

Dinh Quang Pham

Figurer for sesongjustering

© Statistisk sentralbyrå, februar 2011 Ved bruk av materiale fra denne publikasjonen skal Statistisk sentralbyrå oppgis som kilde.	Standardtegn i tabeller	Symbol
ISBN 978-82-537-8055-9 Trykt versjon	Tall kan ikke forekomme	.
ISBN 978-82-537-8056-6 Elektronisk versjon	Oppgave mangler	..
ISSN 1891-5906	Oppgave mangler foreløpig	...
Emne: 00.90	Tall kan ikke offentliggjøres	:
Trykk: Statistisk sentralbyrå	Null	-
	Mindre enn 0,5 av den brukte enheten	0
	Mindre enn 0,05 av den brukte enheten	0,0
	Foreløpig tall	*
	Brudd i den loddrette serien	—
	Brudd i den vannrette serien	
	Desimaltegn	,

Innhold

1	Innledning	5
2	Beskrivelse av figurer	5
2.1	FIGURER FOR OBSERVERTE TALL	7
2.1.1	Figur 1.a. Ujusterte verdier	7
2.1.2	Figur 1.b. Ujusterte verdier for hver av årets tolv måneder	8
2.1.3	Figur 1.c. Ujusterte verdier for hver av årets tolv måneder (etter standardisering av tall for hvert år med gjennomsnitt og standardavvik i året)	8
2.1.4	Figur 1.d. Ujusterte verdier fra år til år for hver enkelt måned	9
2.1.5	Figur 1.e. Boxplott av de ujusterte verdiene for tolv måneder	9
2.1.6	Figur 1.f. Autokorrelasjonsfunksjonen (ACF) av ujusterte verdier	9
2.1.7	Figur 1.g. Spekteret (spektraltetthet)	10
2.1.8	Oppsummering av figurene 1.a-g	10
2.2	FIGURER FOR KORRIGERING AV KALENDEREFFEKTER OG EKSTREMVERDIER	11
2.2.1	Figur 2.a. Ujusterte og korrigerede verdier for effektene av ukedager og ekstremverdier	11
2.2.2	Figur 2.b. Ujusterte og korrigerede verdier for bevegelige helligdager	12
2.2.3	Figur 2.c. Ekstremverdier (intervensjoner)	13
2.2.4	Figur 2.d. Avvikene mellom ujusterte og korrigerede verdier	14
2.2.5	Figur 2.e. Framskrivninger for 12 måneder	14
2.2.6	Figur 2.f. Korrigerede verdier for hver av årets tolv måneder (etter standardisering tall for hvert år med gjennomsnittet og standardavviket i året)	14
2.2.7	Figur 2.g. Spekteret for korrigerede verdier	15
2.2.8	Figur 2.h. Boxplott for tolv måneder for ujusterte og korrigerede verdier	15
2.2.9	Oppsummering av figur 2	15
2.3	FIGURER FOR UJUSTERTE, SESONGJUSTERTE VERDIER OG TREND	15
2.3.1	Figur 3.a. Ujusterte verdiene og trend	18
2.3.2	Figur 3.b. Sesongjusterte verdier og trend	18
2.3.3	Figur 3.c. Sesongjusterte verdier for hver av årets 12 måneder	18
2.3.4	Figur 3.d. Sesongjusterte tall fra år til år for hver enkel måned	18
2.3.5	Figur 3.e og figur 3.f. Endring i prosent fra måned til måned i ujusterte og sesongjusterte verdier	19
2.3.6	Figur 3.g og 3.h. Revisjoner av sesongjusterte verdier og trend	19
2.3.7	Figur 3.i. Avvikene mellom ujusterte og sesongjusterte verdier for tolv måneder	19
2.4	SESONGKOMPONENTEN	20
2.4.1	Figur 4.a. Sesongkomponenten	20
2.4.2	Figur 4.b. Sesongkomponenten for hver av årets 12 måneder	20
2.4.3	Figur 4.c. Sesongkomponenten fra år til år for hver enkelt måned	20
2.4.4	Figur 4.d. Sammenligning sesongmønstrene i figur 1.c og 4.b	20
2.4.5	Oppsummering av figurene 4.a-d	20
2.5	IRREGULÆRKOMPONENTEN (I_t)	21
2.5.1	Figur 5.a. Irregulærkomponenten	21
2.5.2	Figur 5.b. Irregulærkomponenten fra år til år for hver enkel måned	21
2.5.3	Autokorrelasjonsfunksjonen (ACF) av irregulærkomponenten	21
2.5.4	Figur 5.d. Spekteret av sesong- og irregulærkomponenten	21
2.5.5	Figur 5.e. Scatter plott av irregulærkomponenten	21
2.6	SI KOMPONENTEN	21
2.7	DIAGNOSTIKK	21

2.7.1	Figur 7.a. De 11 målene M_1 - M_{11}	22
2.7.2	Figur 7.b. Sliding spans teknikk	23
2.7.3	Figur 7.c. Oppsummering for sliding spans og kvalitetsmålene	24
2.7.4	Kommentarer for kvalitetsmål og sliding spans teknikk	24
2.8	SKATTERPLOTT FOR KOMPONENTENE	24
3	Oppsummering	24

Tabeller

1	<i>Bilsalg. Ujusterte verdier i 1995 og 1996</i>	7
2	<i>Bilsalg. Standardavvik for tolv måneder</i>	8
3	<i>Bilsalg. Ujusterte verdiene for 1990 og 1996 før og etter standardisering</i>	9
4	<i>Spektraltettheten av bilsalg</i>	10
5	<i>Estimeringen for kalendereffekter og ekstremverdier</i>	12
6	<i>Korrigerede faktorer for kalendereffekter og ekstremverdier</i>	12
7	<i>Ujusterte og korrigerede verdiene i tidspunktene for ekstremverdier</i>	14
8	<i>Framskrivinger i 2010 og 95% konfidensintervaller</i>	15
9	<i>Standardavvikene av ujusterte og korrigerede verdier etter standardisering</i>	15
10	<i>Vektene for M_1-M_{11}</i>	23
11	<i>Evalueringen for sesongjustering ved sliding spans teknikken</i>	23
12	<i>Verdiene av M_1-M_{11} etter sesongjustering for bilsalg</i>	23

Sammendrag

Hovedformål med notatet er å presentere figurer som et godt hjelpeverktøy for vurdering av resultater fra en sesongjustering. Vi har laget et program med R slik at brukere kan lage figurer etter å ha kjørt X-12-ARIMA for en tidsserie. Programmet er veldig enkelt å bruke.

1 Innledning

En tidsserie er en samling påfølgende observasjoner i tid. For eksempel måler vi antall arbeidsledige i Norge hver måned. En tidsserie blir dannet. Hvis vi begrenser tallene til de mellom januar 1990 og desember 2009, får vi en tidsserie O_t , der t går fra og med januar 1990 til og med desember 2009. O_t er antall arbeidsledige på tidspunkt t . For å se variasjoner i tidsserien plotter vi verdi mot tidsforløp. Mønsteret i de fleste økonomiske tidsserier gjentar seg fra år til år. For eksempel for hotellstatistikken. Vi ser at antall hotellovernattinger er høyere enn gjennomsnittet i mars, april, juni, juli og august, men mye lavere i desember. Dette gjentar seg fra år til år. Når vi plotter data over tid kan vi se et klart sesongmønster i tidsserien. Ekstremverdier kan oppdages ved et plott. Vi antar at det i tidsserien er en statistisk avhengighet mellom observasjonene på forskjellige tidspunkter og at for hver tidspunkt t er det en sannsynlighetsfordeling $F_t(o)$ for O_t . Vi skriver $F_t(o) = P(O_t < o)$. Vi beskriver tidsserien med ulike momenter. For eksempel middel, varians og autokovarians. Ved å plote autokorrelasjonsfunksjonen (ACF) får vi et godt bilde av sesongmønsteret i tidsserien. En fordel i tidsserieanalyse er at man kan beskrive variasjoner eller egenskaper av en tidsserie med figurer.

Vi sesongjusterer hver måned mange tidsserier som for eksempel produksjonsindeksen, detaljhandelvolumindeksen, hotellovernattinger, bilsalg, Sesongjusterte tall, trend, sesong- og irregulærkomponenten av disse tidsseriene må sjekkes før publisering. Det blir tungt og det krever mye tid å gjøre denne jobben ved å lese bare tall. Derfor lager vi i tillegg figurer og bruker dem som hjelpeverktøy for vurdering av resultater.

US Census har laget X-12-Graph for X-12-ARIMA. Programmet er basert på SAS/Graph og kjøres på PC, Unix og Linux maskiner. Det er enkelt i bruk og krever ikke erfaring i SAS. Det lager mange nyttige og avanserte figurer for ujusterte tall, sesongjusterte tall, trend samt sesong- og irregulærkomponenten. En ulempe ved X-12-Graph er at vi ikke kan lage flere nye figurer i programmet etter behov.

Etter vurdering av X-12-Graph og mange diskusjoner med folk som jobber med sesongjustering i SSB for sesongjustering, lager vi et lignende program med X-12-Graph for sesongjustering i SSB. Vi kan dermed endre eller legge til nye figurer i programmet etter behov. Det blir også enklere å kjøre enn X-12-Graph og det krever ingen erfaring fra brukere.

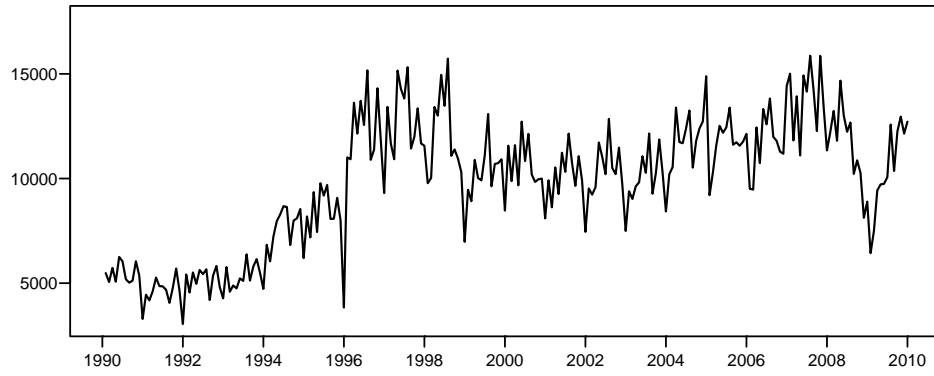
Vi beskriver figurer som er laget av vårt program i notatet ved et konkret eksempel. Det er antall førstegangsregistrerte nye biler samt bruktimporterte biler i Norge. Tidsserien blir kalt "bilsalg". Data er observert fra januar 1990 til desember 2009. I alt er det 240 observasjoner. Vi skal sesongjustere tidsserien og lage figurer for data, sesongjusterte tall, trend, sesong- og irregulærkomponenten.

2 Beskrivelse av figurer

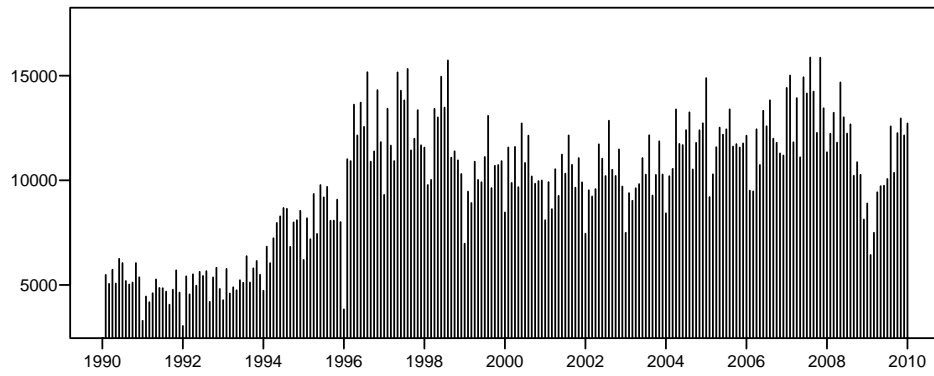
I notatet vil vi kalle rådata for ujusterte data (eller verdier eller tall), for å skille dem fra sesongjusterte verdier. Ofte bruker vi ordet rådata for å vise tall som ikke er behandlet, men det er mange tidsserier i SSB der rådata blir revidert før sesongjustering.

Vi bruker symbolene O_t for ujusterte verdi i tidspunkt t . Det kan være en estimert verdi fra en utvalgsundersøkelse; for eksempel antall arbeidsledige i Norge. Det kan også være en observert verdi; for eksempel antall hotellovernattinger i totalen i Norge. Det er mange faktorer som ikke er sesongvariasjoner eller tilfeldige variasjoner som også påvirker de ujusterte verdiene. Det kan være effektene av kalender (feks, ulike antall arbeidsdager i måneder, bevegelige helligdager som påske, pinse, Kristihimmelfartsdag) eller ekstremverdier. En verdi som er igjen etter å ha fjernet disse effektene fra den ujusterte verdien, kalles en korrigert verdi og betegnes med O_t^K . En trendverdi T_t og sesongverdi S_t er beregnet fra den korrigerede verdien O_t^K . Vi betegner en tilfeldig variasjon med I_t . Det kreves at sesongvariasjonene må være stabile for å få en god sesongjustering.

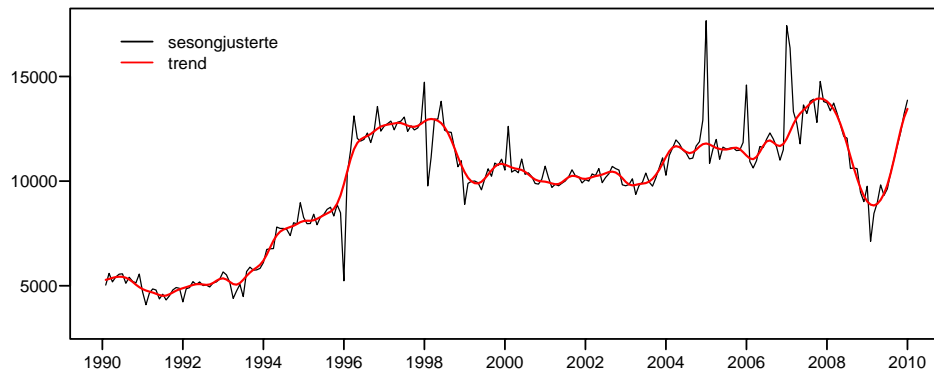
0.a. Ujusterte tall



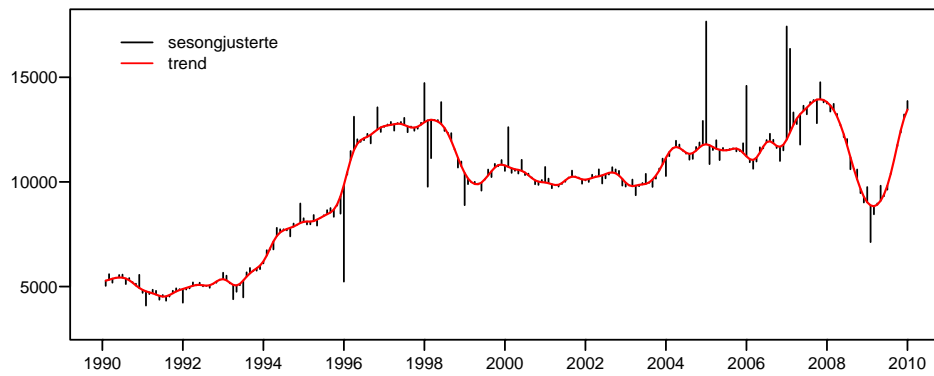
0.b. Ujusterte tall



0.c. Sesongjusterte tall og trend



0.d. sesongjusterte tall og trend



I figur 0 plotter vi ujusterte, sesongjusterte verdier og trend ved å bruke stående linjen og en linje som binder punktene sammen. Figuren 0.a er et vanlig plott for en tidsserie, der verdiene av to tidspunkter t og $t + 1$ er knyttet sammen med en linje. I figuren 0.b plotter vi de ujusterte verdiene med søyler. Figuren 0.a er bedre enn figuren 0.b når man vil vise variasjoner og trend i tidsserien.

Vi plotter sesongjusterte tall på en vanlig måte og med stående linjer i de to siste figurene. Når vi vil vise trend og variasjoner i sesongjusterte tall bruker vi figur 0.c. Når vi vil vise avstander mellom sesongjusterte tall og trend blir figuren 0.d bedre. Merk at sesongjusterte tall er trend og tilfeldige variasjoner i tillegg.

Vi bør velge en metode for å lage figurer etter behov. I de neste avsnittene vil vi presentere figurer for sesongjustering med bilsalget som et eksempel.

2.1 FIGURER FOR OBSERVERTE TALL

En tidsserie i økonomi er ofte beskrevet av en trend som viser en langtidsutvikling av tidsserien, sesongvariasjoner som viser variasjoner som gjentar seg fra år til år i samme tidspunkter eller i samme periode, og tilfeldige variasjoner som beskriver variasjoner som ikke inneholder sesongvariasjoner. Det er mange tidsserier der man ikke klarer å se disse komponentene med en vanlig metode for å lage figur. Men ved en liten endring i metoden kan vi få de figurene som vi ønsker.

2.1.1 Figur 1.a. Ujusterte verdier

La x -aksen presentere tidsforløpet. Det går fra januar 1990 til og med desember 2009 for bilsalget. I alt er det 240 tidspunktene. y -aksen er verdiene for bilsalget for hvert av tidspunktene på x -aksen. Det varierer fra 3044 i desember 1991 til 15869 biler i juli 2007. Vi ser at

- Det er en klar trend. Bilsalget gikk litt langsomt ned og økte kraftig opp i mange år fram til 1997. Deretter gikk bilsalget langsomt ned fram til 2003 og økte det igjen fra 2004. Salget i slutten av 2008 og tidlig 2009 var kraftig redusert på grunn av finanskrisen for hele verden.
- Det finnes noen ekstremverdier. Bilsalget er ganske lav i desember 1995 med 3836 biler mot gjennomsnittet for desember i tidsserien er 9838 biler. Vi får også en lav verdi i januar 2009 med 6134 biler mot 9139 som er gjennomsnittet for januar i tidsserien. Det kan være flere uvanlig lave eller høye verdier i tidsserien. Det krever beregninger for å plukke dem ut. Tabell 1 viser bilsalget i 1995 og 1996. Verdien i desember 1995 er ganske lav i forhold til naboe.

Tabell 1: *Bilsalg. Ujusterte verdier i 1995 og 1996*

	jan.	feb.	mar.	apr.	mai	jun.
	jul.	aug.	sep.	okt.	nov.	des.
1995	8191	7186	9349	7441	9774	9187
	9691	8069	8079	9078	8006	3836
1996	11010	10920	13619	12145	13715	12557
	15168	10893	11383	14312	11828	9305

- Det er store variasjoner. Det kan være store sesongvariasjoner eller store tilfeldige variasjoner. Dette gjør at det blir vanskelig å se sesongmønsteret i tidsserien i figuren.

Metoden som vi bruker for å plote data i figur 1 er ikke god nok til å vise sesongvariasjoner i en tidsserie med store variasjoner.

2.1.2 Figur 1.b. Ujusterte verdier for hver av årets tolv måneder

Vi lurer på om at det er sesongvariasjoner i tidsserien eller ikke. For å svare på dette, anvender vi definisjonen av sesongvariasjoner: Det er variasjoner som gjentar seg fra år til år i samme tidspunkter eller i samme perioder. I figuren plotter vi data for hvert år oppå hverandre for å finne tidspunktene der lave eller høye verdier som gjentar seg fra år til år. Det er ialt 20 kurver for årene 1990-2009. Vi ser et sesongmønster i data. Bilsalget øker mest i juli og synker mest i desember hvert år. Et høyt tall i denne måned skal følge et lavere tall i neste måned. Regelen gjelder for nesten alle årene.

Fordelene ved denne metoden i figuren er at

- Vi kan oppdage bevegelige helligdager som påske, pinse eller Kristihimmelfartsdag i tidsserien, ved å betrakte mønstre av ujusterte tall i mars, april, mai og juni.
- Vi kan se endringer av ujusterte tall i samme måned fra år til år for å evaluere stabilitet av sesongvariasjoner.

Det later til at påske og pinse har påvirket bilsalget. Vi ser store variasjoner som er i april, mai, juli og desember. Vi sjekker dette ved tabell 2 som viser standardavvik for tolv måneder. Standardavvikene av juli og desember er største med 3577 og 3594 biler, henholdvis.

Tabell 2: *Bilsalg. Standardavvik for tolv måneder*

jan.	feb.	mar.	apr.	mai	jun.
jul.	aug.	sep.	okt.	nov.	des.
2795	2641	2993	3135	3209	2935
3577	2829	2562	2883	2701	3594

En ulempe ved denne metoden er at de lave kurvene blir glattere i forhold til de høye kurvene når vi plotter dem sammen. Dermed skal vi ikke se tydelig sesongmønstre av de lave kurvene.

2.1.3 Figur 1.c. Ujusterte verdier for hver av årets tolv måneder (etter standardisering av tall for hvert år med gjennomsnitt og standardavvik i året)

Vi retter opp ulempen i figur 1.b ved å standardisere de ujusterte verdiene for hvert år med deres gjennomsnitt og standardavvik i året. Vi lager en ny tidsserie som ser slik ut, for eksempel for 1990:

$$\frac{O_{jan.90} - mean(1990)}{std(1990)}, \frac{O_{feb.90} - mean(1990)}{std(1990)}, \dots, \frac{O_{des.90} - mean(1990)}{std(1990)},$$

Den nye tidsserien varierer rundt 0-aksen. Sesongmønsteret blir akkurat som før. En fordel etter standardisering er at når vi plotter data for hvert år oppå hverandre slik som vi gjorde i figur 1.b er det ingen kurv som vil bli glattere i forhold til de andre. Tabell 3 viser tall for 1990 og 1996 før og etter standardiseringen. De ujusterte verdiene i 1996 ligger på et annet nivå i forhold til 1990. Deres nivåer etter standardisering blir det samme. Vi ser i figuren en nedgang av bilsalget fra januar til februar, en økning fra februar til mars, osv.

Vi ser i figur 1.c et klarere sesongmønster for bilsalg enn i figur 1.b. Det viser også tydeligere variasjoner i ujusterte tall fra mars til april og fra mai til juni. Påsken påviker bilsalget mer enn pinsen. Det blir også lettere å se variasjoner som ikke er sesongvariasjoner i figuren enn i figur 1.b. Endringene fra år til år i samme måned er store noe som medfører ustabile sesongvariasjoner.

Tabell 3: *Bilsalg. Ujusterte verdiene for 1990 og 1996 før og etter standardisering*

måned	Før		Etter	
	1990	1996	1990	1996
januar	5481	11010	0.2276	-0.7240
februar	5054	10920	-0.3269	-0.7770
mars	5728	13619	0.5484	0.8143
april	5071	12145	-0.3049	-0.0547
mai	6251	13715	1.2277	0.8709
juni	6048	12557	0.9640	0.1881
juli	5191	15168	-0.1490	1.7276
august	5025	10893	-0.3646	-0.7929
september	5117	11383	-0.2451	-0.5040
oktober	6047	14312	0.9628	1.2229
november	5367	11828	0.0795	-0.2416
desember	3289	9305	-2.6195	-1.7292

2.1.4 Figur 1.d. Ujusterte verdier fra år til år for hver enkelt måned

I den opprinnelige tidsserien observerer vi bilsalget med følgende rekke (år \times måned) slik: $O_{jan,1990}, O_{feb,1990}, \dots, O_{des,2009}$. Vi gjør det omvendt i denne figuren. Vi sorterer observasjonene med følgende rekke (måned \times år), slik $(O_{jan,1990}, \dots, O_{jan,2009}), (O_{feb,1990}, \dots, O_{feb,2009}), \dots, (O_{des,1990}, \dots, O_{des,2009})$. Det blir tolv årlige tidsserier for tolv måneder. Vi plotter dem ved siden av hverandre slik som i figuren 1.d. Her ser vi endringer av bilsalget fra det første til det siste året for hver måned. Bilsalget er høyere og endrer seg fortere i de siste årene.

2.1.5 Figur 1.e. Boxplott av de ujusterte verdiene for tolv måneder

Vi beskriver fordelingen av de ujusterte verdiene ved boxplott (Box-Whisker diagram). Det gir en grafisk presentasjon av spredning i verdiene v.h.a 5 tall: minimumsverdi, 1.kvartil, median (2.kvartil), 3.kvartil og maksimumsverdi. I et sortert datasett er 1.kvartil, medianen og 3.kvartil de verdiene som har 25%, 50% og 75% av enhetene under seg, henholdvis. Differansen mellom 3. og 1.kvartil kalles vanligvis for kvartildifferansen. Diagrammet består av en rektangelformet boks og to linjer (whiskers) som strekker seg ut fra endene av denne boksen. Vi kan betrakte medianen som et tyngdepunkt i fordelingen. I figur 1.e plotter vi Box-Whisker diagrammer for tolv måneder. Boxplottet for januar viser at fordelingen blir forskjøvet i retning mot høye verdier. I den øvre halvparten av fordelingen ligger alle enhetene relativt nær tyngdepunktet. I den nedre halvdel er enhetene spredt over et langt større intervall. Boxplott er godt egnet til å sammenligne to eller flere datasett.

2.1.6 Figur 1.f. Autokorrelasjonsfunksjonen (ACF) av ujusterte verdier

Vi antar at det er en statistisk avhengighet mellom observasjonene O_t på forskjellige tidspunkter. Vi antar også at det er for hver t en sannsynlighetsfordeling for O_t og at O_t er en stokastisk prosess. Autokorrelasjonsfunksjonen (ACF) mellom O_t og O_{t+k} er definert slik

$$\rho_k = \text{Corr}(O_t, O_{t+k}) = \frac{\text{Cov}(O_t, O_{t+k})}{[\text{Var}(O_t)\text{Var}(O_{t+k})]^{0.5}}$$

Vi har følgende egenskaper: $\rho_0 = 1$, $\rho_k = \rho_{-k}$ og $|\rho_k| \leq 1$. Autokorrelasjonsfunksjonen er et mål for å måle hvor stor korrelasjon og retning (positiv eller negativ) mellom de to stokastiske variablene O_t og O_{t+k} .

Ved å plote de estimerte autokorrelasjonene $\hat{\rho}_k$ etter lag k , får vi et godt bilde av sesongvariasjonene. Vi ser i figuren at $\hat{\rho}_k$ er signifikant større enn null i lagene 12, 24, 36. Dette viser sesongvariasjoner i bilsalget. Vi bruker kommandoen `acf` i R for å beregne korrelasjoner.

2.1.7 Figur 1.g. Spekteret (spektraltetthet)

Autokorrelasjonsfunksjonen er et verktøy til å studere utviklingen av en prosess i tid. Vi kan også finne en funksjon for å beskrive en tidsseries egenskaper i frekvensrommet, nemlig spektraltetthetsfunksjonen eller spekteret. For enhver stasjonær stokastisk prosess med autokovariansfunksjon γ_k , eksisterer det en monotont økende funksjon $F(\omega)$ via Fourier-Stieltjes integral

$$\gamma_k = \int_{-\pi}^{\pi} e^{i\omega k} dF(\omega) \quad (1)$$

Da kalles funksjonen F for spektralfordelingsfunksjonen til tidsserien. Hvis det finnes en absolutt kontinuerlig ikke-negativ funksjon f slik at $F(\omega) = \int_{-\pi}^{\omega} f(v)dv$ kalles denne for spektraltettheten og 1 kan skrives

$$\gamma(k) = \int_{-\pi}^{\pi} e^{ik\omega} f(\omega)d\omega, \quad k = 0, \pm 1, \dots \quad (2)$$

Vi vil sjekke sesongvariasjoner for bilsalget. Vi bruker kommandoen `spectrum()` i R for å beregne de estimerte verdiene av spektraltettheten. Disse vises i tabell 4. Vi ser at spektraltettheten har en ganske høy verdi i frekvens null. Dette viser en klar trend i tidsserien. Vi ser også at spektraltettheten er tynn og høy rundt frekvensen 0.082164 ($\approx 1/12.17$). Det er et tegn på sesongvariasjoner i ujusterte verdier med en syklus på 12.17 måneder. Merk at en prosess med tilfeldig variasjon har et flatt spekter.

Tabell 4: *Spektraltettheten av bilsalg*

frekvens	periode	spekteret	forklaring
0	∞	664544888	trend
\vdots	\vdots	\vdots	
0.078156313	12.79	11375744	
0.079158317	12.63	11685041	
0.080160321	12.47	11893735	
0.081162325	12.32	11978969	sesongvariasjoner
0.082164329	12.17	11924637	
0.083166333	12.02	11724626	
0.084168337	11.88	11384502	

2.1.8 Oppsummering av figurene 1.a-g

Figurene 1.a-g viser oss om egenskaper ved de ujusterte verdiene. Vi kan se trend, mønsteret av sesongvariasjoner i tidsserien (ved å plote spekteret eller ACF), ekstremverdier (ved Boxplott) og endringer fra tidspunkt til tidspunkt. Dessuten kan effektene av bevegelige helligdager fremheves. Figuren viser også når vi bør sesongjustere tidsserien med *log* transformasjon for ujusterte verdier.

2.2 FIGURER FOR KORRIGERING AV KALENDEREFFEKTER OG EKSTREMVERDIER

Det er sju dager i en uke, fra mandag til søndag. Det er 4 mandager, tirsdager, . . . , søndager i en vanlig februar og en ekstra dag i februar i et skuddår. Det er 31 dager i januar, mars, mai, juli, august, oktober og desember, og 30 dager i de andre månedene. Selv om antall dager i januar 2009, 2010 og mars 2010 er 31 dager er antall ukedager ikke det samme. I en uke er det dager der salget er høyere enn andre dager. Dermed er salget avhengig av bestemte dager i uken og av antall slike dager i måneden.

Påske i Norge er Skjærtorsdag, Langfredag, påskeaften, første påskedag (påskesøndag) og andre påskedag (mandag) for alle årene. Men datoene av disse dagene som faller på varierer fra år til år. Den tidligste datoen for 1.påskedag er 22.mars og den seineste er 25.april. Salget i mars og april varierer fra år til år. Hvis påske hadde falt på en fast dato hvert år skulle vi behandlet effekten av påske som sesongeffekter. Vi har det samme problemet for pinse og Kristihimmelfartsdag. De er bevegelige helligdager som varierer i mai og juni.

Finanskrisen i hele verden i 2008 og 2009 har trukket bilsalget kraftig ned i mange måneder. Tallene ble uvanlige lave. De er å betrakte som ekstremverdier.

Effektene som skyldes ulikt antall ukedager i måneder, bevegelige helligdager og ekstremverdier skal korrigeres fra de ujusterte verdiene for sesongjustering. I de neste avsnittene vil vi vise figurer for denne korrigeringen.

2.2.1 Figur 2.a. Ujusterte og korrigererte verdier for effektene av ukedager og ekstremverdier

X-12-ARIMA utfører en regresjonsanalyse ved en RegARIMA modell for å estimere effektene av ukedager, bevegelige helligdager og ekstremverdier. Vi har brukt metoden i X-12-ARIMA for å korrigere effekten av påske i norske data. Dette har vist at metoden fungerer dårlig, siden påske i USA er kortere enn i Norge, bare 1.påskedag. I notat 97/73 har Solheim og Dinh utviklet en egen metode for å fjerne påskeeffekt i detaljhandelsvolumindeksen. De ujusterte verdiene etter å ha blitt korrigert med vår metode leses inn i X-12-ARIMA for sesongjustering uten å korrigere for påskeeffekt med den innebygde rutinen i programmet. Vi får bedre sesongjusterte tall for mars og april.

ABS (Australian Bureau of Statistics) har fått et lignende problem som i Norge. Grunnen er at påsken i Australia er like lang som i Norge. Xichuan Zhang, Craig H. McLaren og Caler C.S. Leung (2003) har utviklet en metode for påskeskorrigerende av tidsserier i Australia. US Census har laget et lite program som heter Genhol for påskeskorrigerende for landene som Australia og Norge. Man bruker programmet for å lage påskeregressorer. Genhol kan også brukes for å beregne effektene av de andre bevegelige helligdagene som pinse og Kristihimmelfartsdag.

Vi bruker metoden fra US Census og Genhol for å lage et nytt program for korrigerende av bevegelige helligdager. Beskrivelsene av programmet finnes i notat 2007/43.

Estimeringen av effektene av ukedager og ekstremverdier er gitt i tabell 5. Det er signifikante effekter på 5% nivå for ukedager, påske og pinse. Det var mange ekstremverdier i tidsserien. I modellen for estimeringen antar vi at effektene ikke varierer over tid. t -verdiene viser at bilsalget i en måned øker med antall onsdager og fredager og synker med økt antall lørdager. Bilsalget blir redusert i pinsen ($t=-2.16$), men mer kraftig redusert i påske ($t=-7.46$). Dette ser rimelig ut siden det er flere fridager i påske enn pinse.

Etter å ha utført regresjonsanalysen for korrigeringen lister X-12-ARIMA tabeller for korrigererte faktorer for ukedager, bevegelige helligdager og ekstremverdier og en tabell som er en kombinasjon av

disse effektene. Tabell 6 viser korrigerede faktorer for kalender og ekstremverdi i januar 1998. De er 1.0057 og 0.7558, henholdsvis. Observerte verdien er $O_{jan.98}=9776$ biler. Den korrigerede verdien blir $O_t^K = 9776/(1.0057 \times 0.7558) = 12860$.

De ujusterte verdiene og de korrigerede verdiene for bilsalget er vist i figur 2.a. Siden de to kurvene er nesten identiske, blir det bedre å plote dem med farger for å skille dem fra hverandre.

Tabell 5: *Estimeringen for kalendereffekter og ekstremverdier*

Variable	Estimerte parametre	Standard error	t-verdi
mandag	0.0044	0.0089	0.49
tirsdag	0.0079	0.0087	0.90
onsdag	0.0184	0.0089	2.05
torsdag	-0.0063	0.0090	-0.70
fredag	0.0290	0.0088	3.27
lordag	-0.0169	0.0086	-1.97
paske	-0.1412	0.0189	-7.46
pinse	-0.0358	0.0166	-2.16
januar 1998(AO)	-0.2799	0.0582	-4.81
januar 2007(AO)	0.2724	0.0643	4.24
januar 2009(AO)	-0.2526	0.0653	-3.87
desember 1995(AO)	-0.6125	0.0584	-10.47
desember 1998(AO)	-0.1729	0.0584	-2.96
desember 2004(AO)	0.4139	0.0604	6.85
desember 2005(AO)	0.2682	0.0616	4.35
desember 2006(AO)	0.3918	0.0652	6.01

Tabell 6: *Korrigerede faktorer for kalendereffekter og ekstremverdier*

	Januar			Desember				
	1998	2007	2009	1995	1998	2004	2005	2006
kalender	1.0057	1.0311	1.0057	0.9760	1.0202	1.0419	1.0057	0.9760
ekstremverdi	0.7558	1.3132	0.7768	0.5420	0.8412	1.5127	1.3077	1.4797
ujusterte	9776	15010	6434	3836	6974	14886	12135	14417
korrigerede verdi	12860	11085	8236	7251	8126	9444	9227	9983

2.2.2 Figur 2.b. Ujusterte og korrigerede verdier for bevegelige helligdager

Vi vil vise mønstrene av de ujusterte og korrigerede verdiene for påske og pinse. I estimering av effektene av påske og pinse med ny metode trenger vi på forhånd verdiene av parametrene w_1 , w_2 , w_3 og w_4 . De beskriver påskeeffekten i perioden med w_1 dager som ligger før påske og perioden med w_2 dager som ligger etter påske. Tilsvarende for pinseeffekten w_3 og w_4 dager. Vi har ikke laget en metode for å estimere disse dagene. Vi antar at w_i kan få verdiene fra 0 til $N(w_i)$ dager, der $N(w_1) = 7$, $N(w_2) = N(w_3) = 3$ og $N(w_4) = 1$. Estimeringen er utført for 256 kombinasjoner av w_i . Vi plukker den kombinasjonen som har minst *AICC*-verdi. Vi bør snakke med fagfolk om verdiene vi har funnet. Det kan finnes \hat{w}_i som er mer rimeligere.

Vi velger $w_1 = w_2 = w_3 = w_4 = 0$ for bilsalget etter diskusjoner med fagfolk. Dvs at bilsalget ikke blir økt eller redusert før og etter påske. Dette gjelder også for pinse. Vi ser at vi får et bedre sesongmønster for mars, april, mai og juni etter korrigeringen. Det ser ut som en nedgang fra mars til april, en økning fra april til mai og en nedgang fra mai til juni. Det er fortsatt store tilfeldige variasjoner i disse månedene.

2.2.3 Figur 2.c. Ekstremverdier (intervensjoner)

Box and Tiao (1965) har utviklet intervensjonsanalyse for å studere tidsserier som er påvirket av ekstreme begivenheter. Vi tar et eksempel for intervensjoner. Da Kuwait ble invadert av Irak i begynnelse av august 1990, økte prisen på råolje kraftig. I slutten av september ble prisen for Nordsjøolje til over 40 US\$ per fat. Prisen gikk litt opp og ned rundt 40 dollar til midten av oktober. Deretter falt prisen til 30 dollar per fat. Konsekvensen var at verdien for norsk eksport i september og oktober 1990 og flere måneder senere ble svært høy. Vi estimerer effektene av ekstremverdier ved en RegARIMA modell i X-12-ARIMA og ved dummy variablene nedenfor.

1. Additive Outlier (AO) i t_0 . (Et spesielt utslag for et bestemt tidspunkt t_0). Verdier har en brå endring i t_0 . Effekten påvirker tallene bare i t_0 og forsvinner etter t_0 . Vi estimerer effekten ved dummy variabel

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

2. Level Shift (LS) i t_0 . (Nivåskift ved tidspunkt t_0). Verdiene før og etter t_0 ligger i to forskjellige nivåer.

$$LS_t^{(t_0)} = \begin{cases} -1 & \text{for } t < t_0, \\ 0 & \text{for } t \geq t_0 \end{cases}$$

3. Temporary Change i t_0 . Verdien i t_0 ligger på et annet nivå enn de tidlige. Nivået synker etter hvert og forsvinner etter noen tidspunkter.

$$TC_t^{(T_0)} = \begin{cases} 0 & \text{for } t < t_0, \\ \alpha^{t-t_0} & \text{for } t \geq t_0, \text{ and } 0 < \alpha < 1. \end{cases}$$

hvor effekten α ligger mellom 0 og 1.

4. Ramp, t_0 to t_1

$$RP_t^{(t_0, t_1)} = \begin{cases} -1 & \text{for } t \leq t_0, \\ \frac{t-t_0}{t_1-t_0} - 1 & \text{for } t_0 < t < t_1, \\ 0 & \text{for } t \geq t_1 \end{cases}$$

Kritisk verdi for intervensjoner er gitt i tabell 6-14 X-12-ARIMA manualen, versjon 2.10.

Vi finner mange ekstremverdier for bilsalget (se tabell 5). De var i januar 1998, 2007 og 2009, og desember 1995, 1998, 2004, 2005 og 2006. De er "Additive Outlier". En negativ t -verdi vil vise at tall skal bli justert opp og omvendt. Disse ekstremverdiene vises ved høye tyne spisser i figuren. De korrigerede faktorene er 0.7542 (januar 1998), 1.3135 (januar 2007), 0.7808 (januar 2009), 0.5428 (desember 1995), 0.8381 (desember 1998), 1.5006 (desember 2004), 1.2918 (desember 2005) og 1.4568 (desember 2006). Dermed blir den ujusterte verdien i desember 1995 korrigert opp mer enn den ujusterte verdien i januar 1998. Vi finner ikke alltid forklaringer for ekstremverdiene i tidsserien. Lav verdi i desember 1995 har trolig sammenheng med lavere bilavgifter fra januar 1996. Finanskrisen var grunnen at bilsalget gikk kraftig ned i januar 2009.

2.2.4 Figur 2.d. Avvikene mellom ujusterte og korrigerte verdier

Vi fjerner effektene av kalender og ekstremverdier fra de ujusterte verdiene for å få korrigerte verdier. Avvikene mellom de to tidsseriene for bilsalget er plottet i figuren. De korrigerte faktorene er hentet fra tabell A.6 og A.8 i X-12-ARIMA og listet ut i tabell 7. Vi ser at de korrigerte faktorene for kalenderen varierer rundt 1, men for ekstremverdier er det mye lavere eller høyere enn 1 (se kolonne 2 og 3 i tabellen). De ujusterte verdiene blir dermed korrigert mer for ekstremverdier enn kalenderen. Avvikene blir mye større i tidspunktene med ekstremverdier.

Tabell 7: *Ujusterte og korrigerte verdiene i tidspunktene for ekstremverdier*

	Korrigerte faktorer		Ujusterte	Korrigerte verdier		
	Kalender	Ekstrem		kalender	ekstrem	begge
des.1995	0.9759	0.5428	3836	7067	7241	7242
jan.1998	1.0063	0.7542	9776	12962	12880	12881
des.1998	1.0223	0.8381	6974	8321	8139	8139
des.2004	1.0417	1.5006	14886	9920	9522	9523
des.2005	1.0063	1.2918	12135	9393	9335	9334
des.2006	0.9759	1.4568	14417	9896	10140	10141
jan.2007	1.0301	1.3135	15010	11427	11093	11094
jan.2009	1.0063	0.7808	6434	8240	8188	8187

2.2.5 Figur 2.e. Framskrivinger for 12 måneder

Når vi estimerer komponenter i X-11 brukes symmetrisk filter i midten og asymmetrisk filter i starten og slutten av tidsserien. Dette medfører store revisjoner i de siste tidspunktene når en ny observasjon tilføyes. I X-11-ARIMA (1980, Dagum) og seinere X-12-ARIMA innførte man bruk av ARIMA modell for å forlenge tidsserien i begge ender for at et symmetrisk filter kan brukes for alle observasjoner. Vi får mindre revisjoner for de siste tidspunktene i sammenligning med X-11. I X-12-ARIMA innførte man regARIMA modell (linear regression model with ARIMA time series errors) for å estimere effektene av kalender og ekstremverdier med input av ujusterte verdier. Deretter blir modellen brukt for å beregne framskrivninger. ARIMA(0 1 1)(0 1 1) blir brukt for bilsalget. Framskrivningene fra januar til desember 2010 er listet ut i tabell 8. De er plottet i figuren. Vi ser at bilsalget ikke blir lammet av finanskrisen lenger. Siden framskrivninger er beregnet etter å ha fjernet effekten av kalender, blir tallene ikke påvirket av ulike antall ukedager i måned, påske og pinse.

2.2.6 Figur 2.f. Korrigerte verdier for hver av årets tolv måneder (etter standardisering tall for hvert år med gjennomsnittet og standardavviket i året)

Vi laget figur 1.c for at det vil bli lettere å se sesongmønster i ujusterte verdier. Vi vil gjøre det på samme måte for de korrigerte verdiene. Vi ser i figuren at sesongvariasjoner i mars, april, januar og desember blir mer stabile i sammenligning med figur 1.c som følge av korrigerings for påskeeffekt og ekstremverdier. Standardavviket i april blir kraftig redusert fra 0.8422 til 0.4311 og i desember fra 1.2899 til 0.4249 etter korrigerings. Standardavvikene i de andre månedene blir også litt redusert bortsett fra juni og juli. Korrigerings for påskeeffekt og ekstremverdier er helt nødvendig for sesongjustering av bilsalget.

Tabell 8: *Framskrivinger i 2010 og 95% konfidensintervaller*

måned	Estimerte	<i>l</i> 95	<i>u</i> 95
januar	10171	12014	14190
februar	9937	12049	14610
mars	12216	15286	19128
april	10822	13908	17874
mai	10641	14012	18453
juni	11216	15095	20316
juli	12057	16568	22765
august	9829	13768	19285
september	10377	14809	21132
oktober	10275	14923	21673
november	9687	14308	21135
desember	9241	13862	20794

Tabell 9: *Standardavvikene av ujusterte og korrigerte verdier etter standardisering*

	jan.	feb.	mar.	apr.	mai	jun.
Ujusterte	0.9049	0.5480	0.6123	0.8422	0.5594	0.3652
Korrigerte verdi	0.7838	0.4706	0.6077	0.4311	0.4249	0.4309
	jul.	aug.	sep.	okt.	nov.	des.
Ujusterte	0.5977	0.5474	0.4576	0.7278	0.6166	1.2899
Korrigerte verdi	0.6795	0.4734	0.4335	0.7059	0.6388	0.4249

2.2.7 Figur 2.g. Spekteret for korrigerte verdier

Vi plotter spektrene av ujusterte og korrigerte verdier i samme figur. Vi får litt høyere og smalere spisser i lag 12, 4 og 3 etter korrigeringen. Dette viser at sesongvariasjonene blir mer stabile etter korrigeringen.

2.2.8 Figur 2.h. Boxplott for tolv måneder for ujusterte og korrigerte verdier

Boxplottet i desember for de korrigerte verdiene er kortere i forhold til de ujusterte verdiene. Dette viser mindre variasjon for korrigerte verdier i desember. Vi får den samme beskrivelsen for januar. Ellers er det bare små forskjeller mellom boxplottene for de andre månedene.

2.2.9 Oppsummering av figur 2

Vi plotter de korrigerte verdiene sammen med de ujusterte verdiene for å vise endringer av sesongvariasjoner og sesongmønster etter korrigeringen. Vi får mindre variasjoner og klarere sesongmønster for bilsalg etter å ha fjernet effektene av kalender og ekstremverdier.

2.3 FIGURER FOR UJUSTERTE, SESONGJUSTERTE VERDIER OG TREND

De korrigerte verdiene i avsnitt 2.2 blir brukt for å beregne komponentene: trend, sesongjusterte verdier og sesongkomponenten. Brukere kan finne en ganske detaljert beskrivelse av Ladiray og Quenneville (1999).

For å beskrive beregningen enklere antar vi at

- Den originale serien O_t er ikke påvirket av effektene av ukedager, påske og intervensjoner.
- Det er ikke nødvendig å forlenge serien til begge sider for at endepunktene kan beregnes med symmetriske vekter.

De tre modellene for sesongjustering er

$$\begin{array}{ll} \text{Multiplikativ modell (M)} & O_t = T_t \times S_t \times I_t \\ \text{Additiv modell (A)} & O_t = T_t + S_t + I_t \\ \text{Pseudo-Additiv modell (PA)} & O_t = T_t \times (S_t + I_t - 1) \end{array} \quad (3)$$

- Multiplikativ modell brukes når sesongen er en fast andel av nivået.
- Additiv modell brukes når sesongen er en fast verdi.
- Pseudo-Additiv er en ny modell i X12-ARIMA. Den brukes for ikke negative tidsserier hvor observasjonene har ganske lave verdier i samme måneder hvert år.

Komponentene T_t , A_t , S_t og I_t estimeres i tre trinn.

Trinn 1. Initiale estimeringer

(i) Estimerer trend ved det sentrerte 12 måneder glidende gjennomsnittet

$$T_t^{(1)} = \frac{1}{24}O_{t-6} + \frac{1}{12}O_{t-5} + \dots + \frac{1}{12}O_t + \dots + \frac{1}{12}O_{t+5} + \frac{1}{24}O_{t+6}$$

$T_t^{(1)}$ er trenden i den første runden.

(ii) SI_t ratene

$$\begin{array}{ll} \text{(M, PA)} & SI_t^{(1)} = O_t/T_t^{(1)} \\ \text{(A)} & SI_t^{(1)} = O_t - T_t^{(1)} \end{array} \quad (4)$$

(iii) Estimerer innledende skritt for sesongfaktoren $\widehat{S}_t^{(1)}$ ved 3×3 filteret

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

(iv) Sesongfaktor

$$\begin{array}{ll} \text{(M, PA)} & S_t^{(1)} = \widehat{S}_t^{(1)} / (\frac{1}{24}\widehat{S}_{t-6}^{(1)} + \frac{1}{12}\widehat{S}_{t-5}^{(1)} + \dots + \frac{1}{12}\widehat{S}_{t+5}^{(1)} + \frac{1}{24}\widehat{S}_{t+6}^{(1)}) \\ \text{(A)} & S_t^{(1)} = \widehat{S}_t^{(1)} - (\frac{1}{24}\widehat{S}_{t-6}^{(1)} + \frac{1}{12}\widehat{S}_{t-5}^{(1)} + \dots + \frac{1}{12}\widehat{S}_{t+5}^{(1)} + \frac{1}{24}\widehat{S}_{t+6}^{(1)}) \end{array}$$

(v) Sesongjusterte

$$\begin{aligned} \text{(M)} \quad A_t^{(1)} &= O_t/S_t^{(1)} \\ \text{(A)} \quad A_t^{(1)} &= O_t - S_t^{(1)} \\ \text{(PA)} \quad A_t^{(1)} &= O_t - T_t^{(1)} \times (S_t^{(1)} - 1) \end{aligned}$$

Trinn 2 Estimerer sesongfaktor og sesongjusterte tall

(i) Estimerer trend i andre trinn $T_t^{(2)}$ ved

$$T_t^{(2)} = \sum_{j=-H}^H h_j^{(2H+1)} A_{t+j}^{(1)}$$

hvor $h_j^{(2H+1)}$, $-H \leq j \leq H$, $h_j = h_{-j}$ er Hendersons vektene.

(ii) SI -ratene

$$\begin{aligned} \text{(M, PA)} \quad SI_t^{(2)} &= O_t/T_t^{(2)} \\ \text{(A)} \quad SI_t^{(2)} &= O_t - T_t^{(2)} \end{aligned} \tag{5}$$

(iii) Estimerer innledende skritt for sesongfaktoren $\widehat{S}_t^{(1)}$ ved 3×5 filteret

$$\begin{aligned} \widehat{S}_t^{(2)} &= \frac{1}{15} SI_{t-36}^{(2)} + \frac{2}{15} SI_{t-24}^{(2)} + \frac{3}{15} SI_{t-12}^{(2)} + \frac{3}{15} SI_t^{(2)} + \\ &\quad \frac{3}{15} SI_{t+12}^{(2)} + \frac{2}{15} SI_{t+24}^{(2)} + \frac{1}{15} SI_{t+36}^{(2)} \end{aligned}$$

(iv) Sesongfaktor

$$\begin{aligned} \text{(M, PA)} \quad S_t^{(2)} &= \widehat{S}_t^{(2)} / \left(\frac{1}{24} \widehat{S}_{t-6}^{(2)} + \frac{1}{12} \widehat{S}_{t-5}^{(2)} + \cdots + \frac{1}{12} \widehat{S}_{t+5}^{(2)} + \frac{1}{24} \widehat{S}_{t+6}^{(2)} \right) \\ \text{(A)} \quad S_t^{(2)} &= \widehat{S}_t^{(2)} - \left(\frac{1}{24} \widehat{S}_{t-6}^{(2)} + \frac{1}{12} \widehat{S}_{t-5}^{(2)} + \cdots + \frac{1}{12} \widehat{S}_{t+5}^{(2)} + \frac{1}{24} \widehat{S}_{t+6}^{(2)} \right) \end{aligned}$$

(v) Sesongjusterte

$$\begin{aligned} \text{(M)} \quad A_t^{(2)} &= O_t/S_t^{(2)} \\ \text{(A)} \quad A_t^{(2)} &= O_t - S_t^{(2)} \\ \text{(PA)} \quad A_t^{(2)} &= O_t - T_t^{(2)} \times (S_t^{(2)} - 1) \end{aligned}$$

Trinn 3. Estimerer endelig Henderson trend og endelig irregulær komponent

(i) Endelig trend

$$T_t^{(3)} = \sum_{j=-H}^H h_j^{(2H+1)} A_{t+j}^{(2)}$$

(ii) Endelig irregulær komponent

$$\begin{aligned} I_t^{(3)} &= A_t^{(2)} / T_t^{(3)} && \text{for M, PA} \\ I_t^{(3)} &= A_t^{(2)} - T_t^{(3)} && \text{for A} \end{aligned} \quad (6)$$

Vi har

Sesongdekomponering

$$\begin{aligned} \text{(M)} \quad O_t &= T_t^{(3)} \times S_t^{(2)} \times I_t^{(3)} \\ \text{(A)} \quad O_t &= T_t^{(3)} + S_t^{(2)} + I_t^{(3)} \\ \text{(PA)} \quad O_t &= T_t^{(2)} \times (S_t^{(2)} - 1) + T_t^{(3)} \times I_t^{(3)} \end{aligned} \quad (7)$$

2.3.1 Figur 3.a. Ujusterte verdiene og trend

Vi lar X-12-ARIMA å selv få velge lengden av filter for estimering av trend. Programmet har valgt det 13 måneders Hendersons glidende gjennomsnittet. Vi kan angi et tilpassende filter på forhånd for estimeringen. I sammenligning med de ujusterte verdiene er trend glattere. Det viser en nedgang av bilsalget i mange måneder på grunn av finanskrisen. Bilsalget i slutten av 2009 øker igjen til et nivå som er litt lavere enn nivået før finanskrisen.

2.3.2 Figur 3.b. Sesongjusterte verdier og trend

En tidsserie kan sesongjusteres ved en additiv eller multiplikativ modell. For å beskrive forklaringer enklere vil vi anta at det ikke er noen korrigering for ujusterte verdier. Vi skriver $O_t = T_t^a + S_t^a + I_t^a$ for additiv modell og $O_t = T_t^m \times S_t^m \times I_t^m$ for multiplikativ modell, der a og m er symboler for modell. Sesongjusterte verdi for begge modeller blir $A_t^a = T_t^a + I_t^a$ og $A_t^m = T_t^m \times I_t^m$, henholdsvis. For en additiv modell har alle komponentene samme enhet. Men for en multiplikativ modell blir sesong- og irregulærkomponenten beregnet i prosent. Vi tolker irregulærkomponenten som tilfeldige variasjon i tidsserien.

I figuren har vi plottet trend og sesongjusterte verdier med søyler som viser avstander mellom de to komponentene. Det er mange høye søyler som viser ekstremverdier i tidsserien (se tabell 5).

2.3.3 Figur 3.c. Sesongjusterte verdier for hver av årets 12 måneder

Vi lager denne figuren for sesongjusterte verdier på samme måte som vi gjorde på figur 1.c. Formålet er å sjekke sesongvariasjoner i tidsserien. Figuren viser et uklart sesongmønster. Sesongvariasjoner i figur 1.c er fjernet. Variasjoner i januar og desember er større enn de andre månedene.

2.3.4 Figur 3.d Sesongjusterte tall fra år til år for hver enkel måned

Vi lager figuren på samme måte som vi har gjort i figur 1.d, men nå for sesongjusterte verdier. Sammenligning mellom de to figurene viser at gjennomsnittene av ujusterte verdier i tolv måneder danner et sesongmønster av tidsserien. Mønsteret blir fjernet i sesongjusterte verdier. Vi får like store gjennomsnitt for tolv måneder med sesongjusterte verdier (bortsett fra desember som er litt høyere). Dette viser at det ikke lenger finnes sesongvariasjoner i de sesongjusterte verdiene. Vi får store variasjoner av sesongjusterte verdier i januar og desember.

2.3.5 Figur 3.e og figur 3.f. Endring i prosent fra måned til måned i ujusterte og sesongjusterte verdier

Endringen i prosent av ujustert verdi fra tidspunkt $(t - 1)$ til t er gitt ved $100\%(O_t - O_{t-1})/O_{t-1}$. Store endringer i de ujusterte verdiene blir kraftig redusert i de sesongjusterte verdiene. De to figurene viser store variasjoner i ujusterte verdier.

2.3.6 Figur 3.g og 3.h. Revisjoner av sesongjusterte verdier og trend

Man beregner revisjoner av sesongjusterte tall og trend ved

$$R_t^A = \frac{A_{t|t+lag} - A_{t|t}}{A_{t|t}} \quad (\text{for concurrent target, terminologi fra X-12-ARIMA}) \quad (8)$$

$$R_t^A = \frac{A_{t|T} - A_{t|t+lag}}{A_{t|t+lag}} \quad (\text{for final target}) \quad (9)$$

(se tabell 6.13 i X12-ARIMA manualen). Vi forklarer de to formlene for $lag = 1$. Forklaringer er også gyldig for $lag = k$, $k > 1$.

- Ligning (8): I ligningen er $A_{t|t}$ sesongjustert verdi i tidspunkt t for tidsserien (O_1, \dots, O_t) (symbolet $|t$ betyr gitt data til og med tidspunkt t). Da er $A_{t|t+1}$ sesongjusterte tall i tidspunkt t når en ny observasjon O_{t+1} legges til i tidsserien. Da blir $(O_1, \dots, O_t, O_{t+1})$. Selv om de er sesongjustert verdi i samme tidspunkt t , kan de to verdiene være forskjellige. Grunnen er at vi har brukt \hat{O}_{t+1} , en framskriving av O_{t+1} i beregning når vi observerer tidsserien til og med O_t , og O_{t+1} , når denne verdien er observert. Det er et avvik mellom framskrivingen \hat{O}_{t+1} og observert verdi O_{t+1} . Dette skaper en revisjon for sesongjusterte verdi når en ny observasjon tilføyes. Revisjonen beregnes ved ligningen (8).

- Ligning (9): Tidsserien er $O_1, \dots, O_t, \dots, O_T$. der $1 \leq t \leq T$. Da er $A_{t|T}$ sesongjustert verdi i tidspunkt t når vi bruker data til og med T . $A_{t|t+1}$ er definert ovenfor. Antar at $lag = 0$. Da blir det lettere å tolke ligningen. Det er revisjoner av sesongjusterte verdier når nye observasjoner legges til i tidsserien. Ligningen (8) er et tilfelle av (9).

Vi bruker samme formler og tolker på samme måte for trend slik vi har gjort dette for sesongjusterte verdier for trend.

Figurene viser revisjoner av sesongjusterte verdier og trend ved ligning (8) og $lag = 1$. Ofte er det revisjoner i de siste tidspunktene vi er interessert i. Vi får større revisjoner for trend enn sesongjusterte verdier i slutten av tidsserien. Trenden er mer følsom enn sesongjusterte verdier når nye observasjoner legges til i tidsserien.

2.3.7 Figur 3.i. Avvikene mellom ujusterte og sesongjusterte verdier for tolv måneder

Avvikene er definert slik at det er ujusterte verdier som trekkes fra sesongjusterte verdier. De kan få negative eller positive verdier som viser om ujusterte er lavere enn sesongjusterte verdier eller omvendt. I figuren for januar ser vi at avvikene endrer seg litt over tid bortsett fra 2005. Det kan være en ekstremverdi i januar 2005. Vi sjekker dette ved å liste ut noen observerte verdier av januar. Vi får $O_{jan.03}=9390$, $O_{jan.04}=10200$, $O_{jan.05}=9204$, $O_{jan.06}=9508$, $O_{jan.07}=15010$. Vi ser en lav verdi i januar 2005, men ikke veldig lav.

Avvikene i mars og april forandrer seg fra negative til positive og omvendt fortere enn de andre månedene. De går motsatt vei. Når det går opp i mars, blir det ned i april og omvendt. Det er store avvik i 1997 (1.påskedag 30.mars), 2002 (31.mars), 2005 (27.mars), 2008 (23.mars). I disse årene

faller påskedagene i mars eller mellom mars-april. Vi ser at påske har påvirket bilsalget. Testen i tabell 5 viser også dette.

Avvikene mellom ujusterte og sesongjusterte verdier i juli og desember er større enn de andre månedene. Etter sesongjustering blir sesongjusterte verdier lavere enn ujusterte verdier i juli og omvendt i desember.

2.4 SESONGKOMPONENTEN

Vi bruker ordet "sesongfaktor" for å vise størrelsen av sesongvariasjoner i et tidspunkt. Vi kaller tidsserien av sesongfaktorer for sesongkomponenten. Figurene nedenfor viser sesongvariasjoner av bilsalget.

2.4.1 Figur 4.a. Sesongkomponenten

Vi ser at det er endringer i mønsteret og i størrelsen av sesongkomponenten. Det er vanskelig å se hvilke måneder som endrer seg mer eller mindre over tid.

2.4.2 Figur 4.b. Sesongkomponenten for hver av årets 12 måneder

Vi plottet sesongkomponenten for hvert år oppå hverandre. Vi får et bilde av sesongvariasjonene i tidsserien. Det er en oppgang fra januar til juli og deretter er en nedgang. Det er store sesongvariasjoner i januar, juli og desember. Sesongvariasjoner i februar, mars, april, mai og juni er veldig stabile. De blir ustabile i de neste månedene. Høye sesongvariasjoner i juli og lave i desember vil medføre høyt bilsalg i juli og lav i desember.

2.4.3 Figur 4.c. Sesongkomponenten fra år til år for hver enkelt måned

Denne figuren viser mønstrene av sesongvariasjoner fra år til år for tolv måneder. Dette ser vi ikke i figur 4.b. For eksempel blir sesongvariasjonene redusert i januar, mens de øker i desember. For juli øker de fram til 2000 og synker deretter.

2.4.4 Figur 4.d. Sammenligning sesongmønstrene i figur 1.c og 4.b

I figur 1.b og 1.c plottet vi de ujusterte verdiene for hvert år oppå hverandre for å vise sesongmønster i tidsserien før analysen. Dette er en empirisk metode. Etter sesongjustering får vi de estimerte verdiene av sesongvariasjonene som er plottet i figur 4.b.

Vi lurer på om kurvene i figur 1.c og 4.b skal vise det samme sesongmønsteret. Siden det er mange kurver for hver figur blir det litt vanskelig å sammenligne dem mot hverandre. Vi gjør det enklere ved å sammenligne bare gjennomsnittene av kurvene i to figurer. Merk at de to gjennomsnittene ikke har den samme enheten. Det er antall biler i figur 1.c og prosent i figur 4.b. Vi fikser dette ved å standardisere dem. Enheten i begge figurer blir fjernet. Nå kan vi plote dem sammen i samme figur. Det vises i figur 4.b.

Vi ser at de to kurvene faller sammen. Dette viser at den empiriske metoden i figur 1.c uten beregning gir et godt bilde av sesongvariasjoner.

2.4.5 Oppsummering av figurene 4.a-d

Vi får et detaljert bilde av egenskapene ved sesongvariasjonene. Figurene viser størrelsene av sesongvariasjoner (figur 4.a), mønstre i endringer fra år til år for tolv måneder (figur 4.b og 4.c).

2.5 IRREGULÆRKOMPONENTEN (I_t)

Vi plotter irregulærkomponenten for å se restene etter å ha spaltet opp de ujusterte verdiene i trend og sesongkomponent. Figuren kan gi mye informasjon om kvaliteten av sesongjusteringen.

2.5.1 Figur 5.a. Irregulærkomponenten

Denne komponenten viser tilfeldige variasjoner i tidsserien. Siden vi sesongjusterer tidsserien med en multiplikativ modell varierer irregulærkomponenten rundt 100. Vi ser i figuren mange uvanlig lave og høye verdier. De viser tidspunkter for ekstremverdier av type AO.

2.5.2 Figur 5.b. Irregulærkomponenten fra år til år for hver enkel måned

Her plotter vi endringer av irregulærkomponenten fra år til år for hver enkel måned. Vi ser at endringene er størst i januar og desember.

2.5.3 Autokorrelasjonsfunksjonen (ACF) av irregulærkomponenten

Vi vil sjekke om at det er trend eller sesongvariasjoner igjen etter å ha fjernet dem fra tidsserien. Når vi ser tegn på at trenden eller sesongvariasjonen er i den irregulærkomponenten blir kvalitet av sesongjusteringen redusert, siden disse tre komponentene må skilles helt ut fra hverandre. Vi kan sjekke dette ved å plote ACF og vurdere verdiene i lagene 1, 2, 3 (for trend) og 12, 24, 36 (for sesong). Figuren viser at hverken trend eller sesongvariasjon er i den irregulærkomponenten. Vi får en god sesongjustering for bilsalg.

2.5.4 Figur 5.d. Spekteret av sesong- og irregulærkomponenten

Dette er en måte å sjekke trend eller sesongvariasjoner som er igjen etter sesongjusteringen ved spektralanalysen. Vi skal få samme konklusjoner som når vi plotter ACF i figur 5.c. Spekteret av irregulærkomponenten av bilsalg er en linje. Det er mønsteret til en hvit støy prosess.

2.5.5 Figur 5.e. Scatter plott av irregulærkomponenten

Vi vil sjekke korrelasjonen i irregulærkomponenten. Når det finnes en korrelasjon mellom dem kan vi beskrive irregulærkomponenten ved en ARIMA modell. Dermed blir det ikke en hvit støy prosess lenger. Alle tester som er basert den irregulærkomponenten er dermed ikke gyldige. Figuren viser ingen korrelasjon i lag 1,2 og 12 i den irregulærkomponenten.

2.6 *SI* KOMPONENTEN

SI komponenten blir beregnet ved ligningene (4) og (5). Det er de komponenten som er igjen etter å ha fjernet trend fra data. Vi lager plott for *SI* komponenten i trinn 2 ved ligningen (5) sammen med sesongkomponenten S_t for tolv måneder. De to komponentene ligger nær hverandre og det er en god tilpasning mellom dem. Dette viser også at de tilfeldige variasjonene er lave slik at de ikke påvirker prosessen med å fjerne sesongvariasjoner i tidsserien. Figurene viser samtidig variasjoner av sesong og irregulærkomponenten.

2.7 DIAGNOSTIKK

Etter å ha utført en sesongjustering lister X-12-ARIMA opp diagnostikker for å vurdere hvor god sesongjusteringen er. Det er to metoder som blir presentert i X-12-ARIMA. Det er kvalitetsmål og sliding spans teknikken. De to metodene kan gi motsatte konklusjoner for mange tidsserier (se Findley et al. 1990).

2.7.1 Figur 7.a. De 11 målene M_1 - M_{11}

I tabell F.2 lister X12-ARIMA ut de 11 målene M_1 - M_{11} , hvor deres verdier varierer fra 0 til 3. For verdier under 1 er kvaliteten tilfredsstillende. Kvaliteten blir dårligere når M_i går mot 3. Formlene for M_i er gitt i notatet av L. Lothian og M. Morry (1978.c). Vi bruker fortsatt symboler: T_t for trend, A_t for sesongjusterte tall, S_t for sesongkomponent og I_t for irregulærkomponent. De 11 målene beskrives i det følgende

M_1 . *Relativt bidrag fra I_t til den totale variasjonen.* Når $M_1 > 1$ viser det at trend eller sesongkomponent kan bli forstyrret av variasjoner fra irregulærkomponenten. Sesongjusteringen er av lavere kvalitet.

M_2 . *Relativt bidrag fra I_t til den stasjonære delen.* Bidraget av I_t sammenlignes med den originale serien etter å ha fjernet trendkomponenten. Hvis $M_2 > 1$ vil det vise at bidraget fra I_t er høyt.

M_3 . *Endring fra måned til måned i irregulærkomponenten i forhold til endring i trend.* Formålet i sesongdekomponeringen er å fjerne sesongkomponenten fra den originale serien for å få et godt bilde av den sesongjusterte serien. Ikke bare det, trenden og den irregulærkomponenten må også være helt klar. Hvis endringen fra måned til måned i I_t er større enn endringen i trenden blir det vanskelig å skille disse to komponentene. M_3 er laget for denne testen. Når $M_3 > 1$ er variasjonen til I_t for høy.

M_4 . *Mål på autokorrelasjon i irregulærkomponenten.* Når $M_4 > 1$ vil det vise at I_t ikke er en uavhengig tilfeldig prosess. Dette medfører at alle tester som er basert på I_t kan være ugyldige.

M_5 . *Antallet måneder for at endringen i trend skal bli større enn endringen i irregulærkomponenten.* Testen er ikke signifikant når $M_5 > 1$.

M_6 . *Endring fra år til år i den irregulærkomponenten i forhold til endringen i sesongkomponenten.* For å skille de to komponentene S_t og I_t bruker man (3×5) filteret for SI komponenten. Når $M_6 > 1$ betyr det at (3×5) filteret ikke er tilpasset, siden filteret ikke er fleksibelt nok til å følge sesongvariasjoner.

M_7 . *Grad av bevegelig sesongmønster i forhold til stabilt sesongmønster.* Formålet er å måle hvor stor andelen av den stabile delen er i forhold til den bevegelige delen i sesongkomponenten. Når $M_7 > 1$ sier vi at sesongmønsteret ikke er identifisert. Årsakene til at $M_7 > 1$ er

1. Ingen sesong i det hele tatt i den originale serien.
2. Variasjonen i sesongkomponenten er så stor at sesongmønsteret ikke kan identifiseres.
3. Additiv modell er brukt når serien tilpasses bedre med multiplikativ modell.

Ofte vil man ikke sesongjustere en serie når verdien til M_7 er større enn 1.

M_8 . *Grad av fluktasjoner i sesongkomponenten gjennom hele serien.* Når $M_8 > 1$ vil det vise at fluktasjonene er så store at S_t ikke er stabil lenger. Dermed får en dårlig estimat for den sesongjusterte serien.

M_9 . *Grad av linær bevegelse i sesongkomponenten i hele serien.* Når $M_9 > 1$ viser det at fluktasjoner i S_t ikke er tilfeldige.

M_{10} . *Grad av fluktasjoner i sesongkomponenten, i de siste årene.* Denne testen er den samme som M_8 men bare for de siste årene.

M_{11} . *Grad av linær bevegelse i sesongkomponenten i de siste årene.* Denne testen er den samme som M_9 men bare for de siste årene.

Vektene for de 11 målene er gitt i tabell 10.

Tabell 10: *Vektene for M_1 - M_{11}*

	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	Sum
Verdi	0.13	0.13	0.10	0.05	0.11	0.10	0.16	0.07	0.07	0.04	0.04	1

Tilslutt har vi en størrelse Q som er definert slik

$$Q = \sum_{i=1}^{11} w_i M_i \quad (10)$$

Målene som er relativt viktige har større vekter. Vi ser at M_1 , M_2 og M_7 har største vekter fore de er relativt sett viktigere enn de andre. Merk at

- De 11 målene gjelder bare for de fleste tidsseriene, men ikke for alle. De er bare empiriske mål.
- Mål som har større vekt er relativt sett viktigere.
- Hvis har en tidsserie der alle målene større enn 1, vil vi ikke sesongjustere denne tidsserien.
- Ingen mål kan alene bestemme kvaliteten av sesongdekomponeringen.
- Når $M_7 > 1$ skal en være svært forsiktig med å sesongjustere denne tidsserien.

2.7.2 Figur 7.b. Sliding spans teknikk

Formålet med sliding spans teknikken (Findley et al. 1990) er å vurdere stabiliteter for sesongkomponenten S_t og sesongjusterte tall A_t i et bestemt tidspunkt når vi gjentar sesongjusteringsprosedyren på opp til fire forskjellige delvis overlappende delmengder av rådata (såkalt på engelsk). Hver span er å betrakte som en komplett tidsserie og behandles akkurat som den opprinnelige tidsserien.

La $S(\%)$ være prosentandelen måneder hvor sesongkomponenten er definert som upålitelig. $MM(\%)$ være prosentandelen måneder hvor sesongjusterte tall er definert som upålitelig. Evalueringen som er basert på sliding spans tekknikken er gitt i tabell 11. Verdiene av 11 mål for bilsalg er gitt i tabell 12. Vi får fra sliding spans $S(\%)=14.3$ og $MM(\%)=21.2$. Verdiene av kvalitetsmål og sliding spans viser en god sesongjustering for bilsalg.

Tabell 11: *Evalueringen for sesongjustering ved sliding spans teknikken*

$S(\%)$	$MM(\%)$	Konklusjon
$S(\%) \leq 15$	$MM(\%) < 40$	god sesongjustering
$15 < S(\%) \leq 25$	$MM(\%) < 40$	mindre god sesongjustering
$S(\%) > 25$	$MM(\%) > 40$	dårlig sesongjustering

Tabell 12: *Verdiene av M_1 - M_{11} etter sesongjustering for bilsalg*

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	Q
Verdi	0.43	0.10	0.42	0.96	0.55	0.04	0.29	0.68	0.33	0.50	0.45	0.40

2.7.3 Figur 7.c. Oppsummering for sliding spans og kvalitetsmålene

Her beskriver vi grenseverdier for en god sesongjustering ved sliding spans teknikken og kvalitetsmål.

2.7.4 Kommentarer for kvalitetsmål og sliding spans teknikk

Grenseverdiene for kvalitetsmål og sliding spans er basert på simuleringer av tusenvis av tidsserier fra US Census og Statistics Canada. Det gjelder for de fleste tidsserier men ikke alle. Grenseverdiene kan variere fra land til land og fra tidsserier til tidsserier. Brukeren kan endre dem slik at de blir tilpasset tidsserier i sitt eget land. Kunnskap om tidsseriene som skal analyseres er helt nødvendig for å bidra til vurdering av sesongjustering.

2.8 SKATTERPLOTT FOR KOMPONENTENE

Vi lager i figur 8 scatterplott for ujusterte verdier etter å ha justert for kalendereffekter og ekstremverdier O_t^K (men ikke ujusterte verdier O_t , siden O_t^K skal spaltes opp i komponentene), trend T_t , sesong- og irregulærkomponenten, S_t og I_t , for å sjekke korrelasjoner mellom dem. Disse komponentene må være ukorrelerte hver andre.

Vi får for bilsalg $Corr(O_t^K, T_t) = 0.93$, $Corr(O_t^K, A_t) = 0.90$ og $Corr(O_t^K, S_t) = 0.32$. Dette viser at data er dominert mest av trend og i mye mindre grad av sesongvariasjoner. Vi får i tillegg $Corr(T_t, A_t) = 0.96$. Siden sesongjusterte verdier er trend og tilfeldig støy, vil en meget høy korrelasjon mellom sesongjusterte verdier og trend vise ganske lav tilfeldig støy.

Siden korrelasjonene mellom trend, sesong- og irregulærkomponenten er lave $Corr(T_t, S_t) = -0.0044$, $Corr(T_t, I_t) = 0.093$ og $Corr(S_t, I_t) = -0.043$ vil det vise små sammenhenger mellom dem.

3 Oppsummering

I avsnitt 2 har vi presentert figurene for sesongjustering med eksempel fra bilsalg. Vi får

- Figur 1 beskriver ujusterte verdier og sesongvariasjoner i tidsserien. Noen teknikker som autokorrelasjonsfunksjon og spektralanalyse er brukt for å oppdage sesongvariasjoner og boxplott er brukt for å undersøke fordeling av ujusterte verdier for tolv måneder.
- Figur 2 viser tidsserien etter korrigering for kalendereffekter og ekstremverdier. Her kan vi sjekke hvor god korrigeringen for bevegelige helligdager som påske, pinse og Kristihimmelfartsdag er.
- Figur 3 lages for trend, ujusterte og sesongjusterte verdier samt revisjoner av trend og sesongjusterte verdier når nye observasjoner tilføyes.
- Figur 4 er for sesongkomponenten.
- Figur 5 er for irregulærkomponenten.
- Figur 6 viser SI komponenten. Det er komponenten som inneholder både sesongvariasjoner og tilfeldige variasjoner.
- Figur 7 beskriver kvalitetsmål og sliding spans teknikk.
- Figur 8 viser korrelasjoner mellom komponentene.

Vi har laget mange detaljerte figurer for sesongjustering. Vi plotter en komponent med forskjellige måter for at det blir lettere å se egenskaper ved tidsserien. Vi kan

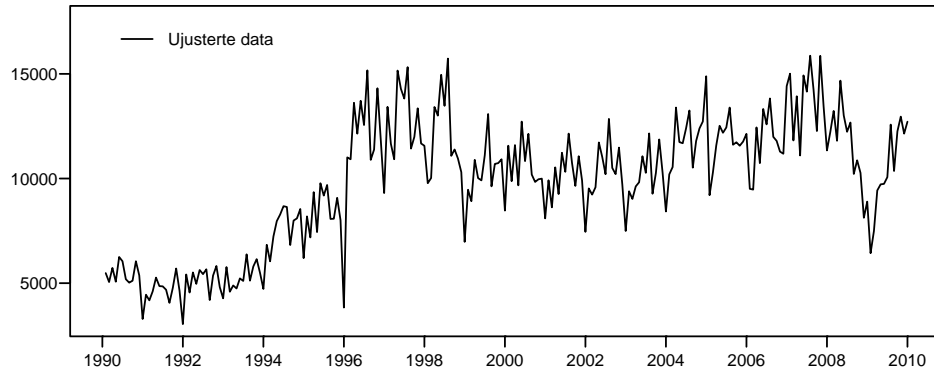
- sammenligne sesongvariasjoner før og etter korrigering for kalendereffekter og ekstremverdier (figur 1.b eller 1.c mot figur 2.f).
- sjekke om sesongmønsteret av ujusterte verdier blir klarere etter å ha korrigert tall med ukedagseffekter og ekstremverdier (figur 1.c mot figur 2.f) eller (figur 1.g mot figur 2.g).
- sammenligne ujusterte og sesongjusterte verdier (figur 3.a mot 3.b), endringer i prosent fra måned til måned i ujusterte og sesongjusterte verdier (figur 3.e mot figur 3.f) og revisjoner av trend og ujusterte verdier når nye observasjoner tilføyes (figur 3.g mot figur 3.h).
- sammenligne sesong og tilfeldige variasjoner (figur 4.a mot figur 5.a) eller (figur 4.c mot figur 5.a).

Vi har laget et program med R for alle disse figurene. Programmet henter tall fra output av X-12-ARIMA. Når vi lager figurer tar vi hensyn til at skala på y -aksen, dvs, verdiene på y -aksen av trend, ujusterte og sesongjusterte verdier er de samme, slik at vi kan legge figurer oppå hverandre for sammenligning. Vi gjør det samme for sesong- og irregulærkomponenten.

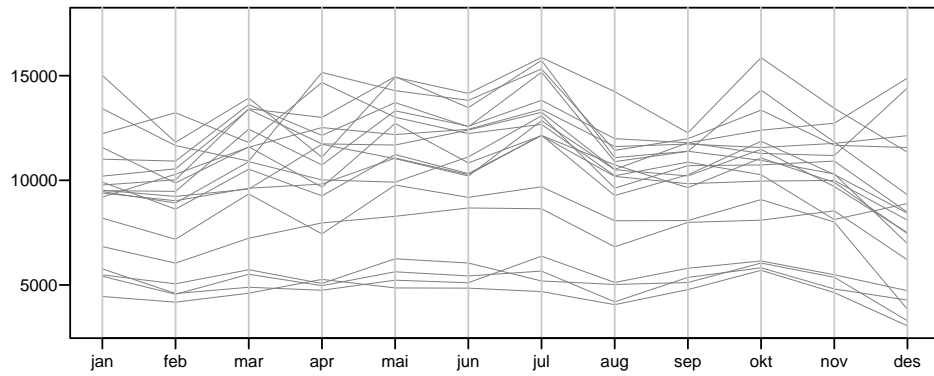
Tolkningen av en sesongjustering blir klarere når vi bruker både tall og figurer.

Vi har skrevet et program for å lage disse figurene etter sesongjustering. Programmet kjøres på UNIX.

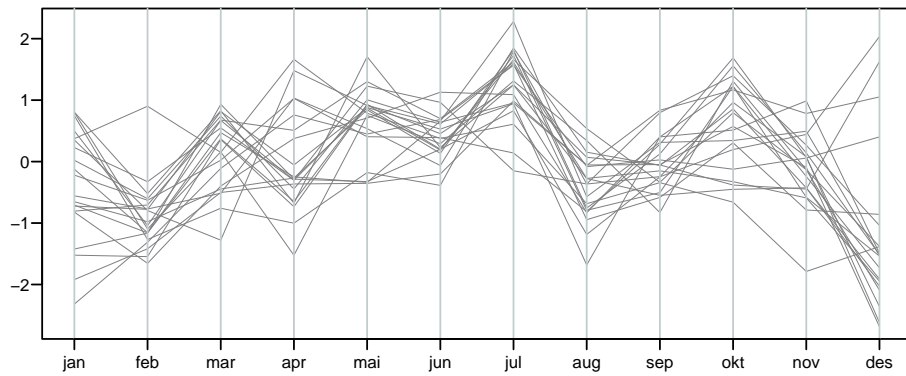
1.a. Ujusterte verdier



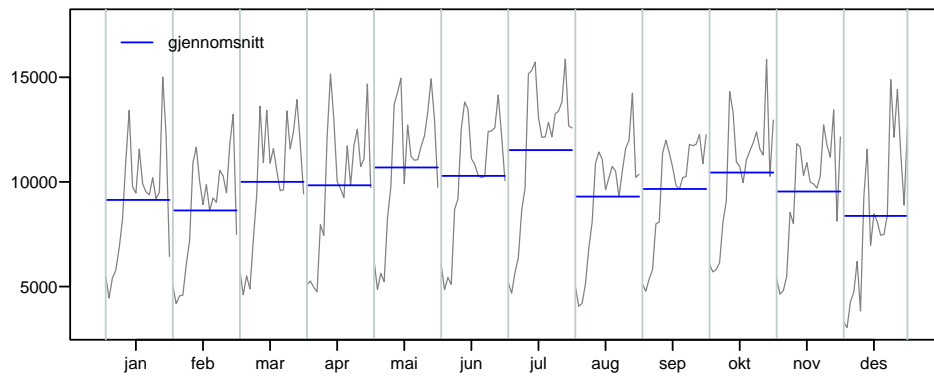
1.b. Ujusterte verdier for hver av årets 12 måneder



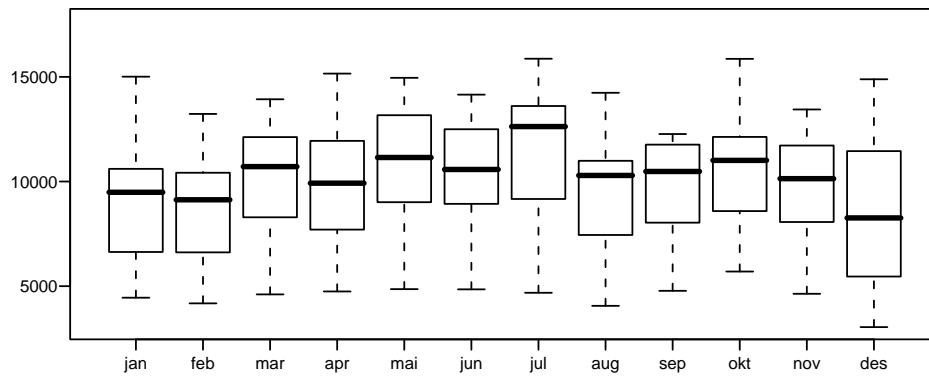
1.c. Ujusterte verdier for hver av årets 12 måneder (etter standardisering)



1.d. Ujusterte verdier fra år til år for hver enkel måned



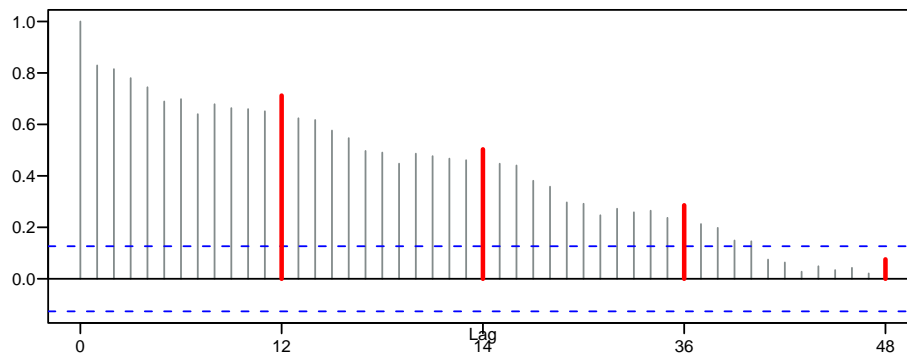
1.e. Boxplot for 12 måneder av ujusterte data



FORKLARING FOR BOX-WHISKER

- Det gir et visuelt bilde av fordelingen
- Det viser både tyngdepunktet i fordelingen (medianen) og spredningsmønsteret rundt dette tyngdepunktet
- Når det gjelder det siste, er det særlig to forhold vi skal legge merke til. Det første er bredden, det andre er eventuelle skevheter

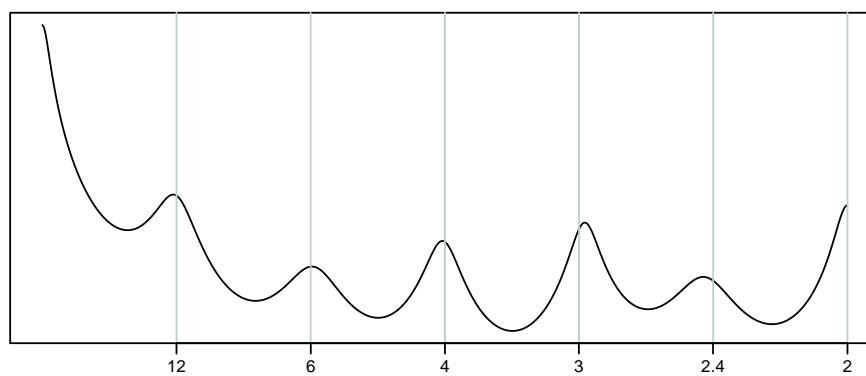
1.f. Autokorrelasjonsfunksjon (ACF)



FORKLARING FOR ACF

- Ved å plote de estimerte ACF etter lag k får vi et godt bilde av sesongvariasjonene
- Når $\rho(k)$ er signifikant større enn null i lagene 12, 24, 36, vil det vise sesongvariasjoner i data

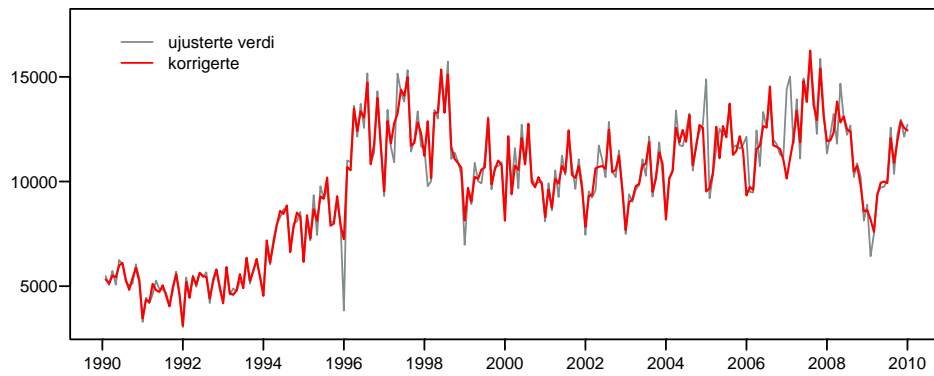
1.g. Spekteret av ujusterte data



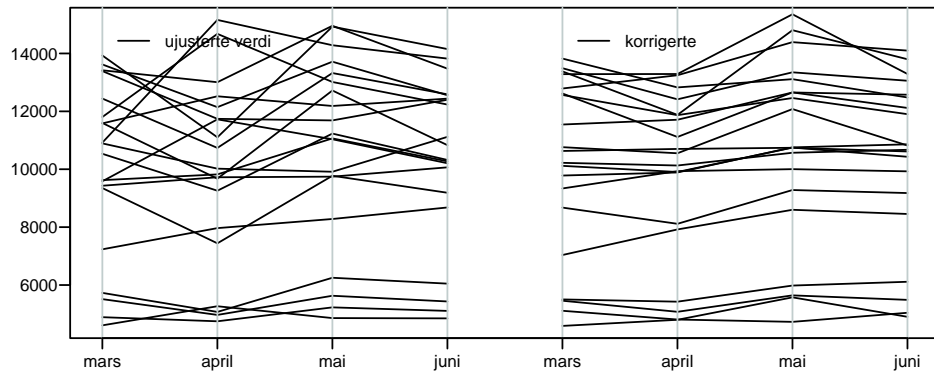
FORKLARING FOR SPEKTERET

- En prosess med tilfeldig variasjon har et flatt spekter
- For en månedlig tidsserie med sesongvariasjoner har spekteret et tynn spiss i periode 12
- En tidsserie hvor trenden er dominerende, er et høy spekter nær null

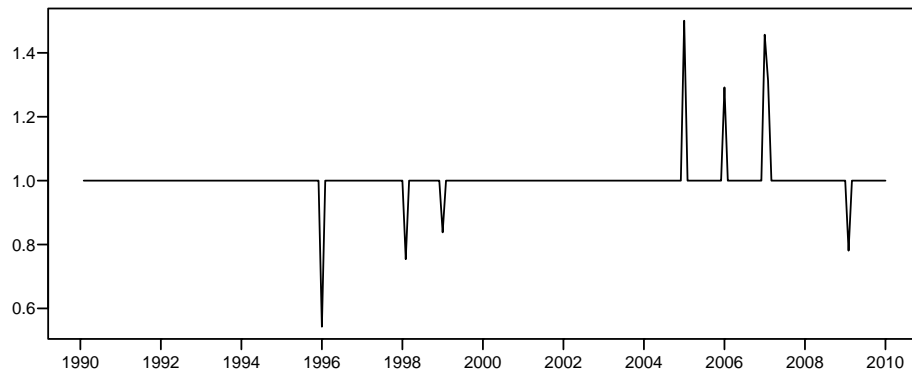
2.a. Ujusterte og korrigerede verdier for kalendereffekter og ekstremverdier



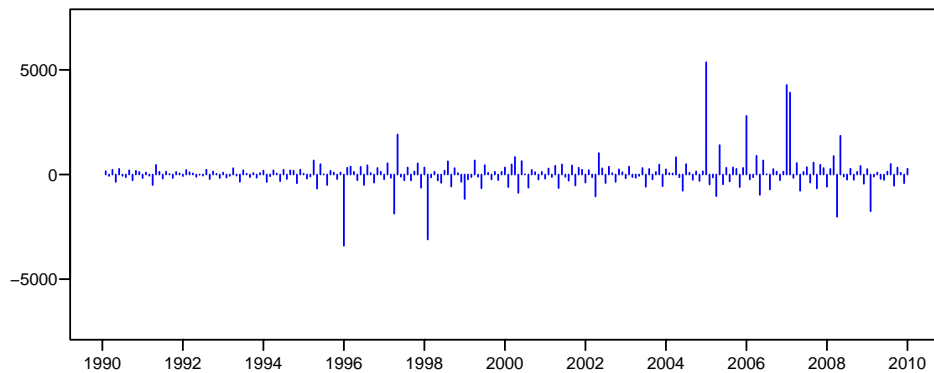
2.b. Ujusterte og korrigerede verdier for mars, april, mai og juni



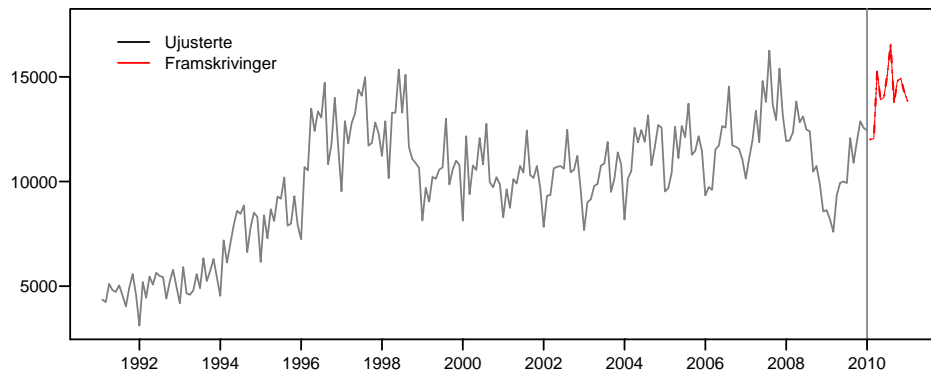
2.c. Korrigerede faktorer for ekstremverdier



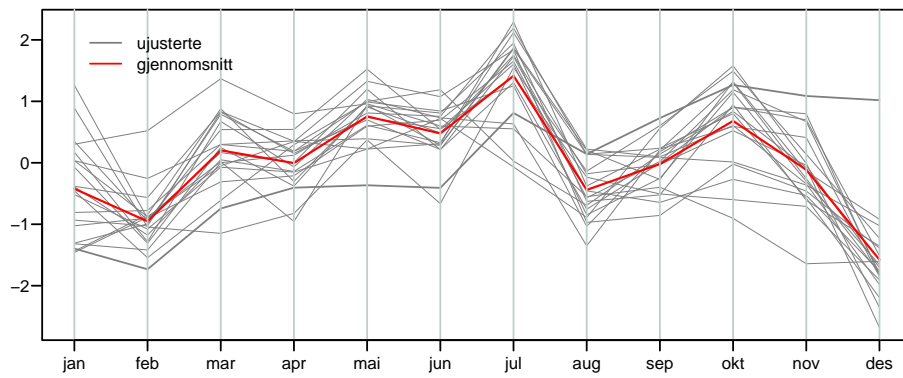
2.d. Avvikene mellom ujusterte og korrigerede verdier



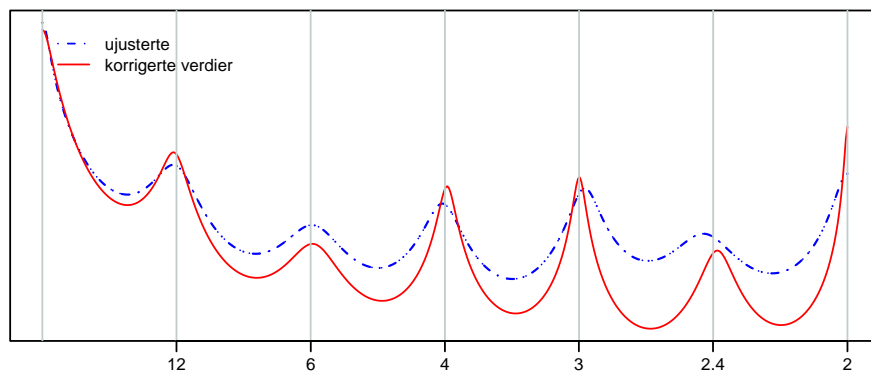
2.e. Ujusterte verdier og 12 måneder fram



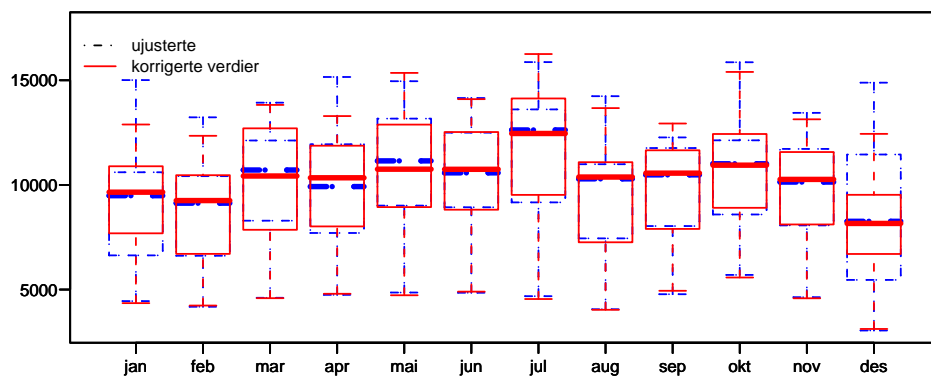
2.f. Korrigerede verdier for hver av årets 12 måneder etter standardisering



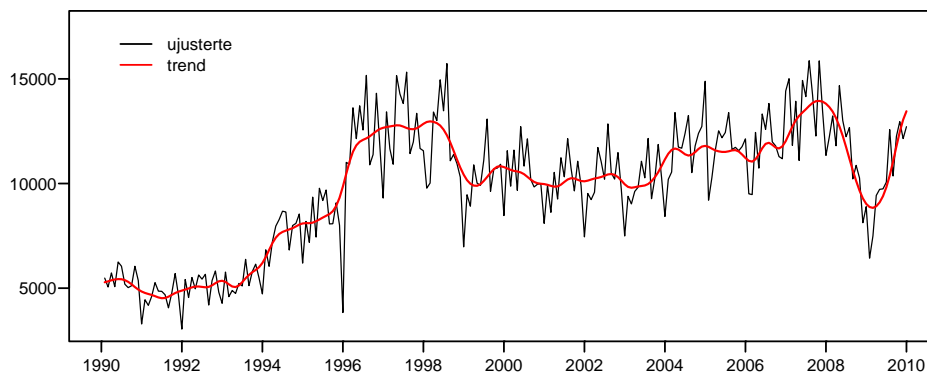
2.g. Spekteret av ujusterte og korrigerede verdier



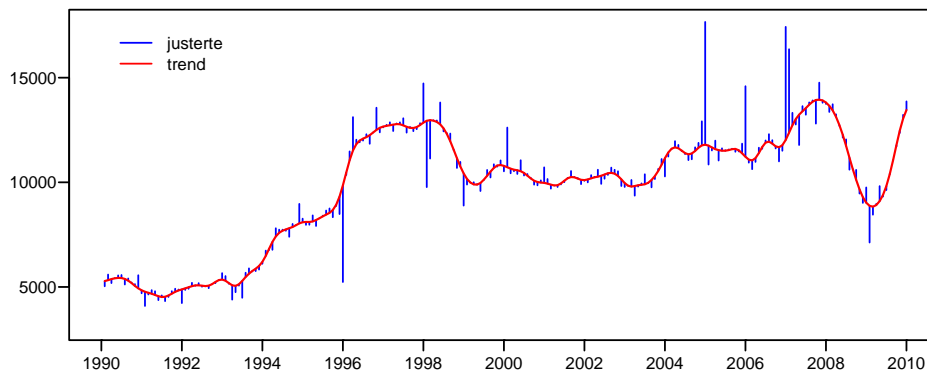
2.h. Boxplot for 12 måneder av ujusterte og korrigerede verdier



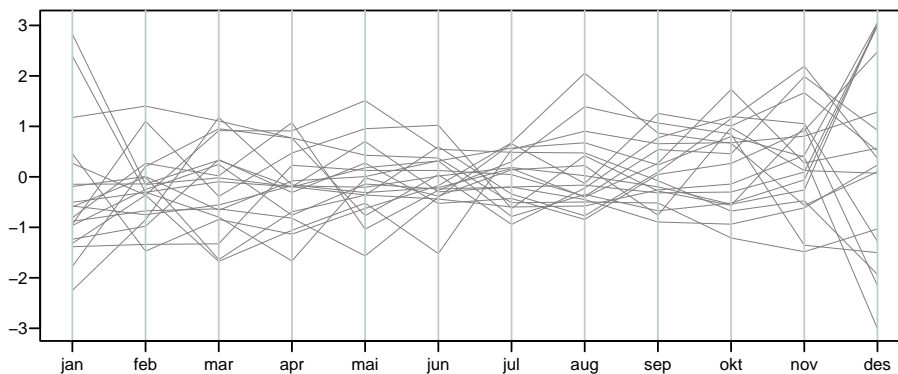
3.a. Ujusterte verdier og trend



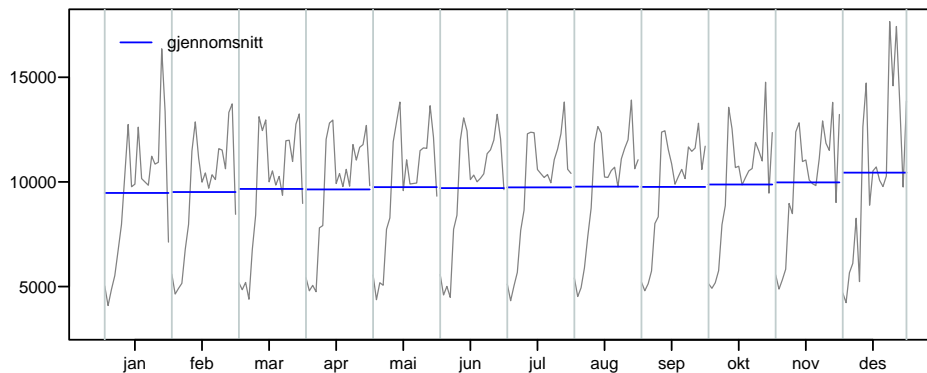
3.b. Sesongjusterte verdier og trend



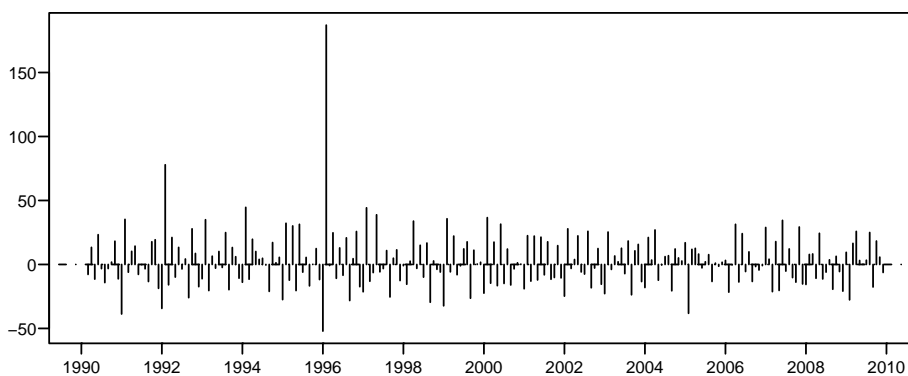
3.c. Se.justerte verdier for hver av årets 12 mnd (etter standardisering)



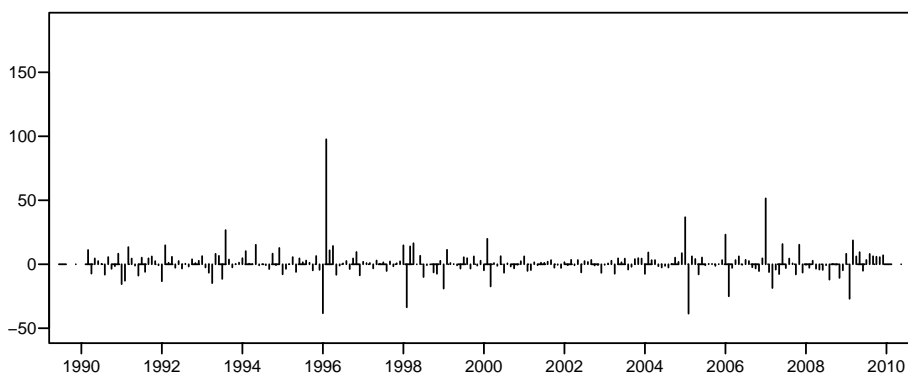
3.d. Sesongjusterte verdier fra år til år for hver enkel måned



3.e. Endring i prosent fra måned til måned i ujusterte verdier



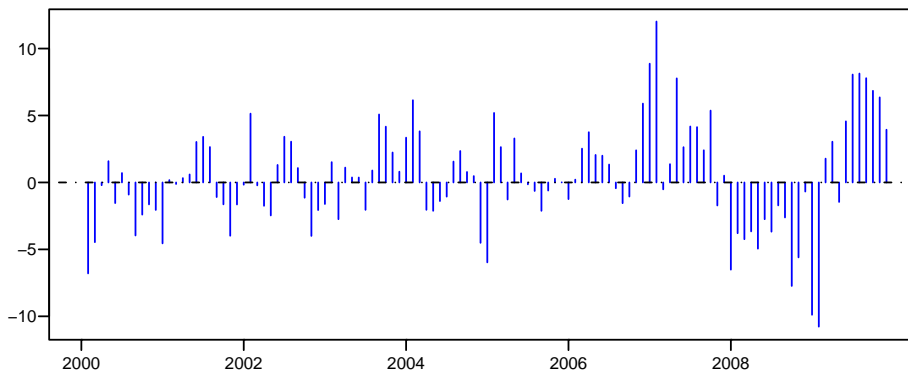
3.f. Endring i prosent fra måned til måned i sesongjusterte verdier



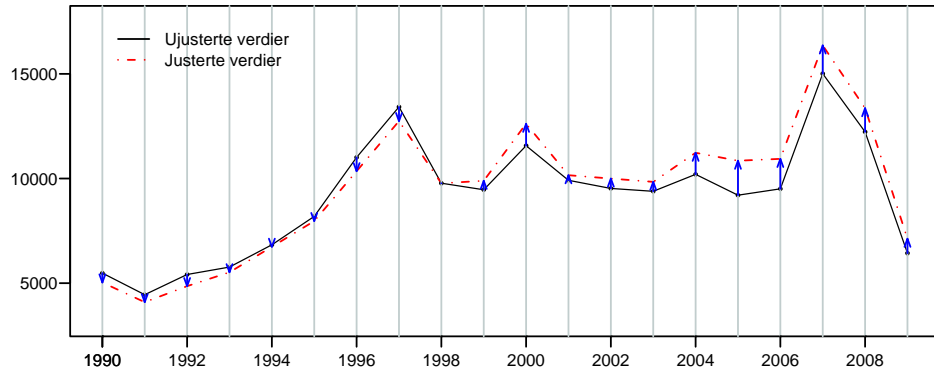
3.g. Revisjoner i prosent for sesongjusterte verdier



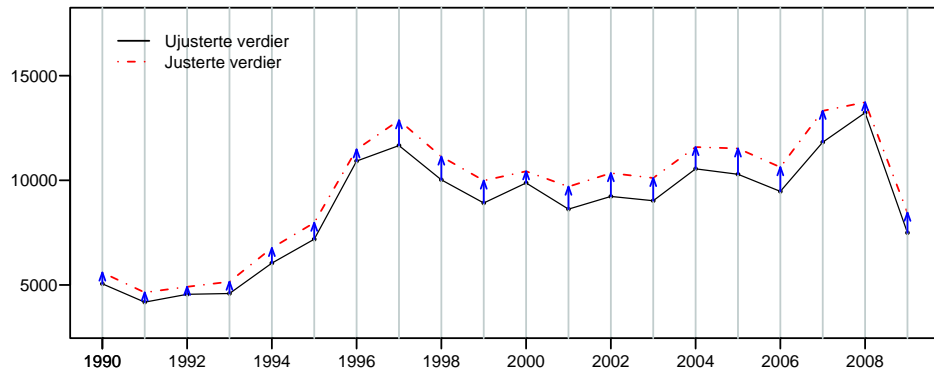
3.h. Revisjoner i prosent for trend



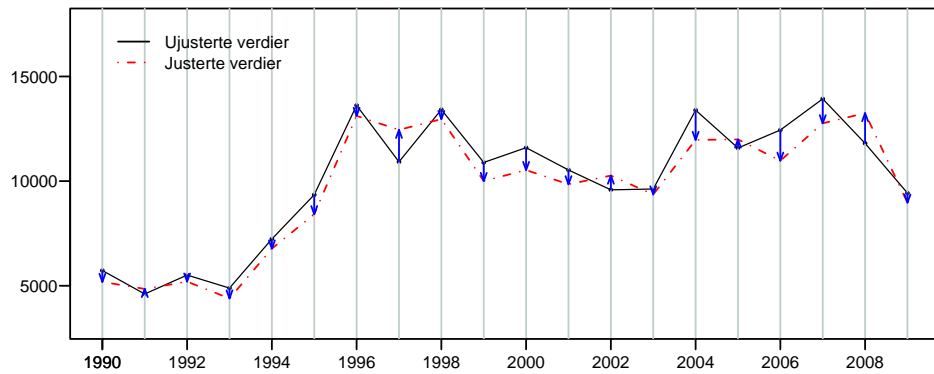
3.1. Avvik mellom ujusterte og sesongjusterte verdier for januar



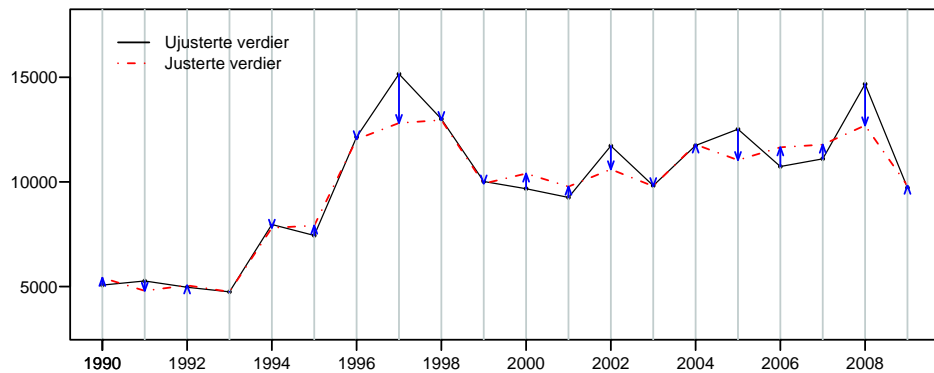
3.1. Avvik mellom ujusterte og sesongjusterte verdier for februar



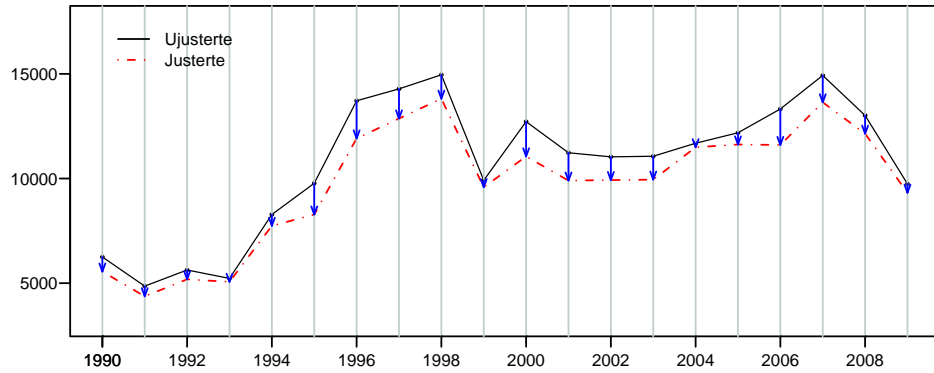
3.1. Avvik mellom ujusterte og sesongjusterte verdier for mars



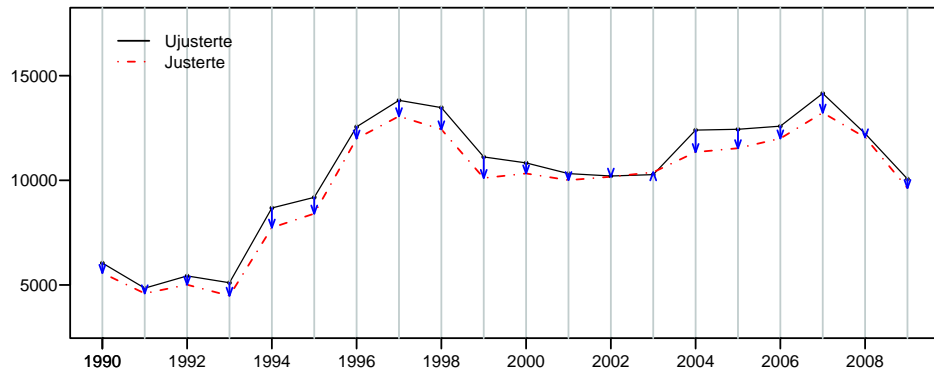
3.1. Avvik mellom ujusterte og sesongjusterte verdier for april



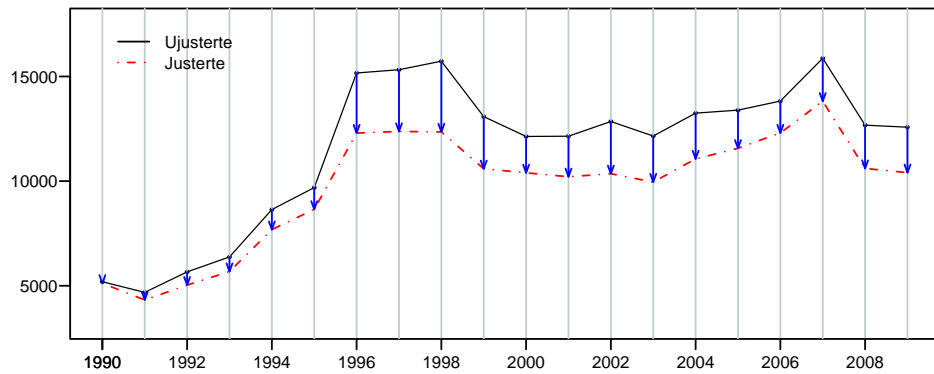
3.1. Avvik mellom ujusterte og sesongjusterte verdier for mai



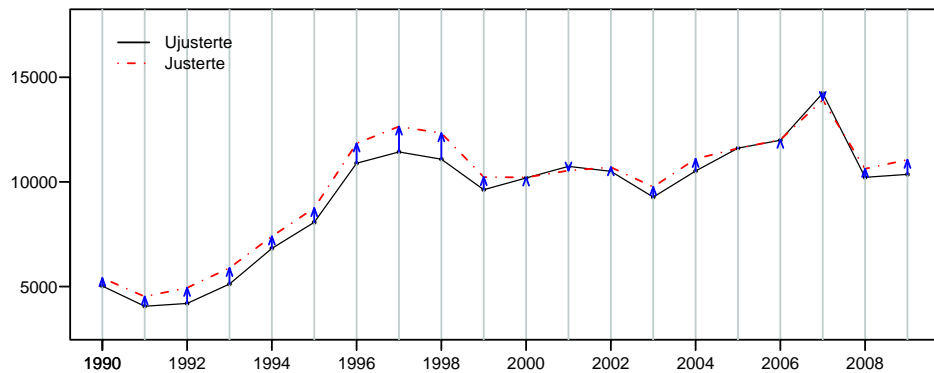
3.1. Avvik mellom ujusterte og sesongjusterte verdier for juni



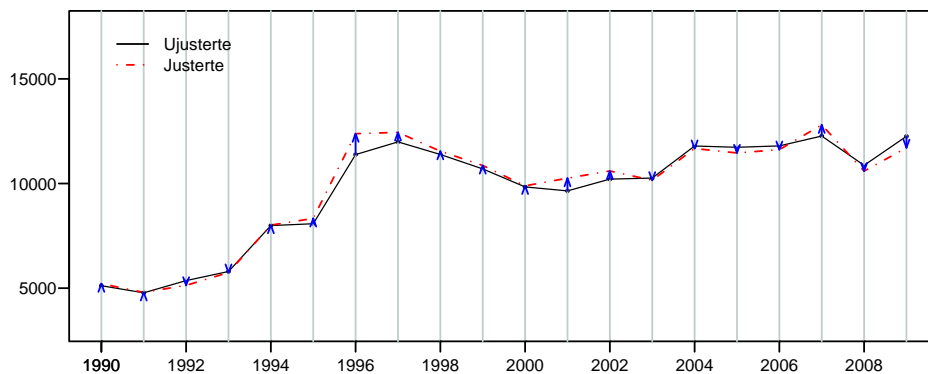
3.1. Avvik mellom ujusterte og sesongjusterte verdier for juli



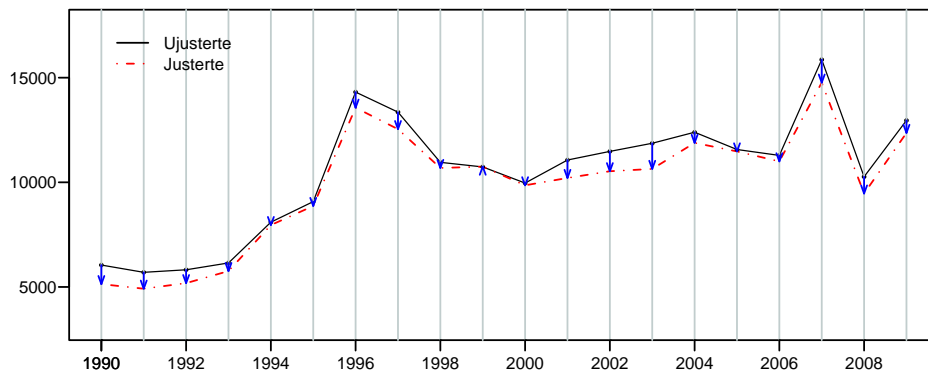
3.1. Avvik mellom ujusterte og sesongjusterte verdier for august



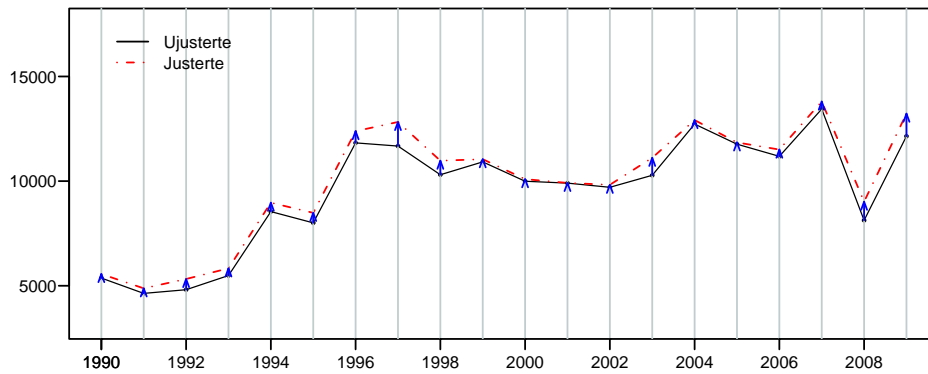
3.I. Avvik mellom ujusterte og sesongjusterte verdier for september



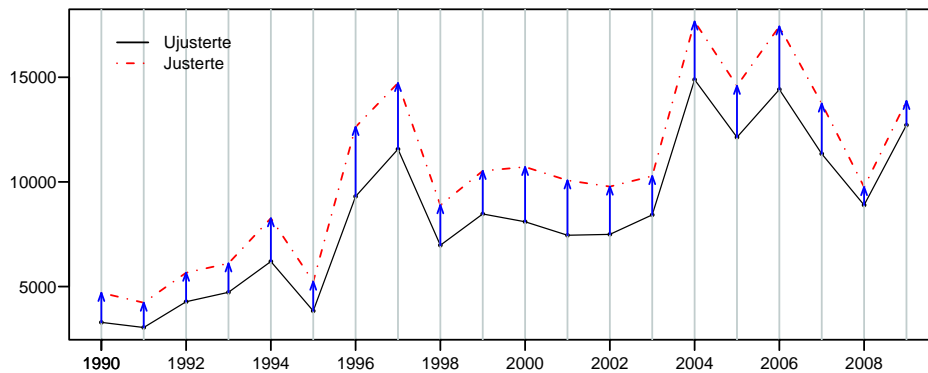
3.I. Avvik mellom ujusterte og sesongjusterte verdier for oktober



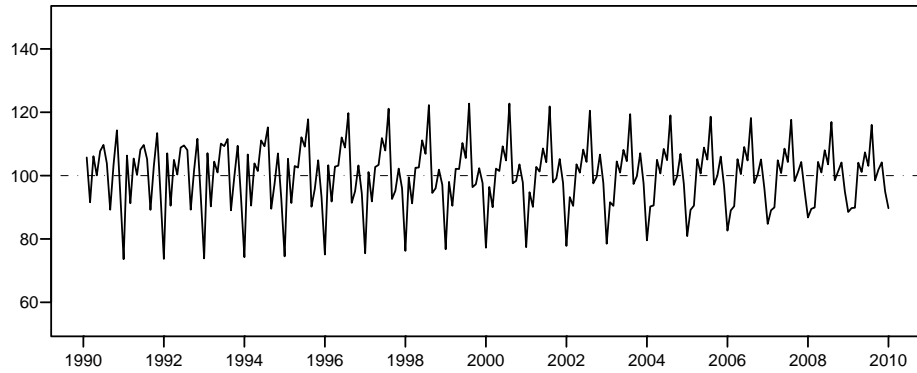
3.I. Avvik mellom ujusterte og sesongjusterte verdier for november



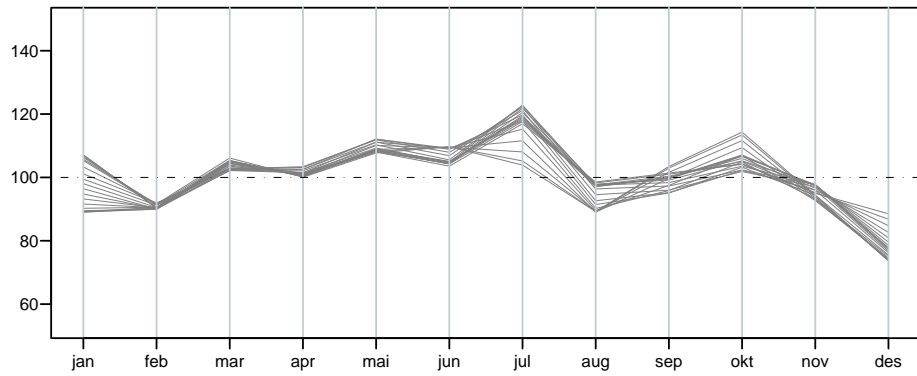
3.I. Avvik mellom ujusterte og sesongjusterte verdier for desember



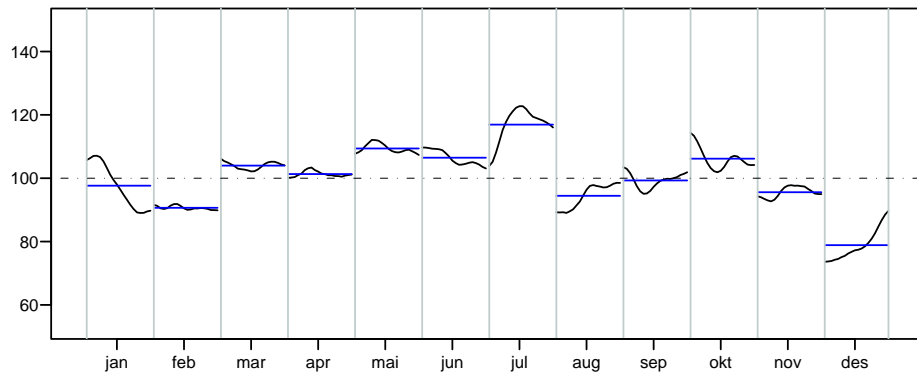
4.a. Sesongkomponenten



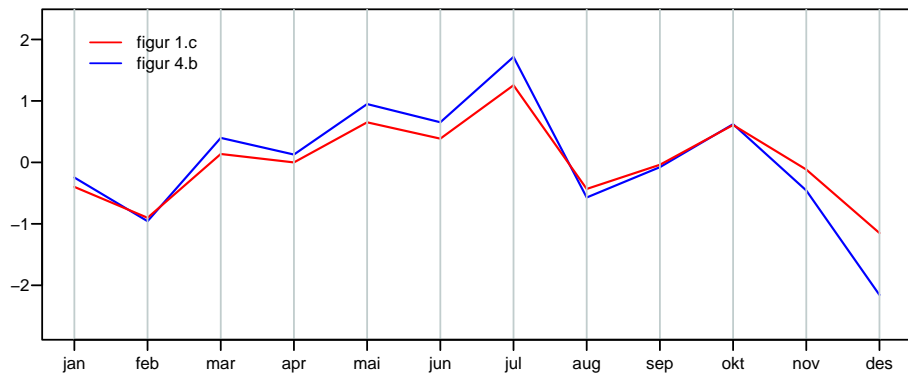
4.b. Sesongkomponenten for hver av årets 12 måneder



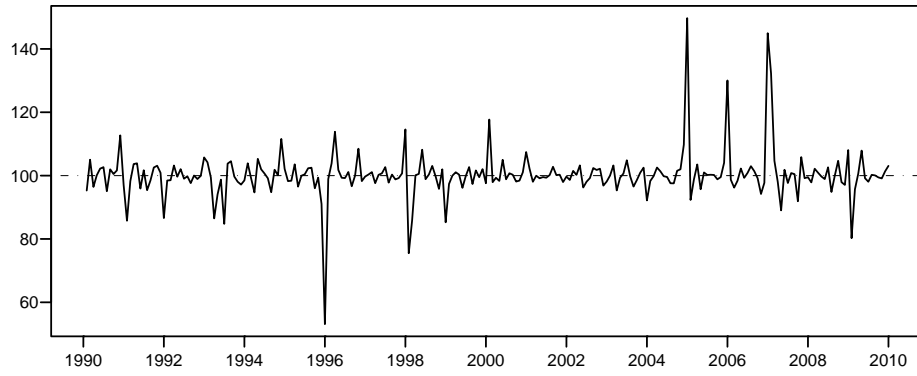
4.c. Sesongkomponenten fra år til år for hver enkel måned



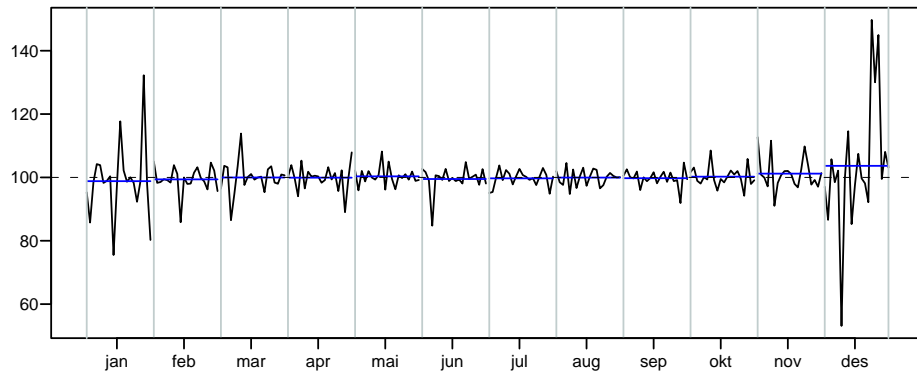
4.d. Gjennomsnittene av de kurvene i figur 1.c og 4.b (etter standardisering)



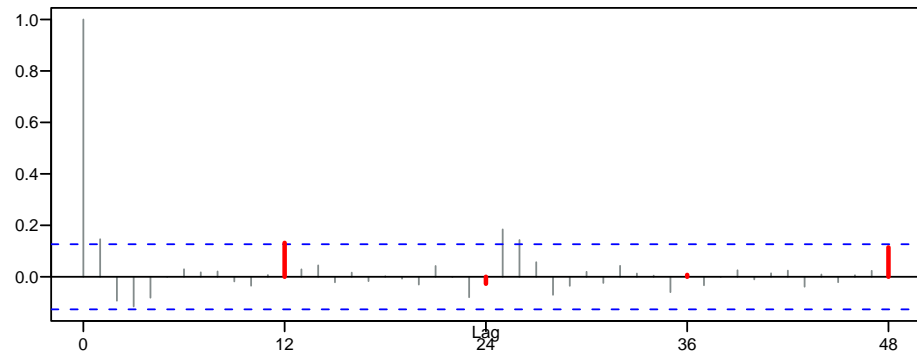
5.a. Irregulærkomponenten



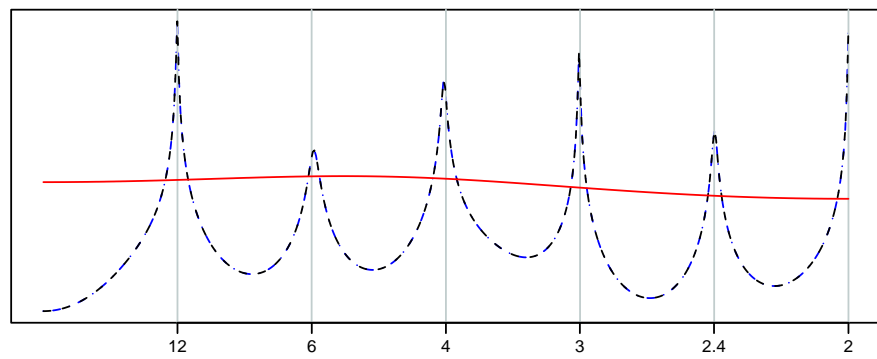
5.b. Irregulærkomponenten fra år til år for hver enkel måned



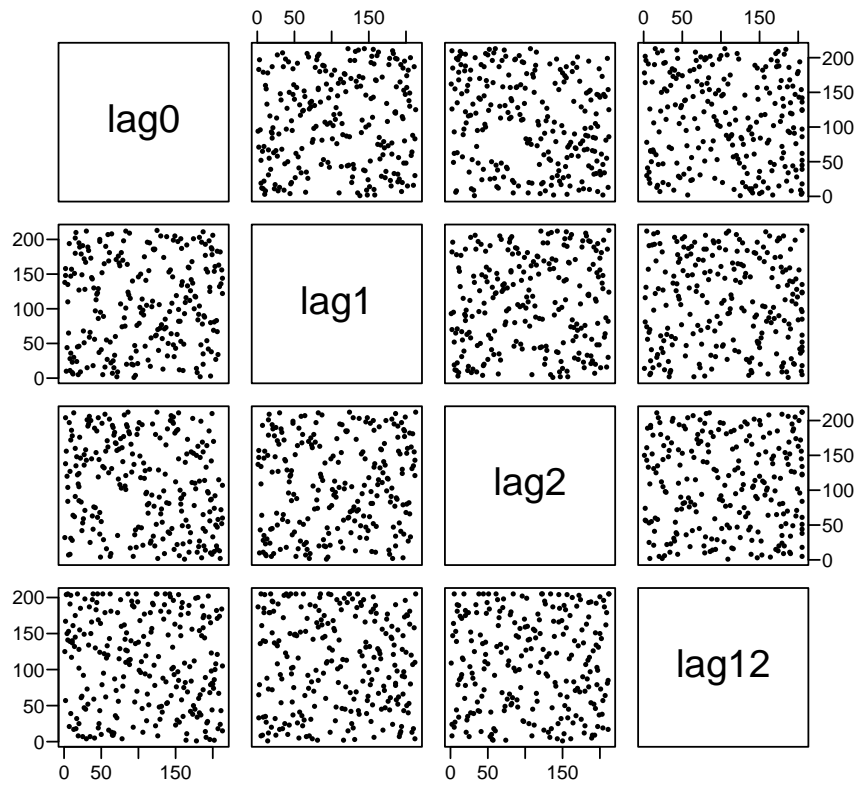
5.c. Autokorrelasjonsfunksjon (ACF)

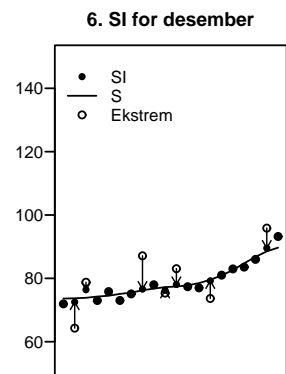
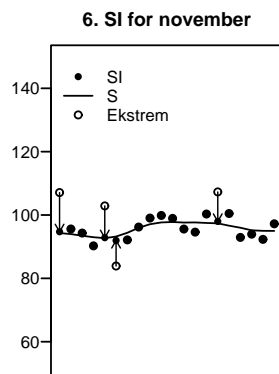
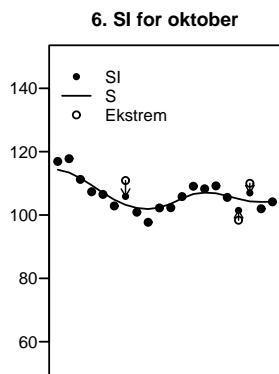
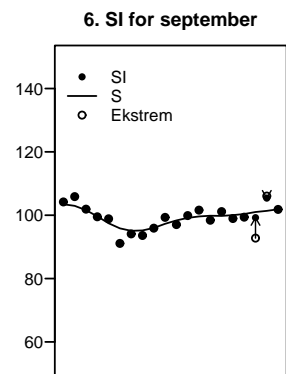
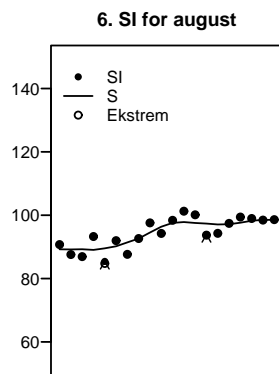
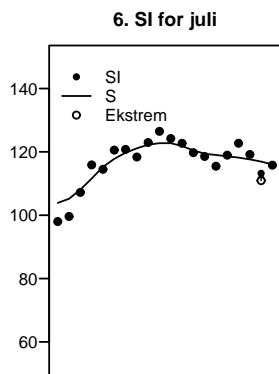
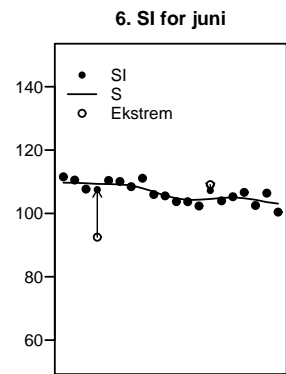
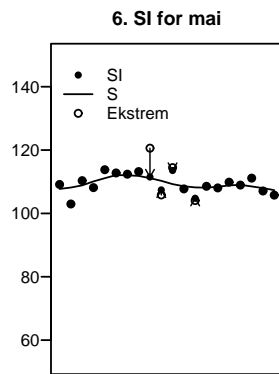
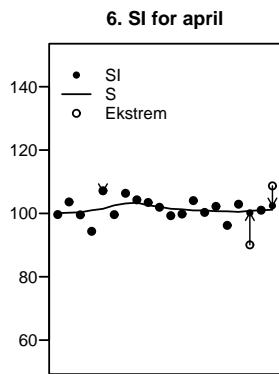
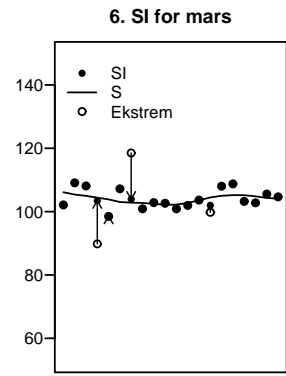
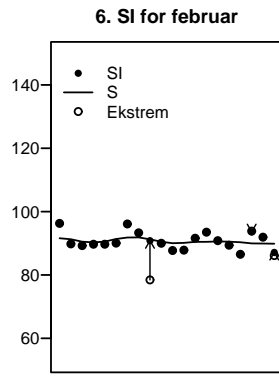
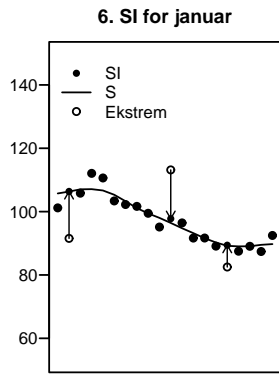


5.d. Spekteret av sesong- og irregulærkomponenten

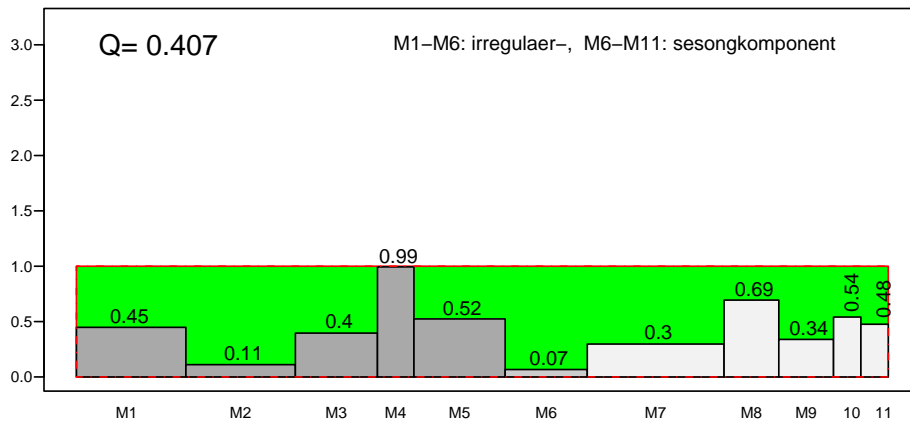


5.e. Scatter plott av irregulærkomponenten

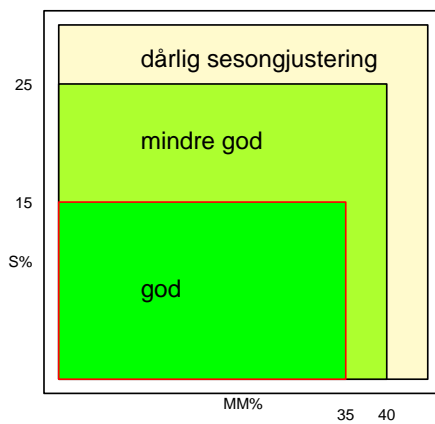




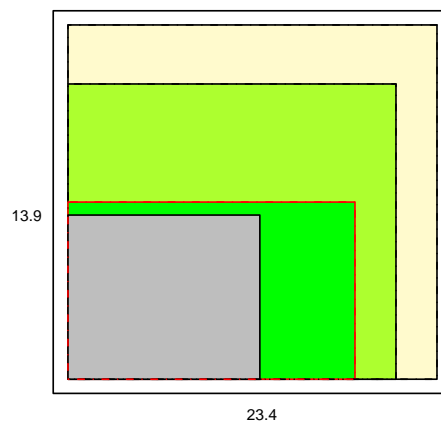
7.a. Kvalitetsmål og Q



7.b. Oppsummering. Sliding spans



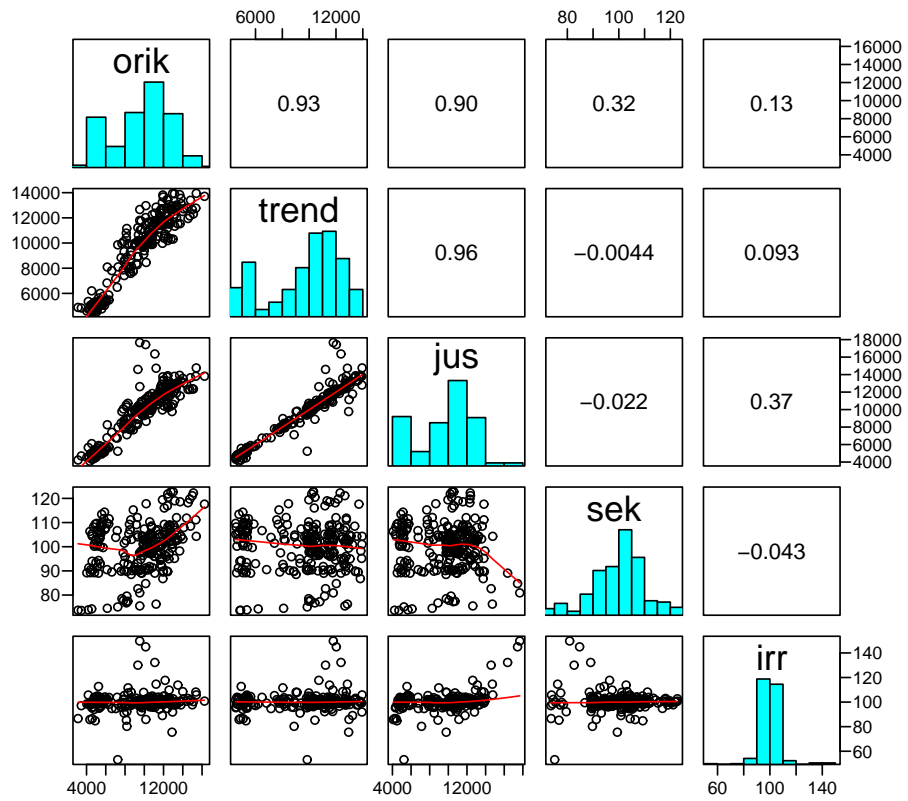
Sliding spans



7.c. Oppsummering. Sliding spans og vektorer for M og Q-verdi

Seasonal Factors	15% is too high	25% is much too high
Month to month changes in SA	35% is too high	40% is much too high
S(%) ≤ 15%	MM(%) ≤ 40%	likely
15 < S(%) ≤ 25%	MM(%) ≤ 40%	less likely
S(%) > 25%	MM(%) > 40%	unlikely
Mål :	M1 M2 M3 M4 M5 M6 M7 M8 M9 M10 M11	
Vekt:	0.13 0.13 0.10 0.05 0.11 0.10 0.16 0.07 0.07 0.04 0.04	

8. Scatter plott for komponentene



Referanser

- [1] Alan Pankratz (1991), *Forecasting with Dynamic Regression Models*, Wiley Interscience.
- [2] Bell W. R. and Hillmer S. C. (1983). "Modelling Time Series With Calendar Variation". *Journal of the American Statistical Association*, 78, 526-534.
- [3] Bureau of the Census. "X-12 ARIMA Reference Manual, Version 0.2.5, October 1, 1999"
- [4] Cleveland W. S. and Susan J. D. (1980), "Calendar Effects in Monthly Times Series: Detection by Spectrum Analysis and Graphical Methods", *Journal of the American Statistical Association*, 75, 487-495.
- [5] Findley D. F., Brian C. Monsell, William R. Bell, Mark C. Otto and Bor-Chung Chen (1998). "New Capabilities and Methods of the X-12 ARIMA Seasonal Adjustment Program", *Journal of Business & Economic Statistics*, 16, 127-177.
- [6] Dagum Estela Bee (1988). "Seasonal Adjustment in the Eighties: Some Problems and Solutions", *The Canadian Journal of Statistics*, Vol 16 , Supplement, 1988, Pages 109-125.
- [7] Dagum Estela Bee, Benoit Quenneville and Brajendra Sutradhar (1992). "Trading-day Variations Multiple Regression Models with Random Parameters", *International Statistical Review*, 60, 57-73.
- [8] Dagum Estela Bee and Marietta Morry (1984). "Basis Issues on the Seasonal Adjustment of the Canadian Consumer Price Index", *Journal of Business & Economic Statistics*, Vol. 2, No. 3 (jul., 1984), pp. 250-259.
- [9] Dagum Estela Bee (1988). "The X11ARIMA/88 Seasonal Adjustment Method Foundations and User's Manual".
- [10] David F. Findley, Kellie Wills, and Brian C. Monsell (2001). "Issues in Estimating Easter Regressors Using RegARIMA Models with X-12-ARIMA", Australian Statistical Publishing Association Inc 2001.
- [11] Hillmer, S. C. and G. C. Tiao (1982) , "An ARIMA Model-Based Approach to Seasonal Adjustment", *Journal of the American Statistical Association*, Vol. 77, No. 377. (Mar., 1982), pp. 63-70.
- [12] John Higginson (1975). "An F Test for the presence of moving seasonality when using census method II-X-11 variant".
- [13] Lars A. Loe (1987). "Framskrivning av tidsseriedata i kvartalsvis nasjonalregnskap", *Notater* 87/1.
- [14] Leiv Solheim og Dinh Quang Pham (1997). "Prekorrigering av påskeeffekten for detaljvolumindeksen 1979-1997", *Notater* 73/97.
- [15] Lothian J. and M. Morry. "A set of Quality Control Statistics for the X-11 ARIMA".
- [16] Maraval A. (2005), "An Application of the TRAMO-SEATS Automatic Procedure; Direct versus Indirect Adjustment", *Documentos de Trabajo*, No. 0524, Banco de Españã.
- [17] Primrose E.J.F. (1951). "The Mathematics of Easter", *The Mathematical Gazette*, Vol. 35, No. 314. (Dec., 1951), pp 225-227.

- [18] Xichuan (Mark) Zhang, Craig H. McLaren and Cales C.S. Leung. "An Easter Proximity Effect: Modelling and Adjustment", *Aust. N. Z. J. Stat* 43(3), 2001, 269-280.
- [19] William S. Cleveland and Susan J. Devlin. "Calendar Effects in Monthly Time Series Modeling and Adjustment", *JASA*, Vol. 77, No. 379 (Sep., 1982), pp. 520-528.
- [20] William S. Cleveland and Irma J. Terpenning. "Graphical Methods for Seasonal Adjustment", *JASA*, Vol. 77, No. 377 (Mar., 1982), pp. 52-62.
- [21] *An Introduction to R, Version 2.5.1*