

Peder Bollingmo
Thomas Toft

Prediksjon av strømpriser

Bacheloroppgave i Økonomi og administrasjon

Veileder: Denis Becker

April 2022

Peder Bollingmo
Thomas Toft

Prediksjon av strømpriser

Bacheloroppgave i Økonomi og administrasjon
Veileder: Denis Becker
April 2022

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
NTNU Handelshøyskolen



Kunnskap for en bedre verden

Forord

Denne oppgaven er den avsluttende oppgaven i vårt bachelorstudium i Økonomi og Administrasjon ved NTNU handelshøyskolen i Trondheim. Oppgaven er skrevet av Thomas Toft og Peder Bollingmo, og innholdet i denne oppgaven står for forfatternes regning. Oppgaven tilsvarer 7,5 studiepoeng, og er knyttet til spesialiseringsretningen innen business analytics.

Vi har valgt å hovedsakelig bruke kunnskapen vi har lært i Essentials of Business Analytics for å svare på om et kunstig nevralt nettverk kan predikere elspot priser mer nøyaktig enn en linær modell. Å jobbe mot en løsning på denne oppgaven har vært interessant og lærerikt, og vi føler virkelig at vi har fått bruk for det vi har lært.

Vi ønsker derfor å rette en takk til vår veileder Denis Becker, som har gjort fagene innen Business Analytics både interessante og lett fordøyelige. Vi er veldig takknemlige for hjelpen vi har fått og de gode svarene på våre spørsmål.

Trondheim 2021,
Thomas Toft og Peder Bollingmo



Sammendrag

De økende kraftprisene har fått mye oppmerksomhet den siste tiden, og mange har uttrykt sin misnøye og uro knyttet til de høye strømprisene. Å kunne finne modeller som predikerer prisene vil kunne gi en økt grad av kontroll, og være med på å gjøre kraftmarkedet mer forutsigbart. Det er mange veier man kan gå for å utvikle gode forklaringer på bevegelsene i kraftmarkedet, men vi har valgt å se på to ganske ulike modeller for å kunne jobbe mot å finne en trend, som kanskje kan bidra til videre forskning.

Problemstillingen vi landet på er «*Kan et kunstig nevralt nettverk predikere historiske elspot priser bedre enn en linære modell?*». Problemstillingen spesifiserer at det handler om forskjellen på de to modellene, selv om det selvsagt også er interessant å se på resultatene isolert sett. Ved bruk av denne problemstillingen valgte vi ut data basert på teori om kraftmarkedet og dens drivere. Her har vi både med de relativt åpenbare direkte faktorene slik som: produksjon, forbruk, temperatur, vindstyrke og mengden vann i vannmagasiner. I tillegg til at vi har med en mengde mindre åpenbare og indirekte faktorer slik som: valutakurs, kullpris, karbonutslippskvoter og prisen på naturgass. Etter å ha skalert disse variablene til sammenlignbare størrelser har vi lagt de inn i våre to modeller. Antagelsen og målet var at det kunstige nevralt nettverket skulle klare å finne sammenhenger som den linære modellen ikke finner, og med det kunne skape bedre og mer presise prediksjoner.

Konklusjonen ble litt annerledes enn forventet, og vi ser at LSTM modellen predikerer bedre enn regresjonsmodellen på treningsdataene. Vi ser derimot at resultatene er litt mer uklare når det kommer til testdataene. MSE og MAPE favoriserer forskjellige modeller, og ved visuell inspeksjon ser vi at ingen av modellene kan predikere spesielt godt. Dette tyder på at begge modellene predikerer dårlig på virkelig data.

Abstract

The rising power prices have received a lot of attention recently, and many have expressed their dissatisfaction and unrest related to the high electricity prices. Being able to find models that predict prices will provide an increased degree of control and help to make the power market more predictable. There are many ways one can go to develop good explanations for the movements in the power market, but we have chosen to look at two quite different models to work towards finding a trend, which may contribute to further research.

The issue we landed on is "*Can an artificial neural network predict historical electricity spot prices better than a linear model?*". The problem specifies that it is about the difference between the two models, although of course it is also interesting to look at the results in isolation. Using this problem, we selected data based on theory about the power market and its drivers. Here we have both the relatively obvious direct factors such as: production, consumption, temperature, wind power and the amount of water in the water reservoirs. In addition to that we have several slightly less obvious and indirect factors such as: exchange rate, coal price, carbon emission quotas and the price of natural gas. After scaling these variables to comparable sizes, we have included them in our two models. The assumption and goal were that the artificial neural network should be able to find connections that the linear model does not find, and with that could create better and more precise predictions.

The conclusion was slightly different than expected, and we see that the LSTM model predict better than the regression model on the training data. On the other hand, we see that the results are a little more unclear when it comes to the test data. MSE and MAPE favor different models, and on visual inspection we see that none of the models can predict particularly well. This suggests that both models predict poorly on real data.

Innholdsfortegnelse

Forord	1
Sammendrag	2
Abstract	3
Innholdsfortegnelse	4
Figurliste	6
Tabelliste	7
Ordforklaringer	8
1. Innledning	9
2. Teori	11
2.1 Kraftbalansen	11
2.1.1 Forholdet mellom import og eksport	12
2.2 Norges kraftflyt	13
2.3 Tilbud og etterspørsel	14
2.4 Vær som påvirkningsfaktor	15
2.5 Vann i magasinene har mye å si	15
2.6 Produksjon som påvirkningsfaktor	15
2.7 CO ₂ , naturgass og kull som påvirkningsfaktor	16
2.8 Valuta som påvirkningsfaktor	16
3. Metode	18
3.1 Datakilder	18
3.1.1 Avhengig variabel	18
3.1.2 Uavhengige variabler	19
3.2 Modellene	32
3.2.1 Lineær regresjon	32
3.2.2 Long Short-Term Memory (LSTM)	33
3.2.3 Skalering	35
3.2.4 Trening og testset	35
3.2.5 Sammenlikne modellene	35
4. Resultater og diskusjon	37
4.1 Lineær regresjon	37
4.2 LSTM-modell	39
4.3 Sammenligning av modellene	41
5. Feilkilder	42
5.1 Feil i dataene	42
5.2 Inntastingsfeil	42

5.3 Overtilpassing	42
5.4 Forutsetninger for linær regresjon	43
5.5 Utelatte variabler	44
5.6 For kort tid	44
5.7 Visuell inspeksjon – tolkning av data	44
5.8 Politikk	45
6. Konklusjon	46
7. Referanser	48

Figurliste

Figur 1 Kraftbalansen (Energidepartementet, 2022)	11
Figur 2 Import og eksport (Pindyck & Rubinfeld, 2013)	12
Figur 3 Kraftflyt (Nord Pool AS, 2022)	13
Figur 4 Tilbud og Etterspørsel (Pindyck & Rubinfeld, 2013).....	14
Figur 5: Elspot price Oslo.....	18
Figur 6 Production NO1.....	19
Figur 7 Production NO2.....	19
Figur 8 Production NO3.....	19
Figur 9 Production NO5.....	20
Figur 10 Production SE3.....	20
Figur 11 Production DK1.....	20
Figur 12 Consumption NO1.....	21
Figur 13 Consumption NO2.....	21
Figur 14 Consumption NO3.....	21
Figur 15 Consumption NO5.....	21
Figur 16 Consumption DK1.....	22
Figur 17 Consumption SE3.....	22
Figur 18 Hydro reservoir NO.....	23
Figur 19 Hydro reservoir SE.....	23
Figur 20 Windpower DK1.....	24
Figur 21 Windpower DK2.....	24
Figur 22 Carbon Emission Future.....	25
Figur 23 Natrual Gas.....	26
Figur 24 Coal Price.....	27
Figur 25 Eur/Nok.....	28
Figur 26 Sek/Nok.....	28
Figur 27 Temperature Norway.....	30
Figur 28 Deep neural network.....	34
Figur 29 LSTM-node.....	34
Figur 30 Prediksjoner treningsdata lineær model.....	38
Figur 31 Prediksjoner testdata lineær model.....	39
Figur 32 Oversikt over LSTM modellen.....	39
Figur 33 Prediksjoner treningsdata LSTMmodell.....	40
Figur 34 Prediksjoner testdata LSTMmodell.....	41

Tabelliste

Tabell 1 Elspot price oslo.....	18
Tabell 2 Production	20
Tabell 3 Consumption.....	22
Tabell 4 Hydro reservoir	23
Tabell 5 Windpower	24
Tabell 6 Carbon Emission Future.....	25
Tabell 7 Natrual Gas	26
Tabell 8 Coal Price	27
Tabell 9 Valuta.....	29
Tabell 10 Month	29
Tabell 11 Temperature Norway	29
Tabell 12 Korrelasjon.....	31
Tabell 13 Koeffesienter fra lineær regresjonsmodel.....	37
Tabell 14 MSE og MAPE.....	41

Ordforklaringer

MSE – Mean Squared Error

MAPE – Mean Absolute Percentage Error

ANN – Artificial neural network

LSTM – Long-short term memory

1. Innledning

Vinterhalvåret 2021-2022 har sett de høyeste strømprisene i Norge noensinne og vi har sett kraftpriser som kryper seg helt opp på 4,5 kroner per kilowatttime, der en varm dusj etter skituren kan ende med å koste deg over 30 kroner (Nord Pool AS, 2022). Prisene har skapt størst problemer for de med trang økonomi, og vi har hørt om studenter som fryser og barnefamilier som sliter med å få økonomien til å gå rundt (Bugge, et al., 2022). De høye strømprisene har gjort folk mer oppmerksomme og interesserte i å finne ut hva det er som gjør at prisene kan bli så høye. Av diverse eksperter, media og strømselskaper har det vært god informasjon om mange faktorer som påvirker kraftprisen. Faktorene som forbruk, nedbørmengde, temperatur, produksjon og kullpris er blant de som påvirker tilbudet og etterspørselen (Norges Energi, 2022).

Å vite hvilke faktorer som er med på å påvirke strømprisene er et godt utgangspunkt for å få en forståelse over kraftmarkedet. Dette er noe vi skal se nærmere på i denne oppgaven. Som vi har sett varierer også prisene veldig av hvor man befinner seg i Norge. Høye priser settes der det er energiknapphet, da produsentene regner med å få solgt det til en bedre pris senere. Dette fører igjen også til lavere produksjon i disse områdene.

Men enda viktigere og mer interessant er nok sammenhengen mellom alle faktorene, og en dypere forståelse av disse vil være verdifull. Isolert sett er det ikke så komplisert å forstå at det er en korrelasjon mellom temperatur og strømpris. Kaldere vær gjør at man bruker mer strøm på oppvarming, som øker forbruket. Kraftprisen er sammensatt av en rekke av disse lineære sammenhengene, og i utgangspunktet tror vi at en linær modell vil være i stand til å predikere elspot prisene med relativt høy treffsikkerhet. På den annen side vil det alltid være andre sammenhenger som strekker seg utenfor de lineære, og det er disse sammenhengene vi tror vi kan finne ved hjelp av et kunstig nevralt nettverk.

Vi har derfor valgt å dykke dypere inn i datamaterialet, og ønsker å finne ut av om et nevralt nettverk kan være i stand til å finne mønstre og sammenhenger på tvers av de lineære sammenhengene. Vi har derfor valgt problemstillingen:

«Kan et kunstig nevralt nettverk predikere historiske elspot priser bedre enn en lineære modell?»

Formålet med å finne ut av dette vil være at man kan finne ut hva som er driverne i kraftmarkedet. Dette kan for eksempel være at man oppdager at en av variablene har større

ringvirkning på de andre variablene, og derfor i større grad påvirker strømprisen. Det er viktig for samfunnet å ha kontroll på faktorene, og en større kontroll gir bedre muligheter for å styre markedet i den retningen man ønsker. Samtidig vil det også være verdifullt gjennom at man i større grad kan vite hva prisen blir om verdensbildet endrer seg.

Vi har avgrenset oppgaven gjennom å bruke historiske data fra og med 2015 til og med 2021. Vi ser for oss at en mulig svakhet kan være at det enten er variabler vi ikke har hensyntatt som påvirket prisen den perioden. Videre selges også kraften over landegrensene, og det kan hende at diverse politiske beslutninger som ikke kan fanges opp gjennom vår modell har påvirket i perioden. Derfor har vi også avgrenset oppgaven til i større grad å gjelde forklaringskraften til de to modellene i forhold til hverandre, og ikke i like stor grad basere suksessen i prosjektet på hvor treffsikre modellene er i prediksjonen av day-ahead- priser. Når det er sagt tror vi at datamengden i forhold til tidsperioden, og mengden variabler skal gi et godt bilde over de mest fundamentale faktorene.

Tidligere forskning har oppnådd gode resultater med å bruke LSTM modeller for å predikere Nord Pool elspot pris. Blant annet har Li og Becker brukt LSTM modeller og variabelseleksjons algoritmer for å predikere elspotprisen (Li & Becker, 2021). Her fant man blant annet ut at en to-steps LSTM modell ga de beste prediksjonene og at variabelseleksjon var essensielt for å kunne gi gode prediksjoner. Dataene som ble brukt på modellene bestod i hovedsak av produksjon, konsum, prognoser for konsum og produksjon, valuta, kraftflyt og kraftflytavvik. Det kan derfor være interessant å undersøke om man kan predikere strøm ved å bruke litt annerledes data som f.eks. kull til å predikere strømprisen.

Videre i oppgaven vil vi ta for oss noe av den mest kjente teorien som bygger opp om det valgte tallmaterialet. Ut ifra den teorien vil vi videre gå inn på data- og metode-delen av oppgaven, som beskriver metoden vi har brukt, presenterer dataene og forklarer de to modellene. Etter denne delen presenteres og analyseres funnene, før vi kommer med en konklusjon som drar frem de relevante forholdene i oppgaven, og helst bidrar til videre forskning.

2. Teori

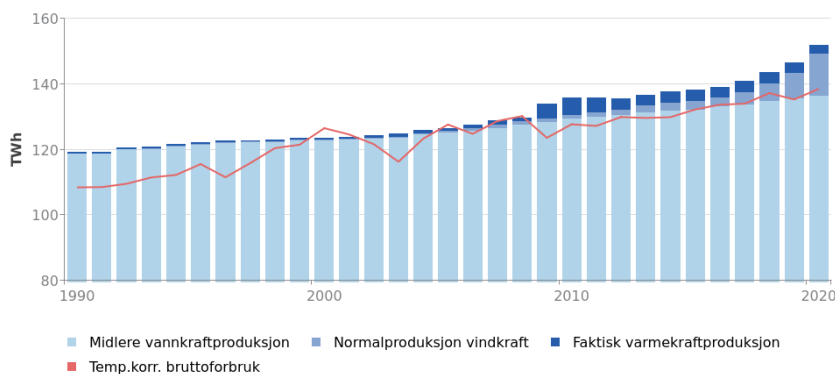
Siden vår problemstilling her er å finne ut hvilken av de to modellene som predikerer de historiske kraftprisene mest presist, er det viktig at datamaterialet gir en så nøyaktig representasjon av kraftmarkedet som mulig. Det er bred kunnskap og opplysning rundt hvilke faktorer som påvirker prisene, og de som tas opp oftest er blant annet: Nedbør, vann i vannmagasinene og Co2 priser. Våre valg av variabler har vært som følger;

Produksjon av kraft (N01, 2,3 og 5, DK1 og SE3), mengde vann i vannmagasiner i både Norge og Sverige, forbruk (N01-NO5, DK1 og SE3), vindstyrke i DK1 og DK2, prisene på; Co2, naturgass, kull, og valuta mot Nok i både SEK og EUR, og til slutt et snitt av temperaturen i Norge.

2.1 Kraftbalansen

Kraftbalansen i Norge omfatter forholdet mellom forbruk og produksjon av kraft (Energidepartementet, 2022). Kraftbalansen sier også om man i enkelt år har eksportert eller importert kraft fra det Norske kraftsystemet. I Norge er vi avhengig av nedbør til vannkraftverkene og vind til vindmøllene for å produsere kraft. Hvis man ikke klarer å produsere mer kraft enn man forbruker i enkelte år, vil dette kunne gi en negativ kraftbalanse, og man må ty til import av kraft.

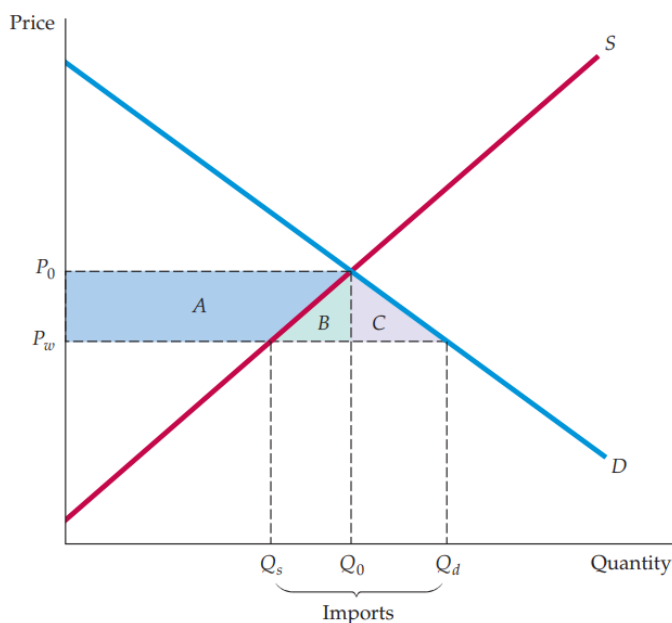
På bildet nedenfor, hentet fra Energidepartementet ser vi at produksjonen av kraft i Norge har økt betraktelig de siste 10 årene. Vi ser også at bruttoforbruket av kraft har gått opp, men på langt nær i like stor grad som produksjonen. Vi kan med dette konkludere med at Norge siden 2009 har hatt en jevn økning på den positive siden av kraftbalansen.



Figur 1 Kraftbalansen (Energidepartementet, 2022)

2.1.1 Forholdet mellom import og eksport

Import er forskjellen mellom innenlands forbruk og innenlands produksjon. Uten kvoter eller tariffen vil et land importere varer når importprisen er lavere enn innlandsprisen (Pindyck & Rubinfeld, 2013). Hvis det ikke er noe import vil verdensprisen legge seg i markedskrysset mellom tilbud og etterspørsel. Om man derimot velger å importere, vil prisen i hjemmemarkedet senkes ned på nivå med verdensprisen, som vil senke den innenlandske produksjonen ned til Q_s , og øke det innenlandske forbruket til Q_d .



Figur 2 Import og eksport (Pindyck & Rubinfeld, 2013)

Med bakgrunn i denne teorien kan man si at Norge, med sin positive kraftbalanse ikke har hatt behov for å importere strøm de siste årene (Energidepartementet, 2022). På den annen side er bruttoforbruket av strøm temperaturkorrigert, som vil si at det allikevel kan komme tider på året der magasinene nærmest er tomme, og temperaturen er lav, som gjør at man må se seg nødt til å importere kraft fra Europa.

2.2 Norges kraftflyt



Figur 3 Kraftflyt (Nord Pool AS, 2022)

På bildet over ser vi kapasiteten på ledningene som sender Norsk kraft både innenlands og utlands, kraften er målt i Mega-watt (Mw) (Nord Pool AS, 2022).

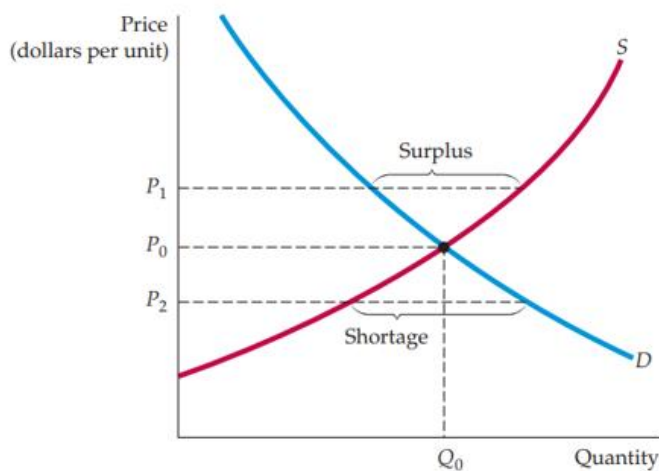
Det kan ikke sendes like mye strøm ut av de nordre delene av Norge, og dermed får de et produksjonsoverskudd om gjør at de kan selge til en lavere pris i disse regionene. Fra Østlandet er det mange ledninger, med mye kapasitet som kan føre kraften videre ut av Norge til resten av Europa. Som Nordpool selv skriver på sine sider er kapasiteten tilgjengelig til alle på samme vilkår (Nord Pool AS, 2022). Dette skaper en økt etterspørsel, som igjen øker prisen på kraften i disse regionene. Eksempelvis ser vi på Nordpool at systemprisen prisen i Trondheim per 10.03.2022 ligger på 139 NOK, sammenlignet med nivået til Oslo på 2 089 NOK.

2.3 Tilbud og etterspørsel

Tilbud og etterspørsel påvirker markedsmekanismene, en påvirkning vi ser gjennom en økning/reduksjon i priser på varer eller tjenester, og tilsvarende i kvantum (Pindyck & Rubinfeld, 2013, p. 22).

Tilbudskurven sier at jo høyere pris tilbyderne tar, jo høyere kvantum har de råd til å tilby. Etterspørselskurven på sin side sier at jo lavere priser varene blir tilbudt til, jo høyere kvantum vil selges fordi forbrukernes kjøpekraft styrkes.

Den optimale pris og mengde blir bestemt av krysningen mellom de to kurvene (tilbudskurven (S) og etterspørselskurven (D)) (Pindyck & Rubinfeld, 2013). Ved et enkelt eksempel kan vi se at dersom prisene økes til P_1 , vil færre ha råd til produktet, og man blir nødt til å senke prisen. Senker man den ned til P_0 vil det forbrukerne er villige til å betale stemme med det tilbyderne har å tilby, og vi vil derfor finne en optimal pris og kvantum.



Figur 4 Tilbud og Etterspørsel (Pindyck & Rubinfeld, 2013)

De fleste av variablene i modellen vår påvirker tilbud og etterspørsel. Dette vil si at de påvirker hvor mye kraft som tilbys på markedet, og hvor mange som etterspør denne kraften. På vinterstid, da temperaturene i Europa er lavere, er det naturlig å tro at tilbyderne kraftprodusentene, må være mer selektive med hvem de selger sin kraft til for å ikke havne i en situasjon der de har svært lite vann i magasinene. Videre, selger de da kraften til de som byr høyest, og ofte kan dette være områder utenfor Norge. For mannen i gata blir det ikke noe varmere i leiligheten, og man har ikke noe annet valg enn å betale for strømmen selv om den er dyr.

2.4 Vær som påvirkningsfaktor

Været påvirker kraftprisen fordi perioder med mer nedbør fyller opp vannmagasinene, også kjent som tilsig, som øker tilbudet og gir lavere priser. Mer spesifikt kan vi si at store nedbørsmengder resulterer i at vannkraftverkene kan bli tvunget til å drive såkalt tvungen kjøring. Dette innebærer at kraftverkene grunnet mengdene med vann må kvitte seg med noe, og heller velger å produsere strøm til en billig penge, enn å la vannet renne forbi (NVE, 2021). I tillegg vil også temperaturen påvirke strømforbruket, da lavere temperaturer vil føre til at folk bruker mer strøm til oppvarming. Den økte etterspørselen vil føre med seg en økning i strømprisene.

For å fange opp hvordan været påvirker har vi valgt å se på snittemperaturen i Norge fra 2015-2021. Man kunne også ha delt opp året i fire sesonger, men en ukentlig snittemperatur tror vi vil gi mer presise resultater.

2.5 Vann i magasinene har mye å si

Hvor mye vann som er i vannmagasinene, har også veldig mye påvirkning på kraftprisen (Energidepartementet, 2022). Vi har nettopp gått inn på hva som skjer ved høye nedbørsmengder, der kraftverkene blir nødt til å produsere billig strøm for å ikke kaste bort kraft. Ser man til motsatte scenario, at nedbørsmengdene er lave, vil dette føre til at produsentene heller velger å holde på vannet til etterspørselen øker. Dette kan for eksempel være i en kaldere periode, der folk er avhengig av oppvarming og derfor er villige til å betale mer for strømmen.

Dette er i stor grad det vi har sett i Norge den siste tiden, der lave nedbørsmengder kombinert med lave temperaturer har tvunget opp kraftprisen til høyeste nivå noensinne.

2.6 Produksjon som påvirkningsfaktor

Hvor mye kraft som produseres vil også naturlig nok påvirke prisene, da økt produksjon gir høyere tilbud og lavere pris for forbrukerne (Energidepartementet, 2022). Denne faktoren har en tett tilknytning til nedbør og mengden vann i magasinene, da det vil produseres mer ved mye tilsig og vann i magasinene, og lite i perioder med lavere nedbørsmengder. Samtidig vil en oversikt over hvor mye som produseres kunne si noe om sesongvariasjonene, og indirekte

kunne fange opp andre faktorer utenom nevnte. Dette kan være midlertidige politiske faktorer, etterspørselen i utlandet osv.

Vi valgte å se på produksjonen i NO1,2,3 og 5, samt DK1 og SE3, og får på denne måten en bredere oversikt over kraftproduksjonen enn kun innenfor Norge, som vil kunne fange opp faktorer i tilbud og etterspørsel utenfor Norges landegrenser. Dette bidrar til et mer helhetlig bilde, da kraftmarkedet er sammensatt og har mange tilbydere, etterspørrere og krafttyper.

2.7 CO₂, naturgass og kull som påvirkningsfaktor

Selv om 90-% av normalproduksjonen i Norge kommer fra vannkraft, er saken en helt annen i resten av Europa (Energidepartementet, 2022). At strømprisene i Norge går opp kan ofte ha en sterk sammenheng med at prisene på alternative energikilder som naturgass og kull har gått opp. Konsekvensen av de økte prisene på alternative energikilder er at de som har gjort seg avhengige av dette ser etter andre alternativer. Dette kan føre til økt etterspørsel av Norske kraft. Denne økte etterspørselen fører igjen til at prisen på norsk kraft øker. Co₂ kvoter påvirker også strømprisen ved at høyere pris på kvotene fører til at det blir dyrere ved å produsere energikilder som slipper ut Co₂. Som sagt fører høyre pris på disse energikildene til høyre etterspørsel av strøm og dermed høyre strømpriser. Dette er logikken bak at co₂ kvoter og energikilder som naturgass og kull påvirker norske strømpriser.

2.8 Valuta som påvirkningsfaktor

For at verdier skal bli overført fra et land til et annet må de bli konvertert fra opprinnelseslandet valuta, til mottakerlandets. Skal man selge varer fra Norge til Tyskland, vil den Tyske parten som handler i euro måtte betale for verdien av NOK i forhold til Euroen. Er den Norske kronen dyr i forhold til euroen, må den tyske parten betale mer i euro for samme andel NOK (Mishkin & Eakins, 2018, p. 44). På den andre siden, hvis den norske kronen er billig i forhold til Euroen er det mer attraktiv å importere fra Norge.

En sentral teori om valuta er at identiske goder skal koste det samme i kjøpekraft uansett hvilken valuta du bruker. Dette konseptet heter «*Law of one price*» og betyr at om transportkostnadene og handelsbarrierene er lave, så vil en homogen vare handles for samme

pris. Samtidig er noen valutaer mer verdifulle enn andre. Det gjør at prisen for en identisk vare i en valuta kan være dyrere eller billigere enn en annen valuta. Forskjellene i pris mellom valutaene kompenserer for den ulike kjøpekraften valutaene representerer.

Konseptet heter kjøpekraftsparitet. Det betyr at valutakurser bestemmes ut ifra prisnivåene mellom land. (Mishkin & Eakins, 2018) Hvis for eksempel et lands prisnivå øker i forhold til et annet vil valutaen deres synke. Det andre landets valuta vil på den andre siden stige. Dette er en viktig dynamikk i hvordan valutakurser bestemmes.

En annen faktor som påvirker valutakursen, er produktivitet (Mishkin & Eakins, 2018). Grunnen til dette er at med høy produktivitet faller ofte prisnivået. Ut ifra kjøpekraftsparitet vil da valutaen styrke seg i forhold til andre valutaer. Handelsbarrierer er også en faktor som påvirker. Ved økt handelsbarrierer pleier valutaen til det landet å styrke seg. En annen påvirkningsfaktor er preferanser for et lands produkt. Om noen prefererer et lands produkt fremfor et annet vil det føre til at valutaen til det foretrekkende produktet styrkes. Det konkurrerende landet vil få en svekket valuta. På denne måten er produktpreferanser med på å styre valutakursen.

Med bakgrunn for denne teorien har vi utformet vår metode som vi skal bruke videre i arbeidet med å utforske hvilke av modellene som vil gi de beste resultatene. Tallmaterialet vi skal bruke for å finne ut om den lineære metoden, eller det nevralt nettverket er mest hensiktsmessig å bruke vil være inspirert av teorien over.

3. Metode

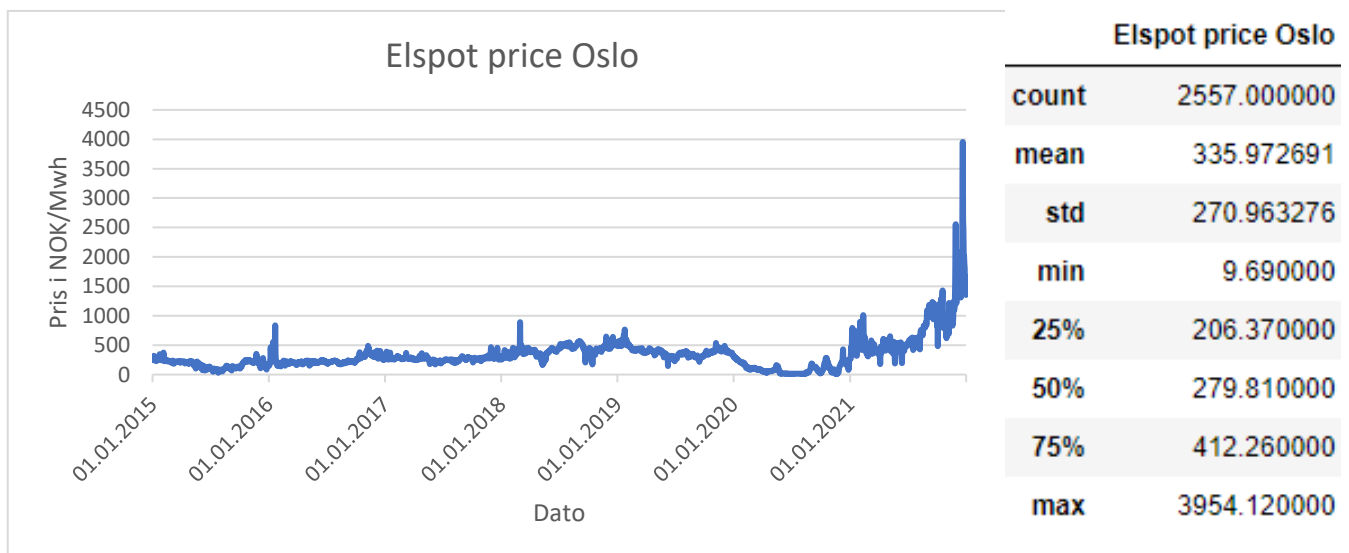
I dette kapittelet skal vi presentere vår metode for å løse vår problemstilling. Vi skal også presenter hvilke data og modeller vi har brukt. Alle tidsseriene vi har brukt i våre modeller starter 01.01.2015 og slutter 31.12.2021.

3.1 Datakilder

3.1.1 Avhengig variabel

Vår avhengig variabel er Elspot prisen for Oslo (NO1). Dette er prisen for hvor mye 1 megawatt time koster. Denne prisen blir satt av Nord Pools kraftbørs med utgangspunkt i tilbudet og etterspørselen i markedet (Nord Pool AS, 2022). Prisen settes 1 dag før den blir levert. Prisen settes også for hver time av døgnet, men i vår analyse har vi brukt den daglige prisen. Denne prisen er regnet ut ved å ta et aritmetisk gjennomsnitt av spotprisen for alle timene av dagen. Vi har hentet denne dataen fra Nord Pools Database.

Tabell 1 Elspot price oslo



Figur 5: Elspot price Oslo

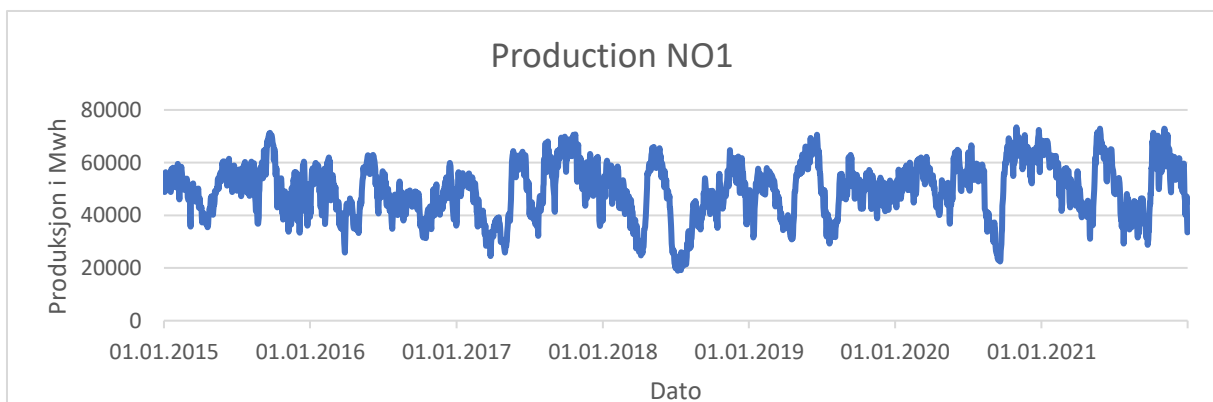
Som vi ser av grafen av spotprisen i Oslo har prisen holdt seg relativt stabil frem til vinteren i 2021. Da ble prisen historisk dyr. Grafen kan også tyde på at det er sesongvariasjoner, ved at prisen generelt sett er billigere om sommeren enn på vinteren.

3.1.2 Uavhengige variabler

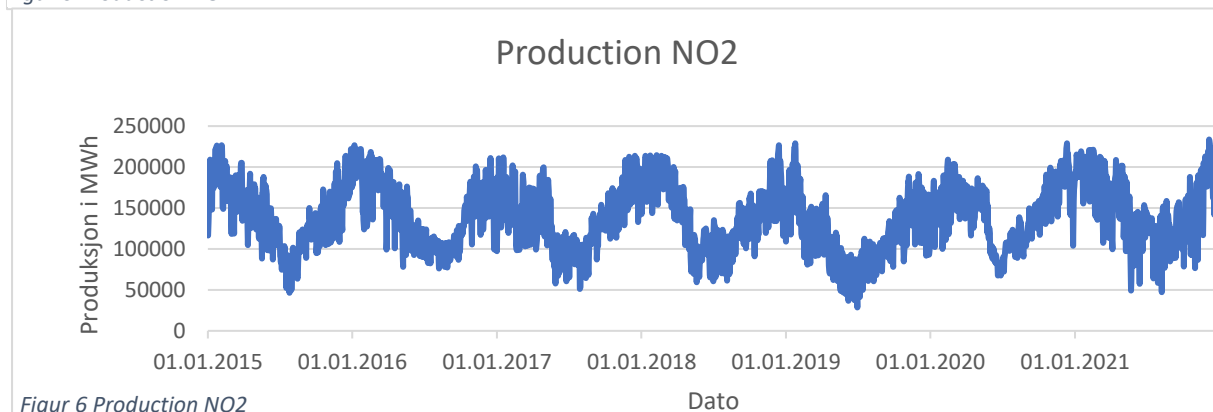
3.1.2.1 Produksjon

Våre første uavhengige variabler er produksjonen for regionene N01, N02, N03, NO5, DK1, SE3. Siden produksjon er en sentral faktor for tilbudet av strøm, er den derfor veldig relevant for å sette spotprisen i kraftmarkedet. Vi ønsker å predikere prisen i Oslo, dvs. i NO1. Det er derfor ganske sannsynlig at produksjonen her påvirker prisen. Grunne til at vi har tatt med produksjon fra de andre områdene er at de er koblet til NO1 området med strømkabel (Nord Pool AS, 2022). Det er derfor trolig at disse områdene påvirker hverandres priser.

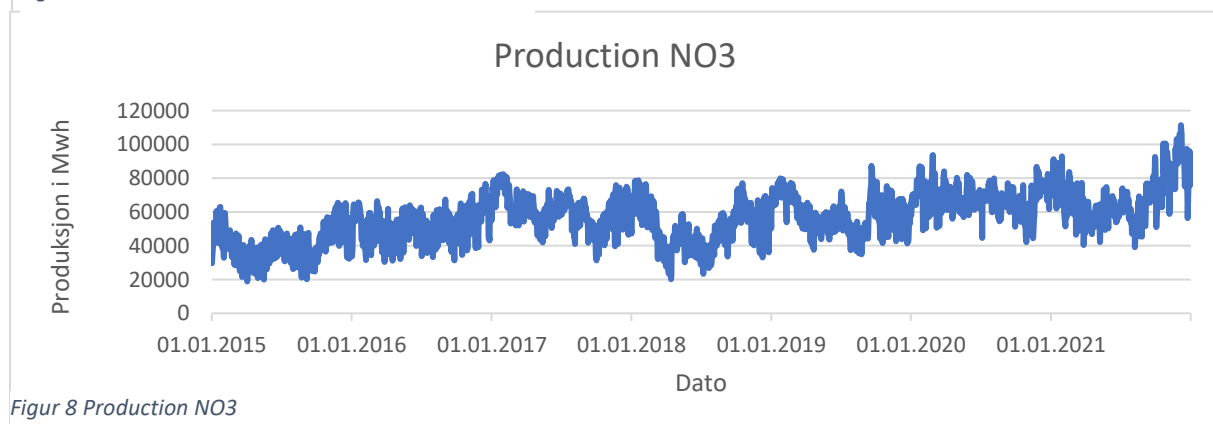
Produksjonen måles i Megawatt timer. Vi har funnet denne dataen fra Nord Pools database.



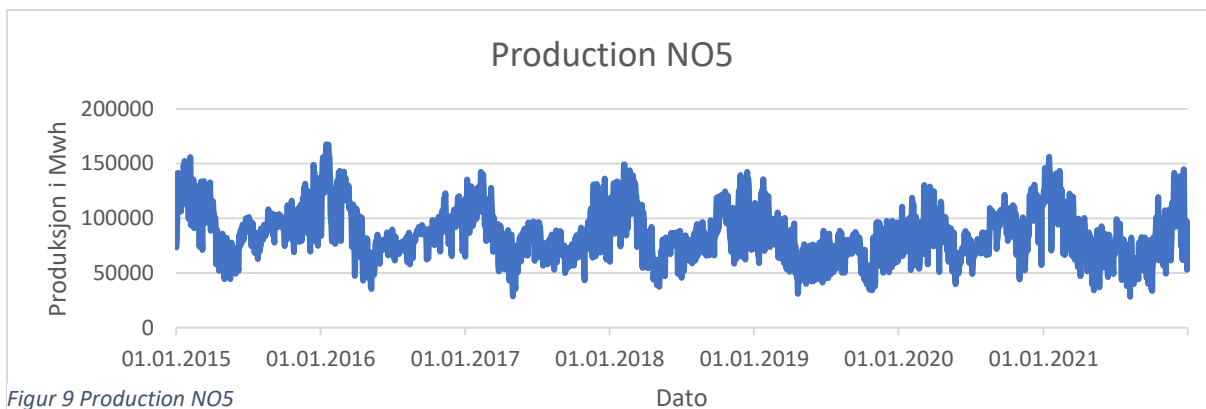
Figur 6 Production NO1



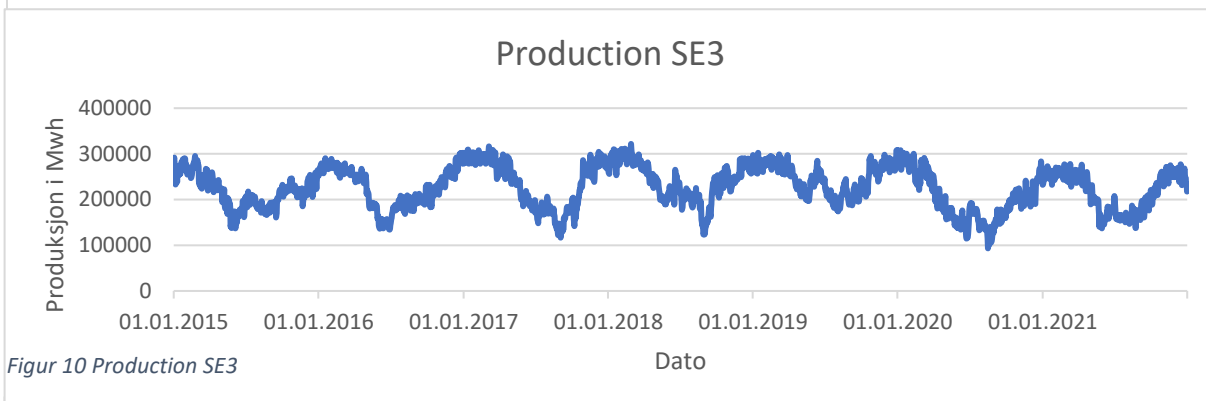
Figur 6 Production NO2



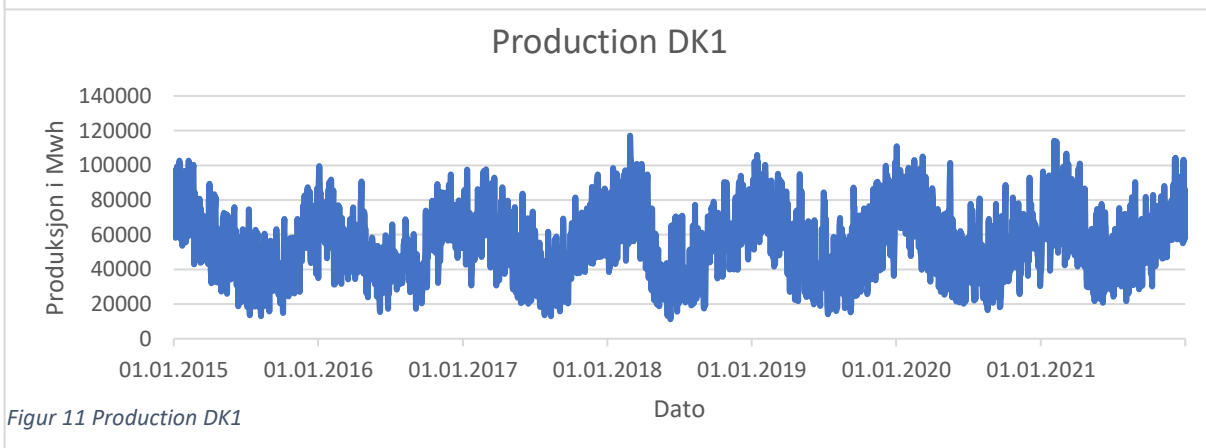
Figur 8 Production NO3



Figur 9 Production NO5



Figur 10 Production SE3



Figur 11 Production DK1

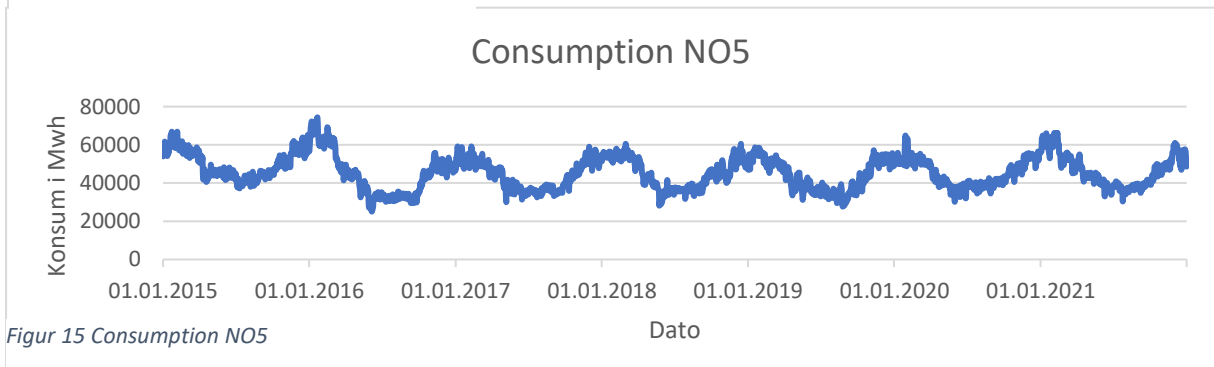
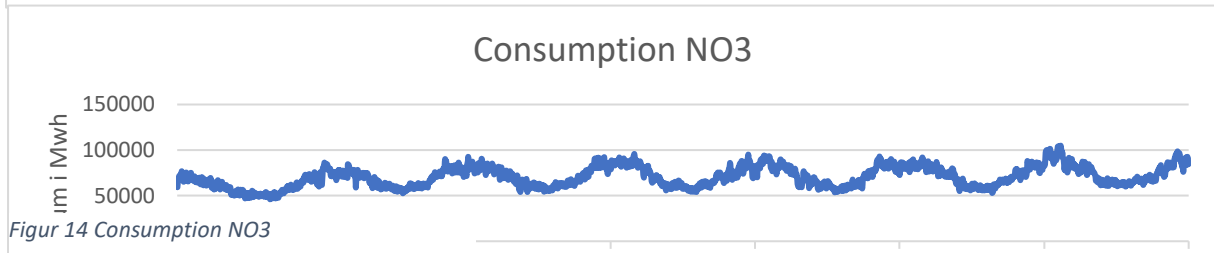
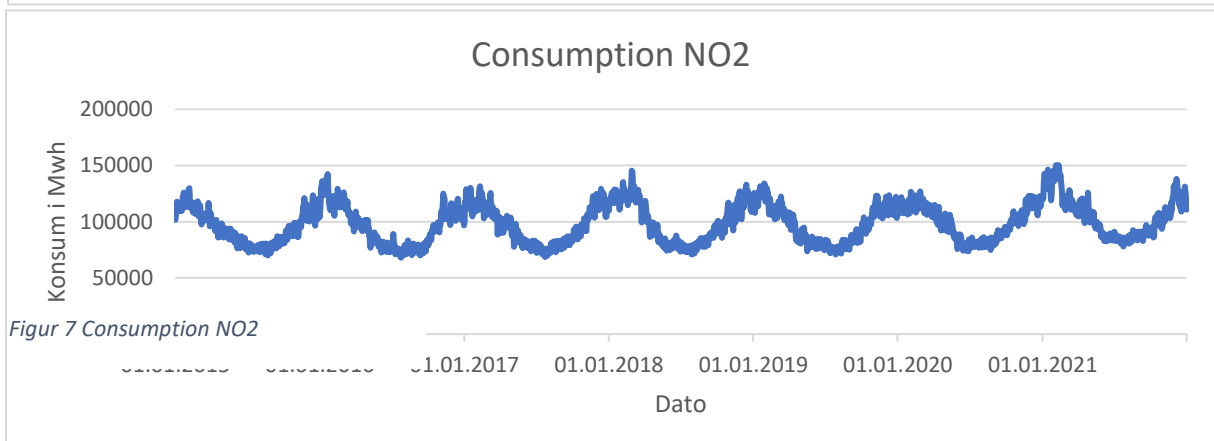
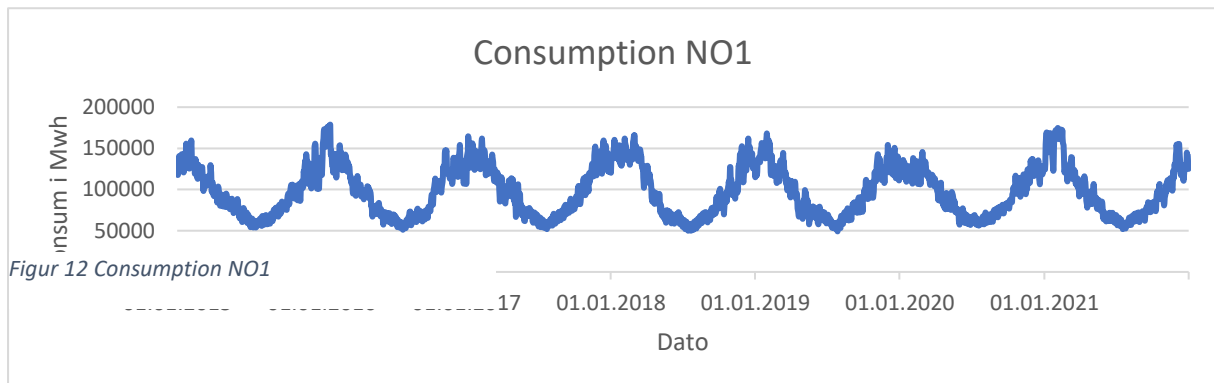
Den følgende tabell 2 viser antall enheter, gjennomsnitt, standardavvik, minimum og maximums verdi og 25%, 50% og 75% persentil av produksjonsvariablene for NO1, NO2, NO3, NO5, DK1 og SE3

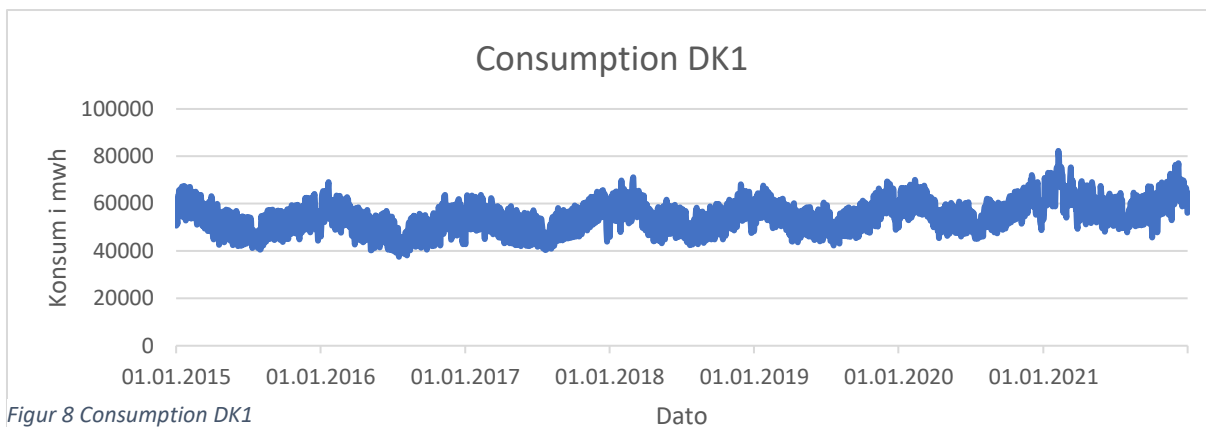
	Production NO1	Production NO2	Production NO3	Production NO5	Production DK1	Production SE3
count	2557.000000	2557.000000	2557.000000	2557.000000	2557.000000	2557.000000
mean	49853.748925	140440.204928	57148.885022	86414.492374	54312.879155	225755.309347
std	10382.849367	39886.816329	14205.468646	23549.667730	19825.434903	45707.663265
min	18967.000000	28577.000000	18881.000000	28335.000000	11216.000000	93226.000000
25%	42796.000000	110054.000000	46999.000000	70897.000000	38783.000000	189822.000000
50%	50544.000000	138480.000000	57784.000000	83216.000000	53882.000000	229751.000000
75%	57545.000000	171253.000000	66765.000000	100844.000000	68353.000000	264386.000000
max	73395.000000	233629.000000	111350.000000	167624.000000	117100.000000	321768.000000

Tabell 2 Production

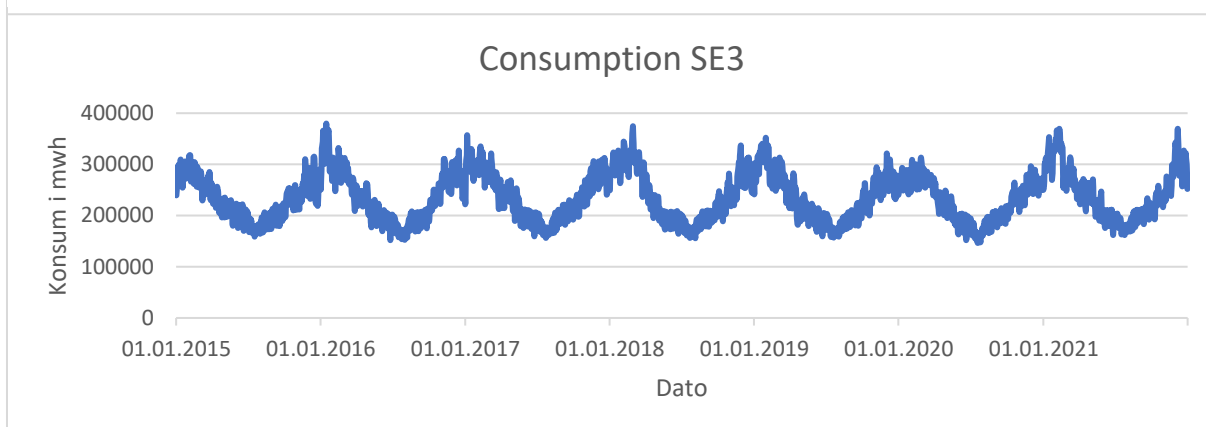
3.1.2.2 Konsum

Vi har også valgt å bruke konsumet i N01, N02, N03, NO5, DK1, SE3 som uavhengige variabler. Som sagt så er disse regionene koblet sammen og det er naturlig å tro at konsumet i de ulike regionene påvirker hverandre. Konsumet er kan på mange måter regnes som etterspørselen av elektrisitet. Konsumet måles også i Megawatt time og vi har funnet dataen fra Nord Pool.





Figur 8 Consumption DK1



Figur 9 Consumption SE3

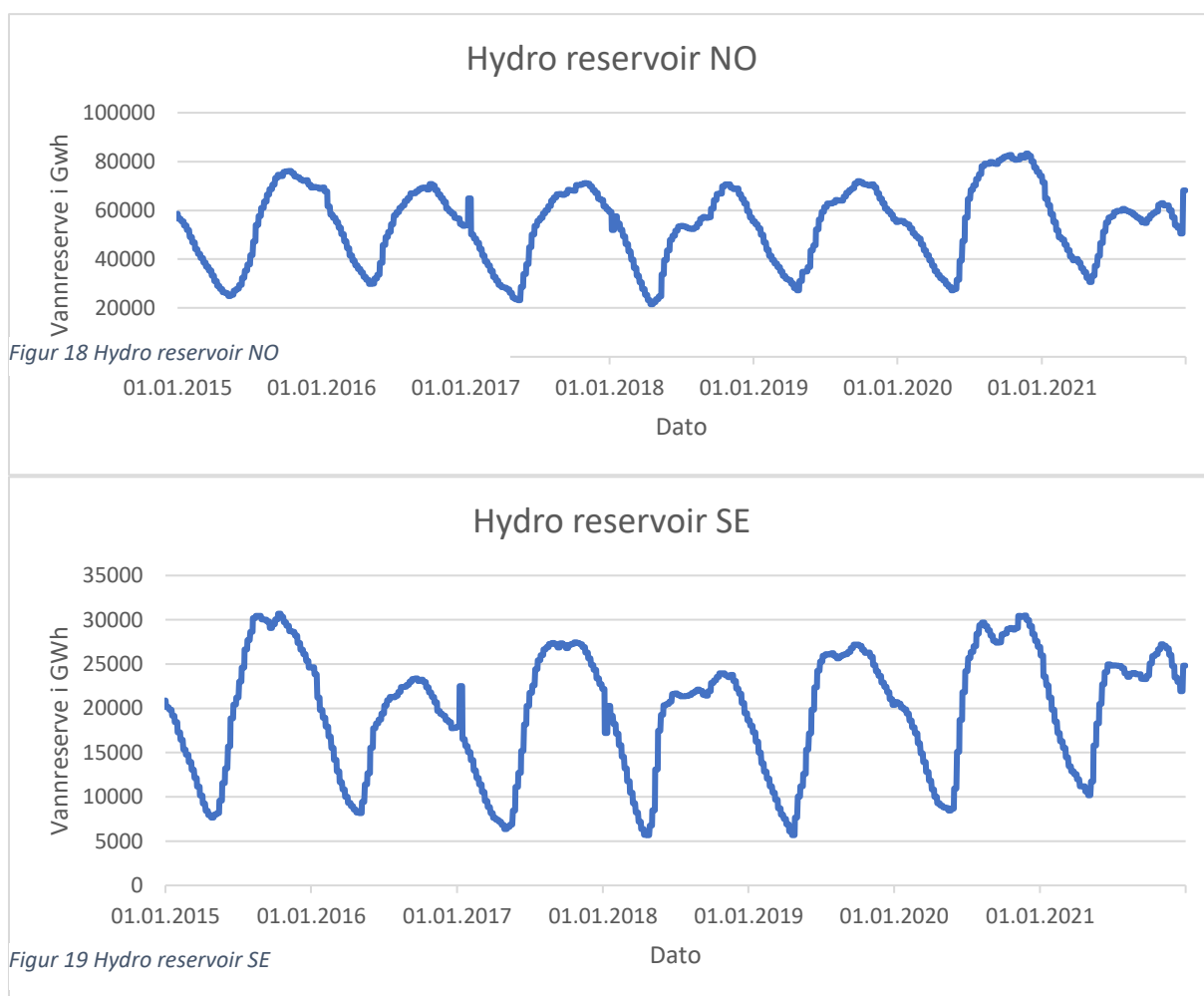
Som vi ser, er konsumet relativt stabilt i perioden 2015 til 2021 i alle regionene. Vi ser også at det er syklisk tendens hvor konsumet øker om vinteren. Dette sammenfaller med produksjonen i regionen fordi produksjonen også ser ut som om er størst på vinteren. Den følgende tabell 3 viser antall enheter, gjennomsnitt, standardavvik, minimum og maximums verdi og 25%, 50% og 75% persentil av konsumvariablene for NO1, NO2, NO3, NO5, DK1 og SE3.

	Consumption NO1	Consumption NO2	Consumption NO3	Consumption NO5	Consumption DK1	Consumption SE3
count	2557.00000	2557.00000	2557.00000	2557.00000	2557.00000	2557.00000
mean	98501.32499	98644.916308	70733.776691	45529.28158	55819.862339	237238.197888
std	29949.83284	17121.254683	11210.978067	8713.79677	7012.979740	46554.089865
min	49218.00000	68002.000000	46015.000000	25055.00000	37558.000000	146463.000000
25%	71472.00000	83268.000000	61499.000000	38166.00000	51000.000000	200861.000000
50%	94797.00000	97296.000000	69837.000000	45106.00000	56035.000000	232417.000000
75%	123146.00000	112412.000000	79851.000000	52110.00000	60808.000000	272795.000000
max	178839.00000	150254.000000	104848.000000	74379.00000	82290.000000	379797.000000

Tabell 3 Consumption

3.1.2.3 Vannmagasiner

Vi har også tatt med energien som er lagret i vannmagasinene i Norge og Sverige. En av de få måtene man kan lagre energi på er i vannmagasiner og om man opplever en høyere etterspørsel kan reservene i vannmagasinene dekke dette behovet. Vi har også her funnet dataen fra Nordpools databaser. Denne dataen blir oppdatert på ukentlig basis og måles i Gigawatt time.



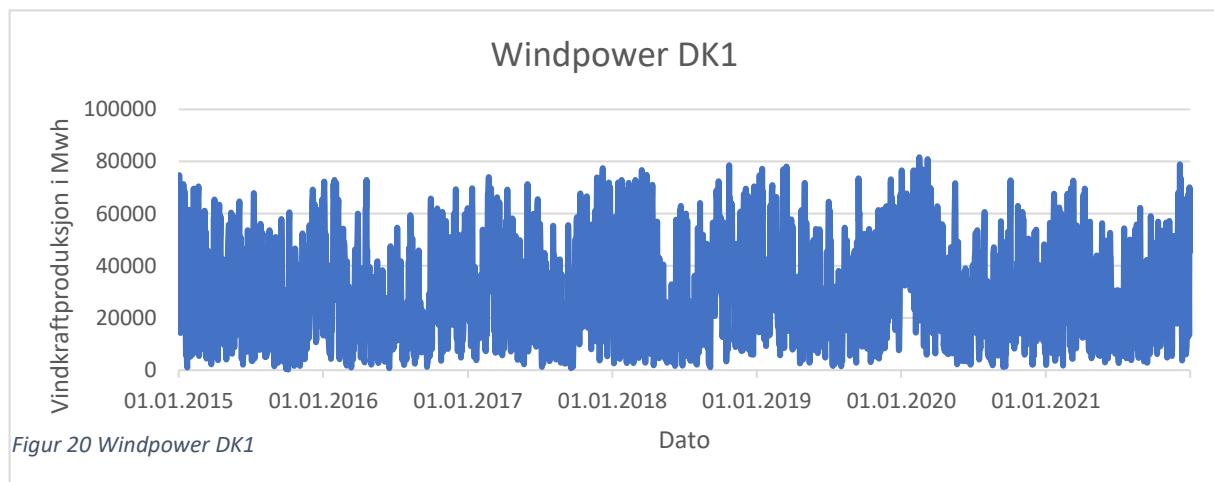
Vi ser at det er klare sesongvariasjoner i hvor mye energi som er lagret i vannmagasinene i både Sverige og Norge. På den andre siden ser det ut som at dette mønsteret er ganske stabilt og at det er liten forskjell mellom topper og bunner.

	Hydro reservoir NO	Hydro reservoir SE
count	2557.000000	2557.000000
mean	53799.226828	19752.901838
std	15589.883594	6868.643130
min	21549.000000	5672.000000
25%	40048.000000	13946.000000
50%	56328.000000	21301.000000
75%	66678.000000	25651.000000
max	83280.000000	30694.000000

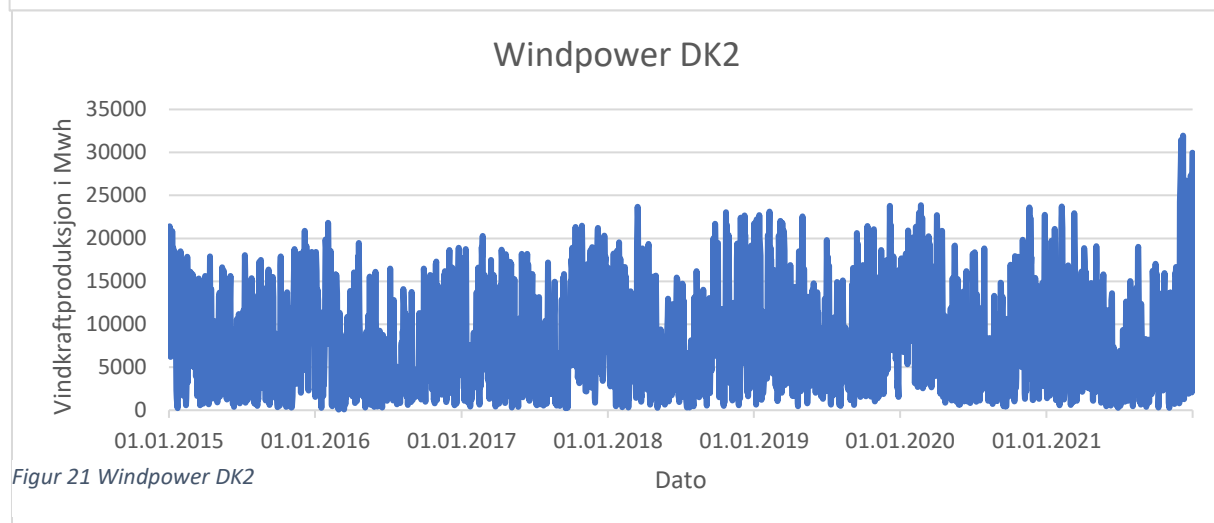
Tabell 4 Hydro reservoir

3.1.2.4 Vindkraftproduksjon

Vindkraft er en viktig energikilde og det er særlig i Europa man produserer strøm ut fra vind. Derfor er det sentralt å ta med en variabel som omhandler vindkraft. Siden Danmark er en stor produsent av vindkraft og er koblet til NO1 med kraftkabel, bruker vi vindkraft fra DK1 og DK2 som uavhengige variabler for vindkraft. Dataene våre vises i Megawatt time og de er målt daglig. Dette er også data som er innhentet fra Nord Pool.



Figur 20 Windpower DK1



Figur 21 Windpower DK2

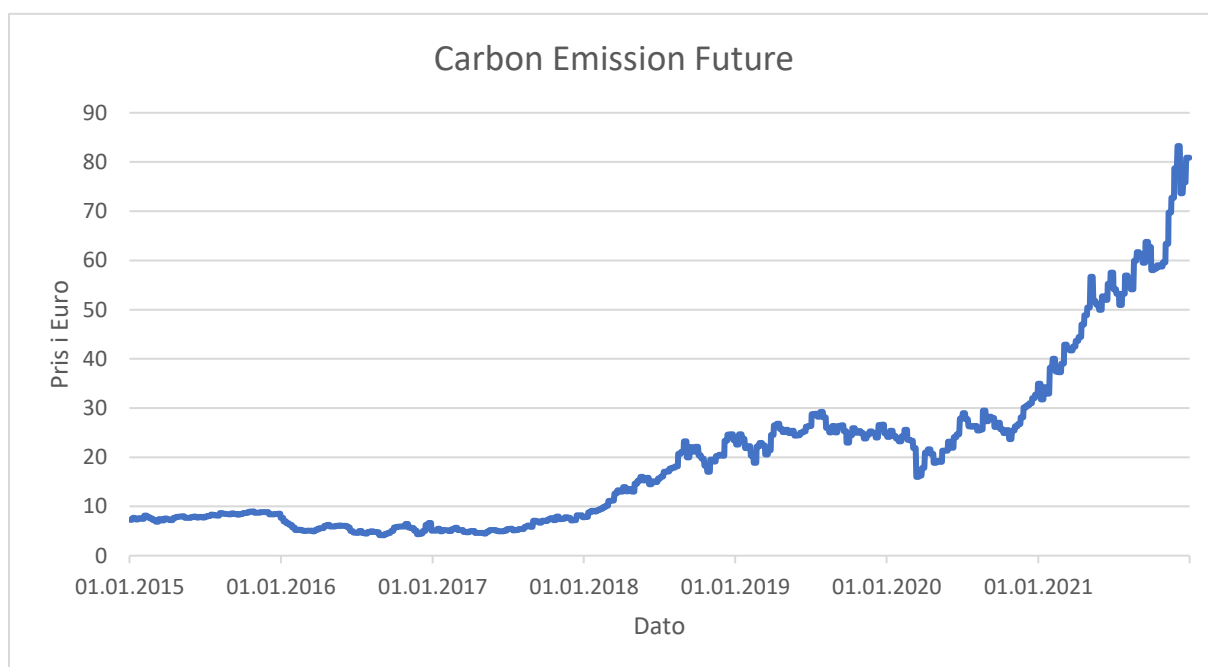
Ut ifra tidsseriene ser man at kraftproduksjonen fra vind er veldig volatil og kan forandre seg mye på kort tid. Man ser også at det ikke er like tydelige sesongvariasjoner som man har sett i de andre variablene.

	Windpower DK1	Windpower DK2
count	2557.000000	2557.000000
mean	29045.491983	7988.395776
std	18828.503938	5871.711878
min	220.000000	77.000000
25%	12830.000000	3031.000000
50%	26202.000000	6727.000000
75%	42292.000000	12146.000000
max	81642.000000	31954.000000

Tabell 5 Windpower

3.1.2.5 Carbon Emission Future

En annen faktor som påvirker strømprisen, er prisen på CO₂-kvoter. CO₂ kvoter handes med futureskontrakter og vår data viser prisen på kontrakten til nærmeste levering. Når en kontrakt går ut, viser dataene vår prisen på neste kontrakt. CO₂ kvotene handles i 1000 Euro og kontraktstørrelsen er på 1000 tonn. Siden futuresmarkedet ikke er åpent alle dager slik som kraftmarkedet, har vi valgt å bruke ukentlige data for at man enklere skal kunne sammenligne prisen i strømmarkedet kontra prisen på CO₂ kvotene. Vi har hentet vår data om CO₂ kvoter fra Investing.com (Investing.com, 2022).



Figur 10 Carbon Emission Future

Ut ifra tidsserien ser man at prisen på CO₂-kvoter har hatt en sterkt stigende trend og har steget ekstremt mye i 2021. Prisen på CO₂ kvoter har gått fra under 10 000 euro per 1000 tonn i 2015 til over 80 000 euro per 1000 tonn på det høyeste i 2021.

Carbon Emission Future	
count	2557.00000
mean	19.89745
std	16.75528
min	4.15000
25%	7.10000
50%	16.00000
75%	25.60000
max	83.24000

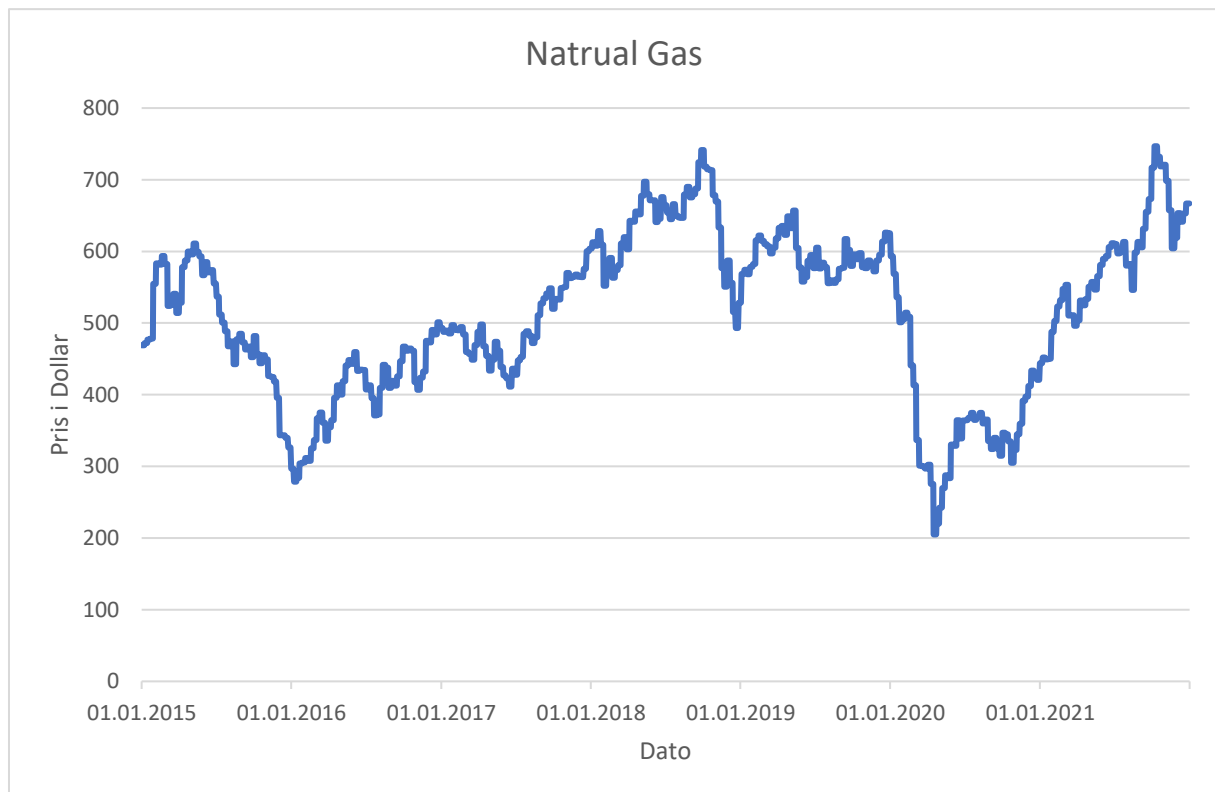
Tabell 6 Carbon Emission Future

3.1.2.6 Pris på naturgass

Prisen på naturgass viser seg også å påvirke strømprisen. Vi har derfor tatt med data fra fremtidskontraktene for naturgass som handles i London (Investing.com, 2022). Også her viser dataen vår prisen på den fremtidskontrakten som er nærmest til levering. Naturgass handles i 100 dollar og størrelsen på hver kontrakt er 100 tonn. Vi har også her brukt ukentlige priser for at man skal kunne sammenligne med kraftprisene. Dette er fordi kraftmarkedet alltid er åpent, mens markedet for naturgass er stengt ved helligdager og helger.

Natrual Gas	
count	2557.000000
mean	511.911126
std	112.382638
min	205.000000
25%	434.500000
50%	525.750000
75%	595.750000
max	746.750000

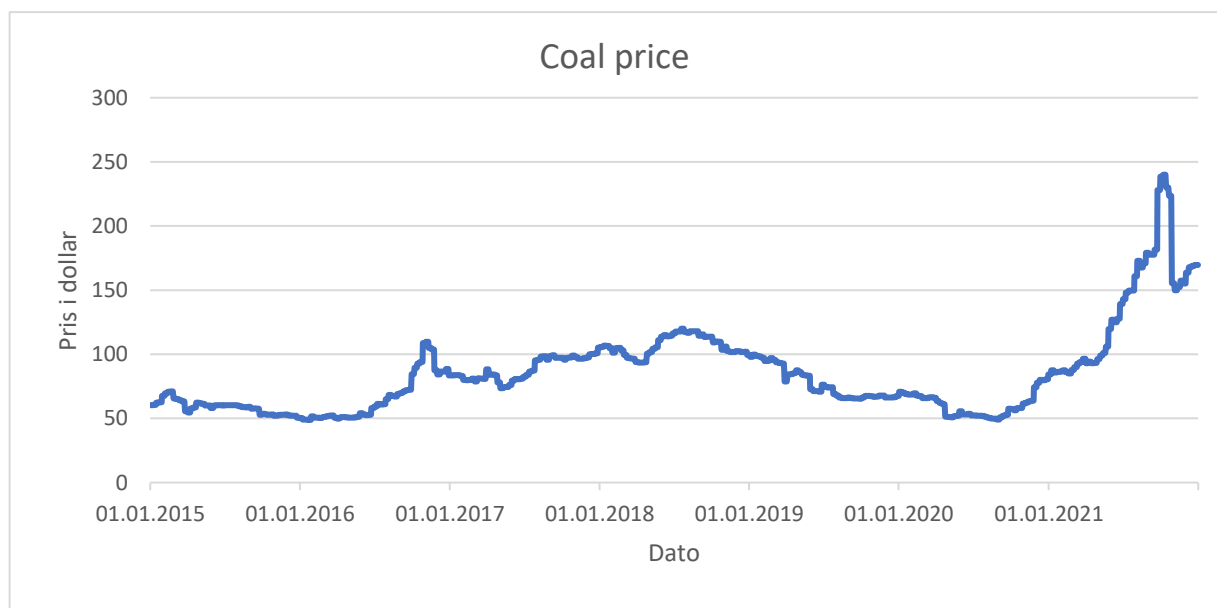
Tabell 7 Natrual Gas



Figur 11 Natrual Gas

3.1.2.7 Kullpris

En annen energikilde som påvirker strømprisen, er kull. Vi har derfor også inkludert kullprisen som en uavhengig variabel. Vi har brukt Newcastle Coal Futures for å gi prisen på kull. Også her er det snakk om rullerende kontrakter, slik som for gas og CO2-kvoter (Investing.com, 2022). Med andre ord ser man på prisen til den kontrakten som er nærmest forfall. Newcastle Coal Futures handles i dollar og man kjøper 1000 tonn kull ved kjøp av en slik kontrakt. Vi har også her brukt ukentlige priser for at man skal kunne sammenligne med kraftprisene.



Figur 12 Coal Price

Av grafene av tidsserien av kullkontrakter ser man at prisen har steget veldig mye 2021. Det kan der for virke som at prisen på kull muligens følger samme trend som prisen på naturgass og CO2 kvoter.

Coal price	
count	2557.000000
mean	85.141865
std	33.248447
min	48.800000
25%	60.400000
50%	80.000000
75%	98.600000
max	240.000000

Tabell 8 Coal Price

3.1.2.8 Valuta

Eksport og import er påvirket av valutakurser. Siden det er mulighet for import og eksport av elektrisitet, er det derfor naturlig å tenke at valutasingninger også påvirker prisen på strøm. Vi har derfor tatt med valutapar som NOK/SEK og NOK/EURO i vår modell. Dette er i hovedsak de valutaen vi eksporterer og importerer fra. Vi eksporterer og importerer også fra Danmark, men siden den danske kronen holdes stabil til Euroen så tenker vi at valutaparet NOK/DKK ikke er nødvendig å ta med (Opstad, 2020). Vi har også her brukt ukentlige priser for at man skal sammenligne med elspotprisen.



Figur 13 Eur/Nok



Figur 14 Sek/Nok

Av tidseriene ser man at Euroen stort sett har styrket seg siden 2015. Man hadde en topp hvor Euroen var på det høyeste i 2020, men har siden stabilisert seg mot normale nivåer. Vi ser også at den Svenske kronen har holdt seg mer stabil mot den Norske kronen, også dette valutaparet hadde en topp i 2020.

	SEK/NOK	EUR/NOK
count	2557.000000	2557.000000
mean	0.971194	9.706742
std	0.042209	0.623957
min	0.892500	8.371300
25%	0.932300	9.311300
50%	0.973500	9.634000
75%	1.002500	10.046800
max	1.132200	12.571600

Tabell 9 Valuta

3.1.2.9 Temperatur og måned

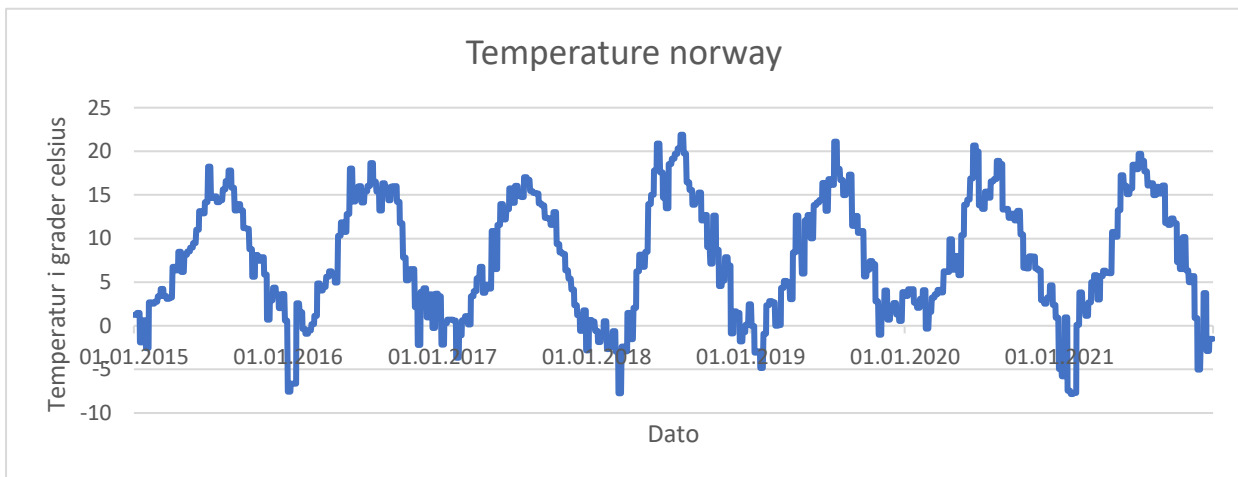
Som vi har forklart tidligere er tilbud og etterspørslene av strøm påvirket av både årstid og temperatur. Derfor tenkte vi at det ville være relevant å ta med variabler som måler dette. Vi har derfor tatt med en variabel måned som er kodet slik at verdien 0 tilsvarer januar, 1 tilsvarer februar osv. Vi har også tatt med en variabel som går på temperatur. Denne variabelen har verdier som er et gjennomsnitt av temperaturen i Norge. Disse dataene er hentet fra Eikons database.

	Month
count	2557.000000
mean	5.522487
std	3.449499
min	0.000000
25%	3.000000
50%	6.000000
75%	9.000000
max	11.000000

Tabell 10 Month

	Temperature norway
count	2557.000000
mean	7.641398
std	6.793032
min	-7.824048
25%	2.460628
50%	6.517821
75%	13.880893
max	21.900613

Tabell 11 Temperature Norway



Figur 15 Temperature Norway

3.1.2.10 Korrelasjon

For å se på sammenhengen mellom den avhengige og de uavhengige variablene kan man regne ut korrelasjonen mellom variablene. På denne måten får man vite om det er en lineær sammenheng eller ikke (Oppen, et al., 2020). Vi har i vår oppgave brukt korrelasjonskoeffisienten Persons r. Denne analysen forutsetter kontinuerlige variabler, men kan også brukes på variabler på ordinalnivå om man har tilstrekkelig med alternativer. Av våre data er de fleste variablene kontinuerlige og de som er på ordinalnivå har mange alternativer. Derfor er Persons R relevant for vår data.

$$\rho_{X,Y} = \frac{COV(X,Y)}{\sigma_X \sigma_Y}$$

Persons r beregnes ved bruk av formelen ovenfor. Man tar kovariansen til X og Y, og deler på standardavviket til X multiplisert med standardavviket til Y.

Generelt sett kan man si at om den absolutte korrelasjonen er over 0,5, så har man en sterk lineær sammenheng (Oppen, et al., 2020). Er absoluttverdien av korrelasjonen mellom 0,3 og 0,5 er det en moderat samvariasjon. Er absoluttverdien lavere enn 0,3 er det en svak samvariasjon. Fortegnet på korrelasjons koeffisienten avgjør om det er en positiv eller negativ sammenheng mellom variablene.

Variabelnavn	Korrelasjonskoeffisient
Elspot price Oslo	<i>1.000000</i>
Coal price	<i>0.771529</i>
Carbon Emission Future	<i>0.682995</i>
Natrual Gas	<i>0.549960</i>
Consumption DK1	<i>0.337181</i>
Consumption NO3	<i>0.335859</i>
Production NO3	<i>0.329388</i>
Consumption NO2	<i>0.250870</i>
Consumption SE3	<i>0.235180</i>
Production NO2	<i>0.183899</i>
Month	<i>0.176742</i>
Consumption NO1	<i>0.173012</i>
Production SE3	<i>0.149968</i>
Production DK1	<i>0.130910</i>
Consumption NO5	<i>0.126791</i>
Production NO5	<i>0.068354</i>
EUR/NOK	<i>0.067895</i>
Hydro reservoir SE	<i>0.063134</i>
Windpower DK2	<i>0.030999</i>
Hydro reservoir NO	<i>-0.009205</i>
Production NO1	<i>-0.025226</i>
Windpower DK1	<i>-0.029537</i>
SEK/NOK	<i>-0.124057</i>
Temperature Norway	<i>-0.165859</i>

Tabell 12 Korrelasjon

Av korrelasjonskoeffisientene ser man at kull har sterkest samvariasjon med elspotprisen. Denne samvariasjonen er positiv noe som betyr at en høyere kullpris varierer sammen med høyere strømpris. Vi ser også at variablene som karbon futures og naturgass har en sterk samvariasjon med elspotprisen i NO1. De andre variablene har middels til svak lineær samvariasjon.

3.2 Modellene

3.2.1 Lineær regresjon

Den første modellen vi skal bruke for å løse vår problemstilling er en lineær multivariat regresjonsmodell. Dette er en av de mest populære modellene og er både relevant for deskriptive analyser, men også prediktive analyser. En lineær regresjonsmodell tar utgangspunkt i uavhengige variabler som multipliseres med en koeffisient for å predikere en verdi på den avhengige variabelen (Oppen, et al., 2020). Man pleier også ha et konstantledd i en slik modell som adderes med summen av koeffisienten som er multiplisert med en uavhengig variabel. Modellen kan også vises matematisk som:

$$y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \dots + \beta_k * x_k + e$$

Her er y den avhengige variabelen, β_0 er konstanten og e er feilleddet. I tillegg har man β_k som er et uttrykk for koeffisienten og x_k som uttrykker de uavhengige variablene.

For å finne de koeffisienten og konstanten som passer best til den avhengige variabelen, bruker man ofte minste kvadratsums metode (Oppen, et al., 2020). Minste kvadratsums metode er en metode som går ut på at man finner den linjen som har den minste gjennomsnittlige kvadrerte avstanden mellom observert verdi og predikert verdi. Dette kalles for Mean Squared Error (MSE). Den matematiske formelen man prøver å minimere er derfor:

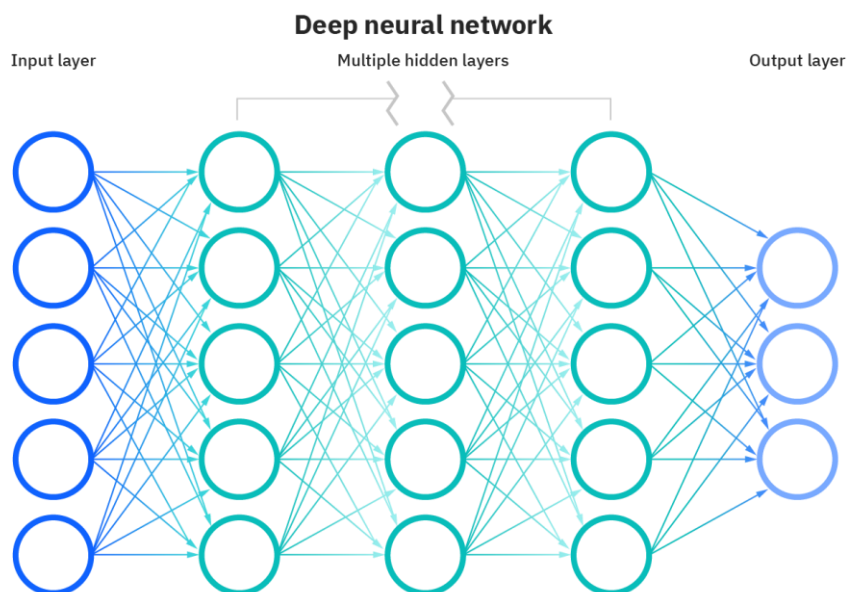
$$MSE = \frac{1}{N} \sum_{i=1}^n (Y_i - \hat{y}_i)^2$$

En lineær regresjonsmodell bygger på en rekke forutsetninger (Oppen, et al., 2020). Den første forutsetningen er at det er en lineær sammenheng mellom de uavhengige variablene og den avhengige variabelen. Den andre forutsetningen er at alle residualene har en forventning lik 0. Residualene er feilleddene. Man forutsetter også fravær av perfekt multikollinearitet. Multikollinearitet betyr at de uavhengige variablene korrelerer med hverandre og ved høy innbyrdes korrelasjon blir koeffisientene ustabile og man kan oppleve høye p-verdier. Man forutsetter også at restleddene er ukorrelerte med hverandre. Det betyr at man ikke skal ha faktorer som korrelerer med de uavhengige eller avhengige variablene og som ikke er inkludert i modellen. Modellen forutsetter også at restleddet har konstant varians og at de er normalfordelte.

Hvis man følger disse forutsetningene, får man en robust og valid estimeringsmetode. Dette gjelder særlig ved om man ønsker en deskriptiv modell. Når det gjelder predikative modeller er det hvor bra prediksjonen til modellen er som er det mest viktige. Derfor er ikke det å tilfredsstille alle forutsetningene like viktig som i en deskriptiv modell. Man kan derfor si at ved prediksjon så er det viktigere at modellen predikerer bra enn at alle forutsetningene er oppfylt.

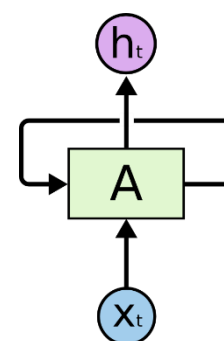
3.2.2 Long Short-Term Memory (LSTM)

Long Short-Term Memory er en modell som brukes innen maskinlæring og bygger på strukturen som er i nevralt nettverk. Et nevralt nettverk er et sett med noder som er koblet sammen og er inspirert av nervevevet i hjernen (IBM Cloud Education, 2020). Man starter med en input layer, deretter blir dataen prosessert gjennom et eller flere hidden layers og til slutt får man den endelige dataen i et output layer. Ved en slik struktur ønsker man at modellen skal gjenkjenne mønster og tolke data slik en menneskehjerne ville gjort. I tillegg til at nevralt nettverk er bygget opp av noder i ulike layers er de også koblet sammen med bindinger som har vekter og aktiveringsfunksjon. En vekt er en parameter som er knyttet til koblingen mellom to noder. Vekten bestemmer hvor mye den ene noden påvirker den andre. Aktiveringsfunksjonen bestemmer hvordan input dataen transformeres til nodens outputdata. Man har flere ulike typer aktiveringsfunksjoner som f.eks. logaritmisk funksjoner og hyperbolske funksjoner. I nevralt nettverk har nodene også en bias. Dette er et konstantledd som blir lagt til vekten. Ved maskinlæring er det vektene og biasene som blir endret på ved trening til å gjenkjenne mønstre. I tillegg til dette har man også en tapsfunksjon. Dette er en funksjon som man ønsker å minimere ved å endre på vektene og biasene i de ulike nodene. I vår LSTM modell bruker vi i likhet med den lineære modellen MSE som tapsfunksjon.



Figur 16 Struktur til et Deep neural network (IBM Cloud Education, 2020)

I en vanlig modell som bruker nevralt nettverk går informasjonen i en retning fra input til output. Det som skiller LSTM modellen fra en slik modell er at nodene også er koblet sammen av løkker slik at informasjonen også bli behandlet tilbakevendende (Olah, 2015). På denne måten kan modellen oppdage mønstre basert på tidligere informasjon. Dette greier ikke vanlige nevralt nettverks modeller. Derfor er LSTM modellen å foretrekke når man ønsker å gjøre prediksjon basert på tidligere informasjon.



Figur 17 LSTM-node (Olah, 2015)

Siden LSTM modellen tar inn tidlige informasjon til å gjøre prediksjoner er strukturen på informasjonen som tas inn i modellen forskjellig fra strukturen i et vanlig nevralt nettverk modell. I et vanlig nevralt nettverk modell er strukturen på informasjonen todimensjonal. Strukturen på inputinformasjonen i en LSTM modell er derimot tredimensjonal og kan representeres i en matrise. Grunnen til at inputinformasjonen er tredimensjonal er fordi man ønsker å ta med tidligere data i tillegg til gjeldende data for å trene opp imot den avhengige variabelen (Olah, 2015). I vår LSTM modell gjør vi prediksjoner ut ifra de forrige 10 dagene. På denne måten gjør datastrukturen det mulig for modellen å oppdage mønstre basert på tidligere data.

3.2.3 Skalering

Ved maskinl ring er det ofte slik at modellen fungerer bedre om dataene er skalert ned til mindre tall (Shmueli, et al., 2020). Treningsprosessen g r ogs  ofte raskere om man har skalert dataene. P  bakgrunn av dette har vi valgt   skalere dataene i LSTM modellen. Ofte blir verdien til variablene skalert ned til et intervall mellom 0 og 1 eller fra -1 til 1. Vi har derfor i v r oppgave valgt   skalere dataene til intervallet mellom 0 og 1. For   presentere resultatene p  en forst kelig m te skalerer vi ogs  dataene tilbake til opprinnelig st rrelse etter at modellen har blitt trent og predikert str mpriser.

3.2.4 Trening og testset

For   lage gode modeller m  man dele opp den dataen man har i et treningsset og et testset. (Shmueli, et al., 2020) Treningssettet er den dataen man bruker for   trene opp modellen. Testsettet er den dataen man bruker for   teste modellen og se hvor god den er til   predikere. Grunnen til at man g r dette er fordi at man pr ver   unng  at modellen blir overtilpasset til dataene v re. Da vil modellen vise veldig gode resultater p  v rt datagrunnlag, men vil ikke gi gode resultater i virkeligheten. I v r oppgave bruker vi dataen fra 01.01.2015 til 31.12.2020 som v rt treningsset. Dataen fra 01.01.2021 til 31.12.2021 er v rt testset.

3.2.5 Sammenlikne modellene

For   sammenlikne resultatene fra modellene v re skal vi bruke Mean Squared Error (MSE) og Mean Absolute Percentage Error (MAPE) i tillegg til visuell inspeksjon. MSE og MAPE er to estimatorer for hvor gode modellene er i forhold til den avhengige variabelen. Disse kriteriene m le hvor god en modell er relativt til en annen og er derfor egent til   sammenligne modeller (Shmueli, et al., 2020). MSE ble beskrevet i avsnittet om line r regresjon. Mean Absolute Percentage Error (MAPE) gir et m l p  den gjennomsnittlige prosentvise feilen mellom predikert og observert verdi. Dette regnes ut ved   bruke denne formelen:

$$MAPE = \frac{100\%}{n} * \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

Her representerer A_t den faktiske verdien og F_t den predikerte verdien. Absoluttverdien av $\frac{A_t - F_t}{A_t}$ blir summert sammen for alle prediksjoner og multipliserer man med $\frac{100\%}{n}$ får man MAPE.

4. Resultater og diskusjon

I følgende kapittel presenterer vi resultatene fra våre modeller og sammenligner dem. Deretter blir resultatene og modellen drøftet og diskutert. Vi begynner med å presentere resultatene for modellen med lineær regresjon og deretter tar vi for oss modellen som bruker long short-term memory.

4.1 Lineær regresjon

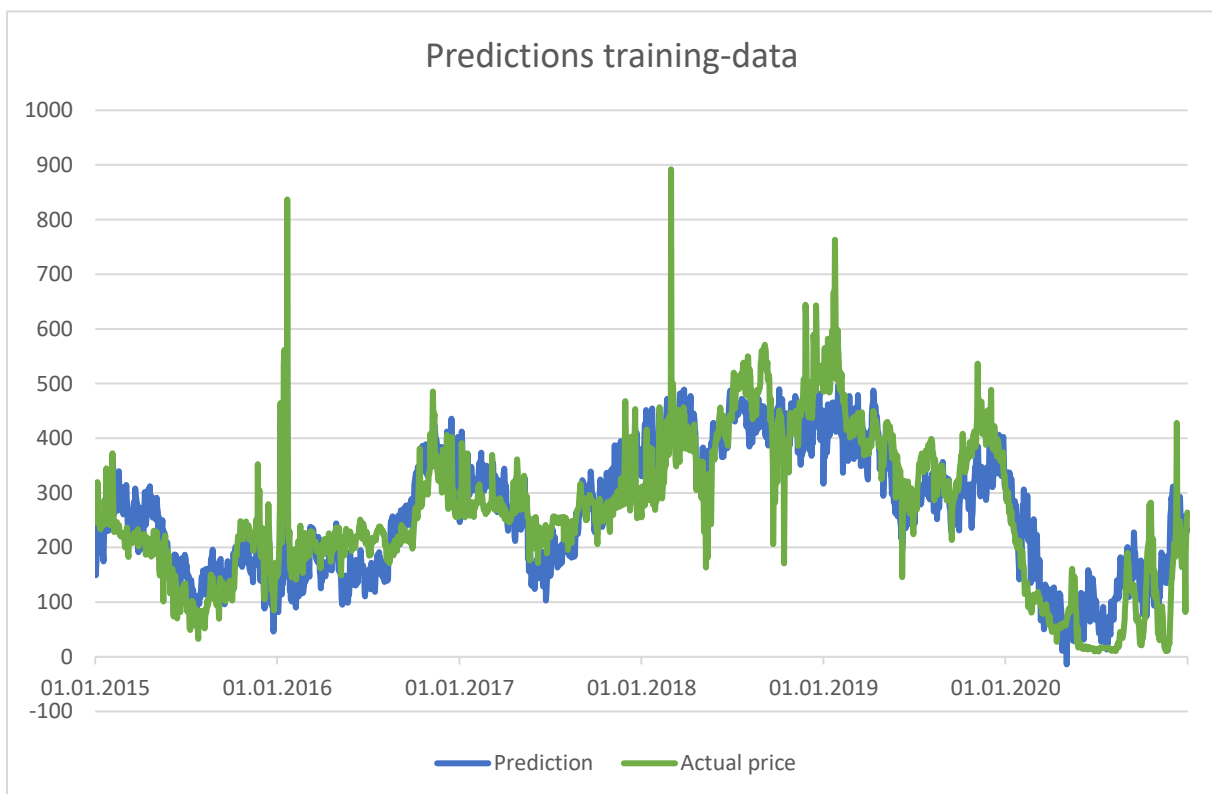
Ut ifra vår lineære regresjon og bruk av minste kvadraters metode får vi en modell med disse koeffisientene:

	coef	std err	t	P> t	[0.025	0.975]
Production NO1	-0.0020	0.000	-12.521	0.000	-0.002	-0.002
Production NO2	-0.0003	8.96e-05	-3.198	0.001	-0.000	-0.000
Production NO3	-0.0014	0.000	-8.151	0.000	-0.002	-0.001
Production NO5	0.0003	0.000	2.766	0.006	9.45e-05	0.001
Production DK1	0.0010	0.000	3.626	0.000	0.000	0.002
Production SE3	3.364e-05	6.64e-05	0.507	0.612	-9.65e-05	0.000
Month	2.5374	0.623	4.075	0.000	1.316	3.758
Hydro reservoir NO	0.0021	0.000	5.438	0.000	0.001	0.003
Hydro reservoir SE	-0.0051	0.001	-5.625	0.000	-0.007	-0.003
Consumption NO1	0.0020	0.000	5.996	0.000	0.001	0.003
Consumption NO2	-0.0027	0.001	-3.947	0.000	-0.004	-0.001
Consumption NO3	0.0007	0.000	1.391	0.164	-0.000	0.002
Consumption NO5	-0.0040	0.001	-7.508	0.000	-0.005	-0.003
Consumption DK1	-0.0010	0.000	-2.355	0.019	-0.002	-0.000
Consumption SE3	0.0017	0.000	9.795	0.000	0.001	0.002
Windpower DK1	-0.0014	0.000	-4.838	0.000	-0.002	-0.001
Windpower DK2	-0.0017	0.001	-3.357	0.001	-0.003	-0.001
EUR/NOK	-19.7314	7.234	-2.728	0.006	-33.918	-5.545
Carbon Emission Future	3.5094	0.439	7.993	0.000	2.648	4.370
SEK/NOK	20.7153	73.132	0.283	0.777	-122.701	164.132
Natural Gas	0.3911	0.024	16.283	0.000	0.344	0.438
Coal price	2.3053	0.129	17.929	0.000	2.053	2.557
Temperature norway	3.3181	0.652	5.093	0.000	2.040	4.596

Tabell 13 Koeffesienter fra lineær regresjonsmodell

Hvis vi legger til grunn et 95% konfidensnivå vil en t-verdi på over 1,96 gjøre koeffisienten statistisk signifikant. Som vi ser ut ifra tabellen over er de fleste koeffisienten i modellen signifikante og har en absoluttverdi på t-verdien over 1,96. Unntakene er koeffisienten for Production SE3, Consumption NO3 og valutaparet SEK/NOK. Disse variablene har da ikke en signifikant effekt på den avhengige variabelen som er elspot prisen i Oslo.

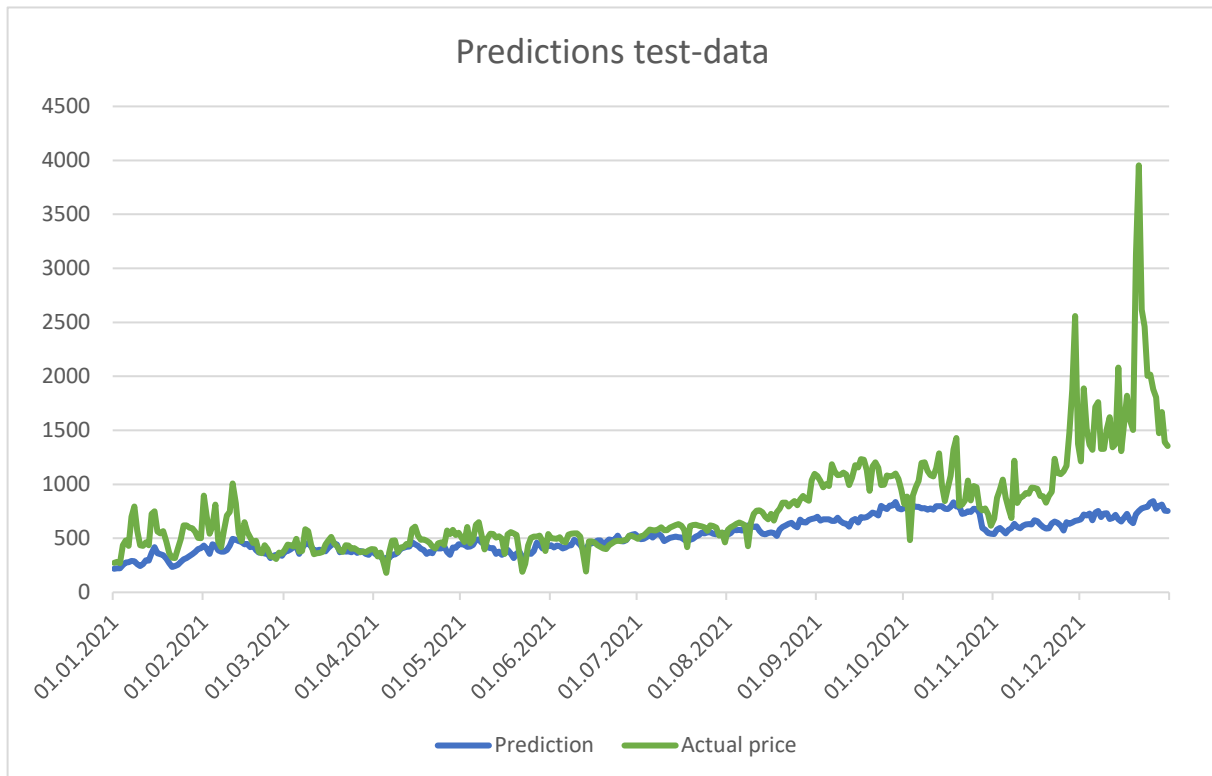
Den viktigste funksjon til våre modeller er å kunne predikere strømpriser. Av grafen under ser man prediksjonene vår lineære modell gjør på dataene i treningssettet. Den blå linjen representerer de predikerte verdiene på elspotprisen i Oslo og den grønne linjen representerer de faktiske strømprisene. Ved visuell inspeksjon ser man at modellen vår treffer ganske bra i forhold til de faktiske strømprisene, men at det likevel er viss forskjell mellom predikert og faktisk verdi. Særlig vanskelig er det for modellen å predikere plutselige og kraftige endringer i strømprisen. Derfor kan man si at den lineære modellen predikerer ganske bra på treningsdataene, men at det fortsatt er variasjon som modellen ikke greier å predikere.



Figur 18 Prediksjoner treningsdata lineær modell

Selv om modellen predikerer nokså bra på treningsdataene er det mest sentrale hvor bra modellen predikerer på testdataene. Disse prediksjonene er illustrert i grafen under. Også her

er den blå linjen de predikerte verdiene, mens den grønne linjen representerer de faktiske verdiene. Av grafen ser man at modellen treffer ganske bra ved sommerhalvåret. Vi ser også at strømprisen blir mer volatile og øker kraftig om vinteren. Dette greier ikke modellen vår å fange opp. Modellen predikerer høyere pris i forhold til prisen den predikerer om sommeren, men den er ikke på langt nær høy nok hvis man sammenligner med den faktiske strømprisen.



Figur 19 Prediksjoner testdata lineær modell

4.2 LSTM-modell

Etter å ha utforsket, forskjellige sammensetninger av noder, hiddenlayers og aktiveringsfunksjoner kom vi frem til denne endelige modellen:

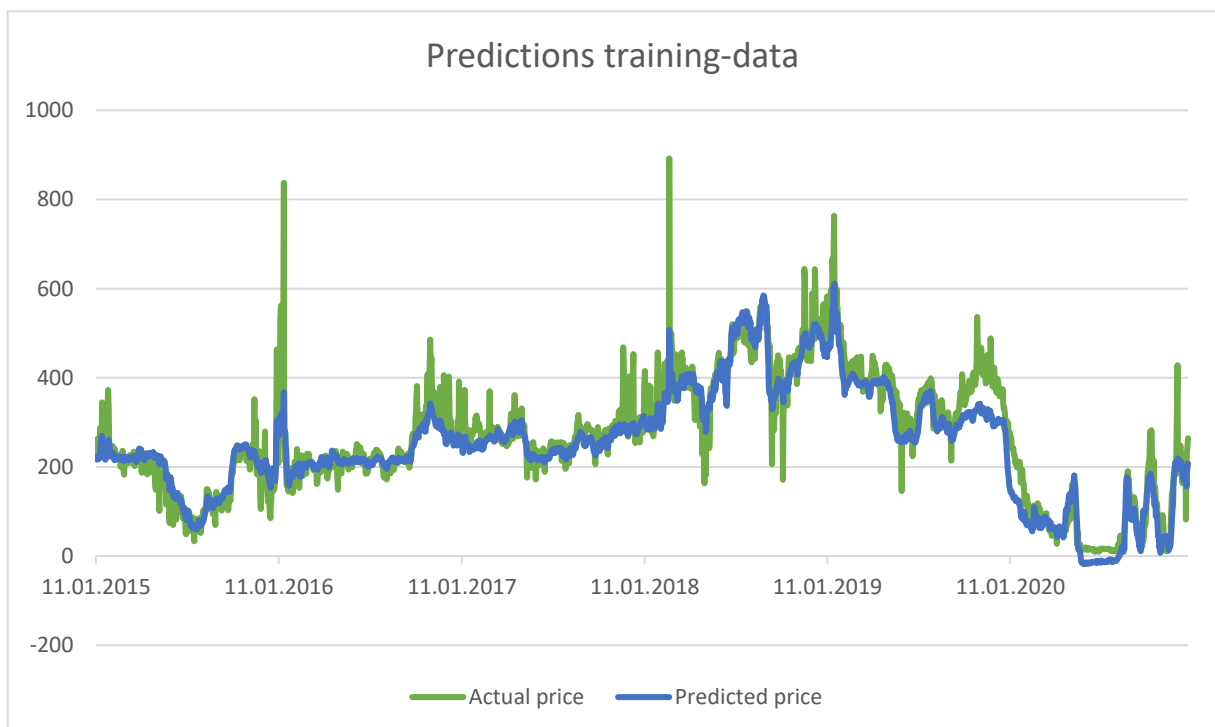
```
model = Sequential()  
model.add(LSTM(60, input_shape=(10,23) , activation='relu', dropout=0.2, return_sequences = True))  
model.add(LSTM(60,activation='tanh'))  
model.add(Dense(1, activation = "linear"))  
model.compile(loss='MSE', optimizer="adam")
```

Figur 20 Oversikt over LSTM modellen

Modellen består altså av 2 hiddenlayers og begge disse lagene inneholder 60 noder hver. Begge lagene har også samme long short-term memomry struktur. Det som skiller disse to lagene er aktiveringsfunksjonen. Det første laget har en ReLu aktiveringsfunksjon. Dette er en

lineær funksjon, som kun kan inneholde positive verdier. Det vil si at man ikke kan få verdier som er lavere enn null. Det andre laget har en hyperbolsk funksjon som aktiveringsfunksjon. Input-layeret tar inn 23 uavhengige variabler fra 10 dager i forveien. Outputlayeret inneholder en node med vanlig nevral nettverk struktur. Her er aktiveringsfunksjonen en lineær funksjon.

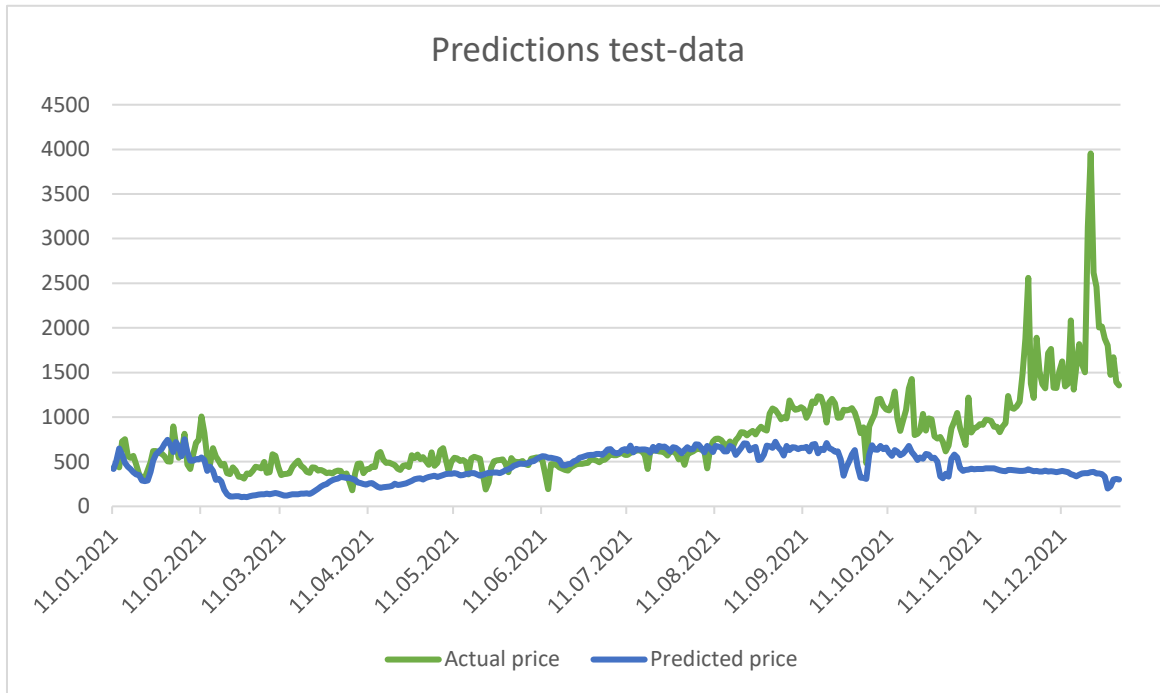
Resultatene til LSTM modellen på treningsdataene er illustrert ved grafen nedenfor. Som tidligere er den faktiske strømprisen representert med en grønn linje. Vår modells prediksjoner ser man ut ifra den blå linjen. Av grafen ser man at modellen predikerer strømprisen ganske godt. Den følger de samme bevegelsene som den faktiske strømprisen. På en annen side greier heller ikke denne modellen å predikere kraftige endringer i strømprisen.



Figur 21 Prediksjoner treningsdata LSTMmodel

Selv om LSTM-modellen predikerer ganske bra på treningsdataene er det mest interessant å se hvordan modellen predikerer på testdataen. Disse resultatene kan man se fra grafen under. Også her er de faktiske prisene illustrert med grønt og de predikerte prisene med blått. Av grafene kan det virke som at også LSTM modellen sliter med å predikere strømpriser som den ikke har trent seg på. Det kan virke som at også denne modellen predikerer best i sommerhalvåret og dårligere om vinteren. Man ser også at modellen ikke greier å henge med

på bevegelsene i strømprisen mot slutten av året. Her er det store forskjeller mellom faktisk og prekikert strømpris.



Figur 22 Prediksjoner testdata LSTMmodell

4.3 Sammenligning av modellene

For å sammenligne modellene bruker vi Mean Squared Error, Mean Absolute Percentage Error og visuell inspeksjon. Av tabellen nedenfor ser man hvor godt de 2 modellene scorer på Mean Squared Error og Mean Absolute Percentage Error.

	MSE Training	MAPE Training	MSE Test	MAPE Test
Lineær regresjon	4225.4191	0.4618	177822.9875	0.2528
LSTM	1766.4556	0.1932	307823.7417	0.38711

Tabell 14 MSE og MAPE

På treningsdataene scorer den lineære regresjonsmodellen dårligere enn LSTM modellen. Dette ser man ved at MSE er høyere hos den lineære modellen enn LSTM. MAPE er også mindre hos LSTM-modellen enn regresjonsmodellen og indikerer også at LSTM modellen predikerer bedre på treningsdataene. Disse resultatene sammenfaller også med tidsserien med faktiske strømpriser og predikerte strømpriser. Ved visuell inspeksjon av disse resultatene ser det også ut som at LSTM modellen følger de faktiske prisene bedre enn regresjonsmodellen.

På testdataene er situasjonen litt mer komplisert. Her er MSE høyere for den lineære regresjonsmodellen, men MAPE-scoren tilsier at LSTM-modellen er den beste. Likevel ser man ved visuell inspeksjon av prediksjonene av modellene og faktiske priser, at ingen av modellene predikerer strømprisen spesielt bra. Begge modellene predikerer spesielt dårlig mot slutten av året, hvor strømprisen økte betydelig. På sommerhalvåret bommer ikke modellene like mye, men det virker ikke som prediksjonene treffer så veldig bra her heller. På bakgrunn av dette kan man si at det er ingen av modellene som predikerer spesielt bra på testdataene.

5. Feilkilder

5.1 Feil i dataene

I vårt forsøk på å predikere strømpriser og finne ut hvilken modell som gir de beste prediksjonene, kan man ha vært utsatt for en rekke feilkilder. En av de første feilkildene er at man har feil i dataene. Selv om vi har undersøkt datasettet vårt og har lett etter feil, kan man ikke være hundre prosent sikker at alle dataene er riktig. Likevel gir ikke undersøkelsen av våre variabler noen grunn til å tro at dataene er helt feil. Om det er noen små avvik i noen verdier, så er de mest sannsynlig ganske små og det er liten risiko for at de er så feil at de ødelegger for modellene våre.

5.2 Inntastingsfeil

En annen feilkilde er inntastningsfeil. I vår analyse har vi prøvd å gjøre ting så riktig som overhodet mulig, men man kan aldri være helt sikker på at man ikke har gjort feil. I oppgaver som datahåndtering, å lage modeller og koding i python er det muligheter for å gjøre feil. Vi har derfor gått igjennom disse oppgavene flere ganger og lett etter feil. Likevel er det ikke sikkert at vi har fått luket ut alle feilene. Derfor er inntastningsfeil en feilkilde som er verdt å nevne.

5.3 Overtilpassing

Når man lager modeller som predikerer tidsserier er overtilpassing et problem. Overtilpassing er når prediksjonene fra en modell er for nærme de faktiske verdiene på treningssettet (Twin, 2021). Da vil modellen kunne predikere godt på treningsdataene, men ikke så godt på data som modellen ikke er trent på. Ut ifra resultatene våre opplever vi at både den lineære

modellen og LSTM modellen predikerer bedre på test dataene. Dette kan tyde på at begge modellene er overtilpasset til treningsdataene. Tiltak man kunne gjort for å minimere sannsynligheten for overtilpassing hadde for eksempel vært å ha trente modellene færre ganger på treningssettet. Dette er særlig aktuelt for LSTM-modellen. Vi hadde også med en god del variabler i våre modeller. Om vi hadde redusert antall variabler kunne man også redusert sannsynligheten for overtilpassing. Selv om disse tiltakene kunne redusert sannsynligheten for overtilpassing, hadde man også økt sannsynligheten for at modellene blir undertilpasset. Undertilpassing er når prediksjonene ikke er tilfredsstillende fordi modellen har trent for lite på treningsdataene. Derfor er det en balansegang å sikre at modellene ikke overtilpasset, men heller ikke undertilpasset.

5.4 Forutsetninger for lineær regresjon

Lineær regresjon bygger på en rekke forutsetninger. Selv om det er nøyaktigheten på prediksjonene som er viktigst ved en predikativ analyse, kan likevel brudd på disse forutsetningene gjøre prediksjonene dårligere. En av forutsetningene er at sammenhengen mellom avhengig variabel og uavhengig variabel er lineær. Denne forutsetningen har vi sjekket ved visuell inspeksjon. På bakgrunn av dette virker det som at sammenhengene mellom den avhengige variabelen og de uavhengige er ganske lineær. Likevel er de fleste sammenhengene ikke perfekt lineær og det kan hende at sammenhengene er ikke-lineære. Likevel virker det som at denne forutsetningen er mest sannsynlig oppfylt.

Andre forutsetninger for lineær regresjon er at alle forklaringsvariablene er ukorrelert med restleddet. Dette er en forutsetning som er vanskelig å kontrollere for siden vi ikke har kontroll over alle de utelatte variablene og derfor ikke har kontroll over om forklaringsvariablene våre korrelerer med restleddet. Om restleddet er korrelert med forklaringsvariablene kan man risikere å få upresise prediksjoner.

Lineær regresjon bygger også på flere forutsetninger som at residualene har forventning lik 0, restleddene er ukorrelerte med hverandre, restleddet har konstant varians og er normalfordelt. I tillegg skal ingen av forklaringsvariablene skrives som perfekt lineær funksjon av noen av de andre forklaringsvariablene. Vi har prøvd å oppfylle disse forutsetningene så godt som mulig, men en del av disse forutsetningene er vanskelig å kontrollere for og man kan ikke være 100% sikker på om de er oppfylt. Ved brudd på disse forutsetningene kan modellen gi dårligere prediksjoner, enn om de hadde vært oppfylt.

5.5 Utelatte variabler

En naturlig kilde til feil når man skal vurdere såpass store datamengder er at man har utelatt enkelte variabler som kunne ha utgjort en stor forskjell for resultatene. Dette har vi prøvd å unngå ved å ta med en mengde forskjellige variabler som kan dekke flere forhold samtidig. Et eksempel på dette er valuta, som indirekte vil fange opp faktorer i samfunnet man ikke nødvendigvis hadde fått dekket med en variabel. Selv om oppgaven inneholder en rekke av disse mer «generelle» variablene som tar opp flere forhold samtidig, må man ta høyde for at noen forhold ikke er med, som kan gjøre modellen mindre presis.

Et eksempel på en slik variabel er nedbør. Strømprisen er blitt sterkt påvirket av nedbør, og det ville antagelig gitt modellen mer presise data å jobbe med om det hadde vært inkludert. Det nærmeste vi kommer å ha med nedbør i vår modell er værdataen som går på temperatur og vann i magasinene, og selv om disse får frem sammenhengen mellom vær og pris, så får man nok ikke en like sterk sammenheng som man ville ha fått ved å ha med nedbør. Man må ta høyde for at det utenom vår kunnskap er variabler som ikke er kommet på, men som helts skulle vært inkludert, men vi tror at vår modell med de variablene den har skal kunne gi et ganske nøyaktig verdensbilde.

5.6 For kort tid

I vår oppgave valgte vi å begrense tidsrommet til å gjelde fra og med 2015, til og med 2021. Dette var et bevisst valg vi tok for å hindre at datamengden ble for store til å håndtere for våre datamaskiner. Samtidig tenkte vi at 6 år med data ville være nok til å gi en relativt god indikasjon på forskjellene i modellene. Dette er allikevel en veldig klar feilkilde, fordi modellene kunne gitt mer presise resultater med mer data.

I tillegg til dette måtte vi også se oss nødt til å omgjøre noe av dataen til verdier på ukesbasis, slik som vann i vannmagasinene, kullpris og valuta. Dette kutter ned datamengden betraktelig, og er en åpenbar feilkilde. Det vil allikevel gi en representativ og god indikasjon, men det er ikke optimalt og kan bidra til at modellene ikke blir riktig dimensjonert.

5.7 Visuell inspeksjon – tolkning av data

En annen feilkilde man må være klar over er den visuelle tolkningen av dataene. Det er en fare for at man kan se sammenhenger i den visuelle representasjonen som ikke er der. Eller i

det minste at man ser dataen med lit subjektive øyne, og tolker den til ønsket retning. Dette har vi forsøkt å være oppmerksom på slik at vi unngår dette, men det er en helt klar feilkilde som man må være klar over.

5.8 Politikk

En annen sentral feilkilde man må ta med i beregningen er politikk. Politiske restriksjoner kan påvirke mange faktorer i samfunnet som igjen vil kunne gi en effekt på strømprisen. Dette kan for eksempel være mindre strømforbruk knyttet til mindre bruk av kontorlokaler i forbindelse med koronarestriksjoner. Eller kanskje mer aktuelt i disse dager, politiske sanksjoner mot Russland i forbindelse med krigen i Ukraina, som øker olje- og kullprisen. Denne prisøkningen vil igjen gi økt pris på strøm, fordi etterspørselen øker.

6. Konklusjon

Ut ifra resultatene våre kan man se at LSTM modellen predikerer bedre enn regresjonsmodellen på treningsdataene. Dette ser man ved visuell inspeksjon, men mest betydningsfullt er resultatene fra Mean Squared Error og Mean Absolute Percentage Error. Begge disse modell-målene tilsier at LSTM modellen er bedre enn regresjonsmodellen på treningsdataene.

På den andre siden er resultatene fra testdataene litt mer komplisert. Ved visuell inspeksjon ser man at begge modellen predikerer best fra januar til september. Fra september og frem til og med desember øker strømprisene voldsomt i forhold til de prisene man har observert før. Ingen av modellene er i nærheten av å predikere denne økningen og avvikene mellom predikert pris og faktisk pris øker betydelig for begge modellene. Sammenligner man modellene ut ifra Mean Squared Error og Mean Absolute Percentage Error, scorer regresjonsmodellen best på Mean Squared Error, mens LSTM scorer best på Mean Absolute Percentage Error. Det at MSE og MAPE favoriserer forskjellige modeller gjør det vanskelig og konkluderer med hvilken modell som er best. Det hjelper heller ikke at begge modellen får store avvik, når de faktiske strømprisene øker betraktelig. Det at man får disse sprikende resultatene kan være en konsekvens av at ingen av modellene predikerer strømprisen på testdataene spesielt godt. Man kan derfor ikke si at en modell er bedre enn den andre på testsettet, men at begge modellene ikke predikerer strømpriser med den nøyaktigheten vi hadde ønsket.

Siden treningsdataene er det datagrunnlaget modellene er trent opp på, er det naturlig å ikke vektlegge resultatene på det datasettet like mye som man vektlegger resultatene fra testsettet. På den andre siden predikerer ingen av modellene tilfredsstillende på testsettet og det er vanskelig og konkluderer med hvilken modell som er den beste. Derfor blir resultatene fra treningssettet avgjørende. Derfor kan man si at Long-short term modellen er den beste modellen, om man må gi en konklusjon. Likevel er det sannsynlige at denne modellen er overtilpasset til treningsdataene og prediksjonene er faktisk ikke bedre enn regresjonsmodellen. Så selv om LSTM modellen kan virke som den beste så er det egentlig usikkert om dette er tilfellet.

Det som særlige resultatene fra testsettet forteller oss, er at begge modellene ikke predikerer strømpriser særlig godt. Man kan derfor si at begge modellene gir for dårlige prediksjoner til at de kan brukes til å forutsi strømpriser. Dette gir grunnlag for å undersøke andre modellere som kan gi bedre prediksjoner. Man kan også prøve å bruke andre data for å undersøke om

det gir bedre resultater. Det som er sikkert, er at våre modeller ikke predikerer strømpriser tilfredsstillende og man bør undersøke videre for å finne modeller som predikerer bedre.

7. Referanser

- Bugge, S., Tørmoen, A. S. T. & Mikalsen, H., 2022. *Verdens Gang*. [Internett]
Available at: <https://www.vg.no/nyheter/innenriks/i/qWEmm0/stroemprisene-foerer-til-kutt-i-annet-forbruk-ogsaa-katten-sherlock-fikk-svi>
[Funnet 20 Februar 2022].
- Energidepartementet, O.-o., 2022. *energifaktanorge.no*. [Internett]
Available at: <https://energifaktanorge.no/norsk-energiforsyning/kraftforsyningen/#vannkraft>
- IBM Cloud Education, 2020. *Neural Networks*. [Internett]
Available at: <https://www.ibm.com/cloud/learn/neural-networks>
[Funnet 23 February 2022].
- Investing.com, 2022. *Commodities*. [Internett]
Available at: <https://www.investing.com/commodities/>
[Funnet Januar 20 2022].
- Li, W. & Becker, D. M., 2021. *Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling*, Trondheim: Science Direct.
- Mishkin, F. S. & Eakins, S. G., 2018. *Financial Markets and Institutions*. Ninth Edition red. s.l.:Pearson.
- Nord Pool AS, 2022. *Historical Market Data*. [Internett]
Available at: <https://www.nordpoolgroup.com/historical-market-data/>
[Funnet 20 Januar 2022].
- Norges Energi, 2022. *Norges Energi*. [Internett]
Available at: <https://norgesenergi.no/stromsmart/dette-pavirker-stromprisen/>
[Funnet 10 Mars 2022].
- NVE, 2021. *nve.no*. [Internett]
Available at: <https://www.nve.no/om-nve/spoer-nve/om-strompriser/sporsmal-og-svar-om-stromprisen/>
- Olah, C., 2015. *Understanding LSTM Networks*. [Internett]
Available at: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
[Funnet 23 Februar 2022].
- Oppen, M., Mørk, B. E. & Haus, E., 2020. *Kvantitative og kvalitative metoder i merkantile fag*. Oslo: Cappelen Damm.
- Opstad, L., 2020. *Innføring i makroøkonomi*. 3. red. Oslo: Cappelen Damm.
- Pindyck, R. S. & Rubinfeld, D. L., 2013. *Microeconomics Eighth Edition*. s.l.:Pearson.
- Shmueli, G., Bruce, P. C., Gedeck, P. & Patel, N. R., 2020. *Data Mining for Business Analytics*. 1 red. Hoboken: John Wiley & Son Inc.
- Twin, A., 2021. *Investopedia*. [Internett]
Available at: <https://www.investopedia.com/terms/o/overfitting.asp>
[Funnet 3 Mars 2022].

