Sæsonkorrigering

DANMARKS
STATISTIK

Sæsonkorrigering

Udgivet af
Danmarks Statistik
Maj 2002
Oplag: 300
Danmarks Statistiks trykkeri

ISBN 87-501-1305-4

Pris 148,00 kr. incl 25 pct. moms

Danmarks Statistik
Sejrøgade 11
2100 København Ø

e-post: dst@dst.dk
www. dst.dk

Forord

Målet med denne publikation er, at give en generel beskrivelse af begrebet sæsonkorrigering. Publikationen giver en dybdegående introduktion til emnet beregnet på de, der beskæftiger sig med sæsonkorrigeret data.

Publikationen indeholder ydermere en beskrivelse af X11ARIMA programmet og X-12 programmet.

Danmarks Statistik anvender X11ARIMA til sæsonkorrigering til og med året 2002, og fra 2003 anvendes X-12 som standard for sæsonkorrigeringen i Danmarks Statistik.

Danmarks Statistik, maj 2002

Otto Andersen

0. Motivation

Et vigtigt spørgsmål

Det vigtigste spørgsmål man i forbindelse med sæsonkorrigering bør stille sig selv er, om der i det hele taget er behov for sæsonkorrigerede tal.

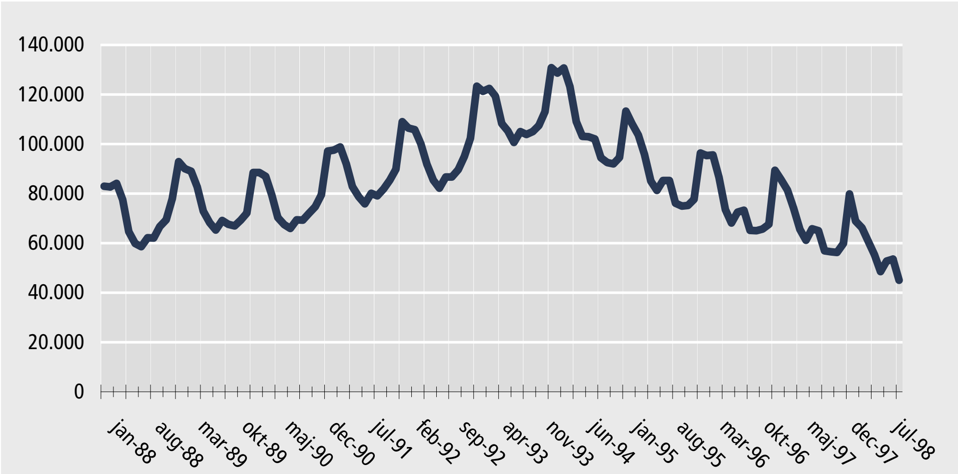
Før dette spørgsmål kan besvares på passende vis, er man nødt til at gøre sig klart, hvorfor man i det hele taget producerer statistik.

Danmark har brug for en stabil produktion af statistik, fordi de politiske og økonomiske beslutningstagere behøver information om samfundets aktuelle tilstand.

Men når man skal vurdere om, et givet nøgletal, som fx ledigheden, er faldet eller steget set i forhold til sidste måned, kan det være svært.

Det er svært, fordi der i ledighedstallene forekommer mere eller mindre regelmæssige sæsonsvingninger som gør sammenligninger fra måned til måned meningsløse, se for eksempel figuren nedenfor.

Figur 0.1 **Ledige mænd 25-59 år i perioden 1988-1998**

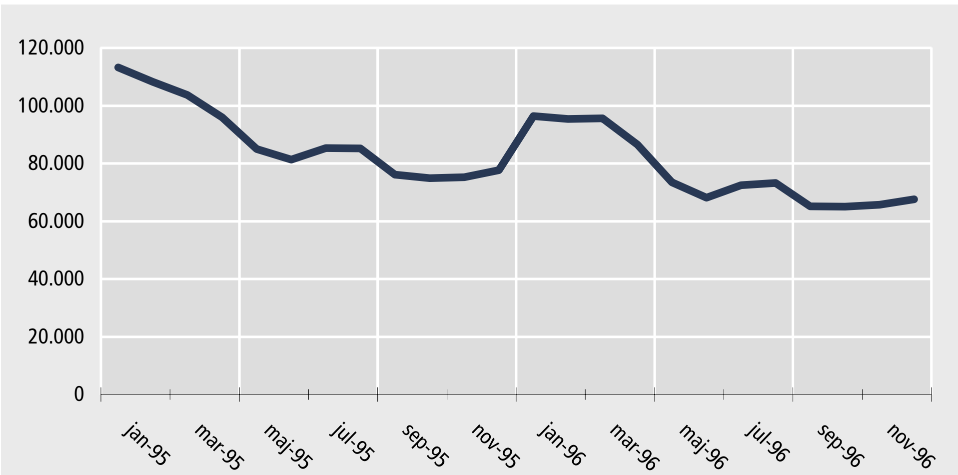


Man ser umiddelbart af figuren, at tallene varierer regelmæssigt efter et fast mønster indenfor det enkelte år, dvs. at der findes en tydelig sæsonvariation i tallene. Ydermere kan man se en variation i niveauet fra år til år.

Hvordan vurderes så udviklingen i ledigheden?

Hvordan kan man af ovenstående figur vurdere udviklingen fra december 1995 til januar 1996? Til dette formål betragtes et udsnit af figuren for perioden 1995-1996:

Figur 0.2 **Ledige mænd 25-59 år for perioden 1995-1996**



Det ses af den detaljerede figur, at kurven stiger brat fra december 1995 til januar 1996, men af den samlede figur ses det også, at denne udvikling finder sted hvert år fra december til januar. Derfor er man nødt til at finde ud af, hvor stor en del af denne stigning som kan skyldes den normale sæsonvariation. Den sæsonrelaterede del må der korrigeres for, således at man kan foretage den ønskede vurdering af ledighedstallene. Sæsonbevægelsen betragtes altså som en forstyrrende faktor, som man ved en konjunkturanalyse må forsøge at korrigere for.

Sæsonkorrigering er nødvendigt

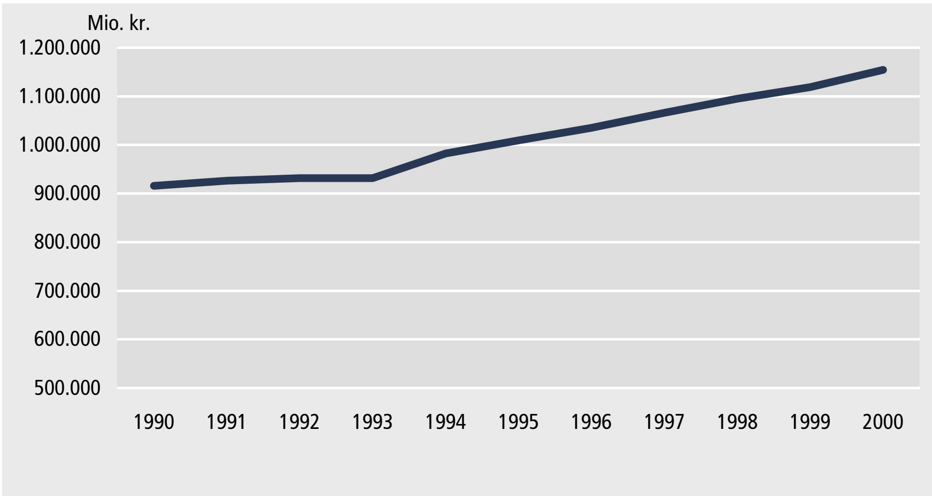
Man må altså konkludere, at der er et væsentligt behov for sæsonkorrigerede tal. Sæsonkorrigerede tal er et redskab, som muliggør en konstatering af vendepunkter i den økonomiske udvikling. For de økonomiske beslutningstagere er den information utrolig vigtig. Jo hurtigere et vendepunkt observeres, jo mindre drastisk behøver et politisk indgreb at være.

Bruttonational-produktet

Et andet eksempel på et økonomisk nøgletal er bruttonationalproduktet:

Eksempel: BNP-tallet
Den nødvendige information, der beskriver om landet er inde i en lav- eller en højkonjunktur, fås af BNP-serien. Informationen fås dog ikke umiddelbart ved at betragte udviklingen i serien. Ved først at rense for de sædvanlige sæsonmæssige udsving kan den ønskede vurdering af landets økonomiske tilstand foretages på meningsfuld vis.

Figur 0.3 **BNP i faste 1995-priser for perioden 1990-2000**



0.1 Tidsserieanalyse ved brug af komponentmodeller

Det er almindeligt at analysere tidsserier ved hjælp af de såkaldte komponentmodeller. I en komponentmodel opsplittes tidsrækken i fire forskellige komponenter.

Hvad er en komponentmodel?

Ved blandt andet at overveje, hvorledes man skulle kunne løse problemet med at skille det regelmæssige sæsonmønster ud fra den konjunkturserie som man ønsker at analysere, har man fundet på at opstille en såkaldt komponentmodel.

De fire komponenter

I komponentmodellen opfatter man tidsrækken som sammensat af fire forskellige komponenter, nemlig;

- **Sæsonkomponenten S**, som er de cykliske periodiske korttidssvingninger, med en periodelængde på et år[1].
- **Konjunkturkomponenten C**, som består af svingninger med en længere periode. Typisk vil disse svingninger være af langt mindre størrelsesorden end sæsonsvingningerne.
- **Trendkomponenten T**, som består af den langsigtede bevægelse i data.

---

[1] [1] Man kan også benytte komponentmodellen på tidsrækker med kortere periodiske svingninger, fx af 1 uges eller 24 timers længde, men det er ikke relevant for arbejdet i Danmarks Statistik

– **Den irregulære komponent I** er det rest-led der fanger de bevægelser op, som ikke er opfanget af modellens øvrige led.

Ekstreme observationer

I modelantagelsen ligger det eksplicit, at de irregulære bevægelser er tilfældige udsving. De irregulære bevægelser indeholder de såkaldte ekstreme observationer.

Ekstreme observationer kan for eksempel skyldes naturkatastrofer eller strejker.

De fire komponenter beskriver konjunkturserien

Disse fire komponenter antages typisk at kunne beskrive konjunkturserien enten ved at denne er lig summen eller produktet af de enkelte komponenter.

Trenden og konjunkturcyklen slås sammen

Da det er svært at skelne imellem trenden og konjunkturcyklen, som begge er ikke observerede, i en tidsserie, slås disse to komponenter ofte sammen og benævnes trendcyklen (TC).

0.1.1 Sæsonkomponenten

Det analysemæssige formål med komponentmodeller er at eliminere de ikke-observerede sæsonbevægelser. Derfor må følgende spørgsmål besvares:

Hvorledes måler man sæsonvariationen og hvis den er til stede, hvorledes fjerner man den så?

Sæsonkomponenten indeholder de kortsigtede regelmæssige bevægelser

Sæsonkomponenten fanger bevægelser, som gentages mere eller mindre regelmæssigt hvert år. For de fleste seriers vedkommende observeres der typisk et ensartet udviklingsforløb hen over et kalenderår.

Sæsonkomponenten indeholder typisk både de såkaldte *klimatiske* faktorer og de såkaldt *institutionelle* faktorer.

Klimatiske faktorer…

De klimatiske faktorer dækker over alle vejrligsbetingede forhold.

…og institutionelle faktorer

De institutionelle forhold dækker over de kalendermæssige konventioner, hvorefter vi vælger at indrette vores samfund. Eksempelvis dækker begrebet over placeringen af religiøse højtider, feriesæson, eller terminer for ind- og udbetalinger fra/til offentlige myndigheder.

0.2 Metoder til sæsonudrensning

Der er ingen entydig metode til sæsonkorrigering…

Når målet er at fjerne sæsonkomponenten fra en given konjunkturserie, findes der ikke én rigtig måde at gøre det på. Dette skyldes, at man ønsker at fjerne et ikke kendt element fra tidsserien.

…der er mange muligheder

Der er derfor mange måder at sæsonudrense på, og metoderne spænder lige fra en simpel udregning af vækstrater til avancerede modelbaserede metoder. Udvalgte metoder vil blive gennemgået nedenfor.

0.2.1 Den årlige vækstrate udregnes

Udviklingen imellem to kalenderår kan udregnes

Den mest enkle måde at fjerne de regelmæssige sæsonsvingninger på er at udregne udviklingen imellem to kalenderår for samme periode.

Metoden kan fungere rimeligt…

Denne enkle metode fungerer rimeligt, når blot sæsonsvingningerne i den givne serie er meget regelmæssige, både med hensyn til den tidsmæssige placering og udsvingenes numeriske størrelse.

…men den har også mangler

Ydermere skal det understreges, at metoden har klare mangler; den giver ingen information om seriens absolutte niveau ej heller om udviklingen igennem året. Hvis man fx har beregnet en stigning i arbejdsløsheden på x % fra juli år t-1 til juli år t, kan der være tale om forskellige situationer:

a. Arbejdsløsheden er inde i en jævnt stigende langtidsbevægelse med x % om året

b. Der har været en stigning i de første måneder, men den er nu ophørt og måske vendt til en nedadgående bevægelse

c. Arbejdsløsheden har i de første måneder været konstant eller måske endog faldende, men nu er der indtrådt en kraftig acceleration i de senere måneder.

0.2.2 Mere avancerede metoder

Der anvendes i reglen langt mere sofistikerede metoder end den ovenfor beskrevne, når der sæsonkorrigeres.

X-11 familien eller TRAMO/SEATS anvendes

Mange statistikbureauer verdenen over benytter en eller anden variant af den såkaldte X-11 familie, andre er skiftet over til den modelbaserede indfaldsvinkel som anvendes i programpakken TRAMO/SEATS.

Demetra

Eurostat har taget initiativ til udviklingen af en brugervenlig windowsbaseret overflade til både X-12 og TRAMO/SEATS, denne brugerflade hedder Demetra. Både X-11 familien, TRAMO/SEATS og Demetra vil blive beskrevet senere.

1. Definitioner og begreber

I dette kapitel gives en introduktion til de basale begreber der benyttes i sæsonkorrigeringssammenhæng.

1.1 Tidsrække

Hvad er en tidsrække?

En tidsrække eller en tidsserie er observationer over tid af den samme størrelse. Man observerer en given størrelse på bestemte tidspunkter med det samme interval imellem observationerne.

Eksempel 1.1.1 tidsrækker

Bruttonationalproduktet (BNP-serien) på kvartaler

Forbrugerprisindekset på måneder

Arbejdsløsheden på måneder

1.2 Dekomposition

Tidsserien dekomponeres

Lad den observerede tidsserie være X_t . En tidsserie dekomponeres, når man er interesseret i at eliminere de sæsonmæssige variationer.

Dekomponering i ikke observerede komponenter

Når en tidsserie dekomponeres, opsplittes den observerede serie i fire ikke-observerede delkomponenter. Man antager, at der er en bestemt type relation mellem disse fire ikke observerede komponenter. Typisk vil man dekomponere efter en såkaldt multiplikativ model (M) eller en såkaldt additiv model (A)[2].

[2] Dette er standardmodellerne, men er langt fra dækkende over alle tidsserier. Man kunne sagtens forestille sig alternative modeller som var mere passende for en given tidsserie.

…men den har også mangler

Ydermere skal det understreges, at metoden har klare mangler; den giver ingen information om seriens absolutte niveau ej heller om udviklingen igennem året. Hvis man fx har beregnet en stigning i arbejdsløsheden på x % fra juli år t-1 til juli år t, kan der være tale om forskellige situationer:

a. Arbejdsløsheden er inde i en jævnt stigende langtidsbevægelse med x % om året

b. Der har været en stigning i de første måneder, men den er nu ophørt og måske vendt til en nedadgående bevægelse

c. Arbejdsløsheden har i de første måneder været konstant eller måske endog faldende, men nu er der indtrådt en kraftig acceleration i de senere måneder.

0.2.2 Mere avancerede metoder

Der anvendes i reglen langt mere sofistikerede metoder end den ovenfor beskrevne, når der sæsonkorrigeres.

X-11 familien eller TRAMO/SEATS anvendes

Mange statistikbureauer verdenen over benytter en eller anden variant af den såkaldte X-11 familie, andre er skiftet over til den modelbaserede indfaldsvinkel som anvendes i programpakken TRAMO/SEATS.

Demetra

Eurostat har taget initiativ til udviklingen af en brugervenlig windowsbaseret overflade til både X-12 og TRAMO/SEATS, denne brugerflade hedder Demetra. Både X-11 familien, TRAMO/SEATS og Demetra vil blive beskrevet senere.

1. Definitioner og begreber

I dette kapitel gives en introduktion til de basale begreber der benyttes i sæsonkorrigeringssammenhæng.

1.1 Tidsrække

Hvad er en tidsrække?

En tidsrække eller en tidsserie er observationer over tid af den samme størrelse. Man observerer en given størrelse på bestemte tidspunkter med det samme interval imellem observationerne.

Eksempel 1.1.1 tidsrækker

Bruttonationalproduktet (BNP-serien) på kvartaler

Forbrugerprisindekset på måneder

Arbejdsløsheden på måneder

1.2 Dekomposition

Tidsserien dekomponeres

Lad den observerede tidsserie være X_t . En tidsserie dekomponeres, når man er interesseret i at eliminere de sæsonmæssige variationer.

Dekomponering i ikke observerede komponenter

Når en tidsserie dekomponeres, opsplittes den observerede serie i fire ikke-observerede delkomponenter. Man antager, at der er en bestemt type relation mellem disse fire ikke observerede komponenter. Typisk vil man dekomponere efter en såkaldt multiplikativ model (M) eller en såkaldt additiv model (A)[2].

---

[2] Dette er standardmodellerne, men er langt fra dækkende over alle tidsserier. Man kunne sagtens forestille sig alternative modeller som var mere passende for en given tidsserie.

Modeldefinitioner

Den multiplikative model er defineret ved

$$X_t = T_t C_t S_t I_t \tag{M}$$

og den additive model er defineret ved:

$$X_t = T_t + C_t + S_t + I_t \tag{A}$$

hvor T er trenden, C er cyklen, S er sæsonkomponenten og I den irregulære komponent, som defineret i afsnit 0.1.

Ikke klart definerede elementer i dekompositionen

Dekomponeringen af en given tidsserie i de forskellige komponenter er en dekomponering af en eneste observeret serie i fire ikke-observerede serier. Derfor findes der ikke nogen former for rettesnor med hensyn til teoretisk basis, hverken når det gælder valget af dekompositionsform eller definition af begreberne sæsonkomponent, trendkomponent eller konjunkturcykelkomponenten.

1.2.1 Eksempel. En additiv model[3]

Et eksempel på en additiv komponeret sæsonserie konstrueres nedenfor. Dette gøres ved at opskrive de tre (der ses bort fra den irregulære komponent) delkomponenter, der tilsammen udgør tidsserien. Det antages, at serien observeres kvartalsvist over en tre årig periode, og de tre led der tilsammen danner tidsserien kan defineres ved:

Tabel. 1.2.1.1

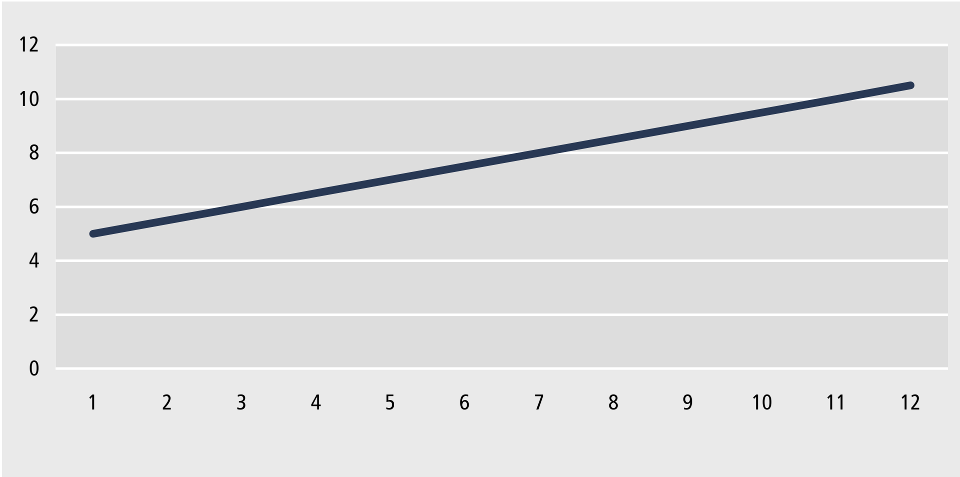
Komponentserier i en additiv model

År	Kvartal	Trend T	Konjunktur C	Sæson S
1	1	5,0	0,0	0,0
	2	5,5	0,1	0,4
	3	6,0	0,3	0,2
	4	6,5	0,4	-0,6
2	1	7,0	0,5	0,0
	2	7,5	0,4	0,4
	3	8,0	0,2	0,2
	4	8,5	-0,1	-0,6
3	1	9,0	-0,3	0,0
	2	9,5	-0,5	0,4
	3	10,0	-0,5	0,2
	4	10,5	-0,3	-0,6

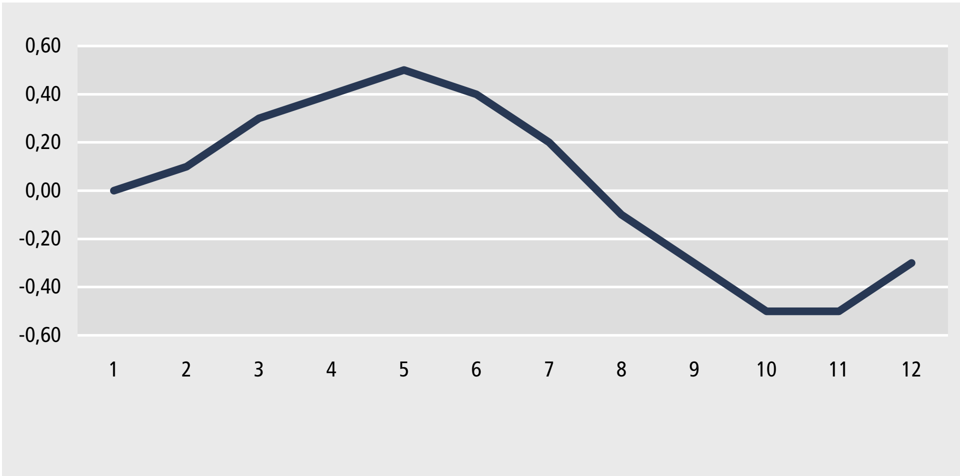
[3] Taleksemplet er taget fra Praktisk Statistik for samfundsvidenskaberne af Lars Thygesen mfl. 4. udgave 1995.

Hvis disse tre faktorer skal afbildes grafisk, ser det således ud:

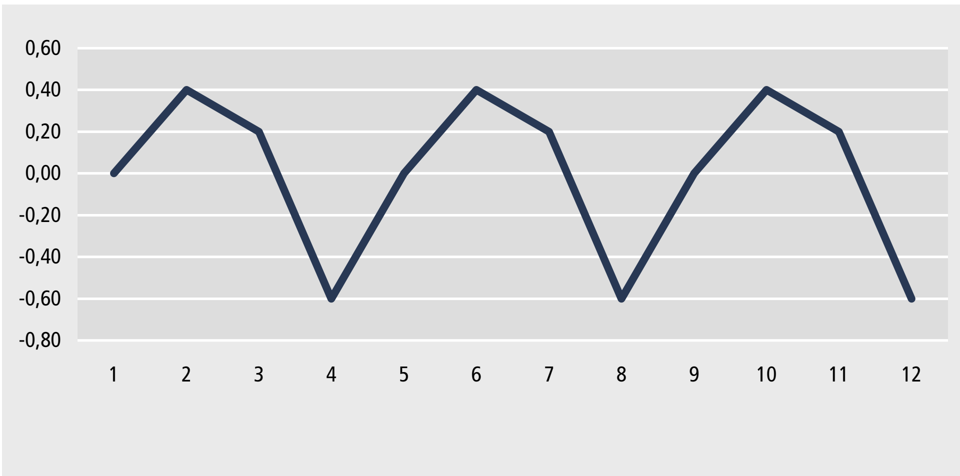
Figur 1.2.1.1 **Trend**



Figur 1.2.1.2 **Konjunktur**



Figur 1.2.1.3 **Sæson**



Disse tre led danner tilsammen selve tidsserien, som er det eneste vi observerer i virkelighedens verden.

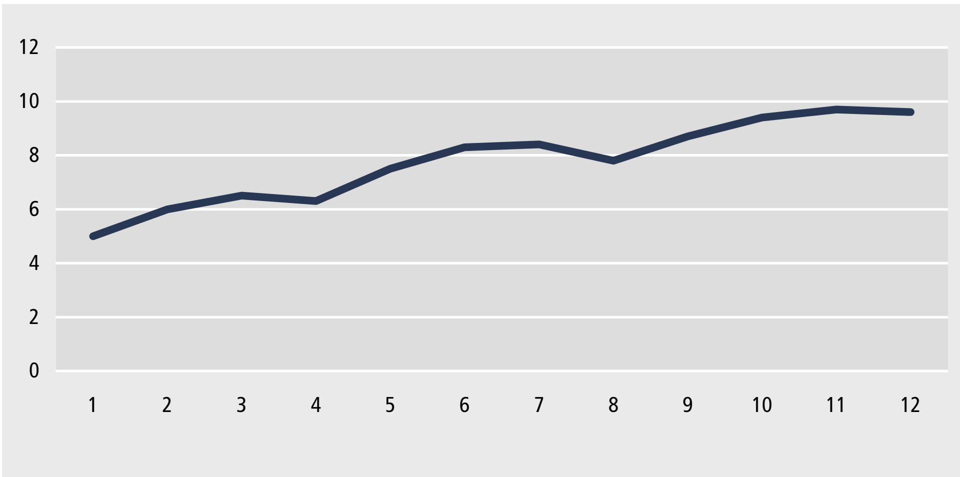
Selve tidsserien X er dannet som en sum af ovenstående tre komponenter.

Tabel 1.2.1.2. **Tidsrække for en additiv model**

År	Kvartal	Tidsrække
		X=T+C+S
1	1	5
	2	6
	3	6,5
	4	6,3
2	1	7,5
	2	8,3
	3	8,4
	4	7,8
3	1	8,7
	2	9,4
	3	9,7
	4	9,6

Hvis tidsrækken skal afbildes i en figur, ser den således ud:

Figur 1.2.1.4 **Tidsrække**



1.2.2 Eksempel. En multiplikativ model[4]

Man kan, svarende til det additive eksempel, konstruere et eksempel på en multiplikativ komponeret sæsonserie.

Dette gøres på samme måde som ovenfor, nemlig ved at opskrive de tre (der ses bort fra den irregulære komponent) delkomponenter, der tilsammen udgør tidsserien.

Serien observeres kvartalsvist over en tre årlig periode, og de tre faktorer, der tilsammen danner tidsserien, kan defineres ved:

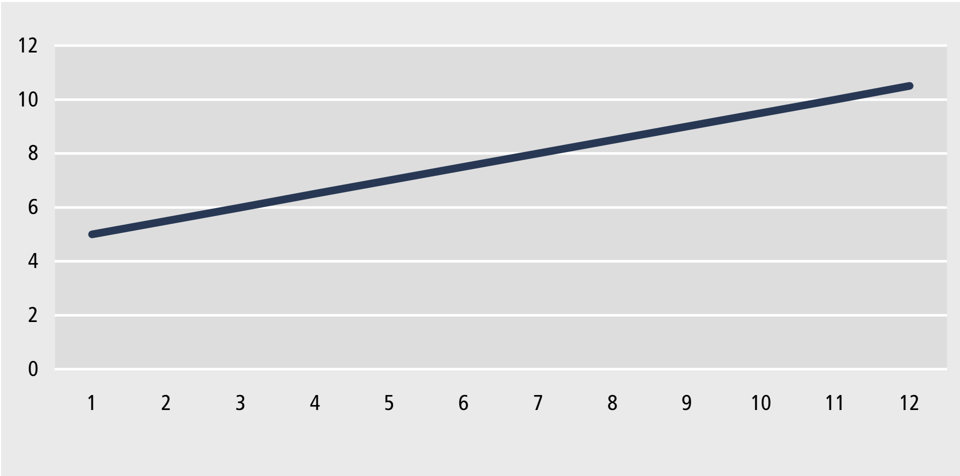
[4] [4] Taleksemplet er taget fra Praktisk Statistik for samfundsvidenskaberne af Lars Thygesen mfl. 4. udgave 1995.

Tabel. 1.2.2.1 **Komponentserier i en multiplikativ model**

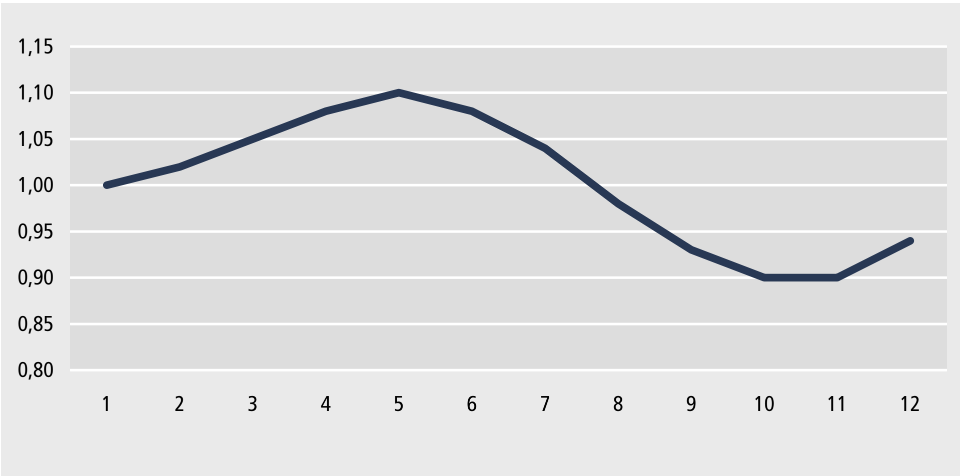
År	Kvartal	Trend T	Konjunktur C	Sæson S
1	1	5,0	1,00	1,00
	2	5,5	1,02	1,08
	3	6,0	1,05	1,04
	4	6,5	1,08	0,88
2	1	7,0	1,10	1,00
	2	7,5	1,08	1,08
	3	8,0	1,04	1,04
	4	8,5	0,98	0,88
3	1	9,0	0,93	1,00
	2	9,5	0,90	1,08
	3	10,0	0,90	1,04
	4	10,5	0,94	0,88

Hvis disse tre faktorer skal afbildes grafisk, ser det således ud:

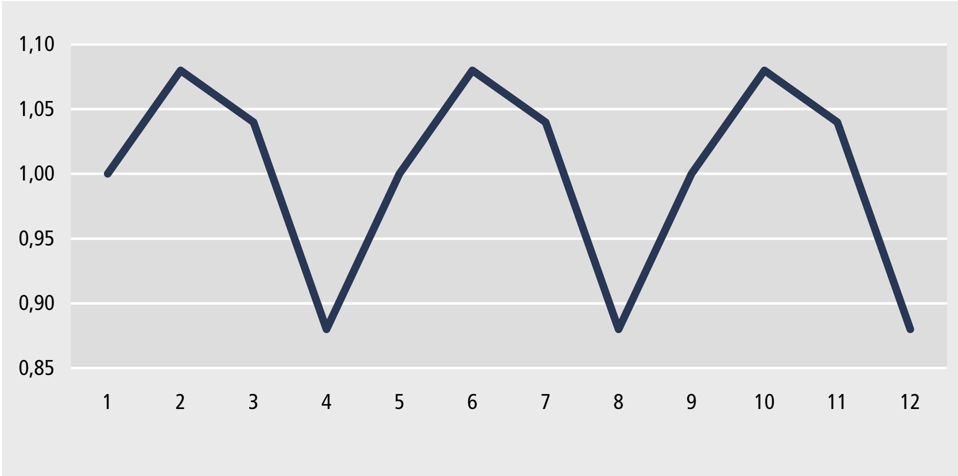
Figur 1.2.2.1 **Trenden**



Figur1.2.2.2 **Konjunkturcyklen**



Figur 1.2.2.3 **Sæsonfaktoren**



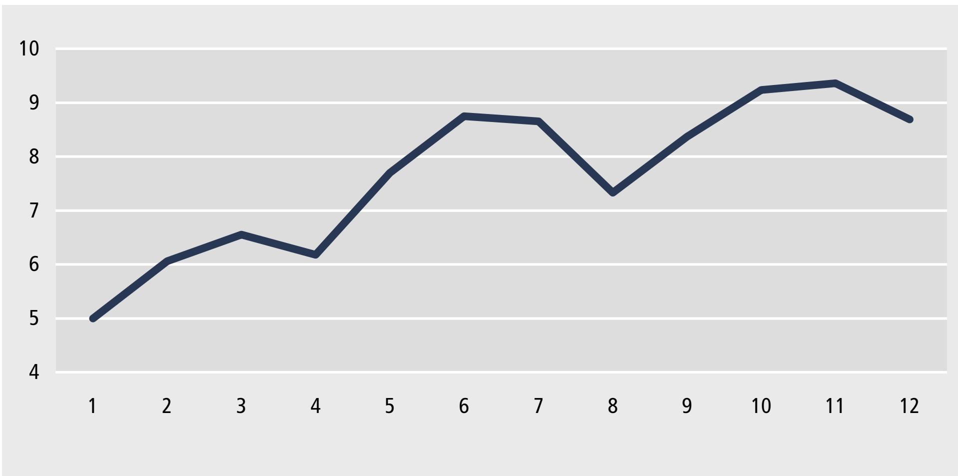
Disse tre faktorer danner tilsammen selve tidsserien, som jo er det eneste vi observerer i virkelighedens verden. Selve tidsserien Y er dannet som et produkt af ovenstående tre komponenter.

Tabel. 1.2.2.2 **Tidsserie fra en multiplikativ model**

År	Kvartal	Tidsrække
		TCS
1	1	5,00
	2	6,06
	3	6,55
	4	6,18
2	1	7,70
	2	8,75
	3	8,65
	4	7,33
3	1	8,37
	2	9,23
	3	9,36
	4	8,69

Hvis tidsrækken skal afbildes i en figur, ser den således ud:

Figur 1.2.2.4 **Tidsrække**



1.2.3 Forskellen på den multiplikative og additive model

Nu er der beskrevet to forskellige komponentmodeller, nemlig den additive model og den multiplikative model. Det er nødvendigt at afklare, hvilken af de to modeller der, synes at beskrive data bedst. Nogle helt generelle overvejelser om seriens struktur er:

Den additive model

I den additive model er sæsonvariationen i den originale serie uafhængig af seriens absolutte niveau, men er tilnærmelsesvis af samme omfang hvert år.

Den multiplikative model

I den multiplikative model har sæsonvariationen tilnærmelsesvis samme relative størrelse år for år, således at sæsonvariationen udgør en bestemt fast procentdel af tidsseriens niveau. Sæsonvariationen har større udsvingningsgrad jo højere seriens niveau er.

1.2.4 Trenden og konjunkturbevægelsen samles

Sondringen imellem de to komponenter trenden og konjunkturbevægelsen er ikke ligetil.

Hvad menes der med trenden?

Trenden er defineret som en jævn langtidsbevægelse. I denne noget løse definition er det nødvendigt at konkretisere, hvad der menes med lang tid.

Trend eller konjunkturcykel?

Hvis vi nu eksempelvis i en tidsrække har konstateret nedgang gennem flere år, hvad er det så egentlig, vi har med at gøre? Er det en aftagende trend, eller er mønstret blot en del af en konjunkturbevægelse af længere varighed?

Komponenterne samles!

Oftest vil man ved analysen af komponentmodellerne behandle disse to komponenter samlet, man kalder den samlede komponent for trend-cyklen.

1.3 Typer af sæsonvariation

Sæsonmønstret ændrer sig mht...

I økonomiske tidsserier, der dækker lange perioder, kan sæsonmønstret ændre sig med hensyn til placeringen indenfor året.

...placeringen

Man vil typisk observere ændringer i sæsonmønstrets placering, når der forekommer institutionelle ændringer såsom ændrede indbetalingsfrister til eller fra den offentlige sektor eller ved ændrede skattefradragsregler.

Ændringer i sæsonmønstrets placering vil også kunne skyldes de klimatiske faktorer. Både vinteren og sommer er herhjemme umulige at knytte til faste måneder. Denne vejrmæssige ustabilitet giver i sig selv et glidende sæsonmønster.

1.4 Glidende gennemsnit

Centrerede glidende gennemsnit

Sæsonkorrigeringen i programmer af den såkaldte X-familie er alle baseret på glidende gennemsnit. Glidende gennemsnit har i sig selv en sæsonudjævnende effekt. Glidende gennemsnit er gennemsnit beregnet med en bestemt observation som udgangspunkt. Typisk vil man i sæsonkorrigeringssammenhænge benytte centrerede glidende gennemsnit.

Eksempler

Et centreret glidende gennemsnit af længde 5, centreret omkring den 7'ende observation, har formen

$$\frac{1}{5}(x_5 + x_6 + x_7 + x_9 + x_{10})$$

Et centreret glidende gennemsnit af længde 6, centreret omkring den 7'ende observation, har formen

$$\frac{1}{6}(\frac{1}{2}\cdot x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + \frac{1}{2}\cdot x_{10})$$

2. Krav til sæsonkorrigeringen

Sæson er ikke et teoretisk defineret begreb

Da begrebet sæson ikke er teoretisk veldefineret, kan man ikke formulere objektive teoretiske krav til den optimale sæsonkorrigeringsmetode.

Krav til optimal sæsonkorrigering må derfor aftestes i praksis

Ethvert metodeforslag må derfor aftestes i praksis, hvorefter det endelige metodevalg delvist må baseres på mere subjektive overvejelser, herunder om resultatet af sæsonkorrigeringen virker troværdig, baseret på andet kendskab til seriens forløb.

Intern arbejdsgruppe-rapport

I 'Rapport vedr. sæsonkorrigering i Danmarks Statistik' udarbejdet af Arbejdsgruppen for fastlæggelse af retningslinier vedr. sæsonkorrigering i Danmarks Statistik i 2001, diskuteres en række krav. Rapportens konklusioner er i det følgende i kort form angivet for hvert enkelt beskrevet emne.

Denne arbejdsgruppe vil fra nu af blot blive betegnet arbejdsgruppen.

2.1 Forhåndskorrektioner

Det kan være nødvendigt at forhåndskorrigere en serie før man foretager den egentlige sæsonkorrigering.

Sæsonkorrigering fjerner den regelmæssige sæson

Når man sæsonkorrigerer, er det for at fjerne det regelmæssige sæsonmønster fra en serie. For at kunne gøre dette skal man først bestemme, hvad det regelmæssige sæsonmønster er. Det kan ikke gøres, hvis der i serien findes observationer, der vil udviske det virkelige sæsonmønster.

Man skal fjerne forstyrrende faktorer

Før en serie sæsonkorrigeres bør det derfor overvejes, om der i serien er et atypisk mønster som der skal korrigeres for på forhånd. Forhåndskorrektioner kan enten være permanente eller midlertidige.

Permanente forhåndskorrektioner

En permanent forhåndskorrektion er en korrektion, der foretages før sæsonkorrigeringen påbegyndes, hvorefter den forhåndskorrigerede serie anvendes som den originale input-serie i X-11 ARIMA. Når der i en serie er foretaget permanente forhåndskorrektioner, er den såkaldte sæsonkorrigerede serie altså en serie, som både er korrigeret for normale sæsonsvingninger, og den 'kendte støj', der bortrenses ved forhåndskorrektionen.

Midlertidige forhåndskorrektioner

En midlertidig forhåndskorrektion fjerner ekstreme værdier, som kan forstyrre estimationen af det normale sæsonmønster, men selve forhåndskorrektionen ophæves i den endelige sæsonkorrigerede serie. Denne vil derfor vise de ekstreme udsving, der var i den originale serie.

2.1.1 Permanente forhåndskorrektioner

Permanente forhåndskorrektioner…

Permanente forhåndskorrektioner forekommer i en række tilfælde.

…sker ved statistikbrud,…

Ved statistikbrud, hvor en talserie som tidsserie betragtet, ikke hænger sammen, er der behov for en permanent forhåndskorrektion. Dette sker fx for indeksserier, når basisperioden ændres.

Et centreret glidende gennemsnit af længde 6, centreret omkring den 7'ende observation, har formen

$$\frac{1}{6}(\frac{1}{2} \cdot x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + \frac{1}{2} \cdot x_{10})$$

2. Krav til sæsonkorrigeringen

Sæson er ikke et teoretisk defineret begreb

Da begrebet sæson ikke er teoretisk veldefineret, kan man ikke formulere objektive teoretiske krav til den optimale sæsonkorrigeringsmetode.

Krav til optimal sæsonkorrigering må derfor aftestes i praksis

Ethvert metodeforslag må derfor aftestes i praksis, hvorefter det endelige metodevalg delvist må baseres på mere subjektive overvejelser, herunder om resultatet af sæsonkorrigeringen virker troværdig, baseret på andet kendskab til seriens forløb.

Intern arbejdsgrupperapport

I 'Rapport vedr. sæsonkorrigering i Danmarks Statistik' udarbejdet af Arbejdsgruppen for fastlæggelse af retningslinier vedr. sæsonkorrigering i Danmarks Statistik i 2001, diskuteres en række krav. Rapportens konklusioner er i det følgende i kort form angivet for hvert enkelt beskrevet emne.

Denne arbejdsgruppe vil fra nu af blot blive betegnet arbejdsgruppen.

2.1 Forhåndskorrektioner

Det kan være nødvendigt at forhåndskorrigere en serie før man foretager den egentlige sæsonkorrigering.

Sæsonkorrigering fjerner den regelmæssige sæson

Når man sæsonkorrigerer, er det for at fjerne det regelmæssige sæsonmønster fra en serie. For at kunne gøre dette skal man først bestemme, hvad det regelmæssige sæsonmønster er. Det kan ikke gøres, hvis der i serien findes observationer, der vil udviske det virkelige sæsonmønster.

Man skal fjerne forstyrrende faktorer

Før en serie sæsonkorrigeres bør det derfor overvejes, om der i serien er et atypisk mønster som der skal korrigeres for på forhånd. Forhåndskorrektioner kan enten være permanente eller midlertidige.

Permanente forhåndskorrektioner

En permanent forhåndskorrektion er en korrektion, der foretages før sæsonkorrigeringen påbegyndes, hvorefter den forhåndskorrigerede serie anvendes som den originale input-serie i X-11 ARIMA. Når der i en serie er foretaget permanente forhåndskorrektioner, er den såkaldte sæsonkorrigerede serie altså en serie, som både er korrigeret for normale sæsonsvingninger, og den 'kendte støj', der bortrenses ved forhåndskorrektionen.

Midlertidige forhåndskorrektioner

En midlertidig forhåndskorrektion fjerner ekstreme værdier, som kan forstyrre estimationen af det normale sæsonmønster, men selve forhåndskorrektionen ophæves i den endelige sæsonkorrigerede serie. Denne vil derfor vise de ekstreme udsving, der var i den originale serie.

2.1.1 Permanente forhåndskorrektioner

Permanente forhåndskorrektioner…

Permanente forhåndskorrektioner forekommer i en række tilfælde.

…sker ved statistikbrud,…

Ved statistikbrud, hvor en talserie som tidsserie betragtet, ikke hænger sammen, er der behov for en permanent forhåndskorrektion. Dette sker fx for indeksserier, når basisperioden ændres.

...ved handelsdags-korrektioner...

Både antallet af de enkelte ugedage og det samlede antal dage varierer fra måned til måned, og det vil ofte påvirke forløbet af en månedsserie.

I produktionsstatistik er det typisk antallet af arbejdsdage i de enkelte måneder, som påvirker seriens forløb, imens omsætningsstatistikken er følsom overfor variationer i både antallet af handelsdage og ugedagsfordelingen.

I alle nyere sæsonkorrigeringsprogrammer er der indbygget en regressionsmodel, der automatisk kan estimere korrektion for handelsdagsvariation (handelsdagskorrektion).

...eller ved påskekorrektion.

På en række serier har påskens månedsplacering i det enkelte år en afgørende betydning, bl.a. på en serie som konsumet af påskeæg. I alle nyere sæsonkorrigeringsprogrammer kan der automatisk korrigeres for påskens beliggenhed via estimation af en såkaldt påskemodel.

2.1.2 Midlertidige forhåndskorrektioner

For visse serier vil strejker, naturkatastrofer og lignende ekstreme atypiske hændelser ødelægge mulighederne for estimation af det normale sæsonmønster. I disse tilfælde foretages midlertidige forhåndskorrektioner.

Sæsonkomponenten må ikke påvirkes

Den sæsonkorrigerede serie skal naturligvis vise det atypiske forløb, men dette må ikke påvirke estimationen af selve sæsonkomponenten.

2.2 Den irregulære komponent

En sæsonkorrigeret tidsserie indeholder både trend-cyklen og den irregulære komponent. Som tidligere nævnt ligger det eksplicit i antagelsen om en komponentmodel, at restleddet (den irregulære komponent) er uden spor af systematik.

Den irregulære komponent må ikke blive for stor

En sæsonkorrigering bliver problematisk, når svingningerne i den irregulære komponent bliver større end svingningerne i sæsonkomponenten.
I en sådan situation er det ikke nemt for sæsonkorrigeringsprogrammet at genkende et stabilt sæsonmønster, eftersom svingningerne i serien overskygges af udsvingene i den irregulære komponent.

2.3 Restsæson

Af og til kan det forekomme, at der i en sæsonkorrigeret serie er systematiske svingninger indenfor et år. Der er altså sæsonbevægelse i den sæsonkorrigerede serie, dette fænomen betegnes restsæson. Det kan fx skyldes, at sæsonbevægelsens mønster gradvis ændrer sig over årene, hvilket igen kunne tænkes at skyldes teknologiske ændringer i samfundet.

Restsæson er et problem…

Dette forhold er problematisk, idet hensigten med sæsonkorrigeringen netop var at fjerne sæsonbevægelsen i den oprindelige serie.

…som skal fjernes

– Man kan i praksis fjerner denne sæsonbevægelse ved at justere på en eller flere af de valgmuligheder der ligger i sæsonkorrigeringsprogrammet, eksempelvis kunne en justering af længden af de glidende gennemsnit som der anvendes i sæsonkorrigeringen fjerne restsæsonen

2.4 Direkte eller indirekte sæsonkorrigering?

Direkte sæsonkorrigering og indirekte sæsonkorrigering diskuteres, når man har en serie, der dannes som en sum eller et gennemsnit af en række af delserier.
Forskellen på de to begreber, ligger i, om aggregatserien eller delserien skal sæsonkorrigeres.

Direkte sæsonkorrigering

Direkte sæsonkorrigering af aggregatserien betyder, at totalen sæsonkorrigeres, dvs. først summeres og dernæst sæsonkorrigeres.

Indirekte sæsonkorrigering

Indirekte sæsonkorrigering af aggregatserien betyder, at de enkelte delserier sæsonkorrigeres, og herefter dannes totalen som summen af de sæsonkorrigerede delserier, dvs. først sæsonkorrigeres og dernæst summeres.

Hvilken metode skal vi anvende?

I litteraturen findes der ingen entydig vejledning med hensyn til, hvilken metode der bør anvendes metodevalget afhænger specifikt af de serier man har med at gøre.

Danmarks Statistik sæsonkorrigerer indirekte

I Danmarks Statistik foretrækkes indirekte korrigering medmindre specielle forhold gør sig gældende.

Indirekte sæsonkorrigering er intuitivt mest forståeligt

Intuitivt set virker indirekte sæsonkorrigering mest fornuftigt. Sæsonmønstre i de enkelte delserier kunne tænkes at udligne hinanden eller føre til et aldeles uklart mønster i en aggregatserie

...men dog ikke altid det der bør bruges

I praksis er der dog det problem, at nogle aggregater kan dannes som summer af forskellige sæt af del-tidsserier: Fx kunne den samlede eksport dannes som summen af eksporttal for alle lande, eller summen af eksporttal for alle varer; indirekte sæsonkorrigering ud fra varer hhv. lande vil sandsynligvis ikke give samme resultat.

2.5 Årsopregning

Når en serie er sæsonkorrigeret gælder der sjældent, at summen (eller gennemsnittet) over et kalenderår af den sæsonkorrigerede serie er lig summen (eller gennemsnittet) over kalenderåret af originalserien. Man kan vælge at fjerne denne forskel ved såkaldt årsopregning.

Danmarks Statistik årsopregner

Danmarks Statistik publicerer årsopregnede sæsonkorrigerede tal for serier der er længere end 5 år.

2.6. Usikkerhed ved sæsonkorrigering

Resultatet af en sæsonkorrigering ændrer sig når originaltidsserien bliver længere, når man får mere information. Dette skyldes at der i X-11ARIMA og i X-12 beregnes en række successivt centrerede glidende gennemsnit af trends og sæsonfaktorer.

Centrerede glidende gennemsnit betyder gæt på fremtiden…

Da man altid beregner centrerede gennemsnit, også for de nyeste observationer, skal man forsøge at spå om fremtiden for at sæsonkorrigere de seneste tal; man skal typisk gætte på værdierne for de kommende 6 måneder eller 2 kvartaler. Dette gøres i X-11ARIMA og X-12 ved at estimere en såkaldt ARIMA-model.

…hvilket medfører revisioner i de seneste sæsonkorrigerede tal

De største revisioner af de sæsonkorrigerede tal ses således også for de seneste observationer. De eneste tilfælde hvor der ikke forekommer revisioner er der, hvor modellen har estimeret udviklingen i fremtiden korrekt.

2.7 Revisioner

Da der er usikkerheder på sæsonkorrigerede tal og da der således kan være behov for væsentlige revisioner, er det nødvendigt at fastlægge en rimelig revisionspraksis.

Revisioner i sæsonkorrigerede tal forekommer tilbage i tiden…

Faktisk vil der typisk forekommer revisioner i de sæsonkorrigerede tal tilbage i tiden. Dette skyldes at man anvender symmetriske glidende gennemsnit i X-11ARIMA og X-12. Således vil sæsonkorrigerede tal for tidligere perioder blive revideret ved forekomsten af nye observationer.

…forklarligt, men umiddelbart uforståeligt

Dét, at de sæsonkorrigerede tal for en given periode ændrer sig selvom de ikke-korrigerede tal er uændrede, kan forekomme uforståeligt for den teknisk ukyndige bruger.

2.7.1 Optimal længde af revisionsperioden

perioden bør være kort…

På den ene side er det klart, at den periode bagud i tid, hvor de sæsonkorrigerede tal revideres, bør være kort for ikke at genere brugerne.

…men ikke for kort

På den anden side må den ikke må være så kort, at udviklingen i den sæsonkorrigerede serie er utroværdig.

Fornyet sæsonkorrigering bør kun medføre beskedne revisioner

Disse to modsatrettede krav fører ikke til et entydigt svar på hvor lang revisionsperioden bør være. En praktisk løsning på overvejelserne er, at revisionsperioden bør være af en sådan længde, at fornyet sæsonkorrigering af en serie kun medfører beskedne ændringer i de tal som ligger udenfor revisionsperioden.

Generel revisionspraksis

Den generelle revisionspraksis er den, at alle serier skal som minimum revideres 13 måneder/5 kvartaler tilbage i tid, og som hovedregel bør der maksimalt revideres 3-4 år tilbage i tid.

Minimumskravet bør overholdes

Minimumskravet bør under alle omstændigheder altid overholdes, mens maksimumskravet kan fraviges, hvis eksempelvis revisionerne længere tilbage i tiden (end revisionsperioden!) er store.

Tallene gøres endelige med årlige intervaller

Sæsonkorrigerede tal erklæres endelige med årlige intervaller.

2.8 Præsentation af sæsonkorrigerede tal

Figurer bør bruges i præsentationen…

Sæsonkorrigerede tal er dannet således, at en sammenligning fra periode til periode har mening. Dermed er sæsonkorrigerede tal meget velegnede til at blive præsenteret i en figurmæssig sammenhæng.

…for sæsonkorrigering er en kompliceret størrelse

Sæsonkorrigering af en tidsserie er dog ikke nogen simpel transformation, som for eksempel udregning af vækstrater, så når man fortolker på de sæsonkorrigerede tal, bør man have dette for øje og gøre sig umage med at give en grundig præsentation.

2.8.1 Udviklingen i de sæsonkorrigerede tal

Der skal kun kommenteres på den periodevise udvikling…

Når der kommenteres på sæsonkorrigerede tal, skal der udelukkende kommenteres på udviklingen i forhold til foregående periode (eller flere forudgående perioder under ét, fx 3-måneders glidende gennemsnit).

…og ikke på den årlige udvikling

At kommentere på udviklingen i de sæsonkorrigerede tal set i forhold til samme periode foregående år, er meningsløst. For det første fordi sæsonkorrigerede tal netop beregnes for at kunne sammenligne udviklingen fra en periode til den næste. For det andet fordi beregningen af årlige vækstrater i sig selv er en primitiv form for sæsonkorrigering.

2.8.2 Tolkning af de sæsonkorrigerede tal

Sæsonkorrigering fjerner kun den normale sæsonbevægelse

Når man tolker på de sæsonkorrigerede tal, er det vigtigt at huske på at sæsonkorrigeringen kun fjerner den normale sæsonbevægelse. Hvis salget af isvafler, sodavand eller øl er særligt stort i en sommermåned, fordi det var ekstremt varmt, så skal det give sig udslag i den sæsonkorrigerede serie.

Sæsonkorrigering skal ikke udglatte mest muligt

Formålet med sæsonkorrigering er netop at fjerne den støj, som skyldes normale sæsonbevægelser, ikke at udglatte serien mest muligt.

2.8.3 Månedsserier

Når man kommenterer på udviklingen fra en periode til den næste, skal man være forsigtig med at drage håndfaste konklusioner.

Skift i udviklingen kan skyldes tilfældigheder

Et skift i udviklingen kan blot skyldes en tilfældighed og behøver således ikke at være et udtryk for, at udviklingen er ved at vende. Denne problemstilling er særligt udtalt for månedsserier, og specielt gælder problematikken for den nyeste sæsonkorrigerede observation.

Længere perioder

I stedet bør man knytte kommentarerne til sammenligninger baseret på lidt længere perioder, for eksempel de seneste tre måneder set i forhold til de foregående tre måneder.

Viden om fortiden er nødvendig

Generelt kan man dog ikke fastsætte retningslinier for hvor lang en periode, der skal anvendes ved sammenligninger. Det skyldes, at der skal tages hensyn både til seriens variabilitet og til usikkerheden ved sæsonkorrigering af den nyeste observation.

Ekstraordinære hændelser

I nogle tilfælde skyldes drastiske ændringer i en sæsonkorrigeret serie fra måned til måned særlige forhold, som er kendte, fx strejker, usædvanligt godt eller dårligt vejr, eller politiske indgreb.

2.8.4 Målet med konjunkturanalysen

Fastlæggelse af vendepunkter…

Det endelige formål med analyse af konjunkturserier er typisk at kunne identificere konjunkturbevægelsens retning, fordi man specielt er interesseret i at fastlægge vendepunkter.

…hertil er sæsonkorrigering et hjælpeværktøj

Sæsonkorrigering er første skridt på vejen, men da den sæsonkorrigerede serie også indeholder en irregulær faktor, kan konjunkturbevægelsen ikke umiddelbart identificeres ved analyse af den sæsonkorrigerede serie.

Det er svært at spå…

Der findes ingen kendt teknik, som kan løse problemet, hvilket naturligvis hænger sammen med at det er temmelig svært at forudsige den økonomiske udvikling.

…men sæsonkorrigerede tal er en slags indikatorer

Analysen baseret på de sæsonkorrigerede tal er det nærmeste, man kommer på en løsning til problemet i praksis, når man vel og mærke kombinerer sin analyse med et dybdegående kendskab til de forhold, som påvirker tidsserien og den metodemæssige del af sæsonkorrigeringen.

Sæsonkorrigering erstatter ikke tankevirksomhed

Sæsonkorrigering er et hjælpeværktøj, som kan lede analytikeren bedre på vej end ved blot at betragte ikke-sæsonkorrigerede tal. Metoden er ingen automatpilot som gør, at man selv slipper for at tænke.

2.9 Manuel sæsonkorrigering

Til illustration af, hvorledes sæsonkorrigering foretages ved hjælp af glidende gennemsnitsmetoder som X-11ARIMA og X-12, vil der her blive gennemregnet sæsonkorrigeringen af de to serier, som blev defineret i afsnittene 1.2.1 og 1.2.2 på den mest simple facon. Eksemplet er hentet fra Fisker, K., K. Nørlund, L. Thygesen, J. Wedebye, R. Østerlund: *Praktisk statistik for samfundsvidenskaberne*, Akademisk Forlag 1995.

2.9.1 Sæsonkorrigering af den additive serie

I den additive model er X=T+C+S+I.

Sæsonkorrigeringen foregår i fem trin.

Første trin I første trin beregnes fire-kvartalers glidende gennemsnit af de oprindelige observationer.
Derved elimineres virkningen af sæsonkomponenten - forudsat at svingningerne er regelmæssige.
Den irregulære komponents positive og negative udsving forudsættes nemlig stort set at ophæve hinanden ved en gennemsnitsberegning, derfor elimineres også virkningen af den irregulære komponent.

De glidende gennemsnit G giver derfor et første skøn over trenden og konjunkturbevægelsen. G=T+C .
Da der skal dannes symmetriske glidende gennemsnit findes der ikke et skøn for de to første og de to sidste kvartaler.

Tabel 2.9.1.1 **Glidende gennemsnit i den additive model.**

År	Kvartal	Tidsrække X	Glidende gennemsnit G
1	1	5,0	…
	2	6,0	…
	3	6,5	6,26
	4	6,3	6,86
2	1	7,5	7,39
	2	8,3	7,81
	3	8,4	8,15
	4	7,8	8,44
3	1	8,7	8,74
	2	9,4	9,13
	3	9,7	…
	4	9,6	…

Eksempelvis beregnes det første fire-kvartalers glidende gennemsnit, dvs. tallet hørende til 3. Kvartal i år 1 som

$$G_{1,3} = \frac{\frac{1}{2}5{,}0 + 6{,}0 + 6{,}5 + 6{,}3 + \frac{1}{2}7{,}5}{4} = 6{,}26$$

Andet trin Andet trin af beregningerne består i, at der dannes en række foreløbige skøn over sæson + irregulær komponent, idet de glidende gennemsnit for hvert kvartal trækkes fra de originale observationer.

Herved fremkommer differenserne D=X-G.

Differenserne danner et skøn over sæsonkomponenten - dog inklusive den irregulære komponent I , idet

D=X-G=(T+C+S+I)-(T+C)=S+I

Tabel 2.9.1.2 **Skøn over sæsonkomponenten i den additive model**

År	Kvartal	Differens D=X-G
1	1	...
	2	...
	3	0,24
	4	-0,56
2	1	0,11
	2	0,49
	3	0,25
	4	-0,64
3	1	-0,04
	2	0,27
	3	...
	4	...

Tredje trin Tredje trin af beregningerne består i at sammenfatte de forskellige skøn (D) for samme kvartal og det sker ved at beregne gennemsnit, S', over årene for hvert enkelt kvartal.

Derved formindskes også påvirkningen af den irregulære komponent, specielt hvis observationsperioden omfatter flere år.

Tabel 2.9.1.3 **Første skøn over sæsonfaktorerne**

	D=X-G			Gennemsnit
Kvartal	År 1	År 2	År 3	S'
1	...	0,11	-0,04	0,04
2	...	0,49	0,27	0,38
3	0,24	0,25	...	0,25
4	-0,56	-0,64	...	-0,60
Sum				0,07

Fjerde trin De endelige sæsontal skal summere til 0. Dette vil dog sjældent være tilfældet for de sæsontal, vi udregner i de tre foregåede trin af beregningerne.

Derfor er det nødvendigt at normere sæsontallene så denne restriktion er opfyldt. Der normeres ved en ligelig regulering af alle kvartalstallene med faktoren

$$\frac{1}{4}0{,}07 = 0{,}02\,.$$

Der gælder altså, at $S = S' - \frac{1}{4}0{,}07$.

Tabel 2.9.1.4 **Endeligt skøn over sæsonfaktorerne**

	D=X-G			Gennemsnit
Kvartal	År 1	År 2	År 3	S
1	...	0,11	-0,04	0,02
2	...	0,49	0,27	0,36
3	0,24	0,25	...	0,23
4	-0,56	-0,64	...	-0,62
Sum				-0,01

Sammenholdes nu denne estimerede sæson med originaldata i den konstruerede tidsrække, da ses det, at den beregnede sæson er identisk med den 'rigtige' sæson, når der regnes med en decimal. Det skal bemærkes, at en sådan sammenligning ikke kan foretages i 'praksis' , hvor den rigtige sæson er ukendt.

Femte trin

Når sæsontallene er beregnet foretages selve sæsonkorrigeringen ved at trække sæsontallene fra de originale observationer.

Målet med sæsonkorrigeringen er at fjerne indvirkningen af den normale sæson, hvorved man får et skøn over trenden plus konjunktur ganske vist påvirket af den irregulære komponent;

$$Y\text{-}S = T + C + S + I\text{-}S = T + C + I$$

De sæsonkorrigerede tal fås således som

Tabel 2.9.1.5 **Sæsonkorrigerede tal i en additiv model**

År	Kvartal	Observation	Sæsonfaktor	Sæsonkorrigeret serie
		X	S	X-S
1	1	5,0	0,0	5,0
	2	6,0	0,4	5,6
	3	6,5	0,2	6,3
	4	6,3	-0,6	6,9
2	1	7,5	0,0	7,5
	2	8,3	0,4	7,9
	3	8,4	0,2	8,2
	4	7,8	-0,6	8,4
3	1	8,7	0,0	8,7
	2	9,4	0,4	9,0
	3	9,7	0,2	9,5
	4	9,6	-0,6	10,2

2.9.2 Sæsonkorrigering af den multiplikative serie

I den multiplikative model er X=TCSI.

Første trin

Først dannes de symmetriske glidende gennemsnit af observationerne ved G=TC som et første skøn over trend og konjunktur.

Tabel 2.9.2.1 **Glidende gennemsnit i den multiplikative model**

År	Kvartal	Observation	4- kvartalers glidende gennemsnit G
1	1	5,0	...
	2	6,1	...
	3	6,6	6,31
	4	6,2	6,98
2	1	7,7	7,56
	2	8,7	7,96
	3	8,7	8,19
	4	7,3	8,34
3	1	8,4	8,49
	2	9,2	8,75
	3	9,4	...
	4	8,7	...

Andet trin

Dernæst fjernes T og C fra tidsrækken, men her skal man dividere, i modsætning til den additive model, dvs. man udregner

$Q = \dfrac{X}{G}$ som et skøn over sæsonen,

idet $Q = \dfrac{X}{G} = \dfrac{TCSI}{TC} = SI$

Tabel 2.9.2.2 **Skøn over sæsonkomponenten I den multiplikative model**

År	Kvartal	Kvotient Q=X/G
1	1	...
	2	...
	3	1,05
	4	0,89
2	1	1,02
	2	1,09
	3	1,06
	4	0,88
3	1	0,99
	2	1,05
	3	...
	4	...

Tredje trin De forskellige sæsontal for samme kvartal sammenfattes ganske som i den additive model ved beregning af gennemsnit over årene, S'.

Tabel 2.9.2.3 **Første skøn over sæsonfaktorerne**

	Q=X/G			Gennemsnit
Kvartal	År 1	År 2	År 3	S'
1	...	1,02	0,99	1,01
2	...	1,09	1,05	1,07
3	1,05	1,06	...	1,06
4	0,89	0,88	...	0,89
Sum				4,02

Fjerde trin Gennemsnittene S' normeres, så disse beregnede sæsonfaktorer varierer omkring 1, dvs. de skal summere til 4. Normeringen sker her ved, at hver sæsonfaktor ganges med 4 og divideres med summen af de fire foreløbige sæsonfaktorer.

$$S = \frac{4S'}{\sum S'}$$

Tabel 2.9.2.4 **Endeligt skøn over sæsonfaktorerne**

	Q=X/G			Normerede Gennemsnit
Kvartal	År 1	År 2	År 3	S
1	...	1,02	0,99	1,00
2	...	1,09	1,05	1,07
3	1,05	1,06	...	1,05
4	0,89	0,88	...	0,88
Sum				4,00

Femte trin Når sæsontallene er beregnet foretages selve sæsonkorrigeringen ved at dividere sæsontallene op i de originale observationer. X/S=TCSI/S=TCI

Tabel 2.9.2.5 **Sæsonkorrigerede tal i en multiplikativ model**

År	Kvartal	Observation	Sæsonfaktor	Sæsonkorrigeret serie
		X	S	X/S
1	1	5,0	1,00	5,00
	2	6,1	1,07	5,70
	3	6,6	1,05	6,29
	4	6,2	0,88	7,05
2	1	7,7	1,00	7,70
	2	8,7	1,07	8,13
	3	8,7	1,05	8,29
	4	7,3	0,88	8,30
3	1	8,4	1,00	8,40
	2	9,2	1,07	8,60
	3	9,4	1,05	8,95
	4	8,7	0,88	9,89

3. Statistisk analyse af tidsrækkemodeller

Statistisk analyse af tidsrækker er det samme som regression

Statistisk analyse af tidsrækkemodeller adskiller sig principielt ikke fra analyse af almindelige regressionsmodeller.

Indhold

En række af de generelle teoretiske begreber gennemgås i afsnit 3.1, i afsnit 3.2 gennemgås de autoregressive processer. I afsnit 3.3 defineres MA-modellen og ARMA-modellen, i afsnit 3.4 gennemgås ARIMA-modellen, og i afsnit 3.5 gennemgås ARIMA-sæsonmodellen. I afsnit 3.6 beskrives identifikation af ARIMA-modeller, og endelig i afsnit 3.7 gennemgås kontrol af ARIMA-modeller.

3.1 Generelle statistiske begreber

Statistisk modellering

Begrebet statistisk modellering dækker over en fastlæggelse af funktionelle sammenhænge imellem en eller flere variabler ved at bruge den information, der ligger i de tilgængelige observationer for variablerne.

Generelle kriterier for modellering er ikke mulige

Desværre er det ikke muligt at opstille generelle kriterier eller retningslinier for udvælgelse af den optimale model. Dog er der en række helt generelle retningslinier, som modellen skal opfylde for at den overhovedet giver mening.

Modellen skal beskrive data...

Modellen skal kunne beskrive variationen i data. Dette krav, som i sig selv synes temmelig overflødigt og banalt, betyder desværre ikke, at den bedste model altid har den højeste forklaringsgrad.

Det er vigtigt, at de forklarende variabler man vælger til sin analyse a priori besidder en eller anden forklaringsgrad.

...den skal være meningsfuld og...

Modellen der benyttes til at beskrive data, skal være meningsfuld i den forstand, at den afspejler virkeligheden.

...have acceptable statistiske egenskaber

Det endelige modelvalg baseres på de statistiske egenskaber ved denne. Det vil sige, at modellen skal have en høj forklaringsgrad, samt at den variation i data som ikke kan forklares ved modellen, er uden information. Selv om disse krav er enkle og ligetil, er processen med at bestemme en optimal model både kompliceret og langvarig.

Modeltyper

De statistiske modeller, der arbejdes med i praksis for at beskrive variationen i en enkelt variabel er ofte regressionsmodeller.

Tabel 2.9.2.5 **Sæsonkorrigerede tal i en multiplikativ model**

År	Kvartal	Observation	Sæsonfaktor	Sæsonkorrigeret serie
		X	S	X/S
1	1	5,0	1,00	5,00
	2	6,1	1,07	5,70
	3	6,6	1,05	6,29
	4	6,2	0,88	7,05
2	1	7,7	1,00	7,70
	2	8,7	1,07	8,13
	3	8,7	1,05	8,29
	4	7,3	0,88	8,30
3	1	8,4	1,00	8,40
	2	9,2	1,07	8,60
	3	9,4	1,05	8,95
	4	8,7	0,88	9,89

3. Statistisk analyse af tidsrækkemodeller

Statistisk analyse af tidsrækker er det samme som regression

Statistisk analyse af tidsrækkemodeller adskiller sig principielt ikke fra analyse af almindelige regressionsmodeller.

Indhold

En række af de generelle teoretiske begreber gennemgås i afsnit 3.1, i afsnit 3.2 gennemgås de autoregressive processer. I afsnit 3.3 defineres MA-modellen og ARMA-modellen, i afsnit 3.4 gennemgås ARIMA-modellen, og i afsnit 3.5 gennemgås ARIMA-sæsonmodellen. I afsnit 3.6 beskrives identifikation af ARIMA-modeller, og endelig i afsnit 3.7 gennemgås kontrol af ARIMA-modeller.

3.1 Generelle statistiske begreber

Statistisk modellering

Begrebet statistisk modellering dækker over en fastlæggelse af funktionelle sammenhænge imellem en eller flere variabler ved at bruge den information, der ligger i de tilgængelige observationer for variablerne.

Generelle kriterier for modellering er ikke mulige

Desværre er det ikke muligt at opstille generelle kriterier eller retningslinier for udvælgelse af den optimale model. Dog er der en række helt generelle retningslinier, som modellen skal opfylde for at den overhovedet giver mening.

Modellen skal beskrive data...

Modellen skal kunne beskrive variationen i data. Dette krav, som i sig selv synes temmelig overflødigt og banalt, betyder desværre ikke, at den bedste model altid har den højeste forklaringsgrad.

Det er vigtigt, at de forklarende variabler man vælger til sin analyse a priori besidder en eller anden forklaringsgrad.

...den skal være meningsfuld og...

Modellen der benyttes til at beskrive data, skal være meningsfuld i den forstand, at den afspejler virkeligheden.

...have acceptable statistiske egenskaber

Det endelige modelvalg baseres på de statistiske egenskaber ved denne. Det vil sige, at modellen skal have en høj forklaringsgrad, samt at den variation i data som ikke kan forklares ved modellen, er uden information. Selv om disse krav er enkle og ligetil, er processen med at bestemme en optimal model både kompliceret og langvarig.

Modeltyper

De statistiske modeller, der arbejdes med i praksis for at beskrive variationen i en enkelt variabel er ofte regressionsmodeller.

Den almindelige regressionsmodel

I den almindelige regressionsmodel forsøger man at beskrive variationen i en variabel (den afhængige variabel eller responsvariablen) ved en eller flere andre uafhængige variable (regressorer). En regressionsligning med en forklarende variabel har formen

$$x_t = \alpha + \beta a_t + \varepsilon_t$$

Hvor x_t er responsvariablen, og a_t er regressoren. α er skæringen med anden aksen, β er hældningen og ε_t er fejlleddene ved modellen.

3.2 Autoregressive modeller

I regressionsmodeller kan der indgå fortidige værdier af den afhængige variabel som forklarende variable. Er dette tilfældet kalder vi modellen for autoregressiv.

Den simpleste autoregressive model

Den simpleste autoregressive model er den såkaldte AR(1) model, defineret ved

$$x_t = \alpha + \beta x_{t-1} + \varepsilon_t \qquad \text{AR(1)}$$

En sådan autoregressiv model giver ikke nogen decideret forklaring på variationen i den afhængige variabel, relationen er blot en påstand af typen 'man gør noget i dag fordi man også gjorde det i går'. Dog er det klart, at dette er en ganske rimelig antagelse til beskrivelse af kortsigtet økonomisk adfærd.

En AR(p)-proces

En autoregressiv proces af orden p (for nemheds skyld antager vi at α=0) er defineret ved:

$$X_t = \gamma_1 X_{t-1} + \gamma_2 X_{t-2} + \ldots + \gamma_p X_{t-p} + \varepsilon_t \qquad \text{AR(p)}$$

Hvor ε_t som sædvanlig er normalfordelt støj der ikke påvirker hinanden over tid.

3.3 Definition af MA-modellen og ARMA-modellen

MA(q)-modellen

En anden type af model definerer den afhængige variabel som en funktion af fejlleddene, nemlig som

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \qquad \text{MA(q)}$$

hvor $\theta_1, \ldots, \theta_q$ er parametre og processen $(\varepsilon_t)_{t \in N}$ er hvid støj.
Modellen kaldes en glidende gennemsnitsmodel af orden q, eller en MA(q)-model (fra engelsk Moving Average).

ARMA(p,q)-modellen

En ARMA-model er en såkaldt blandet model, den er autoregressiv af orden p og med et glidende gennemsnit i fejlleddene af orden q. En ARMA(p,q) model med middelværdi 0 har altså følgende udseende:

$$x_t - \rho_1 x_{t-1} - \ldots - \rho_p x_{t-p} = \varepsilon_t - \alpha_1 \varepsilon_{t-1} - \ldots - \alpha_q \varepsilon_{t-q} \qquad \text{ARMA(p,q)}$$

Formuleret anderledes

Lad lag operatoren L være defineret ved $LX_t = X_{t-1}$.

Formuleret ved hjælp af lagoperatoren L, ser vi at ovenstående er det samme som

$$(1 - \rho_1 L - \ldots - \rho_p L^p) x_t = (1 - \alpha_1 L - \ldots - \alpha_q L^q) \varepsilon_t$$

3.4 Definition af ARIMA-modellen

Processen som betragtes, X_t , kan enten være stationær eller ikke-stationær[5]. Der kan lægges bånd på de indgående parametre så processen er stationær[6]. En ikke-stationær proces kaldes for en integreret proces.

Definition:
En integreret proces

Lad nu $\Delta = 1 - L$.
En proces X_t siges at være integreret af orden d, hvis $\Delta^d X_t = (1-L)^d X_t$ er stationær.

Eksempel: En stationær proces
$X_t = \varepsilon_t$, *hvor* ε_t *som sædvanlig er indbyrdes uafhængige, identiske normalfordelte*.

Eksempel: En ikke stationær proces
$X_t = \sum_{i=1}^{t} \varepsilon_i$ *er ikke stationær, den er integreret af første orden, thi*
$\Delta X_t = X_t - X_{t-1} = \sum_{i=1}^{t} \varepsilon_i - \sum_{i=1}^{t-1} \varepsilon_i = \varepsilon_t$, *som er en stationær proces.*

En ARIMA(p,d,q) model for en tidsserie $(x_t)_{t\in T}$ er en model hvor x_t er integreret af orden d, og at den herved fremkomne stationære serie $(1-L)^d x_t$ modelleres ved en ARMA(p,q) model, dvs. for tidsserien gælder følgende sammenhæng

$$(1-\rho_1 L - \ldots - \rho_p L^p)(1-L)^d x_t = (1-\alpha_1 L - \ldots - \alpha_q L^q)\varepsilon_t \quad \text{ARIMA(p,d,q)}$$

Af definitionen ses det direkte, at ARIMA-modellen er en generalisering af alle de tre modeller som er defineret ovenfor, nemlig af både AR, MA og ARMA-modellerne;

En AR(p)-model er blot en ARIMA(p,0,0)- model, idet ARIMA(p,0,0) dækker over ligningen

$$(1-\rho_1 L - \cdots - \rho_p L^p)x_t = \varepsilon_t \quad \text{, eller}$$
$$x_t = \rho_1 x_{t-1} + \cdots + \rho_p x_{t-p} + \varepsilon_t$$

En MA(q)-model er blot en ARIMA(0,0,q)-model, da ARIMA(0,0,q) dækker over ligningen

$$x_t = (1-\alpha_1 L - \cdots - \alpha_q L^q)\varepsilon_t = \varepsilon_t - \alpha_1 \varepsilon_1 - \cdots - \alpha_q \varepsilon_{t-q}$$

Endelig ses det, at en ARMA(p,q) model blot er en ARIMA(p,0,q)-model, da ARIMA(p,0,q) dækker over ligningen

$$(1-\rho_1 L - \cdots - \rho_p L^p)x_t = (1-\alpha_1 L - \cdots - \alpha_q L^q)\varepsilon_t$$

---

[5] Processer som er integreret af en negativ orden findes også. Processen $\varepsilon_t - \varepsilon_{t-1}$ bliver stationær når den summeres en gang.

[6] Se enten 'Tidsrækkeanalyse for økonomer' af Anders Milhøj for en let gennemgang, eller 'Statistisk analyse af økonometriske modeller' af Søren Johansen for en grundig gennemgang.

3.5 ARIMA-sæsonmodellen

Sæsonmodellen tager højde for den systematiske sæsonvariation…

Den almindelige ARIMA-model, som er defineret ovenfor, kan ikke tage højde for den systematiske sæsonvariation, og derfor er det nødvendigt at indføre den såkaldte ARIMA-sæsonmodel.

…ved at betragte forskelle imellem observationer med sæsonlængden imellem

Den intuitive forskel på de to modeller er den, at i ARIMA-modellen indgår der differensdannelse imellem to på hinanden følgende observationer, mens der i ARIMA-sæsonmodellen indgår differensdannelse imellem to observationer med præcis sæsonlængden imellem sig.

Sæsonlængden

Sæsonlængden vil være fire for kvartalstal og tolv for månedstal. Vi kalder en ARIMA-sæsonmodel for ARIMA(P,D,Q) $_S$, hvor P er ordenen af autoregression i sæsonmodellen, D er ordenen på differensdannelsen, Q er ordenen af glidende gennemsnit i sæsonmodellen og S er sæsonlængden.

En sæson-ARIMA-model

En sæson-ARIMA(P,D,Q) $_S$ model har altså følgende udseende;

$$(1-\beta_1 L^S - \cdots - \beta_P L^{SP})(1-L^S)^D x_t = (1-\phi_1 L^S - \cdots - \phi_Q L^{QS})\varepsilon_t$$

3.6 Multiplikative ARIMA-modeller

En kombination af ARIMA-modellen og sæson-ARIMA-modellen er bedst

Når man vil beskrive økonomiske korttidsserier, er hverken den almindelige ARIMA -model eller sæson-ARIMA-modellen alene velegnet.
Typisk vil der både forekomme trend- og sæsonbevægelse i fællesskab i en given tidsserie. En modeltype der kan anvendes til modellering af en bred klasse af tidsserier, er givet ved en kombination af disse to modeltyper;

Definition

Vi kalder en sådan kombinationsmodel for ARIMA(p,d,q)×ARIMA(P,D,Q) $_S$, og skrevet i formelsprog har denne model følgende form;

$$(1-\rho_1 L - \cdots - \rho_p L^p)(1-\beta_1 L^S - \cdots - \beta_P L^{PS})(1-L)^d(1-L^S)^D x_t$$
$$= (1-\alpha_1 L - \cdots - \alpha_q L^q)(1-\phi_1 L^S - \cdots - \phi_Q L^{SQ})\varepsilon_t$$

Eksempel: Airline modellen
Den såkaldte Box-Jenkins Airline model er det specialtilfælde af den multiplikative ARIMA-model hvor p=P=0 og d=D=q=Q=1 gælder. DVS Airline-modellen har altså følgende udseende;

$$(1-L)(1-L^S)x_t = (1-\alpha_1 L)(1-\phi_1 L^S)\varepsilon_t$$

Denne model er ARIMA(0,1,1)×ARIMA(0,1,1) $_s$.

3.7 Identifikation af ARIMA-modeller

Transformation til en stationær serie

Når en ARIMA-model skal identificeres, skal det fastlægges hvorledes tidsserien transformeres ved differenser til en stationær serie. Dette kan fastlægges ved den såkaldte autokorrelationsfunktion.

Korrelationen

I helt generelle termer er korrelationskoefficienten imellem to variable x og y lig $\frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}}$, som er et tal der altid ligger imellem -1 og 1.

Afbildes autokorrellationsfunktionen grafisk fås en figur der kaldes et korrellogram.

Definition: Autokorrelations-funktion

Autokorrelationsfunktionen (ACF) er en tilsvarende funktion som udregnes for tidsrækker, nemlig som korrelationen mellem x_t og x_{t-j}.

Autokorrelationsfunktionen (hvor n er antallet af observationer) estimeres således ved

$$\rho(j) = \frac{n-1}{n-j-1} \frac{\sum_{i=j+1}^{n} (x_t - \bar{x})(x_{t-j} - \bar{x})}{\sum_{i=1}^{n} (x_t - \bar{x})^2}.$$

her er $\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t$

Ingen sammenhæng, så er autokorrelationen lig 0

Hvis der ikke er nogen sammenhæng mellem observationerne for et givet lag j, er autokorrelationen lig 0.

Hvorledes bruges autokorrelations-funktionen?

Autokorrelationsfunktionen benyttes til at finde ud af, hvor mange gange en proces skal differenses (enten almindelig differens eller sæsondifferens) for at blive stationær, altså til at finde ordenen som den pågældende serie er integreret af.

ACF for en stationær serie

For en stationær serie gælder, at ACF værdierne er små og uden systematik.

ACF fastlægger differensordenen og sæsondifferensordenen

Autokorrelationsfunktionen (ofte suppleret med afledte funktioner heraf) for den transformerede tidsserie benyttes herefter til at identificere ordenen af autoregression og glidende gennemsnit, dvs. fastlæggelse af parametrene p,P og q,Q. Denne identifikationsproces skal dog ikke beskrives nærmere her, da den bygger på en kompleks matematisk begrebsverden, som er uden for rammerne af denne beskrivelse.

3.8 Kontrol af ARIMA-modeller

Først undersøges parametrene...

Første trin i modelkontrollen er en undersøgelse af de estimerede parametre i modellen. Der skal generelt gælde, at summen af parameterestimaterne i henholdsvis den ordinære model og sæsonmodellen hver især numerisk set skal være skarpt mindre end 1. I praksis vil kravet ofte være skærpet til 0,9. Hvis det gælder at summen af disse parameterestimater overskrider denne grænse, betyder det, at der er foretaget for mange differensdannelser i modellen.

...dernæst undersøges modellens forklaringsgrad

Næste trin er at undersøge hvor godt modellen rent faktisk beskriver variationen af data. Modellens forklaringsgrad vurderes typisk ved beregning af determinationskoefficienten R^2. Denne størrelse beregnes ved:

$$R^2 = \frac{\hat{\sigma}^2 - \hat{\sigma}^2_{res}}{\hat{\sigma}^2} = \frac{Variansen\ i\ oprindelig\ serie - residual\ Variansen}{Variansen\ i\ oprindelig\ serie}$$

Determinationskoefficienten ligger mellem 0 og 1, og høje værdier af denne størrelse er overhovedet ikke usædvanligt. Desuden er R^2 den andel af variansen i den afhængige variabel, som kan forklares ved regressionen. Ved at tilføje flere variable, endda nonsens-variable, i regressionen vil R^2 altid stige. Derfor bør man fortolke R^2 med en vis forsigtighed.

residualvariansen undersøges...

En af de helt basale forudsætninger i modellen er, at fejlleddet er hvid støj. Dette kan undersøges ved hjælp af det såkaldte Ljung-Box portmanteau test, som er en funktion af ACF og er givet ved:

$$Q_{LB} = n(n+2)\sum_{j=1}^{k} \frac{\rho_j^2}{n-j}$$

hvor ρ_j er den estimerede residualautokorrelation for lag j, n er antallet af residualer og k er antallet af lags der indgår i estimationen.

Q_{LB} er asymptotisk χ^2-fordelt med et antal frihedsgrader som er lig k fratrukket antal estimerede parametre i modellen.

... og modellens prediktionsevne

En estimeret ARIMA-model anvendes til forudsigelser. Modellens prediktionsevne kan derfor undersøges ved at sammenholde de værdier som modellen predikterer på basis af observationerne op til prediktionsperioden, med de faktiske værdier for en på forhånd given periode fx de to seneste år.

4. Metoder til sæsonkorrigering

Ingen entydigt bestemt måde at sæsonkorrigere på

Da definitionen af begrebet sæsonkorrigering er meget upræcis, findes der ingen entydig bestemt måde at sæsonkorrigere på. Der findes mange forskellige indfaldsvinkler, og disse spænder lige fra blot at tegne en figur, til de avancerede modelbaserede metoder TRAMO/SEATS og X-12ARIMA. I nedenstående kapitel vil metoderne blive præsenteret i den historiske rækkefølge. I afsnit 4.1 vil sæsonkorrigering ved frihåndskurver blive gennemgået. I afsnit 4.2 vil den glidende gennemsnitsmetode blive gennemgået, herunder beskrives X-11 familien. I afsnit 4.3 vil de modelbaserede metoder blive gennemgået, herunder X-12 og TRAMO/SEATS.

4.1 Frihåndskurver

Før computeren for alvor holdt sit indtog i statistikkens verden foretog man også sæsonkorrigering. Original serien blev afbildet grafisk. En person med et detaljeret kendskab til denne konkrete series udvikling indtegnede på skønsmæssig vis en jævn trend-cykel kurve. Denne kurve var ganske simpelt den sæsonkorrigerede serie. Denne metode kan i dag virke noget primitiv, men for det første skal man huske på at denne frihåndskorrigering faktisk blev udført af en ekspert, og desuden var det den eneste mulighed rent teknologisk.

Frihåndskurver er basis

Basis for de første sæsonkorrigeringsprogrammer faktisk var sådanne frihåndskorrigeringer af lange tidsserier.

Metoden er problematisk

Men det er på den anden side også helt klart at metoden er dybt problematisk. Det er aldrig muligt for nogen som helst, ikke engang samme person, at genskabe præcis den samme sæsonkorrigerede serie. Ydermere må metoden siges at være ret ressourcekrævende.

4.2 Glidende gennemsnitsmetoden

Basis for X-11 familien er glidende gennemsnit

Glidende gennemsnit bruges ofte når man antager en komponentmodel for en tidsrække. Anvendelsen af glidende gennemsnit muliggør i dette tilfælde estimation af de enkelte komponenter på en relativ enkel vis, og de anvendes i vid udstrækning i X-11 familien.

X-11 familien er anvendt udbredt, men er udskældt…

X-11 familien er en udbredt sæsonkorrigeringspakke, og anvendes af statistiske bureauer over hele verden. På den anden side er X-11 et af de mest udskældte programmer i den akademiske verden.

…da der ingen underliggende statistisk model eksisterer

Dette skyldes, at det er en sæsonkorrigeringsmetode, som er baseret på forskellige typer af glidende gennemsnit uden en eksplicit underliggende model, og den er udviklet primært ud fra en empirisk baggrund.

$$Q_{LB} = n(n+2)\sum_{j=1}^{k} \frac{\rho_j^2}{n-j}$$

hvor ρ_j er den estimerede residualautokorrelation for lag j, n er antallet af residualer og k er antallet af lags der indgår i estimationen.

Q_{LB} er asymptotisk χ^2-fordelt med et antal frihedsgrader som er lig k fratrukket antal estimerede parametre i modellen.

... og modellens prediktionsevne

En estimeret ARIMA-model anvendes til forudsigelser. Modellens prediktionsevne kan derfor undersøges ved at sammenholde de værdier som modellen predikterer på basis af observationerne op til prediktionsperioden, med de faktiske værdier for en på forhånd given periode fx de to seneste år.

4. Metoder til sæsonkorrigering

Ingen entydigt bestemt måde at sæsonkorrigere på

Da definitionen af begrebet sæsonkorrigering er meget upræcis, findes der ingen entydig bestemt måde at sæsonkorrigere på. Der findes mange forskellige indfaldsvinkler, og disse spænder lige fra blot at tegne en figur, til de avancerede modelbaserede metoder TRAMO/SEATS og X-12ARIMA. I nedenstående kapitel vil metoderne blive præsenteret i den historiske rækkefølge. I afsnit 4.1 vil sæsonkorrigering ved frihåndskurver blive gennemgået. I afsnit 4.2 vil den glidende gennemsnitsmetode blive gennemgået, herunder beskrives X-11 familien. I afsnit 4.3 vil de modelbaserede metoder blive gennemgået, herunder X-12 og TRAMO/SEATS.

4.1 Frihåndskurver

Før computeren for alvor holdt sit indtog i statistikkens verden foretog man også sæsonkorrigering. Original serien blev afbildet grafisk. En person med et detaljeret kendskab til denne konkrete series udvikling indtegnede på skønsmæssig vis en jævn trend-cykel kurve. Denne kurve var ganske simpelt den sæsonkorrigerede serie. Denne metode kan i dag virke noget primitiv, men for det første skal man huske på at denne frihåndskorrigering faktisk blev udført af en ekspert, og desuden var det den eneste mulighed rent teknologisk.

Frihåndskurver er basis

Basis for de første sæsonkorrigeringsprogrammer faktisk var sådanne frihåndskorrigeringer af lange tidsserier.

Metoden er problematisk

Men det er på den anden side også helt klart at metoden er dybt problematisk. Det er aldrig muligt for nogen som helst, ikke engang samme person, at genskabe præcis den samme sæsonkorrigerede serie. Ydermere må metoden siges at være ret ressourcekrævende.

4.2 Glidende gennemsnitsmetoden

Basis for X-11 familien er glidende gennemsnit

Glidende gennemsnit bruges ofte når man antager en komponentmodel for en tidsrække. Anvendelsen af glidende gennemsnit muliggør i dette tilfælde estimation af de enkelte komponenter på en relativ enkel vis, og de anvendes i vid udstrækning i X-11 familien.

X-11 familien er anvendt udbredt, men er udskældt…

X-11 familien er en udbredt sæsonkorrigeringspakke, og anvendes af statistiske bureauer over hele verden. På den anden side er X-11 et af de mest udskældte programmer i den akademiske verden.

…da der ingen underliggende statistisk model eksisterer

Dette skyldes, at det er en sæsonkorrigeringsmetode, som er baseret på forskellige typer af glidende gennemsnit uden en eksplicit underliggende model, og den er udviklet primært ud fra en empirisk baggrund.

4.2.1 Historik for X-11 familien

Census I er færdigudviklet i 1954

– I 1954 var den såkaldte Census I metode færdigudviklet af US Bureau of the Census, og dette skridt har været det utvivlsomt største i sæsonkorrigeringens æra. Allerede i 1955 blev metoden videreudviklet til Census II. Census II var essentielt set en elektronisk version af de manuelle metoder som tidligere var anvendt indenfor sæsonkorrigerring. Metoden blev også kaldt X-0 (eXperimental). Metoden blev kritiseret på flere punkter, bl.a. var det problematisk, at metoden var funderet på empiriske metoder uden noget fundament i statistisk teori. Ydermere fandt man, at Census II blot fjernede de store sæsonudsvingninger, lod de mindre forblive uændrede, og dermed forstyrrede forholdet mellem modellens komponenter. Census II viste sig også i nogle tilfælde at overestimere sæsoneffekten.

X-3,…,X-10

– Blandt andet de ovenfor beskrevne problemer med Census II ledte til videreudviklingen af versionerne X-3 til X-10.

X-11 færdigudviklet i 1965

– Yderligere udvikling ledte i 1965 til den såkaldte X-11 variant, som til dags dato stadigvæk er en af de mest anvendte sæsonkorrigeringsmetoder

X-11ARIMA offentliggøres i 1988

– I 1988 lavede Statistics Canada en udbygget version af X-11, nemlig den såkaldte X-11ARIMA version. Metoden bruger en ARIMA (AutoRegressiv Integreret glidende gennemsnit (Moving Average)) model til at prediktere både fremtid og den fortid som ligger forud for seriens starttidspunkt. Dette er nødvendigt, da sæsonkorrigeringsanalysen baseres på symmetriske glidende gennemsnit. Det originale X-11 program ekstrapolerede disse out-of-sample observationer på en ret tilfældig vis.

X-12 er udviklet

– En ny version, det såkaldte X-12 ARIMA program, er udviklet af US Bureau of the Census. Dog skal man bemærke, at den bagvedliggende indfaldsvinkel i X-12 må siges at være modelbaseret, og dette i en langt højere grad end forgængeren X-11ARIMA. I X-11ARIMA bruges kun en ARIMA-model til prediktionen af fortiden og fremtiden, mens resten af analysen stort set foregår på samme måde som den gjorde i 1954.

4.2.2 Beskrivelse af X-11 metodikken

Den glidende gennemsnitsmetode som anvendes i X-11ARIMA, kan i hovedtræk beskrives ved følgende iterative proces;

Først indlægges forhåndskorrektioner

Første punkt er indlæggelse af eventuelle forhåndskorrektioner.

Dernæst estimeres trend-cyklen

Dernæst estimeres trend-cyklen for den forhåndskorrigerede serie ved beregningen af et centreret glidende gennemsnit[7]. Ved kvartalstal er længden af dette glidende gennemsnit typisk fire kvartaler, mens det typisk er på tolv for månedsserier, netop svarende til en længde som udglatter sæsonen.

Trend-cyklen fjernes

Trend-cyklen fjernes fra den originale serie ved at dividere/fratrække (svarende til den multiplikative hhv. additive model) den estimerede trend-cykel op i/fra den originale serie. Denne for trend-cyklen rensede serie, kaldes for SI-raten (Seasonal and Irregular).

Ekstreme værdier fjernes

Ekstreme værdier af SI-raterne erstattes med mindre ekstreme værdier.

[7] For en mere fyldestgørende gennemgang af de glidende gennemsnit henvises der til appendix 1.

Første skøn for den sæsonkorrigerede serie

Disse modificerede værdier af SI-raterne glattes efterfølgende ud ved anvendelse af glidende gennemsnit.

De herved beregnede sæsonfaktorer divideres op i /fratrækkes den originale serie, og herved fremkommer det første skøn for den sæsonkorrigerede serie.

De første beregningstrin gentages

Med udgangspunkt i dette første skøn over den sæsonkorrigerede serie gentages de første beregningstrin (1.-5.) tre gange i alt.
Formålet er en forbedring, dvs. en udglatning af trend-cykel estimatet, og dermed en mere sikker identifikation af ekstreme værdier.

4.2.3 Vurdering af metoden

Metoden har klare fordele...

Den udbredte anvendelse af glidende gennemsnitsmetoden i praksis skyldes først og fremmest, at en iterativ anvendelse af glidende gennemsnit til trend-cykel estimation er meget fleksibel.

Ydermere må man sige, at der generelt set er et stort antal muligheder i proceduren.

...men også sine ulemper

Et par af de kritikpunkter af X-11ARIMA proceduren, som nævnes i litteraturen (Seasonal adjustment Methods – A comparison, Eurostat, 09.98), skal dog også nævnes her.

1. For det første er test faciliteterne i proceduren problematiske.

   Testene er fx. testet for handelsdagseffekt, F-tests på den estimerede irregulære komponent. Disse F-tests er baseret på antagelsen at den irregulære komponent er normalfordelt hvid støj.

   Dog finder man desværre ofte at der er autokorrelation i de estimerede irregulære faktorer. Dette skyldes filteret som anvendes til at udskille estimatet for den irregulære komponent.

   Antagelserne, der eksplicit ligger bag nulhypotesen om ingen handelsdagseffekt, er derfor ikke opfyldt.

   Statistics Canada er selv opmærksomme på dette problem. I manualen foreslås det at man blot tester på 1/1000 niveau i stedet. Denne fremgangsmåde løser ikke problemet.

2. En anden indvending imod X-11ARIMA er, at behandlingen af ekstreme observationer, specielt ved niveauskift er for simplificeret.

   De glidende gennemsnit er let påvirkelige af et niveauskift og programmet vil forsøge at udbløde denne type af observationer når de første gang forekommer. De vil her blive opfattet som ekstreme observationer.

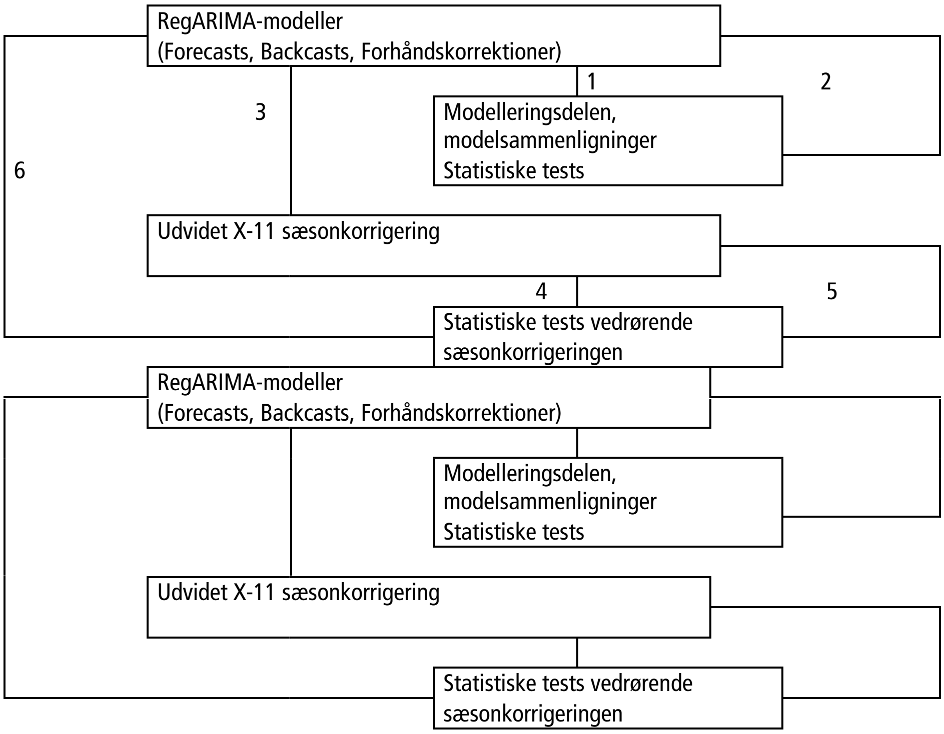
   Kendte man baggrunden for et sådant niveauskift, kunne man pre-korrigere, men desværre findes der ingen værktøjer til at hjælpe os med at estimere effekten af niveauskift.

4.3 De modelbaserede metoder

4.3.1 X-12 Arima

Strukturen i X-12 Regarima, som er det nyeste skud på X-11 stammen, skitseres ved følgende oversigt

Figur 4.3.1.1 **X-12 Regarima**



Forklaring

1. Oversigten skal læses således at man med Reg-ARIMA-modellerne som udgangspunkt går til modelsammenligningen herefter går man tilbage til RegARIMA-delen.
2. Er modelleringsdelen gået godt således at der er valgt en passende model, da fortsættes til sæsonkorrigeringsdelen. Hvis ikke en passende model er valgt går vi tilbage til udgangspunktet indtil det lykkedes.
3. Når der er sæsonkorrigeret vendes tilbage til de statistiske tests vedrørende selve sæsonkorrigeringen.
4. Er testene i orden, stoppes processen hvis ikke gås tilbage og der foretages endnu en sæsonkorrigering eller endnu et modelvalg med efterfølgende sæsonkorrigering.

4.3.1.1 Reg-ARIMA-modellen

En Reg-ARIMA-model

En Reg-ARIMA-model er en udvidelse af ARIMA-modellerne, hvor der beskrives en lineær regressionssammenhæng for processen x_t ved følgende

$$x_t = \sum_i \beta_i y_{it} + z_t$$

hvor y'erne er regressionsvariable observeret samtidig med x, og β'erne er regressionsparamtrene. Fejlleddet z antages at følge en generel ARIMA(p,d,q)-×ARIMA(P,D,Q)$_s$ model.

Denne atypiske regressionsmodel kaldes en RegARIMA-model.
I modellen antages implicit, at de forklarende variabler y kun påvirker den afhængige variabel x igennem nutidige værdier, der er ikke historiske y'er i modelformuleringen.

Udvidelse af ARIMA-modellen

Udvidelsen i forhold til den almindelige ARIMA-model er klar. I modellen ligger der dels en muliggørelse af en regressionssammenhæng i processens middelværdi, og dels en udvidelse af den almindelige regressionsmodel til at tillade, at fejlleddene følger en generel ARIMA proces. Denne klasse af modeller er således temmelig rummelig.

Specifikationskrav

Specifikationen af en Reg-ARIMA-model kræver både at regressionsvariablerne (y_{it}'erne) og ARIMA-modellen for regressionsfejlene z_t specificeres.

4.3.1.2 Standardregressionsvariable i X-12

Adskillige standardregressionsvariable er indbygget i X-12. Se X-12-ARIMA Reference Manual fra US Bureau of the Census for videre forklaring.

Konstantleddet bruges som forklarende variabel…

Den mest basale regressionsvariabel er konstantleddet. Når der ikke defineres differensdannelse i Reg-ARIMA-modellen, er konstantleddet blot det sædvanlige niveau i regressionen. Forekommer der differensdannelse i modeldefinitionen, da vil X-12 ARIMA bruge en trend som regressionsvariabel. Trenden er defineret således, at differenses denne som defineret i Reg-ARIMA-modellen giver leddet en søjle med ettaller, altså et konstantled som ønsket.

... samt en hel række af andre variabler:

Andre typer af standardregressionsvariable er de faste sæsoneffekter, handelsdagseffekter, ferieeffekter og en række specielle dummyer til at modellere midlertidige udsving i serien .

4.3.1.3 Månedsserier

De faste sæsoneffekter

I en månedlig serie kan de faste sæsoneffekter modelleres ved at bruge 12 indikatorvariabler, svarende til en for hver måned. Dette giver en overparametrisering af modellen, hvis man også inkluderer et konstantled, og derfor vælger man typisk at bruge 11 indikatorvariabler (svarende til 11 måneder), og derefter udlede koefficienten til den tolvte måned ud fra de 11 andre og konstantleddet.

Handelsdags-effekter

Handelsdagseffekter er synlige, når en serie påvirkes af et forskelligt antal af de forskellige dage i løbet af en kalendermåned. Handelsdagseffekter modelleres ved variabler som er givet ved (Antal mandage),…,(Antal søndage). Her benyttes kun de seks af dem, idet der indgår et konstantled i modellen.

Arbejdsdags-effekter

Arbejdsdagseffekten er en anden måde at modellere forskellen i ugens dage på. Ved en arbejdsdagskorrektion betragter man serier, der påvirkes af antallet af arbejdsdage i løbet af en måned, hvilket typisk er produktionsserier. Arbejdsdagseffekten modelleres ved de to variabler givet ved (Antal arbejdsdage) og (Antal fridage). Her benyttes kun en af dem, idet der indgår et konstantled i modellen.

Ferie-effekter

Ferieeffekter kan ses i månedsserier, og effekten stammer fra ferieperioder, hvor datoerne varierer over tid, hvis:

- enten den aktivitet, der måles af serien vokser eller aftager mærkbart omkring ferietidspunktet.
- eller dette påvirker mindst to på hinanden følgende måneder i et kalenderår, afhængig af i hvilken måned ferien falder.

Påskeeffekten er den hyppigst fundne ferie under dette fænomen.

Andre midlertidige effekter

X-12 kan også tage højde for pludselige niveauændringer i en serie, hvad enten de er af midlertidig eller permanent karakter, der er indbygget følgende fire typer af regressionsvariabler, der kan tage højde for sådanne ændringer;

– AO(Additive Outliers): Der lægges blot en konstant ind på et givet tidspunkt.
– LS(Level Shift): Der lægges en konstant til fra et givet tidspunkt.
– TC(Temporary Change): Der lægges et niveauskift med en eksponentiel vækst ind.
– Ramp: Der lægges en affin funktion af t (over en begrænset periode) ind. Funktionen er konstant lig med -1 til startniveau, er en lineær funktion i t indtil slutniveauet, hvorefter funktionen konstant er lig 0.

4.3.2 TRAMO/SEATS[8] - den ARIMA baserede model

I TRAMO/SEATS arbejdes der med den samme udgangsmodel som i X-12, nemlig Reg-ARIMA-modellen

$$x_t = \sum_i \beta_i y_{it} + z_t$$

hvor det første led igen blot er deterministiske regressionsvariable, og fejlleddet antages, at følge en ARIMA(p,d,q)×ARIMA(P,D,Q) $_s$ proces.

Ganske som i X-12 kan man lægge handelsdagskorrektioner, påskekorrektioner og korrektioner for enten midlertidige eller permanente niveauskift ind som de deterministiske regressorer.

Hvad sker der i prekorrigeringsdelen TRAMO?

I TRAMO foretages en eksakt Maximum-Likelihood estimation af ARIMA regressionsmodellen. Der udregnes teststørrelser og ekstreme observationer findes og rettes. Der interpoleres ved manglende observationer og der udregnes forecasts.

Hvad sker der i SEATS?

Indeholder den serie som ønskes sæsonkorrigeret enten ekstreme observationer, manglende observationer eller er den påvirket af deterministiske regressorer, skal den først 'renses' af TRAMO. Den 'rensede' og lineariserede serie (output fra ARIMA-modellen) dekomponeres i SEATS til de fire sædvanlige komponenter, nemlig sæson-, trend-, konjunktur- og irregulær komponent .

Fordel ved TRAMO/SEATS

Den indiskutable fordel ved SEATS frem for de ad-hoc baserede metoder er, at den udregner optimale estimatorer og forecasts med hensyn til veldefinerede statistiske modeller og veldefinerede estimationskriterier.

Figur 4.3.2.1 **Oversigt over TRAMO/SEATS**

Den observerede serie
↓
TRAMO
↓
Interpoleret serie
↓

Den stokastiske del: 'Lineariseret' serie (output fra TRAMO)	Regressions effekter
↓	Ekstreme observationer: AO, TC,LS
SEATS	
↓	Handelsdagseffekt
Trend	Påske effekt
+	Ferieeffekt
Seasonal	
+	Regressionsvariabler
(Cykel)	
+	
Irregulær	
↓	↓

Endelige komponenter

4.3.3 Demetra

Demetra er en windowsbaseret brugeroverflade til både X-12 og TRAMO/SEATS. Initiativet til udviklingen af Demetra er taget af Eurostat. Demetra er meget brugervenligt og er velegnet til sæsonkorrigering af mange serier. En del lande i Europa enten anvender eller vil anvende Demetra i sæsonkorrigeringen.

[8] Time series Regression with ARIMA noise, Missing observations and Outliers/Signal Extraction in ARIMA Time Series

5. Om X-11ARIMA

Som det tidligere er nævnt, anvender Danmarks Statistik til og med år 2002 sæsonkorrigeringsprogrammet X-11ARIMA som er udviklet af Statistics Canada. I dette kapitel gives et overblik over beregningsgangen i X-11ARIMA.

5.1 De basale funktioner

I X-11ARIMA foretages overordnet set primært to basale funktioner:

– Forecasting
– Sæsonkorrigering

De basale funktioner i X-11ARIMA er primært forecasting...

Mulighederne i X-11ARIMA med hensyn til forecasting funktionen er

– Indbyggede ARIMA-modeller
– Variabel forecast-længde
– Der backcastes kun for serier som er kortere end syv år
– Autokorrelationsfunktionen for de fittede residualer i de indbyggede ARIMA-modeller udskrives.
– Handelsdags- og påskeeffekter fjernes automatisk (hvis de er tilstede!) før ARIMA-modelleringen.

...og sæsonkorrigeringsfunktionen

I sæsonkorrigeringsfunktionen i X-11ARIMA har man også adskillige muligheder med hensyn til selve sæsonkorrigeringen. Man kan:

– Estimere påskeeffekten
– Øge præcisionen af de asymmetriske Henderson trend-cykel filtre
– Reskalere den originale serie
– Indlægge midlertidige eller permanente forhåndskorrigeringer af originalserien
– Få bruger-definerer print.

ARIMA forecasts er en mulighed

Man vælger selv, om man vil benytte muligheden for ARIMA forecasts eller ej. Fravælger man ARIMA extrapolationen af serien, får man resultater som er tæt på, men ikke nødvendigvis helt magen til de, som fås af X-11. Forskellen skyldes primært forskelle i identifikationen og måden de ekstreme observationer erstattes på, og et bedre filtervalg i X-11ARIMA.

5.2 Forecasting mulighederne i X-11ARIMA

Modeller i X-11ARIMA

Til at ARIMA-modellere tidsserierne uden for endepunkterne (til brug for de symmetriske glidende gennemsnit) kan man i X-11ARIMA vælge mellem følgende multiplikative modeller:

$\text{ARIMA}(0,1,1)\times\text{ARIMA}(0,1,1)_S$

$\text{ARIMA}(0,1,2)\times\text{ARIMA}(0,1,1)_S$

$\text{ARIMA}(2,1,0)\times\text{ARIMA}(0,1,1)_S$

$\text{ARIMA}(0,2,2)\times\text{ARIMA}(0,1,1)_S$

$\text{ARIMA}(2,1,2)\times\text{ARIMA}(0,1,1)_S$

Disse fem modeller testes i den rækkefølge de er opskrevet i, hvilket vil sige at for en given serie er det langt fra sikkert, at alle fem modeller gennemprøves. Hvis den første model (Airlinemodellen) beskriver data godt, så vælges den model og der forsøges ikke med flere.

Variable forecast længde

I X-11ARIMA er der en delrutine der tillader brugeren at vælge en forecast længde på op til tre år, dvs. seksogtredive måneder, eller tolv kvartaler. Vælges der ingen forecast længde, sættes den automatisk til et år. Længden af forecastet hænger sammen med problemstillingen vedrørende minimering af revisioner der skyldes filter ændringer.

Backcasting…

En af de andre muligheder for at tilrette forecast funktionen i X-11ARIMA er muligheden for backcasting. Det er muligt at backcaste serier, hvis længde er mindre end femten år, men der foretages kun backcasting automatisk for serier, hvis længde er mindre end syv år.

…forbedrer kvaliteten af sæsonkorrigeringen

Backcasting forbedrer den generelle kvalitet af sæsonkorrigeringen, men det introducerer til gengæld permanente revisioner i de estimerede komponenter i seriens begyndelse. Ydermere giver backcasting revisioner i de seneste sæsonkorrigerede tal i løbet af 8-10 år.

Påske- og handelsdagseffekter

Hverken handelsdags- eller påskeeffekter kan samles op i en ARIMA-model, idet disse effekter er deterministiske. Derfor skal effekterne fjernes fra tidsserien før den ARIMA-modelleres.

5.3 Sæsonkorrigeringsdelen i X-11ARIMA

Sæsonkorrigeringen beskrives ved at beskrive de indgående komponenter.

Komponenter

De vigtigste indgående komponenter i X-11ARIMA er:

1. estimation af påskeeffekten
2. de asymmetriske Henderson-filtre
3. den automatiske udvælgelse af default sæsonfiltre
4. reskalering
5. forhåndskorrigeringer
6. figurer.

5.3.1 Estimation af Påskeeffekten

Påske er en ikke-månedsfastlagt serie af helligdage. Påsken kan falde enten i marts eller i april, dvs. i enten 1. eller 2. kvartal. Den kan således skabe alvorlige forstyrrelser i både måneds- og kvartalstal.

For at tage påskens skiftende beliggenhed i betragtning er der i X-11ARIMA indbygget en såkaldt påskemodel, der defineres ved følgende variabel:

$$E_i = \frac{1}{2} f(z_i) \left[\frac{\sum_{i \in M} (I_{i,j+i} - I_{i,j})}{n_M} - \frac{\sum_{i \in A} (I_{i,j+i} - I_{i,j})}{n_A} \right]$$

hvor z_i =antallet af dage mellem påskesøndag i det i´te år og den tidligst mulige påskedag i det i´te år som er den 22. marts.

f er en funktion hvor $f(z_i) = \begin{cases} 1, & hvis\ \ Z_i \leq 9 \\ 0, & hvis\ \ Z_i \geq 9 \end{cases}$

i angiver året og j måneden marts (hermed er altså j+1 lig april). I_{ij} er de estimerede residualer fra den første iteration af X-11ARIMA, og de antages at være påvirket af påskeeffekten E_i .

n_M er antallet af år, hvor påsken faldt i marts, og

n_A er antallet af år, hvor påsken faldt i april

5.3.2 Automatisk udvælgelse af default sæsonfiltrene

X-11ARIMA udvælger automatisk et fem-leds (3 x 3) og et syv-leds (3 x 5) vægtet glidende gennemsnit som sæsonfiltre. Et 3 x 5 perioders glidende gennemsnit er en beregning, hvor man først beregner 5 perioders glidende gennemsnit og derefter beregner 3 perioder glidende gennemsnit af disse 5-perioders gennemsnit; i et sådant gennemsnit indgår der 7 perioder, jf. appendiks 1.

Andre typer skal måske vælges

Det kan dog vise sig, at andre typer af glidende gennemsnit er mere passende for en given serie. Dette finder man ud af ved at betragte den såkaldte glidende sæsonmønster kvotient (Moving Seasonality Ratio, MSR), som er givet ved den såkaldte I/S ratio.

Retningslinier

Der er opstillet en række retningslinier (se fx manualen for X-11ARIMA) for hvilke typer af glidende gennemsnit, der bør anvendes.
Eksempelvis bør man for serier med en længde på mindre end 15 år anvende et 3 x 9 glidende gennemsnit, hvis MSR ligger i intervallet [5,2;6,5].

5.3.3 Forhåndskorrigeringer

Man kan i X-11ARIMA foretage både midlertidige og permanente forhåndskorrigeringer.

Midlertidige forhåndskorrigeringer

De midlertidige korrigeringer foretages ved at korrigere originaldata, således at enkelte ekstreme observationer ikke påvirker estimationen af sæsonkomponenterne. Rettelsen annulleres dog i de endeligt sæsonkorrigerede tal, således at de ekstreme observationer slår igennem her.

Permanente forhåndskorrigeringer

De permanente forhåndskorrigeringer er korrektioner, som vi ikke ønsker viser sig i det sæsonkorrigerede data. Begrebet omfatter blandt andet påske- og handelsdagskorrigeringer, og disse forstyrrende faktorer ønsker vi ikke slår igennem i det endeligt korrigerede data.

5.3.4 Hvordan vurderes kvaliteten af sæsonkorrigeringen i X11-ARIMA?

Når der er foretaget en sæsonkorrigering i X11-ARIMA skal man vurdere kvaliteten af sæsonkorrigeringen.

Q-teststørrelsen…

Der udregnes en teststørrelse, der viser hvorledes helhedskvaliteten af sæsonkorrigeringen er. Denne teststørrelse er den såkaldte Q-teststørrelse, der er et vægtet gennemsnit af 11 komponentteststørrelser. Q teststørrelsen er normeret således at den altid ligger imellem 0 og 3. Ligger Q imellem 0 og 1, da er sæsonkorrigeringen som helhed betragtet forløbet acceptabelt, men er Q større end 1 så er der problemer i sæsonkorrigeringen.

...og figurer Der udarbejdes typisk to figurer ved sæsonkorrigeringen af en serie.

1. Den første figur er originalserien mod den sæsonkorrigerede serie.
2. Den anden figur er sæsonfaktoren mod den irregulære faktor.

Den første figur har primært et illustrativt formål, men man bør vurdere udfra denne om der er sæsonsvingninger i de sæsonkorrigerede tal. Det må der ikke være. Den anden figur viser dels om der er estimeret et stabilt sæsonmønster i serien over tid. Det skal der være. Dels viser figuren om den irregulære faktor er stor set i forhold til sæsonfaktoren. Er den irregulære faktor forholdsvis stor da må man konkludere at sæsonkorriugeringen bliver vanskeliggjort af dette.

6. Anvendelser i Danmarks Statistik

I dette kapitel gives en række eksempler på proceduren for sæsonkorrigeringen i Danmarks Statistik.

Udvalgte serier Der er udvalgt en række af de serier, som offentliggøres i sæsonkorrigeret form fra Danmarks Statistik. Serierne løber alle til primo året 2000, og den sæsonkorrigering der er foretaget her, er som den var i året 2000 for de enkelte statistikker.

Grafiske præsentationer Der bliver i dette kapitel primært givet grafiske præsentationer af sæsonkorrigeringen.

Opdeling i grupper De udvalgte serier er opdelt i fire grupper. Den første gruppe er i afsnit 6.1.1 og omhandler uproblematiske sæsonkorrigeringer. Den anden gruppe er i afsnit 6.1.2 og omhandler serier, hvor de estimerede irregulære komponenter er store set i forhold til den estimerede sæsonkomponent. Den tredje gruppe i afsnit 6.1.3 beskriver serier, hvor sæsonkorrigeringen skal undersøges nærmere. Den fjerde og sidste gruppe er i afsnit 6.1.4 og omhandler problematikken vedrørende direkte og indirekte sæsonkorrigering.

Detaljer I appendiks 3 er nærmere angivet modelvalg, teststørrelser osv. i forbindelse med sæsonkorrigeringen.

To figurer for hver serie Der vises for hver enkelt sæsonkorrigering i afsnittene 6.1.1, 6.1.2 og 6.1.3 to figurer. Den første viser originalserien afbildet mod den sæsonkorrigerede serie. Den anden viser serien af estimerede sæsonfaktorer afbildet mod serien af estimerede irregulære faktorer.

Første figur Den første figur er selvforklarende.

Anden figur Den anden figur kan benyttes til at observere to vigtige aspekter af sæsonkorrigeringen. For det første kan man bedømme om det estimerede sæsonmønster er stabilt, og for det andet kan man vurdere forholdet mellem den irregulære faktor og sæsonfaktoren. Viser det sig ud fra teststørrelserne M1-M11 og Q, at sæsonkorrigeringen har en dårlig kvalitet, dvs. at teststørrelserne er større end 1, da kan man med fordel ud fra den anden figur konkludere, hvad problemet med sæsonkorrigeringen eksakt handler om.

Direkte og indirekte sæsonkorrigering For visse serier illustreres problematikken med hensyn til indirekte og direkte sæsonkorrigering. I afsnit 6.1.4 vises der for hver serie en figur, der illustrerer forskellen på den indirekte sæsonkorrigerede serie og den direkte sæsonkorrigerede serie.

...og figurer

Der udarbejdes typisk to figurer ved sæsonkorrigeringen af en serie.

1. Den første figur er originalserien mod den sæsonkorrigerede serie.
2. Den anden figur er sæsonfaktoren mod den irregulære faktor.

Den første figur har primært et illustrativt formål, men man bør vurdere udfra denne om der er sæsonsvingninger i de sæsonkorrigerede tal. Det må der ikke være. Den anden figur viser dels om der er estimeret et stabilt sæsonmønster i serien over tid. Det skal der være. Dels viser figuren om den irregulære faktor er stor set i forhold til sæsonfaktoren. Er den irregulære faktor forholdsvis stor da må man konkludere at sæsonkorriugeringen bliver vanskeliggjort af dette.

6. Anvendelser i Danmarks Statistik

I dette kapitel gives en række eksempler på proceduren for sæsonkorrigeringen i Danmarks Statistik.

Udvalgte serier

Der er udvalgt en række af de serier, som offentliggøres i sæsonkorrigeret form fra Danmarks Statistik. Serierne løber alle til primo året 2000, og den sæsonkorrigering der er foretaget her, er som den var i året 2000 for de enkelte statistikker.

Grafiske præsentationer

Der bliver i dette kapitel primært givet grafiske præsentationer af sæsonkorrigeringen.

Opdeling i grupper

De udvalgte serier er opdelt i fire grupper. Den første gruppe er i afsnit 6.1.1 og omhandler uproblematiske sæsonkorrigeringer. Den anden gruppe er i afsnit 6.1.2 og omhandler serier, hvor de estimerede irregulære komponenter er store set i forhold til den estimerede sæsonkomponent. Den tredje gruppe i afsnit 6.1.3 beskriver serier, hvor sæsonkorrigeringen skal undersøges nærmere. Den fjerde og sidste gruppe er i afsnit 6.1.4 og omhandler problematikken vedrørende direkte og indirekte sæsonkorrigering.

Detaljer

I appendiks 3 er nærmere angivet modelvalg, teststørrelser osv. i forbindelse med sæsonkorrigeringen.

To figurer for hver serie

Der vises for hver enkelt sæsonkorrigering i afsnittene 6.1.1, 6.1.2 og 6.1.3 to figurer. Den første viser originalserien afbildet mod den sæsonkorrigerede serie. Den anden viser serien af estimerede sæsonfaktorer afbildet mod serien af estimerede irregulære faktorer.

Første figur

Den første figur er selvforklarende.

Anden figur

Den anden figur kan benyttes til at observere to vigtige aspekter af sæsonkorrigeringen. For det første kan man bedømme om det estimerede sæsonmønster er stabilt, og for det andet kan man vurdere forholdet mellem den irregulære faktor og sæsonfaktoren. Viser det sig ud fra teststørrelserne M1-M11 og Q, at sæsonkorrigeringen har en dårlig kvalitet, dvs. at teststørrelserne er større end 1, da kan man med fordel ud fra den anden figur konkludere, hvad problemet med sæsonkorrigeringen eksakt handler om.

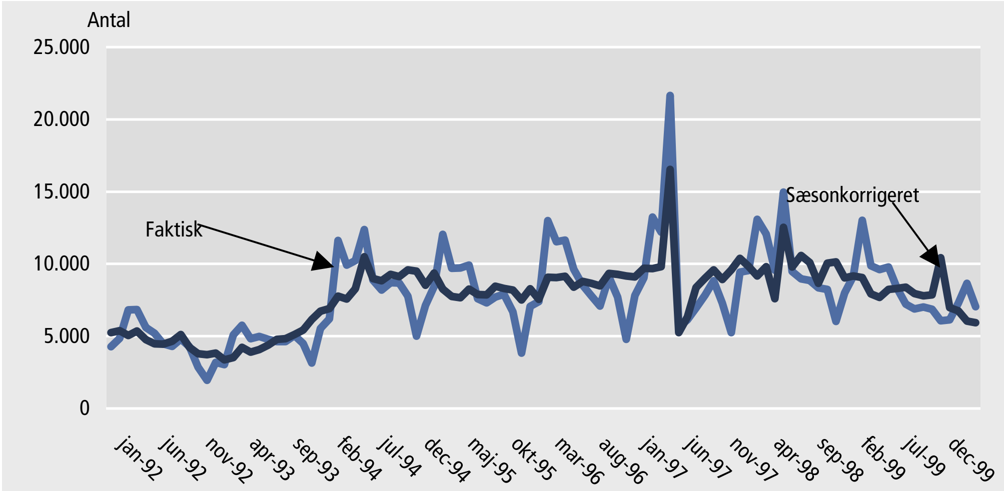
Direkte og indirekte sæsonkorrigering

For visse serier illustreres problematikken med hensyn til indirekte og direkte sæsonkorrigering. I afsnit 6.1.4 vises der for hver serie en figur, der illustrerer forskellen på den indirekte sæsonkorrigerede serie og den direkte sæsonkorrigerede serie.

6.1.1 Ikke-problematiske sæsonkorrigeringer

6.1.1.1 Nyregistrerede biler i husholdningerne

Figur 6.1.1.1.1



Kvaliteten af sæsonkorrigeringen

Sæsonkorrigeringen af de nyregistrerede biler i husholdningerne viser, at serien er velegnet til sæsonkorrigering. Den sammenvægtede teststørrelse Q er lig 0,63.

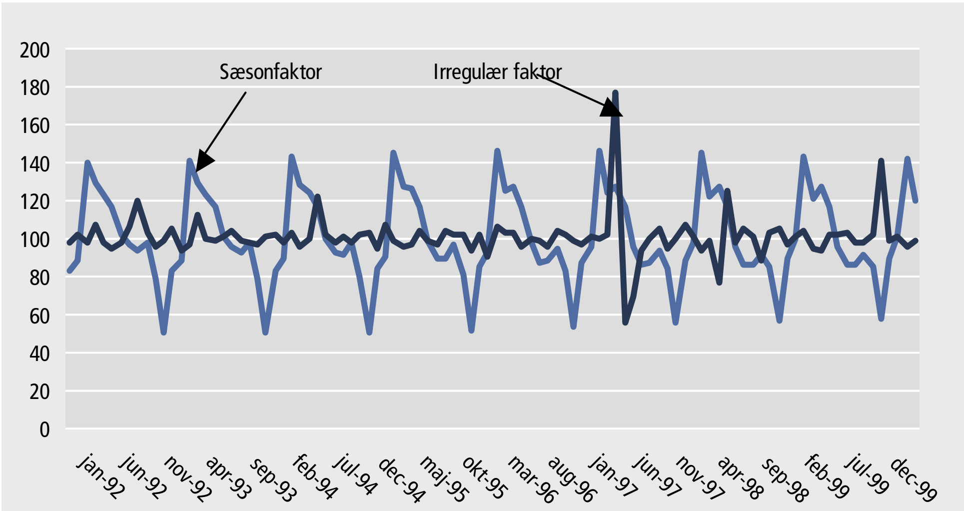
Komponenttest-størrelserne

Af de 11 komponentteststørrelser er der blot en enkelt, nemlig M2, der ligger over en, dvs. er problematisk. Dette skyldes, som man også tydeligt kan se af nedenstående figur, at visse steder er den irregulære faktor stor i forhold til sæsonfaktoren.

Om serien

Serien er en månedsserie observeret over perioden januar 1992 til april 2000, og der er anvendt en multiplikativ model med ARIMA forecasts. Der kan ikke estimeres en ARIMA-model ud fra data. Serien indeholder ingen påskeeffekt, men en klar handelsdagseffekt.

Figur 6.1.1.1.2



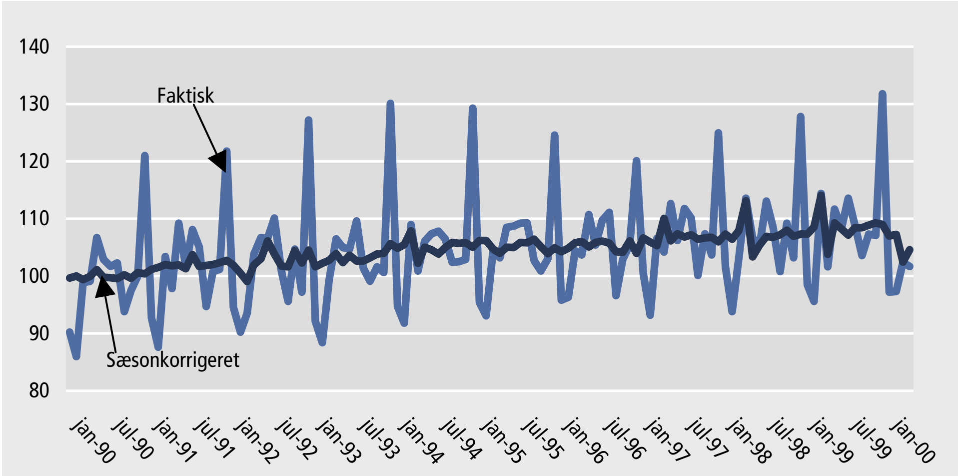
Af ovenstående figur ses det, at der er estimeret et nogenlunde stabilt sæsonmønster i perioden, og desuden ses det at den irregulære faktor er lille set i forhold til sæsonfaktoren.

Konklusion

Sammenholdes resultaterne fra sæsonkorrigeringen ses det, at kvaliteten af sæsonkorrigeringen af de nyregistrerede biler for husholdninger er god. Første kvalitetsindikator Q var mindre end 1, og ovenstående figur viste et stabilt sæsonmønster og ikke-dominerende estimerede irregulære faktorer.

6.1.1.2 Fødevarer og andre drikkevarer, detailomsætningsindekset

Figur 6.1.1.2.1



Kvaliteten af sæsonkorrigeringen

Sæsonkorrigeringen af detailomsætningsindekset for fødevarer og drikkevarer viser, at serien er velegnet til sæsonkorrigering. Den sammenvægtede teststørrelse Q er lig 0,67.

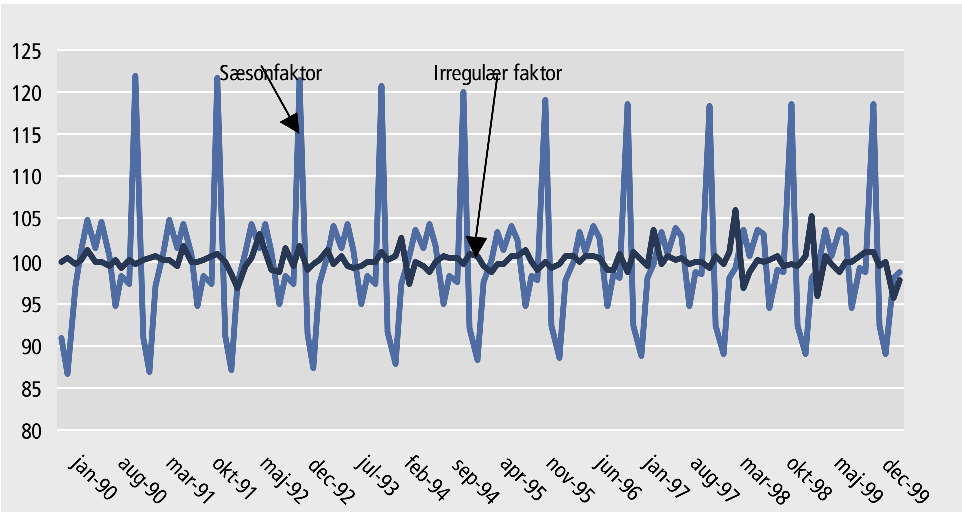
Komponentteststørrelserne

Af de 11 komponentteststørrelser er der blot to, nemlig M3 og M5, der ligger over en, dvs. er problematiske. Dette skyldes formentligt en mindre ikke-stabilitet i sæsonmønstret.

Om serien

Serien er en månedsserie observeret over perioden januar 1990 til april 2000, og der er anvendt en multiplikativ model med ARIMA forecasts. Der kan estimeres en ARIMA-model ud fra data. Serien indeholder ingen påskeeffekt, men en klar handelsdagseffekt.

Figur 6.1.1.2.2



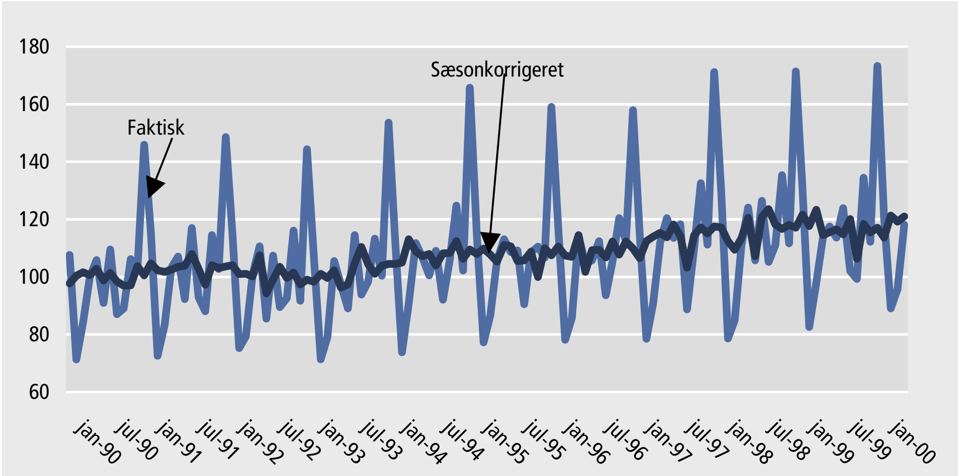
Ovenstående figur afbilder den estimerede sæsonfaktor mod den estimerede irregulære faktor. Denne figur viser både forholdet mellem de to størrelser og om der er estimeret et stabilt sæsonmønster, figuren underbygger de 11 komponentteststørrelser visuelt. Af figuren ses det, at der er estimeret et stabilt sæsonmønster i serien.

Konklusion

Sammenholdes resultaterne fra sæsonkorrigeringen ses det, at kvaliteten af sæsonkorrigeringen af detailomsætningsindekset for fødevarer og drikkevarer er god. Første kvalitetsindikator Q var mindre end 1, og ovenstående figur viste et stabilt sæsonmønster og ikke-dominerende estimerede irregulære faktorer.

6.1.1.3 Beklædning mv., detailomsætningsindekset

Figur 6.1.1.3.1



Kvaliteten af sæsonkorrigeringen

Sæsonkorrigeringen af detailomsætningsindekset for Beklædning mv. viser, at serien er velegnet til sæsonkorrigering. Den sammenvægtede teststørrelse Q er lig 0,86.

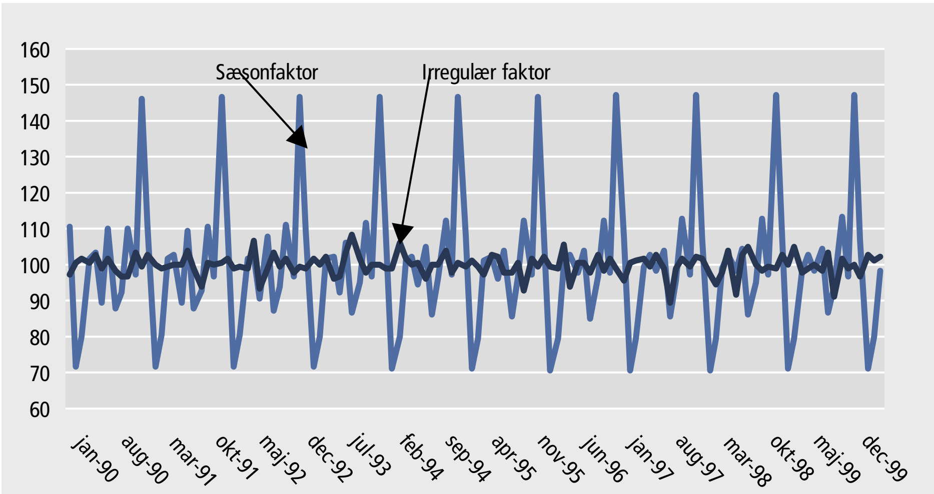
Komponenttest-størrelserne

Af de 11 komponentteststørrelser er der blot to, nemlig M3 og M5, der ligger over en, dvs. er problematiske. Dette skyldes formentligt en mindre ikke-stabilitet i sæsonmønstret.

Om serien

Serien er en månedsserie observeret over perioden januar 1990 til april 2000, og der er anvendt en multiplikativ model med ARIMA forecasts. Der kan estimeres en ARIMA-model ud fra data. Serien indeholder ingen påskeeffekt, men en klar handelsdagseffekt.

Figur 6.1.1.3.2



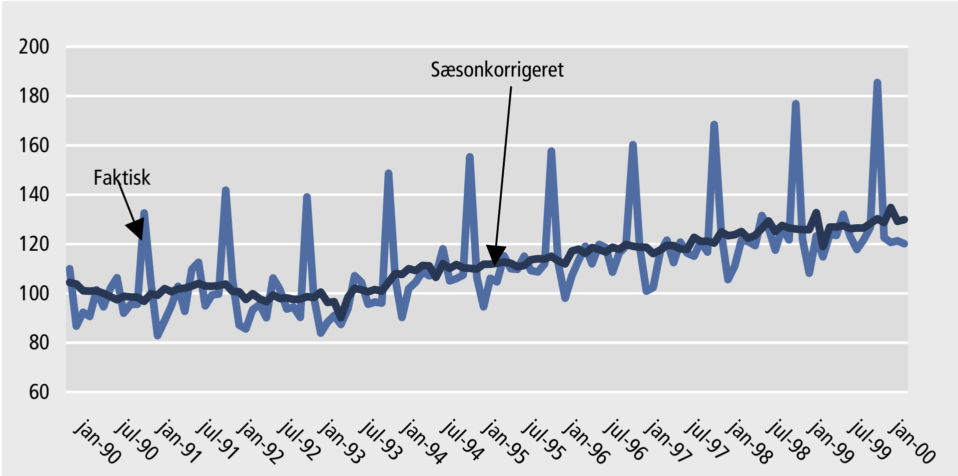
Ovenstående figur afbilder den estimerede sæsonfaktor imod den estimerede irregulære faktor. Denne figur viser både forholdet mellem de to størrelser og om der er estimeret et stabilt sæsonmønster. Figuren underbygger de 11 komponentteststørrelser visuelt. Af figuren ses det, at der er estimeret et stabilt sæsonmønster i serien.

Konklusion

Sammenholdes resultaterne fra sæsonkorrigeringen ses det, at kvaliteten af sæsonkorrigeringen af detailomsætningsindekset for Beklædning mv. er god. Første kvalitetsindikator Q var mindre end 1, og ovenstående figur viste et stabilt sæsonmønster og ikke-dominerende estimerede irregulære faktorer.

6.1.1.4 Andre forbrugsvarer, detailomsætningsindekset

Figur 6.1.1.4.1



Kvaliteten af sæsonkorrigeringen

Sæsonkorrigeringen af detailomsætningsindekset for andre forbrugsvarer viser, at serien er velegnet til sæsonkorrigering. Den sammenvægtede teststørrelse Q er lig 0,32.

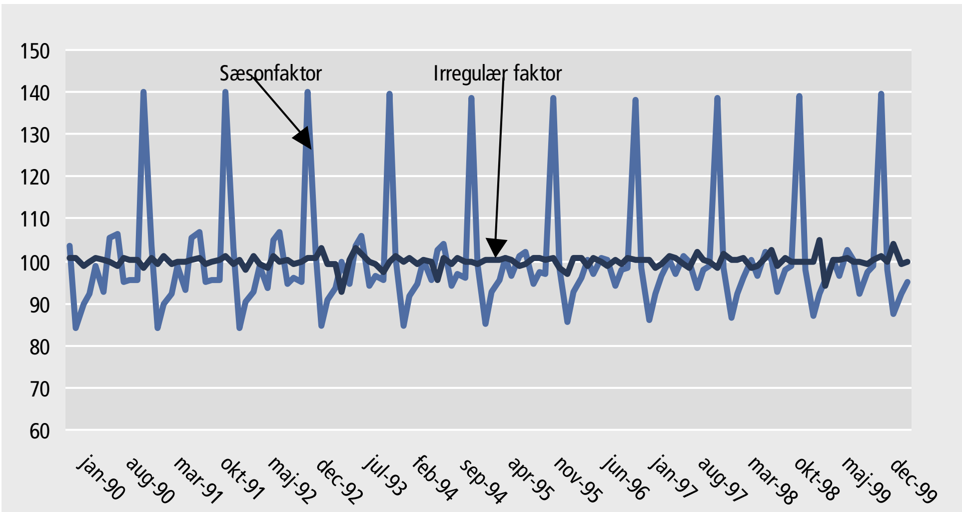
Komponenttest-størrelserne

Af de 11 komponentteststørrelser er der ingen problematiske.

Om serien

Serien er en månedsserie observeret over perioden januar 1990 til april 2000, og der er anvendt en multiplikativ model med ARIMA forecasts. Der kan estimeres en ARIMA-model ud fra data. Serien indeholder ingen påskeeffekt, men en klar handelsdagseffekt.

Figur 6.1.1.4.2



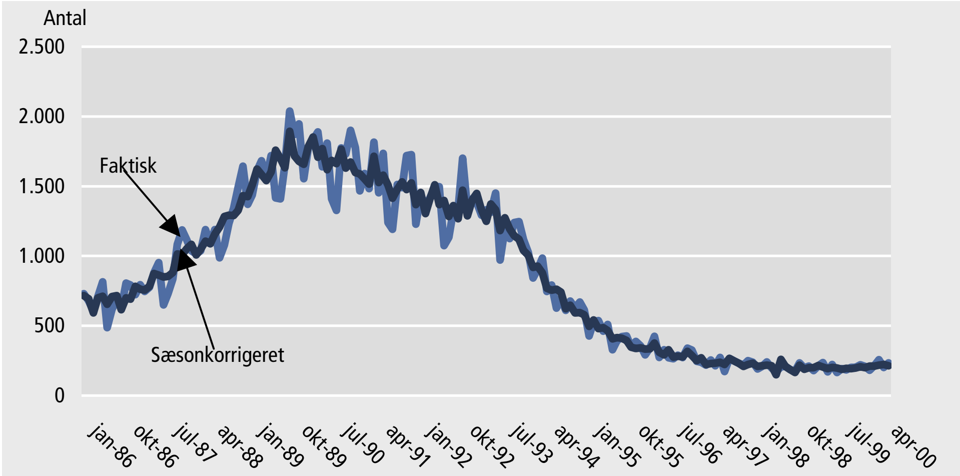
Ovenstående figur afbilder den estimerede sæsonfaktor mod den estimerede irregulære faktor. Denne figur viser både forholdet mellem de to størrelser og om der er estimeret et stabilt sæsonmønster. Figuren underbygger de 11 komponentteststørrelser visuelt. Af figuren ses det, at der er estimeret et stabilt sæsonmønster i serien.

Konklusion

Sammenholdes resultaterne fra sæsonkorrigeringen ses det, at kvaliteten af sæsonkorrigeringen af detailomsætningsindekset for andre forbrugsvarer er god. Første kvalitetsindikator Q var mindre end 1, og ovenstående figur viste et stabilt sæsonmønster og ikke-dominerende estimerede irregulære faktorer.

6.1.1.5 Kundgjorte tvangsauktioner

Figur 6.1.1.5.1



Kvaliteten af sæsonkorrigeringen

Sæsonkorrigeringen af de kundgjort tvangsauktioner viser, at serien er velegnet til sæsonkorrigering. Den sammenvægtede teststørrelse Q er lig 0,55.

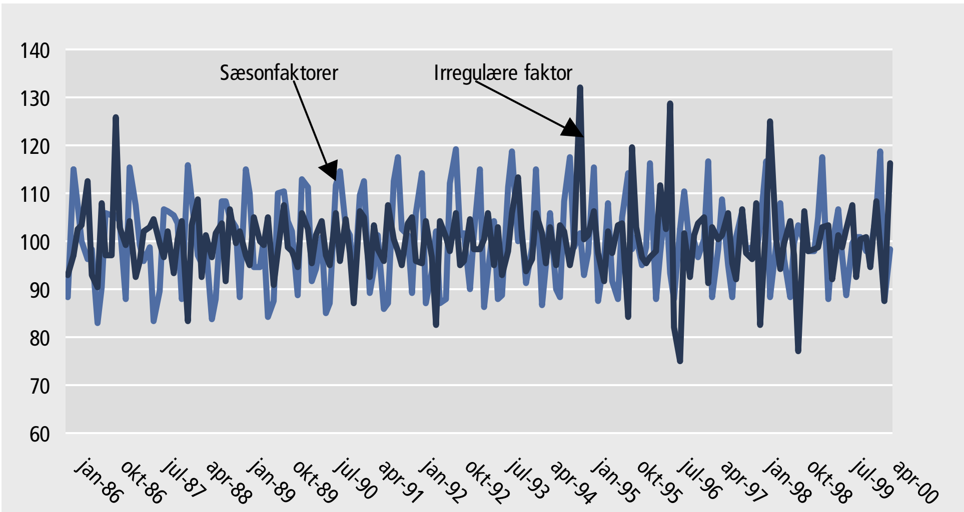
Komponenttest-størrelserne

Af de 11 komponentteststørrelser er blot M1 problematisk.

Om serien

Serien er en månedsserie observeret over perioden januar 1986 til april 2000, og der er anvendt en multiplikativ model med ARIMA forecasts. Der kan ikke estimeres en ARIMA-model ud fra data. Serien indeholder ingen påskeeffekt, men en klar handelsdagseffekt.

Figur 6.1.1.5.2



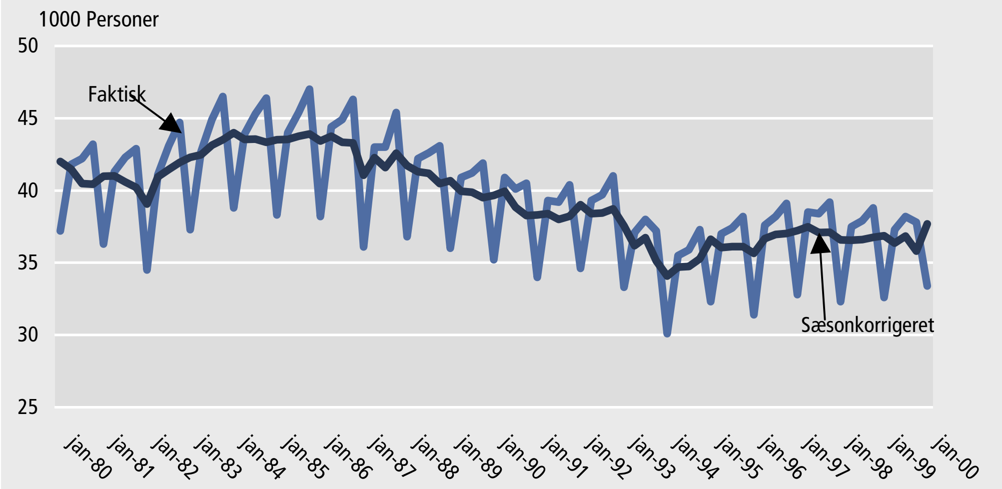
Ovenstående figur afbilder den estimerede sæsonfaktor mod den estimerede irregulære faktor. Denne figur viser både forholdet mellem de to størrelser og om der er estimeret et stabilt sæsonmønster, figuren underbygger de 11 komponentteststørrelser visuelt. Af figuren ses det, at der er estimeret et stabilt sæsonmønster i serien.

Konklusion

Sammenholdes resultaterne fra sæsonkorrigeringen ses det, at kvaliteten af sæsonkorrigeringen af de kundgjort tvangsauktioner er god. Første kvalitetsindikator Q var mindre end 1, og ovenstående figur viste et stabilt sæsonmønster og ikke-dominerende estimerede irregulære faktorer.

6.1.1.6 Atp-beskæftigelsen for landbrug, fiskeri og råstofudvinding

Figur 6.1.1.6.1



Kvaliteten af sæsonkorrigeringen

Sæsonkorrigeringen af atp-beskæftigelsen for landbrug mv. viser, at serien er velegnet til sæsonkorrigering. Den sammenvægtede teststørrelse Q er lig 0,29.

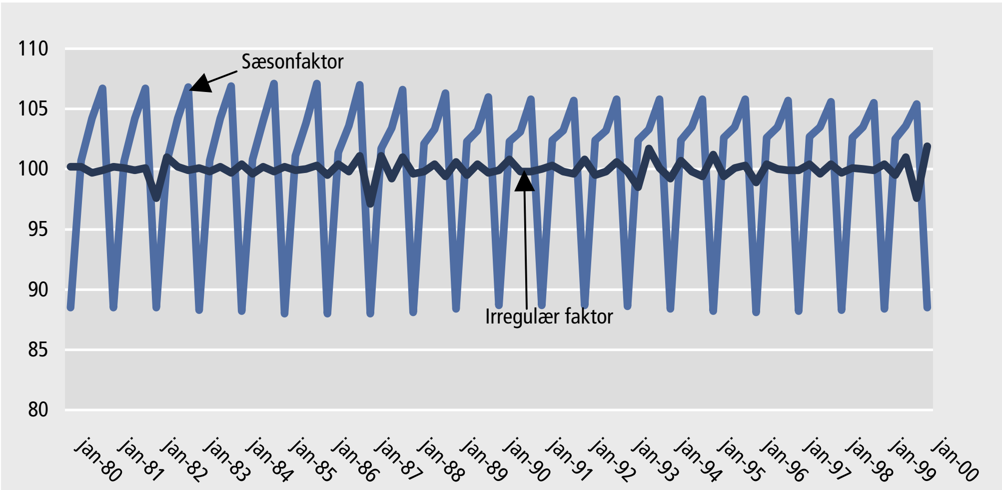
Komponentteststørrelserne

Af de 11 komponentteststørrelser er ingen problematiske.

Om serien

Serien er en kvartalsserie observeret over perioden 1. kvartal 1980 til 2. kvartal 2000, og der er anvendt en multiplikativ model med ARIMA forecasts. Der kan ikke estimeres en ARIMA-model ud fra data.

Figur 6.1.1.6.2



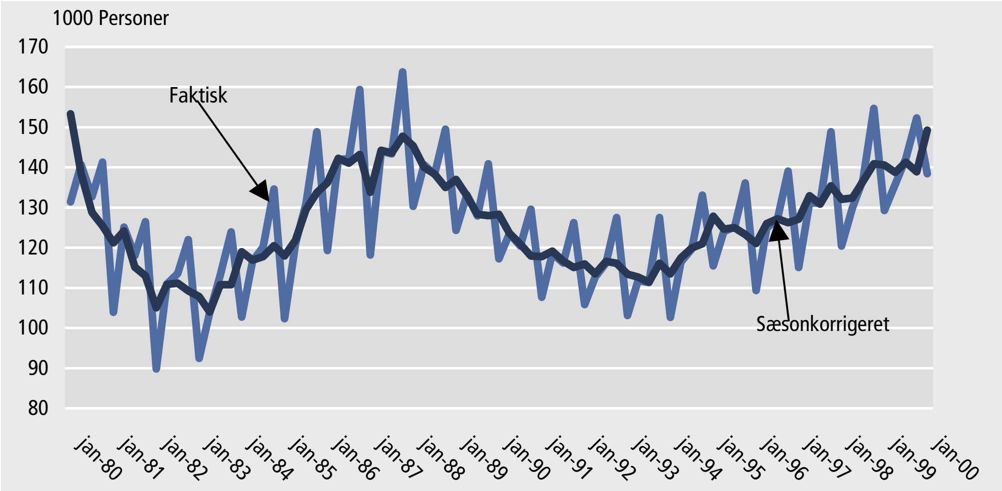
Ovenstående figur afbilder den estimerede sæsonfaktor mod den estimerede irregulære faktor. Denne figur viser både forholdet mellem de to størrelser og om der er estimeret et stabilt sæsonmønster, figuren underbygger de 11 komponentteststørrelser visuelt. Af figuren ses det, at der er estimeret et rimeligt stabilt sæsonmønster i serien.

Konklusion

Sammenholdes resultaterne fra sæsonkorrigeringen ses det, at kvaliteten af sæsonkorrigeringen af atp-beskæftigelsen for landbrug mv. er god. Første kvalitetsindikator Q var mindre end 1, og ovenstående figur viste et stabilt sæsonmønster og ikke-dominerende estimerede irregulære faktorer.

6.1.1.7 ATP-beskæftigelsestal for bygge- og anlægsvirksomhed

Figur 6.1.1.7.1



Kvaliteten af sæsonkorrigeringen

Sæsonkorrigeringen af atp-beskæftigelsen for bygge og anlægsvirksomhed viser, at serien er velegnet til sæsonkorrigering. Den sammenvægtede teststørrelse Q er lig 0,39.

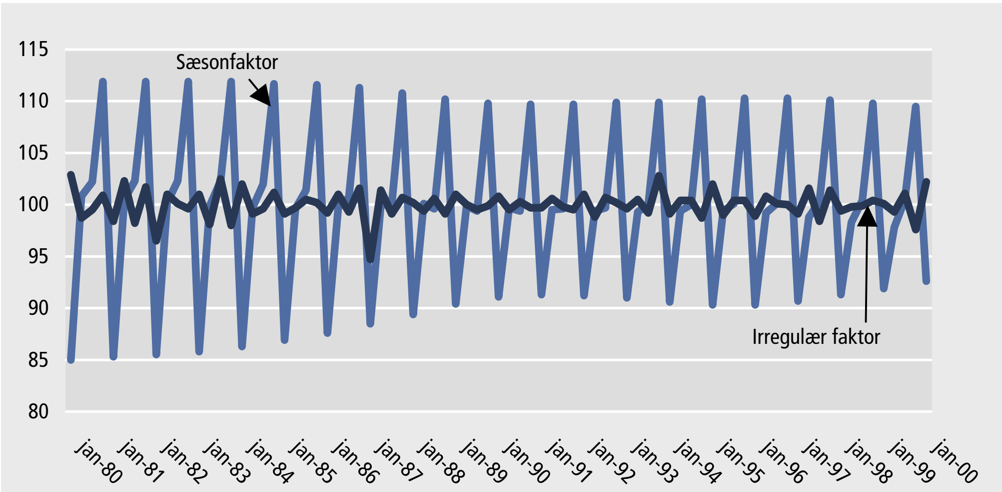
Komponenttest-størrelserne

Af de 11 komponentteststørrelser er ingen problematiske.

Om serien

Serien er en kvartalsserie observeret over perioden 1. kvartal 1980 til 2. kvartal 2000, og der er anvendt en multiplikativ model med ARIMA forecasts. Der kan ikke estimeres en ARIMA-model ud fra data.

Figur 6.1.1.7.2



Ovenstående figur afbilder den estimerede sæsonfaktor mod den estimerede irregulære faktor. Denne figur viser både forholdet mellem de to størrelser og om der er estimeret et stabilt sæsonmønster, figuren underbygger de 11 komponentteststørrelser visuelt.

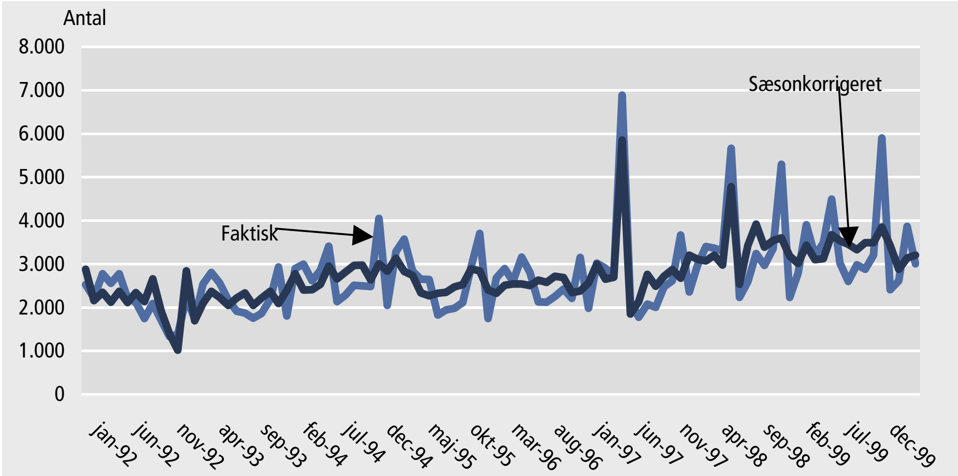
Konklusion

Af figuren ses det, at der er estimeret et rimeligt stabilt sæsonmønster i serien, hvilket vil sige at sæsonkorrigeringen må karakteriseres som rimelig.

6.1.2 Serier med store irregulære komponenter

6.1.2.1 Nyregistrerede biler i erhvervene

Figur 6.1.2.1.1



Kvaliteten af sæsonkorrigeringen

Sæsonkorrigeringen af de nyregistrerede biler i erhvervene viser, at serien umiddelbart ikke er velegnet til sæsonkorrigering, idet den sammenvejede Q-teststørrelse er lig 1,37.

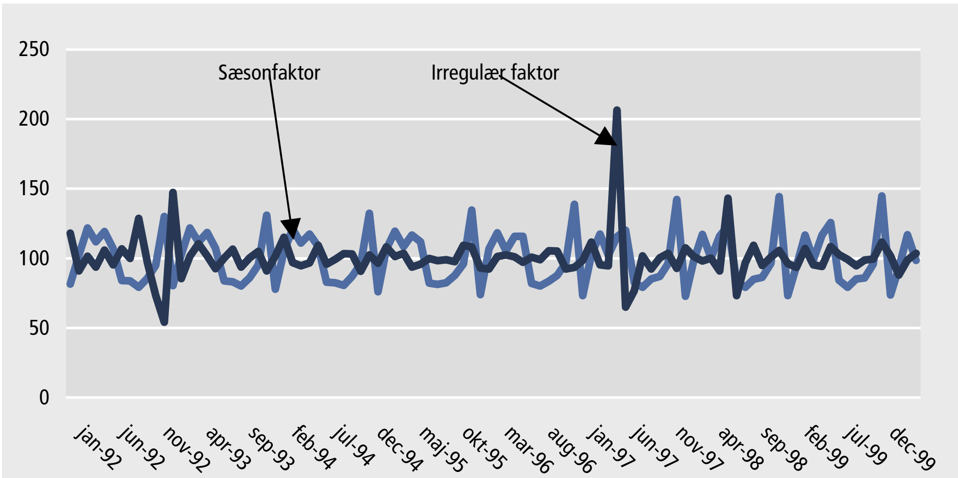
Komponenttest-størrelserne

Af de 11 komponentstørrelser ser vi, at M1,M2,M3 og M5 alle er større end 1. Disse teststørrelser er store, primært fordi den irregulære faktor er stor set i forhold til sæsonkomponenten.

Om serien

Serien er en månedsserie observeret over perioden januar 1992 til april 2000 og der er anvendt en multiplikativ model med ARIMA forecasts. Der kan dog ikke estimeres en ARIMA-model ud fra data. Serien indeholder ingen påskeeffekt, men en klar handelsdagseffekt.

Figur 6.1.2.1.2



Af ovenstående figur ser vi, at den estimerede irregulære faktor er relativ stor set i forhold til den estimerede sæsonfaktor, som vi konkluderede da vi betragtede M teststørrelserne ovenfor.

Konklusion

Serien nyregistrerede biler i erhvervene er ikke velegnet til sæsonkorrigering når hele observationsperioden benyttes.

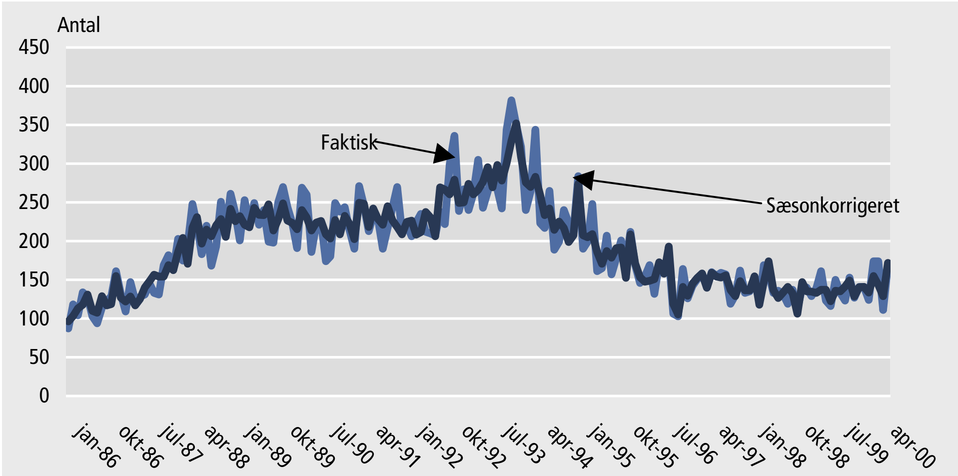
Løsningsforslag

Der synes at være et skift i det estimerede sæsonmønster i 1996-1997, hvilket skyldes et skift i skattelovgivningen. Der er i 1996-1997 sket en udefra kommende påvirkning af bilsalget til erhvervene, idet skattereglerne for køb af firmabiler blev omlagt i denne periode. Denne forklaring kunne give anledning til et skift i det estimerede sæsonmønster. Et sådant skift vil også påvirke Q til at blive større. For at tage højde for dette skift, vil det være oplagt at forsøge sæsonkorrigering på en kortere serie, nemlig en serie for 1996 og frem, hvor tallene alle henfører til den nye skattelovgivning.

I princippet er serien altså alt i alt velegnet til sæsonkorrigering, blot bør man forsøge med en kortere serie, for at tage højde for skatteændringer.

6.1.2.2 Erklærede konkurser

Figur 6.1.2.2.1



Kvaliteten af sæsonkorrigeringen

Sæsonkorrigeringen af de erklærede konkurser viser uniddelbart, at serien er knap så velegnet til sæsonkorrigering. Den sammenvægtede teststørrelse Q er lig 1,08.

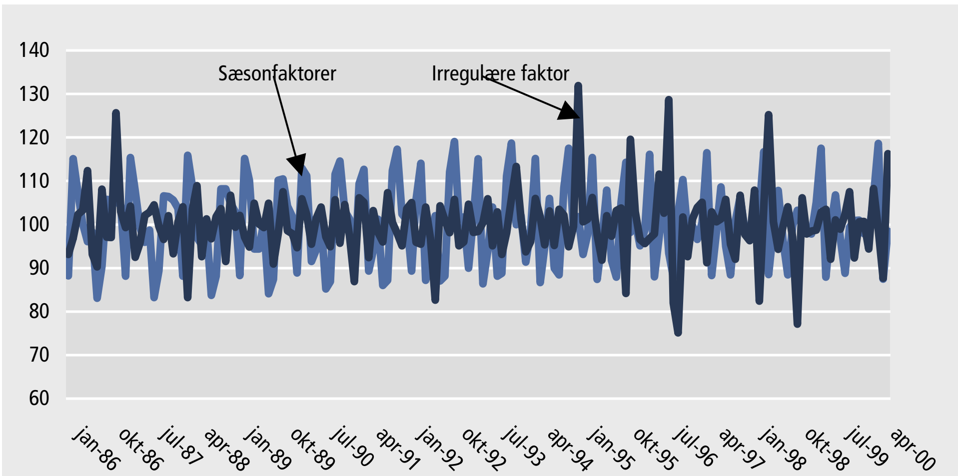
Komponenttest-størrelserne

Af de 11 komponentteststørrelser er M1, M3, M8, M10 og M11 problematiske.

Om serien

Serien er en månedsserie observeret over perioden januar 1986 til maj 2000, og der er anvendt en multiplikativ model med ARIMA forecasts.
Der kan estimeres en ARIMA-model ud fra data. Serien indeholder både en påskeeffekt og en handelsdagseffekt.

Figur 6.1.2.2.2



Ovenstående figur afbilder den estimerede sæsonfaktor mod den estimerede irregulære faktor. Denne figur viser både forholdet mellem de to størrelser og om der er estimeret et stabilt sæsonmønster. Figuren underbygger de 11 komponentteststørrelser visuelt.

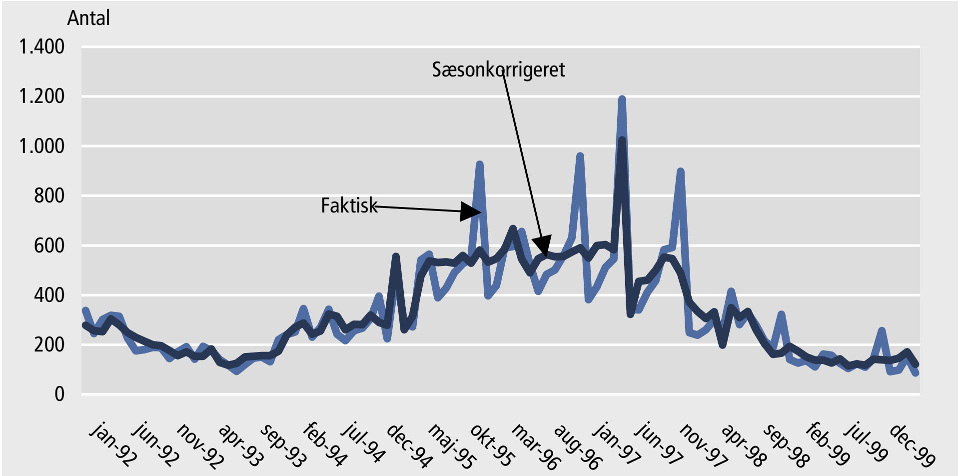
Konklusion

Af figuren ovenfor ses det, at der er estimeret et rimeligt stabilt sæsonmønster i serien, men at den irregulære faktor er høj set i forhold til den estimerede sæsonfaktor. Dette giver sig udslag i komponentteststørrelserne, og dermed i Q.
Alt i alt vil man således sige, at sæsonkorrigeringen af de erklærede konkurser må karakteriseres som god på trods af at Q teststørrelsen er højere end 1.

6.1.3 Serier der bør undersøges nærmere

6.1.3.1 Uidentificerede bilkøb

Figur 6.1.3.1.1



Kvaliteten af sæsonkorrigeringen

Serien uidentificerede bilkøb er en slags restserie, om hvilken vi ikke ved meget. Det er derfor forventeligt, at denne serie ikke umiddelbart er let at sæsonkorrigere. Serien har en Q-teststørrelse på 1,12, og er dermed i princippet en problemserie. Førend den endelige konklusion drages, bør man dog betragte hvilke af M-teststørrelserne der er signifikante.

Samtidig bør det undersøges om de samme konklusioner kan drages af de afbildede estimerede sæsonfaktorer mod de estimerede irregulære faktorer

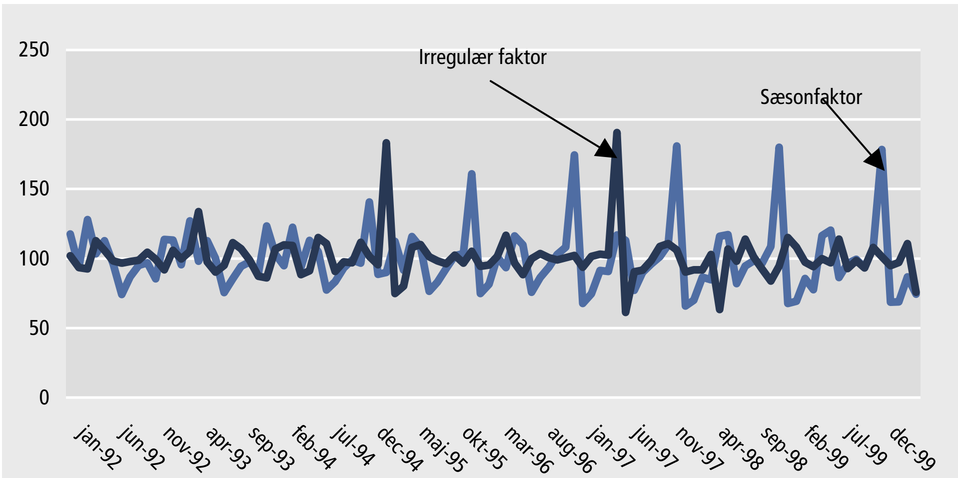
Komponenttest-størrelserne

Af M-teststørrelserne er M1, M7,…,M11 alle større end 1.
Dette indikerer tydeligvis helt generelt problemer i serien, som indeholder både store irregulære faktorer og et ikke-stabilt sæsonmønster.

Om serien

Serien er en månedsserie observeret over perioden januar 1992 til april 2000 og der er anvendt en multiplikativ model med ARIMA forecasts.
Der kan ikke estimeres en ARIMA-model ud fra data. Serien indeholder hverken en påskeeffekt, eller en handelsdagseffekt.

Figur 6.1.3.1.2



Af ovenstående figur ser vi, at som M-teststørrelserne indikerede, er der problemer i serien både med hensyn til store irregulære faktorer og den mest problematiske faktor, nemlig et ustabilt sæsonmønster.

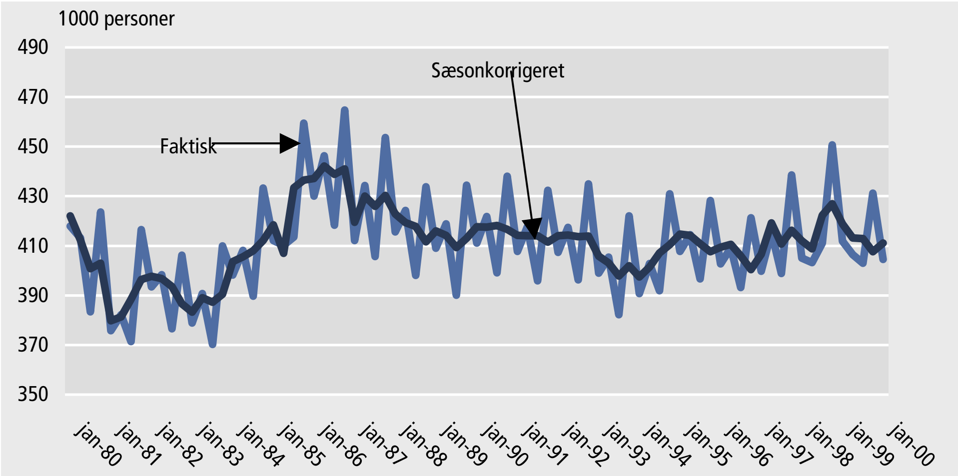
Man bør dog bemærke, at seriens niveau er lavt i forhold til de to andre serier for nyregistrerede biler.
Serien udgør omkring 4 pct. af det samlede antal nyregistreringer.

Konklusion

I sin fulde længde er serien ikke velegnet til sæsonkorrigering. Dette problem kunne man forsøge at løse ved at forkorte serien på samme sted som serien for erhvervene.

6.1.3.2 ATP-beskæftigelsestal for industrien

Figur 6.1.3.2.1



Kvaliteten af sæsonkorrigeringen

Sæsonkorrigeringen af atp-beskæftigelsen for industri viser, at serien er velegnet til sæsonkorrigering. Den sammenvægtede teststørrelse Q er lig 0,33.

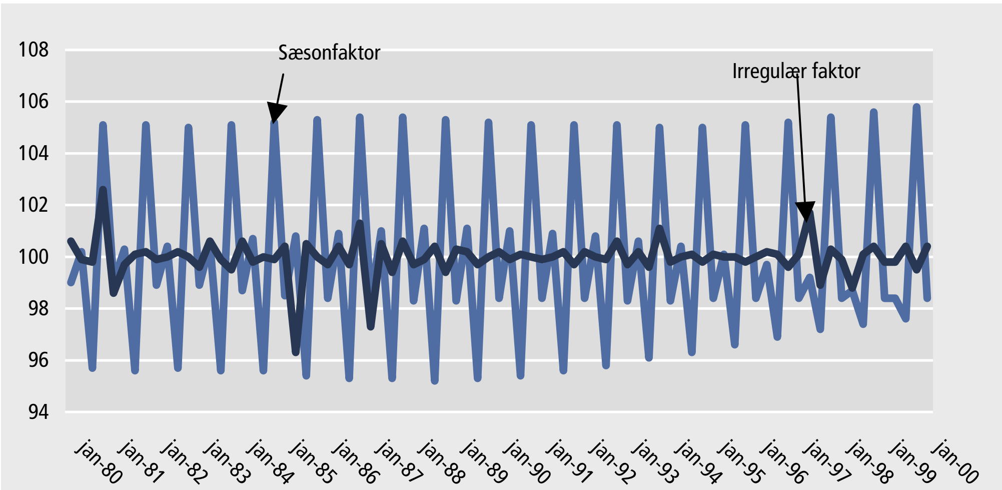
Komponenttest-størrelserne

Af de 11 komponentteststørrelser er ingen problematiske.

Om serien

Serien er en kvartalsserie observeret over perioden 1. kvartal 1980 til 2. kvartal 2000, og der er anvendt en multiplikativ model med ARIMA forecasts. Der kan ikke estimeres en ARIMA-model ud fra data.

Figur 6.1.3.2.2



Ovenstående figur afbilder den estimerede sæsonfaktor mod den estimerede irregulære faktor. Denne figur viser både forholdet mellem de to størrelser og om der er estimeret et stabilt sæsonmønster. Figuren underbygger de 11 komponentteststørrelser visuelt.

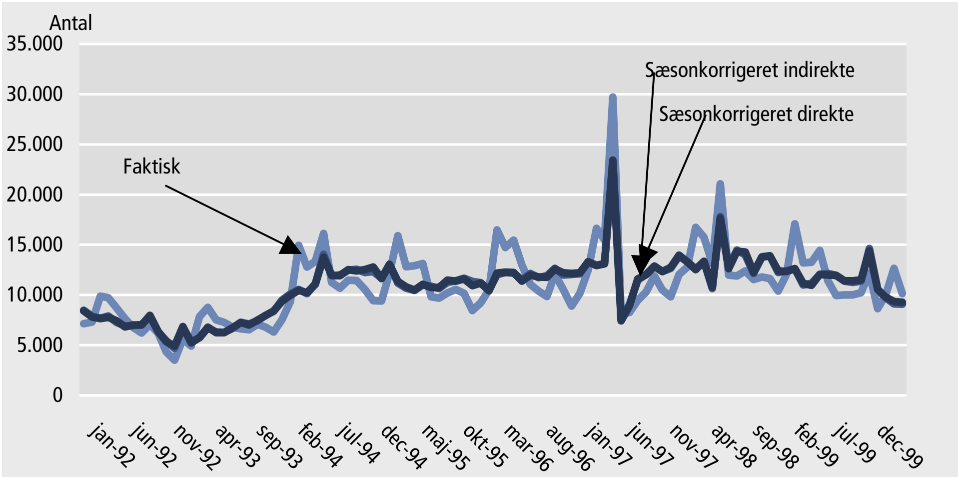
Konklusion

Af ovenstående figur ses det, at der ikke er estimeret et stabilt sæsonmønster i serien, hvilket vil sige at sammenholdt med den lave Q-teststørrelse må kvaliteten af sæsonkorrigeringen af atp-beskæftigelsen for industri undersøges nærmere.

6.1.4 Direkte eller indirekte sæsonkorrigering

6.1.4.1 Nyregistrerede biler, i alt

Figur 6.1.4.1.1



Serien 'Nyregistrerede biler i alt' er sæsonkorrigeret både direkte og indirekte.

Serien og dens delkomponenter

Serien 'Nyregistrerede biler i alt' er dannet som en sum af de tre delkomponenter:

1. nyregistrerede biler i husholdningerne,
2. nyregistrerede biler i erhvervene
3. uidentificeret bilkøb

Direkte og indirekte sæsonkorrigering

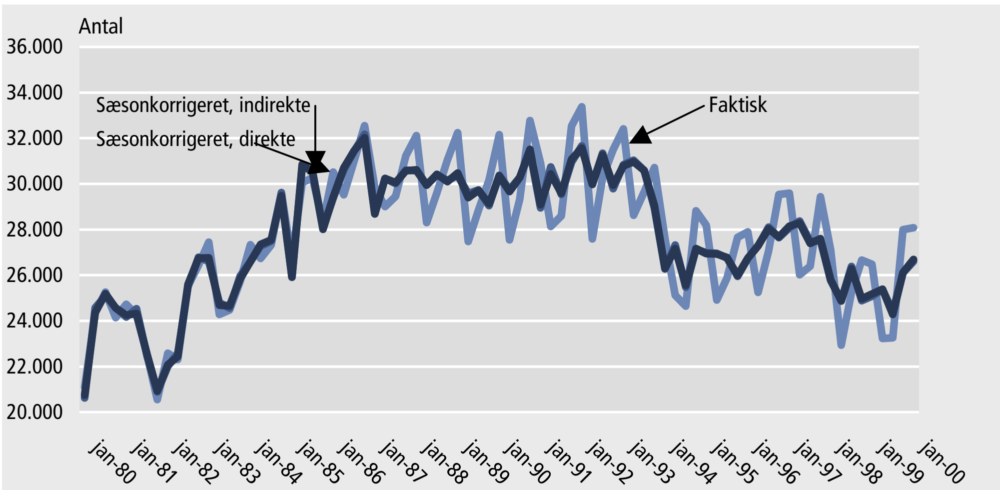
Den direkte sæsonkorrigering er sæsonkorrigeringen af denne sumserie. Den indirekte sæsonkorrigerede serie er lig summen af de sæsonkorrigerede delkomponenter.

Konklusion

Det ses af ovenstående figur at der ingen forskel er mellem den direkte sæsonkorrigerede serie og den indirekte sæsonkorrigerede serie.

6.1.4.2 Indbrud i alt

Figur 6.1.4.2.1



Serien 'Indbrud i alt' er sæsonkorrigeret både direkte og indirekte.

Serien og dens delkomponenter

Serien 'Indbrud i alt' er dannet som en sum af de tre delkomponenter:

1. Indbrud banker mv.
2. indbrud villaer mv.
3. indbrud fritidshuse mv.

Direkte og indirekte sæsonkorrigering

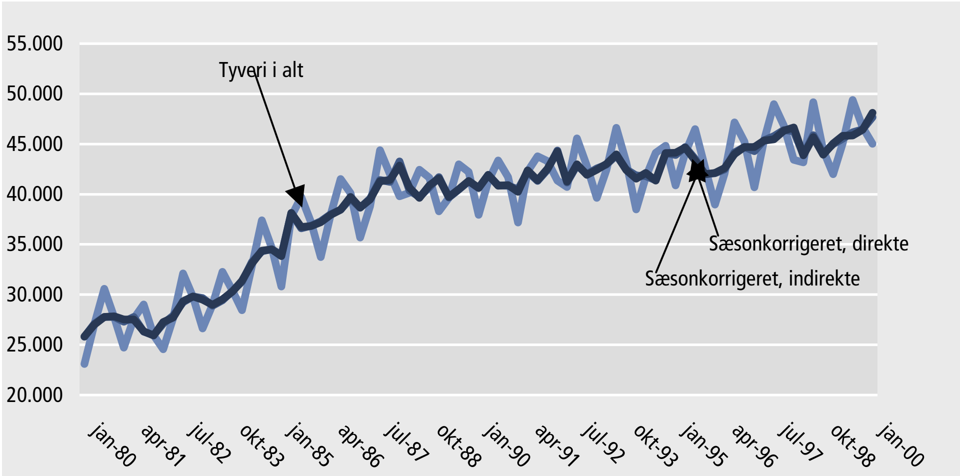
Den direkte sæsonkorrigering fås blot som den sæsonkorrigerede sumserie, mens den indirekte sæsonkorrigerede serie er summen af de sæsonkorrigerede delkomponenter.

Konklusion

Det ses af figuren, at der ingen forskel er mellem den direkte og den indirekte sæsonkorrigerede serie.

6.1.4.3 Tyveri i alt

Figur 6.1.4.3.1



Serien 'Tyveri i alt' er sæsonkorrigeret både direkte og indirekte, og resultaterne ses i figuren ovenfor.

Serien og dens delkomponenter

Serien er dannet som en sum af de tre delkomponenter:

1. Tyveri fra bil
2. butikstyverier
3. tyveri i øvrigt

Direkte og indirekte sæsonkorrigering

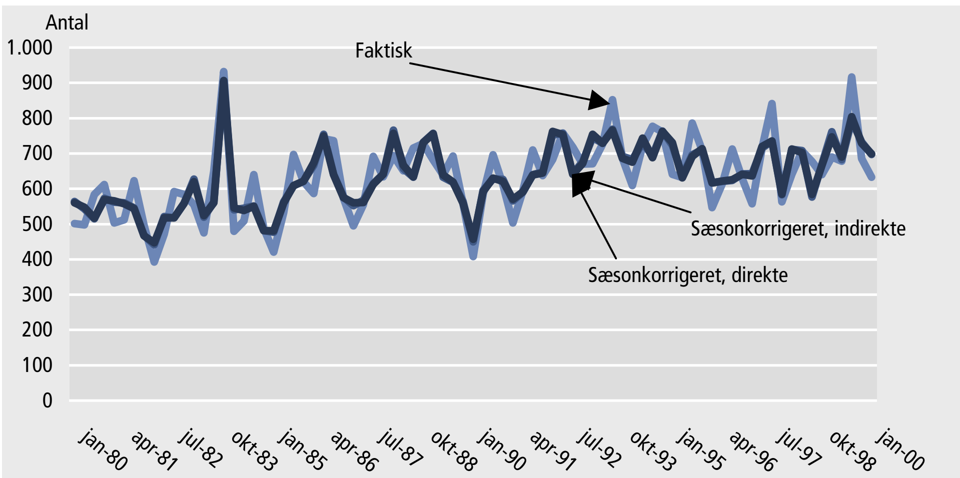
Den direkte sæsonkorrigering fås blot som den sæsonkorrigerede sumserie, mens den indirekte sæsonkorrigerede serie er summen af de sæsonkorrigerede delkomponenter.

Konklusion

Det ses af figuren, at der ingen forskel er mellem den direkte og den indirekte sæsonkorrigerede serie.

6.1.4.4 Sædelighedsforbrydelser i alt

Figur 6.1.4.4.1



Serien 'Sædelighedsforbrydelser i alt' er sæsonkorrigeret både direkte og indirekte og resultaterne ses i figuren ovenfor.

Serien og dens delkomponenter

Serien er dannet som en sum af de tre delkomponenter

1. Voldtægt mv.
2. blufærdighedskrænkelse
3. andre sædelighedsforbrydelser

Direkte og indirekte sæsonkorrigering

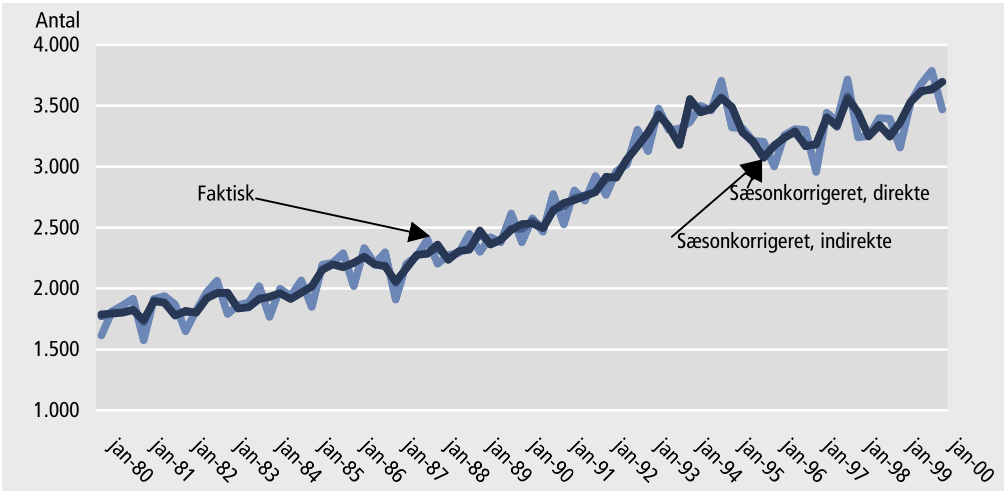
Den direkte sæsonkorrigering fås blot som den sæsonkorrigerede sumserie, mens den indirekte sæsonkorrigerede serie er summen af de sæsonkorrigerede delkomponenter.

Konklusion

Det ses af figuren, at der ingen forskel er mellem den direkte og den indirekte sæsonkorrigerede serie.

6.1.4.5 Voldsforbrydelser i alt

Figur 6.1.4.5.1



Serien 'Voldsforbrydelser i alt' er sæsonkorrigeret både direkte og indirekte og resultaterne ses i figuren ovenfor.

Serien og dens delkomponenter

Serien er dannet som en sum af de fem delkomponenter

1. Vold imod offentlig myndighed
2. drab og forsøg herpå
3. vold imod privatperson
4. trusler
5. andre voldsforbrydelser

Direkte og indirekte sæsonkorrigering

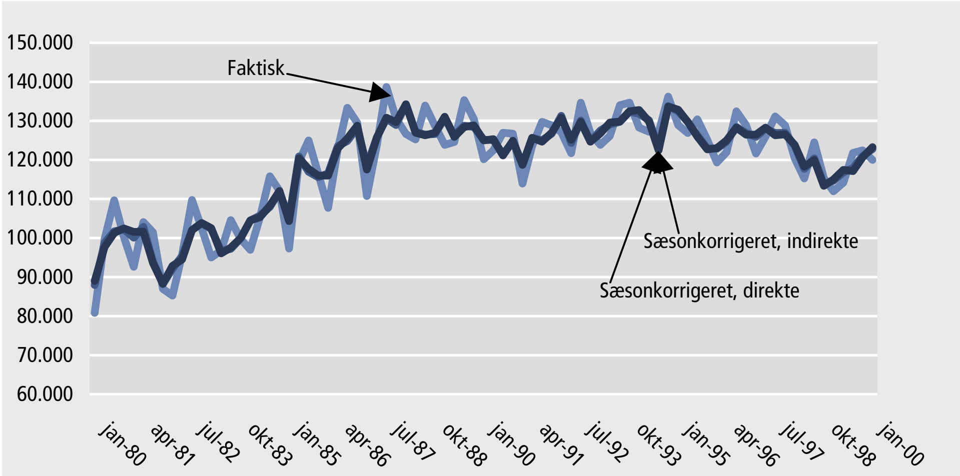
Den direkte sæsonkorrigering fås blot som den sæsonkorrigerede sumserie, mens den indirekte sæsonkorrigerede serie er summen af de sæsonkorrigerede delkomponenter.

Konklusion

Det ses af figuren, at der ingen forskel er mellem den direkte og den indirekte sæsonkorrigerede serie.

6.1.4.6 Ejendomsforbrydelser i alt

Figur 6.1.4.6



Serien 'Ejendomsforbrydelser i alt' er sæsonkorrigeret både direkte og indirekte, resultaterne ses i figuren ovenfor.

Serien og dens delkomponenter

Serien er dannet som en sum af de 15 delkomponenter

1. Dokumentfalsk
2. Indbrud Banker, forretninger mv.
3. Indbrud, Villaer, lejligheder mv.
4. Indbrud, Fritidshuse, garager mv.
5. Tyveri fra bil
6. Butikstyverier mv.
7. Tyveri i øvrigt
8. Tyv/brugstyveri af bil
9. Tyv/brugstyveri af motorcykel/knallert
10. Tyv/brugstyveri af cykel
11. Bedrageri mv.
12. Hæleri
13. Røveri
14. Hærværk
15. andre ejendomsforbrydelser

Direkte og indirekte sæsonkorrigering

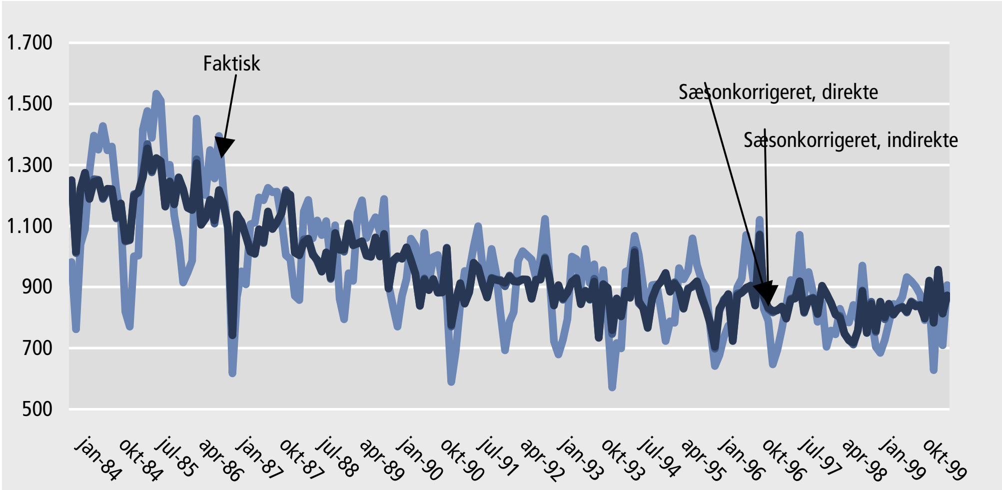
Den direkte sæsonkorrigering fås blot som den sæsonkorrigerede sumserie, mens den indirekte sæsonkorrigerede serie er summen af de sæsonkorrigerede delkomponenter.

Konklusion

Det ses af figuren, at der ingen forskel er mellem den direkte og den indirekte sæsonkorrigerede serie.

6.1.4.7 Skader i personbiler i alt

Figur 6.1.4.7.1



Serien 'Skader i personbiler i alt' er sæsonkorrigeret både direkte og indirekte og resultaterne ses i figuren ovenfor.

Serien og dens delkomponenter

Serien er dannet som en sum af de to delkomponenter

1. Dræbte i personbiler i alt
2. tilskadekomne i personbiler i alt

Direkte og indirekte sæsonkorrigering

Den direkte sæsonkorrigering fås blot som den sæsonkorrigerede sumserie, mens den indirekte sæsonkorrigerede serie er summen af de sæsonkorrigerede delkomponenter.

Konklusion

Det ses af figuren, at der ingen forskel er mellem den direkte og den indirekte sæsonkorrigerede serie.

7. Fremtiden

Det er altid svært at spå om fremtiden. Hvad der vil ske fremover på sæsonkorrigeringsområdet er på ingen måde klart.

7.1 I Danmarks Statistik

Arbejdsgruppe anbefaler X-12...

Danmarks Statistik har haft nedsat en arbejdsgruppe i sæsonkorrigering. Gruppen har anbefalet, at Danmarks Statistik skifter sæsonkorrigeringsmetode til X-12.

... med to hovedbegrundelser

Dette skyldes primært to faktorer: For det første er X-12 et bedre sæsonkorrigeringsprogram, og for det andet er der udviklet den brugervenlige windowsflade Demetra til programmet.

Ønske om brugervenlighed

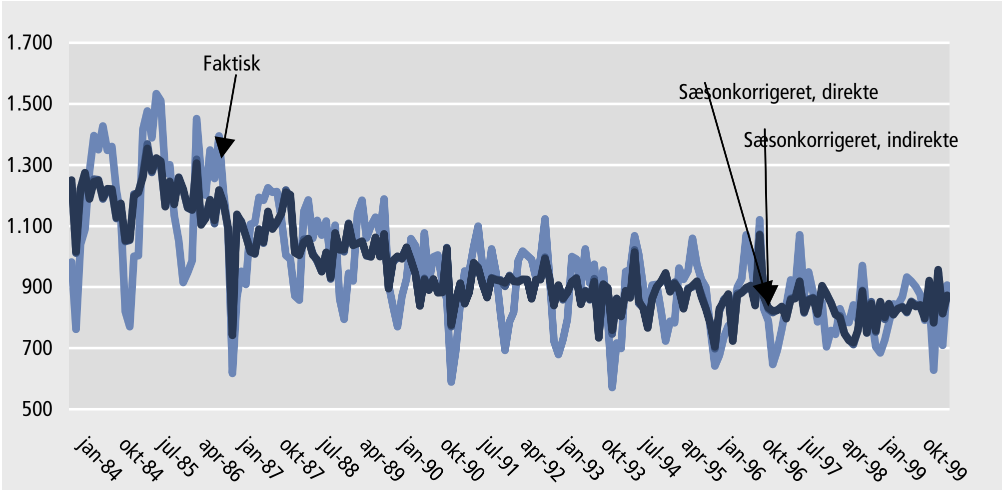
Det er et udtalt ønske for de, der arbejder med sæsonkorrigering, at få et brugervenligt værktøj.

7.2 I internationale sammenhænge

Eurostat har i en Task Force arbejdet med harmonisering af sæsonkorrigeringsmetoder. Denne gruppes arbejde har resulteret i en række anbefalinger på området. Eurostat har opdelt deres anbefalinger i 11 forskellige punkter, og disse anbefalinger er beskrevet nedenfor.

6.1.4.7 Skader i personbiler i alt

Figur 6.1.4.7.1



Serien 'Skader i personbiler i alt' er sæsonkorrigeret både direkte og indirekte og resultaterne ses i figuren ovenfor.

Serien og dens delkomponenter

Serien er dannet som en sum af de to delkomponenter

1. Dræbte i personbiler i alt
2. tilskadekomne i personbiler i alt

Direkte og indirekte sæsonkorrigering

Den direkte sæsonkorrigering fås blot som den sæsonkorrigerede sumserie, mens den indirekte sæsonkorrigerede serie er summen af de sæsonkorrigerede delkomponenter.

Konklusion

Det ses af figuren, at der ingen forskel er mellem den direkte og den indirekte sæsonkorrigerede serie.

7. Fremtiden

Det er altid svært at spå om fremtiden. Hvad der vil ske fremover på sæsonkorrigeringsområdet er på ingen måde klart.

7.1 I Danmarks Statistik

Arbejdsgruppe anbefaler X-12…

Danmarks Statistik har haft nedsat en arbejdsgruppe i sæsonkorrigering. Gruppen har anbefalet, at Danmarks Statistik skifter sæsonkorrigeringsmetode til X-12.

… med to hovedbegrundelser

Dette skyldes primært to faktorer: For det første er X-12 et bedre sæsonkorrigeringsprogram, og for det andet er der udviklet den brugervenlige windowsflade Demetra til programmet.

Ønske om brugervenlighed

Det er et udtalt ønske for de, der arbejder med sæsonkorrigering, at få et brugervenligt værktøj.

7.2 I internationale sammenhænge

Eurostat har i en Task Force arbejdet med harmonisering af sæsonkorrigeringsmetoder. Denne gruppes arbejde har resulteret i en række anbefalinger på området. Eurostat har opdelt deres anbefalinger i 11 forskellige punkter, og disse anbefalinger er beskrevet nedenfor.

1. **Valg af sæsonkorrigeringsmetode.** Eurostat ønsker at harmonisere valget af sæsonkorrigeringsmetoder, således at alle EU-lande bruger samme metode. Der opstilles en række teoretiske kriterier samt en række empiriske kriterier til brug ved vurderingen af de forskellige tilgængelige metoders kvaliteter. På basis af disse kriterier anbefaler Eurostat, at de nationale statistikbureauer anvender modelbaserede sæsonkorrigeringsmetoder: Man anbefaler specielt TRAMO/SEATS. Man ønsker som en følge heraf, at andre metoder som er i brug i de nationale statistikbureauer trækkes tilbage så hurtigt som muligt, og at man derefter implementerer TRAMO/SEATS.
2. **Ændring af sæsonkorrigeringsmetode.** Som udgangspunkt er ændring af sæsonkorrigeringsmetode en ulempe for alle. Derfor bør ændringer af metoden begrænses mest muligt.
3. **Klarhed i de anvendte procedurer.** Sæsonkorrigering er en form for dataanalyse, hvis udfald afhænger stærkt af hvilken metode, der er valgt. Det er derfor vigtigt, at der er fuld åbenhed om analyserne og beslutningerne, som er baseret på sæsonkorrigeringen. Eurostat anbefaler derfor, at man ikke blot publicerer de sæsonkorrigerede tal, men også den bagvedliggende meta-information.
4. **Sæsonkorrigerede data fra medlemslandene.** Eurostat offentliggør sæsonkorrigerede data for de enkelte medlemslande. Således skal Eurostat beslutte, om de selv vil sæsonkorrigere medlemslandenes serier, eller om de skal vælge at bruge de sæsonkorrigerede serier, som kommer fra medlemslandene. Man bør vælge den af fremgangsmåderne, som sikrer den højeste kvalitet af det sæsonkorrigerede data. Sæsonkorrigering kræver typisk en detailviden om den enkelte serie, fx med hensyn til strejker, ferier med variabel beliggenhed og ekstreme observationer. Denne viden ligger i de nationale statistikbureauer og er normalt ikke tilgængelig i Eurostat. Derfor foretrækker Eurostat, at medlemslandene selv sæsonkorrigerer data. På den anden side er det Eurostats ansvar, at de data, som publiceres har en høj kvalitet. Set i dette lys kunne det være ønskværdigt, at Eurostat 'parallelkorrigerer' serierne fra medlemslande for bla. at sikre, at der er konsistens i serierne. Eurostat opdeler derfor inden for dette punkt sine anbefalinger i to trin. På kort sigt accepterer og publicerer Eurostat de sæsonkorrigerede serier fra medlemslandene. På lang sigt ønsker man, at denne arbejdsgruppe tager emnet kvalitetskontrol op, med henblik på en harmoniseret politik på området.
5. **Konsistens ved aggregering.** Man kan vælge at danne det sæsonkorrigerede aggregat som summen af de sæsonkorrigerede komponenter, altså indirekte sæsonkorrigering. Man kan ligeledes vælge at danne aggregatet som summen af de originale komponentserier, og derefter sæsonkorrigere summen, altså direkte sæsonkorrigering. Eurostat har forsøgt entydigt at anbefale direkte sæsonkorrigering, men der er udbredt uenighed både blandt brugere og eksperter i sæsonkorrigerede tal om, hvilken af de to fremgangsmåder der er at foretrække.
6. **Tidskonsistens.** Dette punkt drejer sig om overvejelserne vedrørende årsopregning - bør man foretage denne eksercits eller ej? Eurostat anbefaler, at man ikke årsopregner, fordi denne tidsmæssige restriktion - som kun lægges over et helt kalenderår - ingen videnskabelig dokumentation har. Man bør på sigt informere og vejlede statistikkens brugere i, at en sådan forskel er rimelig nok, når den forekommer.
7. **Publicerede tal, trend-cyklen eller sæsonkorrigerede data?** Originalserien dekomponeres i en trend-cykel komponent, en sæsonkomponent og en irregulær komponent. Trend-cyklen indeholder de langsigtede og cykliske bevægelser fra serien, hvorimod sæsonkomponenten fanger de regelmæssige mønstre som gentages over årene. Eurostat anbefaler, at de nationale statistikbureauer publicerer både trend-cykel data og sæsonkorrigeret data. Begge dele bør publiceres både som tabeller og figurer.
8. **Revisioner.** Når man sæsonkorrigerer, må man træffe et valg vedrørende parametrene i sæsonkorrigeringsproceduren: Skal de re-estimeres ved hver kørsel eller blot en gang om året? Eurostat anbefaler, at man benytter ARIMA forecasts til prediktion, og at man opdaterer sine forecasts, når der er nye observationer til-

gængelige. Forecast-modellen bør dog ikke ændres over kalenderåret. Ændring af modellen finder blot sted en gang årligt.

9. **Anvendelsen af sæsonkorrigeringsprogrammer.** De sæsonkorrigeringsmetoder som Eurostat anbefaler er, som tidligere beskrevet, de modelbaserede. Disse metoder er optimale, når de anvendes af eksperter. Eurostat understreger dog kraftigt, at man aldrig bør sæsonkorrigere efter automatpilot. Der er ingen vej uden om den selvstændige og kritiske vurdering af behandlingen af eksempelvis de ekstreme observationer, modelvalget og en eventuel datatransformation.
10. **Handelsdagskorrektioner.** Handelsdagskorrektioner er en vigtig del af dataanalysen, fordi et forskelligt antal af arbejdsdage i en given periode kan påvirke en tidsserie. Eurostat anbefaler, at handelsdagskorrektioner foretages når de er nødvendige, og at korrektionen foretages ved regression.
11. **Usikkerheder og konfidensintervaller.** Data som publiceres indeholder som regel usikkerheder. Usikkerhederne kan skyldes vidt forskellige ting såsom stikprøvens sammensætning, målefejl, manglende data eller fejlagtige respondentindberetninger. På baggrund af denne usikkerhed bør man gøre sig overvejelser om at formidle denne viden med brugerne. Eurostat anbefaler, at der til de sæsonkorrigerede tal offentliggøres standardafvigelser og konfidensintervaller.

Appendiks 1: Glidende gennemsnit

A1.1 Symmetriske glidende gennemsnit

Et glidende gennemsnit tages over elementer fra en tidsserie, kald denne $x_1,\ldots,x_n$, hvor $n \geq 1$. Lad x'erne være en tidsserie med månedlige observationer. Vi betragter nu udelukkende centrerede glidende gennemsnit, og et tremåneders (centreret) glidende gennemsnit har udseendet

$$\frac{x_{t-1}+x_t+x_{t+1}}{3}, \text{ for } 2 \leq t \leq n-1.$$

Et fire-måneders centreret glidende gennemsnit har følgende form;

$$\frac{\frac{1}{2}x_{t-2}+x_{t-1}+x_t+x_{t+1}+\frac{1}{2}x_{t+2}}{4}, \text{ for } 3 \leq t \leq n-2.$$

Man ser af ovenstående, at formlen for et m-måneders symmetrisk glidende gennemsnit er ret intuitiv og enkel, når m er ulige. Formlens udseende bliver lidt mere indviklet, når m er lige, dette gælder også for den generelt formulerede formel.

Et m-perioders symmetrisk glidende gennemsnit når **m er ulige**, har følgende udseende;

$$\frac{1}{m}\sum_{j=-(\frac{m-1}{2})}^{\frac{m-1}{2}} x_{t+j}, \text{ for } 1+\frac{m-1}{2} \leq t \leq n-\frac{m-1}{2}$$

Et m-perioders symmetrisk glidende gennemsnit når **m er lige**, har følgende udseende;

$$\frac{1}{m}\left[\frac{1}{2}(x_{t-\frac{m}{2}}+x_{t+\frac{m}{2}})+\sum_{j=1-\frac{m}{2}}^{\frac{m}{2}-1} x_{t+j}\right], \text{ for } 1+\frac{m}{2} \leq t \leq n-\frac{m}{2}$$

gængelige. Forecast-modellen bør dog ikke ændres over kalenderåret. Ændring af modellen finder blot sted en gang årligt.

9. **Anvendelsen af sæsonkorrigeringsprogrammer.** De sæsonkorrigeringsmetoder som Eurostat anbefaler er, som tidligere beskrevet, de modelbaserede. Disse metoder er optimale, når de anvendes af eksperter. Eurostat understreger dog kraftigt, at man aldrig bør sæsonkorrigere efter automatpilot. Der er ingen vej uden om den selvstændige og kritiske vurdering af behandlingen af eksempelvis de ekstreme observationer, modelvalget og en eventuel datatransformation.
10. **Handelsdagskorrektioner.** Handelsdagskorrektioner er en vigtig del af dataanalysen, fordi et forskelligt antal af arbejdsdage i en given periode kan påvirke en tidsserie. Eurostat anbefaler, at handelsdagskorrektioner foretages når de er nødvendige, og at korrektionen foretages ved regression.
11. **Usikkerheder og konfidensintervaller.** Data som publiceres indeholder som regel usikkerheder. Usikkerhederne kan skyldes vidt forskellige ting såsom stikprøvens sammensætning, målefejl, manglende data eller fejlagtige respondentindberetninger. På baggrund af denne usikkerhed bør man gøre sig overvejelser om at formidle denne viden med brugerne. Eurostat anbefaler, at der til de sæsonkorrigerede tal offentliggøres standardafvigelser og konfidensintervaller.

Appendiks 1: Glidende gennemsnit

A1.1 Symmetriske glidende gennemsnit

Et glidende gennemsnit tages over elementer fra en tidsserie, kald denne $x_1,\ldots,x_n$, hvor $n \geq 1$. Lad *x*'erne være en tidsserie med månedlige observationer. Vi betragter nu udelukkende centrerede glidende gennemsnit, og et tremåneders (centreret) glidende gennemsnit har udseendet

$$\frac{x_{t-1}+x_t+x_{t+1}}{3}, \text{ for } 2 \leq t \leq n-1.$$

Et fire-måneders centreret glidende gennemsnit har følgende form;

$$\frac{\frac{1}{2}x_{t-2}+x_{t-1}+x_t+x_{t+1}+\frac{1}{2}x_{t+2}}{4}, \text{ for } 3 \leq t \leq n-2.$$

Man ser af ovenstående, at formlen for et m-måneders symmetrisk glidende gennemsnit er ret intuitiv og enkel, når m er ulige. Formlens udseende bliver lidt mere indviklet, når m er lige, dette gælder også for den generelt formulerede formel.

Et m-perioders symmetrisk glidende gennemsnit når **m er ulige**, har følgende udseende;

$$\frac{1}{m}\sum_{j=-(\frac{m-1}{2})}^{\frac{m-1}{2}} x_{t+j}\ , \text{ for } 1+\frac{m-1}{2} \leq t \leq n-\frac{m-1}{2}$$

Et m-perioders symmetrisk glidende gennemsnit når **m er lige**, har følgende udseende;

$$\frac{1}{m}\left[\frac{1}{2}(x_{t-\frac{m}{2}}+x_{t+\frac{m}{2}})+\sum_{j=1-\frac{m}{2}}^{\frac{m}{2}-1} x_{t+j}\right], \text{ for } 1+\frac{m}{2} \leq t \leq n-\frac{m}{2}$$

Et n×m glidende gennemsnit

Ofte opstiller man også et glidende gennemsnit af typen n×m. Den generelle formel vil vi afstå fra at angive her, men blot formulere hvorledes et sådan glidende gennemsnit ser ud i to specialtilfælde. Vi formulerer et 3 × 3-glidende gennemsnit og et 2 × 4 glidende gennemsnit. Begge disse typer glidende gennemsnit har fem led, som det ses af nedenstående;

$$\overline{S}_t^{3\times3} = \frac{1}{3}\left[\frac{s_{t-2}+s_{t-1}+s_t}{3}+\frac{s_{t-1}+s_t+s_{t+1}}{3}+\frac{s_t+s_{t+1}+s_{t+2}}{3}\right]$$

$$= \frac{1}{9}s_{t-2}+\frac{2}{9}s_{t-1}+\frac{1}{3}s_t+\frac{2}{9}s_{t+1}+\frac{1}{9}s_{t+2}$$

$$\overline{S}_t^{2\times4} = \frac{1}{2}\left[\frac{s_{t-2}+s_{t-1}+s_t+s_{t+1}}{4}+\frac{s_{t-1}+s_t+s_{t+1}+s_{t+2}}{4}\right]$$

$$= \frac{1}{8}s_{t-2}+\frac{1}{4}s_{t-1}+\frac{1}{4}s_t+\frac{1}{4}s_{t+1}+\frac{1}{8}s_{t+2}$$

A1.2 Hendersons ideal formel

En ofte anvendt form af glidende gennemsnit er Hendersons glidende gennemsnit, hvor vægtene beregnes efter Hendersons idealformel. Hendersons formel er ideel i den forstand, at kvadratsummen af 3. differenserne mellem HMA(Henderson Moving Averages)-vægtene minimeres. Lad k være længden af det glidende gennemsnit, og de to variabler z og m er givet ved k=2z-3=2m+1. Da er

$$kHMA(x_t)$$

$$= \sum_{i=t-m}^{t+m} \frac{315[(z-1)^2-i^2][z^2-i^2][(z+1)^2-i^2][(3z^2-16)-11i^2]}{8z(z^2-1)(4z^2-1)(4z^2-9)(4z^2-25)} x_i$$

A1.3 Sammenligning af MA-vægte

Når man i sæsonkorrigeringen skal til at udglatte observationer, så findes der ingen optimal type af glidende gennemsnit, og heller ikke en optimal længde af gennemsnit som bør anvendes. Valget vil afhænge af, hvor store udsving serien har, og af analysens formål i øvrigt. Generelt set får man en større udglatning jo længere de glidende gennemsnit er. Udglatningen af en tidsserie afhænger også af den valgte vægtstruktur. Nedenfor er udregnet vægte for forskellige typer fem-perioders glidende gennemsnit (MA for moving averages).

Tabel A.1.3.1 **Vægte for forskellige typer fem-perioders glidende gennemsnit**

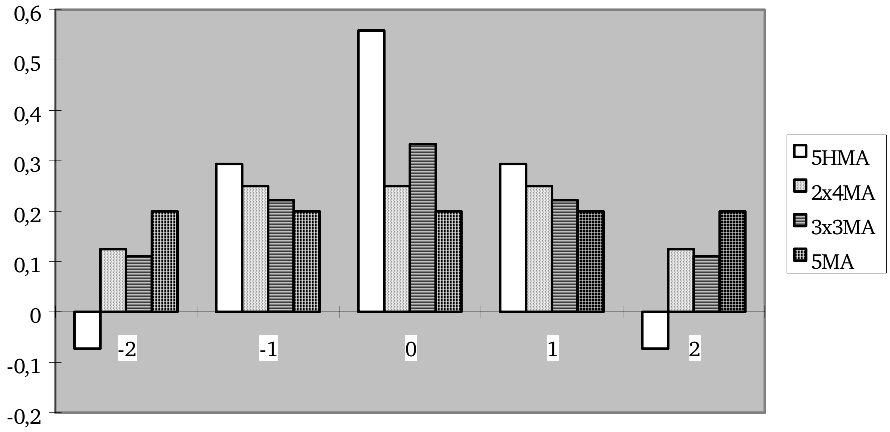
Periode i	5HMA	2x4MA	3x3MA	5MA
-2	-0,073	0,125	0,111	0,2
-1	0,294	0,250	0,222	0,2
0	0,559	0,250	0,333	0,2
1	0,294	0,250	0,222	0,2
2	-0,073	0,125	0,111	0,2

Vægtene i Henderson gennemsnittet med 5 led fås fra formlen i Appendiks A1.2, hvor vi ved at indsætte k=5, får at z=4 og m=2 og derfor vil vægtene for i=-2,-1,0,1,2 have følgende udseende

$$\frac{315(9-i^2)(16-i^2)(25-i^2)(32-11i^2)}{32\times15\times63\times55\times39}$$

Ved at afbilde vægtene for disse fire udvalgte typer af glidende gennemsnit i en figur ser vi, at et Henderson glidende gennemsnit tildeler relativ stor vægt til observationerne omkring midten af gennemsnittet, dvs. af centrum. Denne type af glidende gennemsnit har således ikke en så udglattende effekt, som de andre tre typer af gennemsnit i eksemplet.

Forskellige typer af glidende gennemsnit



Dét, at et Henderson glidende gennemsnit tildeler en stor vægt til de midterste observationer afspejles også i brugen af glidende gennemsnit i X-11ARIMA. Her anvendes Hendersons glidende gennemsnit til beregning af trend-cyklen i serier, som er sæsonkorrigerede. Til beregning af trend-cykler i ikke-sæsonkorrigerede serier benyttes typisk længere og dermed mere udglattende typer, såsom fx 3 × 3 MA eller 2 × 4 MA for kvartalsserier, hvor der er en større udglatningseffekt. For månedsserier vil man typisk vælge et noget længere gennemsnit, nemlig et tretten perioders langt glidende gennemsnit.

Appendiks 2: Om de statistiske tests i X-11ARIMA

Til vurdering af kvaliteten af den foretagne sæsonkorrigering beregner X-11ARIMA elleve forskellige teststørrelser M1,M2,...,M11. Disse sammenvejes (med forskellige vægte) i teststørrelsen Q, der kan ses som et globalt kvalitetsmål. Variationsområdet for alle teststørrelserne er mellem 0 og 3, og teststørrelserne er normeret således at acceptområdet er intervallet [0;1].

M1 Sæsonkomponenten og den irregulære komponent kan ikke separeres tilfredsstillende, hvis den irregulære variation er for høj sammenlignet med variationen i sæsonkomponenten. M1 måler det relative bidrag (til variansen for den procentvise ændring i komponenterne i originalserien) fra den irregulære komponent over et tremåneders interval.

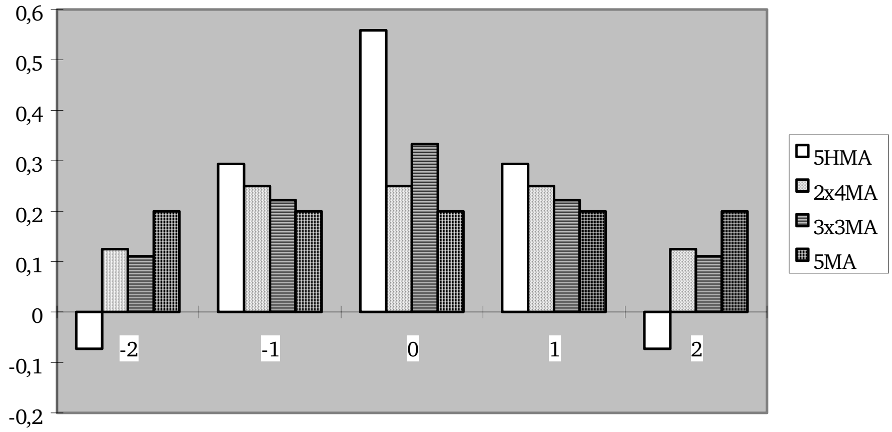
Hvis bidraget fra den irregulære komponent er for stort, er det et udtryk for, at variationen i den irregulære komponent dominerer variationen i sæsonkomponenten.

Vi definerer nu $\bar{I}^2 = \frac{1}{n-1}\sum_{t=2}^{n}(I_t - I_{t-1})^2$, hvor I_t er de endeligt estimerede irregulære faktorer fra tabel D13, og n er antallet af observationer i serien. Udfra samme formel defineres nu $\overline{C}$ og $\overline{S}$ og disse beregnes udfra henholdsvis tabel D12 og tabel D10 i X11-ARIMA outputtet. Vi betragter altså størrelsen:

$$\frac{315\left(9-i^2\right)\left(16-i^2\right)\left(25-i^2\right)\left(32-11i^2\right)}{32\times15\times63\times55\times39}$$

Ved at afbilde vægtene for disse fire udvalgte typer af glidende gennemsnit i en figur ser vi, at et Henderson glidende gennemsnit tildeler relativ stor vægt til observationerne omkring midten af gennemsnittet, dvs. af centrum. Denne type af glidende gennemsnit har således ikke en så udglattende effekt, som de andre tre typer af gennemsnit i eksemplet.

Forskellige typer af glidende gennemsnit



Dét, at et Henderson glidende gennemsnit tildeler en stor vægt til de midterste observationer afspejles også i brugen af glidende gennemsnit i X-11ARIMA. Her anvendes Hendersons glidende gennemsnit til beregning af trend-cyklen i serier, som er sæsonkorrigerede. Til beregning af trend-cykler i ikke-sæsonkorrigerede serier benyttes typisk længere og dermed mere udglattende typer, såsom fx 3 × 3 MA eller 2 × 4 MA for kvartalsserier, hvor der er en større udglatningseffekt. For månedsserier vil man typisk vælge et noget længere gennemsnit, nemlig et tretten perioders langt glidende gennemsnit.

Appendiks 2: Om de statistiske tests i X-11ARIMA

Til vurdering af kvaliteten af den foretagne sæsonkorrigering beregner X-11ARIMA elleve forskellige teststørrelser M1,M2,...,M11. Disse sammenvejes (med forskellige vægte) i teststørrelsen Q, der kan ses som et globalt kvalitetsmål. Variationsområdet for alle teststørrelserne er mellem 0 og 3, og teststørrelserne er normeret således at acceptområdet er intervallet [0;1].

M1 Sæsonkomponenten og den irregulære komponent kan ikke separeres tilfredsstillende, hvis den irregulære variation er for høj sammenlignet med variationen i sæsonkomponenten. M1 måler det relative bidrag (til variansen for den procentvise ændring i komponenterne i originalserien) fra den irregulære komponent over et tremåneders interval.
Hvis bidraget fra den irregulære komponent er for stort, er det et udtryk for, at variationen i den irregulære komponent dominerer variationen i sæsonkomponenten.

Vi definerer nu $\bar{I}^2 = \frac{1}{n-1}\sum_{t=2}^{n}(I_t - I_{t-1})^2$, hvor I_t er de endeligt estimerede irregulære faktorer fra tabel D13, og n er antallet af observationer i serien. Udfra samme formel defineres nu $\overline{C}$ og $\overline{S}$ og disse beregnes udfra henholdsvis tabel D12 og tabel D10 i X11-ARIMA outputtet. Vi betragter altså størrelsen:

$$\frac{\bar{I}^2}{\bar{I}^2 + \bar{C}^2 + \bar{S}^2}$$

Ud fra denne størrelse har Statistics Canada besluttet at fastlægge følgende præmis: Hvis ovennævnte størrelse er større end 0,1, så opfører serien sig ikke tilfredsstillende med hensyn til variationen i den irregulære faktor set i relation til variationen i øvrigt, dvs.

$$M1 = \frac{\bar{I}^2}{\bar{I}^2 + \bar{C}^2 + \bar{S}^2} \times 10.$$

skal være mindre end 1.

M2 M2 er ækvivalent med M1. Eneste forskel er måden man fjerner trenden (for at gøre serien stationær) på. Man fitter en linie[9] til trend-cykel estimaterne fra tabel D12 (fra X11-ARIMA outputtet) for at opnå et trend-estimat. Dette trend-estimat fjernes fra originalserien og således opnås en stationær originalserie, kald denne B1'. Data fra tabel D12 (endeligt estimerede trend-cykel komponenter) de-trendes på samme måde kald denne de-trendede serie D12'. Herefter udregnes kvotienten mellem den af tabel D13 (fra X11-ARIMA outputtet) estimerede varians og den af tabel B1' estimerede varians. Denne størrelse svarer til bidraget fra den irregulære komponent. Hvis denne størrelse er større end 0,1 forkastes hypotesen , dvs. når

$$M2 = 10 \times \frac{Varians\ \ fra\ \ tabel\ \ 13}{Varians\ \ fra\ \ tabel\ \ B1'}$$ er større end 1.

M3 I teststørrelsen M3 sammenholdes størrelsen af den månedlige/kvartalsvise ændring i den irregulære komponent med størrelsen af den månedlige/kvartalsvise ændring i trend-cyklen.

Formålet med at sæsonkorrigere er at udtrække sæsonkomponenten fra originaldata, for at kunne estimere en sæsonkorrigeret serie. Fordi X-11 ARIMA er et iterativt baseret program, er det især vigtigt at ikke kun sæsonfaktoren er estimeret ordentligt i de trin som fører frem til den endelige sæsonkorrigering, men også trend-cyklen og den irregulære komponent skal også være estimeret ordentligt. Hvis bevægelsen fra periode til periode af den irregulære faktor er dominerende for CI-serien, er det svært at separere disse to komponenter, og kvaliteten af sæsonkorrigeringen bliver derfor dårlig.

Vi beregner altså den såkaldte $\tilde{I} / \tilde{C}$ rate;

$$\tilde{I} / \tilde{C} = \frac{\sum_{t=2}^{n} |I_t - I_{t-1}| / I_{t-1}}{\sum_{t=2}^{n} |C_t - C_{t-1}| / C_{t-1}}$$

En høj værdi af $\tilde{I} / \tilde{C}$ raten indikerer så, at variationen i den sæsonkorrigerede serie primært skyldes den irregulære komponent. Værdien af raten anses for værende høj, når den er større end 3, den tilhørende teststørrelse har da formen

$$M3 = |\tilde{I} / \tilde{C} - 1| / 2$$

for månedsserier, og for kvartalsserier er

$$M3 = \left|\tilde{I} / \tilde{C} - \frac{1}{3}\right| / \frac{2}{3}$$

$\tilde{I} / \tilde{C}$ raten fås til beregningen fra tabel D12.

[9] Ved en multiplikativ model fittes en eksponentiel udvikling, og alle komponenterne transformeres med logaritmen.

M4 Teststørrelsen M4 undersøger omfanget af 1. ordens autokorrelation i den irregulære komponent. En af de helt basale antagelser i X-11ARIMA's F-teststørrelser er, at den irregulære komponent er hvid støj, dvs. uafhængige identisk fordelte fejlled med middelværdi nul, konstant[10] varians og kovarians imellem fejlleddene lig nul. X-11ARIMA bruger et fortegnstest kaldet ADR (Average Duration of Run) til at teste for tilfældighed i de endeligt estimerede residualer fra tabel D13. Dette ikke-parametriske test[11] er baseret på basis af antallet af vendepunkter[12]. Testet er designet til at teste tilfældigheden i residualerne imod alternativ hypotesen at fejlleddene følger en AR(1)-proces. For en hvid støj proces (med uendeligt antal observationer) vil ADR være lig 1,5. M4 teststørrelsen er baseret på normalfordelingsapproximationsfomlen af Bradley[13]:

$$M4 = \frac{\left| \frac{n-1}{ADR} - \frac{2(n-1)}{3} \right|}{\sqrt{\frac{16n-29}{90}}} \times \frac{1}{2{,}58}$$

hvor værdien på 2,58 er 1% signifikansniveauet i en normalfordelings tosidet test. Hvis M4 er større end 1 er der autokorrelation i residualerne.

M5 M5 er et mål for antallet af perioder, som det tager den gennemsnitlige absolutte ændring i trend-cyklen at dominere den tilsvarende ændring i den irregulære komponent.
Teststørrelsen er ækvivalent med M3 i den forstand, at den undersøger den relative størrelse af ændringerne i den irregulære faktor og trend-cykel komponenterne. Man udregner for k=1,...,12 (eller k=1,2,3,4 for kvartalsserier) størrelsen

$$\frac{\frac{1}{n-1}\sum_{j=k+1}^{n}(I_t - I_{t-k})^2}{\frac{1}{n-1}\sum_{j=k+1}^{n}(C_t - C_{t-k})^2} \equiv \frac{\bar{I}(k)}{\bar{C}(k)}.$$

Herefter udleder man Monthly Cyclical Dominance (MCD) (for kvartalsserier udledes QCD på samme vis) som

$$\bar{I}(k)/\bar{C}(k) \leq 1 \wedge \bar{I}(k-1)/\bar{C}(k-1) > 1 \Rightarrow MCD = k$$

MCD antager derfor kun heltalsværdier, og er i sin konstruktion lidt 'dum', for den kan ikke aflæse hvor tæt på 1 I/C ratioen rent faktisk er. For at gøre MCD 'klogere', har man konstrueret en ny teststørrelse som skal løse dette problem MCD'. MCD' interpolerer lineært I/C-ratioen for at findes dens skæring med 1. Formelen for MCD' skal ikke angives her, men det skal blot nævnes, at en MCD' værdi på over seks anses som værende uacceptabel. Derfor konstrueres M5 på følgende vis:

$$M5 = \frac{|MCD' - 0{,}5|}{5} \text{ for månedsserier og}$$

$$M5 = \frac{|QCD' - 0{,}17|}{1{,}67} \text{ for kvartalsserier.}$$

---

[10] Over tid

[11] Testet er udviklet af W.A. Wallis og G.H. Moore (1941): '*A Significance Test for Time Series*', National Bureau of Economic Research, Technical Paper no. 1.

[12] Et vendepunkt er defineret ved, at udviklingen fra periode til periode skifter fortegn.

[13] Bradley, James V. (1968): *Distribution-free Statistical tests*. Prentice-Hall, Englewood Cliffs, N.J.

Således er M5>1 når MCD' eller QCD' er uacceptabelt store.

M6 I teststørrelsen M6 undersøges størrelsen af den årlige ændring i den irregulære faktor sammenlignet med størrelsen af den årlige ændring i sæsonfaktoren. Når man tager formålet med sæsonkorrigeringen i betragtning, er det meget vigtigt at kunne identificere sæsonfaktorerne. For at separere den irregulære faktor fra sæsonkomponenten benytter X-11ARIMA et 3 x 5 glidende gennemsnit på SI-ratioen (differensen, ved additiv model). Erfaringer har dog vist, at hvis den årlige ændring i den irregulære faktor er for lille (sammenlignet med den årlige ændring i sæsonfaktoren), altså når I/S-ratioen er lav, så er et 3 x 5 glidende gennemsnit ikke fleksibelt nok til at følge sæsonbevægelsen. Modsat viser det sig, at når I/S-ratioen er for høj, da er 3 x 5 sæsonfilteret for fleksibelt, og de udledte sæsonfaktorer indeholder noget af den irregulære bevægelse. Empiriske studier i Statistics Canada har vist, at det 3 x 5 glidende gennemsnit virker passende når I/S-ratioen er i intervallet [1,56,5]. Dvs.

$$M6 = \frac{\left| \bar{I} / \bar{S} - 4 \right|}{2{,}5}$$

Når M6 er større end 1 forkaster vi hypotesen, men problemet med en høj værdi af M6 kan løses ved at bruge et 3x1 glidende gennemsnit, når I/S-ratioen, som fås fra tabel F2.H er mindre end 1,5, eller ved at bruge det stabile sæsonmønster (stable seasonality option) hvis forholdet er større end 6,5.

M7 M7 tester omfanget af stabil sæson i forhold til bevægelig sæson. Hvis en tidsserie korrigeret for trend-cyklen (dvs. SI-raten) kun udviser en beskeden grad af stabil sæson set i forhold til bevægelig sæson, da vil identifikation af den stabile sæsonvariation være vanskelig. Det anvendte test er en kombination af det F-test anvendt på de endelige SI-ratioer fra tabel D10 som lå i X-11 fra US Bureau of Census, og et test som er udviklet af Statistics Canada af J. Higginson (1975)[14]. Dette test indikerer om sæsonmønstret i serien er identificerbart af X-11ARIMA eller ej. Testet, som lå i X-11 versionen måler størrelsen af stabilt sæsonmønster i tidsserien, kald denne F_s , mens Higgins' test undersøger om der er bevægelig sæson i serien, kald dette test F_m. Sæsonmønstret er identificerbart, hvis den absolutte fejl (eller uro), som ligger i de endelige estimater for sæsonfaktorerne, ikke er for høj. Denne forstyrrelse afhænger af begge ovennævnte F-teststørrelser. Er F_s lav finder vi en høj grad af uro, og er yderligere F_m høj, finder vi mere uro i serien, nemlig den uro som skyldes bevægelig sæson. Man har altså defineret en teststørrelse[15] som indeholder begge ovenfor nævnte F-tests;

$$M7 = \sqrt{\frac{1}{2}\left(\frac{7}{F_s} + 3\frac{F_m}{F_s}\right)}$$

Specielt om M8,..., M11 I den fulde længde af tidsserien kan kun en konstant sæsonkomponent estimeres optimalt. Dette skyldes de anvendte sæsonfiltre i X-11ARIMA. Således vil estimaterne af sæsonfaktorerne indeholde en betragtelig fejl, hvis originalserien indeholder bevægelse fra år til år. Vi skelner mellem to typer af bevægelse, nemlig mellem dem som udviser tilfældige fluktuationer, og dem, hvor ændringerne forekommer i samme retning hen over årene.

Størrelsen af den første type bevægelse kan måles udfra den gennemsnitlige absolutte år-til-år ændring i sæsonfaktorerne, og den anden type bevægelse kan måles ud fra et simpelt aritmetisk gennemsnit af ændringerne. Et sådant gennemsnit vil nemlig give en indikation af størrelsen af systematisk (lineær) bevægelse.

---

[14] '*An F-test for the precence of Moving Seasonality when using Census Method II-X-11 Variant*', Research paper, Seasonal Adjustment and Time Series Staff, Statistics Canada.

[15] En detaljeret udredning af denne teststørrelse findes i Lothian, John and Morry, Marietta (1977): 'The problem of Aggregation; Direct or Indirect' Research Paper, Seasonal adjustment and Time Series Staff, Statistics Canada.

Tilfældige bevægelser måles af teststørrelserne M8 og M10. Teststørrelserne M9 og M11 beskriver størrelsen af den lineære bevægelse. M8 og M10 benytter al data fra tabel D10 til udregningerne, mens der i beregningerne af M10 og M11 kun benyttes data fra de seneste perioder.
Da brugere af sæsonkorrigeret data primært er interesseret i kvaliteten af sæsonkorrigeringen i de nyeste år, blev teststørrelserne M10 og M11 introduceret for at beskrive sæsonbevægelsen i slutningen af serien. Specielt er det vigtigt, at vide om der er en tydelig lineær bevægelse i sæsonfaktorerne for de seneste år, for hvis dette er tilfældet vil estimaterne for sæsonfaktorerne blive forstyrret betragteligt af sæsonfiltrene. Det er den samme forstyrrelse der forhindrer brugen af de seneste års sæsonfaktorer til måling af størrelsesordenen af sæsonbevægelsen. Dette løses ved at undersøge de tre år, der ligger forud for de seneste tre år, hvor man så håber, at sæsonbevægelsen forbliver uændret i disse 'nye' slutår.
Alle værdierne fra tabel D10 normaliseres, når teststørrelserne skal udregnes, dvs. teststørrelserne M8,...,M11 baseres på de normaliserede sæsonfaktorer;

$$S'_t = \frac{S_t - \overline{S}}{\sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(S_t - \overline{S})^2}}$$

M8 Teststørrelsen M8 måler den tilfældige fluktuation af sæsonfaktorerne i hele seriens forløb. En høj værdi vil indikere en høj grad af forstyrrelse i X-11ARIMA estimationen af sæsonfaktorerne. Hvis sæsonfaktorerne for de enkelte år er meget forskellige (og tilfældige), er sæsonkorrigeringen næppe anvendelig, idet vi har fundet et stærkt fluktuerende sæsonmønster. Variationen i sæsonfaktorerne kan måles ved følgende størrelse;

$$\left|\Delta\overline{S}'\right| = \frac{1}{m(T-1)}\sum_{j=1}^{m}\sum_{i=2}^{T}\left|S'_{mi+j} - S'_{m(i-1)+j}\right|$$

hvor m er antallet af observationer på et kalenderår (altså enten lig 4 eller 12) og T er antal år.

Da den gennemsnitlige acceptable værdi for variationen i sæsonfaktorerne sættes til 10 pct., har M8 følgende udseende:

$$M8 = 10\left|\Delta\overline{S}'\right|$$

I beregningen af M8 benyttes data fra tabel D10 dog indgår der kun data fra de år, hvor sæsonfaktorerne er beregnet uden ekstrapolation.

M9 Teststørrelsen anvendes til at teste for den gennemsnitlige lineære bevægelse i sæsonfaktorerne i hele seriens længde. Når man danner gennemsnit for år-til-år ændringerne, måler man mængden af systematisk bevægelse i serien. Hvis der kun er tilfældige fluktuationer fra år til år, vil dette gennemsnit være tæt på nul. Hvis de fleste ændringer er i samme retning, vil den gennemsnitlige absolutte ændring være ret tæt på den gennemsnitlige aritmetiske ændring.
Da det gælder, at (teleskopsum) $\sum_{i=1}^{n-1}\Delta S'_{mi+j} = S'_{m(n-1)} - S'_j$, og med en acceptgrænse på 0,1, får vi at

$$M9 = 100 \times \frac{\sum_{j=1}^{m}\left|S'_{m(n-1)+j} - S'_j\right|}{m(n-1)} \times \frac{1}{10}$$

M10 og M11

Teststørrelserne er identiske med henholdsvis M8 og M9, men disse er blot udregnet for årene n-2, n-3, n-4 og n-5. Brugere er typisk ofte kun interesserede i nyeste data, og derfor giver disse teststørrelser et indblik i kvaliteten af de seneste estimater af sæsonfaktorerne. Til beregningen anvendes alene data fra de seneste år, hvor sæsonfaktorerne er beregnet uden ekstrapolation i enderne.

Q-teststørrelsen

De ovenfor beskrevne elleve teststørrelser M1-M11 sammenvejes til slut til den summariske teststørrelse Q og der anvendes generelt følgende vægtning;

Tabel A.2.1 **Vægte til Q-teststørrelsen for serier længere end 6 år**

M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
13	13	10	5	11	10	16	7	7	4	4

Tabel A.2.2 **Vægte for Q-teststørrelsen for serier kortere end 6 år**

M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
17	17	10	5	11	10	30	0	0	0	0

Bemærk at tabel 2 ikke blot er en reskalering af tabel 1. Vægtene for teststørrelserne M3, M4, M5 og M6 er uændrede, og dette skyldes, at disse teststørrelser ikke måler noget der svarer til det, som måles af M8, M9, M10 og M11.

Appendiks 3: Modelspecifikationer og teststørrelser for de enkelte serier.

Tabel A.3.1 **Modelspecifikationer**

Serienavn	Fre-kvens	Model	Start	Slut	ARIMA-model	Påske-effekt	Handels-dags-effekt
Nyregistrerede biler							
- i husholdningerne	m	M0	1992 01	2000 04	Nej	Nej	Ja
- i erhvervene	m	M0	1992 01	2000 04	Nej	Nej	Ja
- uidentificerede	m	M0	1992 01	2000 04	Nej	Nej	Nej
-Nyregistrerede biler I alt	m	M0	1992 01	2000 04	Nej	Nej	Ja
Detailomsætningsindeks							
Fødevarer og andre dagligvarer	m	M0	1990 01	2000 04	Ja	Nej	Ja
Beklædning mv.	m	M0	1990 01	2000 04	Ja	Nej	Ja
Andre forbrugsvarer	m	M0	1990 01	2000 04	Ja	Nej	Ja
Konkurser og tvangsauktioner							
Erklærede konkurser	m	M0	1986 01	2000 05	Ja	Ja	Ja
Kundgjorte tvangsauktioner	m	M0	1986 01	2000 05	Nej	Nej	Ja
Anmeldte straffelovsovertrædelser							
Indbrud i alt	q	M0	1980 01	2000 01	Ja	Nej	NA
Tyveri i alt	q	M0	1980 01	2000 01	Ja	Nej	NA
Sædelighedsforbrydelser i alt	q	M0	1980 01	2000 01	Ja	Nej	NA
Voldsforbrydelser i alt	q	M0	1980 01	2000 01	Ja	Nej	NA
Ejendomsforbrydelser i alt	q	M0	1980 01	2000 01	Ja	Nej	NA
Færdselsuheld							
Skader i personbiler i alt	m	M0	1984 01	2000 05	Ja	Ja	Ja
Beskæftigelsesopgørelse på grundlag af ATP-indbetalinger							
Landbrug, fiskeri og råstofudvinding	q	M0	1980 01	2000 01	Ja	NA	NA
Industri	q	M0	1980 01	2000 01	Ja	NA	NA
Bygge- og anlægsvirksomhed	q	M0	1980 01	2000 01	Ja	NA	NA

Tabel A.3.2 **Teststørrelser**

Serienavn	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	Q
Nyregistrerede biler												
- i husholdningerne	0,713	1,091	0,497	0,495	0,739	0,215	0,534	0,446	0,392	0,602	0,570	0,63
- i erhvervene	2,340	3,000	1,074	0,031	2,499	0,485	0,760	0,655	0,505	0,863	0,802	1,37
-uidentificerede	2,205	0,578	0,325	0,867	0,616	0,367	1,027	1,678	1,391	1,967	1,835	1,12
-Nyregistrerede biler I alt	1,364	1,658	0,457	0,960	0,941	0,124	0,618	0,440	0,379	0,587	0,570	0,84
Detailomsætningsindeks												
Fødevarer og andre dagligvarer	0,196	0,291	1,257	0,111	3,000	0,375	0,162	0,241	0,214	0,297	0,291	0,67
Beklædning mv.	0,309	0,283	2,588	0,722	3,000	0,309	0,134	0,228	0,144	0,217	0,171	0,86
Andre forbrugsvarer	0,148	0,151	0,644	0,056	0,668	0,536	0,211	0,374	0,258	0,387	0,323	0,32
Konkurser og tvangsauktioner												
Erklærede konkurser	2,766	0,659	1,217	0,422	0,958	0,359	0,714	1,184	0,604	1,395	1,344	1,08
Kundgjorte tvangsauktioner	1,186	0,124	0,607	0,281	0,483	0,754	0,429	0,700	0,533	0,803	0,745	0,55
Anmeldte straffelovsovertrædelser												
Indbrud i alt	2,184	0,610	0,954	1,172	0,809	0,130	0,525	0,831	0,707	0,492	0,372	0,89
Tyveri i alt	0,522	0,212	0,628	1,172	0,648	0,188	0,150	0,390	0,213	0,308	0,307	0,44
Sædelighedsforbrydelser i alt	2,034	1,562	0,752	0,552	0,804	0,605	0,369	0,715	0,371	0,749	0,641	0,87
Voldsforbrydelser i alt	0,971	0,419	0,578	1,172	0,532	0,057	0,255	0,648	0,170	0,813	0,813	0,58
Ejendomsforbrydelser i alt	1,566	0,428	0,902	1,379	0,746	0,165	0,342	0,637	0,548	0,265	0,130	0,72
Færdselsuheld												
Skader i personbiler i alt	1,695	1,691	1,452	0,329	3,000	0,536	0,349	0,633	0,311	0,780	0,716	1,16
Beskæftigelsesopgørelser på grundlag af ATP-indbetalinger												
Landbrug, fiskeri og råstofudvinding	0,107	0,082	0,692	0,862	0,666	0,110	0,058	0,155	0,070	0,089	0,080	0,29
Industri	0,198	0,255	0,356	0,655	0,501	0,342	0,122	0,349	0,192	0,623	0,623	0,33
Bygge- og anlægsvirksomhed	0,300	0,130	0,932	0,862	0,572	0,142	0,141	0,360	0,242	0,235	0,207	0,39