

## Prognoser av elektrisitetsforbruk

I motsetning til de fleste andre produkter kan elektrisitet ikke lagres på en økonomisk forsvarlig måte. Dette innebærer at det til enhver tid må være like mye produksjon som etterspørsel. Gode prognoser for etterspørselen etter kraft er derfor nødvendig.

Liberaliseringen av det Nordiske kraftmarkedet begynte i 1991. Det ble da bestemt at kraftprisene skulle være markedsbestemte, og en nordisk kraftbørs Nordpool ble så opprettet. De ulike aktørene i kraftmarkedet har ulike prognosebehov, noen har kortsiktige andre har langsiktige.

Sluttbrukeraktørene (som Fjordkraft) trenger kortsiktige prognoser, bl.a. fordi forbruk for de neste 24 timene kjøpes på Nord Pool. Dette forbruk vil variere betydelig fra time til time, dels i form av naturlig døgnvariasjon og dels i form av tilfeldig variasjon som avvik fra det normale. Avvik mellom prognostisert bestilt forbruk og faktisk forbruk må kjøpes inn fra et regulerkraftmarked, og medfører vanligvis at man må betale en premie. Man ønsker derfor en forbruksmodell som gir minst mulig avvik fra det faktiske forbruket for det aktuelle leveranseområdet.

Nettselskapene (de selskapene som eier og driver det fysiske nettet) har imidlertid et naturlig monopol, ble det innført regler på hvor mye de kunne ta i nettleie. For tiden får nettselskapene bestemt et øvre tak på inntektene pr år, og disse kommer fra nettleien. Denne nettleien består av en fast og en variabel del. Nettselskapene står fritt til å velge den faste og variable avgiften. For å bestemme den variable delen må man ha en prognose med for forbruket de neste 12-18 månedene der usikkerheten knyttet til tilfeldigheter er tatt med. Man kan så finne hvilket forbruk man med p% sannsynlighet *minst* vil få de neste 12 månedene. Dette konservative estimatet kan så legges til grunn for beregningen av den variable delen av nettavgiften.

*Denne oppgaven går ut på å etablere en prognosemodell for totalforbruk av elektrisitet innen et leveranseområde på kort sik, dvs. innen en uke med timesoppløsning. Flere alternative modellen skal undersøkes med sikte på et valg, der en både tar omsyn til god prediksjonsevne og egnethet for implementering.*

Faktorer som spiller inn på forbruket innenfor en uke er:

- Normalprofil svarende til årstid, uke dag og time i døgnet..
- Temperatur. (dvs. avvik fra normaltemperatur)
- Arbeidsaktivitet. (dvs. helligdager og ferie dager)
- Pris

I praksis er det er mange ting en må ta stilling til ved valg av god modell, og en kortsiktig modell kan godt være helt annerledes enn en lagsiktig modell. Her er noen aktuelle betraktninger:

- Det er vanlig å dele totalforbruket inn i alminnelig forbruk, industriforbruk og annet forbruk, som består av nettap og pumpeforbruk, dvs. kraft brukt til å pumpe vann til høyereliggende magasiner for senere mer lønnsomt bruk. Industriforbruket er i hovedsak avhengig av konjunktorene, og i mindre grad sesong og temperaturavhengig. En kraftleverandør må ta stilling til om det trengs prognoser for hver av de tre komponentene. Hvis ikke, er spørsmålet om en bør lage prognoser for det samlede forbruk direkte, eller studere og lage prognoser for de tre komponentene hver for seg og så legge sammen prognosene.
- En kan spørre seg om det er lurt å estimere en modell som omfatter årsvariasjonen, og korrigerer for det kortsiktige avvik fra denne, eller om en skal estimere en modell basert på data fra de siste ukene og anta at avviket fra normalprofil er reflektert i disse, slik at det i hovedsak dreier seg om å fange opp døgnvariasjonen.
- Hver kraftleverandør har sin kundeprofil, som kan endre seg over tid. Leverandøren har også egne data om forbruket som svarer til profilen. Ved kortsiktige prognoser er det derfor mulig å ta omsyn til endringer som har funnet sted nylig. Det er derfor ikke sikkert at mer data, dvs. gå lenger tilbake i tid, gir bedre prognoser.
- Et spørsmål er om forbruksprognosene blir bedre ved å ta omsyn til pris. Det har vist seg at pris betyr mindre når det gjelder alminnelig forbruk og industriforbruk, men ganske mye når det gjelder pumpeforbruk. Dersom en velger en modell basert på data for kun de siste ukene, er prisen langt på vei reflektert i observert forbruk, og trenger ikke representeres i modellen. En annen sak er om man vet om prisendringer for kommende uke. Da kan det være hensiktsmessig å foreta korrigeringer i prognosene ut fra generell kunnskap om priselastisiteten.
- En leverandør som leverer hovedsakelig til en bestemt region, vil kunne ta omsyn til værvarsler på kort sikt for denne regionen, enten ved å bringe temperatur inn i modellen, eller foreta korreksjoner i prognosene ved forventet avvik fra de siste ukene.
- Et problem er at det er to "sesongkomponenter", en ukesyklus og en døgnsyklus. Den karakteristiske ukesyklus har noenlunde samme forbruksmønster virkedagene mandag til fredag, men noe lavere forbruksnivå lørdag og søndag. De fleste enkle tidsrekkeметoder håndterer bare en sesongkomponent, og vi må lete etter måter å omgå dette på. Kan hende skyldes forskjellen i noen grad industriforbruket, slik at problemet hadde vært løst ved å betrakte dette separat fra alminnelig forbruk. En annen mulighet er å lage en ukedag indeks som måler forholdet til gjennomsnittsnivået. Denne brukes til "å fjerne" ukedag effekten ved divisjon med indeksen, og etterpå til "å tilbakeføre" prognosetallene for respektiv dag ved multiplikasjon med indeksen.

Vi vil betrakte problemet for en kraftleverandør som har sluttbrukere spredd over hele Norge, og der det er meningsfylt å gi prognoser for det totale norske forbruket på timebasis de nærmeste ukene. Vi vil begrense oss til modeller som går rett på det totale forbruk, og ikke splitter opp etter ulike typer forbruk. Prognoser for egne leveranser kan eventuelt lages ved å multiplisere med markedsandelen, som ikke endrer seg på kort sikt.

Vi tar sikte på å lage timeprognoser for dager i første uke av november 2005 (uke 44), med grunnlag i data fra oktober 2005.

For å teste ut ulike prognosemetoder tar vi for oss tilsvarende data fra 2004, dvs. bruker data fra oktober 2004 til å etablere ulike prognosemodeller, for å se hvilke(n) som gjorde det best i å predikere motsvarende dager i november 2004 (uke 45).

## Datakilder og dataproblemer

Statistikk over forbruk og produksjon av elektrisk kraft samles inn og bearbeides av Statistisk Sentralbyrå (SSB) og NordPool, basert på informasjon fra Statnett. Det er vanskelig å gi helt eksakte tall, og rapportert forbruk er derfor (gode) estimater. Mye informasjon kan finnes via nettsidene til Nordpool og SSB. I tillegg rapporterer Norges vassdrags- og elektrisitetsvesen (NVE) i begynnelsen av hver måned en korttidsstatistikk med SSB og NordPool som kilde, men nytter disse på en måte som står for NVE's egen regning (beskrevet i ER-notat 35/99). Dette gjør at tall fra ulike kilder kan avvike noe, uten at de spiller særlig rolle for helheten. Bl.a. rapporterer NVE raskt foreløpige tall, som ofte justeres noe i senere rapporter. Grenseoppgangen mellom alminnelig forbruk og industriforbruk kan også ha betydning. Hos NVE omfatter alminnelig forbruk forsyning til husholdninger, jordbruk, anleggsdrift, tjenesteyting, transport samt forsyning til ikke-kraftintensiv industri, herunder treforedling og bergverk. Det rapporterte industriforbruket er således til kraftintensiv industri alene.

Totalt norsk forbruk i timeoppløsning kan skaffes fra Nord Pool og/eller Statkraft.

Dersom en ønsker å ha med temperatur i prognosemodellen, kan en registrere dagmiddelet og normaltemperaturen i byene Oslo, Bergen, Trondheim, Stavanger og Tromsø. Man kan da beregne en indeks som en vektet sum av avvikene fra normalen i de fem byene. Vi skal ikke gjøre det.

For å redusere feilprognostiseringen kunne det være ønskelig å ha en liste over helligdager og ferier. Denne listen kan brukes i kombinasjon med dummy variabler for å hindre at spesielle omstendigheter forvirr prognosene. Det er imidlertid lite trolig at dette har særlig betydning for vårt formål.

En bør alltid sjekke datafilen for å avdekke eventuelle feil. Plott kan ofte avsløre svært avvikende observasjoner, som viser seg å være feilrapportering pga. registreringsfeil eller skrivefeil. I noen tilfeller kan og bør dette korrigeres. Det skjer ofte at nødvendige enkeltobservasjoner mangler. Dette kan løses på ulikt vis. En mulighet er å ta gjennomsnittet av naboobservasjoner, såfremt dette ikke gjøres i stort omfang og vrir analyseresultatene. I vår situasjon er verdt å være oppmerksom om noe skjer med dataene ved overgangen til sommertid og vintertid.

## Data for prosjektet

Data for prosjektet i MWh er lagret i Excel-filen *Timeforbruk\_NO2004.XLS* og *Timeforbruk\_NO2005\_oktober.XLS* (uke 40-43) er utlagt i mappen Eksamen på kursets hjemmeside. Ytterligere data fra 2005 kan evt. fås via [www.nordpool.com](http://www.nordpool.com) > Market data > Consumption Norway.

Dersom man finner det nødvendig med korreksjoner i data skal dette rapporteres.

Det er hensiktsmessig å føye til ukenummer (kodetall 1-52 evt. 53), ukedagnummer (kodetall 1-7), timenummer (kodetall 1-24), samt "tiden" (1,2,3, .... ) som egne variable.

Vi tenker oss at konteksten for dette prosjektet er en utredning på oppdrag av ledelsen om hvilken prognosemetode som bør velges for praktisk bruk. Utform gjerne besvarelsen som en rapport med et kort "executive summary". Rapporten skal avgis med en knapp tidsfrist, og prosjektgruppen må foreta prioriteringer underveis i arbeidet.

Det er flere vurderingskriterier som kan være aktuelle ved valg av metode:

1. God modelltilpasning til data ("within sample"), f.eks målt med MSD ("Mean Squared Deviation" eller eventuelt MAD ("Mean Absolute Deviation") eller MAPE ("Mean Absolute Percent Error").
2. God prediksjonsevne ("out of sample"), målt med tilsvarende mål.
3. Gir grunnlag for risikovurdering for både enkeltprediksjoner og evt. aggregert over tid, enten av programvaren eller ved beregning på grunnlag av denne.
4. Metoden er enkel å formidle og enkel å implementere med tilgjengelig programvare og/eller lett å programmere f.eks i Excel.
5. Metoden lar seg modifisere til å ta omsyn til å ta omsyn til ulike typer ny informasjon.

Følgende aktuelle prognosemetoder er dekket i programpakken Minitab:

- Enkel tidsrekke-dekomponering (additiv eller multiplikativ)
- Forecasting-metoder (Holt-Winters additive eller multiplikative metode)
- Regresjonsmetoder med trend, sesongdummier og evt. laggede variable
- Tidsrekkeanalyse og prediksjon basert på ARIMA-modell med trend og sesong

For å vurdere disse metodene opp mot hverandre vil vi bruke å data for timeforbruk fra oktober 2004 til å etablere modellene. Prognosene av timeforbruk for ett eller flere døgn utover i november ("out of sample") kan så sammenlignes med det virkelige forbruket. Når vi har funnet en god modell, kan vi bruke denne på tilsvarende data fra år 2005.

Progresjonen i prosjektarbeidet er formulert som konkrete spørsmål på neste side.

Kommentarer ang. arbeidet: Praksis holder seg sjelden helt "innenfor pensum", men dette gir grunnlag for å orientere seg etter behov vha. alternativ litteratur, hjelpefunksjoner i programvaren og søkbar informasjon på nettet. Det er muligens tilfelle også her. Det er heller ikke sikkert at alle spørsmål en stiller seg har entydige riktige svar. I hver av modellene som studeres, pass på følgende: (i) ha den opprinnelige målsetting i tankene ved prioriteringen av hva som tas vare på av utskrifter (ii) ta vare på residualene for å kunne drøfte forutsetningene for analysene.

Kommentarer ang. rapporten:

Prioriter å ha med de utskrifter som er relevante for formålet, og hold alt irrelevant utenfor. Plott er bra dersom budskapet kommer klart fram med relevante kommentarer.

## Prosjektoppgave

1. Betrakt først tidsrekken med timeforbruk for hele 2004 og beskriv grafisk
  - (i) gjennomsnittlig ukeforbruk over året
  - (ii) gjennomsnittlig døgnforbruk over uken
  - (iii) gjennomsnittlig timeforbruk over døgnetDrøft kort hvilke problem plottene antyder mht. for modellvalg og dataanalyse.
2. Justering av timeforbruket for ukedag kan skje ved en "indeks" for ukedag som i gjennomsnitt er lik 1 over uken. Lag en slik indeks, og bekreft at verdiene 1.02 for hver av de fem virkedagene og 0.95 for lørdag og søndag kan passe. Lag så ny variabel, der timeforbruk er justert for ukedag ved divisjon med indeksen.

For å lage prognosemodell har vi nå valget mellom å betrakte tidsrekken for hele 2004, og justere for tid på året på tilsvarende måte vha. ukenummer, eller betrakte tidsrekken for bare en kort tid før prognosetidspunktet 2004. Vi forsøker det siste, og betrakter derfor ukedagjusterte data for oktober 2004.

3. Betrakt først de to metodene en kjenner best til: Enkel multiplikativ dekomposisjon av justert timeforbruk med døgn som sesongperiode og regresjonsanalyse med dummyvariable for time. Tilpass modeller for de to metodene, og vurder MSD "within sample". Vurder spesielt om en bør ha med trendkonstant eller ikke. Hvilken metode ser mest lovende ut?
4. Lag timeprognoser for det første døgnet i november for hver av modellene i pkt. 3, og beregn MSD "out of sample". Er metoden(e) som bød seg fram i pkt. 3 fortsatt best, eller er valget av modell kommet i nytt lys? Spiller valget av tidshorisont "out of sample" noen rolle?
5. Betrakt så mulighetene å lage prognoser basert på en ARIMA-modell. Estimer en  $ARIMA(0|1|1)(0|1|1)_{24}$ -modell basert på ukedagjusterte data fra hele oktober 2004. Lag timeprognoser for det første døgnet i november, og beregn MSD "out of sample". Forsøk, om mulig, å identifisere en bedre ARIMA-modell, og sjekk om denne også er bedre "out of sample". Vurder spesielt om en bør ha med trendkonstant eller ikke.
6. Betrakt så muligheten for å bruke Winters' multiplikative metode med ukedagjusterte data fra hele oktober 2004, med timesprognoser for det første døgnet i november. Vurder metoden med MSD både "within sample" og "out of sample".
7. "Executive summary": Hvilken modell vil prosjektgruppen anbefale ut fra ovenstående analyser og eventuelle andre vurderinger og hvorfor?
8. Vær forberedt på at ledelsen kan etterspørre både forbedringer av ovenstående og enda enklere metoder. Drøft slike muligheter, og gjør eventuelle analyser for å kunne være forberedt på dette.
9. Anta at ledelsen har fulgt prosjektgruppens anbefaling mht. modelltype, og reestimerer denne på grunnlag av data fra oktober 2005, med etterfølgende prognoser av timeforbruket for det første døgnet i november 2005. Rapporter prediksjon, virkelig forbruk og prognosefeil med tilhørende MSD. Sammenlign med hva du ville ha fått dersom du brukte timeforbruk i samme døgn året før som prediksjoner.