

# En teoretisk og empirisk sammenligning af metoder til sæsonkorrektion

Heino Bohn Nielsen\*

Danmarks Nationalbank

November 1999

## Resumé

Dette papir sammenligner de to sæsonkorrektionsmetoder X-11/X-12 og TRAMO/SEATS. Der indledes med en præsentation af metodernes teoretiske grundlag illustreret med eksempler på økonomiske tidsserier. Herefter sammenlignes metoderne empirisk, dels ud fra deres evne til at isolere en kendt sæsonfaktor i simulerede tal og dels ud fra størrelsen af de revisioner som foretages i tidligere sæsonkorrigerede tal, når der kommer nye observationer. Det konkluderes, at den modelbaserede metode TRAMO/SEATS er en godt alternativ til den mere kendte X-12. Teoretisk er metoden overlegen, og i de udførte test kan nogle af fordelene tilsyneladende overføres til praksis. Svaghederne i metoden ligger i den fleksible tilgang, som potentielt gør den ustabil og kan medføre store revisioner i historiske tal.

## 1 Introduktion

For mange økonomiske tidsserier er der ofte mere fokus på sæsonkorrigerede end på ukorrigerede tal. Det illustrerer, at brugere, på trods af mange argumenter mod sæsonkorrektion, generelt finder informationen i de sæsonkorrigerede serier nyttig. Fx er det ofte lettere at analysere kortsigtsvariation i sæsonkorrigerede tal, fordi alle observationer umiddelbart kan sammenlignes, uafhængigt af hvornår på året de optræder, se fx Maravall (1996) for en diskussion.

En grundlæggende idé bag sæsonkorrektion er, at sæsoneffekten er forudsigelig ikke påvirkes af den generelle udvikling. Formålet med sæsonkorrektionen er at estimere sæsoneffekten og fjerne den fra data. Det kan opfattes som en simplifikation, der er nyttig, hvis informationstabt forbundet med proceduren ikke er for stort (Bell og Hillmer, 1984). I frekvensdomæne implicerer sæson, at der er toppe i tidsseriens power-spektrum ved sæsonfrekvenser, fx  $\omega = \frac{2\pi k}{12}$  radianer,  $k = 1, 2, \dots, 6$ , for månedstal. En ideel sæsonkorrektion fjerner sådanne toppe uden at ændre spektrum for andre frekvenser. På den måde kan sæsonkorrektionen opfattes som et signal-extraction problem, der skal fjerne variation ved sæsonfrekvenserne uden at ændre informationsindholdet ved andre frekvenser.

Et centralt problem ved sæsonkorrektion er, at begrebet sæson ikke er særligt præcist defineret. Sæsonen kan føres tilbage til en række forskellige årsager, som er mere eller mindre eksogene, fx; (i) vejret, (ii) forskellige kalendereffekter såsom højtider eller (iii) timing

---

\*Papirets synspunkter og konklusioner er forfatterens egne og deles ikke nødvendigvis af Danmarks Nationalbank. E-mail adresse: hbn@nationalbanken.dk.

af beslutninger, fx placering af ferier og regnskabsår, betalingsterminer, placering af dyrtidsregulering mv. En mulig definition på sæson er givet i Hylleberg (1992) som *en systematisk, men ikke nødvendigvis regulær, bevægelse indenfor året*, men det er uklart, præcis hvor systematisk bevægelserne skal være. Hertil kommer, at sæsonmønsteret kan ændres over tid.

Af særlig interesse for databrugere er, hvor meget tidligere sæsonkorrigerede tal revideres, når der kommer nye observationer. Af praktiske hensyn vil man foretrække så små revisioner som muligt. Revisionerne afspejler imidlertid anvendelsen af ny information, og små revisioner kan dermed blot afspejle, at informationen ikke anvendes til at forbedre en allerede foretagen sæsonkorrektion, se Gómez og Maravall (1998b). Man må derfor afveje hensynet til små revisioner mod hensynet til en optimal sæsonkorrektion givet informationssettet.

I dette papir sammenlignes forskellige metoder til sæsonkorrektion. I præsentationen fokuseres på månedstal, men argumenterne kan let overføres til kvartalstal. Der betragtes to generationer af metoder fra Bureau of the Census, nemlig X-11 og X-12-ARIMA, som begge er baseret på en ad hoc anvendelse af faste filtre. De sammenlignes med TRAMO/SEATS fra Banco de España, som anvender en signal-extraction approach baseret på estimerede Arima-modeller. Generelt er det vanskeligt at sammenligne metoder til sæsonkorrektion. Det skyldes bl.a., at man ikke ved præcist, hvad det er, der skal fjernes, så forskellige metoder i realiteten estimerer forskellige ting, se Planas (1997b). Udgangspunktet for en sammenligning må derfor være, at evaluere metodernes forudsætninger mod data.

Papiret er bygget op som følger: Først præsenteres metoderne og de underliggende statistiske modeller i afsnit 2, og de illustreres med et par eksempler. I afsnit 3 sammenlignes metoderne ud fra deres evne til at isolere en kendt sæsonfaktor i simulerede data. Afsnit 4 betragter nærmere revisioner af de sæsonkorrigerede tal, og afsnit 5 konkluderer.

## 2 Metoder til sæsonkorrektion

Det overordnede formål med sæsonkorrektion er at dekomponere en observeret tidsrække i ikke observerbare komponenter, som indeholder fortolkelig information. Dekomponeringen kan fx have formen

$$Y_t = T_t + S_t + I_t, \quad t = 1, 2, \dots, T \quad (1)$$

hvor  $Y_t$  er den observerede tidsserie, mens  $T_t$ ,  $S_t$  og  $I_t$  er ikke observerede komponenter, der beskriver henholdsvis den trendmæssige udvikling, sæson-komponenten og den irregulær komponent.

Som et led af sæsonkorrektionen estimeres typisk forskellige deterministiske effekter, som fx antallet af handelsdage og placering af påsken. Systematiske effekter, som er knyttet til kalenderen, henregnes typisk til sæsonkomponenten og fjernes dermed fra de sæsonkorrigerede tal, mens andre deterministiske komponenter henregnes trenden eller den irregulære komponent. Man kan enten forhåndskorrigere for sådanne deterministiske komponenter eller på anden måde estimere effekterne, jf. nedenfor<sup>1</sup>.

Sammenhængen i ligning (1) er additiv, men kan gælde for vilkårlige transformationer af data. Hvis  $Y_t = \log(X_t)$  angiver (1) fx en multiplikativ sammenhæng for  $X_t$ . Normalt antages komponenterne i (1) at være ortogonale. Det svarer til, at komponenterne genereres af uafhængige økonomiske mekanismer, se Wallis (1974) eller Bell og Hillmer (1984). I realiteten er det en hård forudsætning, som ikke nødvendigvis er opfyldt, se Hylleberg (1992) for et eksempel.

Følgende præsenteres de to sæsonkorrektionsmetoder; X-11/X-12 og TRAMO/SEATS.

---

<sup>1</sup>I det følgende lægges der ikke meget vægt på de deterministiske komponenter.

## 2.1 Census X-11

Den mest succesfulde og anvendte metode til sæsonkorrektion de sidste 30 år har været Bureau of the Census' metode X-11, se Shiskin et al. (1967). Metoden består i en sekvensiel anvendelse af glidende gennemsnit. Følgende gennemgås rutinens 4 grundlæggende trin, se også Findley et al. (1996) eller Hylleberg (1986).

**Trin 1:** Et første estimat af trend-komponenten,  $T_t^{(1)}$ , findes som et såkaldt  $2 \times 12$  filter, som angiver et 2-perioders glidende gennemsnit af et 12-perioders glidende centreret gennemsnit af tidsserien. Det kan skrives som et vægtet 13-perioders gennemsnit af formen

$$T_t^{(1)} = \frac{Y_{t-6}}{24} + \frac{Y_{t-5}}{12} + \dots + \frac{Y_t}{12} + \dots + \frac{Y_{t+5}}{12} + \frac{Y_{t+6}}{24} = \mu(L)Y_t,$$

hvor  $\mu(\cdot)$  er et polynomium i lagoperatoren  $L$ ,  $LY_t = Y_{t-1}$ . Afvigelsen herfra er det første estimat af sæson og irregulær komponent

$$(S + I)_t^{(1)} = Y_t - T_t^{(1)}.$$

Et første estimat af sæson-komponenten findes med et  $3 \times 3$  filter af månederne hver især<sup>2</sup>. Det kan skrives som et vægtet 5 perioders centreret glidende gennemsnit af månederne

$$\widehat{S}_t^{(1)} = \frac{(S + I)_{t-24}^{(1)}}{9} + \frac{2(S + I)_{t-12}^{(1)}}{9} + \frac{3(S + I)_t^{(1)}}{9} + \frac{2(S + I)_{t+12}^{(1)}}{9} + \frac{(S + I)_{t+24}^{(1)}}{9} = \lambda_1(L)(S + I)_t^{(1)},$$

hvor  $\lambda_1(L)$  er et lagpolynomium. De foreløbige sæsonfaktorer opregnes til (approksimativt) at summe til nul i alle 12 måneders horisonter ved at fratække et  $2 \times 12$  glidende gennemsnit

$$S_t^{(1)} = \widehat{S}_t^{(1)} - \left( \frac{\widehat{S}_{t-6}^{(1)}}{24} + \frac{\widehat{S}_{t-5}^{(1)}}{12} + \dots + \frac{\widehat{S}_t^{(1)}}{12} + \dots + \frac{\widehat{S}_{t+5}^{(1)}}{12} + \frac{\widehat{S}_{t+6}^{(1)}}{24} \right) = (1 - \mu(L)) \widehat{S}_t^{(1)},$$

og det første estimat af henholdsvis den sæsonkorrigerede serie og de irregulære komponent findes som

$$\begin{aligned} A_t^{(1)} &= Y_t - S_t^{(1)} \\ I_t^{(1)} &= (S + I)_t^{(1)} - S_t^{(1)}. \end{aligned}$$

**Trin 2:** I den irregulære komponent,  $I_t^{(1)}$ , identificeres og nedvægtes ekstreme observationer, så det undgås, at de dominerer de glidende gennemsnit. Det gøres ved at sammenholde observationer af den irregulære komponent med en 5 års glidende standardafvigelse. Ekstreme observationer erstattes af et gennemsnit af den ekstreme observation og to observationer på hver side, hvor den ekstreme observation nedvægtes afhængig af den relative størrelse, se Shiskin et al. (1967). Herudover estimeres effekten af forskellige deterministiske komponenter med OLS på den irregulære komponent og effekten fjernes. Herefter genberegnes sæson-komponenten,  $S_t^{(1)}$ , som ovenfor på baggrund af serien med nedvægtede ekstreme observationer.

**Trin 3:** Den resulterende sæsonrensede serie,  $A_t^{(1)}$ , anvendes til at estimere trend-komponenten som et glidende Henderson-gennemsnit

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

---

<sup>2</sup>Der kan i X-11 alternativt vælges filtre af længden  $3 \times 1$ ,  $3 \times 5$  eller  $3 \times 7$ . Valget afhænger af, hvor regulær sæsonen er, og i nyere versioner vælges filteret automatisk.

hvor  $k = \frac{13-1}{2}$  og  $h_j$  er vægte, se Findley et al. (1996) eller Planas (1997a)<sup>3</sup>. Med det nye estimat af trenden genberegnes

$$(S + I)_t^{(2)} = Y_t - T_t^{(2)}.$$

Sæsonfaktorerne findes i trin 3 som et  $3 \times 5$  filter, svarende til 7 perioders vægtet glidende gennemsnit af de enkelte måneder

$$\begin{aligned} \widehat{S}_t^{(2)} &= \frac{(S+I)_{t-36}^{(2)}}{15} + \frac{2(S+I)_{t-24}^{(2)}}{15} + \frac{3(S+I)_{t-12}^{(2)}}{15} \\ &+ \frac{3(S+I)_t^{(2)}}{15} + \frac{3(S+I)_{t+12}^{(2)}}{15} + \frac{2(S+I)_{t+24}^{(2)}}{15} + \frac{(S+I)_{t+36}^{(2)}}{15} = \lambda_2(L)(S + I)_t^{(2)}, \end{aligned}$$

som igen opregnes i forhold til et glidende gennemsnit til de endelige sæsonfaktorer

$$S_t = (1 - \mu(L)) \widehat{S}_t^{(2)}.$$

Den endelige sæsonkorrigerede serie findes nu som

$$A_t = Y_t - S_t.$$

**Trin 4:** Til slut beregnes den endelige trend ud fra de sæsonkorrigerede tal,  $A_t$ , med Henderson-formlen

$$T_t = \sum_{j=-H}^H h_j A_{t+j},$$

og den endelige irregulære komponent findes som

$$I_t = A_t - T_t.$$

Til komponenterne lægges effekten af de deterministiske komponenter, som blev estimeret og fjernet i trin 2.

Den grundlæggende ide bag X-11 er for så vidt simpel, men som det fremgår er beregningerne i praksis relativt komplekse. Et væsentligt problem ved tilgangen er, at den er ad hoc uden et klart teoretisk fundament, og det er umiddelbart vanskeligt at vurdere metodens egenskaber. Nedenfor vises en repræsentation af X-11, som gør vurdering af egenskaberne lettere.

### 2.1.1 Repræsentation af X-11 som et lineært filter

Ses der bort fra behandlingen af outliers mv. i trin 2, er den skitserede metode lineær, og den samlede effekt af de sekventielle glidende gennemsnit er et langt lineært MA-filter. De sæsonkorrigerede data,  $A_t$ , kan således skrives som et 169 perioders glidende gennemsnit

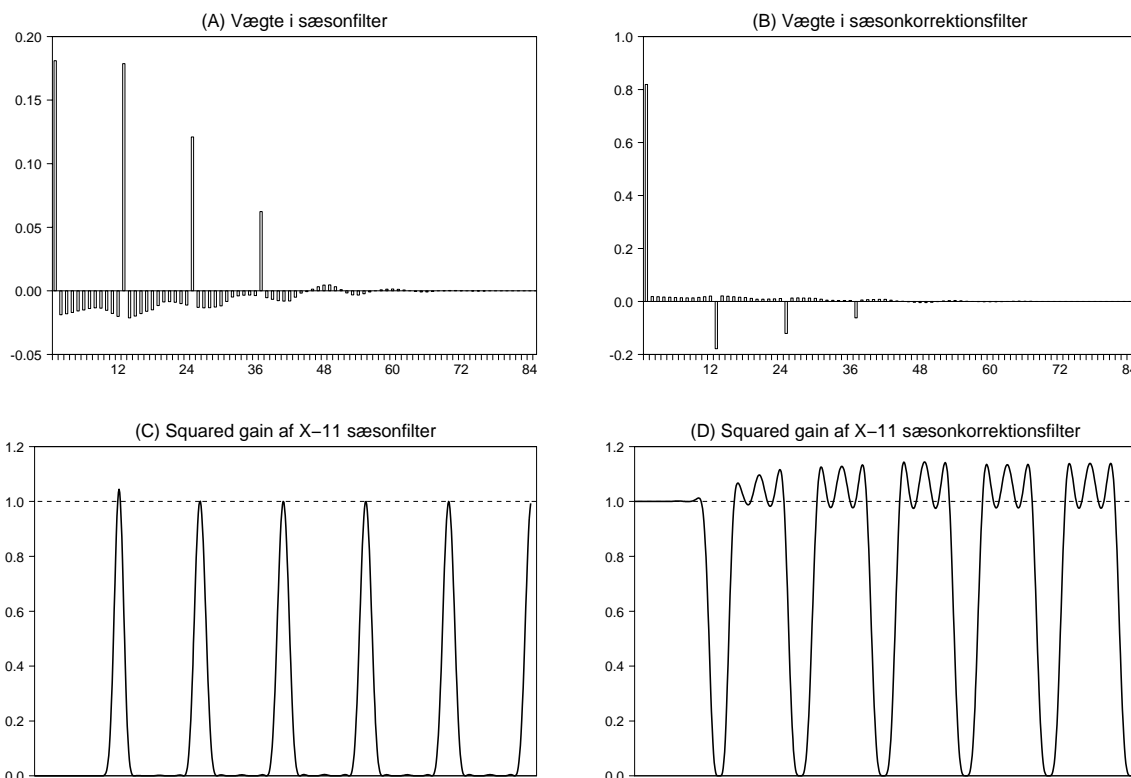
$$A_t = a(L)Y_t = \sum_{j=-84}^{84} a_j Y_{t-j}, \text{ hvor } \sum_{j=-84}^{84} a_j = 1. \quad (2)$$

Lagpolynomiet  $a(L)$  er en funktion af lagpolynomierne  $\mu(L)$ ,  $\lambda_1(L)$ ,  $\lambda_2(L)$  og  $H(L)$  af formen

$$a(L) = (1 - \mu(L)) \cdot \lambda_2(L) \cdot (1 - H(L) \cdot (1 - (1 - \mu(L)) \cdot \lambda_1(L) \cdot (1 - \mu(L))))),$$

---

<sup>3</sup>Længden af gennemsnittet vælges afhængigt af tidsserien, og standardvalget er 13 perioder.



Figur 1: En karakteristik af X-11 filteret.

se fx Bell og Monsell (1992). Filteret  $a(L)$  benytter 84 observationer på hver side af den aktuelle observation. Det betyder, at standardopsætningen af X-11 kun anvender et symmetrisk filter på observationer, der er omgivet af minimum 7 års data til begge sider. For observationer tættere på samplers endepunkter anvendes asymmetriske filtre, og en observation vil dermed blive sæsonkorrigeret med 84 forskellige asymmetriske filtre, inden estimatet af det sæsonkorrigerede tal er endeligt.

Svarende til (2) er filteret for sæsonkomponenten givet som  $1 - a(L)$ , og figur 1 (A) og (B) viser vægtene i henholdsvis sæsonfilteret,  $1 - a(L)$ , og i sæsonkorrektionsfilteret,  $a(L)$ . Den første vægt er til den ulaggede værdi,  $Y_t$ , mens efterfølgende vægte gælder for lags og lead fra  $t \pm 1, \dots, t \pm 84$ .

Sæsonfilteret bestemmer den aktuelle sæsonfaktor som hovedsageligt et gennemsnit af den aktuelle observation og tre laggede og leadede observationer af den samme måned, og samlet summer vægtene til nul. Vægtene i sæsonkorrektionsfilteret summer til én. Der gives en vægt på ca. 0,8 til den aktuelle observation og negative vægte til de tre næste sæsonlags/leads.

Ved at skrive X-11 som et enkelt lineært filter kan man umiddelbart analysere egenskaberne i filteret ved brug af spektralanalyse. Lad  $\omega$  angive frekvensen målt i radianer  $\omega \in [0; \pi]$ . Frekvens-respons-funktionen af  $a(L)$  findes med Fourier-transformationen  $L = e^{-i\omega}$ ;

$$a(\omega) = \sum_{k=-84}^{84} a_k e^{-ik\omega}, \quad (3)$$

hvor  $i = \sqrt{-1}$ . Filteret (3) bestemmer hvor stor en del af variationen i  $Y_t$  ved frekvens  $\omega$ , som transmitteres til variansen i estimatoren  $A_t$ . Sammenhængen mellem spektrum for

input-serien,  $g_Y(\omega)$ , og spektrum for output serien,  $g_A(\omega)$ , er givet ved

$$g_A(\omega) = |a(\omega)|^2 g_Y(\omega),$$

hvor transmissionen  $|a(\omega)|^2$  kaldes squared gain. Et squared gain på nul betyder, at filteret præcis eliminerer sådanne frekvenser, mens et gain på én betyder, at frekvensen passerer uændret igennem filteret.

Squared gain for de to filtre  $|1 - a(\omega)|^2$  og  $|a(\omega)|^2$  er illustreret i figur 1 (C) og (D) for frekvenser  $[0; \pi]$ . Som det fremgår, piler sæsonfilteret  $1 - a(L)$  præcis sæsonfrekvenserne ud af serien, så  $|1 - a(\omega)|^2$  er ca. én i sæsonfrekvenserne og tæt på nul imellem. Modsat er squared gain af sæsonkorrektionsfilteret nul i sæsonfrekvenserne, svarende til at de er helt elimineret. Teoretisk skulle  $|a(\omega)|^2$  være præcis én imellem sæsonfrekvenserne, men det er kun tilfældet for frekvenser lavere end  $\frac{\pi}{6}$ . For højere frekvenser introduceres pukler i  $|a(\omega)|^2$ , og generelt kommer der for mange frekvenser mere variation i output- end i input-serien. Det betyder i praksis, at der for nogle frekvenser er højere varians i den sæsonkorrigerede serie end i den ukorrigerede serie. Det må umiddelbart anses som et ubehageligt træk, og har også være kritiseret, se nedenfor eller i Planas (1996).

I forhold til det lineære filter (2), er der to ting, som kan gøre metoder ikke-lineær. Det første er korrektion for outliers i trin 2. Det andet gælder den multiplikative sæsonkorrektion, som stadig opregner til sum af 12 måneders data og ikke til produkt, så metoder ikke er lineær i logaritmer, se Ghysels, Granger og Siklos (1995).

### 2.1.2 Illustration

Som illustration af X-11 metoden betragter vi en additiv sæsonkorrektion af antal arbejdsløse i Danmark i perioden januar 1980 til august 1999. Sæsonkorrektionen foretages med standardoptioner, og der fjernes i første omgang ikke outliers i proceduren.

Den sæsonkorrigerede serie er vist sammen med de faktiske data i figur 2 (A). Perioden, hvor X-11 anvender det symmetriske filter,  $a(L)$ , er fremhævet med fed. Det er tydeligt, at sæsonkomponenten er af stor betydning for serien, men at sæsonen ikke ser konstant ud.

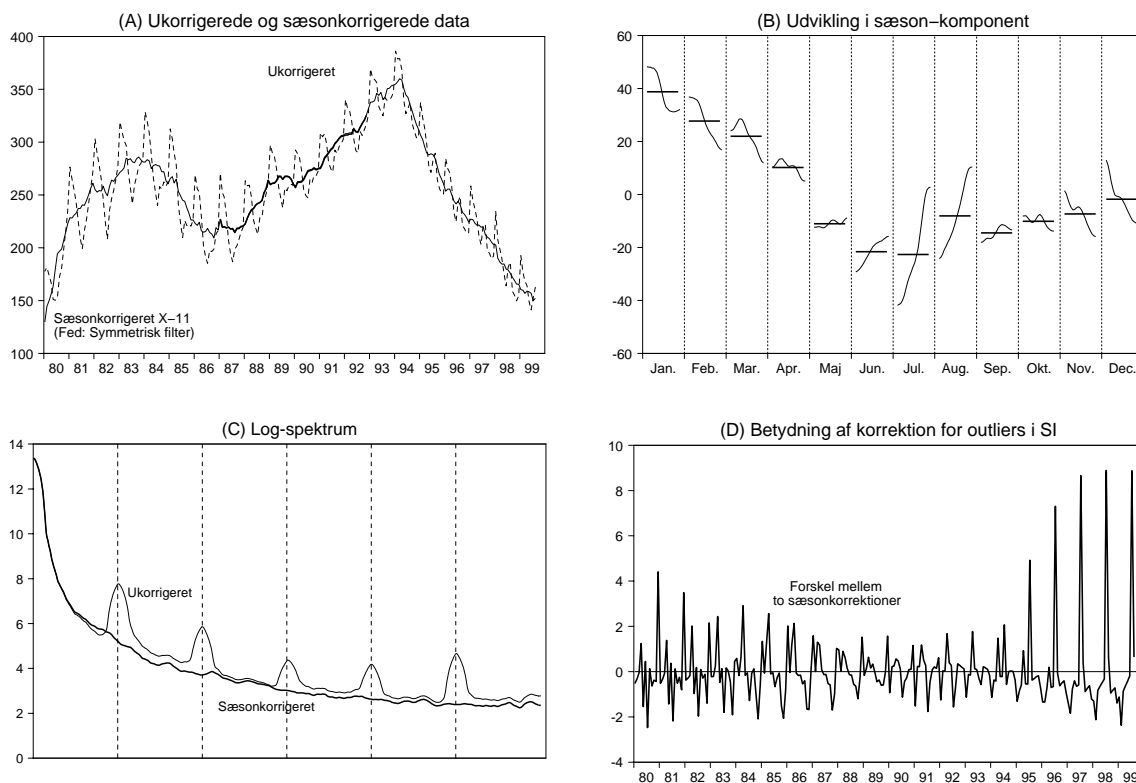
Sæsonkomponenterne findes som et glidende gennemsnit af de enkelte måneders observationer. Det er illustreret i figur (B), som for hver måned viser en tidsserie (af 20 år) for sæsonkomponenten. Den vandrette streg angiver gennemsnit for alle 20 år. Det generelle billede er ikke overraskende, at arbejdsløsheden er højere end gennemsnittet om vinteren og lavere om sommeren, men som det fremgår, har der været betydelige ændringer i sæsonmønstret over perioden, med en overvejende tendens til at sæsonudsvingene bliver mindre, dvs. sæsonfaktorer tættere på nul. Bevægelsen er tydeligst i sommermånederne, og kan bl.a. afspejle indførelsen af AF-ferie og uddannelsesorlov.

Den samlede effekt af filteret, nemlig det estimerede logaritmiske power-spektrum til de ukorrigerede og sæsonkorrigerede tal er vist i figur (C). X-11 lader til at fjerne sæsontoppene i data, og der er ikke store forskelle ved andre frekvenser.

Endelig er betydningen af at korrigere for outliers i  $I_t^{(1)}$ , dvs. graden af ikke-linearitet i trin 2, illustreret i figur (D). Det er tydeligvis ikke ligegyldigt, hvordan korrektionen for outliers sker, og der er relativt store systematiske forskelle mellem de to sæsonkorrektioner.

## 2.2 Census X-12-Arima

X-12-ARIMA er en videreudvikling af X-11. Metoden indeholder præcis samme sæsonkorrektionsrutine som X-11, men er udbygget med en Arima modelleringsfacilitet, REG-ARIMA, og en udvidelse af programmets test og kontrolmuligheder.



Figur 2: Sæsonkorrektion med X-11.

Arima-modellen anvendes dels til at korrigere for bl.a. handelsdage og til at finde outliers i modellen. Arima-modellen, som altså køres inden X-11 sæsonkorrigerer data, er dermed et alternativ til trin 2 ovenfor. Den nye Arima-model er angiveligt betydeligt bedre end OLS estimation på den irregulære komponent, se Findley et al. (1996)<sup>4</sup>.

Derudover anvendes Arima-modellen til at forudsige tidsserien. Det mindsker i princippet problemet med asymmetriske filtre. Hvis Arima-modellen leverer gode out-of-sample forecast af serien, kan det muligvis forbedre X-11s sæsonkorrektion i endepunkterne. Hvis Arima-modellen omvendt er misspecificeret, må resultatet intuitivt være det modsatte.

Følgende præsenteres REG-ARIMA modellen kort.

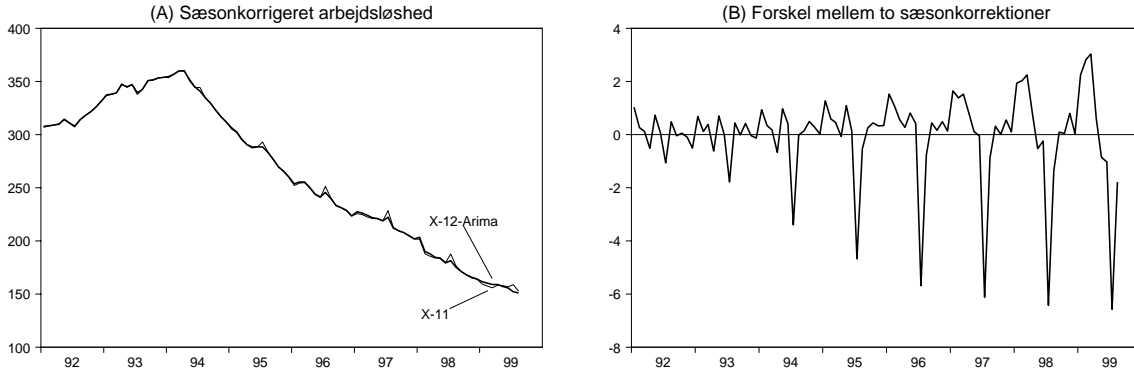
### 2.2.1 Reg-Arima

Betragt en tidsserie,  $Y_1, Y_2, \dots, Y_t, \dots, Y_T$ . Først foretages et test for, om data skal transformeres til logaritmer, svarende til en multiplikativ sæsonkorrektion. Vi betragter videre regressionsmodellen

$$Y_t = \beta' \mathbf{H}_t + Z_t$$

hvor  $\mathbf{H}_t$  er en vektor af (deterministiske) forklarende variable, fx til at tage højde for midelværdi, antal handelsdage, placering af påsken mv., se Bureau of the Census (1999). Det

<sup>4</sup>Det er ikke nødvendigvis en god fremgangsmåde at estimere på den irregulære komponent. Den er normalt præget af autokorrelation, og teststørrelser for signifikans har ikke de sædvanlige fordelinger. Derudover er den allerede korrigeret for sæson og vil naturligt have en stor del af sin varians lokaliseret omkring handelsdagsfrekvenser.



Figur 3: Sæsonkorrektion med X-11 og X-12.

antages, at  $Z_t$ , følger en generel  $(P, D, Q)(BP, BD, BQ)_{12}$  Arima-model

$$\phi(L) \cdot \delta(L) \cdot Z_t = \phi(L) \cdot \delta(L) \cdot (Y_t - \beta' \mathbf{H}_t) = \theta(L) \cdot a_t,$$

hvor  $a_t \sim Niid(0, \sigma_a^2)$  og polynomierne  $\phi(L)$ ,  $\delta(L)$  og  $\theta(L)$  har formen

$$\delta(L) = (1 - L)^D \cdot (1 - L^{12})^{BD} \quad (4)$$

$$\phi(L) = (1 + \phi_1 L + \dots + \phi_P L^P) \cdot (1 + \Phi_1 L^{12} + \dots + \Phi_{BP} L^{12 \cdot BP}) \quad (5)$$

$$\theta(L) = (1 + \theta_1 L + \dots + \theta_Q L^Q) \cdot (1 + \Theta_1 L^{12} + \dots + \Theta_{BQ} L^{12 \cdot BQ}). \quad (6)$$

Polynomiet  $\delta(L)$  indeholder modellens enhedsrødder, og  $D$  og  $BD$  angiver, hvor mange gange der skal tages henholdsvis ordinære og sæsondifferencer af  $Z_t$  for at opnå stationaritet. Polynomiet  $\phi(L)$  har udelukkende stationære autoregressive rødder, og  $\theta(L)$  er en invertibel MA proces.

Identifikationen af Arima modellens orden,  $(P, D, Q)(BP, BD, BQ)_{12}$ , foregår automatisk ud fra en række test på basis af kovariansstrukturen i data. Indenfor Arima-modellen korrigeres der for forskellige typer outliers, som fjernes med dummies. En nærmere præsentation gives i Bureau of the Census (1999) eller i Gómez og Maravall (1997).

### 2.2.2 Illustration

Vi illustrerer REG-ARIMA udvidelsen ved at sammenligne de sæsonkorrigerede tal fra X-11, som tillades at fjerne outliers i den irregulære komponent, med resultatet af X-12-ARIMA. X-12 køres med standardoptioner, se Hood og Findley (1999).

REG-ARIMA starter med at test for, om data skal transformeres til logaritmer. Programmet vælger for de månedlige arbejdsløshedstal ikke at transformere, svarende til at anvende en additiv sæsonkorrektion. REG-ARIMA identificerer modellen  $(0, 1, 1)(0, 1, 1)_{12}$ , den såkaldte Airline model, med parametre

$$(1 - L)(1 - L^{12})Y_t = (1 + 0.2045 \cdot L)(1 - 0.2277 \cdot L^{12})a_t,$$

som har acceptable statistiske egenskaber. I Modellen findes to additive outliers (1985 : 4 og 1987 : 1) med  $t$ -værdier på lidt over 4. Modellen anvendes til at lave forecast for serien, og herpå køres X-11s sæsonkorrektionsrutine.

Som det fremgår af figur 3 (A) og (B), er der specielt i juli måned temmelig stor forskel mod slutningen af samplet. I størrelsesorden 6.000 personer. Det fremgår af figur (A), at der er systematiske hakker i juli i begge sæsonkorrigerede serier, men at de er mindst i X-12-ARIMA. For det aktuelle serie er der dermed indikationer af, at Arima udvidelsen kan forbedre sæsonkorrektionen i slutningen af samplet fx ved at muliggøre en hurtigere tilpasning til de ændrede regler om uddannelsesorlov, men forskellen er ikke dramatisk.

## 2.3 Tramo/Seats

Sæsonkorrektionsmetoden TRAMO/SEATS fra Banco de España er baseret på en dekomponering af estimerede Arima modeller. Ideen er, at man i stedet for at vælge mellem et på forhånd defineret sæt af filtre som i X-11, udleder et optimalt filter givet den observerede tidsrække. Arima modelbaseret sæsonkorrektion er fx præsenteret i Hillmer og Tiao (1982) eller i Planas (1997a).

Proceduren involverer to programmer: TRAMO anvendes til at linearisere data og til at identificere og estimere en Arima-model. Lineariseringen består i en forhåndskorrektion for outliers og deterministiske komponenter,  $H_t$ . Som sådan har TRAMO samme funktion som X-12s REG-ARIMA, og vil ikke blive præsenteret nærmere<sup>5</sup>. Herefter sendes Arima-modellen og de lineariserede data videre til programmet SEATS, som forestår dekomponeringen.

Dekomponeringen i SEATS foregår i to trin. Først specificeres modellerne for komponenterne konsistent med den estimerede model i TRAMO, og derefter bruges en signal extraction metode til at estimere komponenterne.

### 2.3.1 Dekomponering og identifikation

Udgangspunktet er fortsat den ortogonale additive dekomponering (1). Det antages nu, at komponenterne hver især kan beskrives af Arima modeller

$$\phi_S(L) \cdot \delta_S(L) \cdot S_t = \theta_S(L) \cdot a_{St} \quad (7)$$

$$\phi_T(L) \cdot \delta_T(L) \cdot T_t = \theta_T(L) \cdot a_{Tt} \quad (8)$$

$$I_t = a_{It}, \quad (9)$$

hvor polynomierne er af samme form som (4), (5) og (6).  $\phi_S(L) \cdot \delta_S(L) \cdot S_t$  og  $\phi_T(L) \cdot \delta_T(L) \cdot T_t$  er stationære, og  $a_{St}$  og  $a_{Tt}$  er uafhængige normalfordelte variable. De to polynomier  $\phi_S(L) \cdot \delta_S(L)$  og  $\phi_T(L) \cdot \delta_T(L)$  kan have rødder på eller udenfor enhedscirklen, men har ingen fælles rødder. Den antagelse svarer intuitivt til, at de dynamiske egenskaber er forskellige i de ikke-observerede komponenter, som dermed ikke har toppe i spektrum ved samme frekvenser. Endelig antages det, at  $\theta_S(L)$  og  $\theta_T(L)$  ikke har fælles enhedsrødder.

Additiviteten af (1) implicerer, at de observerede data  $Y_t$  er givet ved Arima-modellen

$$\phi(L) \cdot \delta(L) \cdot Y_t = \theta(L) \cdot a_t, \quad (10)$$

hvor polynomierne er relateret til komponentmodellerne som

$$\phi(L) \cdot \delta(L) = \phi_S(L) \cdot \delta_S(L) \cdot \phi_T(L) \cdot \delta_T(L) \quad (11)$$

---

<sup>5</sup>Sammenligning af TRAMO og REG-ARIMA peger på, at modelidentifikationen i TRAMO er bedre og hurtigere end i REG-ARIMA, se Dossé og Planas (1996). Bureau of the Census har planer om at erstatte REG-ARIMA med TRAMO i den næste version af X-12.

og

$$\begin{aligned}\theta(L) \cdot a_t &= \phi_T(L) \cdot \delta_T(L) \cdot \theta_S(L) \cdot a_{St} + \phi_S(L) \cdot \delta_S(L) \cdot \theta_T(L) \cdot a_{Tt} \\ &\quad + \phi_T(L) \cdot \delta_T(L) \cdot \phi_S(L) \cdot \delta_S(L) \cdot a_{It}.\end{aligned}\tag{12}$$

Modellen (10) er en Arima model for den observerede tidsrække, som antages kendt, men som i praksis er estimeret i TRAMO. Det betyder, at vi kender polynomierne  $\phi(L)$ ,  $\delta(L)$  og  $\theta(L)$  sammen med variansen af  $a_t$ . Vi vil nu vise, hvordan modellen for  $Y_t$  kan dekomponeres i modeller for komponenterne  $S_t$ ,  $T_t$  og  $I_t$ .

Som det fremgår af (11) svarer den additive dekomponering af Arima modeller til en faktorisering af det autoregressive polynomium. Det svarer blot til at dele rødderne i polynomiet ud mellem komponenterne. Vi begynder med at se på  $\delta(L)$ , jf. (4). Enhedsrødderne i polynomiet  $(1 - L)^D$  svarer til bevægelser ved nul-frekvens og vil naturligt henregnes trenden,  $T_t$ . Polynomiet  $1 - L^{12}$  har de 12 enhedsrødder  $z = \{\pm 1, \pm \frac{1}{2} \pm \frac{1}{2}i\sqrt{3}, \pm i, \pm \frac{1}{2}\sqrt{3} \pm \frac{1}{2}i\}$  jævnt fordelt på enhedscirklen hørende til frekvenserne  $\frac{2\pi k}{12}$ . Roden  $z = 1$  er identisk med roden i  $1 - L$ , som igen tildeles trenden. Vi faktorerer derfor

$$(1 - L^{12}) = (1 - L)(1 + L + L^2 + \dots + L^{11}) \equiv (1 - L)U(L),$$

og de 11 enhedsrødder i  $U(L)$  tildeles sæsonkomponenten. For at se intuitionen betragtes en simpel model for sæsonen

$$U(L)S_t = \sum_{i=0}^{11} S_{t-i} = 0,$$

som siger at summen af 12 måneders sæsonkomponenter skal være lig nul. Det svarer til deterministiske sæson,  $S_t = S_{t-12}$ . Med stokastik på højresiden  $U(L)S_t = \epsilon_t$ , er sæsonen ikke deterministisk, men fortolkningen er parallel; at 12 måneders sum af sæsonkomponenterne kun må afvige marginalt fra nul og specielt er nul i forventning.

Tilbage er der nu at fordele rødderne i polynomiet  $\phi(L)$ . De går alle til trend-komponenten med den ene undtagelse, at komplekse rødder  $\lambda$ , som ligger tilstrækkeligt tæt på en sæson-frekvens,  $\lambda \in \{\frac{2\pi k}{12} \pm \kappa\}$ , vil tildeles sæsonkomponenten, hvor  $\kappa$  sættes af brugeren<sup>6</sup>.

Dekomponeringen er indtil nu foregået udelukkende som en faktorisering af det autoregressive polynomium. Dvs., at  $\phi_S(L) \cdot \delta_S(L)$  og  $\phi_T(L) \cdot \delta_T(L)$  nu er identificeret. Næste trin er at finde højresiderne af Arima-modellerne for komponenterne, dvs. MA-polynomierne  $\theta_S(L)$  og  $\theta_T(L)$  samt varianserne på  $a_{St}$ ,  $a_{Tt}$  og  $a_{It}$ . Der betragtes kun modeller (7) og (8), som er balanceret, så AR og MA polynomiet er af samme orden. Sammenhængen mellem MA polynomierne er givet i ligning (12), og enhver dekomponering, som opfylder denne ligning givet faktoriseringen af AR polynomiet kaldes tilladelig. Almindeligvis findes der uendelig mange tilladelige dekomponeringer, se Hillmer og Tiao (1982). Vi behøver derfor en arbitrær antagelse til at vælge mellem de tilladelige dekomponeringer.

Det er de autoregressive rødder, der bestemmer toppene i power-spektrum, men uden MA-polynomierne ved vi intet om variansen. Det kan vises, at de tilladelige dekomponeringer kun afviger fra hinanden med hvid støj. Det svarer til, at power-spektrum for de tilladelige dekomponeringer er parallelle. Den antagelse, som anvendes i SEATS, er den kanoniske dekomponering. Den går ud på at maksimere variansen i den irregulære komponent, hvorved power-spektrum for de øvrige komponenter er parallelforskydt så langt

---

<sup>6</sup>Bemærk, at den del af sæsonkomponenten som stammer fra  $\phi(L)$  er lidt usædvanlig og fortolkningsmæssigt anderledes end  $U(L)$ . Det er ikke en enhedsrod, og fanger dermed stationær variation i data, som gentages med regelmæssige mellemrum.

ned som muligt. Fra (1) gælder, at spektra er additive

$$g_Y(\omega) = g_S(\omega) + g_T(\omega) + g_I(\omega),$$

og vi definerer nu minima af komponenternes spektra

$$\epsilon_S = \min_{\omega} \{g_S(\omega)\}, \quad \epsilon_T = \min_{\omega} \{g_T(\omega)\} \quad \text{og} \quad \epsilon_I = \min_{\omega} \{g_I(\omega)\}.$$

Størrelsen  $\epsilon_S + \epsilon_T + \epsilon_I$  kan fortolkes som den mængde ren støj i modellen, som kan omfordeles indenfor de tilladelige dekomponeringer. Vi fjerner nu så meget støj som muligt fra signalerne,  $S_t$  og  $T_t$ , og opnår de nye spektra

$$\begin{aligned} g_S^0(\omega) &= g_S(\omega) - \epsilon_S \\ g_T^0(\omega) &= g_T(\omega) - \epsilon_T \\ g_I^0(\omega) &= g_I(\omega) + \epsilon_S + \epsilon_T. \end{aligned}$$

Det betyder, at spektrene  $g_S^0(\omega)$  og  $g_T^0(\omega)$  nu indeholder et nul og dermed beskriver et ikke-invertibel signal, og at støjen er maksimeret i  $g_I^0(\omega)$  indenfor de tilladelige dekomponeringer.

Den kanoniske dekomponering er unikt identificeret og minimerer variansen på sæsonfaktorer og trend. Enhver anden dekomponering  $\tilde{S}_t$  og  $\tilde{T}_t$  kan skrives som den kanoniske plus støj,  $\tilde{S}_t = S_t + \epsilon_{1t}$ ,  $\tilde{T}_t = T_t + \epsilon_{2t}$ , hvor  $\epsilon_{1t}$  og  $\epsilon_{2t}$  er hvid støj, se Hillmer og Tiao (1982).

Givet kendskab til både AR og MA polynomier i modellerne (7) og (8) og varianserne på  $a_{St}$ ,  $a_{Tt}$  og  $a_{It}$ , er alle modellerne kendt. Betinget på disse modeller kan man med signal extraction metoder finde en MINIMUM MEAN SQUARED ERROR estimator af komponenterne,  $S_t$ ,  $T_t$  og  $I_t$ . Det foregår i spektralt domæne. For  $j = Y, S, T$  kan spektret skrives som

$$g_j(\omega) = \text{Var}[a_j] \cdot \frac{\phi_j(e^{-i\omega}) \cdot \delta_j(e^{-i\omega}) \cdot \phi_j(e^{i\omega}) \cdot \delta_j(e^{i\omega})}{\theta_j(e^{-i\omega}) \cdot \theta_j(e^{i\omega})},$$

se fx Planas (1996)<sup>7</sup>.

Estimationen anvender et WIENER-KOLMOGOROV (WK) filter, som er forholdet mellem spektrum for det ønskede signal og spektrum for den observerede tidsserie, så fx

$$\nu_S(\omega) = \frac{g_S(\omega)}{g_Y(\omega)}.$$

Da vil sæsonkomponenten efter en invers Fourier transformation kunne skrives som det lineære MA-filter

$$S_t = \nu_s(L)Y_t.$$

WK-filteret  $\nu_S(\omega)$  er et uendeligt men konvergent symmetrisk lineært filter, som ligner X-11 filteret,  $a(L)$ . Det kan vises, at  $a(L)$  er approksimativt optimalt i en gruppe af modeller, se fx Burrige og Wallis (1984). For data genereret af sådanne modeller, vil de modelbaserede metoder give stort set samme resultat som X-11. For alle andre modeller vil WK-filteret tilpasses data, mens X-11 vil anvende et fast filter.

---

<sup>7</sup>Strengt taget er spektrum kun defineret for stationære variable, da enhedsrødder implicerer uendelige toppe i spektrum og dermed uendelig varians. Metoden kan imidlertid generaliseres til ikke-stationære variable, hvor  $g_i$  omtales som pseudo-spektrum.

### 2.3.2 Illustration

Som illustration af TRAMO/SEATS betragter vi en sæsonkorrektion af industriens mængdemæssige omsætning for samlet 1985 : 1 – 1999 : 8.

TRAMO tester for specifikation og finder, at data skal transformeres til logaritmer. Det svaret til en multiplikativ sæsonkorrektion. Derefter identificeres Arima-modellen  $(1, 0, 1)$   $(0, 1, 1)_{12}$ , som estimeres til

$$(1 - 0.868L)(1 - L^{12})Y_t = (1 - 0.420L)(1 - 0.575L^{12})a_t.$$

Der estimeres en signifikant effekt fra handelsdage og placering af påsken. Disse deterministiske komponenter fjernes fra data inden dekomponering og lægges derefter tilbage som kalendereffekter i sæsonkomponenten. Der findes ingen outliers i serien, og modellen har rimelige statistiske egenskaber.

Dekomponeringen foregår på de autoregressive polynomier. Polynomiet  $(1 - L^{12})$  faktoriiseres med  $(1 - L)$  til trenden og  $U(L)$  til sæsonkomponenten, mens polynomiet  $(1 - 0.868L)$  går til trenden. Samlet haves modellerne

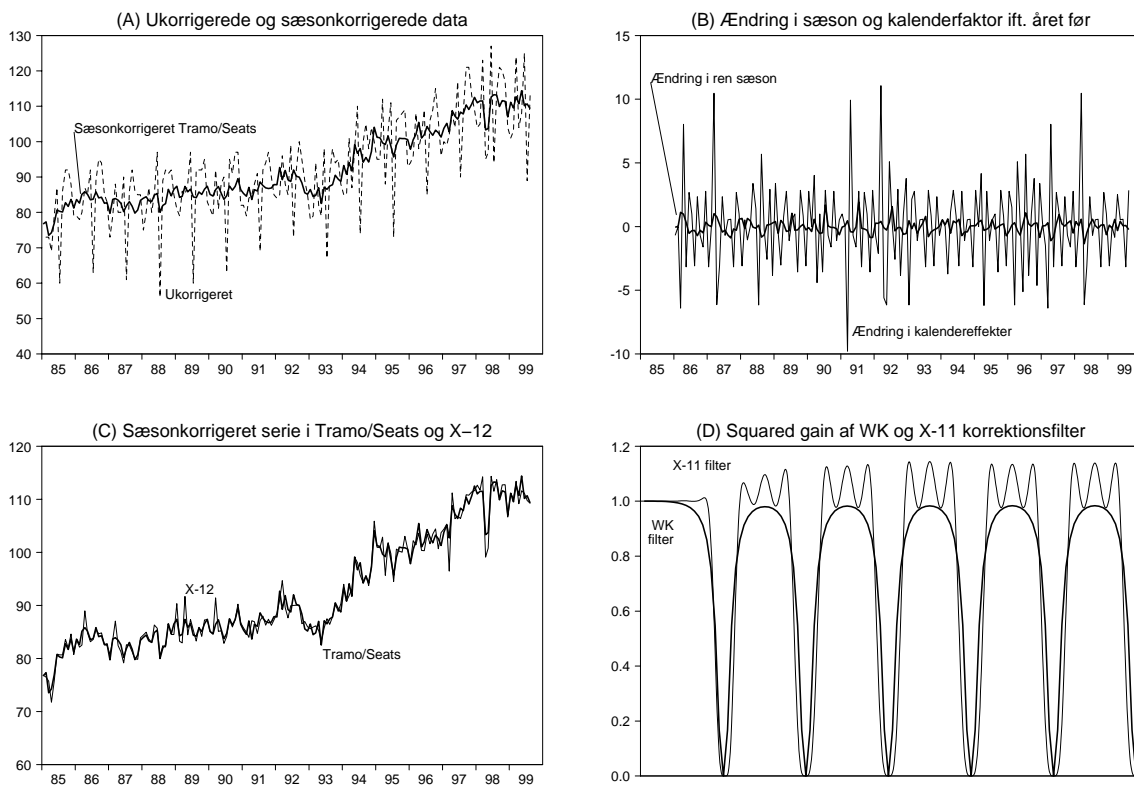
$$\begin{aligned} (1 - 0.868L)(1 - L)T_t &= \theta_T(L)a_{Tt} \\ U(L)S &= \theta_S(L)a_{St} \\ I &= a_{It}, \end{aligned}$$

hvor  $\theta_T(L)$  og  $\theta_S(L)$  er af hhv. orden 2 og 11 ligesom den autoregressive del af modellerne. Dermed er trenden  $I(1)$ , men med en stor stationær autoregressiv rod. Varianserne på innovationerne bestemmes til  $Var(a_{Tt}) = 0.0504$ ,  $Var(a_{St}) = 0.05476$  og  $Var(a_{It}) = 0.35871$ . Givet disse modeller anvendes et WK-filter til at estimere komponenterne.

Resultatet af sæsonkorrektionen er vist i figur 4. (A) viser den ukorrigerede og sæsonkorrigerede serie. Sæsonen er oplagt af stor betydning for industriens omsætning, men ser ikke så stabil ud, og der er tydeligt en del støj i tidsserien. Den irregulære sæson afspejler diverse kalendereffekter, som det fremgår fra figur (B). Selve sæsonkomponenten ændrer sig kun langsomt fra år til år, mens der er store udsving i kalender-effekterne. De største udsving vedrører placeringen af påsken, mens den mere regulære negative autokorrelation vedrører antallet handelsdage i hver måned.

Figur (C) sammenligner den sæsonkorrigerede serie med en tilsvarende sæsonkorrigeret serie fra X-12. Det overordnede indtryk af de to serier er det samme, men der er tydeligt mere variation i den sæsonkorrigerede serie fra X-12 end i TRAMO/SEATS serien. Standardafvigelsen i den logaritmiske ændring i den sæsonkorrigerede serie er således 0.038 i X-12 serie mod 0.026 i TRAMO/SEATS serien.

Det resultat kan umiddelbart undre, eftersom den kanoniske dekomponering præcis minimerer støjen i  $S_t$  og derved maksimerer støjen i  $I_t$  og  $A_t$ . Forklaringen kan imidlertid findes i figur (D), som viser squared gain af X-11 filteret og af WK-filteret for modellen for industriens omsætning. For frekvenser imellem sæsontoppene er squared gain for X-11 filteret flere steder over én, svarende til at der er mere varians i den sæsonkorrigerede serie end i den ukorrigerede serie. Det vil sige, at X-11 (og dermed X-12) overvurderer kortsigtsdynamikken i den sæsonkorrigerede serie, se også Planas (1996). Intuitionen er, at vægtene i figur 1 (B) implicerer, at variansen blæses kunstigt op ved visse frekvenser.



Figur 4: Sæsonkorrektion med TRAMO/SEATS.

### 3 Simulation

Det er ikke nogen enkel sag at sammenligne forskellige sæsonkorrektionsmetoder empirisk. Første problem er, at sæsonkomponenten per definition er ikke-observeret, så metodernes præcision kan ikke umiddelbart testes på faktiske tal. Det kan løses ved at anvende simulerede tal og nedenfor forsøges det at illustrere metodernes præcision ved at isolere en kendt sæsonfaktor<sup>8</sup>.

Der er genereret 15 tidsserier af 120 observationer, som hver især er sammensat af en trend, en sæsonfaktor og en irregulær komponent. Serierne har alle et udgangsniveau på 1000 med en blød stigende trend, men det er ikke af så stor betydning. Dernæst er der genereret 3 sæsonfaktorer

- (I) En konstant deterministisk sæson, som varierer mellem  $-10$  og  $+28$  pct.
- (II) En sæsonkomponent som midt i samplet gradvis ændres.
- (III) En sæsonkomponent som ændres i ét hug midt i samplet.

Den irregulære komponent er specificeret på 5 forskellige måder.

- (i) En normalfordelt variabel med gennemsnit 1 og standardafvigelse 0,02.
- (ii) Normalfordeling med standardafvigelse 0,07.
- (iii) Som (ii) men med 12 måneders autokorrelation,  $I_t^3 = 0.3 \cdot I_t^2 + 0.7 \cdot I_{t-12}^3$ .

<sup>8</sup>Bell og Hillmer (1984) argumenterer for, at simulationseksperimenter ikke nødvendigvis er til meget hjælp, fordi resultatet kan favorisere de metoder, hvis implicite eller eksplicite statistiske model ligger tættest på den model, der har genereret data.

Tabel 1: Afvigelse mellem faktisk og estimeret sæsonkomponent

Simuleret model		Sæsonkorrektionsmetode					
Sæson	Irregulær	OLS	X-11 <sup>(1)</sup>	X-11	X-12	T/S <sup>(2)</sup>	SEATS <sup>(3)</sup>
Stabil	Lille varians	0.900	0.893	0.939	0.970	0.886	<b>0.684</b>
Stabil	Stor varians	2.466	3.071	3.264	3.044	2.601	<b>2.404</b>
Stabil	+ autokor.	<b>2.296</b>	2.963	3.012	2.993	2.806	2.624
Stabil	+ outliers	2.646	3.514	3.337	2.993	2.640	<b>2.591</b>
Stabil	+ niveauskift	2.187	2.763	2.708	2.618	<b>1.931</b>	2.153
Skiftende	Lille varians	1.691	<b>0.999</b>	1.150	1.094	1.052	1.015
Skiftende	Stor varians	2.856	3.104	3.296	3.264	2.842	<b>2.749</b>
Skiftende	+ autokor.	<b>2.712</b>	3.003	3.090	3.105	3.007	2.935
Skiftende	+ outliers	2.997	3.586	3.494	3.327	3.007	<b>2.935</b>
Skiftende	+ niveauskift	2.608	2.857	2.787	2.775	<b>2.301</b>	2.559
Brud	Lille varians	7.330	3.110	3.848	4.154	2.808	<b>2.205</b>
Brud	Stor varians	7.690	4.307	4.762	4.812	4.730	<b>4.295</b>
Brud	+ autokor.	7.636	<b>4.204</b>	4.685	4.770	4.748	4.234
Brud	+ outliers	7.775	<b>4.848</b>	5.608	5.438	5.255	4.958
Brud	+ niveauskift	7.581	4.196	4.395	4.341	<b>3.860</b>	4.113

Noter: <sup>(1)</sup> X-11 filter uden korrektion for outliers. <sup>(2)</sup> TRAMO/SEATS. <sup>(3)</sup> Sæson korrektion med SEATS og en Airline model. Se yderligere i teksten. Fed indikerer mindst afvigelse.

(iv) Som (ii) men med 3 additive outliers.

(v) Som (ii) men med et databrud i form af et niveauskift.

Tidsserierne er genereret for 10 trækninger fra en normalfordeling. Metoderne præcision i at identificere den kendte sæsonfaktor beregnes som den gennemsnitlige kvadrerede afvigelse

$$D = \sqrt{\frac{1}{T} \cdot \sum_{i=1}^T (\hat{S}_t - S_t)^2}. \text{ Der betragtes 6 sæsonkorrektionsmetoder.}$$

1. En simpel OLS sæsonkorrektion. Den estimerer en fleksibel trendkomponent i form af et polynomie i trenden og dummies for de enkelte måneder. Metoden kan anvendes som bench-mark for resultatet af de mere avancerede sæsonkorrektionsmetoder.
2. X-11 i en ren lineær filter-version uden fjernelse af outliers.
3. Default X-11 sæsonkorrektion med automatisk valg af filterlængde mv.
4. Standardopsætning af X-12-ARIMA, se Hood og Findley (1999) for optioner.
5. Default TRAMO/SEATS.
6. SEATS baseret på en simpel Airline model  $(0, 1, 1)(0, 1, 1)_{12}$  og uden korrektion for outliers mv.

Tabel 1 viser for hver af de 15 tidsserier den gennemsnitlige afvigelse af de 10 realisationer. Det generelle indtryk er, at de avancerede metoder ikke er meget bedre end bench-mark OLS i tilfælde med stabil sæson, men at de som ventet er markant bedre ved brud i sæsonmønsteret.

Man kan bemærke, at X-12 lidt overraskende ikke klarer sig væsentlig bedre end de ældre versioner af X-11. Faktisk løser X-12 ikke nogen af opgaverne bedst. Omvendt ser TRAMO/SEATS ud til at klare sig glimrende, og har mindst afvigelser i 10 ud af de 15 modeller. Lidt overraskende ser det ud til, at den simple anvendelse af SEATS Airline model i mange tilfælde er bedre end den fulde anvendelse TRAMO/SEATS.

Man skal generelt være forsigtig med for håndfaste konklusioner på simulationsøvelser som denne, men resultatet ovenfor mellem X-12 og TRAMO/SEATS er alligevel slående. Det skal også ses i lyset af, at data er genereret på en måde som umiddelbart ligger langt fra Arima-tilgangen og dermed må vurderes ikke a priori at favorisere TRAMO/SEATS.

## 4 Revisioner af de sæsonkorrigerede tal

Der er sædvanligvis stor interesse for størrelsen af revisionerne i de sæsonkorrigerede tal, og ud fra et praktisk hensyn vil man foretrække metoder med mindst revisioner. Revisioner udtrykker imidlertid anvendelse af ny information, og der er en modstrid imellem ønsket om en præcis bestemmelse af sæsonfaktoren og små revisioner, se Fischer (1995). Således kan små revisioner udtrykke en dårlig anvendelse af ny information, og dermed en hurtig konvergens mod et inkonsistent resultat.

Nedenfor betragtes revisionerne i henhold til X-11/X-12 modellerne og TRAMO/SEATS. Der er ikke gjort noget for at minimere revisionerne, og udgangspunktet er, at de betragtede konfigurationer skal kunne køres automatisk. Som eksempel betragtes sæsonkorrektionen af den danske pengemængde som den opgøres af Danmarks Nationalbank.

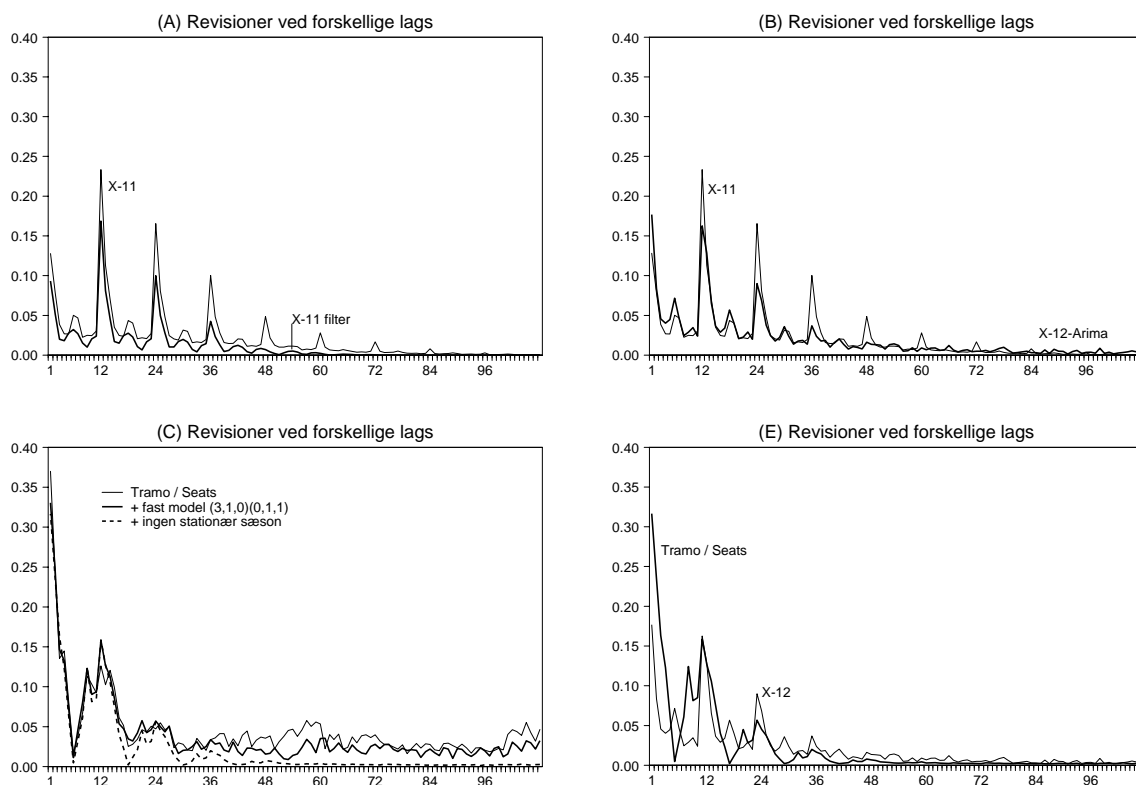
### 4.1 Revisioner i X-11 og X-12

Som præsenteret i afsnit 2.1.1 kan X-11 i grundformen opfattes om et lineært symmetrisk filter med en længde på 84 perioder på hver side. I X-11 er revisionerne derfor udtryk for, at de anvendte ikke-symmetriske filtre ved endepunkterne kommer tættere og tættere på den endelige symmetriske filter. Længden af filteret bestemmer derfor entydigt, hvornår de sæsonkorrigerede data er endelige.

For at illustrere mønsteret i revisionerne er pengemængden sæsonkorrigeret 36 gange for sampler som udvides med en observation af gangen;  $t = 1975 : 4 - T$ , hvor  $T = 1996 : 8 - 1999 : 7$ . Herefter er den gennemsnitlige absolutte revision udregnet. Mønsteret af revisionerne er illustreret i figur 5, som viser hvordan fremkomsten af en ny observation påvirker sæsonkorrektionen af tidligere data.

Som det fremgår af figur (a) afspejler revisionsmønsteret for standardfilteret af X-11 vægtene fra figur 1. Der er således særligt store revisioner ved første lag, i størrelsesorden 1 promille, og ved sæsonlags,  $t - 12$ ,  $t - 24$  og  $t - 36$ . Sammen med den simple X-11 er indtegnet default X-11, hvor der fjernes outliers i den irregulære komponent. Det fremgår, at trin 2 i X-11 generelt øger revisionerne, og at filteret de facto er betydeligt længere end de 84 perioder. At fjernelse af outliers gør filteret ikke-lineært og normalt forlænger filteret er velkendt, se fx Ghysels, Granger og Siklos (1995) samt Hood og Findley (1999).

Sammenlignet med grundformen af X-11 kan Arima-modellen i X-12 mindske revisionerne, fordi tidsserien forecastes, så der kan anvendes et mindre asymmetrisk filter. Modstykket er, at Arima-modellens forecast-fejl afspejles i revisionerne. Hvis forecastfejlene er nul, skal der ikke revideres i de tidligere sæsonkorrigerede tal. Er fejlene omvendt store, kan der forekomme store revisioner. I figur (B) sammenlignes standard X-11 modellen med X-12-ARIMA. Det fremgår, at revisionerne øges for de første lags, mens de for længere lags mindskes. Specielt er noget af systematikken i X-11 ved sæsonlags større end 36 fjernet i X-12. Til gengæld lader det til, at filteret i X-12-ARIMA de facto er væsentlig længere, og der er ingen af de sæsonkorrigerede værdier, som bliver endelige indenfor længden af det betragtede sample.



Figur 5: Revisioner i forskellig sæsonkorrektionsmetoder.

For at sammenligne de samlede revisioner som en sæsonkorrektionsmetode i gennemsnit foretager, udregnes arealerne under kurverne i figur 5. Fortolkningen af arealerne er den procentuelle gennemsnitlige absolutte revision i hele serien som en ekstra observation afstedkommer. Arealerne op til 84 lags og for hele tidsserien er angivet i tabel 2. Som det fremgår, har X-12 mindre revisioner end X-11 indenfor 84 observationer, mens det omvendte er tilfældet for hele samplet. Umiddelbart kan det undre, at X-12 ikke er bedre relativt til X-11, fordi Arima udvidelsen er skabt med henblik på at mindske revisionerne.

Tabel 2: Summeret gennemsnitlig absolut revision

Metode	84 perioder	hele samplet
X-11 (filter)	1.280	1.280
X-11	2.287	2.322
X-12	2.128	2.530
TRAMO/SEATS	4.287	10.496
- Fast model	3.682	7.600
- Fast model + ingen stationær sæson	2.404	2.826
SEATS Airline model	2.212	2.594

Anm.: Tallene angiver arealerne under kurverne i figur 6.

## 4.2 Revisioner i Tramo/Seats

I TRAMO/SEATS er der flere kilder til revisioner. Den første er som i X-12-ARIMA forecast fejl i Arima-modellen. Hertil skal lægges, at Tramo i hver periode identificerer og estimerer en ny Arima-model. I det omfang modellen ændres fra et sample til et andet, vil der kunne opstå (ganske store) ændringer i den sæsonkorrigerede serie. Dermed er revisionsmønsteret i TRAMO/SEATS i højere grad en funktion af data og af den estimerede model.

Revisionerne af den sæsonkorrigerede pengemængde er vist i figur (C). Som det fremgår er revisionsmønsteret ret forskelligt fra X-11, og specielt er den gennemsnitlige revision på første lag betydeligt større, i størrelsesorden 4 promille. Det er tydeligt, at filteret i praksis er meget langt, og der er ingen tendens til, at de gennemsnitlige revisionerne konvergerer mod nul for lange lags.

De store gennemsnitlige revisioner afspejler mange sæsonkorrektioner med små revisioner og nogle få tilfælde med store revisioner i hele samplet for den sæsonkorrigerede pengemængde, som hænger sammen med skift af modellen.

For at se betydningen af modelskift, er det forsøgt at binde modellen til den mest almindeligt identificerede model for pengemængdetallene, nemlig  $(3, 1, 0)(0, 1, 1)_{12}$ . Som det fremgår af figur (C) mindsker det revisionerne marginalt, men billedet er fortsat det samme. Det indikerer, at de identificerede modeller generelt ligger ret tæt på hinanden, men at de estimerede parametre har betydning.

Generelt burde parameterværdierne ikke betyde så meget, men der er en enkelt undtagelse, nemlig hvis en rod i  $\phi(L)$  tildeles sæsonkomponenten. De store revisioner er således udtryk for, at en rod i det autoregressive polynomium hopper frem og tilbage mellem sæsonkomponenten og trenden afhængig af samplelængden. Den autoregressive rod implicerer en top i seriens spektrum. I de tilfælde, hvor den ligger tilstrækkelig tæt på en sæsonfrekvens, tilregnes den sæsonkomponenten. Det betyder, at en af toppene i sæsonkomponentens spektrum tillades at være lidt bredere, så flere frekvenser elimineres fra de sæsonkorrigerede data som derved bliver glattere. De autoregressive rødder kan i visse tilfælde approksimere handelsdagsskorrektion, og resultatet kan indikere, at der er tendens til handelsdagseffekter i pengemængden.

Betydningen er illustreret i figur (C), hvor roden i er bundet til at ligge i trenden, svarende til at undgå en stationær del af sæsonkomponenten. Revisionerne i TRAMO/SEATS, hvor roden er bundet til trenden, er sammeliget med X-12-ARIMA i figur (D). Generelt er revisionerne nu kun lidt større end i X-12, jf. også tabel 1. Den mest markante forskel er første lag, hvor revisionerne i TRAMO/SEATS er næsten dobbelt så store som i X-12. Det antyder, at de seneste sæsonkorrigerede værdier i TRAMO/SEATS er tilsyneladende relativt dårligt bestemt, eller at TRAMO/SEATS anvender mere information i den nyeste observation til at korrigere historiske tal.

Valget af den kanoniske dekomponering i SEATS har betydning for revisionernes størrelse, og det kan vises, at den kanoniske dekomponering maksimerer revisionerne, se Planas (1997b). Andre dekomponeringer har andre egenskaber, og der arbejdes fx på at opstille en dekomponering, som minimerer revisionerne.

Sammenfattende kan man sige, at der er grund til at være opmærksom ved anvendelse af TRAMO/SEATS til sæsonkorrektion af den samme serie måned efter måned. I Gómez og Maravall (1998a) anbefales det at opstille en Arima model en gang om året og derefter binde modellens orden, outliers og lignende det næste år. Ovenstående viser, at man også skal binde eventuelle autoregressive komplekse rødder til at ligge i trenden. I stedet for den lidt langsommelige process med at binde Arima-modellen til at have en bestemt orden, kan man

i stedet anvende Airline modellen  $(0, 1, 1)(0, 1, 1)_{12}$ . Det svarer til at acceptere et mindre fleksibelt filter, som nu afhænger af data ved estimation af to MA-parametre og én varians. Revisionerne i Airline modellen er generelt mindre, jf. tabel 2, og er ret tæt på revisionerne i X-12.

## 5 Konklusion

Dette papir har sammenlignet de to sæsonkorrektionsmetoder X-11/X-12 og TRAMO/ SEATS.

Det er næppe nogen tvivl om, at det teoretiske set-up i TRAMO/SEATS er langt det mest elegante, og metoden har mange teoretiske fordele. De to måske mest centrale er for de første, at metoden tilpasser sig de aktuelle data, og udleder et i princippet optimalt filter. Det leder frem til den anden store fordel, nemlig det klart specificerede udgangspunkt. Filteret er optimalt givet karakteristikken af komponenterne, og med en anden karakteristik findes en anden estimator af de sæsonkorrigerede tal. Det betyder, at forudsætninger og model er klart formuleret. Det gør det lettere at formulere kritik af modellen, og det giver umiddelbart mulighed for at afprøve forskellige antagelser om den statistiske model. Fra dette approach kan der udledes flere testbare implikationer, og ud fra en videnskabsteoretisk synsvinkel, indeholder dette approach givetvis størst udviklingspotentiale.

Det store spørgsmål ved metoden er, om det i praksis er muligt at estimere en velspecificeret Arima-model for data, så de teoretiske fordele kan overføres til praksis. Hvis det ikke er tilfældet, er der ikke meget begrundelse i de teoretiske fordele.

I dette papir blev programmerne sammenlignet ud fra simulerede tal. Konklusionen er oplagt, at TRAMO/SEATS også i praksis synes at være et godt alternativ til X-12. Således var indtrykket, at programmet fint var i stand til at isolere den kendte sæsonkomponent, endda klart bedre end X-11/X-12. Hertil kommer, at X-11 og X-12 har en tendens til at overkorrigere data, så der faktisk er mere variation i den sæsonkorrigerede end i den ukorrigerede tidsserie.

I analysen af revisioner af tallene blev nogle af svaghederne ved metoden illustreret. Der er flere kilder til revisioner, og nogle revisioner har karakter af kvalitative ændringer af de sæsonkorrigerede tal. En del af de problemer kan fjernes på bekostning af fleksibilitet, ved at restrikttere modellen. De mindske revisionerne betragteligt, og den simple anvendelse Seats på Airline modellen var faktisk bedre end TRAMO/SEATS til at identificere den kendte sæson. Det er dog fortsat tendens til, at revisionen af den foregående periodes sæsonkorrigerede tal er betydeligt større end i X-11 og X-12. Det er ikke nødvendigvis problematisk i sig selv, men er noget man bør være opmærksom på.

En foreløbig konklusion må være, at TRAMO/SEATS er et ikke bare meget interessant men også brugbart alternativ til de mere almindelige X-11 og X-12. Generelt kræver programmet nok mere opmærksomhed, men i praksis er det nok også så meget et spørgsmål om at få indbygget mere brugererfaring i metoden.

## Referencer

- [1] Bell, William R. og Steven C. Hillmer (1984): Issues Involved with the Seasonal Adjustment of Economic Time Series, Her fra Svend Hylleberg (Red.): Modelling Seasonality, Oxford (1992).
- [2] ——— og Brian C. Monsell (1992): X-11 Symmetric Linear Filters and their Transfer Functions, Bureau of the Census, Research Report nr. RR 92/15.
- [3] Bureau of the Census (1999): X-12-ARIMA Reference Manual.
- [4] Burridge, Peter og Kenneth F. Wallis (1984): Unobserved-Components Models for Seasonal Adjustmnet Filters, Her fra Svend Hylleberg (Red.): Modelling Seasonality, Oxford (1992).
- [5] Depoutot, Raoul og Christophe Planas (1998): Comparing seasonal adjustment and trend extraction filters with application to model based selection of X-11 linear filters, Eurostat Working Paper.
- [6] Deutsche Bundesbank (1999): Der Übergang vom Saisonbereinigungsverfahren Census X-11 zu Census X-12-ARIMA, Deutsche bundesbank, Monatsbericht 51:9 (september 1999).
- [7] Dosse, J. og C. Planas (1996a): Preadjustment in Seasonal Adjustment Methods: A Comparison of RegArima and Tramo, Eurostat Working Paper.
- [8] ——— og ——— (1996b): Revisions in Seasonal Adjustment Methods: A Comparison of RegArima and Tramo, Eurostat Working Paper.
- [9] Findley, David og Catherine Hood (1999): X-12-ARIMA and its Application to Some Italian Indicator Series, Research Paper, Bureau of the Census.
- [10] ———, D., B. Monsell, W. Bell, M. Otto og B. Chen (1996): New Capabilities and Methods og the X-12-ARIMA Seasonal-Adjustment Program. Journal of Business and Economic Statistics No 16.
- [11] Fischer, Björn (1995): Decomposition of Time Series Comparing Different Methods in Theory and Practice, Eurostat Working Paper.
- [12] Ghysels, Eric, Clive W. J. Granger og Pierre L. Siklos (1995): Is Seasonal Adjustment a Linear or Nonlinear Data Filtering Process ?, Centre interuniversitarie de recherche en analyse des organisations, Scientific Series No. 95s-19.
- [13] Gómez, Victor og Agustín Maravall (1997): Programs TRAMO and SEATS, Instructions for the User. Banco de España.
- [14] ——— og ——— (1998a): Guide for Using the Programs TRAMO and SEATS. Banco de España.
- [15] ——— og ——— (1998b): Seasonal adjustment and signal extraction in economic time series, Working Paper no. 9809, Banco de España.

- [16] Hillmer, S. C. og G. C. Tiao (1982): An ARIMA-Model-Based Approach to Seasonal Adjustment, Her fra Svend Hylleberg (Red.): Modelling Seasonality, Oxford (1992).
- [17] Hood, Catherine og David F. Findley (1999): An Evaluation of TRAMO/SEATS and Comparison with X-12-ARIMA, Research Paper, Bureau of the Census (Draft version).
- [18] Hylleberg, Svend (1986): The X-11 Method, Her fra Svend Hylleberg (Red.): Modelling Seasonality, Oxford (1992).
- [19] ——— (1992): General Introduction, Fra Svend Hylleberg (Red.): Modelling Seasonality, Oxford (1992).
- [20] Maravall, Agustín (1996): Short-term analysis of macroeconomic time series, Working Paper no. 9607, Banco de España.
- [21] ——— (1997): Two discussions on new seasonal adjustments methods, Working Paper no. 9704, Banco de España.
- [22] ——— (1999): An application of Tramo and Seats, Working Paper no. 9914, Banco de España.
- [23] ——— og David A. Pierce (1987): A Prototypical Seasonal Adjustment Model, Her fra Svend Hylleberg (Red.): Modelling Seasonality, Oxford (1992).
- [24] ——— og Christophe Planas (1998): Signal Extraction in Non Invertible ARIMA Models, Eurostat Working Paper.
- [25] Planas C. (1996): Short-Term Variability in Seasonal Adjusted Time Series, Eurostat Working Paper.
- [26] ——— (1997a): Applied Time Series Analysis: Modelling, Forecasting, Unobserved Components Analysis and the Wiener-Kolmogorov Filter, Eurostat Working Paper.
- [27] ——— (1997b): The Analysis of Seasonality in Economic Statistics: A Survey of Recent Developments, Eurostat Working Paper.
- [28] Shiskin, J., A. Young og J. Musgrave (1967): The X-11 Variant of the Census Method II Seasonal Adjustment Program, Bureau of the Census, Technical Report Nr. 15.
- [29] Wallis, Kenneth F. (1974): Seasonal Adjustment and Relations between Variables, Her fra Svend Hylleberg (Red.): Modelling Seasonality, Oxford (1992).