

Aksel Jansen og Mia Åsheim

Fremtidsprediksjon av spotprisen på olje

En sammenligning av prediksjonsevnen til VECM og LSTM

Masteroppgave i Finansiell økonomi

Veileder: Joakim Blix Prestmo

Mai 2021

Aksel Jansen og Mia Åsheim

Fremtidsprediksjon av spotprisen på olje

En sammenligning av prediksjonsevnen til VECM og LSTM

Masteroppgave i Finansiell økonomi
Veileder: Joakim Blix Prestmo
Mai 2021

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi



Kunnskap for en bedre verden

Forord

Masteroppgaven representerer slutten av masterstudiet i Finansiell Økonomi ved Institutt for Samfunnsøkonomi ved Norges Teknisk-Naturvitenskapelige Universitet (NTNU). Oppgaven utgjør 30 studiepoeng. Vi vil takke vår veileder, Joakim Blix Prestmo, for samarbeidet gjennom oppgaven.

Innholdet i denne oppgaven står for forfatterens regning.

Trondheim, Mai 2021

Aksel Jansen

Mia Åsheim

Abstract

In this study, the predictability of two different models will be compared while forecasting the oil price. The models used are a vector error correction model (VECM) based on economic theory and a Long-short term memory (LSTM) model based on machine learning.

The analysis consists of monthly observations from 2001 (1) to 2018 (12) (in-sample period), while the forecasts are over the time interval 2019 (1) to 2021 (3) (out-of-sample period). The comparison is based on the results from the future predictions of one and three months, where the models will be evaluated based on the evaluation criteria's root mean square error (RMSE) and mean absolute percentage error (MAPE).

Two future forecasts-periods will be used in this study, a rolling forecast of one month, $t + 1$, and of three months, $t + 3$. The comparison is based on the results from the future predictions of one and three months ahead, where the models will be evaluated based on the evaluation criteria RMSE and MAPE.

From the analysis of the future predictions, we found that the models performed very similarly. For the $t + 1$ predictions, LSTM outperformed VECM, with a difference of 0.55 USD/barrel for RMSE and 0.57% for MAPE. While for the $t + 3$ predictions, VECM outperformed LSTM. The difference here was only 0.08 USD/barrel for RMSE and 0.90% for MAPE.

We do not consider the results for the $t + 1$ and $t + 3$ predictions to be satisfactory, neither for the VECM nor the LSTM model. The reason for this is that the deviations are of a significant size, given RMSE and MAPE.

A possible explanatory variable for the models' results may be the choice of time-period. Based on this, we choose to divide the out-of-sample period into two new evaluation periods, to rule out that the time-period plays a significant role. These periods are further called pre and post Covid-19". Here we find that new results are better for the $t + 1$ forecasts compared to the whole out-of-sample period, while the results for $t + 3$ are worse, given RMSE and MAPE. This may indicate that it is the model specification that has the greatest effect on both models' predictions, and not the time-period itself.

The conclusion about which model performs best is ambiguous, as none of the models stands out significantly. Considering the six different time horizons the forecast is performed on, it is LSTM that gives the smallest deviations four of these six periods, given RMSE and MAPE. Based on this, it may indicate that the LSTM model provides the best future predictions on the oil prices. However, we would like to point out that the differences are small, and it is recommended to make further analyzes before a conclusion is made.

Sammendrag

I denne studien skal vi sammenligne prediksjonsevnen til to forskjellige modeller på oljeprisen. Modellene som benyttes er en vector error correction modell (VECM) basert på økonomisk teori og en Long-short term memory (LSTM) modell som er basert på maskinlæring.

Analysen tar utgangspunkt i månedlige observasjoner fra 2001(1) til 2018(12) (in-sample periode), mens prognosene er over tidsintervallet 2019(1) til 2021(3) (out-of-sample periode). Det vil bli benyttet to fremtidsprognoser i denne studien, en rullerende prognose for én måned frem i tid, $t + 1$, og en for tre måneder frem i tid, $t + 3$. Sammenligningen baserer seg på resultatene fra fremtidsprediksjonene på én og tre måneder, der modellene vil bli evaluert ut fra evalueringskriteriene root mean square error (RMSE) og mean absolute percentage error (MAPE).

Fra analysen av fremtidsprediksjonene ga modellene tilnærmet like resultater. Gitt resultatene for $t + 1$ prediksjonene gjør LSTM det bedre enn VECM, med en differanse på 0.55 USD/fat for RMSE og 0.57% for MAPE. For $t + 3$ prediksjonene gjør VECM det bedre enn LSTM, differansen var her kun på 0.08 USD/fat for RMSE og 0.90% for MAPE.

Vi anser ikke resultatene for $t + 1$ og $t + 3$ prediksjonene å være tilfredsstillende, verken for VECM eller LSTM modellen. Grunnen til dette er at avvikene er av en betydelig størrelse, gitt RMSE og MAPE.

En mulig forklaringsvariabel til modellenes resultater kan være valg av tidsperiode. Med bakgrunn i dette velger vi å dele out-of-sample perioden i to nye evalueringsperioder, for å utelukke at det er tidsperioden som spiller en vesentlig rolle. Disse periodene blir videre kalt ”pre og post Covid-19”. Her finner vi at nye resultatene blir bedre for $t + 1$ prognosene sammenlignet med hele perioden, mens resultatene for $t + 3$ blir dårligere. Dette kan tyde på at det er modellspesifikasjonen som gir størst utslag på prediksjonene for begge modellene, og ikke tidsperioden i seg selv.

Konklusjonen om hvilken modell som presterer best er tvetydige, da ingen av modellene skiller seg vesentlig ut. Tatt i betraktning de seks forskjellige tidshorisontene prediksjon-

nene er utført på, er det LSTM som gir de minste avvikene fire av disse seks periodene, gitt RMSE og MAPE. Med bakgrunn i dette kan det tyde på at LSTM modellen gir de beste fremtidsprediksjonene på oljeprisen. Vi ønsker likevel å gjøre oppmerksom på at differansene er små og det anbefales å gjøre ytterlige analyser før en endelig konklusjon tas.

Innhold

Tabeller	ix
Figurer	xi
1 Innledning	1
1.1 Problemstilling	1
2 Litteratur	3
3 Metode Vector error correction modell	6
3.1 Stasjonaritet	6
3.2 Kointegrasjon	7
3.3 Valg av antall laggede verdier	8
3.4 VAR	9
3.5 VECM	11
3.6 Diagnosetester	13
3.7 Fremtidsprognoser	14
3.8 Evalueringskriterier	15
4 Metode Long-short term memory modell	17
4.1 Nevrale nettverk	17
4.1.1 Tapsfunksjon	19
4.2 LSTM modellen	19
4.3 Trening av nettverk	21

4.4	Valideringsdata	22
4.5	Fremtidsprognoser	23
4.6	Modellseleksjon	24
4.6.1	Komponenter	24
4.6.2	Optimeringprosessen	27
5	Data	29
5.1	Valg av variabler	29
5.2	Deskriptiv statistikk	31
6	Empiriske resultater og diskusjon	35
6.1	Vector error correction modell	35
6.1.1	Stasjonaritet	35
6.1.2	Kointegrasjon	36
6.1.3	Valg av laggede verdier	37
6.1.4	Diagnosetester	38
6.1.5	Kortsiktige og langsiktige forhold	40
6.2	LSTM	43
6.3	Fremtidsprognoser	45
6.3.1	RMSE og MAPE	45
6.3.2	RMSE og MAPE pre- og post Covid-19	48
7	Oppsummering og konklusjon	51
8	Videre forskning og begrensninger	53

Referanser	55
Appendiks	58
A Tabeller	58

Tabeller

1	Deskriptiv statistikk av variablene i nominelle verdier	33
2	Deskriptiv statistikk av variablene i differensierte verdier	33
3	Korrelasjonsmatrise av variablene i nominelle verdier	34
4	Korrelasjonsmatrise av variablene differensierte verdier	34
5	Augmented Dickey-Fuller test nominelle verdier	35
6	Augmented Dickey-Fuller test differensierte verdier	36
7	Johansen kointegrasjonstest	37
8	Valg av laggede verdier	37
9	Lagrange-multiplier test for autokorrelerte residualer	38
10	Jarque-Bera test for normalfordelte residualer	40
11	Modellseleksjonen fra GA-optimeringen	44
12	Antall vektorer i LSTM-nettverket	44
13	RMSE og MAPE for hele tidsperioden	45
14	Jarque-Bera's test for normality	47
15	RMSE og MAPE $t + 1$ før og etter Covid-19	49
16	RMSE og MAPE $t + 3$ før og etter Covid-19	50
17	Oversikt over variabel-koder og datakilder.	58
18	Kortsiktige forhold VECM	58
19	ECT & Johansen normalisering av langsiktige forhold	59
20	Operatortypene anvendt i GA-optimering	59
21	Gyldige allele-verdier i GA-optimering	59

22	Parameterene til GA-optimeringen av de to LSTM-strukturene som selek- teres.	60
23	Dimensjonen til $LSTM_{t+1}$	60
24	Dimensjonen til $LSTM_{t+3}$	61
25	Deskriptiv statistikk av modellenes prediksjonsavvik	61

Figurer

1	Oppbygning til et kunstig nevralt nettverk	18
2	Illustrasjon av de forskjellige aktiveringsfunksjonene	19
3	Oppbygning til et LSTM-nettverk.	21
4	Overfitting og underfitting	22
5	Inndeling av data-periodene.	23
6	Illustrasjon av et kromosom av varianten bit-string.	25
7	Flytdiagram av optimeringsprosessen til den genetiske algoritmen.	28
8	Variablenes utvikling fra 2001(4)-2021(3) i nominelle verdier	31
9	Variablene utvikling fra 2001(4)-2021(3) i differensierte verdier	32
10	Egenverdi stabilitet VECM	39
11	Grafisk illustrasjon av individene under GA-optimeringen	44
12	Modellenes fremtidsprediksjoner og prediksjonsavvik	46
13	Fordelingen til modellenes prediksjonsavvik.	47

Forkortelser

ADF – Augmentetd Dickey-Fuller

AIC – Akaike informasjons-kriterium

ANN – Artificial neural network (*norsk, kunstig nevralt nettverk*)

ANFIS – Adaptive neural-based fuzzy inference system

ARDL – Autoregressive distributed lag model

ARIMA – Autoregressive integrated moving average

ARMA – Autoregressive–moving-average

BCF – Brent crude oil futures price

BCS – Brent crude oil spotprice

CBOE – Chicago board option exchange

EA – Evolutionary algorithm (*norsk, evolusjonær algoritme*)

GA – Genetisk algoritme

GARCH – Generalized autoregressive conditional heteroskedasticity

GOP – Global oljeproduksjon

HL – Hidden layer (*norsk, skjult nivå*)

HQIC – Hannan-Quinn informasjons kriterium

IC – Informasjonskriterium

IL – Input layer (*norsk, input-nivå*)

LSTM – Long short-term memory

MAPE – Mean absolute percentage error

NN – Neural network (*norsk, nevralt nettverk*)

OL – Output layer (*norsk, output-nivå*)

OPEC – Organization of the Petroleum Exporting Countries

RMSE – Root mean squared error

RNN – Recurrent neural network (*norsk, tilbakevendende nevralt nettverk*)

VAR – Vector autoregression

VECM – Vector error correction model (*norsk, vektor-feilkorrigeringsmodell*)

VIX – Volatility index

YCI – Monthly year-to-year change in oil inventory

1 Innledning

Olje står i dag for den største delen av verdens energiforbruk, noe som gjør at oljeprisen har en betydelig effekt på verdensøkonomien. Uventede svingninger kan føre til uønskede økonomiske konsekvenser for alle involverte parter, slik vi blant annet så i forbindelse med oljekrisen i 2014. Det samme så vi igjen i våren 2020 da Covid-19 utviklet seg til en verdenspandemi som resulterte i ekstreme lave oljepriser og uroligheter i markedet. I forbindelse med dette hadde var oljeprisen (også kjent som brent råolje eller brent crude) nede i en sluttkurs på 9.12 USD/fat den 21 april, Blessing (2021).

I Norge har petroleumsvirksomheten siden funnet av Ekofisk i 1969 og frem til i dag, vært en av de viktigste næringene, målt i statlige inntekter, verdiskapning og finansiering av det norske velferdssamfunnet. For 2021 er det anslått at statens samlede netto kontantstrøm fra denne sektoren er på hele 99 milliarder kroner, Regjeringen.no (2020). Selv om inntektene fra petroleumssektoren er redusert betraktelig de siste årene, er dette fortsatt en av de større inntektspostene for staten.

Gode fremtidsprognoser på oljeprisen er av denne grunn av stor nasjonal og internasjonal interesse, der mange forskere og økonomer har viet mye tid til å utvikle forskjellige modeller til akkurat dette formålet. Tidligere studier, for eksempel Alquist, Kilian og Vigfusson (2013), har basert seg på alt fra enkle til mer avanserte økonomiske modeller som inkluderer flere variabler som varierer over tid. Det har også i senere tid blitt svært vanlig å bruke modeller basert på maskinlæring i slike prognoser. Valg av modell og metode har nå blitt en like viktig oppgave som selve fremtidsprediksjonen. Oljeprisen er imidlertid en av de vanskeligste og mest komplekse handelsvarene å gi nøyaktige prognoser på. Grunnene til dette er at oljeprisen er har høy volatilitet og har en ikke-lineære sammenheng.

1.1 Problemstilling

Motivasjonen for denne studien er at resultatene vi har funnet fra tidligere forskning som sammenligner slike modeller har vært misvisende. I lignende studier har det blitt sammenlignet modeller, der en av modellene ikke har vært egnet til analyser av mer komplekse tidsserier. Denne studien tar utgangspunkt i en Long-short term memory (LSTM) modell og en Vector error correction model (VECM). Grunnen til at disse modellene er valgt er

at de har relativt lik struktur og oppbygning, men samtidig forskjellige da de baserer seg på forskjellige teorier og metoder. Formålet med studien er å evaluere, og sammenligne prediksjonsevnen til de to forskjellige modellene, for så å konkludere hvilken modell som er best.

I motsetning til LSTM har VECM flere betingelser som må være oppfylt for at modellen skal være valid og korrekt spesifisert. En av VEC-modellens betingelser er at variablene som inngår i modellen må være stasjonære, noe som ikke er tilfellet for LSTM. Variablene som inngår i studien, blir av denne grunn valgt slik at betingelsene for VECM er oppfylt. På denne måten kan modellenes resultater sammenlignes på likt variabel grunnlag.

Gitt studiens problemstilling: ***Fremtidsprediksjon av spotprisen på Nordsjøolje - En sammenligning av prediksjonsevnen til VECM og LSTM***, håper vi å kunne gi en mer korrekt sammenligning av modellene prediksjonsevne.

Oppgavens oppbygging er som følger: I kapittel 2 går vi gjennom tidligere forskning og resultatene fra disse. I kapittel 3 gjennomgås metoden for den økonomiske modellen, mens i kapittel 4 blir metoden til maskinlærings-modellen gjennomgått. Variablene og datasettet blir presentert i kapittel 5. I kapittel 6 blir analysen av begge modellene presentert, før resultatene fra fremtidsprediksjonene sammenlignes og diskuteres. En oppsummering og studiens konklusjon blir gitt i kapittel 7, før forslag til videre forskning og studiens begrensninger blir gjennomgått i kapittel 8.

2 Litteratur

Forskere og økonomer har lenge vært opptatte av å identifisere driverne av oljeprisen, og det er skrevet mange studier om hvordan disse driverne påvirker oljeprisen på kort og lang sikt. Hovedformålet til samtlige av disse studiene har vært å lage gode fremtidsprediksjoner på oljeprisen, der modellene som har blitt benyttet til prognosene har variert.

I studien til McCallum og Wu (2005), sammenlignes det blant annet to modeller for å predikere oljeprisen. Én modell som kun baserer seg på futures, og én modell som baserer seg på spredning mellom futures- og spotprisen. Der studien konkluderer med at sistnevnte modell gir de beste prognoser på både kort og lang sikt. Videre konkluderes det med at det å inkludere oljefutures i analyser kan gi nyttig informasjon om fremtidige bevegelser i oljeprisen. Futures inneholder nyttig informasjon om investorers forventinger om den fremtidige oljeprisen. Dette er begrunnet med at partene (både kjøper og selger) er enige om prisen for leveringstidspunktet når avtalen inngås.

Alquist og Kilian (2010) sammenlignet i sin studie ble flere forskjellige modeller basert på oljefutures benyttet til å predikere oljeprisen. I studien bruker de en random walk modell som benchmark-modell. Denne benchmarken ble så sammenlignet med en modell som bruker oljefutures direkte, og flere modeller som bruker spredningen mellom futures- og spotprisen. I studien konkluderte de med at futures-modellen, og spredningsmodellen der spredningen var satt til en, slår random walk modellen på alle tidshorisonter.

I likhet med studiene over finner Manescu og Van Robays (2014) at en direkte futures-modell gir bedre fremtidsprediksjoner enn en random walk modell.

Problemet med random walk-, no-change-, og univariate modeller, er at de ikke fanger opp dynamikken i markedet. Vector autoregressiv (VAR) og VECM har av denne grunnen blitt godt anvendte modeller til slike komplekse analyser da de kan inneholde mange innsatsfaktorer og fange opp deres samlede påvirkning på oljeprisen. I en studie av Alquist, Kilian og Vigfusson (2013) ble det benyttet ulike VAR modeller til å predikere oljeprisen, der disse ble sammenlignet med enkle futures-modeller som diskutert over. I studien ble det konkludert med at VAR-modellene ga mer presise prediksjoner enn random walk modellene, opp til en tidshorison på 6 måneder. Etter dette ga ikke modellene signifikant

bedre resultater.

Baumeister og Kilian (2014) sammenligner i sin studie ulike VAR modeller for å predikere oljeprisen på kvartalsvis data. Studien bygger på den tidligere studien av Alquist, Kilian og Vigfusson (2013). Resultatene viser at VAR-modellene kun gjorde det bedre enn random walk modellen for prediksjoner på en og to kvartaler frem i tid.

I en studie gjort av Beckers og Beidas-Strom (2015) ble flere ulike VAR-modeller estimert til å gi prognoser på oljeprisen. Resultatene viste forbedringer på evalueringskriteriene sammenlignet med random walk modellen. Benchmark VAR-modellene var bedre på alle tidshorisonter med unntak av 24 måneder, mens den utvidede VAR-modellen var bedre på kort sikt, 1-3 måneder.

Resultatene fra samtlige av studiene varierer, men de har til felles at VAR-modellene er mer presise enn random walk og no-change-modellene på kort sikt. Det må her tas i betraktning at de ulike studiene benytter ulike innsatsfaktorer, laggede verdier og tidshorisonter.

De senere årene har det blitt mer og mer vanlig å benytte modeller som baserer seg på maskinlæring i studier som omhandler fremtidsprognoser. Flere av disse studiene sammenligner også disse modellene med tradisjonelle tidsseriemodeller basert på økonomisk teori.

I en studie gjort av Sagheer og Kotb (2019), blir flere forskjellige maskinlæringsmodeller og en autoregressive integrated moving average (ARIMA) modell sammenlignet med en Deep Long-short term memory (DLSTM) modell. Formålet med studien er å predikere den fremtidige oljeproduksjonen, der modellene blir sammenlignet basert på root mean square error (RMSE). Studien konkluderte med at ARIMA hadde de største avvikene, mens DLSTM ga de beste fremtidsprognoser, sammenlignet med alle de andre modellene.

Wang mfl. (2009) benyttet i en studie flere varianter av maskinlæringsmodeller, som artificial neural network (ANN), og adaptive neural-based fuzzy inference system (ANFIS) for å nevne noen, og en autoregressive–moving-average (ARMA) modell til å gi prognoser på månedlige utslipp for vannkraft. ARMA modellen ble i denne studien benyttet som benchmark, og gitt evalueringskriteriene var det denne som ga de dårligste prediksjone-

ne. Studien konkludert med at samtlige av modellene var overlegen sammenlignet med benchmark-modellen.

I en studie gjort av Kristjanpoller og Minutolo (2016) sammenlignes prediksjonsevnen til en ANN og en generalized autoregressive conditional heteroskedasticity (GARCH) modell, med en ANN-GARCH modell. Studien viser at en kombinasjon av den økonomiske modellen og det nevralt nettverket ga de beste fremtidsprognosene. Studien la lite vekt på om det var ANN- eller GARCH modellen som presterer best, da disse ikke blir sammenlignet med hverandre.

Verstyuk (2020) sammenligner i sin studie en VAR og et multivariate recurrent neural network (MRNN). Formålet med studien var å se hvilken modell som ga de beste prediksjonene på en rekke makroøkonomiske variabler. Begge modellene fikk tildelt det samme datasettet, der denne informasjonen ble optimert gitt modellenes rammeverk. Studien brukte MSE som evalueringskriterium, og konkluderte med at MRNN i snitt slo VAR med 8% på samtlige prediksjoner.

Problemet med flere av studiene over er at de sammenligner økonomiske modeller som ikke er egnet til komplekse tidsserieanalyser på lik linje med modeller som ANN og DLSTM. Resultatene fra studiene er av den grunn ikke overraskende, da modellene ikke gir mening å sammenligne.

Med denne studien bidrar vi med følgende til litteraturen:

Sammenligningen baserer seg på to modeller, som har like forutsetninger til å gi den beste fremtidsprediksjonen på oljeprisen. Både VECM og LSTM er meget godt egnet til å analysere komplekse tidsserier.

Ulikt fra tidligere litteratur som har sammenlignet maskinlærings- og økonomiske modeller, vil modellene i denne studien bruke de samme input-variablene. Vi vil så forsøke å gjøre begge modellene så gode som mulig, gitt det rammeverket de har.

Målet med denne studien er å kunne gi en mer realistisk sammenligning av slike modeller, og fylle et hull vi mener mangler i litteraturen.

3 Metode Vector error correction modell

Formålet med denne studien er å finne ut hvilken modell som predikerer den fremtidige oljeprisen best. Dette kapitlet tar for seg rammeverkene og oppbyggingen av VECM, mens kapittel 4 tar for seg rammeverket og oppbyggingen av LSTM.

3.1 Stasjonaritet

En av forutsetningene for fremtidsprediksjoner med VAR og VECM er at tidsseriene må være stasjonære. Noen av problemene som kan forekomme ved ikke-stasjonære tidsserier er feilaktige regresjoner og t-verdier som ikke følger t-fordelingen, Brooks (2019). Variabler som er stasjonære, er mer egnet til analyser da disse ikke inneholder sesongeffekter eller trender, noe som gjør dem mer forutsigbare å predikere.

Alexander (2008) definerer en variabel y_t som stasjonær hvis følgende kriterier er oppfylt

- $E(y_t)$ er endelig og uavhengig av t
- $Var(y_t)$ er endelig og uavhengig av t
- $Cov(y_t, y_s)$ er en endelig funksjon av $|t-s|$, men ikke av t eller s alene

Augmentet Dickey-Fuller (ADF) testen er en av de mest kjente og brukte testene for å sjekke om en tidsserie er stasjonær eller ikke. Dette er en type test som går under betegnelsen unit root test (test for enhetsrøtter). ADF er en modifisert versjon av den opprinnelige Dickey og Fuller (1979) testen, som bedre klarer å behandle modeller som inneholder store mengder data. ADF testen er en statistisk signifikanstest, hvor nullhypotesen er at tidsserien ikke er stasjonær og den alternative hypotesen er at tidsserien er stasjonær. Testen produserer test-statistisk og p-verdier til de inkluderte variablene, disse brukes så til å avgjøre om nullhypotesen kan forkastet eller ikke.

Fra den opprinnelige DF testen har vi:

$$y_t = \Phi y_{t-1} + u_t$$

Der nullhypotesen er at $\Phi = 1$, mot den alternative hypotesen er at $|\Phi| < 1$.

Denne kan omskrives til

$$\Delta y_t = y_t - y_{t-1} = \psi y_{t-1} + u_t$$

Null hypotesen er nå $\psi = 0$ (der $\Phi - 1 = \psi$), mens den alternative hypotesen er gitt ved $\psi < 0$.

Fra dette kan ADF testen defineres som:

$$\Delta y_t = \psi y_{t-1} + \sum_{t=1}^p \alpha_p \Delta y_{t-p} + u_t \quad (1)$$

For ADF testen gjelder samme kritiske verdier og hypotese som for DF testen, Brooks (2019).

3.2 Kointegrasjon

Modellen blir testet for kointegrasjon for å avgjøre om det skal benyttes en VAR eller VEC modell til å gjennomføre analysen. Engle og Granger (1987) oppdaget at tidsserier som var ikke-stasjonære, men som var integrert av samme orden $I(d)$, fortsatt kunne ha lineær kombinasjoner som var stasjonære. Ved påvist kointegrasjon vil en VECM bli benyttet, og det er da mulig å si noe om både de kortsiktige og langsiktige forholdet mellom variablene som inngår i modellen.

Metoden som vil bli benyttet i denne studien for å sjekke modellen for kointegrasjon er kointegrasjonstesten av Søren Johansen (1988) kointegrasjonstest. Dette er en forbedring av den bivariate modellen til Engle og Granger (1987), som klarer å tar for seg kointegrasjon mellom flere tidsserier.

Johansen kointegrasjonstest benytter både λ_{trace} og λ_{maks} , der trace-testen er en samlet test for alle egenverdiene (λ), mens maks-testen utfører separate tester på de enkelte egenverdiene. Vi vil her kun ta for oss trace-testen. Basert på en studier gjort av Cheung og Lai (1993); Soren Johansen og Juselius (1990) er denne mer robust sammenlignet med maks-testen. Grunnen til dette er at trace-testen tar i betraktning alle egenverdiene, til og

med de minste.

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^m \ln(1 - \hat{\lambda}_i) \quad (2)$$

Her er T antall observasjoner, r antall kointegrerte vektorer under nullhypotesen og $\hat{\lambda}_i$ er de estimerte egenverdiene. Nullhypotesen er at det ikke er mer enn r kointegrerte vektorer, mens alternativhypotesen er at det er flere enn r kointegrerte vektorer, Brooks (2019).

Først tester vi nullhypotesen $r = 0$, mot $r \geq 1$ for å avklare om det eksisterer minst ett kointegrert forhold. Hvis vi ikke kan forkaste nullhypotesen, konkluderer vi med at det ikke eksisterer noen kointegrerte forhold, og vi kan bruke en differensiert VAR-modell.

Hvis nullhypotesen $r = 0$ forkastes finnes det minst ett kointegrerte forhold i modellen. Vi må så teste nullhypotesen $r = 1$ mot $r \geq 2$. Hvis denne hypotesen også forkastes fortsetter vi med $r = 2$ mot $r \geq 3$, og så videre. Antall kointegrerte forhold blir bestemt når vi ikke lengre kan forkaste nullhypotesen. Så lenge $r < m$, der m er antall inkluderte variabler, estimerer vi en VECM med r kointegrerte forhold.

Ved $r = m$ har vi full rang, og vi estimerer en VAR modell på nivåform.

3.3 Valg av antall laggede verdier

Når vi skal velge antall laggede verdier (p) som optimerer modellen, kan dette gjøres med hjelp av informasjonskriterier (IC). De mest brukte informasjonskriteriene er Akaike informasjons kriterium (AIC), Schwarz Bayes informasjons kriterium (SBIC) og Hannan-Quinn informasjons kriterium (HQIC).

$$\begin{aligned} MAIC &= \ln|\hat{\Sigma}| + 2N/T \\ MSBIC &= \ln|\hat{\Sigma}| + \frac{N}{T} \ln(T) \\ MHQIC &= \ln|\hat{\Sigma}| + \frac{2N}{T} \ln(\ln(T)) \end{aligned}$$

Her er $\hat{\Sigma}$ varians-kovarians matrisen til residualene, T er antall observasjoner, og N er antall parametere i modellen. N er lik $n^2p + n$, hvor n er antall ligninger i systemet og p

er antall lags, Brooks (2019). Alle informasjonskriteriene består av to komponenter, der det første leddet er logaritmen av varians-kovariansmatrisen og det andre leddet er hvor hardt de forskjellige IC straffes for å inkludere parametere.

Vi vil i denne oppgaven bruke AIC som informasjonskriterium til valg av optimalt antall p , da denne egner seg best når formålet er prediksjoner, Shmueli mfl. (2010).

3.4 VAR

Vector autoregression (VAR) modellen har siden Sims (1980) introduserte denne på 80-tallet vært en av de mest brukte verktøyene til å analysere multivariat tidsseriedata. I motsetning til enkle lineære regresjonsmodeller klarer VAR-modellen å fange opp forhold mellom flere variabler som varierer over tid. Denne egenskapen har gjort modellen svært populær til analyser av finansielle tidsserier og fremtidsprognoser.

Modellen har en enkel oppbygging, der nåverdien av de inkluderte variablene blir forklart av de tidligere verdiene (lags) av de samme variablene.

Det enkleste systemet er en VAR av orden 1, også kjent som en VAR(1), Pesaran (2015). Matematisk kan dette fremstilles som:

$$\mathbf{y}_t = \Phi \mathbf{y}_{t-1} + \mathbf{u}_t, \quad t = 1, 2, \dots \quad (3)$$

Der \mathbf{y}_t er en $m \times 1$ vektor av valgte variabler, Φ er en $m \times m$ koeffisientmatrise og \mathbf{u}_t er en $m \times 1$ vektor med støy som tilfredsstiller følgende tre antakelser:

Antakelse 1: $E(\mathbf{u}_t) = \mathbf{0}$, for alle t

Antakelse 2: $E(\mathbf{u}_t \mathbf{u}_s) = \Sigma$, der Σ er en positiv bestment matrise

Antakelse 3: $E(\mathbf{u}_t \mathbf{u}_s) = \mathbf{0}$, for alle $t \neq s$

Enkelt forklart er \mathbf{u}_t en vektor med serielt ukorrelerte tilfeldige variabler, med null i gjennomsnitt, bedre kjent som hvit støy. S og t er her to forskjellige tidspunkt.

For en VAR(p) modell med m variabler og p lags kan ligningen skrives som

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \dots + \Phi_p \mathbf{y}_{t-p} + \mathbf{u}_t \quad (4)$$

Her gjelder samme symbolforklaring som for en VAR(1), ligning (3).

For enklere å gjøre analyser av en VAR(p) modell kan man omskrive ligning (4) til en VAR(1). Dette gjør man ved å uttrykke VAR(p)-modellen på kompanjong-form.

Vi får da følgende modell:

$$\mathbf{Y}_t = \Phi \mathbf{Y}_{t-1} + \mathbf{U}_t \quad (5)$$

Når er \mathbf{Y}_t en $mp \times 1$ vektor med valgte variabler, Φ er en $mp \times mp$ kompanjong koeffisientmatrise, og \mathbf{U}_t er en $mp \times 1$ vektor med støy.

Ved å skrive ligning (5) på matriseform får vi følgende, Pesaran (2015):

$$\mathbf{Y}_t = \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+2} \\ y_{t-p+1} \end{pmatrix}$$

$$\Phi \mathbf{Y}_{t-1} = \begin{pmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ I_m & 0 & \dots & 0 & 0 \\ 0 & I_m & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_m & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p+1} \\ y_{t-p} \end{pmatrix}$$

$$\mathbf{U}_t = \begin{pmatrix} u_t \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

Kompanjongmatrisen Φ vil blant annet bli brukt for å sjekke modellen for stabilitet i kapittel 6.

3.5 VECM

Vector error correction (VEC) modellen er en utvidet versjon av VAR-modellen som tar hensyn til kointegrerte forhold mellom variablene, Engle og Granger (1987). Forskjellen mellom VECM og VAR, er at VECM inkluderer ett feil-korrigeringsledd. Denne egenskapen gjør modellen i stand til å fange opp variablenes avvik fra det langsiktige likevektsforholdet, Lütkepohl (2005). VECM er også kjent som en kointegrert VAR.

Antar følgende VAR(p) modell:

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \dots + \Phi_p \mathbf{y}_{t-p} + \mathbf{u}_t = \sum_{i=1}^p \Phi_i \mathbf{y}_{t-i} + \mathbf{u}_t \quad (6)$$

VAR modellen fra ligning (6) kan så bli omskrevet til VECM-form:

$$\Delta \mathbf{y}_t = \Pi \mathbf{y}_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta \mathbf{y}_{t-j} + \mathbf{u}_t \quad (7)$$

Der \mathbf{y}_t er en K -vektor av $I(1)$ variabler, $\Pi = \sum_{j=1}^p \Phi_j - I_m$, og $\Gamma_j = -\sum_{i=j+1}^p \Phi_i$, for $j = 1, 2, \dots, p-1$, Pesaran (2015).

Hvis variabel Φ_j er integrert av første orden $I(1)$, har matrise Π en rang r gitt $0 \leq r < m$. Når en VECM har en slik rang betyr det at det eksisterer minst én kointegrert vektor. Videre kan vi skrive matrise $\Pi = \alpha \beta'$, der α er en $m \times r$ matrise og β' er $r \times m$ matrise. α eller error correction term (ECT) kan betraktes som ”justerings koeffisienten” til modellen, denne forteller hvor fort variablene korrigerer seg tilbake når de kommer ut av likevekt.

β inneholder parameterne til den kointegrerte ligningen.

For eksempel anta at $m = 5$, slik at systemet inneholder 5 variabler. Elementene i Π kan da skives som

$$\mathbf{\Pi} = \begin{pmatrix} \Pi_{11} & \Pi_{12} & \Pi_{13} & \Pi_{14} & \Pi_{15} \\ \Pi_{21} & \Pi_{22} & \Pi_{23} & \Pi_{24} & \Pi_{25} \\ \Pi_{31} & \Pi_{32} & \Pi_{33} & \Pi_{34} & \Pi_{35} \\ \Pi_{41} & \Pi_{42} & \Pi_{43} & \Pi_{44} & \Pi_{45} \\ \Pi_{51} & \Pi_{52} & \Pi_{53} & \Pi_{54} & \Pi_{55} \end{pmatrix}$$

Hvis $r = 1$, har vi en kointegrerte vektorer, og α og β vil være en 5×1 matrise gitt ved

$$\mathbf{\Pi} = \mathbf{\alpha}\mathbf{\beta}' = \begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \\ \alpha_{14} \\ \alpha_{15} \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} & \beta_{15} \end{pmatrix}$$

Inkluderer vi y_{t-p} i uttrykket over, og skriver denne litt om får vi

$$\mathbf{\Pi}\mathbf{y}_{t-p} = \mathbf{\alpha}\mathbf{\beta}'\mathbf{y}_{t-p} = \begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \\ \alpha_{14} \\ \alpha_{15} \end{pmatrix} \begin{pmatrix} \beta_{11}y_1 & \beta_{12}y_2 & \beta_{13}y_3 & \beta_{14}y_4 & \beta_{15}y_5 \end{pmatrix}_{t-p} \quad (8)$$

Her er $\beta'y_t$ en $1 \times r$ vektor med de kointegrerte forhold, også kjent som de langsiktige forholdene, Pesaran (2015). En VAR modell inkluderer ikke $\alpha\beta'y_t$, som gjør at vi ikke klarer å fange opp de langsiktige forholdene.

Gitt matrisen i ligning (8) kan vi skrive ut separate ligninger for hver variabel Δy_t . Det er også vanlig å normalisere disse ligningene til bestemte variabler. Hvis vi for eksempel

normaliserer med hensyn på y_1 , vil det kointegrerte uttrykket for Δy_1 bli

$$\alpha_{11} \left(y_1 + \frac{\beta_{12}}{\beta_{11}} y_2 + \frac{\beta_{13}}{\beta_{11}} y_3 + \frac{\beta_{14}}{\beta_{11}} y_4 + \frac{\beta_{15}}{\beta_{11}} y_5 \right)_{t-p}$$

Denne formen for normalisering vil bli benyttet senere når vi skal se på langsiktige forhold mellom variablene i modellen.

3.6 Diagnosetester

Diagnosetester blir brukt for å sjekke påliteligheten til modellen. Diagnosetestene som vil bli gjennomført på modellen er autokorrelasjon, stabilitet og normalfordeling.

Autokorrelasjon

Autokorrelasjon er når en tidsserie er lineært relatert til tidligere verdier av seg selv. Hvis dette er tilfellet kan analysen av tidsserien lede til misvisende resultater. For å sjekke residualene for autokorrelasjon vil det bli benyttet en Lagrange-multiplikator test. Her er nullhypotesen at det ikke eksisterer autokorrelasjon i lag i , der $i = 1, 2, \dots, p$. Mot den alternative hypotesen at det eksistere autokorrelasjon. H_0 forkastes når $p < 0.05$.

Stabilitet

Som nevnt over forutsetter VAR modeller stasjonaritet, men det samme gjelder for stabilitet. En VAR-modell er stabil (kovarians-stabil) når effekten av \mathbf{u}_t forsvinner over tid. Kompanjongmatrisen fra ligning (5) er gitt ved

$$\Phi = \begin{pmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ I_m & 0 & \dots & 0 & 0 \\ 0 & I_m & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_m & 0 \end{pmatrix}$$

Ved stabilitet er egenverdiene til kompanjongmatrisen mindre enn en i absoluttverdi, $\lambda < |1|$. Dette kan uttrykkes som

$$|\Phi - \lambda I| = 0$$

I VECM brukes stabilitetsbetingelsen til å sjekke om antall kointegrerte forhold er spe-

sifisert riktig. Den kan også brukes til å sjekke om antatte stasjonære variabler er ikke-stasjonære. For å teste VECM for stabilitet skrives modellen på VAR-form, der man sjekker egenverdiene til VAR modellens kompanjongmatrise. Gitt en VECM med M variabler og r kointegrerte forhold, vil denne ha $M - r$ egenverdier i kompanjongmatrisen. De resterende egenverdiene må være mindre enn 1 for at modellen skal være stabil å bevege seg mot sin langsiktige likevekt.

Normalfordeling

For å sjekke om tidsserien har normalfordelte residualer vil Jargue-Bera testen bli benyttet. Nullhypotesen er en samlet hypotese om at det ikke eksisterer skjevhet eller kurtose, som betyr at serien er normalfordelt. Den alternative hypotesen er at det eksisterer skjevhet og/eller kurtose. H_0 forkastes når $p < 0.05$. For at tidsserien skal være normaltfordelt må den forventede skjevheten og kurtosen være null, Lütkepohl (2005). For kurtosen tilsvarer dette en verdi på 3.

3.7 Fremtidsprognoser

En av de mest vanlige formålene med VAR og VECM er å estimere fremtidsprognoser. Vi har to forskjellige metoder som kan benyttes til dette formålet, dynamiske- og statiske prognoser.

Dynamiske prognoser vil si at den predikerte verdien y^* på tidspunkt $t + h$, i vårt tilfelle spotprisen på olje, avhenger av den predikerte verdien på tidspunkt $t + h - 1$. Altså den predikerte verdien to dager frem i tid, $t + 2$, avhenger av den predikerte verdien for i morgen, $t + 1$.

Ved statiske prognoser avhenger den predikerte verdien y^* på tidspunkt $t + h$, av den faktiske verdien y på tidspunkt $t + h - 1$. Dette betyr at vi er avhengig av de faktiske verdiene på oljeprisen for å lage fremtidsprognosene.

I denne studien sammenlignes to fremtidsprognoser. Den første prognosen for er en rullerende statistisk prognose som predikerer en måned frem i tid, $t + 1$. Den andre, er en rullerende dynamisk prognose som predikerer tre måneder frem i tid, $t + 3$. Prognosen for $t + 3$ for LSTM er ikke lik som for VECM, og vil bli gjennomgått i kapittel 4.

En VAR(p) modell, se ligning (4), kan estimeres h -steg frem $y_{t+h|T}^*$ på følgende måte, Pesaran (2015):

$$\begin{aligned} y_{t+h|T}^* &= \Phi_1 y_{t+h-1|T}^* + \dots + \Phi_p y_{t+h-p|T}^* \\ &= \sum_{p=1}^p \Phi_p y_{t+h-p|T}^* \end{aligned} \quad (9)$$

Fremtidsprognoser basert på VECM tar utgangspunkt i VAR-modellens fremtidsprognose.

3.8 Evalueringskriterier

For å evaluere og avgjøre hvilken modell som gir de beste fremtidsprediksjonene vil det i denne studien bli benyttet to forskjellige evalueringskriterier; RMSE (root mean square error), og MAPE (mean absolute percentage error). Der både RMSE og MAPE gir et mål på avstanden mellom predikerte verdier og de faktiske verdiene, og vil benyttes i sammenligningen av både VECM og LSTM.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{T - (T_1 - 1)} \sum_{t=T_1}^T |y_{t+h} - y_{t,h}^*|^2} \quad (10)$$

$$MAPE = \frac{100}{T - (T_1 - 1)} \sum_{t=T_1}^T \left| \frac{y_{t+h} - y_{t,h}^*}{y_{t+h}} \right| \quad (11)$$

I ligning (10) og (11) er fremtidsprediksjonen h -steg fremover gitt ved $y_{t,h}^*$, og den virkelige verdien på tid t er gitt ved y_t . T er de totale observasjonene (in-sample + out-of-sample), og T_1 er den første fremtidsprediksjonen, Brooks (2019).

RMSE er kvadratroten av det gjennomsnittlige kvadrerte avviket. Det vil si at modellene som evalueres, straffes eksponentielt med størrelsen til de enkelte avvikene. Som det kommer frem av ligning (10), så kalkuleres gjennomsnittet av de kvadrerte avvikende, før kvadratroten av gjennomsnittsverdien oppnås. At det er de kvadrerte verdiene som anvendes i gjennomsnittsberegningen, er årsaken til at RMSE er et eksponentielt vektet

evalueringsmål. Dette betyr at ett enkelt avvik som er veldig stort, er nok til å gi en dårlig RMSE-verdi for hele evalueringsperioden. Av denne grunn, så inkluderes det også et likevektig evalueringsmål, kalt MAPE. Som det framgår av ligning (11), så er MAPE definert som den gjennomsnittlige absoluttverdien til det prosentvise avviket. Det vil si at modellene som evalueres, blir straffet proporsjonalt i forhold til størrelsen til de enkelte avvikene. Med RMSE og MAPE, så er det ønskelig med lavest mulig verdi. Desto lavere verdien er på RMSE og MAPE, desto bedre er fremtidsprognosene. Med andre ord, så velges det i denne studien å anvende ett likevektig, og ett eksponentielt vektet mål for å vurdere fremtidsprognosene.

Den numeriske verdien til RMSE og MAPE vil i denne studien tolkes på følgende vis; en numerisk verdi på 1, gitt RMSE og MAPE, betyr dette at modellenes fremtidsprediksjon bommer med henholdsvis 1 USD/fat olje og 1%.

4 Metode Long-short term memory modell

Ett Recurrent Neural Network (RNN), kjent som Long short-term memory (LSTM) er den andre modellen som vil bli benyttet i denne studien. Teorien bak RNN er basert på det tidligere arbeidet av Rumelhart, Hinton og Williams (1986), men ble ikke et populært analyseverktøy før på 2000-tallet. Grunnen til dette var at modellen ikke hadde et fungerende rammeverk før LSTM ble oppfunnet av Hochreiter og Schmidhuber (1997). Modellen egner seg spesielt godt til analyser av data, der sekvensen av de tidligere hendelsene har innvirkning på fremtidige hendelser. Dette er årsaken til at LSTM-modellen er aktuell ved prediksjon av tidsserier.

Utvelgelsen av LSTM-strukturen skjer ved hjelp av en genetisk algoritme (GA). Dette er en optimeringsalgoritme som søker etter den beste LSTM-modellen med en prøv- og feil-metodikk (heuristikk).

Før vi går inn på teorien til LSTM og optimering prosessen, vil et enkelt kunstig nevralt nettverk (ANN) forklares. Deretter vil strukturovergangen fra et ANN til en LSTM bli forklart og illustrert.

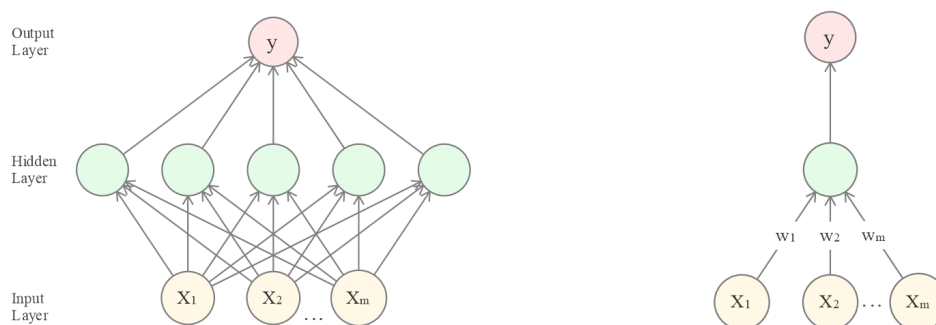
4.1 Nevrale nettverk

Et nevralt nettverk (NN) er et sett med algoritmer som forsøker å etterligne en biologisk hjerne og er designet til å gjenkjenne mønster. Slik som hjernen oppfatter og tolker impulser fra sansene våre, tolker det nevralt nettverket sansedata gjennom maskinoppfattelse, Nielsen (2015). For at et NN skal kunne tolke data må all virkelig informasjon oversettes til vektorer som inneholder numeriske mønstre i binær form.

Et ANN består av et stort antall sammensatte behandlingselementer, kalt nevroner, som jobber sammen for å løse ett eller flere problemer. Disse nevronene opererer parallelt og er ordnet i nivåer, Nielsen (2015). I det første nivået blir den ubehandlede informasjonen mottatt og behandlet, før utgående informasjon blir sendt videre til neste nivå (senere kalt *layer*). Denne prosessen gjentas helt til det siste nivået, som produserer det endelige resultatet. En slik ANN blir kalt et feed forward network, Goodfellow, Bengio og Courville (2016).

Nevronet, også kalt *node*, er den grunnleggende byggeklossen i et nevralt nettverk. En ANN visualiseres ofte som i figur 1a, der nettverket består av tre typer noder som er sortert i nivåer basert på informasjonsflyten (nedenfra og opp). Sirklene representerer nodene i nettverket, og fargekodene gul, grønn og rød, representerer henholdsvis nodene i Input Layer (IL), Hidden Layer (HL), Output Layer (OL). Strekene mellom nodene representerer synapseforbindelsene i hjernen.

Figur 1: Oppbygning til et kunstig nevralt nettverk

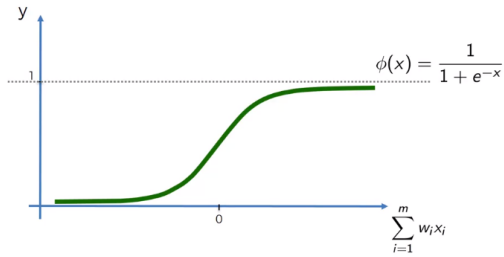


(a) Enkel fremvisning av en ANN med ett hidden layer nivå.

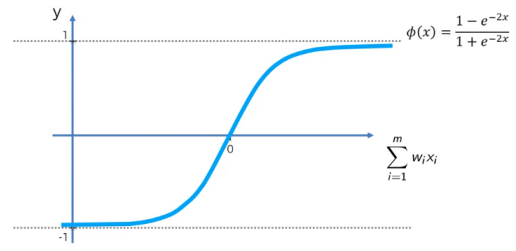
(b) Illustrasjon av vektningen w , mellom nodene i et nevralt nettverk.

En biologisk hjerne vil forsøke å lære ved å styrke/svekke synapseforbindelsene mellom nevronene i hjernen. For å etterligne biologiens synapseforbindelser, så vektet signalene som beveger seg mellom nodene i nettverket. Dette er illustrert i figur 1b, der w er vektene som justeres, når modellen trener og tilpasser seg treningsdata (in-sample). Styrken på signalene påvirker hvilken grad nodene aktiveres, som igjen påvirker hvordan nodene i de neste nivåene blir aktivert, Goodfellow, Bengio og Courville (2016). Metoden de inngående signalene blir behandlet på i HL-nodene, defineres ut fra hvilken aktiveringsfunksjon, ϕ som anvendes. De vanlige aktiveringsfunksjoner er illustrert i figur 2.

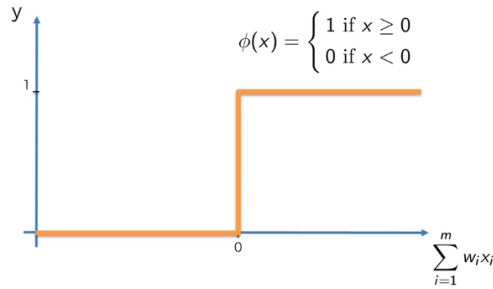
Figur 2: Illustrasjon av de forskjellige aktiveringsfunksjonene



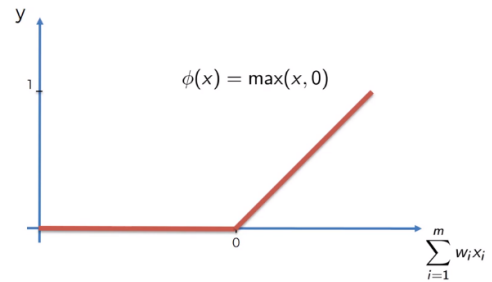
(a) Sigmoid-aktiveringsfunksjon.



(b) Hyperbolic tangent-aktiveringsfunksjon.



(c) Threshold-aktiveringsfunksjon.



(d) Rectifier-aktiveringsfunksjon.

4.1.1 Tapsfunksjon

Tapsfunksjonen i output-noden defineres ut fra problemet nettverket skal forsøke å løse. Disse funksjonene er ofte tilsvarende som i parametriske modeller, Goodfellow, Bengio og Courville (2016). I en punktprediksjon, slik som denne studien forsøker med spotprisen på olje, så er det ønskelig at prediksjonsavviket minimeres. Tapsfunksjonen som anvendes i LSTM-modellen er derfor RMSE, definert av ligning (10).

4.2 LSTM modellen

En RNN-modell har en nettverksstruktur som gir et såkalt ”internt” minne. I et enkelt nevralt nettverk, blir informasjonen kun behandlet i den ene retningen, fra input til output. Mens i en RNN er informasjonen tilbakevendende, derav navnet *Recurrent Neural Network*. En versjon av RNN er en LSTM-RNN. Dette er en modifisert versjon av den opprinnelige RNN-modellen, som lettere klarer å identifisere tidligere hendelser og ta disse i bruk. Hovedmekanismen til LSTM er at den løser ”the vanishing gradient problem” til RNN-modellen. Teorien og matematikken som omhandler dette vil ikke bli gjennomgått i denne studien, men det henvises til følgende artikler; Hochreiter og Schmidhuber (1997); Hochreiter (1998).

En av hovedforskjellene mellom en generell ANN og en LSTM, er strukturen og dens dimensjon. Dette er siden de laggede verdiene til variablene nå behandles i modellen. I en ANN er input-strukturen én vektor for hvert tidssteg. Det vil si at en ANN-modell tar inn én vektor med variabelobservasjoner som input-signaler til det nevralt nettverket. Input-strukturen til LSTM-modellen er derimot en matrise. Den består da av en vektor for hver eneste laggede periode som er inkludert i modellen. Forskjellen i inputstrukturen fra en generell ANN til LSTM vises av ligning (12) og (13).

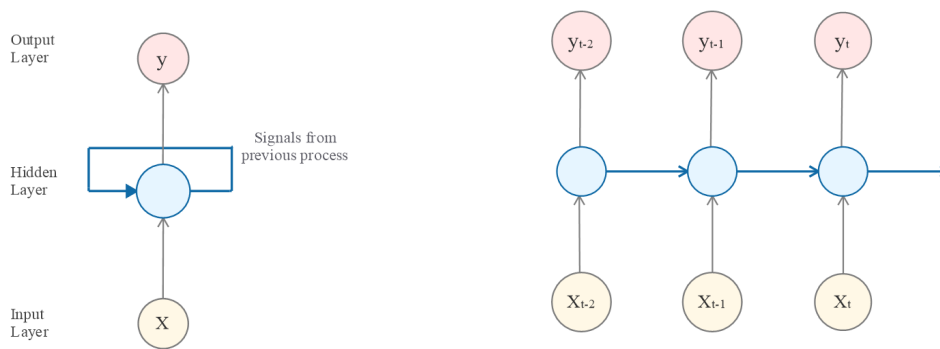
$$ANN \text{ Input struktur} = \begin{pmatrix} x_{1,t} & x_{2,t} & x_{3,t} & \dots & x_{n,t} \end{pmatrix} \quad (12)$$

$$LSTM \text{ Input struktur} = \begin{pmatrix} x_{1,t} & x_{2,t} & x_{3,t} & \dots & x_{m,t} \\ x_{1,t-1} & x_{2,t-1} & x_{3,t-1} & \dots & x_{n,t-1} \\ x_{1,t-2} & x_{2,t-2} & x_{3,t-2} & \dots & x_{n,t-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{1,t-nlags} & x_{2,t-nlags} & x_{3,t-nlags} & \dots & x_{m,t-nlags} \end{pmatrix} \quad (13)$$

Proessen til LSTM er presentert i figur 3. Slik som før er den blå sirkelen nodene i HL-nivået, den gule sirkelen nodene i IL og røde sirkelen nodene i OL. En LSTM består av flere slike strukturer, der hver struktur korresponderer til de ulike laggede periodene som er inkludert i modellen. Signalflyten går fortsatt fra $IL \rightarrow HL \rightarrow OL$, men i en LSTM, så blir signalene i den ene prosessen til input i den neste. Det vil si at strukturen som tar inn variabelinformasjonen fra X_t , mottar også signalene fra HL-nivået i forrige prosess, som tar inn verdier fra X_{t-1} , Hua mfl. (2019). Dette vises av den blå pilen i figur 3a. Sett fra et annen perspektiv kan dette illustreres som i figur 3b.

På denne måten blir de sekvensielle sammenhengene mellom variablene identifisert, og modellen anvender dermed sekvensen av tidligere hendelser for å estimere en fremtidig prognose på $t + h$ steg frem i tid.

Figur 3: Oppbygning til et LSTM-nettverk.



(a) Illustrerer den tilbakevendende prosessen til et LSTM-nettverk.

(b) En utfoldet illustrasjon av figur (a), som viser de ulike prosessene som sender informasjonen fra HL videre til neste prosess.

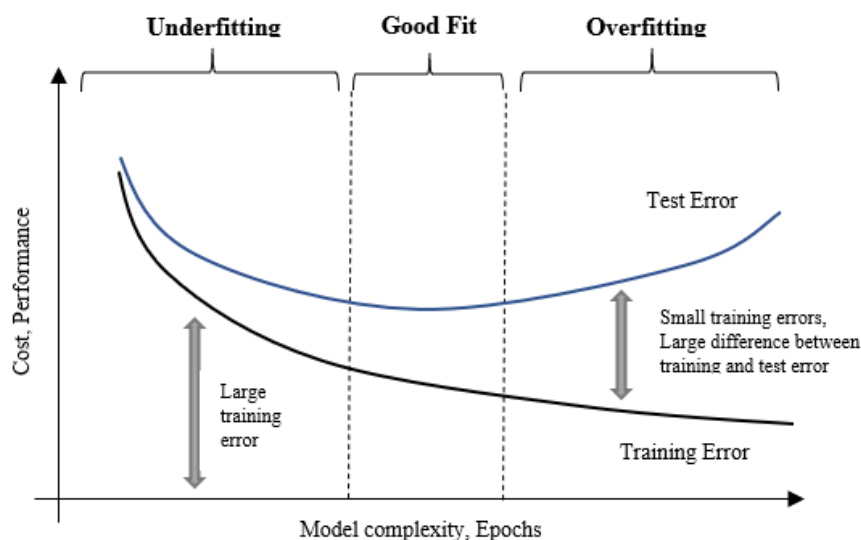
4.3 Trening av nettverk

Det nevralt nettverket trener ved hjelp av en metode kalt backpropagation, Rumelhart, Hinton og Williams (1986). Når signalene beveger seg gjennom nettverket med en viss kombinasjon på vektene, så evalueres denne vekt-kombinasjonen med tapsfunksjonen av nodene i OL, som nevnt i delkapittel 4.1.1.

Informasjonen fra tapsfunksjonen sendes tilbake i nettverket der vektene w_i , mellom nodene i nettverket justeres, Goodfellow, Bengio og Courville (2016). Hvor ofte vektene oppdateres under trening bestemmes av det som kalles *batch size*. Størrelsen på denne angir hvor mange observasjoner modellen evaluerer med tapsfunksjonen før vektene oppdateres. Slik fortsetter modellen å trene ved å iterere igjennom modellens treningsdata. Antall iterasjoner av treningsdataen kalles for *epoch*.

Et veldig kjent problem ved trening av nevralt nettverk, er det som kalles for *overfitting* og *underfitting*. Enkelt forklart, så betyr det at modellen enten er for bra, eller for dårlig tilpasset in-sample-dataen. Resultatet av begge tilfellene avhenger av graden, men kan resultere i en modell som gir sub-optimale eller dårlige prediksjoner, Goodfellow, Bengio og Courville (2016). Denne problematikken er illustrert i figur 4.

Figur 4: Overfitting og underfitting



Figuren viser hvordan overfitting og underfitting oppstår i et nevralt nettverk. Modellens kompleksitet og hvor lenge den trenes gir utslag på hvordan modellen presterer helhetlig, men også på forholdet mellom in-sample og out-of-sample.

Underfitting oppstår gjerne når nettverket ikke er stort nok, eller får riktig informasjon i forhold til problemet modellen forsøker å løse. Dette gjør at modellen har utilstrekkelig kapasitet i forhold til problemets kompleksitet. Med andre ord, så er problemet for komplisert i forhold til informasjonen som blir gitt til modellen.

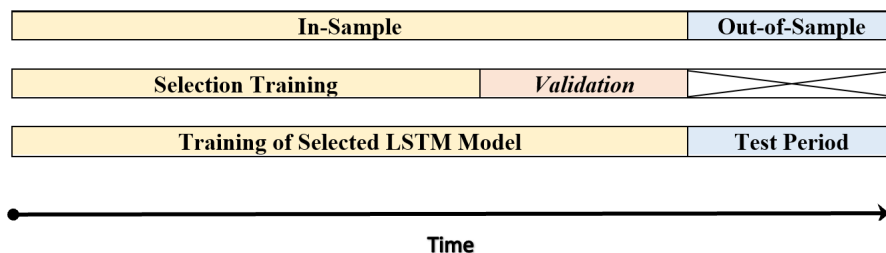
For å unngå overfitting, så benyttes en metode kalt *dropout*. Metoden fungerer slik at en andel av signalene mellom de ulike HL-nivåene fjernes. Signalene som fjernes er tilfeldig valgt for hvert steg i treningen. Dette medfører at modellen ikke mottar de eksakt samme signalene, for hver epoch når modellen trenes. Ytterligere detaljer om denne metoden er gitt av Gulli og Pal (2017).

4.4 Valideringsdata

For å kvalitetssikre modellen implementeres en valideringsperiode for trening, evaluering og seleksjon. Det betyr at det konstrueres en "out-of-sample"-periode fra in-sample-perioden. Potensielle LSTM-modeller blir vurdert ut ifra en periode som ikke inngår i treningsperioden, slik at vektene i større grad blir tilpasset de generelle sammenhengene mellom variablene for en vilkårlig periode, Bishop (2006). Denne inndelingen vises i figur

5. Dette blir dermed et bidrag til å redusere problematikken med overfitting.

Figur 5: Inndeling av data-periodene.



Figuren viser inndelingen av datasett for seleksjon og vurdering av de ulike LSTM-modellene i seleksjonsprosessen.

Den første raden i figuren representerer in-sample og out-of-sample som defineres i kapittel 6. Rad to illustrerer de datasettene som anvendes i seleksjonsprosessen av LSTM-modellen. Den nederste raden viser trenings- og testperioden for den valgte LSTM-modellen.

4.5 Fremtidsprognoser

I denne studien vil en fremtidsprognoser for én og tre måneder blir sammenlignet med hverandre. Det vil i forbindelse med denne dette bli trent to LSTM nettverk, der begge fremtidsprediksjonene vil være en statisk rullerende prognose. Dette er forskjellig fra VECM sin 3-måneders prognose, som har en dynamisk metodikk. Statisk og dynamisk prognoser ble utdypet og forklart i delkapittel 3.7.

4.6 Modellseleksjon

En genetisk algoritme (GA), går under kategorien evolusjonære algoritmer som er inspirert av Charles Darwins' evolusjonsteori, Bäck, Fogel og Michalewicz (2018). GA er en heuristisk optimeringsmetode som effektivt kan oppnå en tilnærmet optimal løsning, der det er veldig mange mulige kombinasjoner av parameter-verdier i utvalget. Desto bredere og dypere et nevralt nettverket er, desto lenger tid tar det å trene modellen. Hvis hastigheten på prosessering av data er begrenset, kan det ta lang tid å justere alle vektene i modellen. Dette er ofte tilfellet ved trening av nevrale nettverk på en vanlig privat datamaskin, som er hovedårsaken til at en genetisk algoritme anvendes i seleksjonen av LSTM-modellens parameterverdier.

Disse algoritmene forsøker å gjenspeile den utviklingen vi observerer i naturen, der organismer endrer seg ved hjelp av genmutasjon over flere generasjoner. Mutasjoner av organismens genetikkk vil da føre til endringer i organismens egenskaper og attributter. Disse endringene i genetikkk kan både være en fordel og en ulempe, gitt problemet organismen står overfor. Under evolusjonsteoriens naturlige seleksjon, så er organismens optimeringsproblem hovedsakelig overlevelse og/eller videreføring av gener, Brøgger (2020). Mens innen programmering av GA, kan seleksjons-kriteriet for videreføring av gener defineres. Individets egenskaper kan videre bli definert av en fitness-funksjon som indikerer dens egnethet, gitt seleksjonskriteriet som er valgt.

En GA er en metaheuristisk optimeringsmetode. Dette betyr at det er en overordnet heuristikk (Evolusjonær Algoritme), som kontrollerer andre heuristiske prosesser. Algoritmen søker etter mulige løsninger som ikke alltid framgår av tradisjonelle optimeringsmetoder, Glover og Kochenberger (2006). Disse prosessene blir som oftest tilpasset problemet ved å anvende ulike variasjoner av heuristikk.

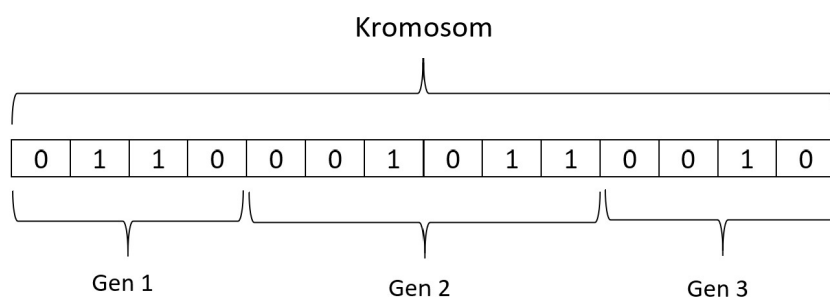
4.6.1 Komponenter

Genetiske algoritmer benytter mye av terminologien fra naturlig genetikkk for de komponentene algoritmen forsøker å representere. I dette avsnittet vil GA-komponentene som benyttes i denne studien defineres.

Individets oppbygning

I en genetisk algoritme er et *individ* betegnelsen på én spesifikk løsning til det problemet man forsøker å optimere. Som nevnt tidligere, vil den genetiske algoritmen forsøke å finne den optimale LSTM-strukturen og dens treningsparametere. I modellen korresponderer derfor ett individ til én enkel LSTM-struktur. Det vil si at hvert individ holder på parameterinformasjonen som er nødvendig, for å konstruere og trene den spesifikke LSTM-strukturen individet representerer. Denne informasjonen er definert ut ifra individets *kromosom*. I denne studien benyttes en bit-string variant som representasjon på et kromosom. Individets ulike gener er definert på ulike posisjoner langs kromosomets bit-string, der genene representerer de parametere i LSTM-modellen man ønsker å optimere. Et generelt bit-string-kromosom er illustrert i figur 6.

Figur 6: Illustrasjon av et kromosom av varianten bit-string.



Figuren viser oppbygningen til et generelt bit-string kromosom. De ulike genene representerer de ulike parametere en ønsker å optimere. Den numeriske tallverdien kan tolkes ut ifra den binære tallrekken.

Tallverdien til hvert gen kan tolkes ut ifra de binære verdiene til genet, og er definert som *allele*. Alle allele-verdiene til et individ utgjør til sammen én spesifikk tallverdi-kombinasjon av parameteren til LSTM-modellen.

Populasjon og generasjon

Begrepet populasjon blir brukt som beskrivelse på et gitt antall individer (LSTM-modeller) som er gruppert sammen. Generasjon er definert som tids-dimensjonen, og beskriver de ulike populasjonene sekvensielt gjennom optimeringsprosessen.

Evolusjonær Algoritme (EA)

Dette er den overordnede algoritmen som formaterer den evolusjonære loopen (optimeringsprosessen), og GA-operatorene. EA utfører selve optimeringen og returnerer det mest optimale individet, som algoritmen har oppnådd gjennom den evolusjonære loopen.

Varianten som anvendes i denne studien er en av de mindre komplekse, og er blant klassen av evolusjonære algoritmer som først ble analysert og foreslått av Holland mfl. (1992). Algoritmen benytter en stokastisk seleksjon av foreldrepopulasjonen, altså individene som blir brukt til å konstruere neste generasjons populasjon Bäck, Fogel og Michalewicz (2018). Dette er hensiktsmessig for å unngå problemer med lokale minimumspunkt, siden metoden søker stokastisk rundt i mulighetsområdet av potensielle individer.

GA-Operatører

Dette er algoritmens prosesser, som benytter ulike logiske operasjoner som input-parametere i operasjonene. Det tekniske aspektet ved de ulike operatorene vil ikke bli tatt for seg i denne studien, men henvises til dokumentasjonen for modulen som er anvendt, Fortin mfl. (2012). Navn på metodene som korresponderer med dokumentasjonen er gitt av tabell 20 i appendiks. Operatorene som styres av EA er som følger:

Evaluation – implementerer en fitness-funksjon som er konstruert i forhold til det optimeringsproblemet man står overfor. Fitness-funksjonen evaluerer et individ ved hjelp av individets egne allele-verdier og returnerer et numerisk mål på individets egnethet i forhold til de andre individene. Dette gjøres ved at individene (LSTM-modellene) konstrueres og trenes, i henhold til deres allele-verdier. Individene trenes og testes på valideringsdata som forklart i delkapittel 4.4. RMSE-verdien fra valideringsperioden er definert som individenes fitness-score.

Selection – er operasjonen som selekterer individene basert på deres egenskaper. For det optimeringsproblemet den genetiske algoritmen blir anvendt på, så er det kun én egenskap som er i fokus, individenes fitness score (RMSE). Seleksjonskriteriet for individene er derfor $\min(\text{fitness score})$, uavhengig av generasjonen individet er fra. Løsningen, altså det beste individet, vil da være det beste individet som er oppnådd gjennom alle generasjoner i optimeringsprosessen.

Crossover – er en operator som tar kromosomet fra ett eller flere individ (foreldrene), for deretter å konstruere nye individ (barn), ved å krysse kromosomene. Metoden

som benyttes for å krysse kromosomene vil ha direkte utslag på individene i hver generasjon. Denne operatoren har som hensikt å gjøre små endringer på individer fra forrige generasjon som har blitt evaluert. Dermed kan det testes ut små variasjoner på relativt gode løsninger i søk etter enda bedre individ.

Mutation – er en operator som gjennomføres for å skape diversifisering blant individene i populasjonen. Dette hindrer at individene i populasjonen blir for like, og at man dermed fanges” i et lokalt minimumspunkt. Derfor implementeres en mutasjonsmetode kalt *bit-flip* på kromosomet til en andel av populasjonen. Denne mutasjonsvarianten benyttes siden den skaper store endringer, som er en ønsket egenskap.

4.6.2 Optimeringsprosessen

Algoritmen som finner en tilnærmet optimal parameterkombinasjon til LSTM forklares nedenfor. Altså er dette prosessen som selekterer parametrene-verdiene til LSTM-modellen, som skal sammenlignes med den korresponderende VECM. Parametrene og mulighetsområdet er gitt av tabell 21 i appendiks.

Betegnelser

P_g = populasjonen til generasjon g

$cspb$ = *crossover-rate*, der $0 < cspb < 1$

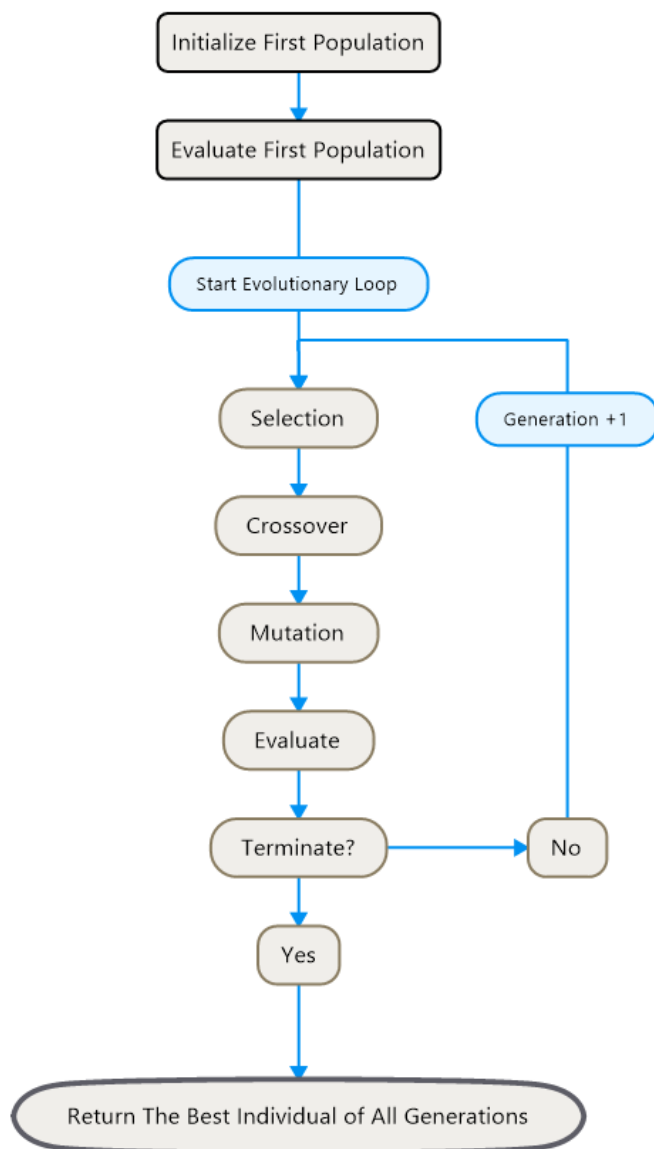
$mutpb$ = *mutation-rate*, der $0 < mutpb < 1$

$ngen$ = maks antall generasjoner

Algoritmen starter innledningsvis med å konstruere en tilfeldig populasjon P_0 , som blir evaluert. Deretter begynner den evolusjonære loopen, der den første hoved-operasjonen er seleksjons-operatoren. Denne operatoren konstruerer populasjonen, P_{g+1} , ut ifra de evaluerte individene i forrige generasjon. *Crossover* blir anvendt på en andel av P_{g+1} uttrykt av parameteren $cspb$, og returnerer en modifisert populasjon P'_{g+1} . Denne populasjonen blir videre modifisert til P''_{g+1} ved at en andel, $mutpb$, av populasjonen P'_{g+1} , muteres. Den siste sekvensen i loopen er at P''_{g+1} evalueres, og en ny runde i den evolusjonære loopen påbegynnes. Dette medfører en oppdatering av tidsdimensjonens verdi, $g = g + 1$. Slik fortsetter EA å søke etter de optimale individene helt til termineringkriteriet, $g = ngen$, tilfredsstilltes. I tillegg, så avsluttes optimeringsprosessen dersom det beste individet ikke forbedres tilstrekkelig i løpet av et gitt antall generasjoner. Ytterligere detaljer om EA-

typen som anvendes refereres til varianten *eaSimple* fra dokumentasjonen *Algorithms — DEAP 1.3.1 documentation* (2021). Flytdiagram over prosessen vises nedenfor i figur 7.

Figur 7: Flytdiagram av optimeringsprosessen til den genetiske algoritmen.



5 Data

I dette kapittelet vil vi presentere variablene som blir brukt i de to modellene. For å velge ut potensielle variabler som forklarer endringer i oljeprisen har vi tatt utgangspunkt i tidligere litteratur og studier med lignende problemstillinger.

Da formålet med denne studien er å sammenligne prediksjonsevnen til to forskjellige modeller settes det noen begrensninger for at disse kan evalueres på likt grunnlag. Siden VAR/VECM har flere betingelser som må være oppfylt sammenlignet med LSTM, har vi tatt utgangspunkt i disse når variablene skal bestemmes. For å unngå problemer med overparametrisering blir det forsøkt å velge så få variabler som mulig, som samtidig forklarer det meste av svingningene i oljeprisen.

5.1 Valg av variabler

De endelige variablene som inngår i denne studien ble valgt ved hjelp av en ”uni-directional forward selection”. Dette er en stegvis regresjonsprosedyre som baserer seg på p-verdien til variablene. Seleksjonsprosessen gjennomføres på et utvalg av variabler som antas å ha en signifikant lineær påvirkning på oljeprisen. Seleksjonsprosessen starter med å velge en endogen variabel, i dette tilfellet brent crude spot. Prosessen velger så ut ønskede eksogene variabler, der den første variabelen som blir tatt med i modellen er den variabelen med lavest p-verdi. Den neste variabelen som inkluderes er den med lavest p-verdi, gitt resterende utvalg av variabler. Seleksjonsprosessen fortsetter helt til det kun er én variabel igjen med en $p - verdi < 0.05$ i utvalget. For de spesielt interessert anbefales Tao og Brooks (2019) for den tekniske gjennomgangen av seleksjonsprosessen og teorien bak denne.

Etter gjennomført forward selection satt vi igjen med følgende fem variabler som vil benyttes i denne studien.

BCS - Brent crude spotpris

BCS er den globale spotprisen på brent nordsjøolje. Da formålet med denne studien er å predikere den fremtidige oljeprisen, vil det gi mye informasjon å inkludere tidligere verdier av denne. Når man inkluderer laggede verdier av oljeprisen kan man få en bedre

forståelse av hvordan oljeprisen beveger seg, samt hvordan den reagerer på endringer og sjokk i økonomien.

BCF - Brent crude electronic energy futures

Futures prisen på olje sier noe om aktørenes forventinger om oljeprisen. Ved handel av futures har både kjøper og selger akseptert en pris de mener gjenspeiler markedets forventninger for de gjeldene varen. Vi mener at oljefutures er en viktig variabel å inkludere da aktørenes forventinger kan si noe om både aktivitetsnivå og utviklingen i markedet. BCF sier også noe om den fremtidige etterspørselen av olje.

GOP - Global oil & gas exploration and production total return index

Den globale oljeproduksjonen er en god indikasjon på tilbud og aktivitet i markedet. Der perioder med dårlig økonomi og lav aktivitet resulterer i redusert produksjon, mens perioder med god økonomi og større aktivitet i markedet resulterer i det motsatte. Lav oljepris er forbundet med det første eksempelet, og høy oljepris er forbundet med det andre eksempelet.

VIX - Market volatility index

VIX er et mål på markedets forventninger om fremtidig volatilitet eller risiko i S&P 500, og publiseres av Chicago Board Option Exchange (CBOE). Volatilitetsindeksen er også kjent som fryktindeksen da den er et mål på finansiell uro. I denne studien vil VIX-en bli brukt som en indikator for utroligheter i markedet.

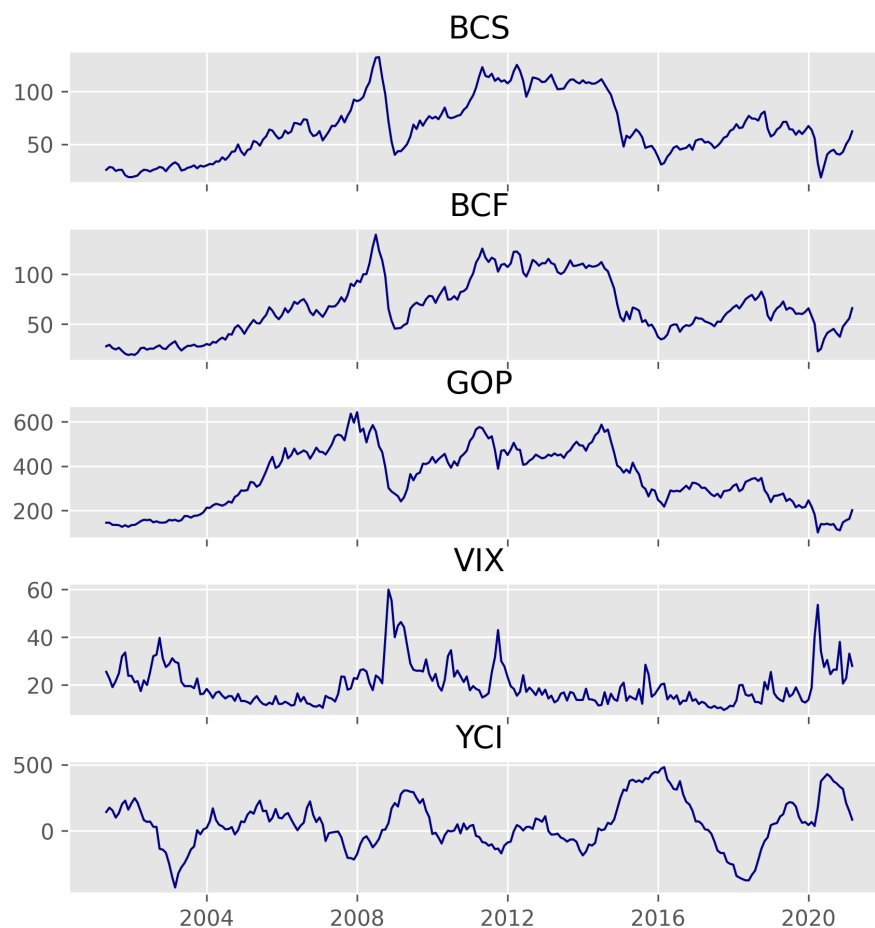
YCI - Monthly year-to-year change in oil inventory

YCI måler de månedlige endringene i oljelagrene, fra år til år. Sagt på en annen måte forteller YCI hvor store endringer det er i de globale oljelagrene per måned sammenlignet med året før. En av verktøyene OPEC (Organization of the Petroleum Exporting Countries) har for å påvirke oljeprisen bortsett fra produksjonen, er oljelagrene. Når oljeprisen er veldig høy, kan stormaktene fylle opp oljelagrene for å presse prisen ned. YCI vil av denne grunn kunne gi mye informasjon om svingningene i oljeprisen og tilbud av olje.

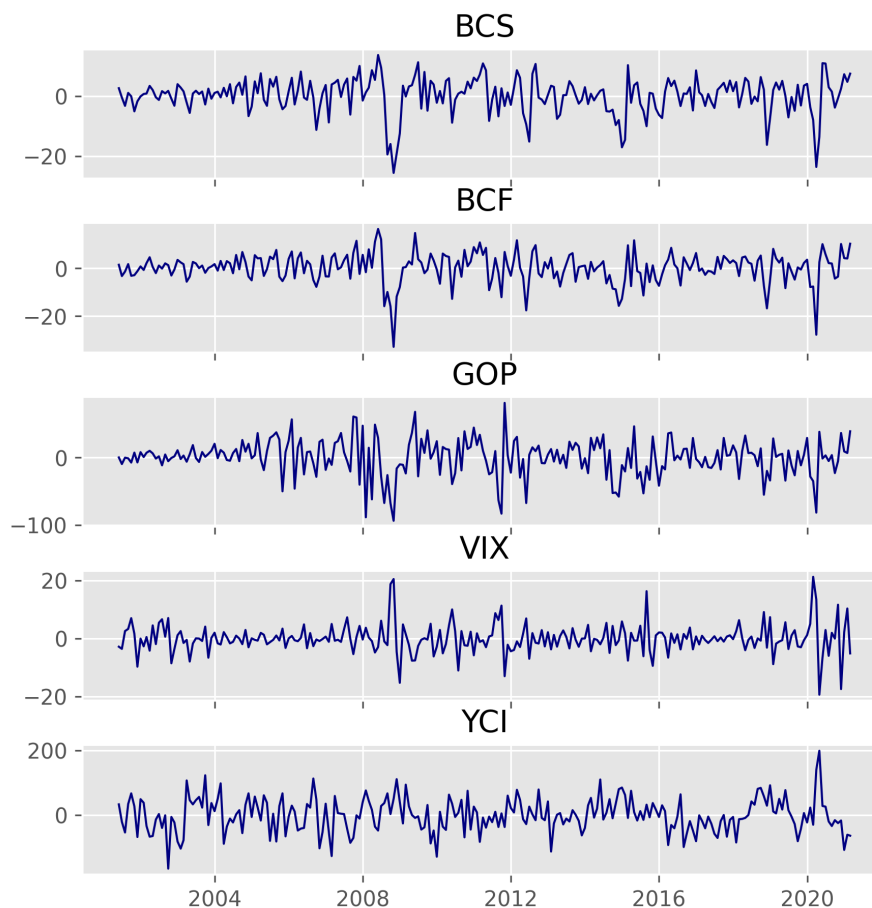
5.2 Deskriptiv statistikk

De finansielle databasene brukt i denne studien er Eikon og FRED. Variablene med tilhørende koder og hvilke databaser de er hentet fra er oppsummert i tabell 17 i appendiks. Vi har innhentet månedlig data i tidsperioden 2001(4) til 2021(3), for de fem variablene presentert over. Daten er delt inn i to deler, in-sampel og en out-of-sample. In-sample består av data fra 2001(4) til 2018(12), og vil bli brukt til å trene og analysere modellene på. Out-of-sample består av data fra 2019(1) til 2021(3), og vil bli brukt til å sammenligne fremtidsprognosene med. Figur 8 og 9 viser hvordan variablene har utviklet seg gjennom hele tidsperioden.

Figur 8: Variablenes utvikling fra 2001(4)-2021(3) i nominelle verdier



Figur 9: Variablene utvikling fra 2001(4)-2021(3) i differensierte verdier



Tidsperioden inneholder flere hendelser som hadde store konsekvenser for den globale økonomien og oljeprisen. Perioden inkluderer blant annet terrorangrepet 9/11 i 2001, finanskrisen i 2008 og oljekrisen i 2014 og Covid-19 pandemien som brøt ut i 2020.

Tabell 1 og 2 presenterer den deskriptive statistikken for variablene i henholdsvis nominelle og differensierte verdier.

Tabell 1: Deskriptiv statistikk av variablene i nominelle verdier

	BCS	BCF	GOP	VIX	YCI
Obs	240	240	240	240	240
Mean	65.65	66.46	342.70	19.84	51.02
Std	9.08	29.02	135.68	8.45	184.90
Skew	0.40	0.36	0.04	1.76	0.05
Kurt	2.16	2.18	1.90	6.96	2.90
Min	18.38	19.14	100.75	9.51	-430
Max	132.72	139.83	642.78	59.89	482

Tabell 2: Deskriptiv statistikk av variablene i differensierte verdier

	Δ BCS	Δ BCF	Δ GOP	Δ VIX	Δ YCI
Obs	239	239	239	239	239
Mean	0.17	0.15	0.27	-0.03	-0.85
Std	6.02	6.29	27.34	5.08	53.39
Skew	-1.17	-1.18	-0.59	0.47	0.05
Kurt	5.67	7.26	4.28	7.07	3.65
Min	-25.65	-32.85	-93.68	-19.39	-166
Max	13.73	16.42	80.88	21.27	199

Gitt tabellen for de differensierte verdiene ser man at YCI har det største standardavviket, etterfulgt av GOP. BCS, BCF og GOP har et positivt gjennomsnitt, mens VIX og YCI har et negativt gjennomsnitt. VIX har den minste gjennomsnittlige endringen per måned, mens YCI har den største gjennomsnittlige endringen per måned. Videre kan man merke seg at det kun er YCI som har en skjevhet og kurtose tilnærmet null (for kurtose tilsvarer dette en verdi på 3).

Tabell 3 og 4 viser parvis korrelasjon mellom variablene i henholdsvis nominelle og differensierte variabler.

Tabell 3: Korrelasjonsmatrise av variablene i nominelle verdier

	BCS	BCF	GOP	VIX	YCI
BCS	1.0000				
BCF	0.9944*	1.0000			
GOP	0.8384*	0.8548*	1.0000		
VIX	-0.1528*	-0.1702*	-0.2853*	1.0000	
YCI	-0.3469*	-0.3388*	-0.2771*	0.1029	1.0000

* markerer parvis korrelasjon som er signifikante på 1%.

Tabell 4: Korrelasjonsmatrise av variablene differensierte verdier

	Δ BCS	Δ BCF	Δ GOP	Δ VIX	Δ YCI
Δ BCS	1.0000				
Δ BCF	0.7564*	1.0000			
Δ GOP	0.4110*	0.7276*	1.0000		
Δ VIX	-0.1720*	-0.3603*	-0.5015*	1.0000	
Δ YCI	-0.3167*	-0.2928*	-0.2021*	0.0065	1.0000

* markerer parvis korrelasjon som er signifikante på 1%.

I begge tilfellene har BCS positiv korrelasjon med BCF og GOP, og negativ korrelasjon med VIX og YCI. Variablene markert med * indikerer at det er en parvis korrelasjon mellom variablene som er signifikant på 1%. Fra tabellene ser man at det kun er VIX og YCI som ikke har en signifikant parvis korrelasjon, dette gjelder for både de nominelle- og differensierte variablene.

6 Empiriske resultater og diskusjon

Følgende kapittel tar for seg de gjennomførte testene og resultatene. Kapittelet er delt inn i tre deler, i første del blir analysen av VECM gjennomgått, i del to blir LSTM gjennomgått, mens i del tre blir modellene sammenlignet og diskutert.

6.1 Vector error correction modell

I denne delen vil resultatene av ADF-testen, Johansen kointegrasjonstest, valg av laggede verdier, og diagnostestene blir gjennomgått. Etterfulgt av analysen av VECM, der vi vil diskutere modellens kortsiktige- og langsiktige forhold.

6.1.1 Stasjonaritet

For at variablene kan benyttes i VAR/VECM må disse være stasjonære. Ved hjelp av ADF-testen kan vi kontrollere om tidsserien tilfredsstiller dette kravet eller ikke. Testens nullhypotese antar at det eksisterer en unitroot, mot alternativhypotesen som antar at det ikke eksisterer en unitroot.

Tabell 5: Augmented Dickey-Fuller test nominelle verdier

	t-stat	p-value	nobs	1%	5%
BCS	-2.589	0.0953	238	-3.464	-2.881
BCF	-2.471	0.1228	238	-3.464	-2.881
GOP	-1.830	0.3657	238	-3.464	-2.881
VIX	-4.845*	0.0000	238	-3.464	-2.881
YCI	-2.953**	0.0414	238	-3.464	-2.811

ADF-test på alle variablene i nominelle verdier.

*,** viser variabler som er stasjonære på henholdsvis 1% og 5% signifikantnivå.

Gitt resultatene i tabell 5 ser vi at VIX og YCI er stasjonær, $I(0)$, på henholdsvis 1% og 5% signifikantnivå når testen bruker nominelle verdien, som betyr at de ikke inneholder

en unit-root. Resterende variabler er ikke stasjonær, og vi må finne ut hvilken orden de er integrerte av. Først blir variablene differensiert, før ADF-testen gjennomføres på nytt. Dette er for å sjekke om de er integrert av første orden, $I(1)$. Flere studier som Sims (1980); Sims, Stock og Watson (1990), argumenterer mot bruken av differensiering. Grunnen til dette er at det kan lede til tap av viktig informasjon omhandlende hvordan variablene beveger seg sammen. Da stasjonaritet er en av forutsetningene for modellene, har vi valgt å se bort fra dette, og regner med at tapet av informasjon er av mindre betydning.

ADF-testen av de differensierte verdiene konkluderte med at alle variablene er stasjonære ved 1% signifikantnivå, tabell 6.

Tabell 6: Augmented Dickey-Fuller test differensierte verdier

	t-stat	p-value	nobs	1%	5%
Δ BCS	-9.320*	0.000	237	-3.464	-2.881
Δ BCF	-9.406*	0.000	237	-3.464	-2.881
Δ GOP	-9.406*	0.000	237	-3.464	-2.881
Δ VIX	-13.738*	0.000	237	-3.464	-2.881
Δ YCI	-8.492*	0.000	237	-3.464	-2.881

Alle variablene er differensiert av første orden.

* viser variablene som er stasjonære på 1% signifikantnivå.

Gitt resultatet fra ADF-testen på de differensierte variablene kan vi konkludere med BCS, BCF og GOP er $I(1)$.

6.1.2 Kointegrasjon

Modellen med nominelle verdier er testet for kointegrasjon med Johansen kointegrasjons-test. H_0 antar at antall kointegrerte forhold er r , mot H_1 som antar at antall kointegrerte forhold er flere enn r . Resultatene fra Johansen kointegrasjonstest er gitt i tabell 7.

Tabell 7: Johansen kointegrasjonstest

<i>rank</i>	λ	Trace	5%
0	.	139.87	68.52
1	0.364	43.95*	47.21
2	0.091	23.75	29.68

Variablene som inngår i modellen er BCS, BCF, GOP, VIX og YCI.

* markerer antall kointegrerte forhold basert på trace-statistikken.

Tabellen viser egenverdiene λ , trace statistikken og de kritiske verdi på 5% signifikantnivå. Av tabellen ser man at H_0 blir akseptert for $r = 1$, og en kan konkludere med at modellen inneholder ett kointegrerte forhold. Gitt dette resultatet vil det bli brukt en VECM, med ett kointegrert forhold, videre i denne studien.

6.1.3 Valg av laggede verdier

Tabell 8 viser valg av antall laggede verdier for VAR modellen, basert på AIC, HQIC og SBIC. Testen er en samlet test for alle variablene, med en maksimal lag-lengde satt til 12. Som argumentert for i kapittel 3 vil vi i denne studien bruke AIC som informasjonskriterium. Fra tabellen ser vi at antall lags for VAR modellen er 2. Antall lags for VECM er alltid en mindre enn i den underliggende VAR-modellen, som i dette tilfellet er $2 - 1 = 1$.

Tabell 8: Valg av laggede verdier

Lags	AIC	HQIC	SBIC
0	45.88	45.91	45.96
1	35.71	35.91*	36.20*
2	35.64*	36.01	36.54
3	35.70	36.23	37.01

Inkluderte variabler er BCS, BCF, GOP, VIX og YCI.

* markerer antall laggede verdier gitt IC.

6.1.4 Diagnosetester

Gitt resultatene fra de innledende testene, vil det bli estimert en VEC(1) modell, med ett kointegrert forhold, og de fem variablene presentert i kapittel 5. Etter VECM er kjørt på in-sample datasettet må modellen evalueres, dette blir gjort med de tre diagnosetestene forklart i metode-kapittelet.

Autokorrelasjon

Resultatet fra Lagrange-Multiplier testen er gjengitt i tabell 9. Nullhypotesen er at residualene ikke er autokorrelerte på gitt laglengde, mens alternativhypotesen er at residualene er autokorrelerte. Nullhypotesen blir forkastet når $p < 0.05$. Gitt resultatene fra testen konkluderer vi med at H_0 ikke kan forkastes, og at det ikke eksisterer autokorrelerte residualer i modellen.

Tabell 9: Lagrange-multiplier test for autokorrelerte residualer

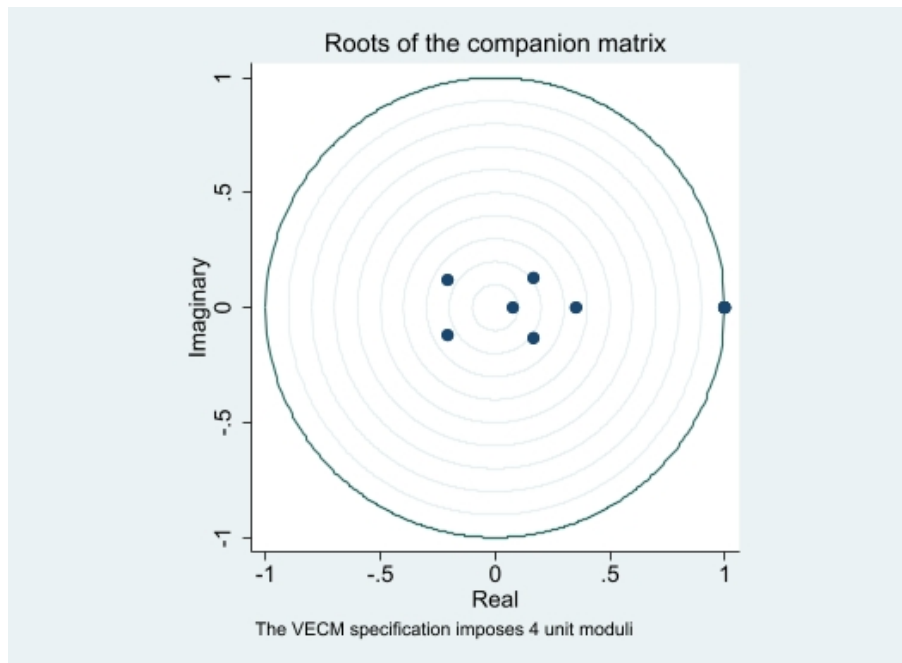
Lags	chi2	df	Prob>chi2
1	35.2321	25	0.0841
2	34.4698	25	0.0982

Testen er en samlet test på alle variabler og deres laggede verdier.

Stabilitet

En VECM er stabil hvis $M - r$ egenverdiene er mindre enn en. For modellen i denne studien som inkluderer fem variabler og ett kointegrert forhold betyr dette modellen inneholder $M - r = 5 - 1 = 4$ egenverdier i kompanjongmatrisen. Der de resterende egenverdiene i modellen må være mindre enn 1. Gitt resultatet fra stabilitets testen, figur 10, ser man at alle de resterende egenverdiene ligger innenfor enhets sirkelen. Dette betyr at VECM tilfredsstillter betingelsen om stabilitet og konvergerer mot langtidslikevekten.

Figur 10: Egenverdi stabilitet VECM



Normalfordeling

Det vil bli benyttet en Jarque-Bera test for å sjekke normalfordelingen til residualene. Nullhypotesen er at residualene er normalfordelte, mot den alternative hypotesen at residualene ikke er normalfordelt. H_0 forkastes når $p < 0.05$. Gitt resultatet i tabell 10 ser man at det kun er ligningen for YCI og BCF som ikke forkaster H_0 . Alle de resterende variablene forkaster H_0 på 1% signifikantnivå, og konkluderer med at residualene ikke er normalfordelt. Dette er også tilfellet for den samlede testen for alle variablene.

Et kjent problem med finansdata er at residualene ikke følger en normalfordeling, noe som ikke er et unntak i denne studien. I likhet med standard regresjonsteori, argumenterer Juselius (2006) for at utvalgsstørrelsen må være tilstrekkelig stor for at residualene ikke skal inneholde skjevhet og kurtose. Når dette er tilfellet, vil residualene følge en normal-distribusjon.

Tabell 10: Jarque-Bera test for normalfordelte residualer

Ligning	chi2	df	Prob > chi2
Δ BCS	20.393	2	0.0000
Δ BCF	5.883	2	0.0528*
Δ GOP	19.295	2	0.0000
Δ VIX	54.179	2	0.0000
Δ YCI	0.682	2	0.7110*
Δ ALL	137.065	10	0.0000

* markerer variabler som er normalfordelte.

En mulig løsning for å få residualene normalfordelt kan være å re-spesifisere modellen. Dette kan gjøres ved å inkludere dummy variabler for ekstremverdier eller strukturelle endringer i tidsserien. Vi vil i denne studien ikke legge særlig vekt på at residualene ikke er normalfordelte, da dette som sagt er et kjent problem med finansdata.

6.1.5 Kortsiktige og langsiktige forhold

I dette avsnittet vil resultatet fra VEC modellen bli analysert. Modellen er basert på resultatene fra de innledende testene, der antall lags og kointegrerte forhold ble bestemt. Videre baserer analysen seg på månedlige observasjoner i perioden 2001(4) - 2018(12), tidligere spesifisert som in-sample delen av datasettet. Da hovedfokuset i denne studien er oljeprisen vil dette bli hovedfokuset i analysen av VECM.

Fra kapittel 3 har vi ligningen til VECM gitt som:

$$\Delta \mathbf{y}_t = \alpha \beta' \mathbf{y}_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta \mathbf{y}_{t-j} + \mathbf{u}_t$$

Her er det første leddet er de langsiktige forholdene, bestående av justerings koeffisienten og den kointegrerte ligningen. Andre leddet er de kortsiktige forholdene. VECM ble modellert i Stata, gitt formelen over kan vi skrive ligningen for BCS på følgende måte:

$$\Delta BCS_t = -0.5571^* \left(1 BCS \quad -1.0259 BCF^* \quad 0.0069 GOP^* \quad 0.0345 VIX \quad 0.0022 YCI^{**} \right)_{t-1} \\ + \left(-0.0473 \Delta BCS \quad 0.3779 \Delta BCF^{**} \quad 0.0096 \Delta GOP \quad -0.1465 \Delta VIX^{**} \quad -0.0062 \Delta YCI \right)_{t-1}$$

* og ** markerer variablene som er signifikante på henholdsvis 1% og 5%. Alle variablenes kortsiktige forhold og justeringskoeffisienter er vedlagt i appendiks, tabell 18 og 19.

Langsiktige forhold

For å kunne tolke justeringskoeffisienten, α , tilhørende BCS må denne være negativ og signifikant. Gitt resultatet i ligningen over, ser man at oljeprisens ECT er -0.5771 og signifikant på 1%. Dette betyr at avvik i oljeprisen som har oppstått i tidligere perioder (altså avvik fra langtidslukevekten), blir korrigert for i inneværende måneden med 0.5771 USD/fat.

De langsiktige forholdene mellom variablene er utledet ved hjelp av Johansen normalisering. Ved å normalisere vektoren på BCS presenteres en ligning som viser hvordan endringer i BCF, GOP, VIX og YCI påvirker BCS på langsikt. β kan altså tolkes som at det eksistere et langsiktig likevektsforhold mellom variablene. Den kointegrerte vektoren til BCS er gitt ved:

$$\beta' = (1.000 BCS, -1.026 BCF, 0.007 GOP, 0.035 VIX, 0.002 YCI)$$

Fra den normaliserte vektoren kan man se at BCS og BCF nesten har et perfekt langsiktig forhold, som kan skyldes variablenes naturlige sammenheng. Dette bekreftes også ved å sammenligne BCS og BCF i figur 8 og 9 i kapittel 5.

Når resultatet skal tolkes blir fortegnet til koeffisienten endret. Og vi får følgende langsiktige forhold mellom BCS og de andre variablene:

$$BCS_t = 1.026 BCF_t - 0.007 GOP_t - 0.035 VIX_t - 0.002 YCI_t \quad (14)$$

Ligning 14 vil bli tolket for hver variabel, og dens påvirkning på BCS.

BCF - Brent crude futures

Brent crude futures er målt i USD/fat olje. Når BCF øker med 1 USD vil dette resultere i en langsiktig prisøkning på 1.0256 USD/fat, alt annet likt. Dette forholdet gir mening da BCF kan ses på som et mål for etterspørsel av olje. Når en avtale om kjøp og salg av oljefutures blir inngått, gjenspeiler dette markedets forventninger om den fremtidige oljeprisen. Dette gir igjen en pekepinn på den fremtidige etterspørselen, og når etterspørselen etter olje øker, øker også prisen.

GOP - Global oljeproduksjon

Global oljeproduksjon er målt i millioner fat med olje. Når GOP øker med en million fat vil dette resultere i en langsiktig prisreduksjon på 0.0069 USD/fat, alt annet likt. Dette kan virke som en stor endring i oljeproduksjonen, men i et globalt perspektiv er det ikke det. Norge hadde alene en oljeproduksjon på 1 773 000 fat med olje i mars 2021, Oljedirektoratet (2021). Hvis Norge øker sin produksjon med 37% vil dette resultere i prisreduksjonen over. Videre gir dette forholdet mening. Økt oljeproduksjon fører til at oljelagrene kan fylles opp, som betyr at tilbudet øker. Økt tilbud resulterer på sin side i at oljeprisen reduseres.

VIX - Volatilitet indeks

Volatilitetsindeksen er målt i poeng. Når VIX øker med ett poeng vil dette resultere i en langsiktig prisreduksjon på 0.0345 USD/fat, alt annet likt. Økte forventninger om at uroligheter i markedet, kan resultere i en nedgang i aktivitetsnivået i markedet. Når markedsaktivitet går ned resulterer dette i lavere etterspørsel etter handelsvarer slik som olje. Når etterspørselen etter olje går ned, går også oljeprisen ned. Videre kan det diskuteres om denne effekten er signifikant på lang sikt, da VIX kun er signifikant på 10%. Forholdet mellom BCS og VIX gir mening på kort sikt, da uforutsette kriser og uroligheter påvirker VIX med en gang, som igjen påvirker BCS slik man så i mars 2020 på grunn av Covid-19.

YCI - Månedlig år-til-år endring i oljelager

De månedlige år-til-år endringene i oljelagrene er mål i millioner fat med olje. Når YCI øker med en million fat vil dette resultere i en langsiktig prisreduksjon på 0.0022 USD/fat, alt annet likt. På lik linje med eksemplet under GOP, betyr dette at Norge alene øker sitt

oljelager med 1 millioner fat sammenlignet med året før. Når lagerbeholdningen øker, økes tilbudet som igjen reduserer oljeprisen.

Kortsiktige forhold

Fra ligning 15 ser man den kortsiktige dynamikken mellom variablene og BCS.

$$\begin{aligned} \Delta BCS_t = & -0.0473\Delta BCS_{t-1} + 0.3779\Delta BCF_{t-1}^{**} + 0.0096\Delta GOP_{t-1} \\ & - 0.1465\Delta VIX_{t-1}^{**} - 0.0062\Delta YCI_{t-1} \end{aligned} \quad (15)$$

** markerer kortsiktige forhold som er signifikante på 5%.

Man kan her legge merke til variablene har samme påvirkning på oljeprisen på kort- og lang sikt. Videre er det kun BCF og VIX som har en signifikant påvirkning på oljeprisen på kort sikt, med følgende virkningsforhold: En økning i BCF på 1 USD/fat, vil resultere i en kortsiktig økning i BCS på 0.3779 USD/fat, mens en økning i VIX på ett poeng, vil resultere i en reduksjon i BCS på 0.1465 USD/fat, alt annet likt.

De kortsiktige forholdene mellom BCF, VIX og BCS følger samme resonnement som for de langsiktige forholdene.

6.2 LSTM

Ved seleksjon av LSTM-struktur, så ble dybden til nettverket begrenset til tre HL. Et dype-re nettverk enn dette resulterer i problemer med overfitting som forklart av delkapittel 4.3. I figur 11 nedenfor, vises utviklingen i seleksjonsprosessen av de to LSTM-strukturene. Hvert individ i optimeringen har en kromosomlengde på 42 bits, og en populasjonsstørrelse på 20 individer i hver generasjon. Ved seleksjon av $LSTM_{t+1}$, så ble det konstruert og evaluert 59 generasjoner før termineringskriteriet ble oppfylt. Mens for $LSTM_{t+3}$ ble det iterert 120 generasjoner.

Figur 11: Grafisk illustrasjon av individene under GA-optimeringen



(a) GA optimering av LSTM-modellens struktur og treningsparametre for $(t+1)$ -prognosene.

(b) GA optimering av LSTM-modellens struktur og treningsparametre for $(t+3)$ -prognosene.

$LSTM_{t+1}$ og $LSTM_{t+3}$ består av en dybde på henholdsvis 2 og 3 hidden layers. Lag-periodene som inkluderes er på 1 og 6, mens dropout-raten mellom nivåene i nettverket er på 5% og 27%. Dette vises av tabell 11, hvor treningsparametrene, batch size og epochs, for de to LSTM-strukturene er like. Antall vektorer som justeres under trening av $LSTM_{t+1}$ er på 1.141.546, mens antallet for $LSTM_{t+3}$ er på 1.397.351. Ytterligere informasjon om hvert enkelt nivå er gitt av tabell 23 og 24.

Tabell 11: Modellseleksjonen fra GA-optimeringen

	Lags	Hidden Layers	Epochs	Batch Size	Dropout %
$LSTM_{t+1}$	1	2	2	1	5
$LSTM_{t+3}$	6	3	2	1	27

Tabell 12: Antall vektorer i LSTM-nettverket

	Total Params	Trainable Params
$LSTM_{t+1}$	1.141.546	1.141.546
$LSTM_{t+3}$	1.397.351	1.397.351

6.3 Fremtidsprognoser

I denne delen vil resultatene fra modellenes prediksjoner analyseres og sammenlignes. Sammenligningen baserer seg på resultatene fra fremtidsprediksjonene på én og tre måneder, der modellene vil bli evaluert ut fra evalueringskriteriene RMSE og MAPE. Videre vil modellenes prediksjonsavvik og normalfordeling for hele out-of-sample perioden, 2019(1)-2021(3), bli sammenlignet.

6.3.1 RMSE og MAPE

Resultatene fra RMSE og MAPE viser gjennomsnittet for alle fremtidsprediksjonene, for $t + 1$ og $t + 3$ tilsvarer dette gjennomsnittet av 26 og 24 observasjoner.

Tabell 13: RMSE og MAPE for hele tidsperioden

Model	RMSE	MAPE
$VECM_{t+1}$	4.952135	8.846951
$LSTM_{t+1}$	4.398797*	8.280921*
$VECM_{t+3}$	15.620113*	29.663968*
$LSTM_{t+3}$	15.699350	30.566945

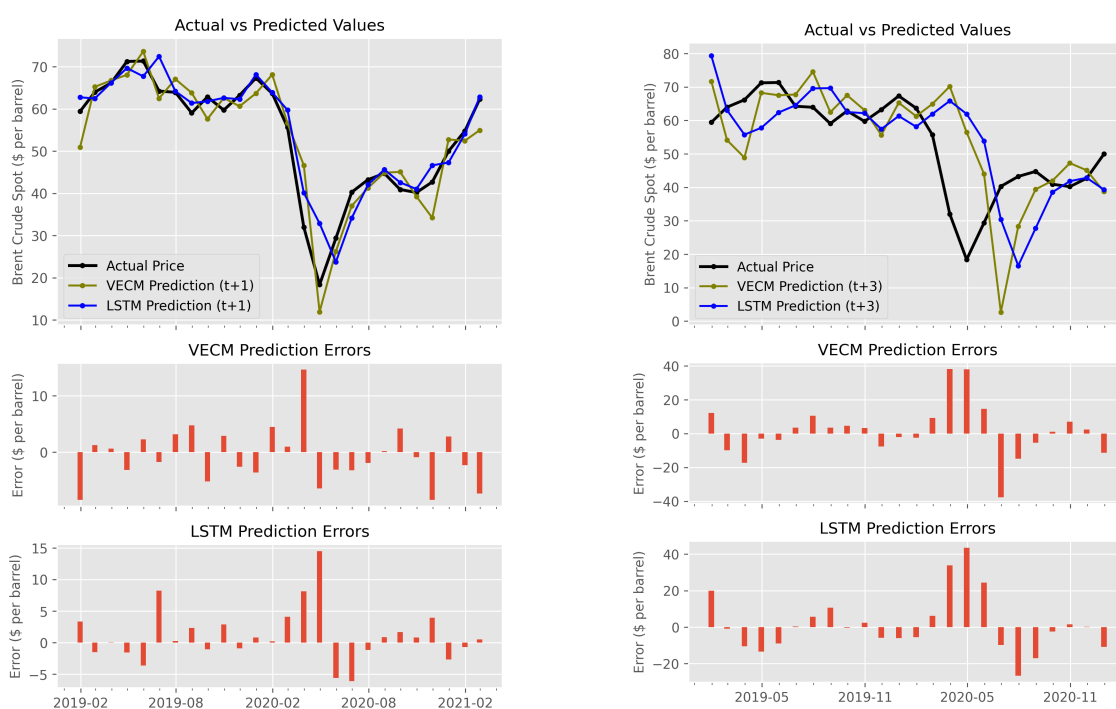
RMSE og MAPE viser gjennomsnittet av alle predikerte verdier for begge modellene i perioden 2019(1)-2021(3).

* markerer den modellen med det minste avviket.

For fremtidsprognosene for en måned frem i tid er det LSTM som gjør det best, og har det lavest avviket gitt RMSE og MAPE. Fra tabellen ser man at VECM og LSTM i snitt bommer med henholdsvis 4.95 USD/fat og 4.40 USD/fat gitt RMSE, og 8.85% og 8.28% gitt MAPE. Modellens fremtidsprediksjoner og den virkelige verdien på BCS er illustrert i figur 12a. Frem til 2020(2) ser det ut til at begge modellene treffer spotprisen forholdsvis godt, med unntak av noen litt større avvik. I perioden 2020(2)-2020(6) sliter begge modellene med å treffe spotprisen, og man kan se at det er her de største avvikene oppstår. Modellenes treffsikkerhet forbedrer seg igjen mellom 2020(6)-2020(11), før modellene igjen opplever endel avvik ut tidsperioden.

For prognosen tre måneder frem i tid er det VECM som gjør det best, og har det laveste avviket gitt RMSE og MAPE, altså motsatt fra $t + 1$. Her bommer VECM i snitt med 15.62 USD/fat og LSTM med 15.70 USD/fat gitt RMSE, og 29.66% og 30.57% gitt MAPE. Modellens fremtidsprediksjoner og den virkelige verdien er illustrert i figur 12b. Til forskjell fra $t + 1$ prognosene sliter modellene gjennom hele perioden, og treffer kun på spotprisen noen få ganger. Man kan fortsatt merke seg at begge modellene opplever de største avvikene i tidsperioden 2020(3)-2020(10), før de stabiliserer seg litt igjen mot slutten av perioden.

Figur 12: Modellenes fremtidsprediksjoner og prediksjonsavvik



(a) VECM og LSTM-modellens ($t + 1$)-prognoser på out-of-sample data.

(b) VECM og LSTM-modellens ($t + 3$)-prognoser på out-of-sample data.

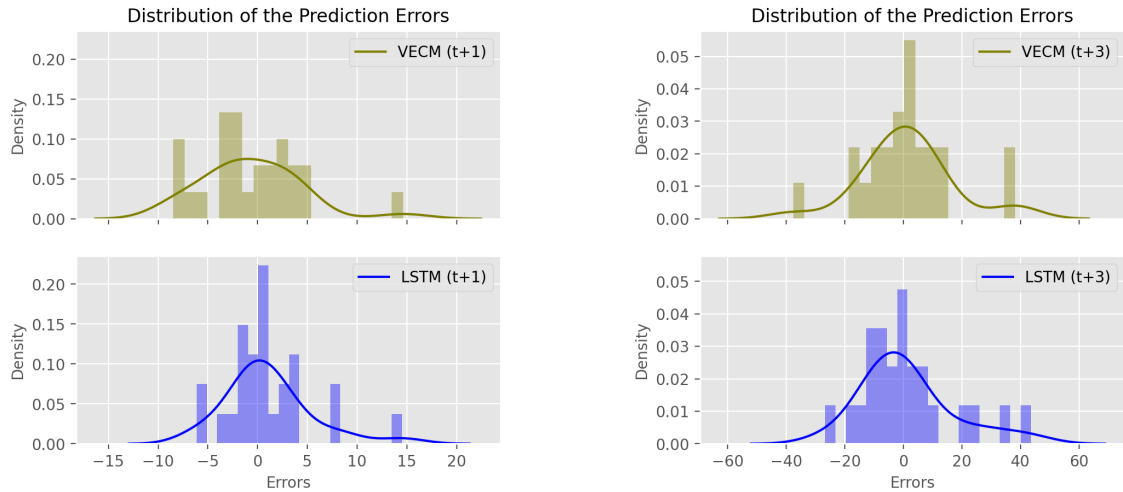
Videre kan vi se på modellenes fordeling av residualene for prognosene.

Fordelingen til modellenes avvik for $t + 1$ kan ses i figur 13a. Ut fra denne kan vi se at LSTM bærer preg av kurtose og er mer høyre-skjev sammenlignet med VECM, som blir bekreftet av Jarque-Beras normality test. Testen konkluderer videre med at prediksjonsavvikene til VECM er normalfordelt noe som ikke er tilfellet for LSTM. Resultatene fra testen finnes i tabell 14.

Fordelingen til modellenes $t + 3$ prediksjonene er illustrert i figur 13b. Fra denne ser man at LSTM er litt mer høyre-skjev sammenlignet med VECM, i likhet med prediksjonen for $t + 1$. I motsetning til $t + 1$ prognosene har nå begge modellene normalfordelte residualer.

VECM har altså normalfordelte residualer på begge prediksjonene, selv om dette ble forkastet når modellen ble testet på in-sampel perioden.

Figur 13: Fordelingen til modellenes prediksjonsavvik.



(a) Fordelingen til VECM og LSTM-modellens prediksjonsavvik for $(t + 1)$ -prognoser på out-of-sample data.

(b) Fordelingen til VECM og LSTM-modellens prediksjonsavvik for $(t + 3)$ -prognoser på out-of-sample data.

Tabell 14: Jarque-Bera's test for normality

	JB-stat	Chi^2 two-tail prob.	Skew	Kurtosis
$VECM_{t+1}$	5.3729	0.0681*	0.809605	4.5289747
$LSTM_{t+1}$	10.0890	0.0064	1.1446	5.0181
$VECM_{t+3}$	2.4780	0.2897*	0.3217	4.4367
$LSTM_{t+3}$	4.7557	0.0927*	0.9990	3.8739

H_0 : residualene er normalfordelt, forkaster H_0 når $p < 0.05$.

* markerer normalfordelte residualer

6.3.2 RMSE og MAPE pre- og post Covid-19

Verdiene til RMSE og MAPE for $t + 1$ og spesielt $t + 3$, er større enn ønskelig for begge modellene. Dette kan tyde på at fremtidsprediksjonene inneholder enkelte avvik som skiller seg vesentlig fra den virkelige verdien. Av figur 12 ser man at begge modellene sliter i perioder med store avvik etter 2020(3) og frem til 2021(3). Dette kan ses i sammenheng med flere hendelser som skjedde i denne tidsperioden: Våren 2020 ble Covid-19 til en verdenspandemi som slo ut på den globale økonomien. Rundt det samme tidspunktet erklærte Saudi-Arabia en priskrig på olje mot Russland. Videre ble tilbudet av olje alt for stort grunnen den svekkede økonomien, som igjen presset prisene ned. Overtilbudet av olje kom ikke under kontroll før OPEC gikk med på å redusere oljeproduksjonen sommeren 2020, Blessing (2021).

For å sjekke om det er tidsperioden som er grunnen til de store prediksjonsavvikene eller om det er modellene, deles out-of-sample perioden i to. De to nye tidsperiodene blir kalt pre- og post Covid-19, og er gitt ved 2019(1)-2020(2) og 2020(3)-2021(3).

Prediksjonsavvikene for $t + 1$ for de to nye tidsperiodene er gitt i tabell 15. Her er det LSTM som gjør det best både før og etter Covid-19, og har det lavest avviket gitt RMSE og MAPE. Av resultatene kan det se ut til at hypotesen om at pandemien ga utslag på prognosene stemte. Differansen mellom pre- og post Covid-19 for RMSE er på 2.24 USD/fat for VECM og 2.55 USD/fat for LSTM. Mens differansen for MAPE er på 8.71% for VECM og 9.66% for LSTM.

Videre ser man fra de nye resultatene at RMSE og MAPE pre- og post Covid-19 ikke er like hele out-of-sample perioden. For pre Covid-19 og hele perioden er det en differanse for RMSE på 1.16 USD/fat for VECM og 1.36 USD/fat for LSTM, mens for MAPE er det en differanse på 3.72% for VECM og 4.76% for LSTM.

For post Covid-19 og hele perioden er det en differanse for RMSE på 1.08 USD/fat for VECM og 1.18 USD/fat LSTM, og for MAPE er den på 4.98% for VECM og 4.90% for LSTM.

Tabell 15: RMSE og MAPE $t + 1$ før og etter Covid-19

Model	RMSE	MAPE
$VECM_{pre}$	3.787926	5.131098
$LSTM_{pre}$	3.037023*	3.519320*
$VECM_{post}$	6.032809	13.836121
$LSTM_{post}$	5.582381*	13.182114*

Pre er tiden før Covid-19, 2019(1)-2020(2)

Post er tiden etter Covid-19 inntraff, 2020(3)-2021(3).

* markerer den modellen med det minste avviket.

Prediksjonsavvikene for $t + 3$ for de to nye tidsperiodene er gitt i tabell 15. Her er det VECM som gjør det best før Covid-19, mens LSTM gjør det best etter, gitt RMSE og MAPE. Differansen for pre- og post Covid-19 er her mye større sammenlignet med de nye prediksjonene for $t + 1$. For RMSE er denne nå på 14.3 USD/fat for VECM og 13.36 USD/fat for LSTM, mens for MAPE er den nå på 47.60% for VECM og 45.54% for LSTM.

Også her ser man at RMSE og MAPE pre- og post Covid-19 avviker fra hele perioden.

For pre Covid-19 og hele perioden er det en differanse for RMSE på 7.64 USD/fat for VECM og 7.02 USD/fat for LSTM, mens for MAPE er det en differanse på 19.09% for VECM og 19.72% for LSTM.

For post Covid-19 og hele perioden er det en differanse for RMSE på 6.66 USD/fat for VECM og 6.34 USD/fat LSTM, og for MAPE er den på 28.50% for VECM og 25.82% for LSTM.

Tabell 16: RMSE og MAPE $t + 3$ før og etter Covid-19

Model	RMSE	MAPE
$VECM_{pre}$	7.980894*	10.571991*
$LSTM_{pre}$	8.687494	10.857963
$VECM_{post}$	22.280001	58.159519
$LSTM_{post}$	22.042356*	56.392736*

Pre er tiden før Covid-19, 2019(1)-2020(2)

Post er tiden etter Covid-19 inntraff, 2020(3)-2021(3).

* markerer den modellen med det minste avviket.

Av resultatene fra de to nye tidsperiodene kan det se ut til at de opprinnelige evalueringstidsperiodene kan ha noe av grunnen til avvikene i prediksjonene. Begge modellenes prediksjoner gjør det bedre pre-Covid-19 sammenlignet med den opprinnelige tidsperioden, mens post Covid-19 prediksjonene gjør det dårligere. RMSE og MAPE er nå av akseptabel størrelse for $t + 1$ pre Covid-19 for begge modellene, men for post Covid-19 og $t + 3$ prognosene er de fortsatt av en betydelig størrelse. Det kan av dette tyde på at det er modellene som gir størst utslag på prediksjonene, og ikke tidsperioden i seg selv. Videre kan det diskuteres om det er variablene som inngår i modellen som ikke klarer å fange opp endringene, eller om det er modellspesifikasjonene som er brukt.

7 Oppsummering og konklusjon

Hensikten med denne studien er å gi en mer realistisk sammenligning av prediksjonsevnen til en økonomisk modell og en maskinlæringsmodell. Modellene som blir sammenlignet er en VECM og en LSTM.

Analysen tar utgangspunkt i et datasett med månedlige observasjoner fra 2001(4)-2021(3). Basert på tidligere studier ble data for tidsperioden innhentet på en rekke variabler, der de endelige variablene som inngår i modellene blir valgt ut via en forward-selection. I analysen av VECM finner vi kun oljefutures og VIX til å ha en signifikant påvirkning på oljeprisen på kort sikt. Mens på lang sikt har oljefutures, global oljeproduksjon og endringer i oljelagrene en signifikant virkning på oljeprisen. En slik analyse er ikke gjennomført på LSTM, da prosessen til modellen er en "black-box", med andre ord noe som skjer i det skjulte. Her blir istedenfor modellens struktur og optimeringsprosess gjennomgått.

Gitt modellenes fremtidsprediksjoner for én og tre måneder, viser det seg at modellene gjør det relativt likt for begge tidsperiodene. For den gjennomsnittlige $t + 1$ prediksjonen er det LSTM modellen som gir det minste avviket, mens for den gjennomsnittlige $t + 3$ prediksjonen er det VECM som gjør det best. Selv om modellene nesten gir de samme prediksjonene, er avvikene av en betydelig størrelse. Som et resultat av dette velger vi å dele opp evalueringssettet i to. De to nye tidsperiodene blir inndelt i tiden før og etter Covid-19 ble en verdenspandemi. Grunnen til dette er for å se om det er tidsperioden eller modellene som er årsaken til de store avvikene. De nye resultatene bekreftet at modellene fortsatt presterte tilnærmet likt. Der vi også ser en forbedring i begge prognosene for tiden før Covid-19.

Gitt RMSE og MAPE er resultatene som følge: For hele tidsperioden og pre Covid-19 perioden er det LSTM som gjør det beste for $t + 1$ prediksjonene, mens VECM gjør det best for $t + 3$. Mens for post Covid-19 perioden gjør LSTM det best for både $t + 1$ og $t + 3$. Når det kom til avvikene i fremtidsprediksjonen kan det virke som at hovedproblemet ligger i modellenes evne til å fange opp svingningene i oljeprisen. Om det er variablene som blir brukt, eller optimeringen av modellene, eller en kombinasjon av disse to, får være en problemstilling til videre arbeid.

Sammenlignet med tidligere studier som har sammenlignet økonomiske modeller og maskinlæringsmodeller, finner vi i denne studien motstridende resultater. Det viste seg at når to jevnge modeller ble sammenlignet på likt grunnlag, her i form av likt variable grunnlag, er det små differanser som skiller modellene fra hverandre. Dette resultatet skiller seg ut fra tidligere studier som Verstyuk (2020); Sagheer og Kotb (2019); Wang mfl. (2009), der alle konkluderte med at maskinlæringsmodeller er overlegene økonomiske modeller.

Konklusjonen om hvilken modell som presterer best er tvetydige, da ingen av modellene skiller seg vesentlig ut. Tatt i betraktning de seks forskjellige tidshorizontene prediksjonene er utført på, er det LSTM som gir de minste avvikene fire av disse seks periodene, gitt RMSE og MAPE. Med bakgrunn i dette kan det tyde på at LSTM modellen gir de beste fremtidsprediksjonene på oljeprisen. Vi ønsker likevel å gjøre oppmerksom på at differansene er små og det anbefales å gjøre ytterlige analyser før en endelig konklusjon tas.

8 Videre forskning og begrensninger

Arbeidet med denne studien har vært tidkrevende og vi har tilegnet oss mye kunnskap helt til siste slutt. Dette har gitt oss en bedre forståelse for både økonomiske modeller og maskinlæringsmodeller. Dette har resultert i at vi har blitt bevisst på begrensningene med studien, og videre forskning på lignende problemstillinger.

Antall variabler som inkluderes i modellene er noe begrenset. Av dette ser det ut til at modellene ikke klarer å fange opp alle svingningene i oljeprisen. Vår anbefaling til videre forskning er å inkludere flere variabler som kan ha signifikant påvirkning på oljeprisen. Her kan man for eksempel inkludere variablene som inngår i Alquist, Kilian og Vigfusson (2013).

En av begrensningene med denne studien er at det ikke ble inkludert eksogene eller dummy-variabler i modellene. Ved å inkludere dummy-variabler for ekstremverdier og strukturelle skift, kunne vi unngått modellspesifikke problemer med VECM. En av problemene med modellen er at residualene ikke er normalfordelte, dette ble gjennomgått i 6.1.4. Når residualene ikke er normalfordelt er ikke VECM velspesifisert, noe som kan resultere i misvisende resultater.

En av LSTM modellens begrensninger er at det ikke er mulig å gi en økonomisk tolkning av denne. Det er umulig å analysere forholdene mellom variablene som inngår i modellen, og deres påvirkning på oljeprisen.

I videre forskning kan det vært interessant å inkludere en no-change modell som benchmark. En sammenligning av VECM og LSTM opp mot en no-change modell, vil gi et mer objektive mål på hvor gode fremtidsprediksjonene er. Man kan da få et mål på om modellene faktisk gir bedre prognoser enn hvis man antar at prisen ikke endrer seg.

Denne studien skulle opprinnelig benytte dagligdata i analysen. Grunnen til at dette ble valgt bort var at VECM ikke var valid gitt diagnosetesten. Høy-frekvent data inneholder mye støy, noe som er vanskelig å håndtere i en VECM da det ofte resulterer i at residualene er autokorrelerte og ikke-normalfordelte. Denne begrensningen gjelder ikke for LSTM, og det ville vært interessant å få sammenlignet modellene på daglig data, gitt at man klarte å få en velspesifisert VECM.

En annen mulighet kan være å bruke kvartalsvis data, man får da mulighet til å inkludere andre makro-indikatorer.

En annen begrensning med denne oppgaven er at det ikke er utført impulsrespons analyser. Hadde dette blitt gjennomført kunne man sagt noe om påvirkningen på oljeprisen gitt ett sjokk i de andre variablene.

I videre forskning kan det også vært mulig å sammenligne prediksjonsevnen til en ANN og en auto-regressive distributed lag (ARDL), kontra en VECM og en LSTM. Disse modellene har en lik informasjonsstruktur, noe som resulterer i at de kan sammenlignes på likt grunnlag.

Referanser

- Alexander, Carol (2008). *Market risk analysis, practical financial econometrics*. Bd. 2. John Wiley & Sons.
- Algorithms — DEAP 1.3.1 documentation* (mai 2021). <https://deap.readthedocs.io/en/master/api/algo.html#module-deap.algorithms>. (Accessed on 05/21/2021).
- Alquist, Ron og Lutz Kilian (2010). «What do we learn from the price of crude oil futures?» I: *Journal of Applied econometrics* 25(4), s. 539–573.
- Alquist, Ron, Lutz Kilian og Robert J Vigfusson (2013). «Forecasting the price of oil». I: *Handbook of economic forecasting*. Bd. 2. Elsevier, s. 427–507.
- Baumeister, Christiane og Lutz Kilian (2014). «What central bankers need to know about forecasting oil prices». I: *International Economic Review* 55(3), s. 869–889.
- Beckers, Benjamin og Samya Beidas-Strom (2015). «Forecasting the Nominal Brent Oil Price with VARs-One Model Fits All?» I:
- Bishop, Christopher M (2006). *Pattern recognition and machine learning*. springer.
- Blessing, Elizabeth (jan. 2021). *What Happened to Oil Prices in 2020*. <https://www.investopedia.com/articles/investing/100615/will-oil-prices-go-2017.asp>. (Accessed on 05/30/2021).
- Brooks, Chris (2019). *Introductory econometrics for finance*. Cambridge university press.
- Brøgger Anton; Voje, Kjetil Lysne (sep. 2020). *Utviklingslæren*. <https://snl.no/utviklingslaeren>. Hentet 9. mai 2021.
- Bäck, Thomas, David B Fogel og Zbigniew Michalewicz (2018). *Evolutionary computation 1: Basic algorithms and operators*. CRC press.
- Cheung, Yin-Wong og Kon S Lai (1993). «Finite-sample sizes of Johansen's likelihood ratio test for cointegration». I: *Oxford Bulletin of Economics and statistics* 55, s. 3.
- Dickey, David A og Wayne A Fuller (1979). «Distribution of the estimators for autoregressive time series with a unit root». I: *Journal of the American statistical association* 74(366a), s. 427–431.
- Engle, Robert F og Clive WJ Granger (1987). «Co-integration and error correction: representation, estimation, and testing». I: *Econometrica: journal of the Econometric Society*, s. 251–276.

-
- Fortin, Félix-Antoine mfl. (jul. 2012). «DEAP: Evolutionary Algorithms Made Easy». I: *Journal of Machine Learning Research* 13, s. 2171–2175.
- Glover, Fred W og Gary A Kochenberger (2006). *Handbook of metaheuristics*. Bd. 57. Springer Science & Business Media.
- Goodfellow, Ian, Yoshua Bengio og Aaron Courville (2016). *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press.
- Gulli, Antonio og Sujit Pal (2017). *Deep learning with Keras*. Packt Publishing Ltd.
- Hochreiter, Sepp (1998). «The vanishing gradient problem during learning recurrent neural nets and problem solutions». I: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 6(02), s. 107–116.
- Hochreiter, Sepp og Jürgen Schmidhuber (1997). «Long short-term memory». I: *Neural computation* 9(8), s. 1735–1780.
- Holland, John Henry mfl. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press.
- Hua, Yuxiu mfl. (2019). «Deep learning with long short-term memory for time series prediction». I: *IEEE Communications Magazine* 57(6), s. 114–119.
- Johansen, Soren og Katarina Juselius (1990). «Maximum likelihood estimation and inference on cointegration—with appucations to the demand for money». I: *Oxford Bulletin of Economics and statistics* 52(2), s. 169–210.
- Johansen, Søren (1988). «Statistical analysis of cointegration vectors». I: *Journal of economic dynamics and control* 12(2-3), s. 231–254.
- Juselius, Katarina (2006). *The cointegrated VAR model: methodology and applications*. Oxford university press.
- Kristjanpoller, Werner og Marcel C Minutolo (2016). «Forecasting volatility of oil price using an artificial neural network-GARCH model». I: *Expert Systems with Applications* 65, s. 233–241.
- Lütkepohl, Helmut (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.
- Manescu, Cristiana og Ine Van Robays (2014). «Forecasting the Brent oil price: addressing time-variation in forecast performance». I:

-
- McCallum, Andrew H og Tao Wu (2005). «Do oil futures prices help predict future oil prices?» I: *FRBSF Economic Letter*(2005-38).
- Nielsen, Michael A (2015). *Neural networks and deep learning*. Bd. 25. Determination press San Francisco, CA.
- Oljedirektoratet (apr. 2021). *Produksjonstall mars 2021*. <https://www.npd.no/fakta/nyheter/Produksjonstal/2021/produksjonstal-mars-2021/>. Hentet 16. mai 2021.
- Pesaran, M Hashem (2015). *Time series and panel data econometrics*. Oxford University Press.
- Regjeringen.no (okt. 2020). *Statsbudsjettet 2021: Statens inntekter og utgifter - regjeringen.no*. <https://www.regjeringen.no/no/statsbudsjett/2021/statsbudsjettet-2021-statens-inntekter-og-utgifter/id2768898/>. Hentet 02/01/2021.
- Rumelhart, David E, Geoffrey E Hinton og Ronald J Williams (1986). «Learning representations by back-propagating errors». I: *nature* 323(6088), s. 533–536.
- Sagheer, Alaa og Mostafa Kotb (2019). «Time series forecasting of petroleum production using deep LSTM recurrent networks». I: *Neurocomputing* 323, s. 203–213.
- Shmueli, Galit mfl. (2010). «To explain or to predict?» I: *Statistical science* 25(3), s. 289–310.
- Sims, Christopher A (1980). «Macroeconomics and reality». I: *Econometrica: journal of the Econometric Society*, s. 1–48.
- Sims, Christopher A, James H Stock og Mark W Watson (1990). «Inference in linear time series models with some unit roots». I: *Econometrica: Journal of the Econometric Society*, s. 113–144.
- Tao, Ran og Chris Brooks (2019). «Python Guide to Accompany Introductory Econometrics for Finance». I: *Available at SSRN 3475303*.
- Verstyuk, Sergiy (2020). «Modeling multivariate time series in economics: From auto-regressions to recurrent neural networks». I: *Available at SSRN 3589337*.
- Wang, Wen-Chuan mfl. (2009). «A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series». I: *Journal of hydrology* 374(3-4), s. 294–306.

Appendiks

A Tabeller

Tabell 17: Oversikt over variabel-koder og datakilder.

VARIABEL	DATAKILDE	KODE
BCS	FRED	MCOILBRENTU
BCF	Refinitiv Eikon	LCOc1
GOP	Refinitiv Eikon	.TRXFLDGLPUOILE
VIX	Refinitiv Eikon	.VIX
YCI	Refinitiv Eikon	Finn koden!

Tabell 18: Kortsiktige forhold VECM

	cons	ΔBCS_{t-1}	ΔBCF_{t-1}	ΔGOP_{t-1}	ΔVIX_{t-1}	ΔYCI_{t-1}
$\Delta BCS_t =$	0.1995	-0.0473	+0.3779**	0.0096	-0.1465**	-0.0062
$\Delta BCF_t =$	-0.0034	-0.0114	0.4019***	0.0108	-0.2461**	-0.0119
$\Delta GOP_t =$	-0.0650	1.0802***	-0.0237	0.0151	-1.1151**	-0.0504
$\Delta VIX_t =$	0.0553	0.0181	-0.2160	0.0163	-0.0359	-0.0048
$\Delta YCI_t =$	-0.3044	1.5394	-2.2940	0.0903	0.2201	0.1834**

*, **, *** representerer signifikante variabler på henholdsvis 1%, 5% og 10%.

Tabell 19: ECT & Johansen normalisering av langsiktige forhold

	α	β
BCS	-0.5771*	1
BCF	0.5773**	-1.0259*
GOP	2.5993**	0.0069*
VIX	-0.4056**	0.0345***
YCI	-1.0132	0.0021**
Con		-0.4291

*, **, *** representerer signifikante variabler på henholdsvis 1%, 5% og 10%.

Tabell 20: Operatortypene anvendt i GA-optimering

Operator	Type
Evolutionary Algorithm	<i>OneMin()</i>
Fitness Score	RMSE
Evaluation	<i>min(fitness)</i>
Selection	Tournament
Crossover	2 point crossover
Mutation	bitflip

Tabell 21: Gyldige allele-verdier i GA-optimering

	Lag Size	Hidden Layers	Units	Epochs	Batch Size
Min Value	1	1	50	1	1
Max Value	12	3	500	20	3

Tabell 22: Parameterene til GA-optimeringen av de to LSTM-strukturene som selekteres.

	Population Size	Max Generations	Chromosome Length
Optimization Parameters	20	120	42

Tabell 23: Dimensjonen til $LSTM_{t+1}$

Layer (type)	Output Shape	Param #
lstm_1	(None, 1, 444)	799200
dropout_1	(None, 1, 444)	0
lstm_2	(None, 1, 145)	342200
dropout_2	(None, 1, 145)	0
dense_1	(None, 1, 1)	146
Total Params:	1.141.546	
Trainable params:	1.141.546	

Tabell 24: Dimensjonen til $LSTM_{t+3}$

Layer (type)	Output Shape	Param #
lstm_1	(None, 6, 257)	270364
dropout_1	(None, 6, 257)	0
lstm_2	(None, 6, 225)	270364
dropout_2	(None, 6, 225)	0
lstm_3	(None, 6, 318)	691968
dropout_3	(None, 6, 318)	0
dense_1	(None, 6, 1)	319
Total Params:	1.397.351	
Trainable params:	1.397.351	

Tabell 25: Deskriptiv statistikk av modellenes prediksjonsavvik

	$VECM_{t+1}$	$LSTM_{t+1}$	$VECM_{t+3}$	$LSTM_{t+3}$
count	26	26	24	24
mean	-0.6388	1.0621	1.3820	1.2779
std	5.0080	4.3532	15.8935	15.9838
min	-8.4801	-6.1242	-37.6092	-26.6850
25%	-3.2166	-1.1418	-5.8968	-9.1862
50%	-1.3373	0.3945	1.7184	-0.6199
75%	2.6448	2.7554	7.5970	5.7961
max	14.6380	14.4976	38.1329	43.5435
skew	0.8600	1.2159	0.3435	1.0669
kurt	2.1414	2.7395	2.0868	1.3863

