



DET TEKNISK-NATURVITENSKAPELIGE FAKULTET

BACHELOROPPGAVE

Studieprogram/spesialisering:
Matematikk og fysikk

Vårsemesteret, 2022

Åpen / Konfidensiell

Forfatter:
Benjamin Sigbjørnsen

Benjamin Sigbjørnsen
(signatur forfatter)

Veileder:
Alex Bentley Nielsen

Tittel på bacheloroppgaven:
En statistisk analyse av strømprisen i NO₂

Engelsk tittel:
A statistical analysis of the price on electricity in NO₂

Studiepoeng: 20

Emneord:
Statistisk læring, Dataanalyse,
R, Strømpriser, Nord Pool,
Modellering

Sidetall: 48

Stavanger, 13.mai 2022

Kapittel 1

Sammendrag

Denne oppgaven ser på hva som påvirker strømprisen i prisområdet NO2 på strømbørsen Nord Pool. NO2 er prisområdet til Sørvest-Norge. Strømprisen i Norge bestemmes av tilbud og etterspørsel. Tilbudet av strøm avhenger av hvor mye strøm som produseres. I Norge så kommer strømmen fra fornybare energikilder som i all hovedsak er væravhengige, og derfor blir strømprisen påvirket av været. Etterspørselen etter strøm bestemmes av forbruket. I tillegg så påvirkes tilbud og etterspørsel av strømprisene i andre prisområder på grunn av overføring av strøm. Flere forklaringsmodeller ble tilpasset og prøvd på valideringsdata fra 2022. Det ble prøvd lineære og additive modeller med ulikt datagrunnlag. Det ble hentet data fra og med 2017 til og med 2021. Den beste modellen til å forklare strømprisen ble funnet å være en lineær modell med data fra 2021 og logaritmen til strømprisen som responsvariabel. Modellene ble også brukt til å si noe om når et vannkraftverk bør selge strøm.

Kapittel 2

Forord

Først og fremst må jeg si takk til veilederen min Alex Bentley Nielsen. Har fått god veiledning og mange gode innspill. Oppgaven var hans forslag, og selv om den egentlig var beregnet for en masterstudent, så ble den skalert ned og gitt til meg. Strømmermarkedet er komplisert. Jeg skjønner hvorfor oppgaven var beregnet på en masterstudent, men jeg har lært mye og fått utfordret meg i noe jeg er interessert i. Takk også til Jan Terje Kvaløy, Tore Selland Kleppe og Magnus Landstad for deres synspunkt og hjelp som har hjulpet meg videre i oppgaven. Til slutt må jeg si takk til Nord Pool for tilgang til dataene deres.

Innhold

1	Sammendrag	1
2	Forord	2
3	Introduksjon	5
4	Strømmarkedet	7
4.1	Historien til strømmarkedet i Norge	7
4.2	Nord Pool	8
4.3	Prediktorer for strømprisen i NO2	9
4.3.1	Vann- og vindkraft	9
4.3.2	Overføring av strøm	10
5	Metodikk	13
5.1	Innsamling av data	13

<i>INNHold</i>	4
5.2 Modeller	17
5.3 Analyse av resultater	20
6 Analyse	25
6.1 Variasjon i strømprisen i løpet av året	25
6.1.1 Resultater og diskusjon	25
6.1.2 Hva ville dette betydd for et lite vannkraftverk?	30
6.2 Forklare variasjon i strømprisen best mulig	33
6.2.1 Resultater og diskusjon	34
6.2.2 Hva ville dette betydd for et lite vannkraftverk?	41
7 Konklusjon	43
Referanser	45
8 Tillegg	48
8.1 R-kode	48

Kapittel 3

Introduksjon

Strøm, strømpris og strømmarkedet har de fleste av oss et forhold til. Vi er avhengige av strøm, og bruker strøm nesten uansett hva vi gjør. Samfunnet i dag hadde ikke fungert uten strøm. Siden vi bruker så mye strøm, så er prisen veldig viktig. Strømmen holder alltid lik kvalitet, men prisnivået svinger. I mange år har prisen på strøm vært lav i hele Norge, hele året rundt. Dette virker det å være slutt på nå (Malkenes Hovland, 2022), ihvertfall midlertidig. I Norge bruker vi i gjennomsnitt 16 000 kWh i året per husholdning og betaler i gjennomsnitt 17400 NOK i året per husholdning (Esparza, 2019). Det er derfor ikke rart at folk bryr seg om strømprisen. En endring på bare 10 prosent i strømprisen tilsvarer nesten 2000 kr for en gjennomsnittsfamilie i Norge. I 2022 har mange fått en mye større strømregning, og mange har slitt så mye at staten så seg nødt til å betale deler av strømregningen (Olje- og energidepartementet, 2021). Det er derfor naturlig at det blir støy rundt strømprisen og strømmarkedet.

Hvordan utvikler strømprisen seg i framtiden? En måte å beregne dette på

er ved å se på tidligere data. Det vil da være hensiktsmessig å se på faktorer som har med tidspunkt å gjøre, slik at vi kan predikere prisen til bestemte tidspunkt. Dette er derimot ikke det eneste som påvirker strømprisen. Det vil være mulig å predikere strømprisen bedre med mer data.

Det er nyttig å vite hvordan strømprisen utvikler seg hvis en skal selge eller kjøpe strøm, og tidspunktet for kjøp og salg ikke betyr noe. Dette vil for eksempel være hensiktsmessig å vite for et vannkraftverk. Vannet kan ofte lagres helt til det er fordelaktig å produsere strøm i turbinene. Det er også nødvendig å vite hvilke faktorer som påvirker strømprisen hvis man skal prøve å gjøre noe med de store svingningene i prisen på strøm.

I denne oppgaven så skal jeg identifisere forklaringsvariabler for spotprisen i NO2, og se på hvordan prediktorene påvirker prisen. Denne informasjonen skal brukes til å si noe om hvordan et vannkraftverk bør regulere strømproduksjonen.

Kapittel 4

Strømmarkedet

4.1 Historien til strømmarkedet i Norge

Det moderne strømmarkedet ble på mange måter til i 1991. Da kom energiloven og endret spillereglene. Nord Pool ble opprettet og markedet ble liberalisert. “The Nordic power market is among the most liberalized and competitive power markets in the world and has experienced several deregulations since it was liberalized in 1991” (Haugom, 2018, Side 1).

Energiloven regulerer produksjon, omforming, overføring, omsetning, fordeling og bruk av energi. Målet er at alt dette skal foregå på en måte som gagnar samfunnet, både i det lange løp og økonomisk her og nå. Hovedpoenget med loven var å liberalisere strømmarkedet slik at det ikke var staten, men tilbud og etterspørsel som styrte prisene. Etter at energiloven ble innført, så hadde Norge et av de mest liberale kraftmarkedene. EU og andre har senere fulgt etter med liknende tankegang (Askheim og Brænd, 2022).

Markedet har fungert bra lenge med lave og forutsigbare strømpriser. I 2021 ble prisen på strøm mye høyere. Nå etterlyses det endringer blant folk og bedrifter (Næringslivsmagasinet Rosenkilden, 2022).

4.2 Nord Pool

Følgende materiale om Nord Pool i seksjon 4.2 er hentet delvis fra Store norske leksikon (Askheim, 2019).

Nord Pool er en børs for kjøp og salg av strøm mellom prisområder i og rundt Norden. Norge, Sverige og Danmark er delt inn i flere prisområder, mens resten av landene som handler strøm på Nord Pool ikke er delt opp i prisområder. Når Nord Pool kom, ble den verdens første internasjonale børs for handel av strøm/kraft. Det er tilbud og etterspørsel på denne børsen som bestemmer strømprisen. Nord Pool har tradisjonelt sett blitt eid av landene som selger strøm på Nord Pool, men i 2019 kjøpte den europeiske børsen Euronext opp 66 prosent av selskapet.

Hvis du ønsker å selge eller kjøpe strøm, så er det forskjellige måter å gjøre dette på. Én type handel på strømbørsen er kontrakter som varer i en time. Prisen fastsettes for et døgn om gangen, og endrer seg fra time til time. Prisen settes slik at tilbud og etterspørsel blir balansert. Dette markedet kalles spotmarkedet. Strømvtale som baserer seg på spotpris har vist seg å være den rimeligste strømvtaalen over tid for folk flest (Kaldestad, 2021). Når strømprisen blir nevnt, i denne oppgaven eller i samfunnet ellers, så snakkes det som regel alltid om spotprisen. Strømprisen du betaler er ikke lik spotprisen. Du betaler også nettleie og en avgift til strømselskapet ditt.

En annen type strømavtale er å kjøpe strøm til fastpris. Fastpris går ut på at kjøper får en fast pris på strømmen mot at han binder seg til en avtale. Opsjoner og grønne sertifikater blir også brukt i krafthandelen.

4.3 Prediktorer for strømprisen i NO2

Det som styrer strømprisen er, som nevnt tidligere, i all hovedsak tilbud og etterspørsel. Tilbud fra vannkraft og vindkraft og prisene i strømmarkedene rundt NO2 er viktigst for strømprisen i NO2 (M. B. Landstad, personlig kommunikasjon, 4. januar 2022).

4.3.1 Vann- og vindkraft

I November i 2021 kom 90,2 prosent av produksjonen i Norge fra vannkraft, mens 8,8 prosent kom fra vindkraft og 1 prosent fra varmekraft (SSB, u.å). Dette gjør vannkraft og vindkraft til de klart viktigste energikildene for strømproduksjon i Norge. Det vil også være rimelig å anta at energiproduksjonen fra jordvarme er konstant. Både vannkraft og vindkraft er avhengige av været, og derfor vil strømprisen være avhengig av været. For å predikere strømprisen vil det være nødvendig å ha gode prediksjoner av været framover i tid, og data bakover i tid, for å lage gode modeller av strømprisen.

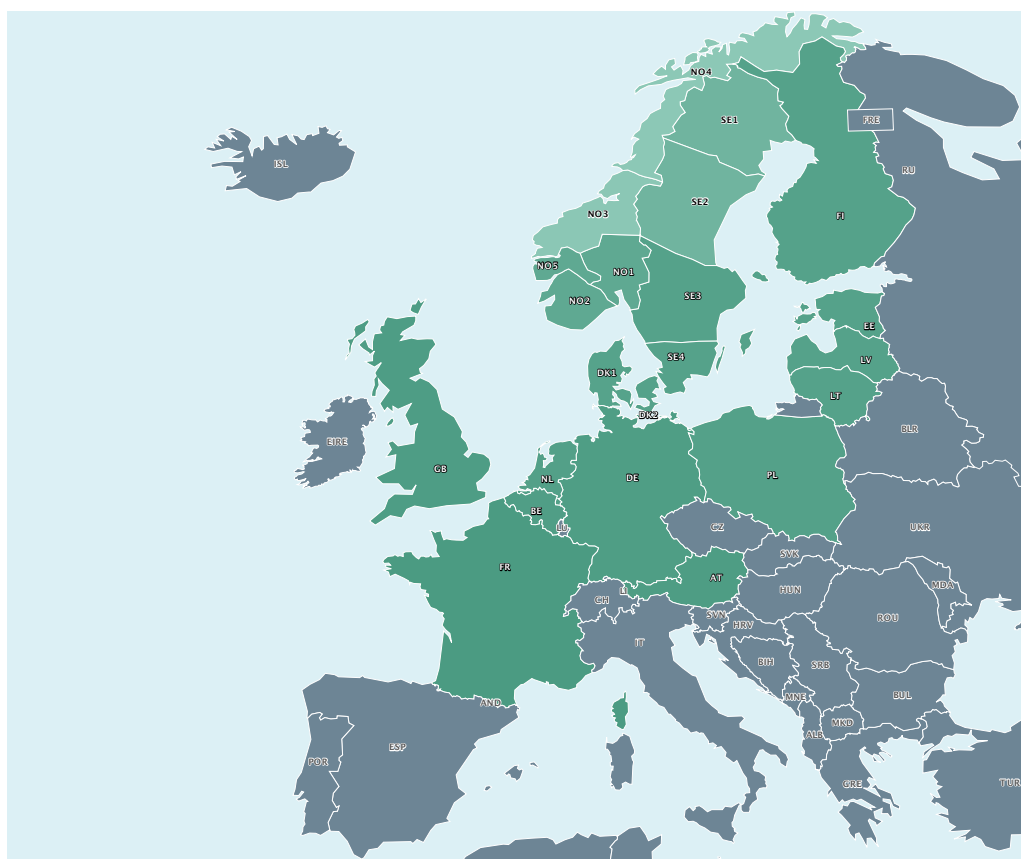
Mengden strøm som kan produseres fra vannkraft vil være proporsjonal med mengden nedbør. Man kan produsere mer strøm jo mer vann det er som renner gjennom turbinen. Dette vil gjelde for de aller fleste vannkraftverkene så lenge det ikke er store nedbørsmengder og vannet bare må slippes utenom

turbinen for å unngå flom. Det som er fint med vannkraft er at vannet kan lagres i magasiner. Selv om det ikke har kommet nedbør på en god stund, så kan vi fortsatt få strøm fra vannkraft. Det er den totale mengden nedbør over en lengre periode som er viktig. Vannkraft er den eneste fornybare energikilden som kan lagres uten å involvere batterier. Hvor mye vann som er lagret i magasinene, kalt fyllingsgraden, vil derfor også være viktig for strømprisen.

Vindkraft derimot må brukes med en gang den produseres. Vinden kan på ingen måte lagres så produksjonen fra vindkraft er proporsjonal med vindstyrken der og da. Med mindre det blåser for mye. Blåser det for mye, kan ikke vindmøllene produsere strøm på grunn av risiko for ødeleggelser.

4.3.2 Overføring av strøm

Overføring av strøm mellom ulike markeder er det som gjorde at Nord Pool ble startet. Tanken er at markeder der det er stort tilbud på strøm og lite etterspørsel kan selge strøm til plasser med større etterspørsel. Dette gjør at tilbud og etterspørsel balanseres og at prisen blir likere i markedene som utveksler strøm med hverandre. Overføring blir begrenset av flaskehals (Abrahamsen, 2014, side 8), og derfor er det viktig å bygge ut overføringskapasiteten for å bedre flyten.



Figur 4.1: Prisområder Nord Pool (Nord Pool, 2022)

Norge er delt inn i 5 regionale strømmarked med ulike prisnivå som vist i Figur 4.1. Sørøst-Norge (NO1), Sørvest-Norge (NO2), Midt-Norge (NO3), Nord-Norge (NO4) og Vest-Norge (NO5). I tillegg til at det overføres strøm mellom regionene i Norge, så overføres det strøm mellom Norge og andre land. Fra NO2 går det kabler til Nederland, Tyskland, Danmark og England. Det overføres også strøm til Sverige, Finland og Russland fra andre deler av Norge. Prisen i NO2 vil dermed avhenge av alle strømmarkedene rundt, ved at det vil overføres strøm gjennom kabler hvis prisen er høyere på andre siden av kabelen. Ofte vil dette være tilfelle fra Norge til utlandet. Norge har

historisk sett hatt lave strømpriser å grunn av stort tilbud av strøm fra all vannkraften.



Figur 4.2: Dette er alle Norges utenlandskabler og -linjer per oktober 2021 (Magnussen, 2021).

Kapittel 5

Metodikk

5.1 Innsamling av data

Alle data som omhandler responsvariabelen, strømprisen i NO₂, er hentet fra Nord Pool (Nord Pool, u.å.) i 2022 på nettsiden

<https://www.nordpoolgroup.com/historical-market-data/>

Daglig spotpris i Kristiansand ble brukt. Kristiansand ligger i NO₂, og daglig spotpris betyr at dataene viser dagsgjennomsnittet til spotprisen. Dataene ble i slutten av semesteret, før jeg skulle levere, skjult bak innlogging.

Data fra og med år 2017 til og med år 2021 er brukt til modellering, mens dataene fra 2022 er brukt som verifiseringsett. Tanken bak å bruke de siste fem årene som treningsdata er at det gir et stort datagrunnlag, uten å gå langt tilbake i tid.

Videre ble det hentet data fra ulike prediktorer. Informasjon om årstall, dato, årstid og dager (nummer) etter nyttår ble lagt til. Det ble også lagt til infor-

masjon om datoene er i en helg, i fellesferien eller på en helligdag (søndag er ikke med). Dette er informasjon som er lett tilgjengelig i en kalender. Fellesferie er det forskjellige definisjoner på, her ble det valgt 4 uker som er omtrent midt på sommeren. Slik ser de øverste linjene av dataene ut:

```
i..SYS SE1 SE2 SE3 SE4 FI DK1 DK2 Oslo Kr.sand Bergen Molde Tr.heim TromsA. EE
285.86 273.38 273.38 273.38 378.47 274.18 505.50 505.50 272.60 272.60 272.60 267.89 267.89 236.81 378.47
313.20 329.82 329.82 329.82 450.92 342.71 528.10 528.10 280.41 280.41 280.41 320.49 320.49 253.47 450.92
269.14 269.28 269.28 269.28 285.23 269.75 297.11 313.06 269.28 269.28 269.28 269.28 269.28 229.56 330.46
406.74 440.48 440.48 440.48 458.35 452.89 440.48 458.35 434.61 434.61 434.61 440.48 440.48 204.81 455.05
456.62 493.48 493.48 493.48 493.48 493.48 493.42 493.48 480.22 480.22 480.22 493.48 493.48 209.59 493.48
413.62 430.00 430.00 430.00 430.00 430.00 430.00 430.00 430.00 430.00 430.00 430.00 430.00 251.35 430.00
LV LT Nummer Dato Helg Aarstid Aar
378.47 378.47 1 01.01.2021 0 0 0
450.92 450.92 2 02.01.2021 1 0 0
330.46 330.46 3 03.01.2021 1 0 0
455.05 458.35 4 04.01.2021 0 0 0
493.48 493.48 5 05.01.2021 0 0 0
430.00 430.00 6 06.01.2021 0 0 0
```

Figur 5.1: Bilde fra utskrift i Rstudio ved bruk av R

Årstid er en kategorisk variabel, der 0 er vinter, 1 er vår, 2 er høst og 3 er sommer. År er kodet som antall år før 2021 og er brukt som kategorisk variabel. Etter at modellen for 2021 var klar så ble det lagt til data fra 2017 til 2020 i modellen for å få et større datagrunnlag.

Deretter ble det lagt til info om tilbud og etterspørsel for både modellen med data fra 2021 og modellen med data fra de fem siste årene. Tilbud ble representert ved nedbørsmengde siste 30 dager, middel av middelvind samme dagen, mengde vann i magasinene i Norge, overføring til NO2 og gjennomsnittstemperatur siste 30 dager. Etterspørselen ble tatt hensyn til ved at data om overføring fra NO2 og forbruk i NO2 ble brukt.

Data om nedbørsmengde, vindstyrke og temperatur ble hentet fra Norsk Klimaservicesenter (Norsk Klimaservicesenter, u.å) i 2022 på nettsiden <https://seklima.met.no/>

Tidsoppløsningen som ble brukt var “daglig”. Målestasjonen som ble brukt for målingene av temperatur og nedbør var Sinnes i Sirdal. Grunnen til dette

er at i Sirdal kommune er det mye strømproduksjon fra vannkraft. Det var noen få datoer som manglet målinger. Dette ble løst ved å bruke lineær interpolering. Det ble også hentet noe data fra 2016 slik at vi kan se på nedbør og temperatur de siste 30 dagene for datoer i starten av 2017. 30 dager ble brukt siden mange vannkraftverk kan lagre en del vann, og derfor vil ikke nødvendigvis værendringer fra dag til dag spille inn. 30 dager er bare tilfeldig valgt, så det kan være bedre måter å løse det på. Nedbør siste 100 dager, eller siste 365 dager kan være andre gode alternativer. Nedbør siste 7 dager ble også prøvd, uten at det gjorde modellen bedre. Se R-koden i vedlegget.

Målestasjonen som ble brukt for målingene av vind er Obrestad i Hå kommune som ligger på Jæren. På Jæren er det flere store vindmølleparker, og det vil derfor være relevant å vite noe om vinden på Jæren. Her manglet det også noen få datoer, og valgte derfor å bruke interpolering her også. Antar at det vil være en sterk korrelasjon mellom temperaturen og nedbøren på Sinnes i Sirdal og vindstyrken på Obrestad på Jæren og været i resten av NO2 siden avstandene ikke er så store.

Informasjon om vannmengde i magasinene, overføring til og fra NO2 og forbruk ble også hentet fra Nord Pool, inne på samme nettside som data om strømprisene ble funnet. Dataene om fyllingsgrad til vannmagasin virket å ha noen feil i de første og de siste ukene av datasettene. Det virket som om årene hadde blitt blandet. Gjorde derfor noen endringer slik at det så mer riktig ut, og at det ble en jevn overgang fra et år til et annet. Uten endringene var det store hopp opp og ned igjen i dataene om fyllingsgraden. Kategoriene som ble brukt var daglig overføring og forbruk og ukentlig data om magasin størrelse.

Slik ser de øverste linjene av dataene ut etter at de har blitt samlet i et datasett:

```
head(daily2021)
i..SYS SE1 SE2 SE3 SE4 FI DK1 DK2 Oslo Kr.sand Bergen Molde Tr.heim TromsÅ. EE
285.86 273.38 273.38 273.38 378.47 274.18 505.50 505.50 272.60 272.60 272.60 267.89 267.89 236.81 378.47
313.20 329.82 329.82 329.82 450.92 342.71 528.10 528.10 280.41 280.41 280.41 320.49 320.49 253.47 450.92
269.14 269.28 269.28 269.28 285.23 269.75 297.11 313.06 269.28 269.28 269.28 269.28 269.28 229.56 330.46
406.74 440.48 440.48 440.48 458.35 452.89 440.48 458.35 434.61 434.61 434.61 440.48 440.48 204.81 455.05
456.62 493.48 493.48 493.48 493.48 493.48 493.42 493.48 480.22 480.22 480.22 493.48 493.48 209.59 493.48
413.62 430.00 430.00 430.00 430.00 430.00 430.00 430.00 430.00 430.00 430.00 430.00 430.00 251.35 430.00
LV LT Nummer Dato Helg Aarstid Aar Forbruk Overføring Magasin Vind Nedbor Temperatur
378.47 378.47 1 01.01.2021 0 0 0 119602 -70038 71506 3.1 196.9 -0.08666667
450.92 450.92 2 02.01.2021 1 0 0 120586 -74149 71506 2.8 196.1 -0.38666667
330.46 330.46 3 03.01.2021 1 0 0 120911 -65940 71506 2.5 194.5 -0.97666667
455.05 458.35 4 04.01.2021 0 0 0 132902 -69299 68153 3.7 187.0 -1.58666667
493.48 493.48 5 05.01.2021 0 0 0 137098 -78287 68153 3.5 186.9 -2.10000000
430.00 430.00 6 06.01.2021 0 0 0 142606 -66672 68153 2.4 186.8 -2.55000000
```

Figur 5.2: Bilde fra utskrift i RStudio ved bruk av R.

Alle verdiene som omhandler strøm har enheten MWh, bortsett fra mengden vann i magasinene som har enheten GWh. Negativ verdi på overføring betyr at NO2 eksporterer strøm. Vindstyrken er målt i m/s og nedbørmengden i mm. Nedbør står for sum nedbør siste 30 dager, og temperaturen er målt i celsius og står for gjennomsnitt siste 30 dager.

Til slutt ble det hentet data fra 2022 for å teste hvor godt modellene fungerer på framtidig data, og for å finne ut hvilken av modellene som presterer best. Data fram til og med 13. mars ble brukt siden det var så mye data jeg fikk hentet før Nord Pool endret formatet til dataene.

En mulig feilkilde kan være at dataene om vannmengde i vannmagasinene gjelder for hele landet. Det kan være variasjoner i magasin størrelsen i NO2 som ikke vises like godt når vi ser på magasin størrelsen til landet som helhet. Selv om det vil være korrelasjon mellom været forskjellige steder i NO2, så vil det ikke være 100 prosent korrelasjon, og derfor vil dette være en feilkilde. Endringene jeg har gjort i dataene vil også gjøre modellene mindre nøyaktige. I tillegg kan det være feil jeg ikke har sett, og som kan påvirke modellen.

5.2 Modeller

I denne oppgaven er det hovedsaklig brukt lineære modeller for å forklare strømprisen.

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon \quad (5.1)$$

Lineære modeller er hensiktsmessige å bruke fordi det er lett å trekke slutninger rundt tilpasset modell. Som responsvariabel er logaritmen til strømprisen brukt. Dette gjør at vi ikke kan få en negativ strømpris, og at prisen øker eksponentielt med endring i prediktorene. Dette er vanlig i økonomiske modeller, og fungerer ofte mye bedre enn hvis du ikke tar logaritmen til prisen (T. S. Kleppe, personlig kommunikasjon, 21. januar 2022). Det har hendt at strømprisen er negativ (Malkenes Hovland, 2020), men det skjer sjeldent. Det har uansett aldri skjedd at gjennomsnittlig timespris gjennom en dag er negativ. Siden dataene vi jobber med er gjennomsnitt gjennom dagen, så vil det være problemfritt å anta at strømprisen ikke kan bli negativ. Ulempen med å bruke logaritmen av responsvariabelen er at du ikke kan lese like enkelt fra modellen hvordan prediktorene påvirker responsvariabelen.

Den lineære modellen (5.1) som blir brukt blir tilpasset ved minste kvadraters metode. For at vi skal kunne gjøre dette så antar vi at residualene til modellen er normalfordelte og uavhengige av hverandre, at variansen er konstant og at det er lineær sammenheng mellom responsvariabelen og prediktorene. Koeffisientene blir, ved bruk av minste kvadraters metode, valgt slik at summen av kvadratet av residualene (RSS) blir minimert, se (5.2). Ved å derivere RSS med hensyn på betakoeffisientene og sette lik null så

finner vi koeffisientene.

$$\text{RSS} = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}))^2 \quad (5.2)$$

Det er også brukt additive modeller. Ofte er ikke relasjoner mellom variabler lineære, og det er derfor nødvendig med utvidelser av den lineære modellen til en additiv modell (5.3).

$$Y = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_p(x_p) + \epsilon \quad (5.3)$$

I denne oppgaven er det brukt smoothing spline for å prøve å fange opp ikke-lineære effekter. En smoothing spline minimerer (5.4):

$$\sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int f''(t)^2 dt, \quad (5.4)$$

“f” er høyresiden av (5.3). λ er en tuning-parameter som bestemmer hvor bølgete funksjonen skal være. λ lik uendelig vil gi en lineær modell, mens λ lik 0 vil gi en funksjon som interpolerer alle observasjonene.

Noen ganger kan det være hensiktsmessig å kjøre en “best subset” algoritme til å velge ut den beste delmengden av prediktorer til å forklare og predikere responsvariabelen. Dette er blitt gjort i denne oppgaven for å se på hva som påvirker strømprisen i løpet av året. For å finne den beste delmengden av prediktorer så må man gjøre bias-variance tradeoff (nøyaktighet-usikkerhet avveining). Målet er å få bedre prediksjoner av framtidig data, selv om du får dårligere prediksjoner av treningsdata, ved å fjerne prediktorer. Ved å fjerne prediktorer får man mer feil i prediksjonen av treningsdata, men samtidig

får man mindre usikkerhet knyttet til modellen. Mindre usikkerhet vil si at resultatet vil bli mer likt fra gang til gang hvis man får nye treningsdata. Modelleringen av beste delmengde av prediktorer skjer ved metoden forward selection. Her blir det lagt til en og en prediktor, og det er den prediktorer som minsker MSE mest som blir lagt til først. Etterpå kan man se på kriteriene BIC, adjusted R-squared og Cp for å finne antall prediktorer som er best å bruke.

$$\begin{aligned} \text{Adjusted}R^2 &= 1 - \frac{\text{RSS}/(n-d-1)}{\text{TSS}/(n-1)} \\ &= 1 - \frac{n-1}{n-d-1}(1-R^2) \end{aligned} \quad (5.5)$$

$$C_p = \frac{1}{n}(\text{RSS} + 2d\hat{\sigma}^2) \quad (5.6)$$

$$\text{BIC} = \frac{1}{n}(\text{RSS} + \log(n)d\hat{\sigma}^2) \quad (5.7)$$

“d” er antall prediktorer, “n” er størrelsen på utvalget og “TSS” betyr “total sum of squares”, summen av de kvadrerte avstandene mellom enkeltobservasjoner og gjennomsnittlig observasjonsverdi. En modell som gir høy verdi for adjusted R-squared og lav verdi for Bic og Cp er å foretrekke. Adjusted R-squared, Bic og Cp er brukt fordi de, i motsetning til R-squared, straffer modeller med mange prediktorer. Det er ønskelig med en modell med færrest mulig prediktorer, og som i størst mulig grad forklarer variansen.

Svakheten til en best subset-algoritme er at den, ved bruk av interaksjonsledd som forklaringsvariabel, kan ta bort en av prediktorene som er i bruk i interaksjonsleddet før den tar bort selve interaksjonsleddet. Dette er ikke ønskelig, og det er grunnen til at ikke best subset-algoritmen blir brukt flere

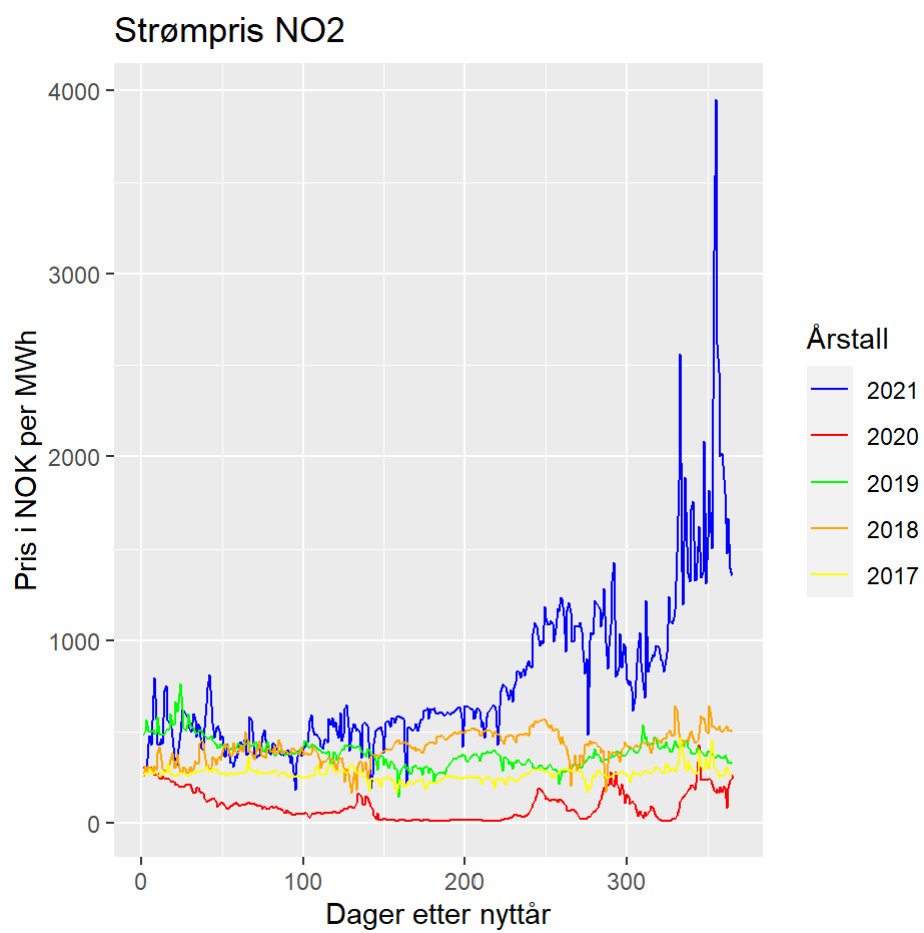
ganger. Et interaksjonsledd er en forklaringsvariabel som består av to uavhengige variabler.

5.3 Analyse av resultater

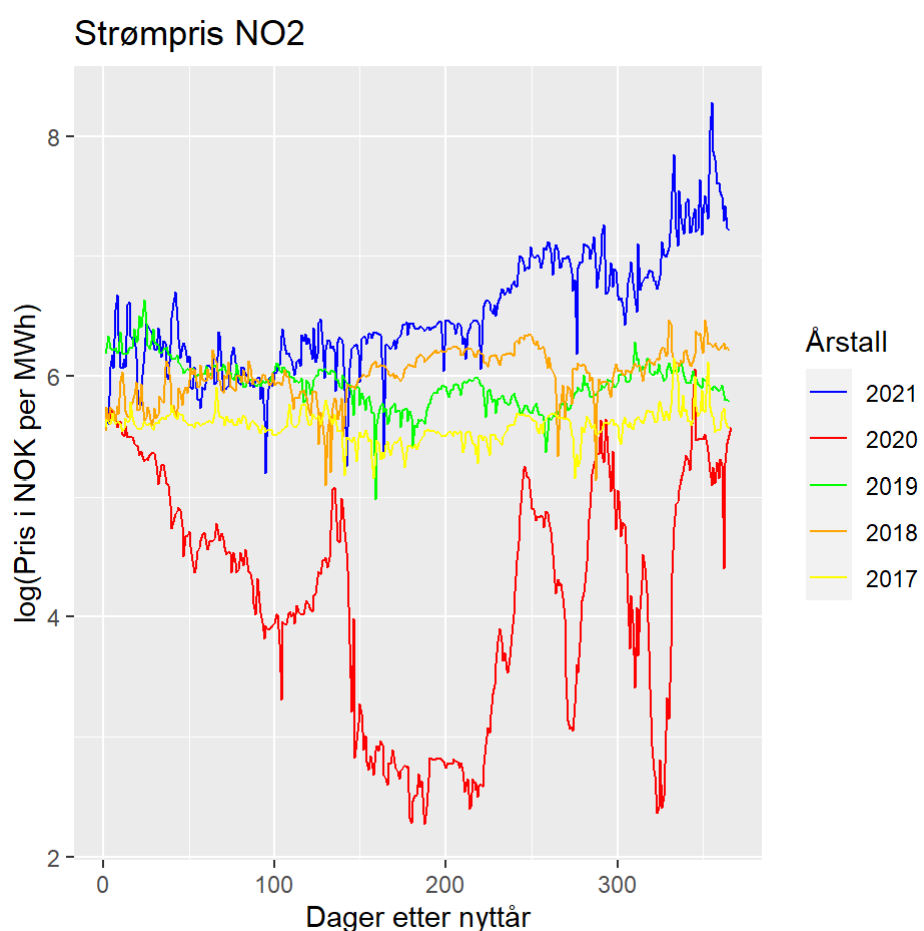
Analysen av resultatene er gjort i R-studio ved å bruke programmet R. Det er også brukt pakker i R ved hjelp av RTools. R-fila ble delt opp i to deler. En del som omhandler variasjon på grunn av tid på året, mens den andre delen skal prøve å forklare strømprisen best mulig ved hjelp av alle dataene som ble hentet inn.

Det første som ble gjort i begge deloppgavene var å laste inn dataene fra 2021. Alle dataene ble lagret som csv-filer. Csv-filer blir lest av R med kommandoen `read.csv2()`. Deretter ble de innledende modellene tilpasset ved å bruke kodene `lm()` og `gam()`. Kommandoen for `gam` må hentes fra `library(gam)`. De endelige modellene ble funnet ved å se på utskriftene til de innledende modellene, og ved bruk av `regsubsets` der det er hensiktsmessig. Denne kommandoen finner man i `library(leaps)`. For å finne informasjon om de tilpassede modellene så er kommandoen `summary()` nyttig. Her finner man blant annet p-verdien til testene H_0 : Koeffisienten til prediktoren er lik null, H_1 : Koeffisienten til prediktorene er ikke lik null i høyre kolonne og R-squared som forteller hvor mye av variansen til dataene som modellen forklarer.

Med data fra flere år vil modellene se anderledes ut. Strømprisene var mye høyere i 2021 enn de var fra 2017 til 2020, og koeffisientene vil derfor se anderledes ut siden de ikke vil spå like høye priser.



Figur 5.3: Plott av strømprisen.



Figur 5.4: Plott av logaritmen til strømprisen.

Strømprisen i 2021 var høyere enn de foregående årene. Ser også at Figur 5.4 viser at 2020 avviker fra andre år. Dette betyr at det vil være sannsynlig at en modell, med data fra flere år tilbake i tid, ikke vil være like god til å forutse hvordan strømprisen vil bli framover på grunn av endrede betingelser. Dette gjelder spesielt for modellene som ser på variasjon i løpet året, modellene som har mer info har potensial til å forklare noen av de endrede betingelsene. Dette er derimot ikke noe vi kan vite, og derfor vil det være hensiktsmessig å lage flere modeller og teste dem på et valideringssett (data fra 2022) for å finne

ut hvilken modell som er best. Dette gjøres ved kommandoen “`predict()`” og ved å se på MSE til de forskjellige prediksjonene.

Har i koden også valgt å lage modeller med data fra de 2 siste årene, de 5 siste årene og med data fra 2017 til 2020. Se vedlagt R-kode for detaljer. Har også laget modeller med årstall som kategorisk prediktor, og modeller uten. Fordelen med årstall som kategorisk prediktor er at inne i årstallene kan det ligge mye informasjon om endrede betingelser, slik at modellen forklarer mer av variansen totalt sett. Ulempen er at det ikke går an å teste modellen på framtidige år.

Når vi tilpasser dataene, antar vi at residualene til modellen med logaritmen til strømprisen som prediktor er normalfordelte. Dette kan vi sjekke om stemmer ved qq-plott, histogram og tidsserieplott av residualene. Qq-plottet skal være nærmest mulig ei linje, histogrammet skal indikere normalfordeling og det skal ikke være noe mønster i tidsserieplottet. Dette er en subjektiv vurdering (J. T. Kvaløy, personlig kommunikasjon, 2020). Vi bør også sjekke om det er korrelasjon mellom prediktorene. Dette sjekkes ved å se om VIF er større enn 10. Dette gjøres i R ved å bruke kommandoen “`vif()`” som ligger i “`library(car)`” og sjekke om “GVIF opphøyd i $(1/(2*Df))$ ” ganget med seg selv blir større enn 10 (J. T. Kvaløy, personlig kommunikasjon, 4. mars 2022). En annen måte å se etter korrelasjon mellom prediktorer er å bruke kommandoen `ggpairs()` som ligger i `library(GGally)`. Denne kommandoen plotter prediktorer mot hverandre, og vil også kunne undersøke relasjon mellom respons og hver enkelt prediktor.

Til slutt brukes den beste egnede modellen til å se på hvordan to aktuelle eksempler, elektrifisering av norsk sokkel (Nyhus og Siem, 2022) og North Sea Link (Viseth, 2021), kan påvirke strømprisen. Elektrifiseringen av norsk

sokkel er et av de viktigste klimatiltakene vi kan gjøre, men det vil føre til at det norske strømforbruket vil stige med 10 prosent. North Sea Link er en overføringskabel fra Norge til England. Den ble delvis satt i drift i slutten av 2021 og bidrar til robustheten til det europeiske kraftmarkedet. Norge tjener også store penger på den, men den vil føre til dyrere strøm siden den vil øke etterspørselen etter strøm. Overføringen av strøm kan øke med opp til 50 prosent. Dette kommer man fram til ved å se på tall på hvor mye kablen kan overføre (Viseth, 2021), og ved å se på historiske data for overføring.

Kapittel 6

Analyse

6.1 Variasjon i strømprisen i løpet av året

Et interessant perspektiv vil være å se på hvordan strømprisen varierer gjennom året. Dette ville vært nyttig å vite for et lite vannkraftverk med en produksjon som er så liten at den ikke påvirker strømprisen av betydning. Denne informasjonen vil vannkraftverket kunne bruke til å lage en modell av når de bør selge strøm. Alle bildene i analysedelen er hentet fra utskriftene til R-koden som ligger i vedlegget.

6.1.1 Resultater og diskusjon

Den første forklaringsmodellen som ble prøvd med logaritmen til strømprisen som responsvariabel var en lineær modell med data fra 2021 og med interaksjonsledd som vist i Figur 6.1. Vi kan se på F-testen at det er relasjon

mellom responsen og prediktorene. En F-test er en hypotesetest der H_0 er at alle prediktorene er 0, og H_1 er at minst en av prediktorene er forskjellig fra null. Vi kan også se på p-verdiene at det vil være fordelaktig å forenkle modellene. Fellesferie, helligdag og interaksjonsleddene ser spesielt utsatte ut med høye p-verdier.

```

      Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.666112   0.047054 141.669 < 2e-16 ***
Helg         -0.212037   0.087510  -2.423  0.01589 *
Helligdag    0.007257   0.137050   0.053  0.95780
Aarstid1    -0.532385   0.066483  -8.008  1.69e-14 ***
Aarstid2    -0.193371   0.069601  -2.778  0.00576 **
Aarstid3     0.276917   0.066254   4.180  3.68e-05 ***
Fellesferie -0.054153   0.097173  -0.557  0.57769
Helg:Aarstid1 0.051379   0.123683   0.415  0.67809
Helg:Aarstid2 0.095396   0.123654   0.771  0.44094
Helg:Aarstid3 0.138014   0.123581   1.117  0.26484
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.376 on 355 degrees of freedom
Multiple R-squared:  0.41,    Adjusted R-squared:  0.395
F-statistic: 27.4 on 9 and 355 DF,  p-value: < 2.2e-16

```

Figur 6.1: Koeffisientene til prediktorene til den innledende modellen.

Den beste lineære modellen til å forklare variasjon i strømprisen i løpet av året er vist nedenfor i Figur 6.2. Disse fire prediktorene ble valgt siden det var 4 prediktorer som ga best score blant kriteriene som måler hvilken modell som er best, og det var disse 4 som best subset-algoritmen plukket ut. Det er også nokså stor forskjell mellom 3 og 4 prediktorer, så det virker ikke som om det hadde vært noe å hente ved å forenkle modellen. Dette tyder på at strømprisen er forskjellig i de forskjellige årstidene og at det er forskjell på strømprisen i ukedagene og i helgene. Vi kan ikke konkludere med at fellesferien og helligdager har betydning for strømprisen. Plott av scoren til tre forskjellige kriterier er vist i Figur 6.3 nedenfor.

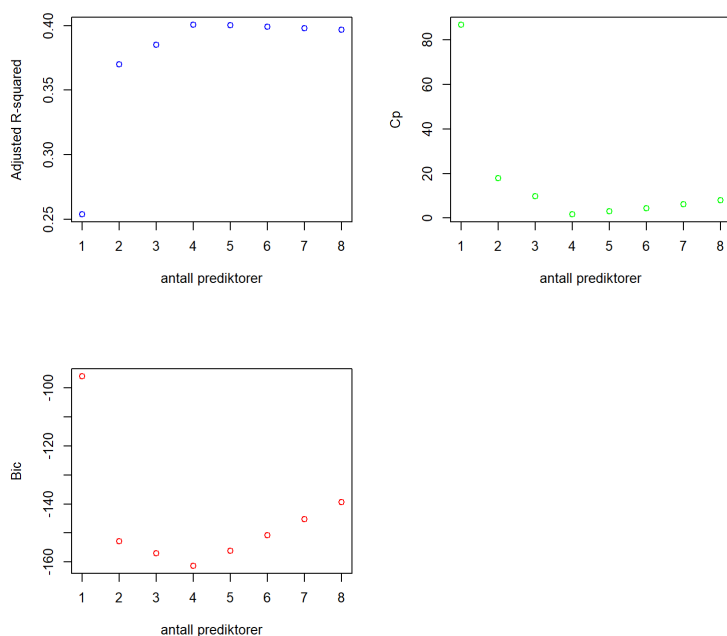
```

---
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.64540    0.04140 160.509 < 2e-16 ***
Helg         -0.13979    0.04341  -3.220  0.00140 **
Aarstid1     -0.51710    0.05550  -9.318 < 2e-16 ***
Aarstid2     -0.17730    0.05550  -3.195  0.00152 **
Aarstid3      0.31642    0.05565   5.686  2.69e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3743 on 360 degrees of freedom
Multiple R-squared:  0.4071,    Adjusted R-squared:  0.4005
F-statistic: 61.8 on 4 and 360 DF,  p-value: < 2.2e-16

```

Figur 6.2: Koeffisientene til prediktorene til den beste modellen.



Figur 6.3: Plott av score til kriterier for beste modell.

De to modellene som ble tilpasset med data fra 2017 til 2021 er vist nedenfor i Figur 6.4 og 6.5. 6.4 viser modellen med årstall som kategorisk prediktor. Denne modellen forklarer en større andel av variansen og har lavere p-verdier enn modellen som ikke har med år som prediktor. Det er tydelig at strømprisen varierer fra år til år siden årene er signifikante. Modellene

for strømprisen med data fra 2020-2021 og 2017-2020 kan finnes i R-koden i vedlegget.

```

      Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.80209    0.03188 213.338 < 2e-16 ***
Helg         -0.08862    0.02428  -3.649 0.00027 ***
Aarstid1     -0.35768    0.03105 -11.519 < 2e-16 ***
Aarstid2     -0.55320    0.03105 -17.817 < 2e-16 ***
Aarstid3     -0.15166    0.03113  -4.871 1.21e-06 ***
AarF1        -2.34086    0.03466 -67.536 < 2e-16 ***
AarF2        -0.56976    0.03468 -16.427 < 2e-16 ***
AarF3        -0.50051    0.03468 -14.430 < 2e-16 ***
AarF4        -0.92267    0.03468 -26.602 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4686 on 1817 degrees of freedom
Multiple R-squared:  0.7556, Adjusted R-squared:  0.7546
F-statistic: 702.3 on 8 and 1817 DF, p-value: < 2.2e-16

```

Figur 6.4: Koeffisientene til modell med data fra 2017-2021.

```

      Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.93178    0.04556 130.209 < 2e-16 ***
Helg         -0.08763    0.04782  -1.832  0.0670 .
Aarstid1     -0.35441    0.06115  -5.796 8.01e-09 ***
Aarstid2     -0.54993    0.06115  -8.993 < 2e-16 ***
Aarstid3     -0.14839    0.06132  -2.420  0.0156 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

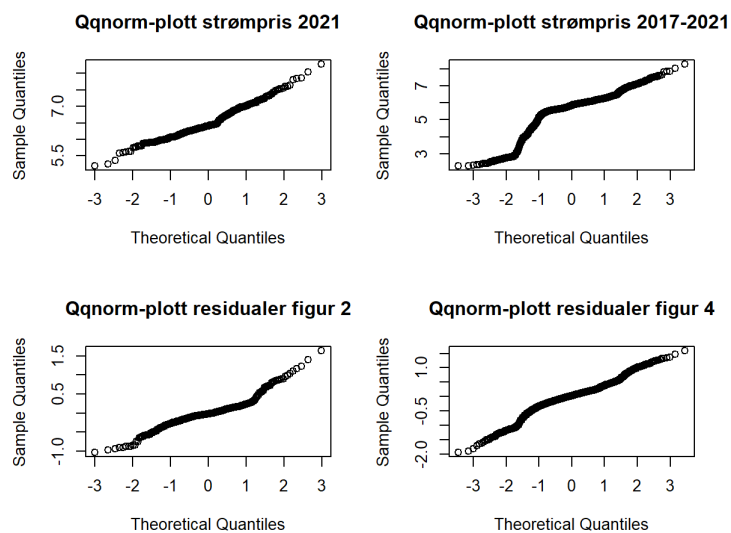
Residual standard error: 0.9228 on 1821 degrees of freedom
Multiple R-squared:  0.05005, Adjusted R-squared:  0.04796
F-statistic: 23.99 on 4 and 1821 DF, p-value: < 2.2e-16

```

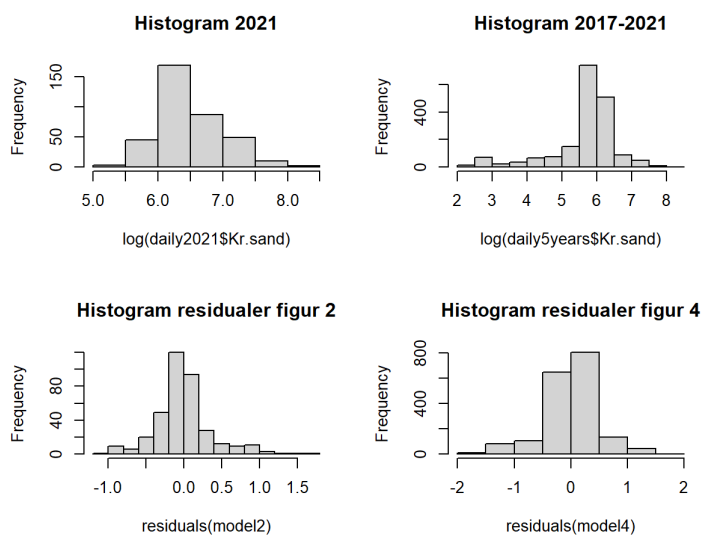
Figur 6.5: Koeffisientene til modell med data fra 2017-2021.

Er det ok å anta at logaritmen til strømprisen og residualene til modellene er normalfordelte? Plottene nedenfor viser at dataene ser ut til å være nokså normalfordelte, men det er problemer. Det ser ut som om strømprisene ikke er uavhengige når vi ser på tidsserieplottet av dataene. Dette er ikke forenelig med normalfordeling. Qq-plottene og histogrammene indikerer at dataene er

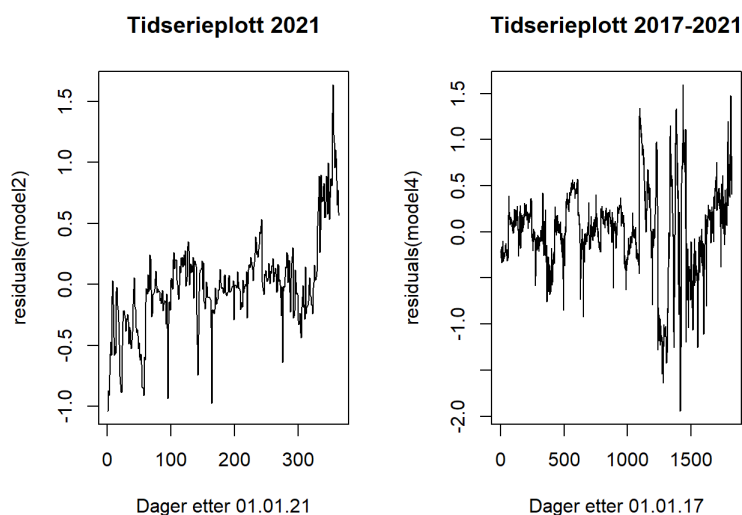
noenlunde normalfordelte.



Figur 6.6: Qq-plott. Skal ideelt sett være en rett linje.



Figur 6.7: Histogram. Skal ideelt sett være lik kurven til normalfordelingen.

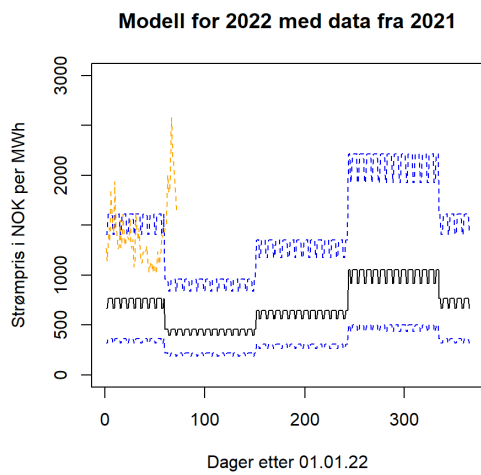


Figur 6.8: Tidsserieplott. Skal ideelt sett ikke vise noe mønster.

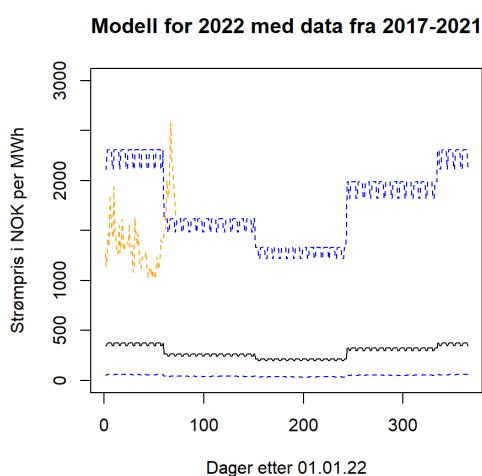
6.1.2 Hva ville dette betydd for et lite vannkraftverk?

Figurene 6.9 og 6.10 viser en prediksjon av strømprisen for 2022 med data fra 2021 og 2017-2021. De blå linjene som er tegnet inn viser et 95 prosent-prediksjonsintervall for prediksjonene. Det vil si at, gitt at betingelsene for 2022 er uendret, så vil 95 prosent av målingene falle innenfor prediksjonsintervallet. Det er større avstand til øvre grense enn til nedre grense, og dette kommer av at logaritmen til strømprisen er brukt som respons. Strømprisen blir predikert til å være høyere med data fra 2021, noe som er naturlig siden prisnivået var høyere i 2021 enn resten av årene. Strømprisen er høyere enn øvre grense i prediksjonsintervallene til begge modellene når våren kommer.

Dette betyr at det har skjedd endringer i strømmarkedet. Krigen i Ukraina er en av hendelsene som har ført til endringer.



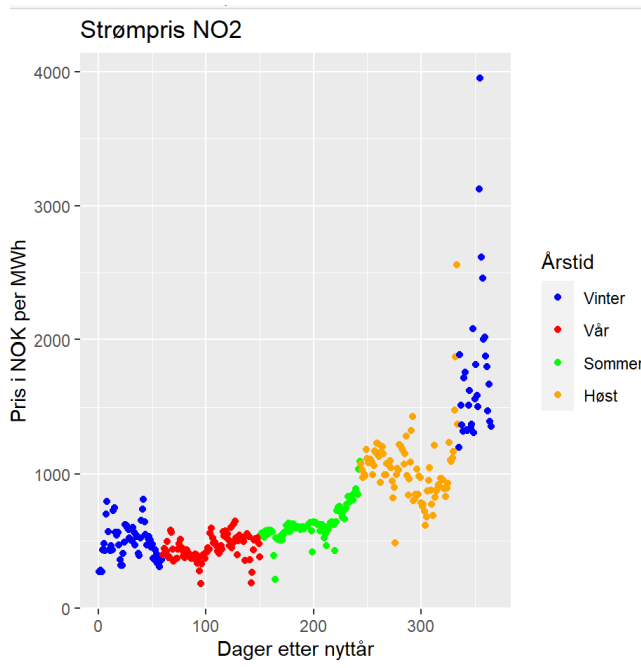
Figur 6.9: Plott av prediksjon med data fra 2021. Gul linje er hva strømprisen var i 2022 fram til og med 13.mars. Svart linje er prediksjon. Blå linjer viser prediksjonsintervallet.



Figur 6.10: Plott av prediksjon med data fra 2017-2021. Gul linje er hva strømprisen var i 2022 fram til og med 13.mars. Svart linje er prediksjon. Blå linjer viser prediksjonsintervallet.

Dette vil altså si at ifølge figur 6.9 så vil det være mest å tjene på å produsere og selge strømmen på høsten, og minst å tjene ved å selge strømmen på våren. Vi kan også, ved se på figur 6.2, si at strømmen er dyrere på hverdager. Derfor er linjene bølgete. Det kan være grunn til å stille spørsmål ved denne konklusjonen. Plottet i Figur 6.11 viser hvordan strømprisen utviklet seg i 2021. Det ser ut som om strømprisen vokser (eksponentielt) jo lengre ut i året vi kommer. Dette er ikke blitt tatt hensyn til i modellen, som antar at residualene er uavhengige, og man bør derfor være forsiktig med å trekke konklusjoner. 6.10 viser at strømprisen er dyrest om vinteren og billigst om sommeren. Dette kan kanskje være mer nærliggende å tro. Begge plottene viser imidlertid at strømprisen er høyere om høsten og vinteren enn sommeren og våren. Vi trenger flere prediktorer for å kunne spå strømprisen bedre, og

for å kunne spå hvorfor strømprisen steg gjennom 2021.



Figur 6.11: Plott av strømprisen i NO2 i 2021.

6.2 Forklare variasjon i strømprisen best mulig

En modell med informasjon om vær og tilbud og etterspørsel etter strøm som prediktorer vil potensielt kunne gi en modell som både forklarer en større andel av variasjonen i strømprisen og som gir bedre prediksjonen av strømprisen i nær framtid, ved å se på værmeldinger og andre prediksjonen av variabler.

6.2.1 Resultater og diskusjon

Den første modellen som ble tilpasset med de innsamlede dataene fra 2021 med logaritmen til strømprisen som responsvariabel vises under i Figur 6.12. F-testen viser at vi kan konkludere med at det er sammenheng mellom responsen og prediktorene. P-verdiene viser at det ikke kan konkluderes med at det trengs informasjon om helg for å forklare strømprisen. Det vil si at effekten av helg kan bli forklart ved de andre variablene. Det vil være nærliggende å tenke at forbruket blir påvirket av om det er helg eller ikke. Interaksjonsleddet mellom temperaturen og sommeren (Aarstid2) har også en nokså høy verdi, og dette kan tyde på at effekten av temperaturen er lik for sommer og vinter (grunnivået).

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.452e+00  3.230e-01  26.170 < 2e-16 ***
Helg         -6.540e-03  3.159e-02  -0.207  0.83613
Aarstid1     -1.750e+00  9.330e-02 -18.756 < 2e-16 ***
Aarstid2     -2.236e+00  1.973e-01 -11.330 < 2e-16 ***
Aarstid3      6.885e-01  1.194e-01   5.768  1.77e-08 ***
Overforing   -3.372e-06  5.594e-07  -6.027  4.22e-09 ***
Vind         -1.121e-02  4.242e-03  -2.643  0.00859 **
Nedbor       -2.467e-03  4.463e-04  -5.529  6.30e-08 ***
Temperatur   1.285e-01  7.426e-03  17.305 < 2e-16 ***
Forbruk      1.027e-05  2.304e-06   4.457  1.12e-05 ***
Magasin      -4.214e-05  4.319e-06  -9.758 < 2e-16 ***
Aarstid1:Temperatur -1.268e-01  1.324e-02  -9.578 < 2e-16 ***
Aarstid2:Temperatur  2.809e-02  1.468e-02   1.914  0.05647 .
Aarstid3:Temperatur -1.808e-01  1.562e-02 -11.572 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2399 on 351 degrees of freedom
Multiple R-squared:  0.7626,    Adjusted R-squared:  0.7538
F-statistic: 86.72 on 13 and 351 DF,  p-value: < 2.2e-16

```

Figur 6.12: Sammendrag av innledende modell.

Ved å fjerne helg som prediktor, siden koeffisienten ikke er signifikant, får vi den endelige lineære modellen for prediksjon av strømprisen med data fra 2021 som vist i Figur 6.13.

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.433e+00  3.089e-01  27.302 < 2e-16 ***
Aarstid1    -1.749e+00  9.297e-02 -18.808 < 2e-16 ***
Aarstid2    -2.232e+00  1.964e-01 -11.368 < 2e-16 ***
Aarstid3     6.898e-01  1.190e-01  5.795 1.52e-08 ***
Overforing  -3.400e-06  5.409e-07  -6.287 9.61e-10 ***
Vind        -1.109e-02  4.199e-03  -2.642  0.0086 **
Nedbor      -2.472e-03  4.452e-04  -5.553 5.55e-08 ***
Temperatur   1.287e-01  7.364e-03  17.476 < 2e-16 ***
Forbruk      1.042e-05  2.179e-06   4.782 2.55e-06 ***
Magasin     -4.221e-05  4.302e-06  -9.812 < 2e-16 ***
Aarstid1:Temperatur -1.266e-01  1.316e-02  -9.617 < 2e-16 ***
Aarstid2:Temperatur  2.817e-02  1.465e-02   1.923  0.0553 .
Aarstid3:Temperatur -1.806e-01  1.556e-02 -11.601 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2396 on 352 degrees of freedom
Multiple R-squared:  0.7625, Adjusted R-squared:  0.7545
F-statistic: 94.2 on 12 and 352 DF, p-value: < 2.2e-16

```

Figur 6.13: Sammendrag av endelig modell.

Med data fra de siste fem årene, blir den endelige modellen for strømprisen med årstall som kategorisk variabel som vist nedenfor i Figur 6.14. Helg er fjernet fra den innledende modellen. Informasjonen om nedbør siste 30 dager kunne kanskje også med fordel blitt fjernet, men velger å beholde den siden den ble beholdt i modellen med data fra 2021. R-squared er omtrent like stor for begge de endelige modellene. R-squared er 0.81 og 0.76 for modellene med data fra henholdsvis 2017-2021 og 2021. Modellen med data fra 2017-2021, men uten årstall som kategorisk variabel er vist nedenfor i 6.15. Her er R-squared mye lavere enn for de andre modellene. Det er noen betingelser som har endret seg gjennom årene som denne modellen ikke kan forklare.

```

                Estimate Std. Error t value Pr(>|t|)
(Intercept)    6.842e+00  2.217e-01  30.868 < 2e-16 ***
Aarstid1      -8.320e-01  4.904e-02 -16.965 < 2e-16 ***
Aarstid2      -1.556e+00  1.235e-01 -12.594 < 2e-16 ***
Aarstid3      -3.269e-01  5.748e-02  -5.687 1.51e-08 ***
Overforing    -2.897e-06  4.973e-07  -5.825 6.77e-09 ***
Vind          -9.116e-03  3.202e-03  -2.847 0.00446 **
Nedbor        -1.089e-04  1.627e-04  -0.669 0.50328
Forbruk       8.678e-06  1.854e-06   4.680 3.09e-06 ***
Magasin       -1.607e-05  1.463e-06 -10.986 < 2e-16 ***
Temperatur    1.344e-01  7.836e-03  17.152 < 2e-16 ***
AarF1         -2.329e+00  3.694e-02 -63.044 < 2e-16 ***
AarF2         -5.091e-01  3.568e-02 -14.269 < 2e-16 ***
AarF3         -5.982e-01  3.459e-02 -17.291 < 2e-16 ***
AarF4         -9.056e-01  3.689e-02 -24.552 < 2e-16 ***
Aarstid1:Temperatur -1.046e-01  9.736e-03 -10.746 < 2e-16 ***
Aarstid2:Temperatur -3.305e-02  1.182e-02  -2.796 0.00523 **
Aarstid3:Temperatur -8.944e-02  9.890e-03  -9.044 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.416 on 1776 degrees of freedom
(33 observations deleted due to missingness)
Multiple R-squared:  0.8082,    Adjusted R-squared:  0.8065
F-statistic: 467.8 on 16 and 1776 DF,  p-value: < 2.2e-16

```

Figur 6.14: Sammendrag av endelig modell.

```

                Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.756e+00  3.752e-01  12.674 < 2e-16 ***
Helg           7.013e-02  4.596e-02   1.526 0.12724
Aarstid1      -1.110e+00  8.761e-02 -12.669 < 2e-16 ***
Aarstid2      -1.573e+00  2.276e-01  -6.912 6.65e-12 ***
Aarstid3       5.290e-01  1.051e-01   5.033 5.31e-07 ***
Overforing    2.502e-06  8.795e-07   2.844 0.00450 **
Vind          -1.592e-02  6.040e-03  -2.636 0.00845 **
Nedbor        -2.140e-03  2.945e-04  -7.266 5.51e-13 ***
Forbruk       3.762e-05  3.273e-06  11.494 < 2e-16 ***
Magasin       -4.786e-05  2.530e-06 -18.918 < 2e-16 ***
Temperatur    9.577e-02  1.474e-02   6.496 1.07e-10 ***
Aarstid1:Temperatur -2.198e-03  1.794e-02  -0.123 0.90250
Aarstid2:Temperatur 1.011e-01  2.169e-02   4.661 3.38e-06 ***
Aarstid3:Temperatur 3.194e-03  1.790e-02   0.178 0.85836
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7853 on 1779 degrees of freedom
(33 observations deleted due to missingness)
Multiple R-squared:  0.3152,    Adjusted R-squared:  0.3102
F-statistic: 62.99 on 13 and 1779 DF,  p-value: < 2.2e-16

```

Figur 6.15: Sammendrag av endelig modell.

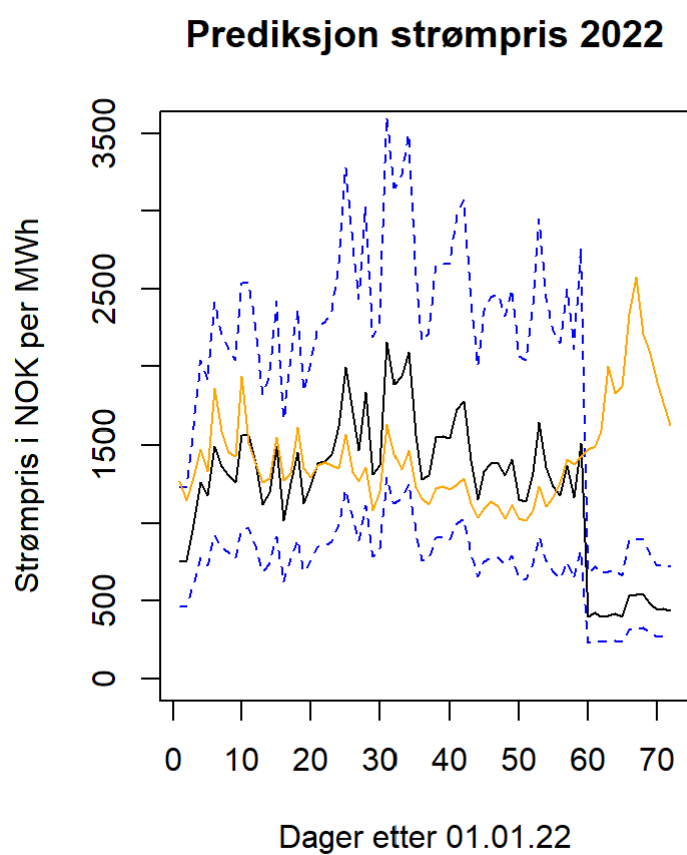
Vi har nå flere forskjellige modeller, i tillegg til to additive modeller med data fra 2021 og 2017 til 2021, som det er ønskelig å teste på data fra 2022. Tabellen nedenfor i Figur 6.16 viser at modellene med data fra 2021 presterer best. Den lineære modellen med mest mulig data fra 2021 presterer aller best.

Den presterer også bedre enn de additive modellene, noe som tyder på at det er lineær sammenheng mellom logaritmen til strømprisen og prediktorene. I 6.17 vises prediksjon av strømprisen i 2022 ved bruk av modell som ble tilpasset med data fra 2021.

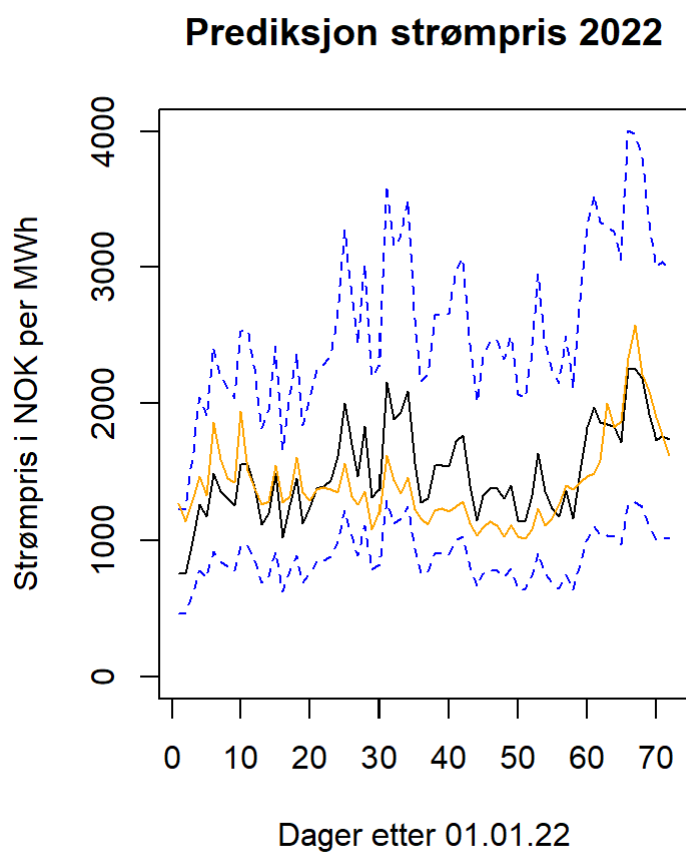
Når det blir vår så avviker modellen. Dette kan forklares med Figur 6.11, strømprisen steg gjennom 2021, og derfor endret betingelsene seg fra starten til slutten av året. Ser at prediksjonene kan bli bedre om våren får samme prisnivå som vinteren. 6.18 viser prediksjon av strømprisen der alle datoene blir regnet som vinter. Dette stemmer mye bedre, kanskje har årstidene mindre å si i 2022? Trenger mer data for å konkludere.

Modell	MSE
Figur 6.2	537919.19
Figur 6.5	1146604.28
Figur 6.13	32863.24
Figur 6.15	736386.23
Gam 2021	52937.35
Gam 2017-2021	1264968.22

Figur 6.16: MSE til noen av modellene som ble tilpasset.



Figur 6.17: Prediksjon av strømprisen. Svart linje = prediksjon, gul linje = faktisk strømpris og blå linjer = prediksjonsintervall. Prediksjonsprisen faller når det blir vår, rundt dag 60 etter nyttår.



Figur 6.18: Prediksjon av strømprisen. Svart linje = prediksjon, gul linje = faktisk strømpris og blå linjer = prediksjonsintervall. Predikerer at prisen er uavhengig av årstid.

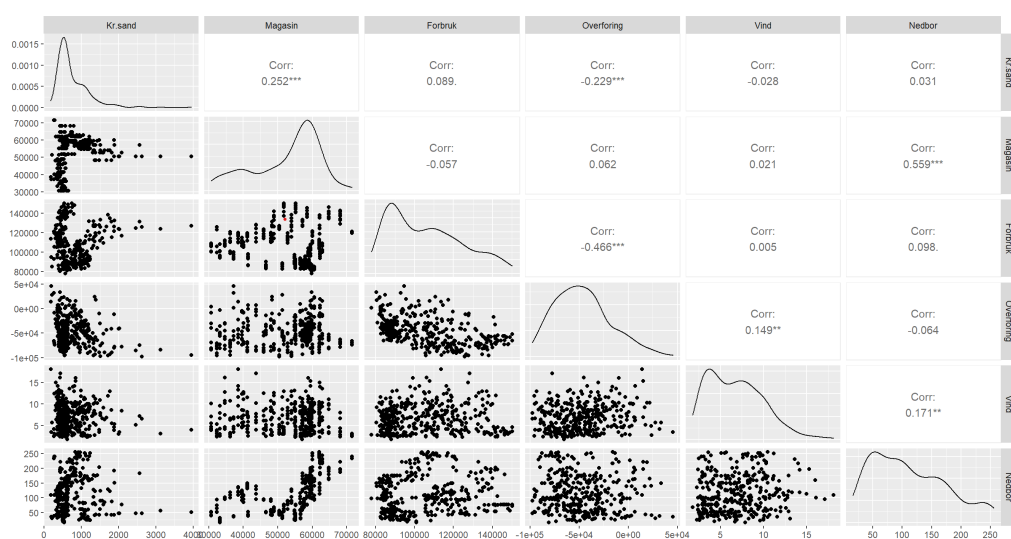
Figur 6.19 viser VIF-faktorene, og at det er korrelasjon mellom prediktorene. Dette er problematisk. Spesielt temperatur har en høy VIF-faktor. Temperaturen vil være avhengig av hvilken årstid det er, og forbruket og temperaturen vil være korrelerte siden det vil bli større etterspørsel etter energi for å lage varme når det er kaldt. 6.20 viser også at det er problemer.


```

> vif(modell12)
              GVIF Df GVIF^(1/(2*Df))
Aarstid      3535.910885 3      3.903165
Overforing   1.481412  1      1.217133
Vind         1.130415  1      1.063210
Nedbor       4.974928  1      2.230455
Temperatur   19.846781  1      4.454973
Forbruk      10.457708  1      3.233838
Magasin      10.894695  1      3.300711
Aarstid:Temperatur 1278.198924 3      3.294325
> vif(modell14)
              GVIF Df GVIF^(1/(2*Df))
Aarstid      501.403593 3      2.818586
Overforing   1.990554  1      1.410870
Vind         1.091305  1      1.044656
Nedbor       1.817905  1      1.348297
Forbruk      10.839019  1      3.292267
Magasin      5.348655  1      2.312716
Temperatur   25.630806  1      5.062688
AarF         2.578190  4      1.125678
Aarstid:Temperatur 720.780934 3      2.994336
> vif(gam1)
              GVIF Df GVIF^(1/(2*Df))
Helg         1.276229  1      1.129703
Aarstid      48.761321  3      1.911375
s(Overforing) 1.507405  1      1.227764
s(Vind)      1.127417  1      1.061799
s(Nedbor)    2.713247  1      1.647194
s(Temperatur) 12.370091  1      3.517114
s(Forbruk)   11.368788  1      3.371763
s(Magasin)   9.854709  1      3.139221
> vif(gam2)
              GVIF Df GVIF^(1/(2*Df))
Helg         1.201887  1      1.096306
Aarstid      15.269290  3      1.575082
s(Overforing) 1.718711  1      1.310996
s(Vind)      1.083281  1      1.040808
s(Nedbor)    1.492972  1      1.221872
s(Temperatur) 10.182987  1      3.191079
s(Forbruk)   7.977340  1      2.824418
s(Magasin)   4.429331  1      2.104598

```

Figur 6.19: VIF-faktor til 6.13 og 6.14 og begge de additive modellene.



Figur 6.20: Plott av noen av variablene mot hverandre. Variablene plottet mot seg selv viser relativ frekvens.

6.2.2 Hva ville dette betydd for et lite vannkraftverk?

Med gode prediksjoner for overføring, vær og forbruk så vil et vannkraftverk kunne ta avgjørelser rundt salg av strøm i nær framtid. Lengre framover går det ikke an å lage mye bedre prediksjoner enn modellene i del 1, så lenge det ikke skjer noen betydelige endringer i kraftmarkedet.

Modellene i del 2 kan også bli brukt til å si noe om hva som vil skje med strømprisen om det kommer en endring i en av prediktorene. Et anslag for prosentvis økning i strømprisen ved elektrifisering av norsk sokkel og etter at kabelen til Storbritannia åpnet er vist nedenfor i Figur 6.21. 10 prosent høyere forbruk vil, ifølge den beste modellen i denne oppgaven, gi en økning i strømprisen på 12 prosent. 50 prosent økning i overføring vil gi en økning i strømprisen på 8 prosent. Analyser gjort for nesten 10 år siden viser det

samme, strømprisen forventes å bli nesten 10 prosent dyrere etter åpningen av kabel til England (Lie, 2014). Forbruket i NO2 er totalt sett større enn netto overføring ut av NO2, og derfor vil en prosentvis endring i forbruket ha mer å si enn en prosentvis endring i overføring ut av NO2.

```
> faktorOverføring^1.5/faktorOverføring*100-100  
[1] 8.249688  
> faktorForbruk^1.1/faktorForbruk*100-100  
[1] 11.6667
```

Figur 6.21: Estimert økning i strømprisen i prosent.

For et vannkraftverk, så vil dette bety at det kan være lurt å vente med å selge mest mulig strøm til etter at en utenlandskabel har åpnet, eller til etter oljeplattformene blir elektrifisert. Planene om nye utenlandskabler og elektrifiseringen av norsk sokkel ligger trolig et stykke fram i tid på grunn av de høye strømprisene, men forbruket og overføringskapasiteten vil variere. Kabler blir midlertidig stengt, og dette vil kunne påvirke strømprisen. Bli for eksempel North Sea Link stengt så spår modellene at prisen på strøm vil synke.

Kapittel 7

Konklusjon

I del 1 så var konklusjonen at prisen på strøm i NO₂ er lavere i helgene, og at den varierer mellom årstidene. Forholdet mellom årstidene er forskjellig med forskjellige datagrunnlag, men strømprisen virker å være høyere når det er vinter eller høst enn når det er vår eller sommer (se Figur 6.9 og 6.10). Den beste modellen til å forklare strømprisen ble funnet ved hjelp av data fra 1. januar 2022 til og med 13. mars 2022. Dette viste seg å være en lineær modell med logaritmen til strømprisen som respons og prediktorer som sier noe om forbruk, overføring, magasiner, nedbør, årstid, vind og temperatur. Helg ble også prøvd som prediktor, men ble ikke tatt med siden koeffisienten ikke var signifikant. Likningen til den endelige modellen vil være på formen til (5.1) der “Y” er logaritmen til strømprisen, x-ene er verdien til de forskjellige prediktorene og der betaene er koeffisientene vist i kolonnen “Estimate” til Figur 6.13.

Overføres det mer strøm ut av NO₂ så blir prisen på strøm dyrere (og hvis det overføres til NO₂ så blir prisen lavere), men en prosentvis endring i forbruket

gir en større endring i prisen siden forbruket står for en større andel av etterspørselen enn det overføringen gjør. Elektrifisering av norsk sokkel vil gi rundt 10 prosent høyere forbruk, og dette vil gi en økning i strømprisen på 12 prosent, ifølge den beste modellen. Den beste modellen spår også at en 50 prosent økning i overføring ut av NO₂, som den nye kabelen til England potensielt kan gi, vil gi en økning i strømprisen på 8 prosent.

Dette vil si at, for et vannkraftverk, så vil det vil være lurt å selge mest mulig strøm når det ikke er helg, når det er vinter og høst og mens forbruket og overføringen til utlandet er høyest mulig. For å kunne predikere strømprisen i framtiden best mulig, så vil det være nødvendig med prediksjoner av forbruk, overføring, magasin størrelse og vær. En annen utvidelse av oppgaven, som kunne vært interessant å se på, ville vært å prøve og forklare mer av variansen med flere prediktorer. Spesielt endringene fra år til år sliter modellene i denne oppgaven med å forklare. Med data fra 2017 til 2021 ble modellen mye bedre med årstall som kategorisk variabel, men det er nærliggende å tro at det ligger noe mer bak endringene enn at årstallet blir annerledes. Ved å se på plott og sum av variabler (se R-koden i vedlegget), så er det vanskelig å konkludere med hva som har gjort strømprisen så dyr i NO₂ i 2021. Det virker å være en blanding av mindre nedbør, mindre vind, lavere nivå på magasinene, mer overføring til andre prisområder og høyere forbruk som har ført til høyere strømpriser i 2021 enn tidligere.

Referanser

- Abrahamsen, K. L. (2014). *Elektrisitetspriser – En empirisk og teoretisk analyse av tilbud og etterspørsel* (Masteroppgave). UiO. <https://www.duo.uio.no/handle/10852/40972>
- Askheim, L. O. (2019, 30. desember). Nord Pool. *Store norske leksikon*. http://snl.no/Nord_Pool
- Askheim, L. O. & Brænd, T. J. (2022, 30. mars). Energiloven. *Store norske leksikon*. <http://snl.no/energiloven>
- Esparza, H. Ø. (2019, 4. desember). *Sjekk om du har et gjennomsnittlig strømforbruk* [strøm]. <https://xn--strm-ira.no/gjennomsnittlig-str%C3%B8mforbruk>
- Haugom, E., Hoff, G. A., Molnár, P., Mortensen, M. & Westgaard, S. (2018). The Forward Premium in the Nord Pool Power Market. *Emerging Markets Finance and Trade*, 54(8), 1793–1807. <https://doi.org/https://doi.org/10.1080/1540496X.2018.1441021>

- Kaldestad, Ø. H. (2021, 15. desember). *Bør du velge fastpris eller spotpris?* [forbrukerradet]. <https://www.forbrukerradet.no/siste-nytt/bor-du-velge-fastpris-eller-spotpris/>
- Lie, Ø. (2014, 20. august). *Norges kraftoverskudd forsvinner til England* [Teknisk Ukeblad]. <https://www.tu.no/artikler/norges-kraftoverskudd-forsvinner-til-england/230246>
- Magnussen, K. (2021). *Her er alle Norges utenlandskabler* [Illustrasjon]. <https://www.tu.no/artikler/her-er-alle-norges-utenlandskabler/513908>
- Malkenes Hovland, K. (2020). I natt fikk du penger for å bruke strøm: – Andre gang i historien. *E-24*. <https://e24.no/i/VqqaJr>
- Malkenes Hovland, K. (2022). NVE-sjefen advarer: Frykter dyr strøm ut hele året. *E-24*. <https://e24.no/i/mrEMWO>
- Nord Pool. (u.å.). *Historical Market Data*. Hentet 26. januar 2022, fra <https://www.nordpoolgroup.com/historical-market-data/>
- Nord Pool. (2022). *See market data for all areas* [Illustrasjon]. <https://www.nordpoolgroup.com/en/Market-data1/>
- Norsk Klimaservicesenter. (u.å.). *Observasjoner og værstatistikk* [seklima.met]. Hentet 3. mars 2022, fra <https://seklima.met.no/>
- Nyhus, H. & Siem, B. (2022). Forslag om å stanse elektrifiseringa av sokkelen opp i Stortinget. *NRK*. <https://www.nrk.no/vestland/forslag-om-a-stanse-elektrifiseringa-av-sokkelen-opp-i-stortinget-1.15794338>

Næringslivsmagasinet Rosenkilden. (2022). Kraftkrisen ryster næringslivet.

<https://www.naeringsforeningen.no/magasin/2022-01/index.html>

Olje- og energidepartementet. (2021, 13. desember). *Nå hjelper vi folk med*

strømregningen [Regjeringen]. [https://www.regjeringen.no/no/aktuelt/](https://www.regjeringen.no/no/aktuelt/na-hjelper-vi-folk-med-stromregningen/id2891984/)

[na-hjelper-vi-folk-med-stromregningen/id2891984/](https://www.regjeringen.no/no/aktuelt/na-hjelper-vi-folk-med-stromregningen/id2891984/)

SSB. (u.å.). *Elektrisitet*. Hentet 18. januar 2022, fra [https://www.ssb.no/](https://www.ssb.no/energi-og-industri/energi/statistikk/elektrisitet)

[energi-og-industri/energi/statistikk/elektrisitet](https://www.ssb.no/energi-og-industri/energi/statistikk/elektrisitet)

Viseth, E. S. (2021, 6. oktober). *Montel: Englandskabelen har tjent inn 130*

millioner på en uke [Teknisk Ukeblad]. [https://www.tu.no/artikler/](https://www.tu.no/artikler/montel-englandskabelen-har-tjent-inn-130-millioner-pa-en-uke/514025)

[montel-englandskabelen-har-tjent-inn-130-millioner-pa-en-uke/514025](https://www.tu.no/artikler/montel-englandskabelen-har-tjent-inn-130-millioner-pa-en-uke/514025)

Kapittel 8

Tillegg

8.1 R-kode

<https://github.com/BenSig00/Statistical-analysis-of-the-power-market-in-NO2.git>