



*Magne Holstad og Dinh Quang Pham*

# Notater

**Sesongjustering av strøm-  
forbruket for alminnelig  
forsyning i månedlig  
elektrisitetsstatistikk**  
Metode og resultater



## **Forord**

Dette notatet beskriver noen resultater og metode for sesongjustering av elektrisitetsforbruket til alminnelig forsyning i månedlig elektrisitetsstatistikk. Prosjektarbeidet er ledet av Magne Holstad ved Seksjon for energistatistikk. Beskrivelsen av metodene er delvis basert på Dinh Quang Phams notat: Innføring i tidsserier – sesongjustering og X-12-ARIMA. Dinh Quang Pham har også bidratt med nyttige kommentarer, forklaringer, IT-løsninger og inspirasjon under arbeidet med sesongjustering av strømforbruket innenfor alminnelig forsyning. En stor takk til Tom Andersen Langer som også har vært involvert i arbeidet. Hans kunnskap og forståelse av sesongjustering har vært til stor hjelp. Bjørn Bleskestad og Marius Bergh har gitt nyttige kommentarer til dette dokumentet. Marius Bergh har også vært med underveis i prosjektet og hans innsikt i kraftmarkedet har vært verdifull. En takk til Eirik Håland Linstad ved Seksjon for nasjonalregnskap for beregning av produksjonsandeler for tjenesteytende sektor og annen industri enn den kraftintensive i perioden 1993-2005.

## Sammendrag

Forbruk i alminnelig forsyning påvirkes av sesongvariasjon gjennom året. Den viktigste kilden til sesongvariasjon er temperatur. Siden forbruket har et systematisk mønster gjennom året, kan vi ikke sammenlikne tall fra måned til måned på en meningsfull måte. Gjennom statistiske metoder kan vi renske rådata for all systematisk sesongvariasjon. De sesongjusterte tallene gir oss informasjon om en økning (nedgang) i rådata bare skyldes sesongvariasjon, eller om vi har en generell økning (nedgang) i strømforbruket. Sesongjusterte tall beskriver derfor den underliggende utviklingen.

Dette notatet består av tre deler. I del I presenterer vi rådata. Plott av rådata viser tydelig sesongmønster. Strømforbruket har høye verdier gjennom vinteren, avtar gjennom sommermånedene for så å stige gjennom høstmånedene. Autokorrelasjonsdiagram og plott av rådata for hvert år over hverandre understøtter også at forbruk i alminnelig forsyning har et veldefinert sesongmønster. I tillegg indikerer autokorrelasjonsdiagrammet at rådata er en ikke-stasjonær prosess. Dette tester vi om stemmer ved å bruke en Dickey-Fuller test. I Dickey-Fuller testen beholder vi  $H_0$  hypotesen, som sier at rådata er en ikke-stasjonær prosess.

I del II forklarer vi hvordan rådata prekorrigeres ved en ARIMA modell. Kalendereffekter skyldes ulikt antall arbeidsdager fra måned til måned eller bevegelige helligdager som påske, pinse og Kristi Himmelfartsdag. Da sammensetningen av ukedager i en måned kan ha betydning for strømforbruket, tester vi først i en økonometrisk modell om alle variabler som fanger opp kalendereffekter og bevegelige helligdager er signifikant. Deretter tar vi ut den variabelen som er minst signifikant og estimerer på nytt. Til slutt står vi igjen med en modell hvor bare ukedagseffekten er signifikant på 5 prosent nivå. I denne modellen antar vi at hverdager påvirker strømforbruket i alminnelig forsyning med lik vekt, mens lørdag og søndag påvirker med en lavere vekt. Denne spesifiseringen av ukedagseffekten er rimelig, fordi enkelte forbruksgrupper innenfor tjenesteyting vil bruke mindre strøm i helgen på grunn av ingen eller redusert åpningstid. Dette kan også gjelde for noen forbruksgrupper i industrien utenom kraftintensiv industri, men forbruket innenfor treforedling, som utgjør om lag 34 prosent av dette forbruket, går jevnt gjennom hele året. Parameterestimaterne i modellen uttrykker hvor mye logaritmen til rådata er prekorrigert, med motsatt fortegn.

I siste del av notatet viser vi resultatene fra sesongjusteringen og ulike indikatorer som måler hvor god sesongjusteringen er. Sesongjusterte tall er satt sammen av de to komponentene trend og irregulær komponent. Den irregulære komponenten reflekterer tilfeldig variasjon som ikke gjentar seg fra år til år og kan derfor tolkes som et støyledd. Ved å fjerne den irregulære komponenten fra den sesongjusterte serien framkommer trenden, som beskriver den langsiktige utviklingen i data. Grafen med sesongjusterte tall viser at vi ved å glatte ut sesongsvingninger gjennom året kan sammenlikne tall fra måned til måned på en meningsfull måte. Plottet av trenden indikerer en positiv underliggende vekst i strømforbruket på nittitallet, mens det ser ut til at forbruket flater ut i perioden 2000(1)-2008(5) per måned på rundt 6630 GWh. Kvalitetsmålene uttrykker at vi alt i alt oppnår en god sesongjustering. M3 og M5 (se tabell på side 18) har riktignok høyere verdier enn 1, noe som ikke er gunstig. Dette kommer av at vi da har store utslag i den irregulære komponenten. Q, som er en vektet samleindikator for M1-M11 bortsett fra M2, er lik 0,46 og det er svært bra. Kritisk verdi for Q er 1. Stabiliteten for sesongkomponenten og sesongjusterte tall er tilfredsstillende.

# Innholdsfortegnelse

<b>1. Innledning .....</b>	<b>4</b>
<b>2. Presentasjon av rådata.....</b>	<b>5</b>
<b>3. Prekorrigering av rådata.....</b>	<b>7</b>
3.1 Ukedagseffekt/trading day effect .....	7
3.2 Bevegelige røde dager og inneklemdager .....	8
3.3 Bevegelige helligdager .....	8
3.4 Ekstremer .....	9
3.5 RegARIMA modell.....	9
3.6 Resultater.....	10
<b>4. Sesongjustering av prekorrigert serie .....</b>	<b>13</b>
4.1 Sesongjustert serie, trend, irregulær komponent og sesongkomponent .....	14
4.2 Tolkning av sesongjusterte tall.....	16
4.3 Kvalitetsindikatorer for sesongjustering .....	17
4.4 Konklusjon .....	19
<b>Kildehenvisning .....</b>	<b>20</b>
<b>Appendix I: Utledning av Dickey-Fuller test for en autoregressiv prosess av orden 4.....</b>	<b>21</b>
<b>Appendix II: De fem ARIMA-modellene som X12arima vurderer .....</b>	<b>22</b>
<b>Appendix III: Utledning av modell I hvis støyleddet følger en ARIMA modell.....</b>	<b>23</b>
<b>Appendix IV: Initiale beregningsmetoder for sesong, trend og irregulær komponent .....</b>	<b>24</b>

# 1. Innledning

Månedlig elektrisitetsstatistikk går tilbake til 1979 og har som formål å gi en oversikt over produksjon og forbruk av elektrisk kraft. Statistikken omfatter blant annet kraftproduksjon, eksport og import samt forbruk innenfor kraftintensiv industri og alminnelig forsyning. Seksjon for energistatistikk vurderte å sesongjustere alle disse seriene, men valgte til slutt kun å sesongjustere strømforbruket innenfor alminnelig forsyning. Dette skyldes at tidsseriene for eksport, import og kraftintensiv industri ikke har et veldefinert sesongmønster. Tidsserien for produksjon inneholder sesongvariasjon, men det kan ses i sammenheng med at den langt på vei blir indirekte bestemt av forbruket innenfor alminnelig forsyning. Det vises her for øvrig til at publiseringen av produksjonsindeksen for industrien også omfatter sesongjustert produksjonsindeks for kraftforsyning.

Forbruk i alminnelig forsyning er nettoforbruk<sup>1</sup> av elektrisk kraft fratrukket forbruk i kraftintensiv industri. Husholdninger, tjenesteytende sektor og annen industri enn den kraftintensive står for det aller meste av dette forbruket. I årlig elektrisitetsstatistikk 2006 utgjorde husholdninger, tjenesteytende sektor og annen industri enn den kraftintensive henholdsvis 42, 34 og 20 prosent av det totale forbruket innenfor alminnelig forsyning.

Forbruket i alminnelig forsyning påvirkes sterkt av sesongvariasjon gjennom året. Den klart viktigste kilden til sesongvariasjon er temperatur, men det kan også være andre sesongfaktorer som påvirker forbruket gjennom året. For eksempel kan fellesferien i juli, som er knyttet til institusjonelle forhold, bidra til som regel lavere strømforbruk i denne måneden sammenliknet med juni. Selv om forbruket innenfor annen industri enn den kraftintensive i svært liten grad er temperaturavhengig, påvirkes de to største forbruksgruppene klart av temperaturforhold.

Sesongvariasjon gjentar seg systematisk fra måned til måned år etter år. Gjennomsnittstemperaturer fra Meteorologisk institutt viser at mai og juni i perioden 1979-2007 alltid er varmere enn foregående måned, mens september, oktober og november alltid er kaldere enn foregående måned. Systematiske temperaturforskjeller mellom månedene har bidratt til at det alltid har vært lavere strømforbruk i mai og juni enn i foregående måned og tilsvarende for september, oktober og november bare med motsatt fortegn. For å kunne sammenlikne tall fra måned til måned på en meningsfull måte i tillegg til å få et bilde på den underliggende veksten i strømforbruket, må vi fjerne den systematiske sesongvariasjonen. Dette kan vi gjøre gjennom statistiske metoder.

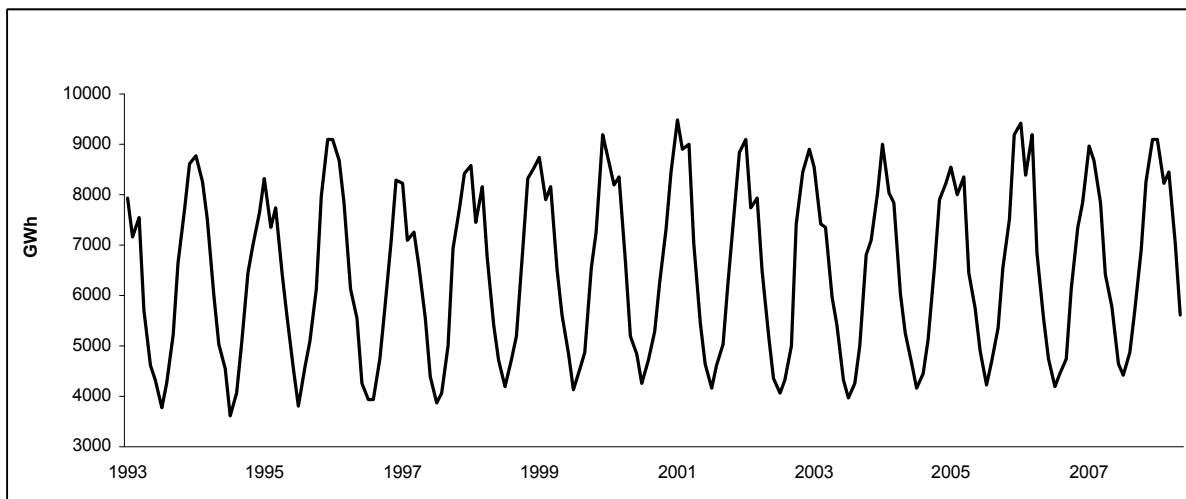
Den empiriske analysen i dette dokumentet er utført med PcGive 10.0, Excel og SAS.

---

<sup>1</sup> Nettoforbruk er bruttoforbruk av elektrisk kraft fratrukket pumpekraftforbruk, annet eget forbruk og beregnet tap i linjenettet.

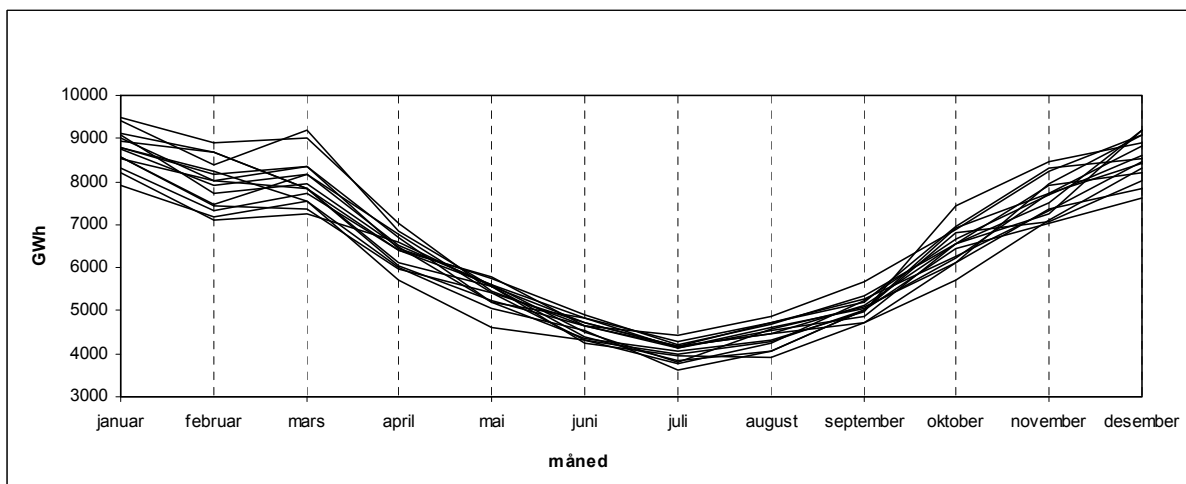
## 2. Presentasjon av rådata

Figur 2.1: Plott av rådata



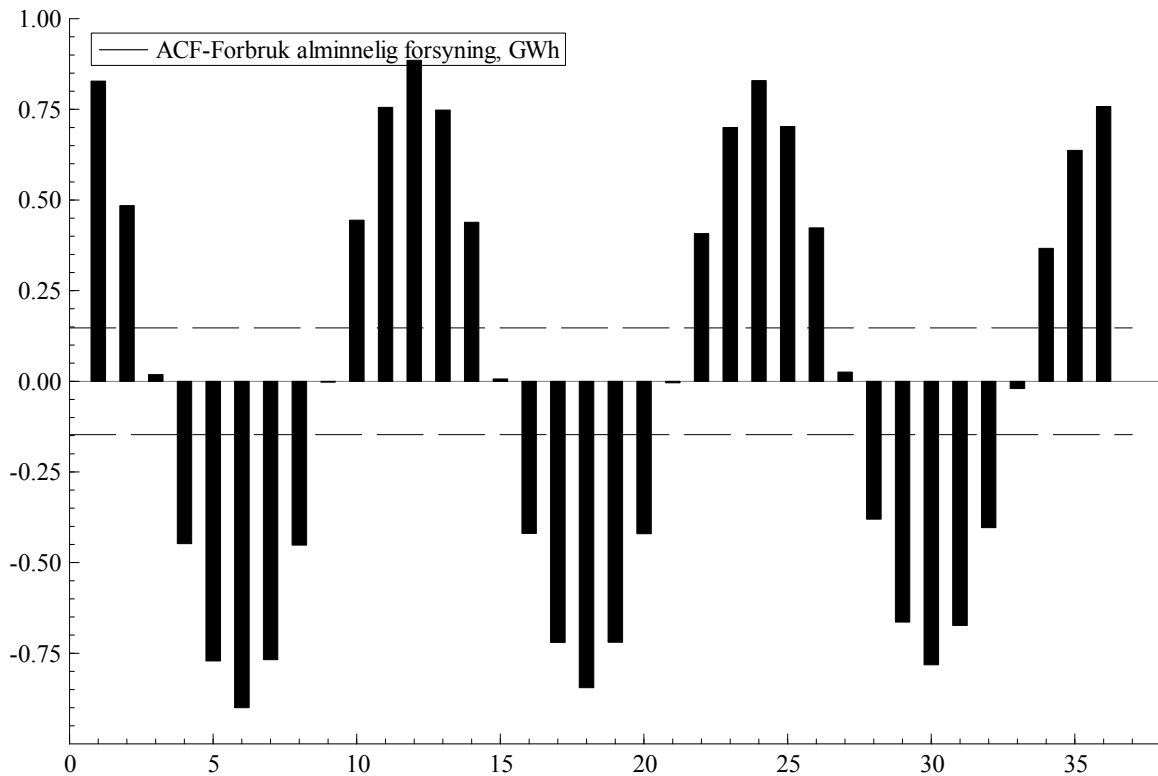
Vi har ovenfor plottet forbruk i alminnelig forsyning i perioden 1993(1)-2008(5). Det er i alt 185 observasjoner. Vi ser av grafen at rådata har et tydelig sesongmønster gjennom året som gjentar seg systematisk fra år til år. Strømforbruket har høye verdier gjennom vinteren, avtar gjennom sommermånedene for så å stige gjennom høstmånedene. Dette kan man visualisere tydeligere ved å plote rådataene for hvert år over hverandre.

Figur 2.2: Rådata plottet for hvert år over hverandre



Figur 2.2 viser at rådata har et tydelig sesongmønster. Vi ser av grafen at observasjonspunktene i januar, februar, mars, oktober, november og desember har større variasjon enn de andre månedene. Dette kan ses i sammenheng med at strømforbruket er mer temperaturfølsomt i disse månedene.

Figur 2.3: Autokorrelasjonsdiagram av rådata



Autokorrelasjonsdiagrammet viser at lag 12, 24 og 36 er signifikant på 5 prosent nivå, noe som viser sesongvariasjon i data. Lag 12 uttrykker korrelasjon eller avhengighet mellom observasjon  $t$  og  $t-12$ , dvs. at det for eksempel er korrelasjon mellom verdien i januar 2007 og verdien i samme måned året før.

Autokorrelasjonsdiagrammet indikerer også at rådata er en ikke-stasjonær prosess siden de aller fleste lagene er signifikant på 5 prosent nivå. Dette kan vi teste om understøttes av en Dickey-Fuller test. Box-Jenkins metoden indikerer at rådata,  $O_t$  følger en autoregressiv orden av fjerde grad. Vi kan da spesifisere Dickey-Fuller testen med følgende likning<sup>2</sup>:

$$\Delta O_t = \gamma_1 O_{t-1} + \gamma_2 \Delta O_{t-1} + \gamma_3 \Delta O_{t-2} + \gamma_4 \Delta O_{t-3} + u_t$$

$u_t$  er et white noise støy-ledd,  $u_t \sim iid(0, \sigma^2)$

Vi estimerer likningen ovenfor med OLS. Tabell 2.1 på neste side viser at  $t$ -verdien for  $\gamma_1$  er lik -1,48. Derfor beholder vi  $H_0$  hypotesen om at rådata er en ikke-stasjonær prosess.

<sup>2</sup> Se appendix I for utledning



Tabell 2.1

	<b>Coefficient</b>	<b>Std.Error</b>	<b>t-value</b>	<b>t-prob</b>	<b>Part.R^2</b>
O(t)_1	-0,0131589	0,008872	-1,48	0,14	0,0123
DO(t)_1	0,512127	0,06987	7,33	0	0,2329
DO(t)_2	0,275092	0,07653	3,59	0	0,068
DO(t)_3	-0,305109	0,07105	-4,29	0	0,0944
<b>Periode</b>	1993(1)-2008(5)				

### 3. Prekorrigering av rådata

Metoden vi bruker for å sesongjustere forbruk innenfor alminnelig forsyning består av to deler. Først prekorrigerer vi rådata gjennom en regresjonsanalyse, og deretter sesongjusterer vi den prekorrigerte serien blant annet ved å benytte ulike beregningsmetoder for glidende gjennomsnitt. Kalendereffekter skyldes ulikt antall arbeidsdager fra måned til måned eller bevegelige helligdager som påske, pinse og Kristi Himmelfartsdag. Hvis det er slik at sammensetningen av ukedager i en måned har betydning for strømforbruket, må vi korrigere for denne effekten for å få informasjon om hva som er den reelle utviklingen i strømforbruket.

#### 3.1 Ukedagseffekt/trading day effect

Selv om både januar og mars har 31 dager, vil sammensetningen av hverdager og helgedager være forskjellig mellom de to månedene. I januar 2008 var det 23 hverdager og 8 helgedager, mens det i mars var 21 hverdager og 10 helgedager. Hvis vi bruker mer strøm i en måned med flere hverdager (bare fordi den har flere hverdager) enn en annen, vil vi trekke feil konklusjoner om den underliggende veksten hvis vi ikke korrigerer for dette. Når det gjelder forbruk innenfor alminnelig forsyning, kan følgende to modellspesifikasjoner for ukedagseffekt være rimelig:

##### Modell (i):

Mandag=tirsdag=.....=...fredag: Påvirker strømforbruket med lik vekt, mot lørdag og søndag som påvirker med en lavere vekt.

##### Modell (ii):

Mandag=tirsdag=.....=lørdag: Påvirker strømforbruket med lik vekt, mot søndag som påvirker med en lavere vekt.

Enkelte forbruksgrupper innenfor tjenesteyting, blant annet offentlig forvaltning og privat tjenesteyting, vil bruke mindre strøm i helgen på grunn av ingen eller redusert åpningstid. Dette kan også gjelde for noen forbruksgrupper i industrien utenom kraftintensiv industri, men forbruket innenfor treforedling, som utgjør om lag 34 prosent av dette forbruket, går jevnt gjennom hele året. Modell I er derfor en rimelig spesifisering av ukedagseffekten. Siden antakeligvis flere forbruksgrupper har åpent på lørdag og stengt på søndag, blant annet varehandel, tester vi også modell II. Når det gjelder husholdningens strømforbruk, virker det mest rimelig at ukedager og helgedager har lik vekt. Dinh Pham jobber med et notat om hvordan dummy variabler for ukedagseffekt lages. Det vil bli lagt ut en lenke til dette dokumentet med en gang det er ferdig.

### 3.2 Bevegelige røde dager og inneklemdager

Hvis modell (i) eller (ii) har en signifikant effekt på strømforbruket, må vi også korrigere for bevegelige røde dager og inneklemdager. Disse dagene faller av og til på en hverdag og andre ganger i helgen. Røde dager er 1.januar, 1.mai, 17.mai, 25.desember og 26.desember, mens inneklemdager er 27.-31.desember. Vi lager dummy variabler for å korrigere for dette. Valg av modell (i) eller (ii) legger grunnlaget for hvordan vi skal lage dummy variablene. I første estimering kjører vi med dummy variabler for alle røde dager og inneklemdager. Deretter tar vi ut den dummy variabelen som er minst signifikant og estimerer på nytt. Siden inkludering av 30. og 31.desember vil føre til modelltekniske problemer, er disse dummy variablene utelatt. Vi korrigerer kun for dummy variabler som er signifikante og hvor fortegnet på estimatet er negativt, altså i tråd med våre antakelser om ukedagseffekten.

Eksempel på dummy variabler:

Under modell (i):

$$D_{17.mai} = \begin{cases} 1 & \text{hvis 17.mai er en mand, tirs, ..., fredag og } t=\text{mai} \\ 0 & \text{ellers} \end{cases}$$

$$D_{24.des} = \begin{cases} 1 & \text{hvis 24.des er en mand, tirs, ..., fredag og } t=\text{desember} \\ 0 & \text{ellers} \end{cases}$$

Under modell (ii):

$$D_{17.mai} = \begin{cases} 1 & \text{hvis 17.mai er en mand, tirs, ..., lørdag og } t=\text{mai} \\ 0 & \text{ellers} \end{cases}$$

$$D_{24.des} = \begin{cases} 1 & \text{hvis 24.des er en mand, tirs, ..., lørdag og } t=\text{desember} \\ 0 & \text{ellers} \end{cases}$$

### 3.3 Bevegelige helligdager

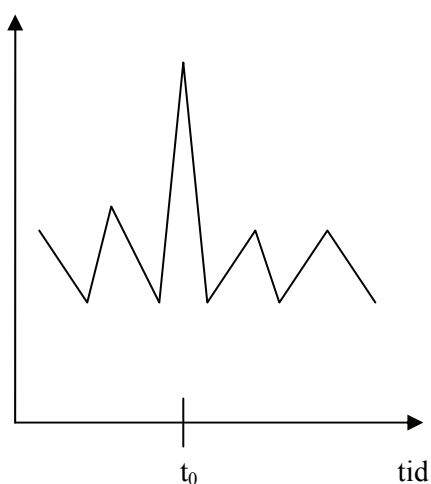
Kristi Himmelfartsdag og pinsen er alltid henholdsvis 39 og 49 dager etter 1.påskedag. Denne systematikken gjør at dummy variablene for påske, pinse og Kristi Himmelfartsdag henger sammen. Vi korrigerer kun for disse helligdagene når påsken faller i mars eller deler seg mellom mars og april. Vi betrakter påsken som sesongvariasjon når den faller i april, siden påsken vanligvis faller i denne måneden. Jamfør diskusjonen ovenfor om ukedagseffekt forventer vi negative parameterestimater for påske, pinse og Kristi Himmelfartsdag. I påsken drar om lag 10-15 prosent av husholdingene til fjells og virksomhet innenfor tjenesteyting og annen industri enn den kraftintensive har stengt. Dette trekker i retning av et lavere strømforbruk. Effekten begrenses nok

likevel mye av at oppvarming i bygninger og boliger ikke slås av og at mye av industrien går. I tillegg har mange hytter innlagt strøm. Vi korrigerer kun for påske, pinse og Kristi Himmelfartsdag hvis variablene er signifikante og parameterestimaterne er i tråd med vår forkunnskap. Dinh Pham har skrevet et notat om hvordan dummy variablene for påske, pinse og Kristi Himmelfart lages. Vedlagt er lenke til dette notatet: [http://www.ssb.no/emner/00/90/notat\\_200743/notat\\_200743.pdf](http://www.ssb.no/emner/00/90/notat_200743/notat_200743.pdf)

### 3.4 Ekstremer

Vi lar programmet X12arima teste for de to ekstremene additiv uteligger (AO) og nivåskifte (LS). Utifra plottene av rådata forventer vi at X12arima ikke vil identifisere noen ekstremer.

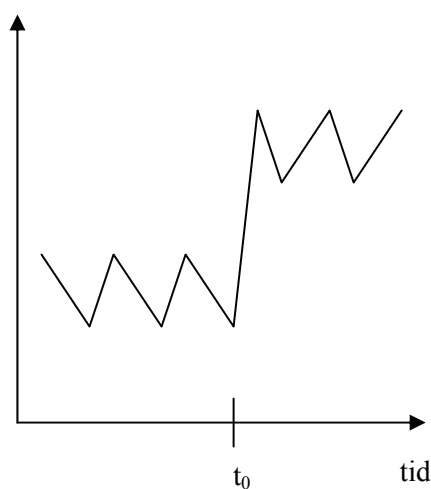
Figur 3.1: Additiv uteligger



Figur 3.1 viser at additiv uteligger skyldes et bestemt utslag i tidspunkt  $t_0$ . For å korrigere for denne effekten, lager X12arima følgende dummy variabel:

$$AO_t = \begin{cases} 1 & \text{for } t=t_0 \\ 0 & \text{for } t \neq t_0 \end{cases}$$

Figur 3.2: Nivåskifte



Figur 3.2 viser at ved  $t_0$  begynner tidsserien å fluktuere rundt et nytt og høyere nivå. For å korrigere for nivåskifte, lager X12arima følgende dummy variabel:

$$LS_t = \begin{cases} 1 & \text{for } t \geq t_0 \\ 0 & \text{for } t < t_0 \end{cases}$$

### 3.5 RegARIMA modell

I tidsseriedata er ofte restleddet autokorrelert. Autokorrelasjon vil si at restleddene har et systematisk mønster, for eksempel at vi har en oppsamling av positive eller negative residualer etter hverandre (positiv autokorrelasjon). T-verdier og standardavvik er ikke pålitelig i en modell

hvor restleddet er autokorrelert, men parameterestimaterne er forventningsrette. Vi løser problemet med autokorrelasjon ved at vi antar at restleddet følger en ARIMA-modell og lar programmet X12arima bestemme en passende ARIMA spesifisering. Når X12arima velger modell, er det tre kriterier som skal være oppfylt:

- (1) Gjennomsnittlig absolutt prognosefeil i prosent i de siste tre årene skal være mindre eller lik 15 prosent.
- (2) Residualene skal være ukorrelerte, mer presist  $\chi^2$  skal være større enn 5 %.
- (3) Det er ingen tegn til "overdifferencing" ( $\sum_i \theta_i > 0,9$  eller  $\sum_i \Theta_i > 0,9$ )

X12arima velger mellom de fem ARIMA-modellene i appendix II og kontrollerer først om kriteriene på forrige side er oppfylt for modell (a). Hvis kriteriene er oppfylt, velger programmet modellen, ellers går den videre til modell (b). Av og til oppfyller ingen av modellene alle kriteriene. Da velges modell (a), Airline-modellen, ved default. Airline modellen er den enkleste av ARIMA modellene og har følgende struktur  $(p, d, q)(P, D, Q) = (0, 1, 1)(0, 1, 1)$ .

For å finne ut hvilke variabler vi skal prekorrigere for, estimerer vi først en modell hvor alle variabler er inkludert. Deretter tar vi ut den variabelen som er minst signifikant og estimerer på nytt. Vi korrigerer kun for variabler som er signifikant på 5 prosent nivå og hvor parameterestimaterne er forankret i våre forkunnskaper.

Da vi har to plausible spesifiseringer for ukedagseffekt, testes disse separat i modell I<sup>3</sup> og II. I både modell I og II har vi utelatt konstantledd og alle variablene er transformert med log, fordi dette gir lavest verdier for AICC og Hannan Quinn. Jo, lavere verdiene er for AICC og Hannan Quinn, jo bedre er modellspesifiseringen.

### 3.6 Resultater

Modell I:

$$O_t = \beta_0 \text{leapyear} + \beta_1 \text{fjorpa\ss ke}_t + \beta_2 \text{pa\ss ke}_t + \beta_3 \text{pinse}_t + \beta_4 \text{etterpinse}_t + \beta_5 \text{KH}_t + \beta_6 \text{TD15}_t + \beta_7 D_{1.\text{jan}} + \beta_8 D_{1.\text{mai}} + \beta_9 D_{17.\text{mai}} + \beta_{10} D_{\text{des}24} + \beta_{11} D_{\text{des}25} + \beta_{12} D_{\text{des}26} + \beta_{13} D_{\text{des}27} + \beta_{14} D_{\text{des}28} + \beta_{15} D_{\text{des}29} + z_t$$

hvor:

estimeringsperiode: 1993(1)-2008(5)

alle variabler er log transformert

leap year: Skuddår

KH: Kristi Himmelfartsdag

TD15: Mand=tirs=...=fredag: lik vekt, mot lørdag og søndag.

$z_t \sim \text{ARIMAmodell}$

Vi estimerer modell I ved hjelp av maximum likelihood. Tabell 3.1 på neste side viser estimeringsresultatene. Vi ser av resultatene at det bare er ukedagseffekten som er signifikant både på 5 og 10 prosent nivå. Parameterestimaterne har riktig fortegn og uttrykker at vi bruker mer strøm på hverdager enn i helgen. Alle dummy variablene for bevegelige røde dager og inneklemdager er ikke-signifikant og har svært lave t-verdier. Påske er heller ikke signifikant, men har riktig fortegn. X12arima velger airline modellen ved default.

<sup>3</sup> Se appendix III for utledning av modell I

Tabell 3.1

Variabel	Parameterestimat	Standardfeil	T-verdi
leap year	0,03	0,02313	1,30
Førpåske	-0,0056	0,02017	-0,28
Påske	-0,012	0,02296	-0,52
Pinse	-0,0304	0,02403	-1,26
Etterpinse	0,0303	0,0198	1,53
KH	0,0478	0,02834	1,69 (*)
TD15	0,0025	0,00114	2,20 (**)
D <sub>1.jan</sub>	0,004	0,02275	0,18
D <sub>1.mai</sub>	-0,007	0,02804	-0,25
D <sub>17.mai</sub>	-0,0329	0,027	-1,22
D <sub>24.des</sub>	0,0145	0,07302	0,20
D <sub>25.des</sub>	-0,0021	0,0372	-0,06
D <sub>26.des</sub>	0,0003	0,06349	0,00
D <sub>27.des</sub>	0,0477	0,05917	0,81
D <sub>28.des</sub>	-0,0067	0,04037	-0,17
D <sub>29.des</sub>	-0,032	0,07115	-0,45
$\theta_1$	0,4747	0,06128	
$\Theta_1$	0,9999	0,0523	

(\*), (\*\*), (\*\*\*) indikerer signifikans på henholdsvis 10, 5 og 1 prosent nivå.

#### Modell II:

$$O_t = \gamma_0 \text{leapyear}_t + \gamma_1 \text{førpåske}_t + \gamma_2 \text{påske}_t + \gamma_3 \text{pinse}_t + \gamma_4 \text{etterpinse}_t + \gamma_5 \text{KH}_t + \gamma_6 \text{TD16}_t + \gamma_7 D_{1.\text{jan}} + \gamma_8 D_{1.\text{mai}} + \gamma_9 D_{17.\text{mai}} + \gamma_{10} D_{24.\text{des}} + \gamma_{11} D_{25.\text{des}} + \gamma_{12} D_{26.\text{des}} + \gamma_{13} D_{27.\text{des}} + \gamma_{14} D_{28.\text{des}} + \gamma_{15} D_{29.\text{des}} + u_t$$

#### hvor:

estimeringsperiode: 1993(1)-2008(5)

alle variabler er log transformert

TD16: Mand=tirs...=lørdag: lik vekt, mot søndag

$u_t \sim \text{ARIMA modell}$

Vi estimerer også modell II ved hjelp av maximum likelihood. Tabell 3.2 på neste side viser estimeringsresultatene. Vi ser av disse resultatene at det bare er Kristi Himmelfartsdag som er signifikant både på 5 og 10 prosent nivå, men fortegnet er ikke riktig i forhold til våre forventninger. Ukedagseffekten er ikke signifikant, men har riktig fortegn. Alle dummy variablene for røde dager og innklemdager er fortsatt ikke-signifikant og har svært lave t-verdier. X12arima velger ARIMA modellen (0,1,2)(0,1,1).

Tabell 3.2

Variabel	Parameterestimat	Standardfeil	T-verdi
leap year	0,0358	0,02183	1,64 (*)
førpåske	-0,0167	0,01815	-0,92
Påske	0,0023	0,01985	0,12
Pinse	-0,0231	0,01885	-1,23
etterpinse	0,0327	0,01803	1,81 (*)
KH	0,0499	0,02429	2,06 (**)
TD16	0,0014	0,00088	1,61
D <sub>1.jan</sub>	-0,0231	0,03265	-0,71
D <sub>1.mai</sub>	0,0171	0,03099	0,55
D <sub>17.mai</sub>	0,0131	0,04118	0,32
D <sub>24.des</sub>	0,0179	0,03419	0,52
D <sub>25.des</sub>	0,0033	0,04105	0,08
D <sub>26.des</sub>	-0,0177	0,035	-0,51
D <sub>27.des</sub>	0,0485	0,04738	1,02
D <sub>28.des</sub>	0,0283	0,03846	0,74
D <sub>29.des</sub>	-0,0224	0,03747	-0,6
$\theta_1$	0,4	0,06439	
$\theta_2$	0,3725	0,06406	
$\Theta$	1	0,05523	

I både modell I og II faller alle variabler til slutt vekk bortsett fra ukedagseffekten når vi tar ut en og en variabel som er ikke-signifikant. Siden det ikke er helt klart utifra forkunnskap hvilken ukedagseffekt som passer best, velger vi den modellen som oppnår lavest AICC og Hannan Quinn.

Modell I kun med TD15:

$$O_t = \beta_0 TD15 + z_t$$

$z_t \sim ARIMA$  modell

Vi estimerer likningen ovenfor med maximum likelihood og oppnår følgende resultater:

Tabell 3.3

Variabel	Parameterestimat	Standardfeil	T-verdi
TD15	0,0023	0,00094	2,5 (**)
$\theta_1$	0,4505	0,0649	
$\theta_2$	0,3542	0,06455	
$\Theta$	1	0,05721	
<b>Informasjonskriterier</b>	<b>Verdi</b>		
AICC	4860,6334		
Hannan Quinn	4866,657		

Vi ser av regresjonsresultatene at ukedagseffekten er signifikant på 5 prosent nivå og har riktig fortegn. Estimatet uttrykker at vi bruker mer strøm på hverdager enn i helgen. X12arima velger (0,1,2)(0,1,1)-modellen.

Modell II kun med TD16:

$$O_t = \gamma_0 TD16 + u_t$$

$u_t \sim ARIMA$ modell

Tabell 3.4

Variabel	Parameterestimat	Standardfeil	T-verdi
TD16	0,0016	0,00081	1,99 (**)
$\theta_1$	0,4577	0,06507	
$\theta_2$	0,3477	0,06475	
$\Theta$	1	0,05647	
<b>Informasjonskriterier</b>	<b>Verdi</b>		
AICC	4862,8401		
Hannan Quinn	4868,8637		

I denne modellen er også ukedagseffekten signifikant på 5 prosent nivå og har riktig fortegn. (0,1,2)(0,1,1)-modellen er også valgt her.

Da AICC og Hannan Quinn har lavere verdier i modellen med TD15, velger vi å bruke denne spesifikasjonen for ukedagseffekt. Selv om vi antakeligvis bruker litt mindre strøm på en søndag enn på en lørdag, er nok likevel strømforbruket lavere på en lørdag sammenliknet med hverdagene. Dette forklarer muligens hvorfor TD15 gir lavere AICC og Hannan Quinn. **Parameterestimatene i modell I uttrykker hvor mye logaritmen til rådata er prekorrigert, med motsatt fortegn.**

For å oppnå stabilitet, velger vi å kjøre med modell I og kun med regressoren TD15 i perioden 2008(5)-2009(5). En annen viktig grunn for dette er også at det vil være tidkrevende å vurdere nye modeller fra måned til måned, og konklusjonene ovenfor vil neppe bli endret når vi får flere månedlige observasjoner.

## 4. Sesongjustering av prekorrigert serie

Den prekorrigerte serien,  $NyO_t$  er beskrevet gjennom følgende log additiv modell:

$$NyO_t = \log(S_t) + \log(T_t) + \log(I_t)$$

Hvor:

$S_t$ : Sesongkomponent

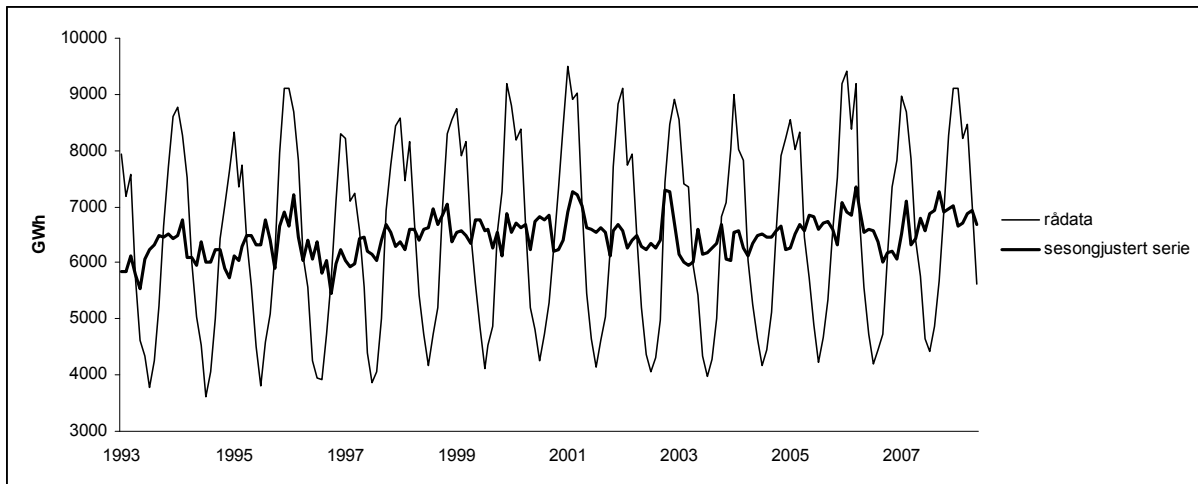
$T_t$ : Trendkomponent

$I_t$ : Irregulær komponent

Sesong-, trend- og irregulær komponent samt sesongjusterte tall,  $A_t$ , estimeres gjennom flere trinn. I appendix IV vises det hvordan de initiale estimeringene utføres.

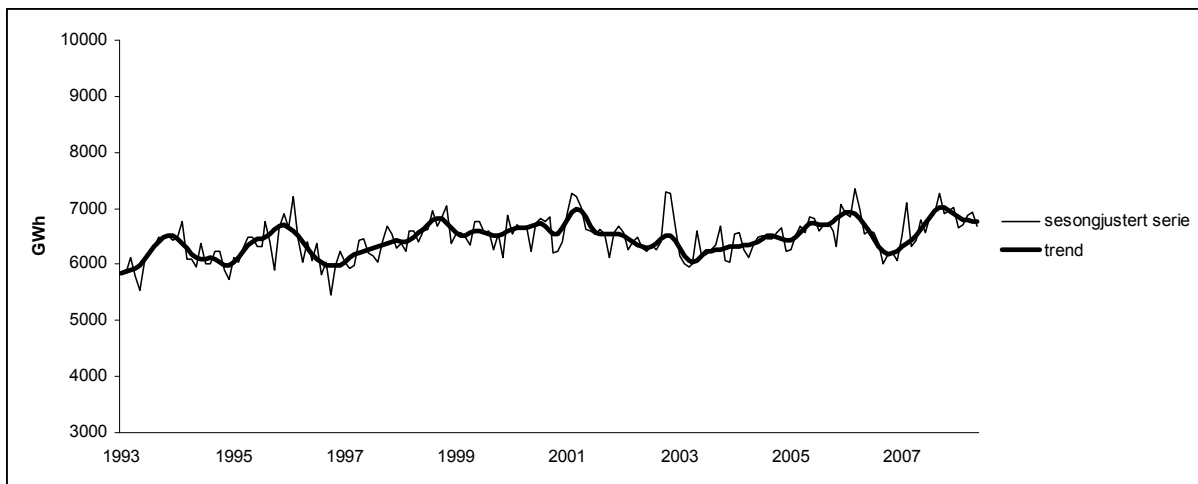
## 4.1 Sesongjustert serie, trend, irregulær komponent og sesongkomponent

Figur 4.1: Sesongjustert serie og rådata



Grafen over viser estimert sesongjustert serie for perioden 1993(1)-2008(5) basert på alle trinnene i beregningsmetoden samt rådata i samme periode. Sesongjusterte tall er rensket for all systematisk sesongvariasjon og er satt sammen av de to komponentene trend og irregulær komponent. Den irregulære komponenten reflekterer tilfeldig variasjon som ikke gjentar seg fra år til år og kan derfor tolkes som et støyledd i den log additive modellen. Ved å fjerne den irregulære komponenten fra den sesongjusterte serien framkommer trenden, som beskriver den langsiktige utviklingen i data. Figur 4.1 viser nytteverdien av sesongjustering. Vi kan nå på en meningsfull måte sammenlikne alle måneder med hverandre, for eksempel januar med juli.

Figur 4.2: Sesongjustert serie og trend



Figur 4.2 viser estimert trend for forbruk i alminnelig forsyning i perioden 1993(1)-2008(5). Sammenliknet med figur 4.1 av rådata, ser vi at grafen med trend gir bedre og mer informasjon om den langsiktige utviklingen i forbruket i alminnelig forsyning. Grafen indikerer en positiv underliggende vekst i strømforbruket på nittitallet, mens det ser ut til at forbruket flater ut i perioden 2000(1)-2008(5) på rundt 6630 GWh. En utflating av strømforbruket på 2000-tallet kan ha sammenheng med flere forhold.



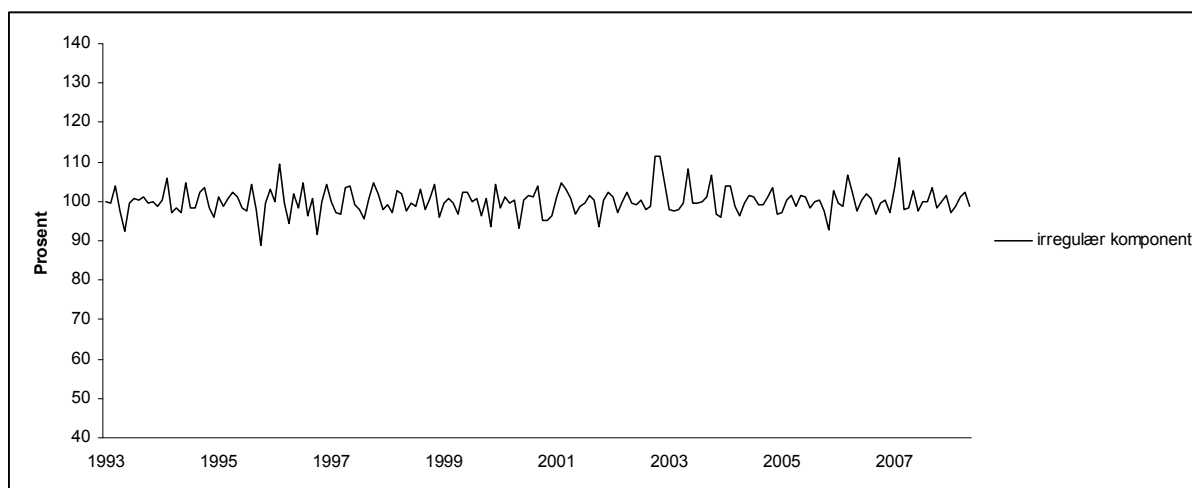
For det første; større prisvolatilitet på 2000-tallet og høye strømpriser i 2003, 2004 og 2006 kan ha utløst investeringer i ENØK-utstyr (sparedusjer, varmepumper, bedre isolasjon i bygninger etc.). Dette kan ha dempet veksten i forbruket.

For det andre, en mulig ekspansjon i mindre kraftintensive næringer og kontraksjon i mer kraftintensive sektorer kan også ha bidratt til en utflating i strømforbruket innenfor alminnelig forsyning.

For det tredje, økt fokus på miljøvern og global oppvarming kan også ha hatt en dempende effekt på forbruket.

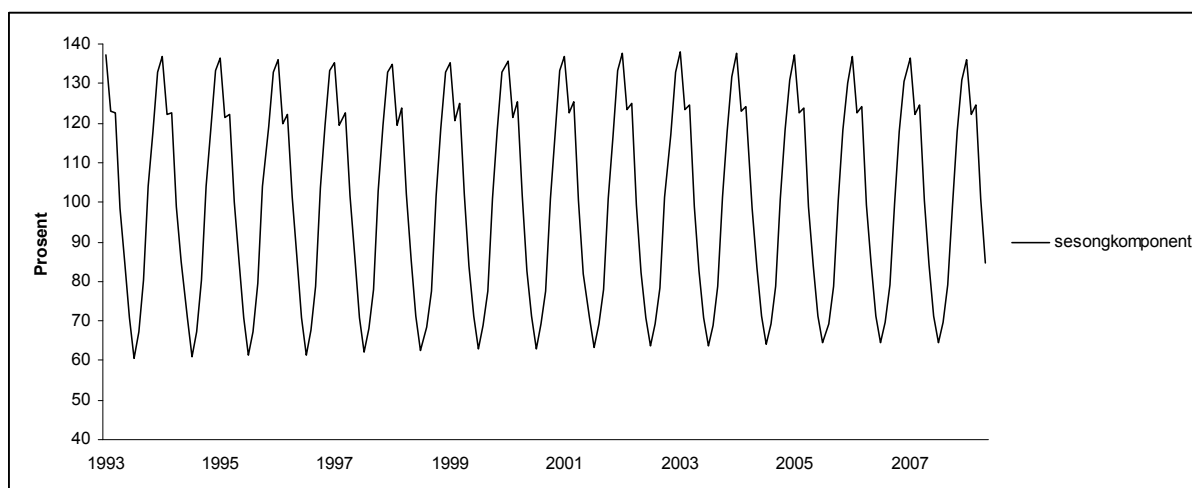
Avvikene mellom trend og sesongjustert serie reflekterer den irregulære komponenten.

Figur 4.3: Irregulær komponent



Av figur 4.3 ser vi at den irregulære komponenten preges av flere kraftige utslag i perioden 1993(1)-2008(5). Den irregulære komponenten framkommer ved å se på forholdet mellom sesongjusterte tall og trend.

Figur 4.4: Sesongkomponent



Den tidsserien som dannes av sesongfaktorene kalles sesongkomponenten. Sesongfaktorene for januar, februar,... og desember uttrykker hvor mye strøm som brukes i prosent av den mengden som trenden indikerer for disse månedene. I januar og juli 1993 er sesongfaktoren lik henholdsvis

137,1 og 60,7 prosent. Disse tallene har følgende tolkning: I januar og juli bruker vi henholdsvis 137,1 og 60,7 prosent av den mengden som trenden indikerer.

## 4.2 Tolkning av sesongjusterte tall

Boksene under viser fire forskjellige tilfeller på månedlige endringer i ujustert og sesongjustert serie. Generelt for alle tabellene gjelder at sesongjustert vekst fra en måned til en annen må tolkes med forsiktighet, og man kan ikke legge for stor vekt på enkelttall. Dette kommer av at sesongjusterte tall også inneholder et støyledd, som fanger opp tilfeldig variasjon. X12arima fjerner den systematiske (gjennomsnittlige) sesongvariasjonen fra rådata, men tar ikke ut sjokk. For eksempel var mars 2007 den varmeste mars som noen gang er registrert for Norge som helhet. Vanligvis er strømforbruket høyere i mars enn i februar, men på grunn av svært høy temperatur i mars 2007 var det en nedgang i det ujusterte forbruket til alminnelig forsyning fra februar til mars på hele 9,7 prosent. Dette bidro også til at det ble en stor nedgang i det sesongjusterte forbruket, som var 10,9 prosent. For å kunne avgjøre om den underliggende veksten er positiv eller negativ, må man derfor se sesongjusterte tall over tid og i sammenheng med forhold som har betydning for strømforbruket.

<b>Tilfelle I</b>			
	<b>August 2007</b>	<b>September 2007</b>	<b>Endring i prosent</b>
Forbruk i alminnelig forsyning, ujustert	4859	5679	16,9
Forbruk i alminnelig forsyning, sesongjustert	6945	7256	4,5

Av økningen i det ujusterte forbruket på 16,9 prosent kan 12,4 prosent ses i sammenheng med sesongvariasjon (matematisk litt upresist). Den underliggende veksten fra august til september er positiv.

<b>Tilfelle II</b>			
	<b>August 2006</b>	<b>September 2006</b>	<b>Endring i prosent</b>
Forbruk i alminnelig forsyning, ujustert	4450	4738	6,5
Forbruk i alminnelig forsyning, sesongjustert	6363	6011	-5,5

Hele økningen i det ujusterte forbruket kan ses i sammenheng med sesongvariasjon og den underliggende veksten er negativ (matematisk litt upresist). I dette tilfellet har vi en positiv vekst i det ujusterte forbruket, mens sesongjustert vekst er negativ. Dette henger sammen med at det ujusterte forbruket i alminnelig forsyning vanligvis øker mer enn 6,5 prosent fra august til september. I 2000, 2001, 2002, 2003, 2004 og 2005 var veksten i det ujusterte forbruket fra august til september henholdsvis 12,0, 9,0, 15,7, 17,0, 14,9 og 14,1 prosent.

<b>Tilfelle III</b>			
	<b>April 2008</b>	<b>Mai 2008</b>	<b>Endring i prosent</b>
Forbruk i alminnelig forsyning, ujustert	7034	5622	-20,1
Forbruk i alminnelig forsyning, sesongjustert	6944	6673	-3,9

Av nedgangen i det ujusterte forbruket på 20,1 prosent kan 16,2 prosent ses i sammenheng med sesongvariasjon (matematisk litt upresist). Den underliggende veksten er negativ.

<b>Tilfelle IV</b>			
	<b>April 2007</b>	<b>Mai 2007</b>	<b>Endring i prosent</b>
Forbruk i alminnelig forsyning, ujustert	6407	5770	-9,9
Forbruk i alminnelig forsyning, sesongjustert	6418	6790	5,8

Hele nedgangen i det ujusterte forbruket på 9,9 prosent kan ses i sammenheng med sesongvariasjon og den underliggende veksten er positiv (matematisk litt upresist). I dette tilfellet har vi en negativ vekst i det ujusterte forbruket, mens sesongjustert vekst er positiv. Dette henger sammen med at det ujusterte forbruket vanligvis faller mer enn 9,9 prosent fra april til mai.

### 4.3 Kvalitetsindikatorer for sesongjustering

X12Arima presenterer en rekke kvalitetsindikatorer som uttrykker hvor god sesongjusteringen og metoden er, og i dette avsnittet vil vi presentere flere av disse målene.

#### (i) M1, M2,..., M11 og Q

Tabell 4.1 viser 11 mål på hvor god sesongjusteringen er. Verdiene kan variere mellom 0 og 3, men kvaliteten er kun tilfredsstillende for verdier under 1. I tillegg inneholder tabellen en Q-verdi, som er en vektet samleindikator for alle de 11 målene bortsett fra M2. Forkastningsområde er  $Q > 1$ . Jo, lavere Q-verdi jo bedre.

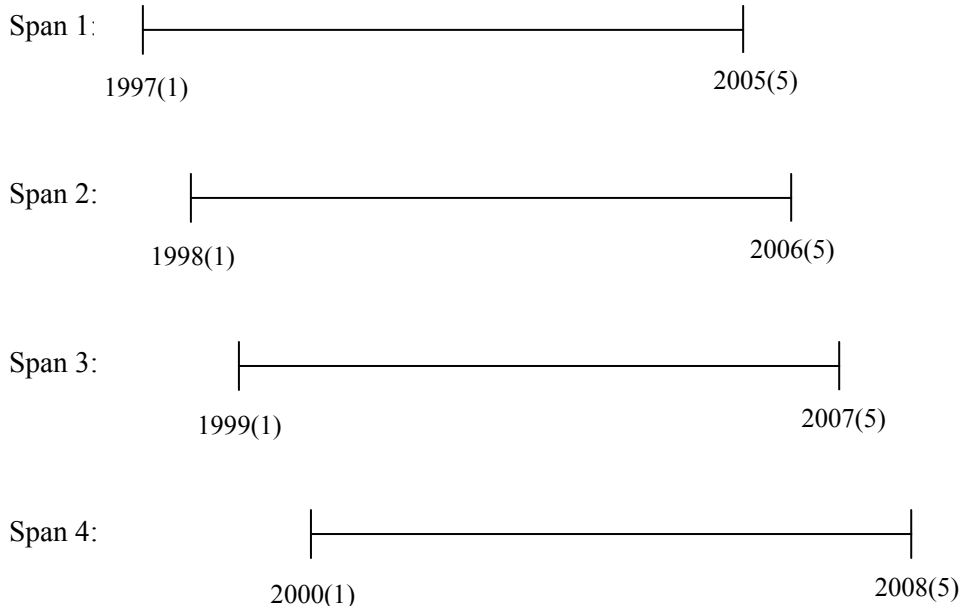
Tabell 4.1

<b>Mål</b>	<b>Verdi</b>	<b>Beskrivelse av indikator</b>
M1	0,093	Måler relativt bidrag fra irregulær komponent i den totale variasjonen
M2	0,1	Måler relativt bidrag fra irregulær komponent i den stasjonære delen
M3	1,189	Måler endring fra måned til måned i irregulær komponent i forhold til måned til måned endring i trenden
M4	0,544	Måler autokorrelasjon i irregulær komponent
M5	1,065	Måler antall måneder for at endring i trenden skal bli større enn endring i irregulær komponent
M6	0,78	Måler endring fra år til år i irregulær komponent i forhold til endring i sesongkomponenten
M7	0,093	Måler grad av bevegelig sesongmønster i forhold til stabilt sesongmønster
M8	0,163	Måler grad av fluktuasjoner i sesongkomponenten gjennom hele serien
M9	0,046	Måler grad av lineær bevegelse i sesongkomponenten i hele serien
M10	0,141	Måler grad av fluktuasjoner i sesongkomponenten de siste årene
M11	0,12	Måler grad av lineær bevegelse i sesongkomponenten de siste årene
Q	0,46	Vektet samleindikator for alle målene bortsett fra M2

Resultatene viser at vi alt i alt oppnår en god direkte sesongjustering av forbruk innenfor alminnelig forsyning. Verdiene for de to viktigste målene, M7 og Q, er svært bra. Både M3 og M5 ligger høyere enn kritisk verdi, noe som ikke er gunstig. Dette kommer av at vi da har kraftige utslag i den irregulære komponenten.

### (ii) Sliding spans

Det er sliding span testen som statistikkbyrået i USA, US Census vektlegger tyngst i vurderingen av hvor god sesongjusteringen er. I sliding span testen deles utvalget opp i fire delutvalg. Hvert delutvalg eller span inneholder 101 observasjoner og behandles akkurat som den originale tidsserien.



Formålet med denne testen er å vurdere stabiliteten for sesongkomponenten og sesongjusterte tall. Dette gjøres ved at det estimeres sesongkomponenter og sesongjusterte tall for hver måned i alle delutvalgene. Hvis prosentvis differanse mellom høyeste og laveste verdi for sesongkomponenten i delutvalg 1, 2, 3 og 4 for samme måned og år overstiger 3 %, er dette ikke gunstig. For å vurdere stabiliteten til sesongjusterte tall, beregnes relativ endring i prosent fra måned til måned i alle delutvalgene. Deretter ser man på differansen mellom høyeste og laveste verdi i delutvalg 1, 2, 3 og 4 for samme måned og år. Også her er kritisk verdi 3 %. Tabell 4.2 viser resultatene fra sliding span testen.

Tabell 4.2

Mål	Verdi	Beskrivelse av indikator
S(%)	14.2 %	Prosentandelen måneder hvor sesongkomponenten er definert som upålitelig
MM(%)	17.9 %	Prosentandelen måneder hvor sesongjusterte tall er definert som upålitelig
		<b>Kritisk verdi for S(%) og MM(%) er hhv 15 og 35%</b>

Vi ser av resultatene at stabiliteten for sesongkomponenten og sesongjusterte tall er tilfredsstillende.

### (iii) F-tester for sesongmønster og sesongbevegelse

Disse statistiske testene gir informasjon om sesongmønsteret og sesongbevegelsen er signifikant. I figur 2.2 på side 6 så vi at sesongmønsteret er karakterisert ved en U-form gjennom året. For å oppnå en god sesongjustering, er det viktig å ha et veldefinert sesongmønster og at det ikke er bevegelse i sesongmønsteret. Hvis sesongmønsteret endrer seg over tid, vil det skape inkonsistens i

sesongjusteringen. Statistiske tester viser at sesongmønsteret er signifikant på 5 prosent nivå med F-verdi lik 681,203, mens sesongbevegelsen er ikke signifikant på 5 prosent nivå med F-verdi lik 1,576.

#### 4.4 Konklusjon

Alle tester for sesongvariasjon viser at strømforbruket innenfor alminnelig forsyning har et veldefinert sesongmønster gjennom året. Et plott av rådata for hvert år over hverandre i perioden 1993-2007 viser at sesongmønsteret er karakterisert ved en U-form gjennom året. Da det er systematiske forskjeller i strømforbruket mellom mange av månedene, som gjentar seg fra år til år, vil en sammenlikning av tall fra måned til måned ikke si noe om den reelle utviklingen i forbruket. For å kunne følge strømforbruket gjennom året på en god måte, rensker vi rådata for sesongvariasjon. Det er av stor verdi å kunne sammenlikne tall fra måned til måned i tillegg til å følge den underliggende utviklingen i strømforbruket.

Ulike kvalitetsindikatorer for sesongjustering viser at vi alt i alt oppnår en god sesongjustering. Dette henger sammen med det veldefinerte sesongmønsteret i rådata. M7, som måler graden av bevegelig sesongmønster i forhold til stabilt sesongmønster, er lik 0,093, mens det veide målet for alle M-indikatorer, Q er lik 0,46. Da kritisk verdi for begge disse målene er 1, er dette svært gode resultater. Bare to av de elleve M-målene ligger over kritisk verdi. M3 og M5 er lik henholdsvis 1,189 og 1,065, noe som ikke er gunstig. Dette kommer av at vi da har kraftige utslag i den irregulære komponenten. Siden månedstemperaturer kan svinge kraftig fra år til år, vil dette også gi stor variasjon i forbruket (se figur 2.2 og kommentarer på side 6) og markante utslag i den irregulære komponenten. Sesongjusterte tall er satt sammen av trend og irregulær komponent. Derfor må sesongjusterte tall ses over tid og i sammenheng med forhold som har betydning for strømforbruket.

## **Kildehenvisning**

Hamre Jørn Ivar (2004): Sesongjustering av hovedseriene i AKU. Dokumentasjon av ny metode og resultater

Quang Pham Dinh (2001): Innføring i tidsserier – sesongjustering og X-12-ARIMA

Løvås Gunnar G (1999): Statistikk – for universiteter og høyskoler

Harris R. (1995): Cointegration Analysis in Econometric Modelling

Hill R.C, Griffiths W.E, og Judge G. G (2001): Undergraduate Econometrics

Gourieroux C og Jasiak J. (2001): Financial Econometrics

## Appendix I: Utleddning av Dickey-Fuller test for en autoregressiv prosess av orden 4

Det partielle autokorrelogrammet indikerer at rådata,  $O_t$  er en autoregressiv prosess av orden 4. Derfor kan vi skrive følgende:

$$O_t = \phi_1 O_{t-1} + \phi_2 O_{t-2} + \phi_3 O_{t-3} + \phi_4 O_{t-4} + u_t$$

$u_t$  er et white noise støyledd

Ved å trekke fra  $O_{t-1}$  på begge sider av likhetstegnet, får vi følgende:

$$\Delta O_t = (\phi_1 - 1)O_{t-1} + \phi_2 O_{t-2} + \phi_3 O_{t-3} + \phi_4 O_{t-4} + u_t$$

Siden  $\Delta O_t = O_t - O_{t-1}$  kan vi sette inn følgende uttrykk for  $O_{t-2}$ ,  $O_{t-3}$  og  $O_{t-4}$ :

$$\Delta O_t = (\phi_1 - 1)O_{t-1} + \phi_2(O_{t-1} - \Delta O_{t-1}) + \phi_3(O_{t-1} - \Delta O_{t-1} - \Delta O_{t-2}) + \phi_4(O_{t-1} - \Delta O_{t-1} - \Delta O_{t-2} - \Delta O_{t-3}) + u_t$$

Ved å samle like ledd kan likningen over skrives som:

$$\Delta O_t = (\phi_1 + \phi_2 + \phi_3 + \phi_4 - 1)O_{t-1} + (-\phi_2 - \phi_3 - \phi_4)\Delta O_{t-1} + (-\phi_3 - \phi_4)\Delta O_{t-2} + (-\phi_4)\Delta O_{t-3} + u_t$$

La:

$$(\phi_1 + \phi_2 + \phi_3 + \phi_4 - 1) = \gamma_1$$

$$(-\phi_2 - \phi_3 - \phi_4) = \gamma_2$$

$$(-\phi_3 - \phi_4) = \gamma_3$$

$$(-\phi_4) = \gamma_4$$

Vi ender da opp med samme likning som på side 7:

$$\Delta O_t = \gamma_1 O_{t-1} + \gamma_2 \Delta O_{t-1} + \gamma_3 \Delta O_{t-2} + \gamma_4 \Delta O_{t-3} + u_t$$

## Appendix II: De fem ARIMA-modellene som X12arima vurderer

Modell (a):

(0,1,1)(0,1,1)

$$O_t = O_{t-1} + O_{t-2} - O_{t-3} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \Theta_1 \varepsilon_{t-2} + \theta_1 \Theta_1 \varepsilon_{t-3}$$

Modell (b):

(0,1,2)(0,1,1)

$$O_t = O_{t-1} + O_{t-2} - O_{t-3} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \Theta \varepsilon_{t-2} + \theta_1 \Theta \varepsilon_{t-3} + \theta_2 \Theta \varepsilon_{t-4}$$

Modell (c):

(2,1,0)(0,1,1)

$$O_t = O_{t-1} + \phi_1 O_{t-1} - \phi_1 O_{t-2} + \phi_2 O_{t-2} - \phi_2 O_{t-3} + O_{t-2} - O_{t-3} - \phi_1 O_{t-3} + \phi_1 O_{t-4} - \phi_2 O_{t-4} + \phi_2 O_{t-5} + \varepsilon_t - \Theta \varepsilon_{t-2}$$

Modell (d):

(0,2,2)(0,1,1)

$$O_t = 2O_{t-1} - O_{t-2} + O_{t-2} - 2O_{t-3} + O_{t-4} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \Theta \varepsilon_{t-2} + \theta_1 \Theta \varepsilon_{t-3} + \theta_2 \Theta \varepsilon_{t-4}$$

Modell (e):

(2,1,2)(0,1,1)

$$O_t = O_{t-1} + \phi_1 O_{t-1} - \phi_1 O_{t-2} - \phi_2 O_{t-3} + O_{t-2} - O_{t-3} - \phi_1 O_{t-3} + \phi_1 O_{t-4} - \phi_2 O_{t-4} + \phi_2 O_{t-5} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \Theta \varepsilon_{t-2} + \theta_1 \Theta \varepsilon_{t-3} + \theta_2 \Theta \varepsilon_{t-4}$$



### Appendix III: Utledning av modell I hvis støyledet følger en ARIMA modell

Modell I kan skrives mer kompakt med følgende likning:

$$O_t = \sum_{i=0}^{15} \beta_i X_{i,t} + z_t$$

$\beta_i$  : regresjonsparametere ,  $X_{i,t}$  : regresjons variable . Siden  $z_t$  følger en arima modell, gjelder følgende generelle likning:

$$z_t = \frac{\theta(B)\Theta(B^s)}{\phi(B)\Phi(B^s)\Delta^d\Delta_s^D} \varepsilon_t \quad , \quad \varepsilon_t \sim iid(0, \sigma^2)$$

Hvor:

$$\Delta^d = (1-B)^d$$

$$\Delta_s^D = (1-B^{12})^D$$

$$\phi(B) = (1-\phi_1 B - \dots - \phi_p B^p)$$

$$\Phi(B^s) = (1-\Phi_1 B^{12} - \dots - \Phi_p B^{12 \cdot p})$$

$$\theta(B) = (1-\theta_1 B - \dots - \theta_q B^q)$$

$$\Theta(B^s) = (1-\Theta_1 B^{12} - \dots - \Theta_Q B^{Q \cdot 12})$$

B er kun en lag operator

I Airline modellen er ARIMA(p,d,q)(P,D,Q)=ARIMA(0,1,1)(0,1,1). Ved å anta at  $z_t$  følger Airline modellen, kan den første likningen skrives som:

$$(1-B)(1-B^{12})O_t = \sum_{i=0}^{15} \beta_i (1-B)(1-B^{12})X_{i,t} + (1-\theta_1 B)(1-\Theta_1 B^{12})\varepsilon_t$$

Ved å multiplisere ut parentesene får vi følgende:

$$(1-B-B^{12}+B^{13})O_t = \sum_{i=0}^{15} \beta_i (1-B-B^{12}+B^{13})X_{i,t} + (1-\Theta_1 B^{12}-\theta_1 B+\theta_1 \Theta_1 B^{13})\varepsilon_t$$

Dette kan vi skrive som:

$$O_t - O_{t-1} - O_{t-12} + O_{t-13} = \sum_{i=0}^{15} \beta_i (X_{i,t} - X_{i,t-1} - X_{i,t-12} + X_{i,t-13}) + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \Theta_1 \varepsilon_{t-12} + \theta_1 \Theta_1 \varepsilon_{t-13}$$

$\beta_i, \theta_1, \Theta_1$  blir bestemt gjennom kompliserte iterasjoner og maximum likelihood estimering.

Den nye høyresidevariabelen  $O_t - O_{t-1} - O_{t-12} + O_{t-13}$  er stasjonær.

## Appendix IV: Initiale beregningsmetoder for sesong, trend og irregulær komponent

### Trinn I. Initiale estimeringer.

- (i) Vi estimerer trend ved et sentrert 12 måneders glidende gjennomsnitt

$$T_t^1 = \frac{1}{24} NyO_{t-6} + \frac{1}{12} NyO_{t-5} + \dots + \frac{1}{12} NyO_t + \dots + \frac{1}{12} NyO_{t+5} + \frac{1}{24} NyO_{t+6}$$

$T_t^1$  er trenden i første estimeringsrunde

- (ii) Vi beregner  $SI_t$  rater

$$SI_t^1 = \frac{NyO_t}{T_t^1}$$

- (iii) Vi estimerer innledende skritt for sesongfaktoren  $\hat{S}_t^1$  ved 3x3 filteret

$$\hat{S}_t^1 = \frac{1}{9} SI_{t-24}^1 + \frac{2}{9} SI_{t-12}^1 + \frac{3}{9} SI_t^1 + \frac{2}{9} SI_{t+12}^1 + \frac{1}{9} SI_{t+24}^1$$

- (iv) Beregner så sesongfaktor på følgende måte

$$S_t^1 = \frac{\hat{S}_t^1}{\frac{1}{24} \hat{S}_{t-6}^1 + \frac{1}{12} \hat{S}_{t-5}^1 + \dots + \frac{1}{12} \hat{S}_{t+5}^1 + \frac{1}{24} \hat{S}_{t+6}^1}$$

- (v) Vi kan nå beregne forløpige sesongjusterte tall,  $A_t^1$ , ved følgende formel

$$A_t^1 = \frac{NyO_t}{S_t^1}$$