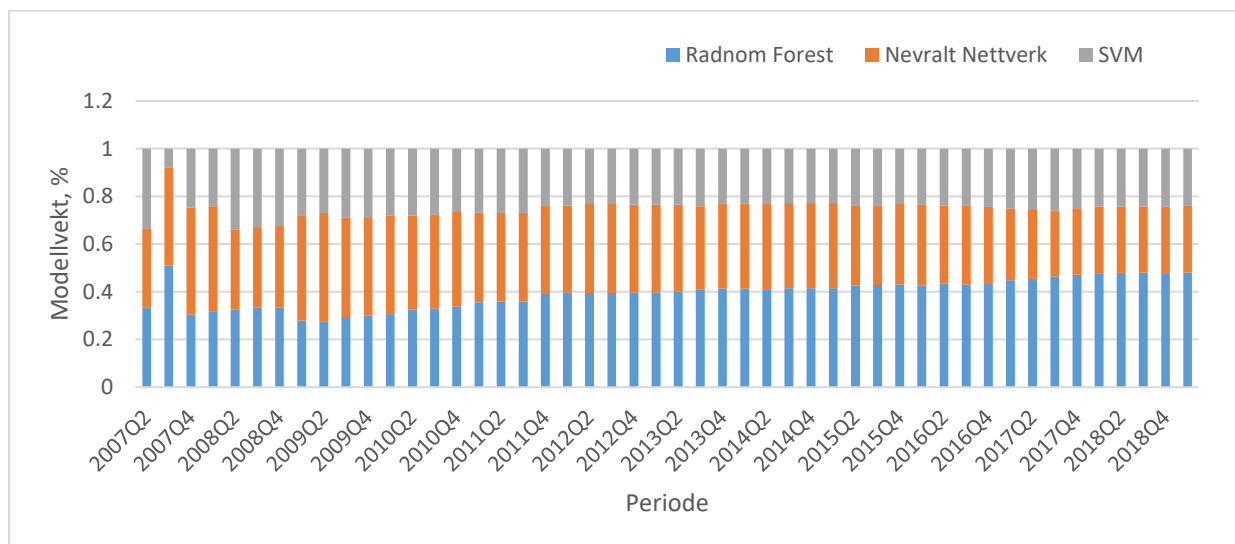
Tabell 2: Feilledd for samlede benchmark og maskinlæringsmodeller

Figur 6 viser hvordan benchmarkmodellen er vektet med invers MSE metoden over perioden. Variasjonen i begynnelsen av perioden kommer av at det historiske sammenligningsgrunnlaget begynner i første periode, og jobber seg mot et jevnere gjennomsnitt jo lenger utover vi kommer.



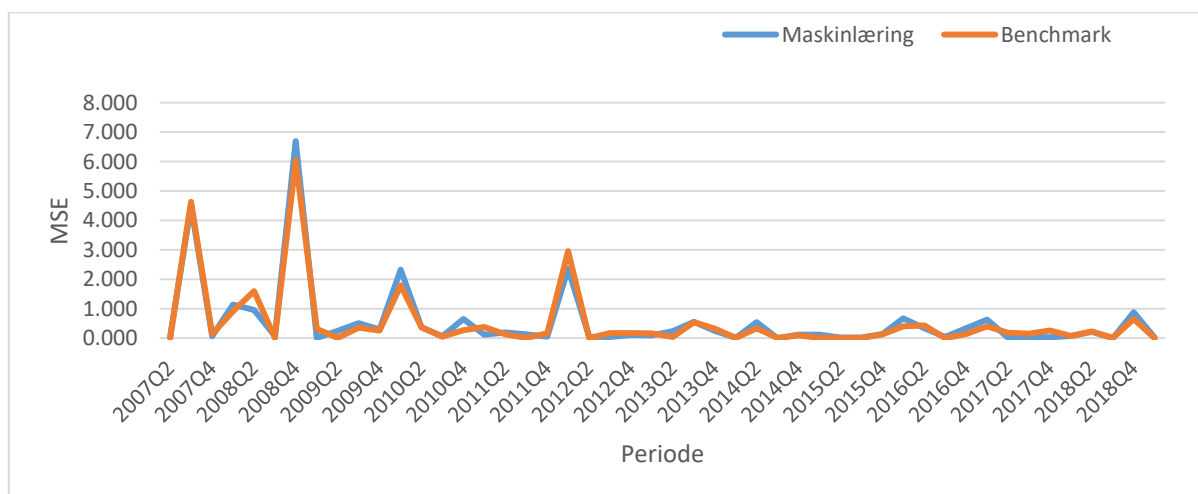
Figur 6: Vekting benchmarkmodeller i prosent over perioden

I slutten av perioden er vektfordelingen nesten tilsvarende et enkelt gjennomsnitt. Tilsvarende for maskinlæringsmodeller ser vi en utjevning i figur 7, men at SVM prioriteres lavere enn de to andre modellene, mens random forest som viste seg å være ganske konservativ over perioden, blir vektet høyere. En samlemodell kan tenkes å inneha samme evne til å følge skarpe endringer i BNP-vekst som vi så i nevralt nettverk og SVM, samtidig som den flates ut gjennom random forest.

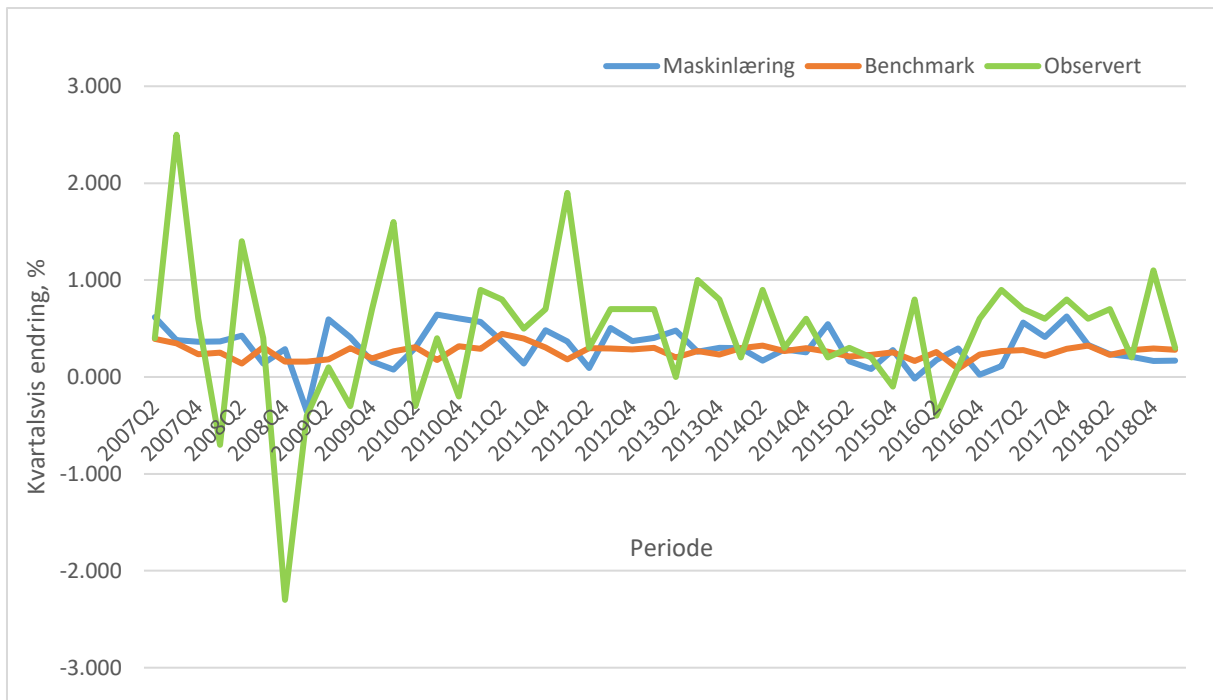


Figur 7: Vekting maskinlæringsmodeller i prosent over perioden.

Setter vi de to samlemodellene opp motser vi at de matcher hverandre ganske likt.

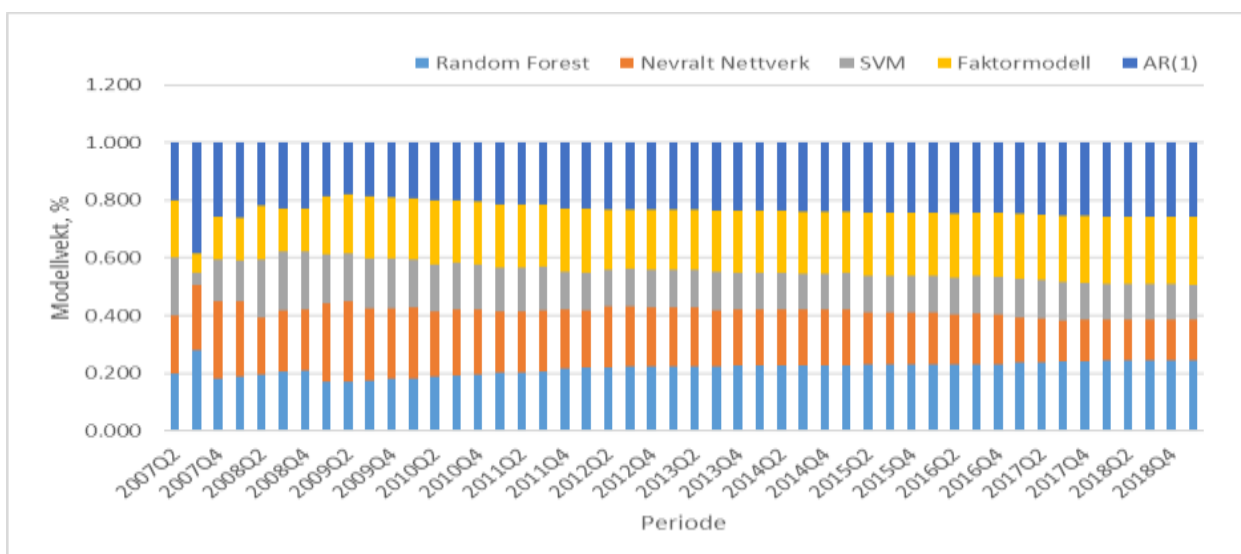


Figur 8: Feilledd for to samlemodeller, benchmark og maskinlæring.

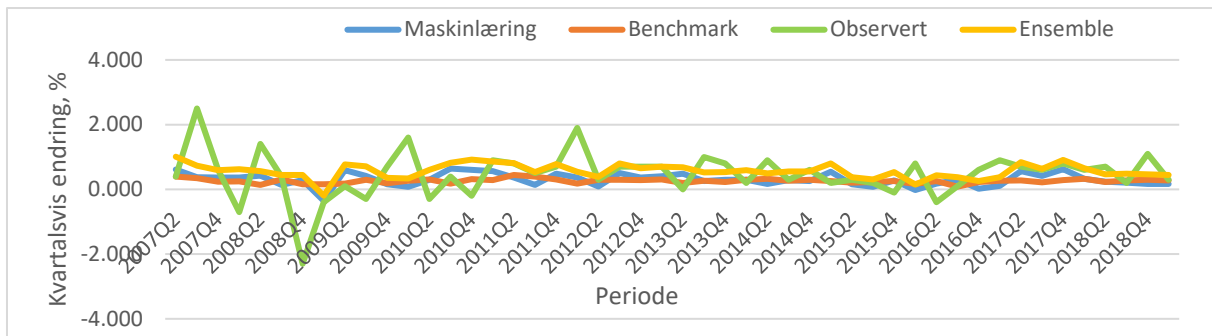


Figur 9: Prediksjon ved maskinlæring og benchmark, mot observerte verdier.

Figur 9 illustrer de to samlemodellenes prediksjon over perioden, som ventet er begge modellen ganske flate, men maskinlæring svinger mer enn benchmarkmodellen, en nyttig egenskap når vi slår alt sammen til en komplett modell. Slår vi sammen benchmark og maskinlæring vektet AR(1) og random forest høyere enn tidligere i forhold til de andre modellene, mens nevral nettverk og SVMs viktighet fortsatt reduseres (figur 10).



Figur 10: Modellvektning i komplett modellensemble. Random forest og AR(1) høyest vektet.



Figur 11: Tre ulike samlemodeller mot observert BNP-vekst.

Vi ønsker å se om maskinlæringsmodeller kan forbedre benchmarkmodellen. Derfor prøver vi oss frem med ulike kombinasjoner av benchmarkmodeller og enkeltmodeller fra maskinlæringsgruppen (tabell 3). Den beste får vi ved å slå sammen benchmark( AR(1), faktormodell) og random forest. Vi kan likevel få en enda bedre modell ved å slå sammen alle modellene, og inkludere en naiv random walk.

### Prediksjonsfeil

Modell	Vekt	1. måned	
		MSE	%
Ensemble	IMSE	0.546	45.99 %
	Avg	0.547	46.11 %
BM + RF	IMSE	0.540	45.50 %
	Avg	0.534	44.97 %
BM + NN	IMSE	0.547	46.10 %
	Avg	0.555	46.76 %
BM + SVM	IMSE	0.552	46.51 %
	Avg	0.572	48.14 %
Ensemble + RW	IMSE	0.518	43.6 %
	Avg	0.511	43.1 %

Tabell 3: Feilledd for ulike sammensetninger av samlemodeller.



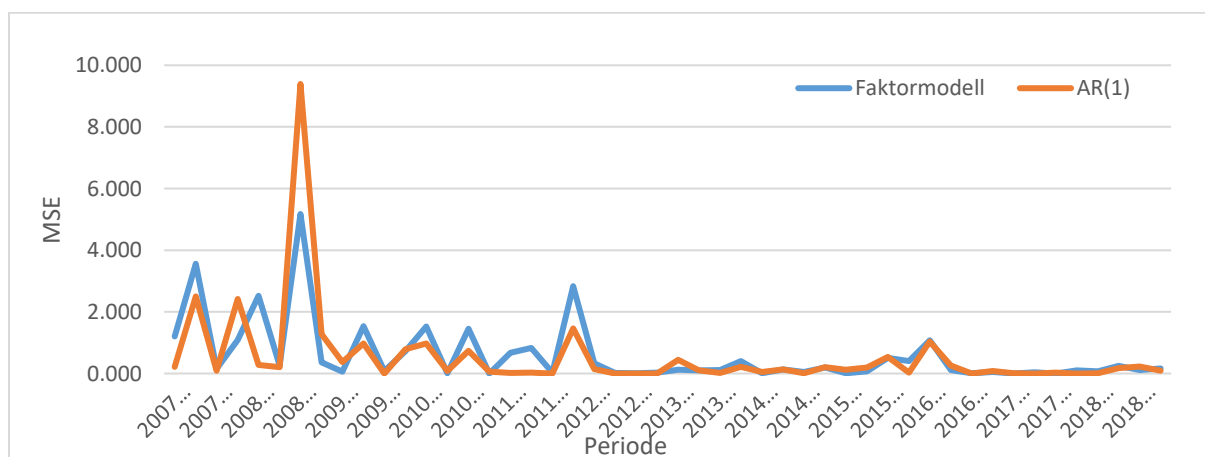
## 4.2 Andre måned

I tabell 4 har vi gjennomsnittlig prediksjonsfeil for hver enkeltmodell over perioden 2007Q2-2019Q1.

Prediksjonsfeil		
Modell	2 måned	
	MSE	%
Random Forest	0.541	55.67 %
Nevralt Nettverk	0.886	91.20 %
Support Vector Regression	1.068	109.93 %
Faktormodell	0.516	53.14 %
AR(1)	0.557	57.31 %
Random Walk	0.971	100.00 %

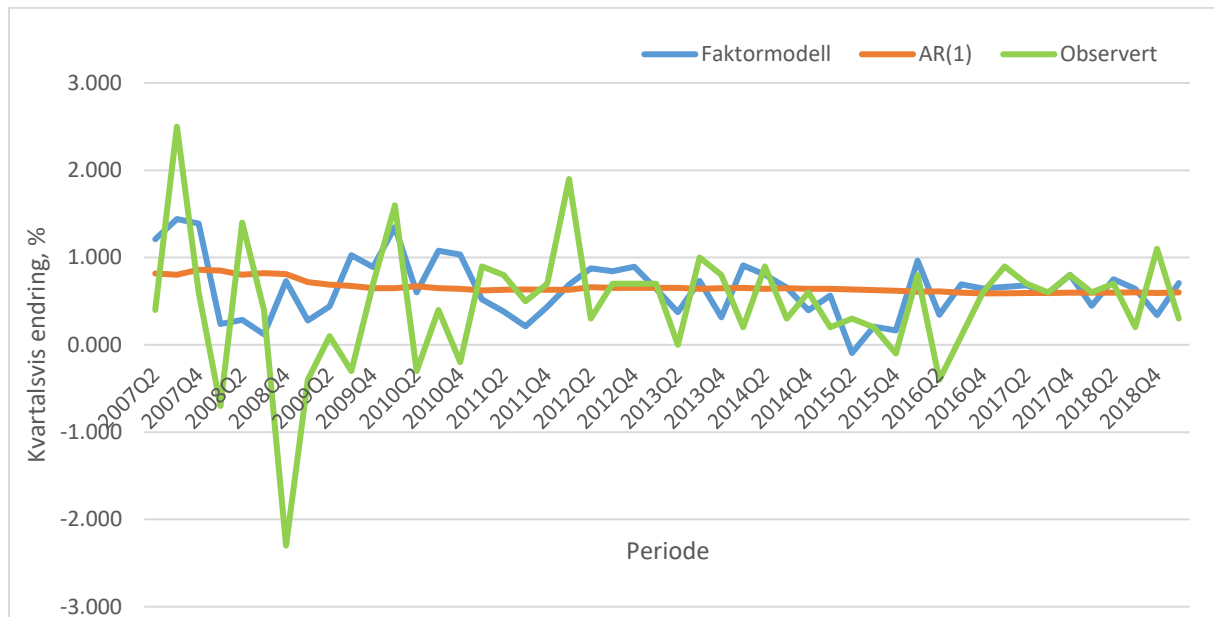
Tabell 4: Gjennomsnittlig feilledd for enkeltmodeller 2000-2018.

Resultatene for prediksjon i andre måned har mye fellestrekk med første måned. Random forest, faktormodell og AR(1) er fortsatt de mest nøyaktige over perioden, men nevralt nettverk og SVM har forbedret seg noe. Nytt i andremånedanalyse er at har mer oppdatert data, der makrotall (utenom industriproduksjon) er kommet inn for kvartalets første måned.



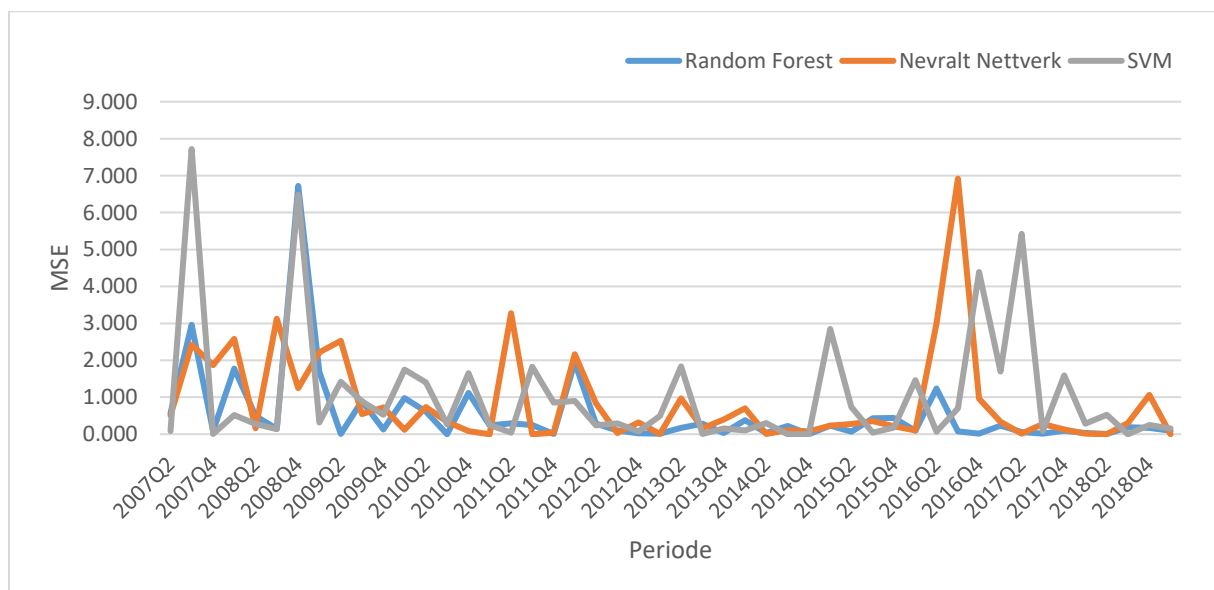
Figur 12: Feilledd over perioden for faktormodell og førsteordens autoregresjon.

I figur 12 er begge modellenes feilledd lavere enn i første måned, og flater ut fortere over perioden. AR(1) predikson er forholdsvis flat over hele perioden.

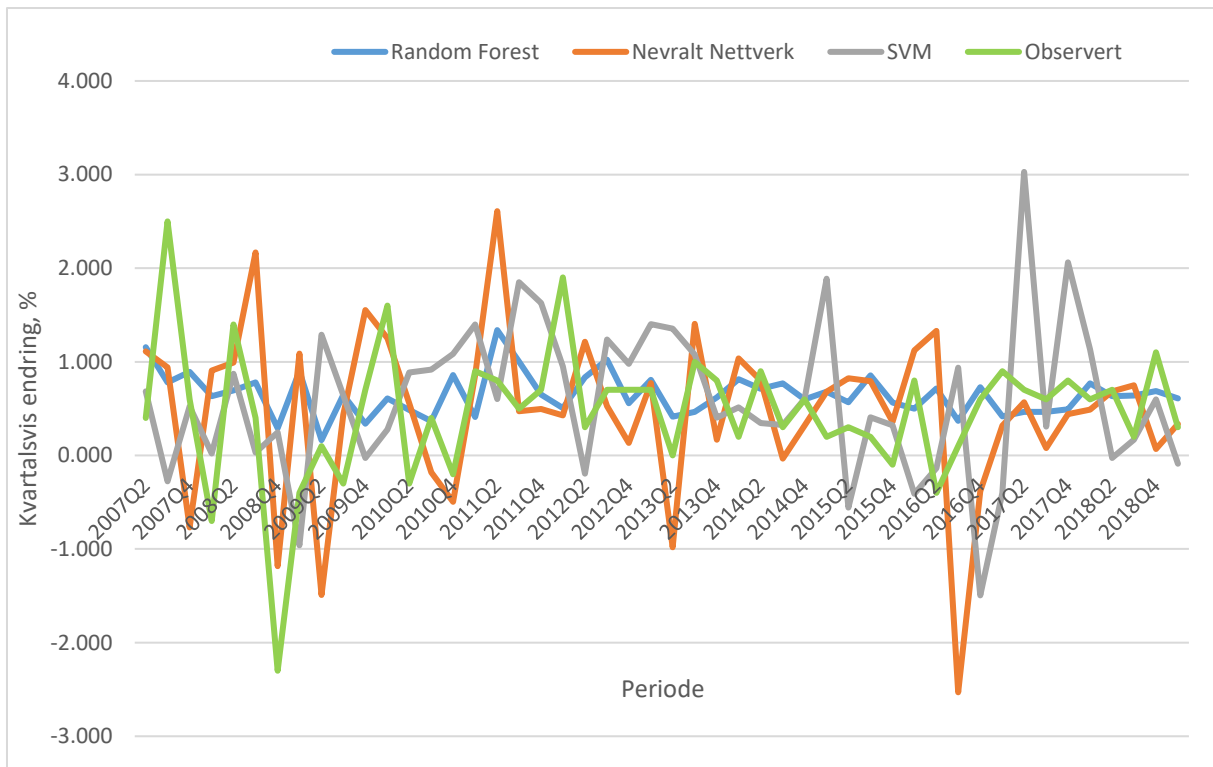


Figur 13: Predikert økonomisk vekst, mot observert. Faktormodell og førsteordens autogregresjon.

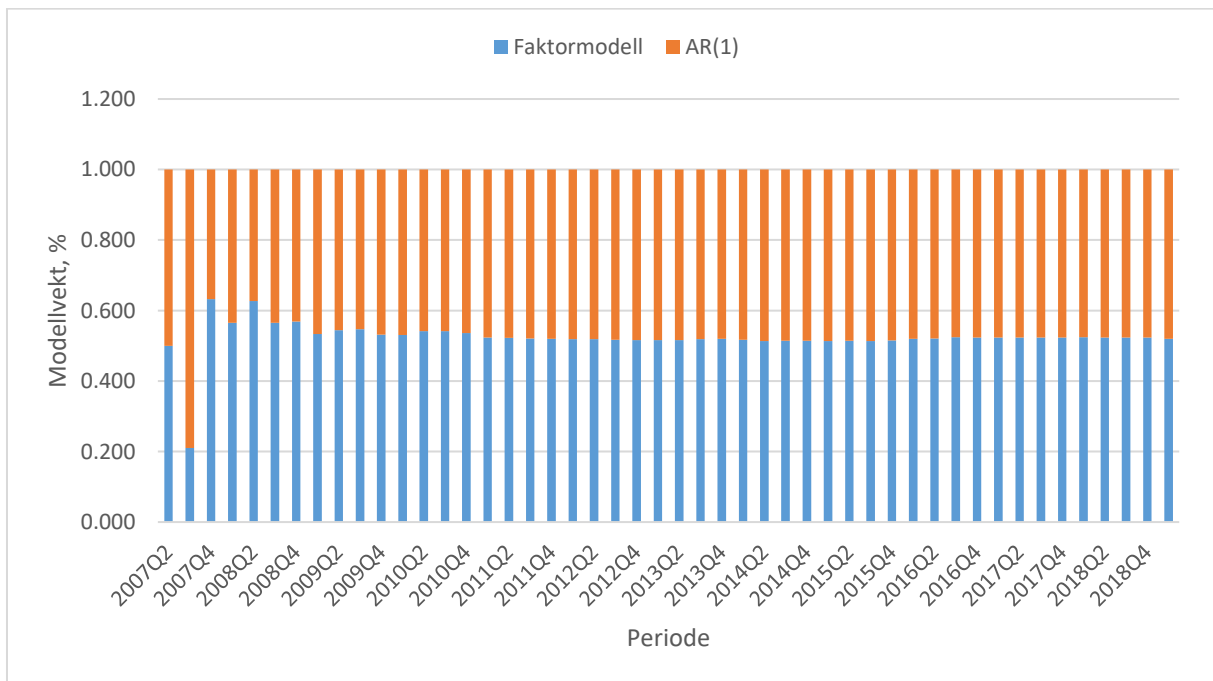
For maskinlæringsmodellene er det ganske uendret siden første måned, det kan virke som om modellene ikke greier å bruke de oppdaterte dataene like effektivt som benchmarkmodellene.



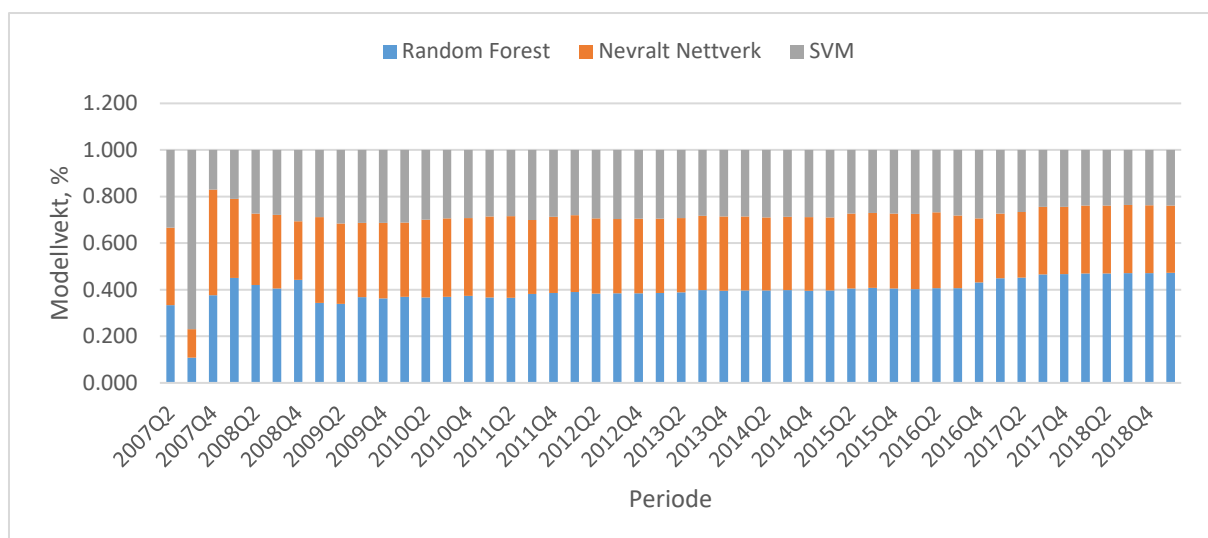
Figur 14: Feilledd over perioden for Random forest, nevralt nettverk og SVM.



Figur 15: Predikert økonomisk vekst, mot observert. Random forest, Nevralt Nettverk, SVM.



Figur 16: Vekting benchmarkmodeller i prosent over perioden



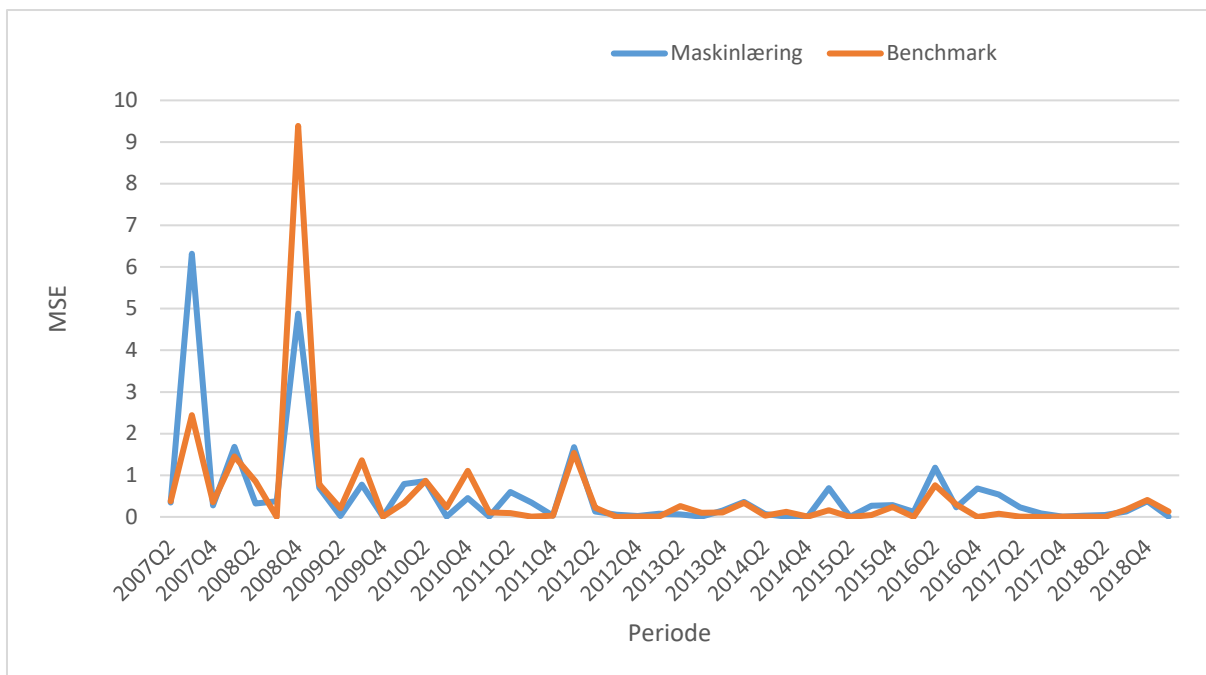
Figur 17: Vekting maskinlæringsmodeller i prosent over perioden.

### Prediksjonsfeil

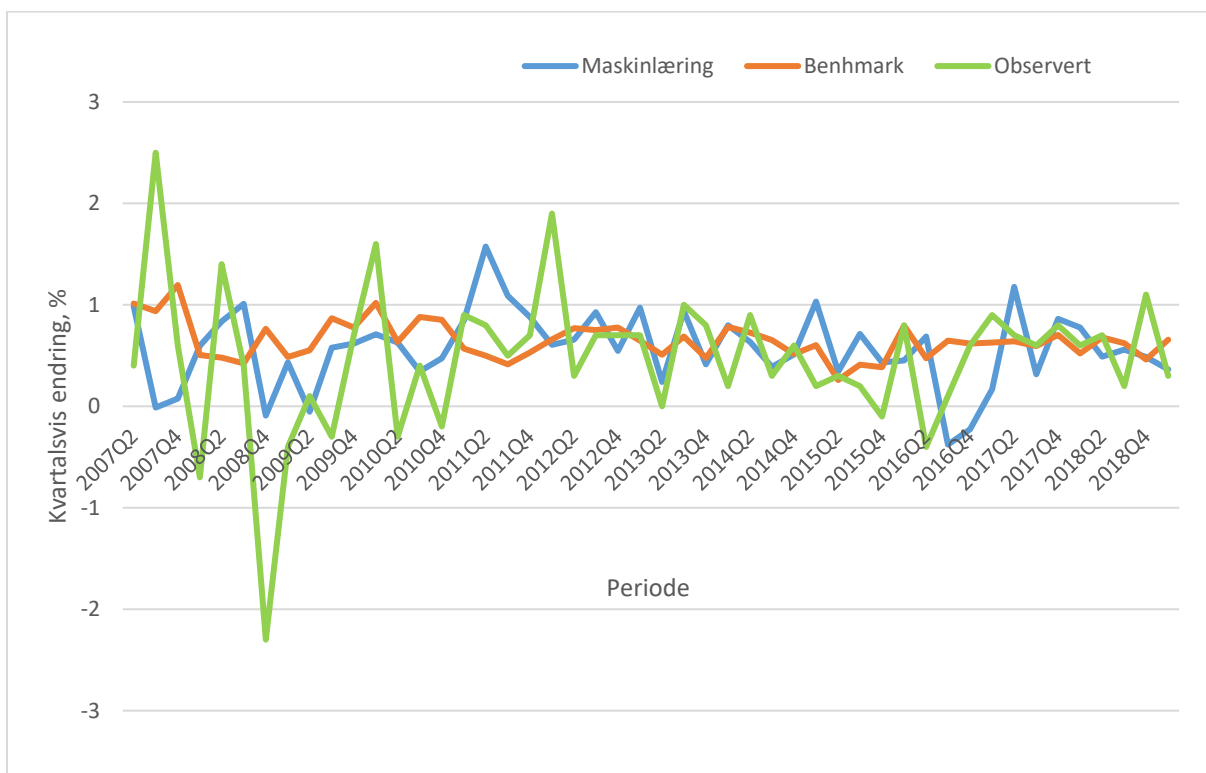
Modell	Vekt	2 måned	
		MSE	%
Benchmark	IMSE	0.522	53.70 %
	Avg	0.509	52.36 %
Maskinlæring	IMSE	0.548	56.37 %
	Avg	0.491	50.55 %

Tabell 5: Prediksjonsfeil for samlemodellene benchmark og maskinlæring.

Nytt i andre måned er også at feilledene for de to samlemodellene varierer over perioden noe mer enn i første måned. Det kan være tegn på at modellene benytter seg av de nye dataene ulikt.

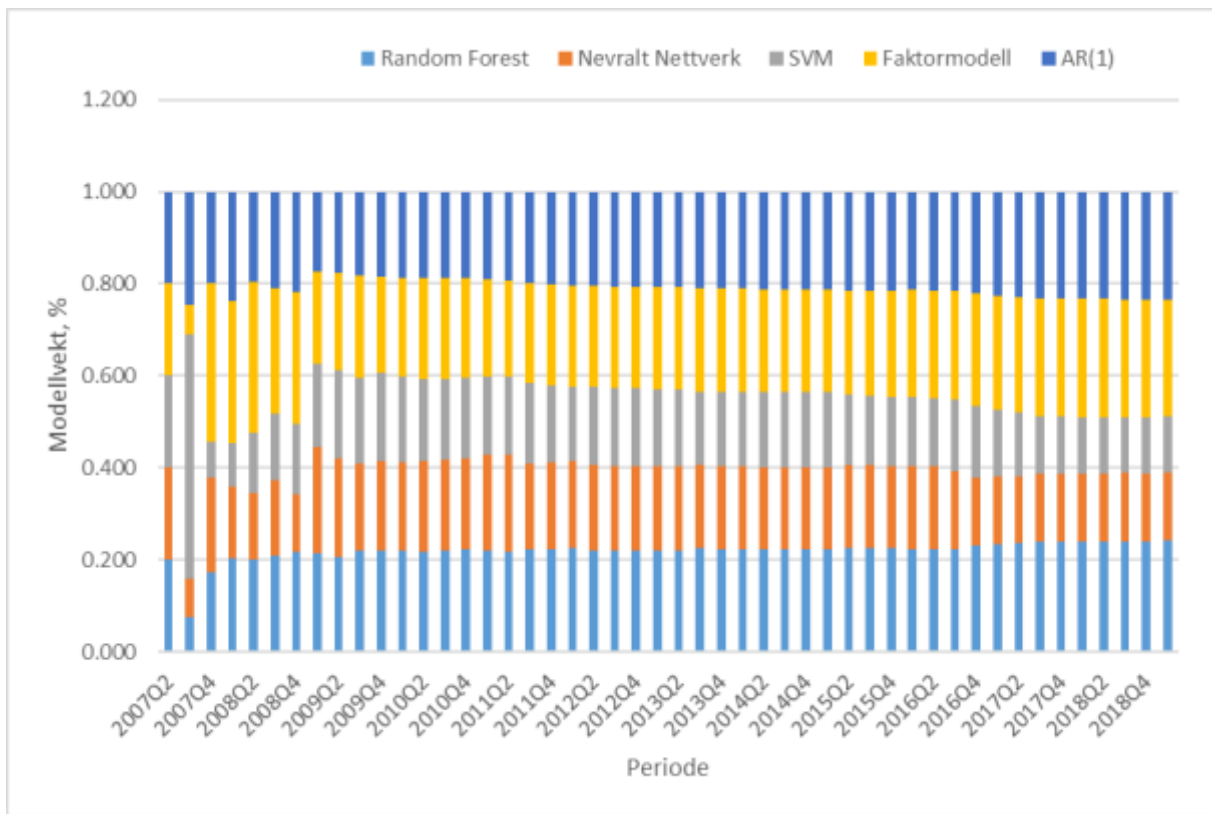


Figur 18: Feilledd for to samlemodeller, benchmark og maskinlæring.



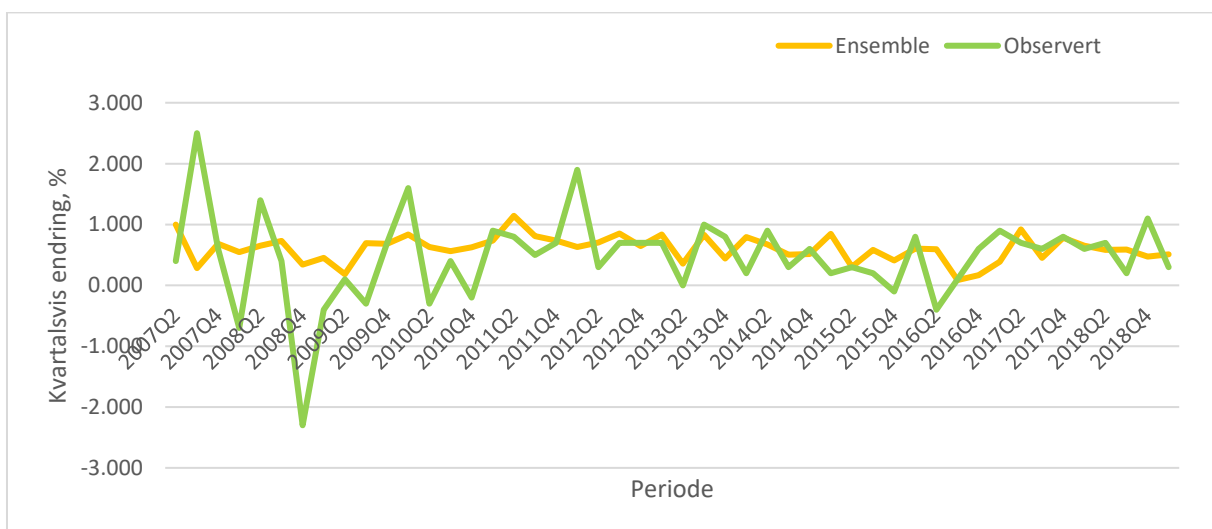
Figur 19: Prediksjon ved maskinlæring og benchmark, mot observerte verdier..

Som i første måned slår vi sammen alle modellene og vurderer sammensetningen.



Figur 20: Modellvektning ensemble.

Vektingen av modellen er ganske likt som i første måned, AR(1) og random forest, de «flate» modellene er relativt høyere vektet enn tidligere. I sum får vi også et flatt ensemble.



Figur 21: Ensembleprediksjon mot observert BNP-vekst.

Overraskende nok er nevralt nettverk forslått som den beste maskinlæringsmodellen å slå sammen med benchmarkmodellene. En mulig forklaring kan være at random forest, som har gjort det best av maskinlæringsmodellene, ligner for mye på benchmark, og derfor ikke bidrar noe til prediksjonen. Nevralt nettverk har vært unøyaktig, men kan bidra til å predikere de svingingene i vekst som benchmarkmodellene ikke får til. Enda en overraskelse er at random walk fortsatt kan forbedre ensemblet med sin inntreden.

### Prediksjonsfeil

Modell	Vekt	2 måned	
		MSE	%
Ensemble	IMSE	0.503	51.77 %
	Avg	0.450	46.36 %
BM + RF	IMSE	0.512	52.74 %
	Avg	0.502	51.72 %
BM + NN	IMSE	0.469	48.33 %
	Avg	0.446	45.90 %
BM + SVM	IMSE	0.578	59.48 %
	Avg	0.541	55.68 %
Ensemble + RW	IMSE	0.486	50.1 %
	Avg	0.436	44.9 %

Tabell 6: Feilledd for ulike sammensetninger av samlemodeller.

### 4.3 Tredje måned

I slutten av kvartalets tredje måned sitter en på finansvariabler fra samme måned, makrovariabler fra andre måned, og industriproduksjon fra første måned i kvartalet.

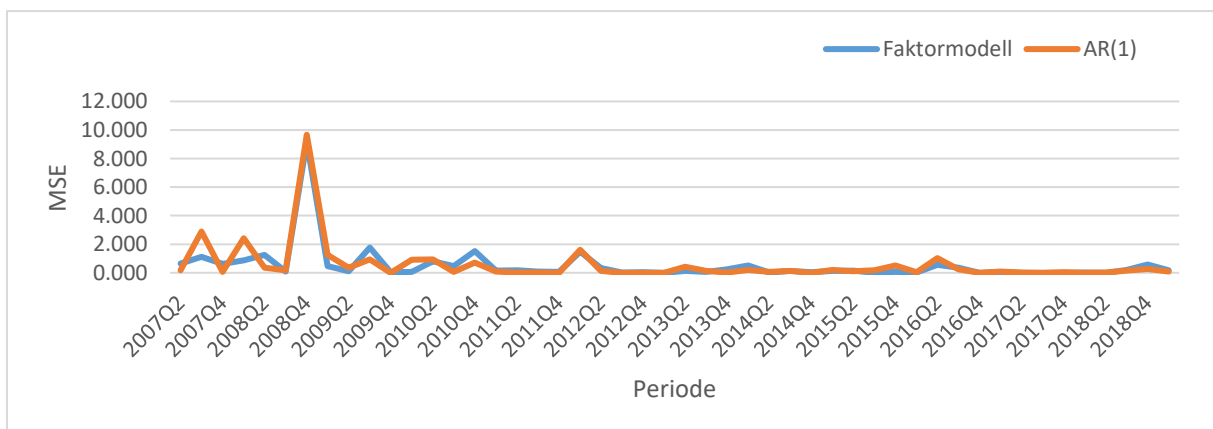
#### Prediksjonsfeil

Modell	3 måned	
	MSE	%
Random Forest	0.524	54.01 %
Nevralt Nettverk	0.609	62.70 %
Support Vector Regression	0.934	96.20 %
Faktormodell	0.483	49.78 %
AR(1)	0.593	61.10 %
Random Walk	0.971	100.00 %

Tabell 7: Gjennomsnittlig feilledd for enkeltmodeller 2000-2018.

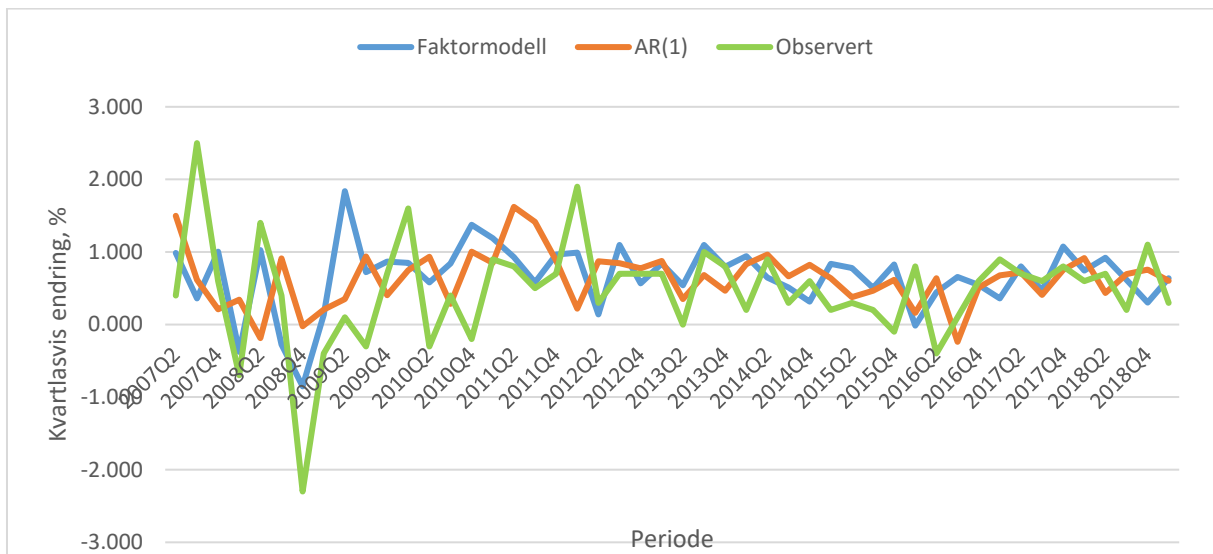
Når en predikerer i slutten av tredje måned har en så oppdaterte data en kan få. Fra tabell 7 er samtlige modeller med unntak av AR(1) mer nøyaktige enn i de tidligere periodene.

Resultatene for prediksjon i andre måned har mye fellestrekk med første måned. Random forest, faktormodell og AR(1) er fortsatt de mest nøyaktige over perioden, men nevralt nettverk og SVM har forbedret seg noe. Nytt i andremånedanalyse er at har mer oppdatert data, der makrotall (utenom industriproduksjon) er kommet inn for kvartalets første måned.



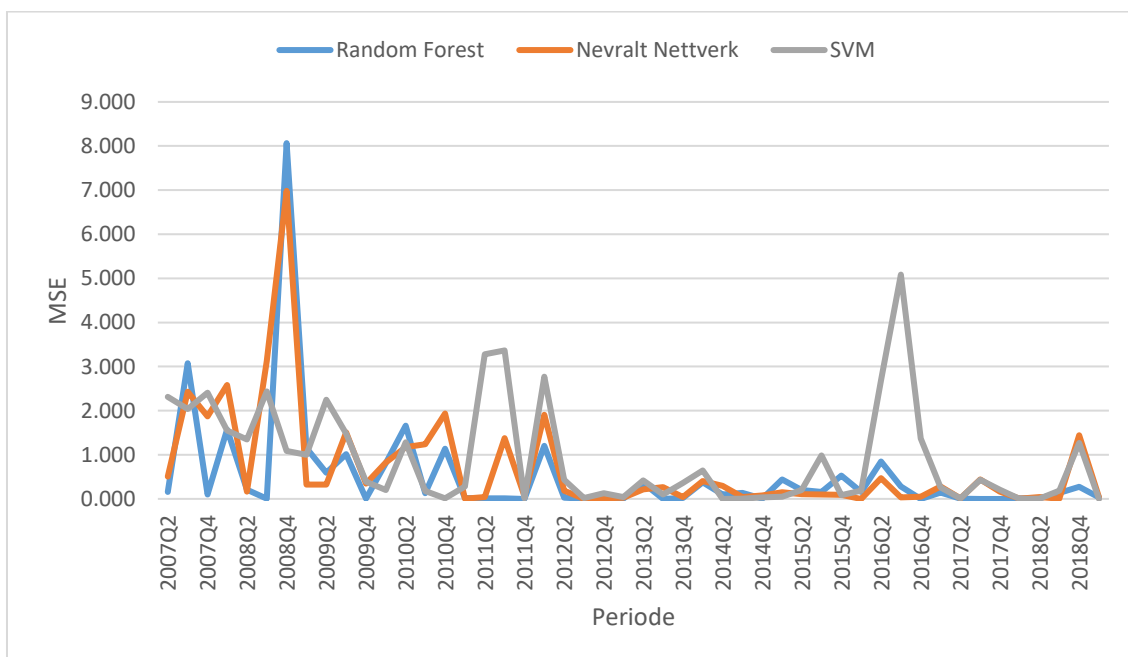
Figur 22: Feilledd over perioden for faktormodell og førsteordens autoregresjon.



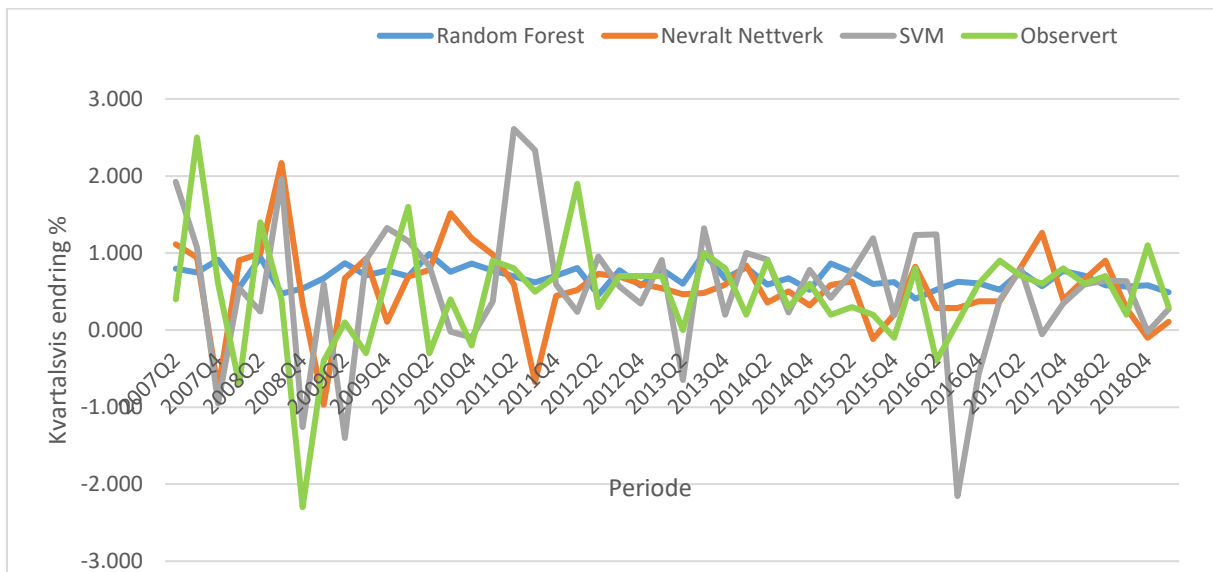


Figur 23: Predikert økonomisk vekst i kvartalets tredje måned, mot observert verdi. Faktormodell og førsteordens autogresjon.

I forhold til tidligere måneder varierer faktormodellen mindre. Maskinlæringsmodellene greier fortsatt ikke predikere finanskrisen med de data vi har brukt, men er mer nøyaktig enn i tidligere måneder. Spesielt nevralt nettverk har lavere feilledd over perioden, mens SVM fortsatt sliter med store avvik.

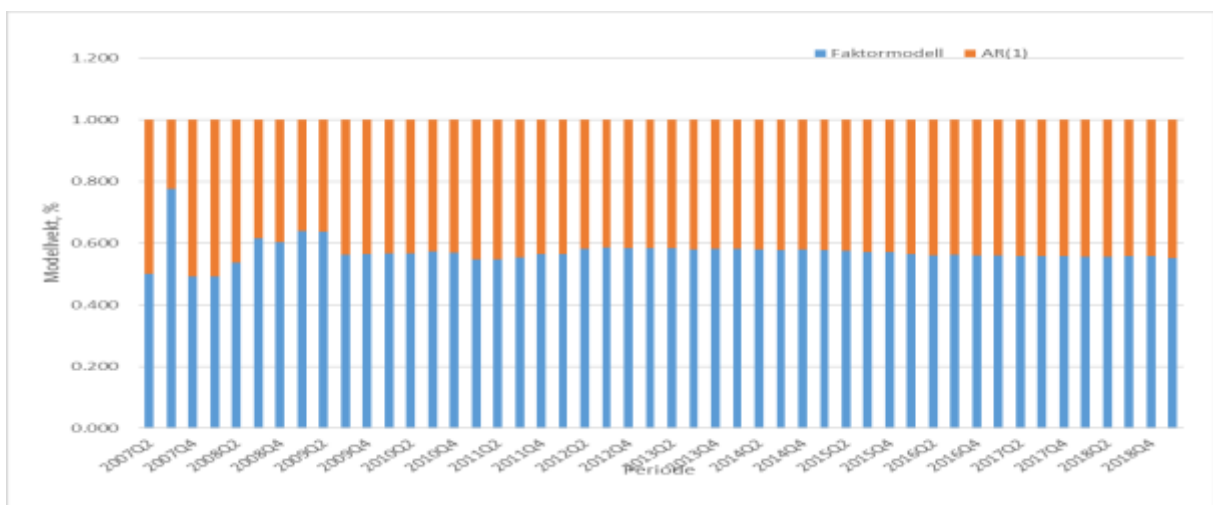


Figur 24: Feilledd over perioden for Random forest, nevralt nettverk og SVM.

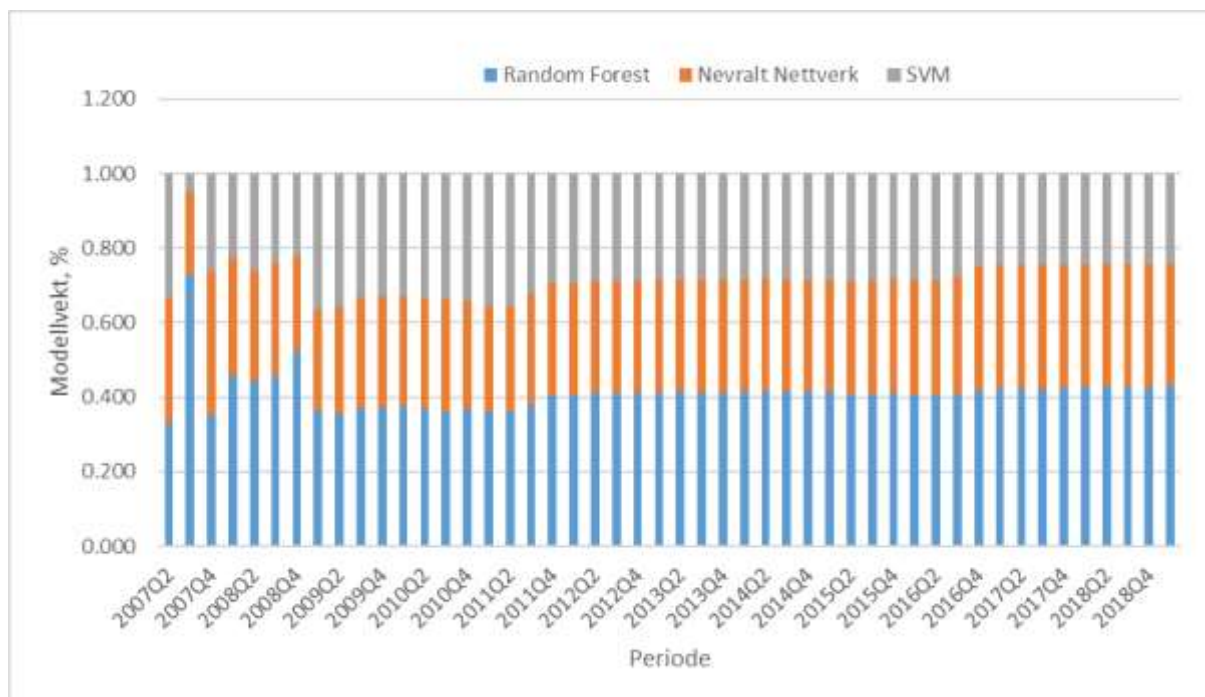


Figur 25: Predikert økonomisk vekst, mot observert. Random forest, Nevralt Nettverk, SVM. Tredje måned.

I tredje måned greier modellene i større grad å henge med når BNP-vekst svinger sterkt. SVM er modellen som er nærmest i å predikerer finanskrisen i 2008, men svinger varierer også mye når observerte verdier ellers holdes stabile. Også nevralt nettverk er blitt bedre i tredje måned, og har for første gang en gjennomsnittlig MSE som er noe lik benchmarkmodellene. Vi slår sammen modellen i hver sine grupper, og ser at vektingen er mye lik som tidligere.



Figur 26: Modellvekting for benchmark for prediksjon i tredje måned. Prosent over perioden

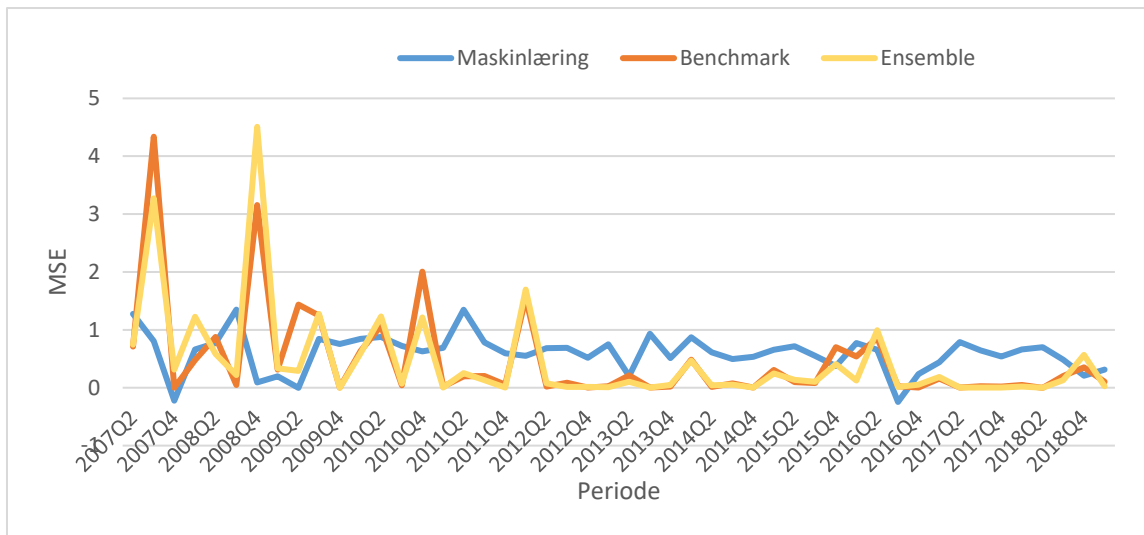


Figur 26: Modellvektning for maskinlæringsmodeller samlet for prediksjon i tredje måned. Prosent over perioden

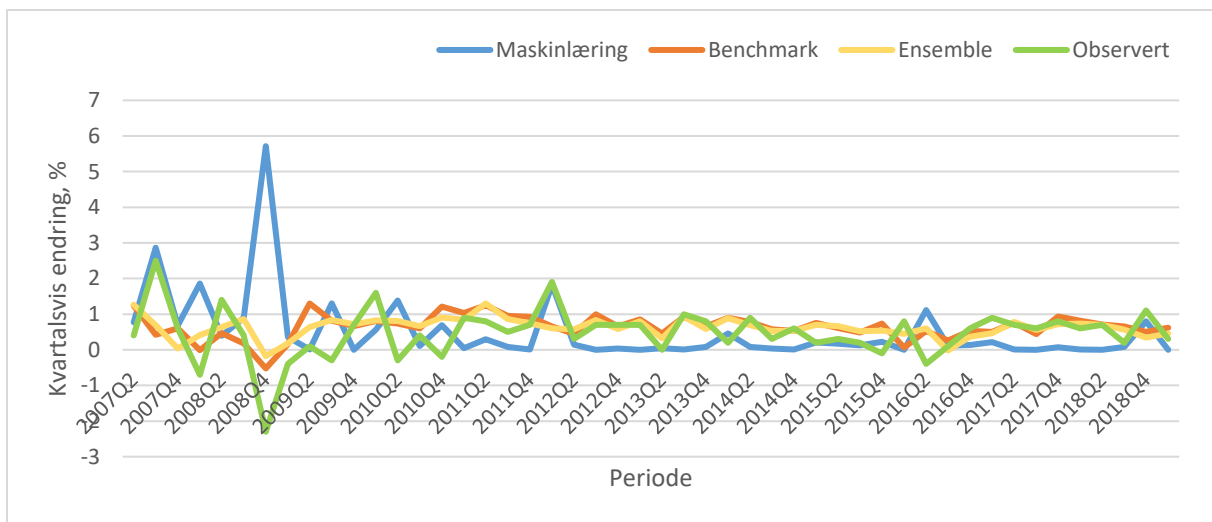
### Prediksjonsfeil

Modell	Vekt	3 måned	
		MSE	%
Benchmark	IMSE	0.476	49.03 %
	Avg	0.470	48.39 %
Maskinlæring	IMSE	0.500	51.49 %
	Avg	0.490	50.46 %

Tabell 8: Prediksjonsfeil for benchmark og maskinlæringsgruppen over perioden, tredje måned.

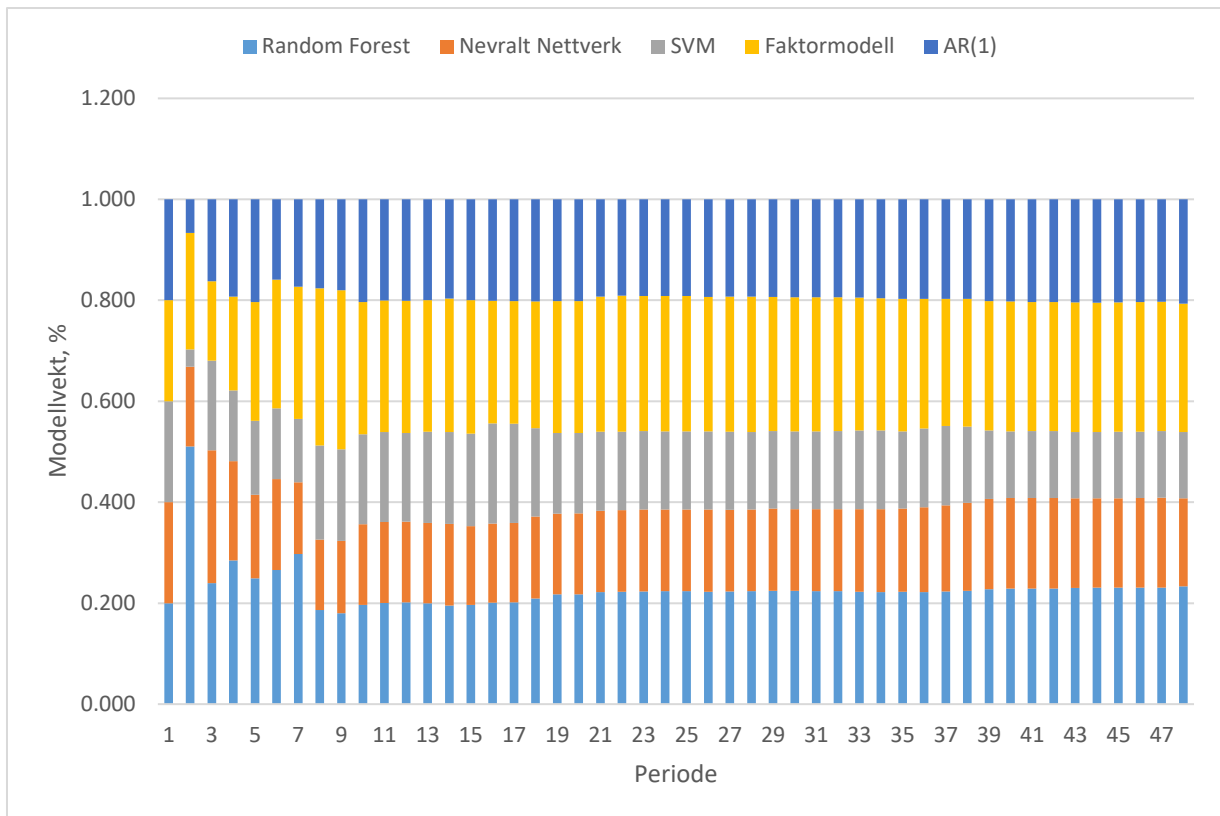


Figur 27: Feilledd for to samlemodeller, benchmark og maskinlæring.



Figur 28: Prediksjon ved maskinlæring, benchmark og ensemblemodell, mot observert BNP-vekst.

Maskinlæringsmodellen samlet er forbedret siden første og andre måned som forventet, men må fortsatt se seg slått av benchmarkgruppen. Når vi kombinerer modellene på ulike måter, greier vi å forbedre prediksjon noe. I figur 27 vises en ensemblemodellens feilledd over perioden, der vi ser at den ligger tett opp mot benchmarkgruppen. Samtidig ser vi at ensemble-prediksjonen ligger tettere opp mot observerte verdier enn før (figur 28).



Figur 29: Modellvekt ensemble, tredje måned.

Til tross for at feilledet i ensemble-modellen er så nær identisk til benchmark, vektene modellene fortsatt noe likt, nesten halvparten benchmark og halvparten maskinlæringsmodeller. SVM og nevral nettverk vektet fortsatt relativt lavt til tross for deres høyere nøyaktighet i tredje måneds prediksjon.

Vi lister opp flere kombinasjoner av samlemodeller for prediksjon i tredje måned. Vi ser at igjen er et fullstendig ensemble med random walk fortsatt den beste kombinasjonen av modeller. Benchmarkgruppens opprinnelige gjennomsnittlige feilled for perioden så vi var 0,476, eller 49,30 % av random walk. Sammenlignet kan benchmark forbedres ved å legge til nevral nettverk eller SVM i tredje måned, mens random forest ikke bidrar til forbedring. Fjerner vi derimot random forest fra ensemblet, øker feilledet til 47,10 % av RW.

### Prediksjonsfeil

Modell	Vekt	3 måned	
		MSE	%
Ensemble	IMSE	0.455	46.85 %
	Avg	0.447	46.04 %
BM + RF	IMSE	0.476	49.00 %
	Avg	0.471	48.50 %
BM + NN	IMSE	0.469	48.26 %
	Avg	0.467	48.10 %
BM + SVM	IMSE	0.469	48.32 %
	Avg	0.472	48.57 %
Ensemble + RW	IMSE	0.438	45.1 %
	Avg	0.428	44.1 %

Tabell 9: Prediksjonsfeil for ulike kombinasjoner av modeller i tredje måned.

#### 4.4 Oppsummering

Vi har predikert kvartalsvis vekst i slutten av hver måned innenfor det aktuelle kvartalet. Som ventet får vi lavere prediksjonsfeil for modellene, med unntak av autoregresjon og random walk, som ikke benytter seg av oppdateringer i forklaringsvariablene. Selv om random forest gjør det bra, er det faktormodellen som jevnt over gjør det best. De andre maskinlæringsmodellene viste seg å være offer for overfitting, der de følger svingninger godt, men samtidig svinger mye ellers.

### **Prediksjonsfeil**

Modell	1. måned		2. måned		3 måned	
	MSE	%	MSE	%	MSE	%
Random Forest	0.576	59.30 %	0.541	55.67 %	0.525	54.01 %
Nevralt Nettverk	0.982	101.06 %	0.886	91.20 %	0.609	62.70 %
Support Vector Regression	1.147	118.12 %	1.068	109.93 %	0.935	96.20 %
Faktormodell	0.595	61.27 %	0.516	53.14 %	0.484	49.78 %
AR(1)	0.542	55.80 %	0.557	57.31 %	0.594	61.10 %
Random Walk	1.187	100.00 %	0.971	100.00 %	0.971	100.00 %

Tabell 10: Prediksjonsfeil for enkeltmodeller over for alle prediksjonshorisonter.

For begge modellgruppene forbedres prediksjon betraktelig i slutten tredje måned, når data for inneværende kvartal for første gang er tilgjengelig for alle variabelgrupper. Selv om noen av modellene gjorde det være enn en naiv random walk, greier vi likevel å forbedre maskinlæringsmodellen ved å ta gjennomsnittet av prediksjonene deres. Det er ikke åpenbart om en bør bruke invers kvadrert residual som basis for vektning, eller gjennomsnittet når en samler modellene, dette varierer over perioden, samt modellgrupper.

### **Prediksjonsfeil**

Modell	Vekt	1.måned		2. måned		3. måned	
		MSE	%	MSE	%	MSE	%
Benchmark	IMSE	0.525	54.07 %	0.522	53.70 %	0.476	49.03 %
	Avg	0.549	56.58 %	0.509	52.36 %	0.47	48.39 %
Maskinlæring	IMSE	0.546	56.23 %	0.548	56.37 %	0.5	51.49 %
	Avg	0.542	55.82 %	0.491	50.55 %	0.49	50.46 %

Tabell 11: Prediksjonsfeil for benchmark og maskinlæringsmodeller for alle prediksjonshorisonter.

## Prediksjonsfeil

Modell	Vekt	1. måned		2. måned		3. måned	
		MSE	%	MSE	%	MSE	%
Ensemble	IMSE	0.546	45.99 %	0.503	51.77 %	0.455	46.85 %
	Avg	0.547	46.11 %	0.45	46.36 %	0.447	46.04 %
BM + RF	IMSE	0.54	45.50 %	0.512	52.74 %	0.476	49.00 %
	Avg	0.534	44.97 %	0.502	51.72 %	0.471	48.50 %
BM + NN	IMSE	0.547	46.10 %	0.469	48.33 %	0.469	48.26 %
	Avg	0.555	46.76 %	0.446	45.90 %	0.467	48.10 %
BM + SVM	IMSE	0.552	46.51 %	0.578	59.48 %	0.469	48.32 %
	Avg	0.572	48.14 %	0.541	55.68 %	0.472	48.57 %
Ensemble + RW	IMSE	0.518	43.60 %	0.486	50.10 %	0.438	45.10 %
	Avg	0.511	43.10 %	0.436	44.90 %	0.428	44.10 %

Tabell 12: Prediksjonsfeil for ensemblemodeller for alle prediksjonshorisonter.

Ved å samle alle modellene i analysen, greier vi å få prediksjonsnøyaktighet som er bedre enn hvilken som helst av enkeltmodellene. Dette gjelder i større grad i kvartalets siste måned, mens de i første og andre varierer litt om en bør kombinere benchmark med kun én av maskinlæringsalgoritmene. Kanskje det mest overraskende resultatet er at random walk kombinert med de andre modellene i analysen skulle gi det beste resultatet over alle prediksjonshorisontene, for gjennomsnittet for hele perioden. En annen overraskelse er at samlemodellene gjør det dårligere i andre måned, enn i første, spesielt for maskinlæring. Det kan tyde på overfitting, at maskinlæringsalgoritmene legger for stor vekt på variabler som er blitt publisert i andre måned, blant annet realøkonomiske tidsserier fra forrige kvartal. Samtidig finner også Aastveit og Trovik (2012) at å inkludere internasjonale makrovariabler gir dårligere prediksjon, en mulig forklaring på det vi ser her. Samtidig kan de gode resultatene fra første måned forklares ved finansvariablers fremoverskuende egenskaper, da finansverdier ikke bare beskriver nåtid, men en forventning om fremtiden.

## 5.0 Konklusjon

Vi har gjennomgått og evaluert treffsikkerheten til seks ulike modeller på simulert sanntidsdata, til å predikere kvartalsvis BNP-vekst. Prediksjonstidspunkt har funnet sted i



slutten av kvartalets første, andre og tredje måned. Datasettet har bestått av 128 tidsserier, som skal fange opp ulike sider ved norsk økonomi. Publiseringstidspunktet for ulike tiddserier for en periode varierer veldig, vi har derfor tatt hensyn til dette ved å evaluere modellene etter hvor oppdatert datasettet er. Vi har sett at det hjelper på nøyaktigheten å være nærmere kvartalsslutt. At maskinlæringsalgoritmer kan forbedre prediksjonen stemmer med funn hos Richardson (2018).

Vi har funnet at maskinlæringsmodeller ikke er noe hellige gral som garanterer gode prediksjoner. Selv om maskinlæringsmodellene gjerne plukker opp variasjoner i kvartalsvis endring bedre enn tradisjonelle modeller, bommer de også kraftigere. Istedenfor finner vi at maskinlæringsmodeller passer godt til å kombinere med tradisjonelle modeller, og kan gi modellkombinasjoner som gir bedre prediksjoner enn tradisjonelle modeller greier på egen hånd.

I en fremtidig analyse bør variabelviktigheten studeres med ved å predikere hver gang en variabel publiseres i løpet av måneden. På den måten kan en se i hvilken grad modellene benytter seg av data ulikt. En kan også inkludere sentimentvariabler, «myke data» fra spørreundersøkelser og forventingsindekser.

## Referanseliste

- Aastveit, K. A., Gerdrup, K. R., & Jore, A. S. (2011) Short-term Forecasting of GDP and Inflation in Real-time: Norges Bank's System for Averaging Models. Norges Bank Staff Memo 9/2011.
- Aastveit, K. A., & Trovik, T. (2012) Nowcasting norwegian GDP: the role of asset prices in a small open economy. *Empirical Economics*, 42 (1), 95-119.
- Aastveit, K. A., Gerdrup, K., Jore, A. S., & Thorsrud, L. (2014) Nowcasting GDP in real time: a density combination approach. *Journal of Business and Economic Statistics*, 32(1), 48-68
- Ahmed, N. K., Atiya, A. F., Gayar, N. E. & El-Shishiny, H. (2010) 'An empirical comparison of machine learning models for time series forecasting', *Econometric Reviews* 29(56), 594-621.
- Baffigi, A., Golinelli, R. & Parigi, G. (2004) Bridge models to forecast the euro area GDP, *International Journal of Forecasting*, 20(3) 447-460.
- Bates, J. & Granger, C. (1969) The Combination of Forecasts. *Journal of the Operational Research Society*, 20(4), 451-468.
- Breiman, L. (2001) Random Forests. *Machine Learning*, 45(1), 5–32.
- Claeskens G., Magnus J., Vasnev A., Wang W. (2016) The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting*, 32 (3), 754-762. <https://doi.org/10.1016/j.ijforecast.2015.12.005>
- Chakraborty, C., & Joseph, A., Machine Learning at Central Banks (September 1, 2017). Bank of England, Working Paper, 674. <https://dx.doi.org/10.2139/ssrn.3031796>

- Doz C., Reichlin L. (2011) A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Econometrics*, 164 (1), 188.  
<https://doi.org/10.1016/j.jeconom.2011.02.012>
- Galton, F. (1907). Vox Populi. *Nature*, 75, 450–451.
- Ghysels, E., Santa-Clara, P. & Valkanov, R. (2004) The MIDAS Touch: Mixed Data Sampling Regression Models . *Cirano Working Paper Series*.
- Hastie, T., Tibshirani, R. & Friedman, J. (2008) The Elements of Statistical Learning. Springer, 2008. ISBN 978-0-387-84858-7
- Hendry, D., & Clements, M. (2004). "Pooling of Forecasts." *Econometrics Journal*, 7(1), 1-31.
- Sax C., Eddelbuettel D. (2018). Seasonal Adjustment by X-13ARIMA-SEATS in R. *Journal of Statistical Software*, 87(11), 1–17. doi:10.18637/jss.v087.i11.
- Makridakis S., Spiliotis E. &Assimakopoulos V. (2018) Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE* 13(3):  
<https://doi.org/10.1371/journal.pone.0194889>
- Richardson, A., Mulder, T. & Vehbi, T. (2018) Nowcasting New Zealand GDP using machine learning algorithms. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3256578
- Rosenblatt, F. (1957) The Perceptron--a perceiving and recognizing automaton. Report 85 460-1, Cornell Aeronautical Laboratory.
- Rünstler, G., Barhoumi, K., Benk, S., Cristadoro, R., den Reijer, A., Jakaitiene, P., Jelonok, A., Rua, K., Ruth, K., van Nieuwenhuize, C. (2009). Short-term forecasting of GDP using large datasets: a pseudo real-time forecast evaluation exercise, *Journal of Forecasting*, 28(7), 595-611.
- Schumacher C, Breitung J (2008) Real-time forecasting of german GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting* 24(3):386–398.
- Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222.

Teräsvirta, T., van Dijk, D. & Medeiros, M. (2005) Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: a re-examination, *International Journal of Forecasting* 21(4), 755 – 774.

Timmermann, A. (2006). *Forecast Combinations. Handbook of Economic Forecasting*, vol 1. Tilgjengelig fra <https://www.elsevier.com/books/book-series/handbook-of-economic-forecasting> (Hentet 18.03.18).

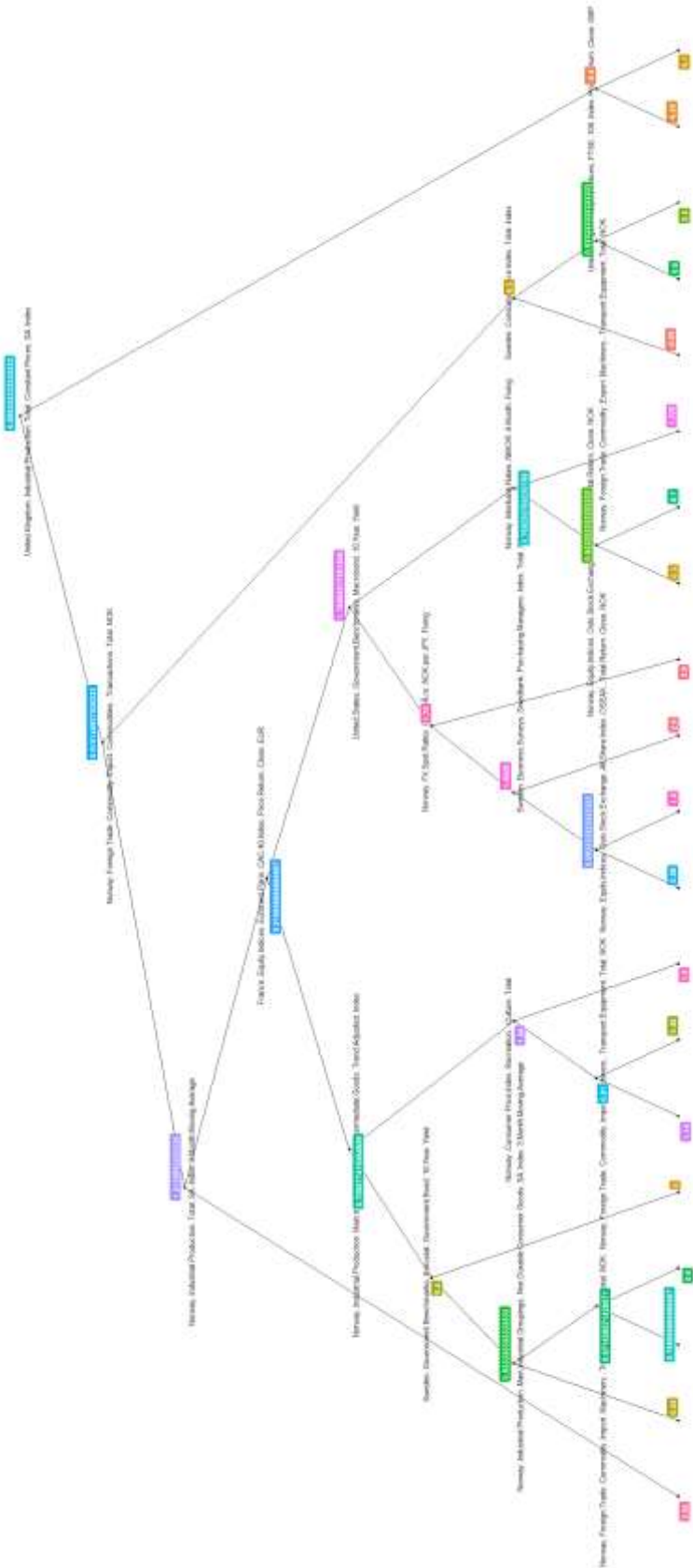
Vapnik, V. (1995) *The Nature of Statistical Learning Theory*. New York: Springer.

Varian, H. (2014) Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, 28(2), 3- 28.

Watson, M. & Stock, J. (2016) *Factor Models and Structural Vector Autoregressions in Macroeconomics*. Princeton, 2016. Tilgjengelig fra: <https://www.princeton.edu/mwatson/papers/StockWatsonDFMHOM030916.pdf> (Hentet: 23.03.18)

# Appendix

## A.1 Variabelviktighet



## A.2 Tidsserietransformasjon

Tidsseriene er transformert for å gjøre de stasjonær.

1 = ingen transformasjon

2 = førstedifferansen,

3 = førstedifferanse på logaritmeform

	Tidsserie	Trans
BNP	Gross Domestic Product, Mainland, Constant Prices, Change P/P	1
Finans	Central Bank of Norway, Trade Weighted Index, TWI	3
	Central Bank of Norway Import Weighted Krone Index, I44	3
	OBX All Share Index, Total Return, Close, NOK	3
	OBX Consumer Discretionary, Index, Total Return, Close, NOK	3
	OBX Consumer Staples, Index, Total Return, Close, NOK	3
	OBX Energy, Index, Total Return, Close, NOK	3
	OBX Financials, Index, Total Return, Close, NOK	3
	OBX Health Care, Index, Total Return, Close, NOK	3
	OBX Industrials, Index, Total Return, Close, NOK	3
	OBX Information Technology, Index, Total Return, Close, NOK	3
	OBX Materials, Index, Total Return, Close, NOK	3
	OBX Telecommunication Services, Index, Total Return, Close, NOK	3
	OBX, Utilities, Index, Total Return, Close, NOK	2
	FX Spot Rates, NOK per EUR, Fixing	3
	FX Spot Rates, NOK per GBP, Fixing	3
	FX Spot Rates, NOK per USD, Fixing	3
	FX Spot Rates, NOK per SEK, Fixing	3
	FX Spot Rates, NOK per SDR, Fixing	3
	FX Spot Rates, NOK per JPY, Fixing	3
	FX Spot Rates, NOK per DKK, Fixing	3
Sentiment	Leading Indicators, Retriever Norway AS, Financial News Index, FNI	2
Finans, utland	United States, S&P500 Index, Price return, Close, USD	3
	United States, Interbank rates, LIBOR, 3 month, Fixing	2
	Sweden, Interbank rates, STIBOR, 3 month, Fixing	2
	Euro Area, STOXX 50, Index, Price Return, Close, EUR	3
	United Kingdom, FTSE 100, Index, Price Return, Close, GBP	3
	Germany, Deutsche Boerse, DAX 30, Index, Total Return, Close, EUR	3
	France, Euronext Paris, CAC 40, Index, Price Return, Close, EUR	3
	Sweden, Nasdaq OMX, Stockholm Index, Price Return, Close, SEK	3
	United Kingdom, FTSE 350, Index, Dividend Yield	2
	United States, Government Benchmark, 3 Month, Yield	3
	United States, Government Benchmark, 10 Year, Yield	3
	United States, Government Benchmarks, Eurostat, Government Bond, 10 Year, Yield	2
	Sweden, Government Benchmarks, Eurostat, Government Bond, 10 Year, Yield	2
	Euro Area, Government Benchmarks, Eurostat, Government Bond, 10 Year, Yield	2
Rente	Interbank Rates, NIBOR, 1 Month, Fixing	2
	Interbank Rates, NIBOR, 3 Month, Fixing	2
	Interbank Rates, NIBOR, 6 Month, Fixing	2
	Government Benchmarks, 2 Year, Yield, End of Period	2
	Government Benchmarks, 5 Year, Yield, End of Period	2
	Government Benchmarks, 10 Year, Yield, End of Period	2
	Government Benchmarks, 10 Year, Yield, End of Period	2
Råvarepriser	Leading Indicators, Citi, Commodity Terms of Trade Index	3
	Crude Oil, Brent, Spot, North Sea, Close, USD	3
	World, Aluminum, LME Official Prices, USD	3
Arbeidsmarked	Unemployment, Registered, Males & Females	2
	Employment, Employed Persons (LFS), Males & Females, By Age, Total 15-74 Years	3

Industriproduksjon	Industrial Production, Main Industrial Groupings, Intermediate Goods, Index	3
	Main Industrial Groupings, Energy Goods, Index	3
	Main Industrial Groupings, Energy Goods, Index	3
	Main Industrial Groupings, Intermediate Goods, Index	3
	Main Industrial Groupings, Non-Durable Consumer Goods, Index	3
	Main Industrial Groupings, Capital Goods, Index	3
	Industrial Production, Total, Index	3
	Manufacturing, Printing, Reproduction, Index	3
	Manufacturing, Rubber, Plastic & Mineral Products, Index	3
	Manufacturing, Furniture & Manufacturing N.E.C., Index	3
	Manufacturing, Ships, Boats & Oil Platforms, Index	3
	Manufacturing, Fabricated Metal Products, Index	3
	Manufacturing, Textiles, Wearing Apparel, Leather, Index	3
	Manufacturing, Refined Petroleum, Chemicals, Pharmaceuticals, Index	3
	Manufacturing, Basic Metals, Index	3
	Manufacturing, Computer & Electrical Equipment, Index	3
	Manufacturing, Transport Equipment N.E.C, Index	3
	Manufacturing, Wood & Wood Products, Index	3
	Manufacturing, Non-Ferrous Metals, Index	3
	Manufacturing, Food Products, Index	3
	Manufacturing, Basic Chemicals, Index	3
	Manufacturing, Paper & Paper Products, Index	3
	Manufacturing, Food, Beverages & Tobacco, Index	3
	Manufacturing, Repair, Installation of Machinery, Index	3
	Manufacturing, Total, Index	3
Konsumpriser	Consumer Price Index, Total, Index	3
	Consumer Price Index, Transport, Total	3
	Consumer Price Index, Communication, Total	3
	Consumer Price Index, Recreation & Culture, Total	3
	Consumer Price Index, Restaurants & Hotels, Total	3
	Consumer Price Index, Miscellaneous Goods & Services, Total	3
	Consumer Price Index, Food & Non-Alcoholic Beverages, Total	3
	Consumer Price Index, Alcoholic Beverages & Tobacco, Total	3
	Consumer Price Index, Clothing & Footwear, Total	3
	Consumer Price Index, Housing, Water, Electricity, Gas & Other Fuels, Total	3
	Consumer Price Index, Health, Total	3
	Leading Indicators, Citi, Inflation Surprise Index	2
Blandet, utland	United States, Business Surveys, Federal Reserve Bank of Philadelphia, Business Outlook Survey, Manufactur	2
	United States, Consumer Surveys, Conference Board, Consumer Confidence Index, Total	3
	Euro Area, Economic Surveys, DG ECFIN, Economic Sentiment Indicator, Balance, SA	2
	Sweden, Business Surveys, Swedbank, Purchasing Managers' Index, Total, SA	2
	Euro Area, HICP, All-Items, Index	3
	United Kingdom, Consumer Price Index, Total, Index	3
	United.States..Business.Surveys..ISM..Report.on.Business..Non.Manufacturing..Purchasing.Managers..Index.	2
	United.States..Business.Surveys..ISM..Report.on.Business..Manufacturing..Purchasing.Managers..Index..SA	2
	United States, Industrial Production, Industry Group, Manufacturing, Total (SIC), SA, Index	3
	Sweden, Manufacturing, Total, Calendar Adjusted, Constant Prices, SA, Index	3
	Euro Area 19, Industrial Production, Total Excluding Construction, Calendar Adjusted, SA, Index	3
	United States, Consumer Price Index, All Urban Consumers, U.S. City Average, All Items Less Food & Energy, S	3
	Sweden, Consumer Price Index, Total, Index	3
	United Kingdom, Industrial Production, Total, Constant Prices, SA, Index	3

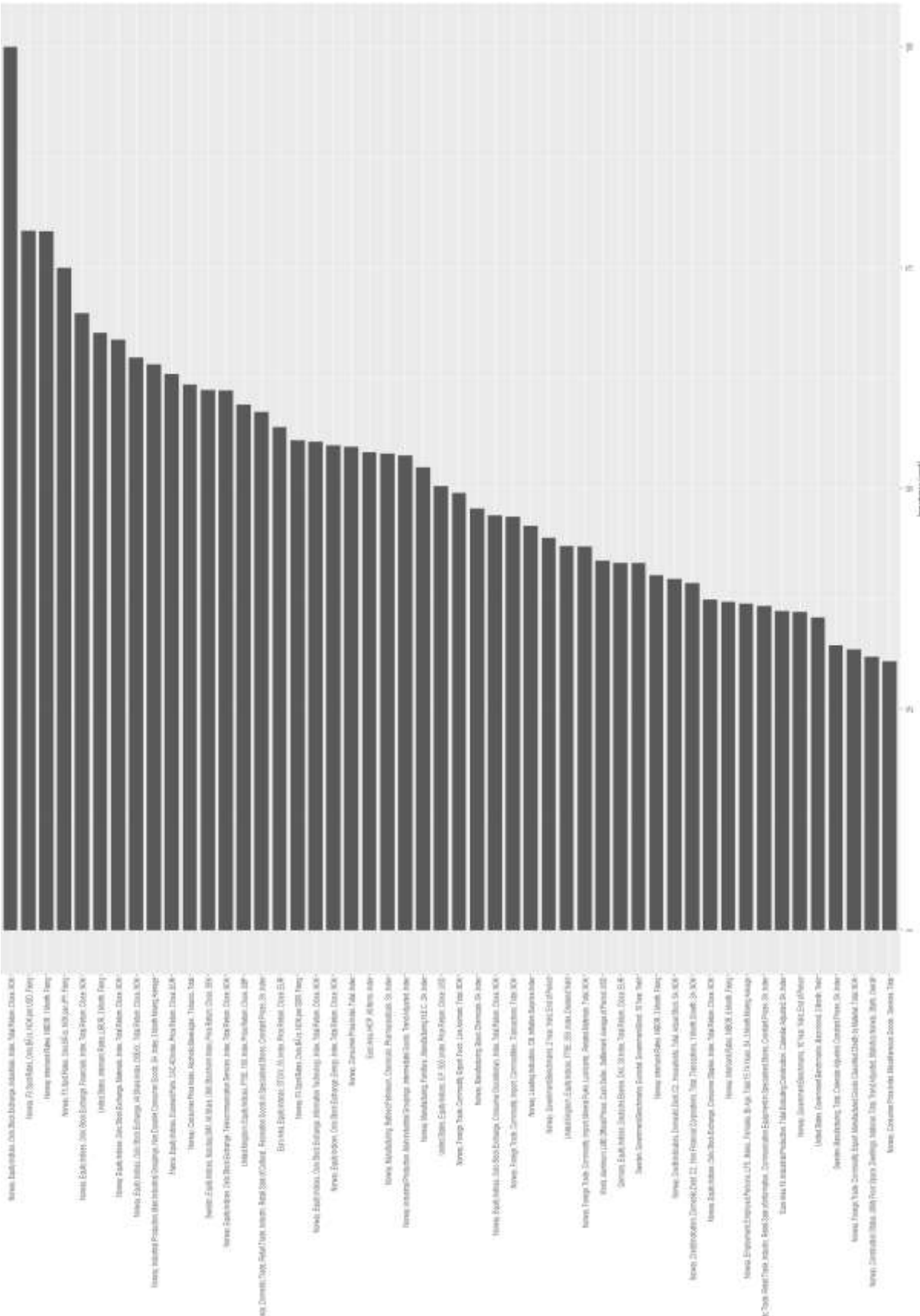
Eksport & import	Foreign Trade, Commodity, Export, Animal & Vegetable Oils, Fats & Waxes, Total, NOK	3	
	Foreign Trade, Commodity, Export, Beverages & Tobacco, Total, NOK	3	
	Foreign Trade, Commodity, Export, Commodities & Transactions, Total, NOK	3	
	Foreign Trade, Commodity, Export, Crude Materials, Inedible, except Fuels, Crude Fertilizers & Crude Mineral	3	
	Foreign Trade, Commodity, Export, Food & Live Animals, Total, NOK	3	
	Foreign Trade, Commodity, Export, Machinery & Transport Equipment, Total, NOK	3	
	Foreign Trade, Commodity, Export, Manufactured Goods Classified Chiefly by Material, Total, NOK	3	
	Foreign Trade, Commodity, Export, Mineral Fuels, Lubricants & Related Materials, Total, NOK	3	
	Foreign Trade, Commodity, Export, Miscellaneous Manufactured Articles, Total, NOK	3	
	Foreign Trade, Commodity, Export, Crude Oil, SA, NOK	3	
	Foreign Trade, Commodity, Import, Animal & Vegetable Oils, Fats & Waxes, Total, NOK	3	
	Foreign Trade, Commodity, Import, Beverages & Tobacco, Total, NOK	3	
	Foreign Trade, Commodity, Import, Commodities & Transactions, Total, NOK	3	
	Foreign Trade, Commodity, Import, Crude Materials, Inedible, except Fuels, Total, NOK	3	
	Foreign Trade, Commodity, Import, Food & Live Animals, Total, NOK	3	
	Foreign Trade, Commodity, Import, Manufactured Goods Classified Chiefly by Material, Total, NOK	3	
	Foreign Trade, Commodity, Import, Machinery & Transport Equipment, Total, NOK	3	
	Foreign Trade, Commodity, Import, Mineral Fuels, Lubricants & Related Materials, Total, NOK	3	
	Boligbygging	Construction Status, Number, Dwellings, National, Total, Statistics Norway, Starts, Overall	3
		Construction Status, Utility Floor Space, Dwellings, National, Total, Trend Adjusted, Statistics Norway, Starts, (	3
Detailhandel	Total, except of Motor Vehicles & Motorcycles, Constant Prices, SA, Index.x	3	
	Retail Sale of Automotive Fuel in Specialised Stores, Constant Prices, SA, Index	3	
	Retail Sale of Information & Communication Equipment in Specialised Stores, Constant Prices, SA, Index	3	
	Retail Sale of Other Household Equipment in Specialised Stores, Constant Prices, Index	3	
	Retail Sale of Cultural & Recreation Goods in Specialised Stores, Constant Prices, SA, Index	3	
	Industry, Retail Sale of Other Goods in Specialised Stores, Constant Prices, SA, Index	3	
	Industry, Retail Trade Not in Stores, Stalls & Markets, Constant Prices, SA, Index	3	
Kredittindikator	Credit Indicators, Domestic Debt in NOK (C1), Transactions, 1-Month Growth, SA, NOK	2	
	Credit Indicators, Domestic Debt (C2), Non-Financial Corporations, Total, Transactions, 1-Month Growth, SA, N	2	
	Credit Indicators, Domestic Debt (C2), Households, Total, Actual Stock, SA, NOK	3	
	Credit Indicators, Domestic Debt (C2), All Sectors, Total, Actual Stock, SA, NOK	3	

### A.3 Prosent av total varians forklart ved komponenter i faktormodell

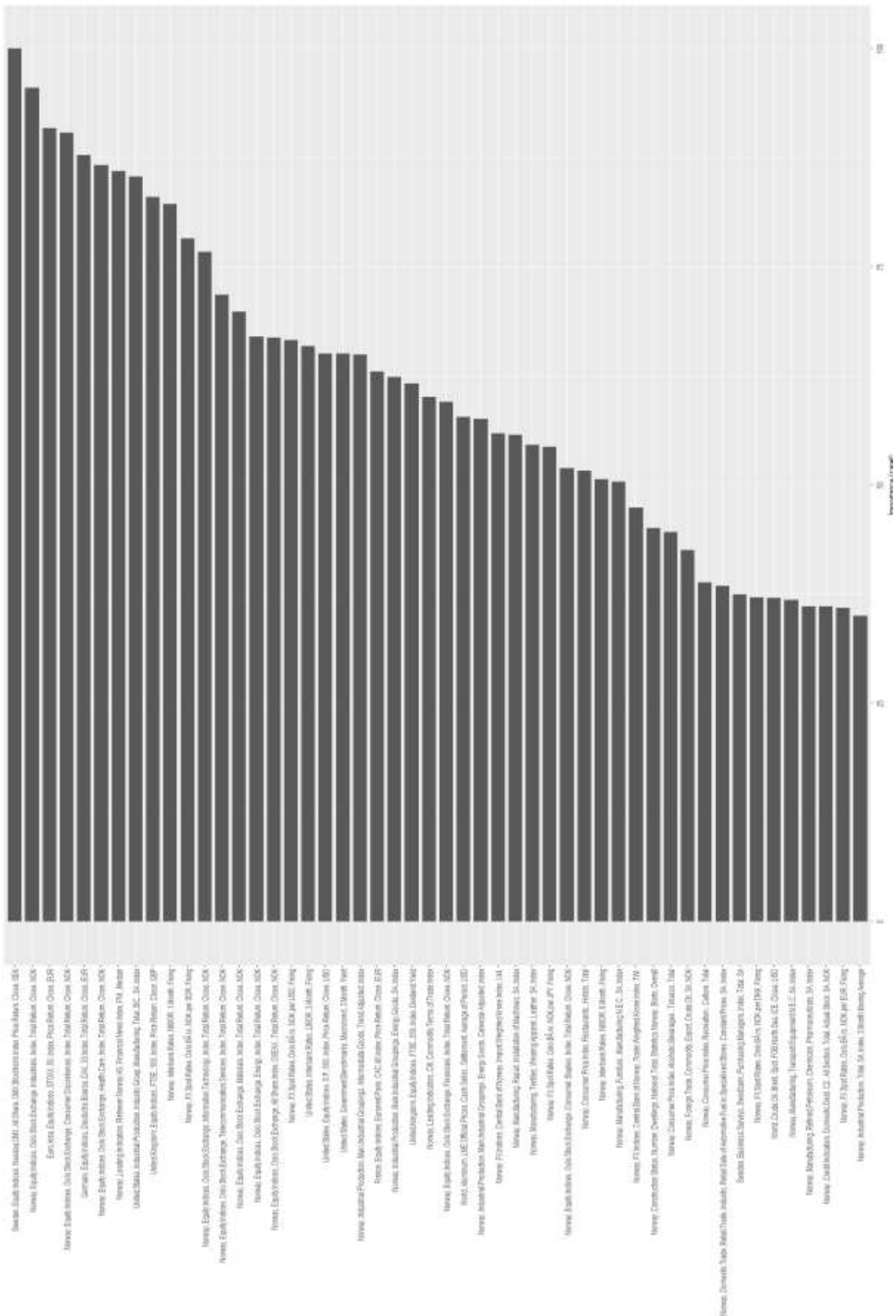
Faktorantall	1	2	3	4	5	6	7	8	9	10
Andel av varians	0.1376	0.06737	0.05873	0.04736	0.03807	0.03613	0.0333	0.02948	0.0276	0.02585
Akkumulert	0.1376	0.20497	0.2637	0.31106	0.34913	0.38526	0.41856	0.44804	0.47564	0.50149



# A.4.1 Variabelviktighet i random forest, første måned



## A.4.2 Variabelviktighet i random forest, andre måned



### A.4.3 Variabelviktighet i random forest, tredje måned

