

*Dinh Quang Pham*

**Å simulere revisjoner for  
sesongjusterte tall og trend  
fra X-12-ARIMA**

Notater

# Innhold

<b>1</b>	<b>Innledning</b>	<b>1</b>
<b>2</b>	<b>Ulemper med Hendersons vekter</b>	<b>1</b>
<b>3</b>	<b>Revisjoner</b>	<b>3</b>
<b>4</b>	<b>Simulering for varians til revisjonen av sesongjusterte tall</b>	<b>4</b>
<b>5</b>	<b>Eksempler</b>	<b>6</b>
5.1	Antall førstegangs registrerte personbiler . . . . .	6
5.2	Detaljvolumindeksen, SNN52.IVL.U . . . . .	7
5.3	Passasjerer på utenlandsruter totalt . . . . .	8
<b>6</b>	<b>Simuleringsstudie</b>	<b>9</b>
<b>7</b>	<b>Oppsummering</b>	<b>9</b>

## Figurer

1	<i>Diagrammet for simuleringen . . . . .</i>	5
2	<i>Histogrammene for sesongkomponenten, sesongjusterte tall og trend for juni 2001, ved 1000 simuleringer . . . . .</i>	7

# Å simulere revisjoner for sesongjusterte tall og trend fra X-12-ARIMA<sup>1</sup>

## 1 Innledning

X-12-ARIMA er en ikke parametrisk metode basert på glidende gjennomsnitt, med Hendersons vektorer for å spalte opp en tidsserie i tre komponenter: sesongkomponenten  $S_t$ , trenden  $T_t$  og den irregulære komponenten  $I_t$ . Med unntak av begynnelsen og slutten av tidsserien er filteret symmetrisk, i endene av serien brukes det et asymmetrisk filter (X-11). Det gir ofte store endringer mot slutten av serien for sesongjusterte tall og trenden når en ny observasjon kommer. For å redusere denne effekten forlenger man tidsserien et år med framskrivninger slik at de siste observasjonene også beregnes med et symmetrisk filter (X-11- og X-12-ARIMA). Usikkerheten i framskrivningene medfører endringer av de tre komponentene når nye observasjoner tilføyes, men endringene er som i regel mindre enn ved bruk av det asymmetriske filteret.

Bell (1999) har utviklet en metode for å beregne revisjoner for sesongjusterte tall fra X-12-ARIMA. Den er basert på varianser av tilbakeskrivingene og framskrivingene. Det krever mye matriseberegninger. Metoden som vi presenterer i neste avsnitt er enklere.

Før vi presenterer metoden vil vi beskrive ulemper ved filtre med Hendersons vektorer i X-12-ARIMA som medfører revisjoner for sesongjusterte tall og trend.

## 2 Ulemper med Hendersons vektorer

La  $\{O_t\}$ ,  $t = 1, \dots, N$  være observasjonene. Vi vil beregne  $Y_t$  ved

$$Y_t = \omega_{-2}O_{t-2} + \omega_{-1}O_{t-1} + \omega_0O_t + \omega_1O_{t+1} + \omega_2O_{t+2} \quad (1)$$

der vektene  $\omega_j$  er gitt i tabell 1. Vi får

$$\begin{aligned} - \text{ for } t = N, & \quad Y_t = -0.073O_{N-2} + 0.403O_{N-1} + 0.670O_N \\ - \text{ for } t = N - 1, & \quad Y_t = -0.073O_{N-3} + 0.294O_{N-2} + 0.522O_{N-1} + 0.257O_N \end{aligned}$$

$$\begin{aligned} - \text{ for } 3 \leq t \leq N - 2, & \quad Y_t = -0.073O_{t-2} + 0.294O_{t-1} + 0.558O_t + 0.294O_{t+1} - 0.073O_{t+2} \end{aligned} \quad (2)$$

og på samme måte, er

$$\begin{aligned} - \text{ for } t = 1, & \quad Y_t = 0.670O_1 + 0.403O_2 - 0.073O_3 \\ - \text{ for } t = 2, & \quad Y_t = 0.257O_1 + 0.522O_2 + 0.294O_3 - 0.073O_4 \end{aligned}$$

---

<sup>1</sup> ./x12a\_problem/y020214b.tex

Tabell 1: 5-ledds Hendersons vektene ( $\omega_j$ )

	$N - 4$	$N - 3$	$N - 2$	$N - 1$	$N$
$N$	0	0	-0.073	0.403	0.670
$N - 1$	0	-0.073	0.294	0.522	0.257
$N - 2$	-0.073	0.294	0.558	0.294	-0.073

Vi ser at asymmetriske vekter også blir brukt for  $t = N - 1$  og  $t = N$ , mens i midten er vektene symmetriske. Vi bruker symbolene:

$Y_{t|N}$  for å betegne  $Y_t$  gitt observasjonene fram til og med tidspunkt  $N$  i beregningen, og

$\hat{O}_N(k)$  er  $k$ -trinns framskrivingen for  $O_{N+k}$  gitt  $O_1, \dots, O_N$ .

Vi får fra tabell 1

$$Y_{N|N} = -0.073O_{N-2} + 0.403O_{N-1} + 0.670O_N$$

Når  $O_{N+1}$  er observert, blir

$$Y_{N|N+1} = -0.073O_{N-2} + 0.294O_{N-1} + 0.522O_N + 0.257O_{N+1}$$

Endringen av  $Y_N$  som vi kaller en revisjon, når  $O_{N+1}$  tas med, er

$$r_N = Y_{N|N+1} - Y_{N|N} = -0.109O_{N-1} - 0.148O_N + 0.257O_{N+1} \quad (3)$$

$r_N$  er stor når  $O_{N+1}$  ligger langt fra  $O_N$  og  $O_{N-1}$ .

Framskrivingen med minste kvadratfeil for  $O_{N+1}$ , er gitt ved

$$\hat{O}_N(1) = E(O_{N+1}|O_N, O_{N-1}, \dots, O_1) \quad (4)$$

Vi forlenger tidsserien med  $\hat{O}_N(1), \hat{O}_N(2), \dots, \hat{O}_N(12)$ , ved hjelp av en ARIMA modell. Vi får fra tabell 1

$$\hat{Y}_{N|N} = -0.073O_{N-2} + 0.294O_{N-1} + 0.558O_N + 0.294\hat{O}_N(1) - 0.073\hat{O}_N(2)$$

og

$$\hat{Y}_{N|N+1} = -0.073O_{N-2} + 0.294O_{N-1} + 0.558O_N + 0.294O_{N+1} - 0.073\hat{O}_{N+1}(1)$$

Revisjonen blir

$$\hat{r}_N = \hat{Y}_{N|N+1} - \hat{Y}_{N|N} = 0.294(O_{N+1} - \hat{O}_N(1)) - 0.073(\hat{O}_{N+1}(1) - \hat{O}_N(2)) \quad (5)$$

Merk at  $\hat{O}_N(2)$  er framskrivingen for  $O_{N+2}$  med dataene  $O_1, \dots, O_N$ , mens  $\hat{O}_{N+1}(1)$  er framskrivingen for  $O_{N+2}$  med dataene fram til og med tidspunkt  $N + 1$ . Når vi har en god tilpasning mellom data og ARIMA modellen, blir  $|\hat{r}_N|$  mindre enn  $|r_N|$ , siden  $\hat{O}_N(1) \approx O_{N+1}$  og  $\hat{O}_{N+1}(1) \approx \hat{O}_N(2)$ .

Variansen for  $\widehat{r}_N$  er

$$\begin{aligned} Var(\widehat{r}_N) &= 0.294^2 Var(\widehat{O}_N(1)) + 0.073^2 Var(\widehat{O}_N(2)) - \\ &\quad 2(0.294)(0.073)Cov(\widehat{O}_N(1), \widehat{O}_N(2)) + 0.073^2 Var(\widehat{O}_{N+1}(1)) \end{aligned}$$

I X-12-ARIMA, for å unngå store revisjoner for sesongjusterte tall og trend ved å bruke asymmetriske vektorer i slutten av tidsserien, forlenger man dataene med framskrivninger for et år slik at sesongkomponenten, sesongjusterte tall og trend for den observerte tidsserien beregnes med symmetriske vektorer, mens asymmetriske vektorer brukes på den framskrevne tidsserien.

### 3 Revisjoner

Vi antar at  $O_t$  kan spaltes opp i to komponenter, med og uten sesongvariasjoner,  $S_t$  og  $A_t$ , henholdsvis.

$$O_t = S_t + A_t \quad (6)$$

La  $\nu(B)$  være filteret for å beregne sesongkomponenten. Vi får

$$\widehat{S}_t = \nu(B)O_t \quad (7)$$

$\nu(B)$  har symmetriske vektorer ( $\nu_{-j} = \nu_j$ ) og skrives slik

$$\nu(B) = \dots + \nu_{-2}B^2 + \nu_{-1}B + \nu_0 + \nu_1F + \nu_2F^2 + \dots$$

hvor  $FO_t = O_{t+1}$ . For SEATS er  $\nu(B)$  Wiener-Kolmogorov filter og for X-12-ARIMA er det  $3 \times 3$ ,  $3 \times 5$  og  $3 \times 9$  glidende gjennomsnitt filter.

Vi antar at  $O_t$  beskrives ved en generell ARIMA modell

$$O_t = \frac{\theta(B)\Theta(B^s)}{\phi(B)\Phi(B^s)\nabla^d\nabla_s^D}\epsilon_t = \epsilon_t + \psi_1\epsilon_{t-1} + \dots = \Psi(B)\epsilon_t \quad (8)$$

hvor  $s$  er sesongperioden,  $\{\epsilon_t\}$  er hvit støy prosess med forventning 0 og varians  $\sigma_\epsilon^2$ .  $\nabla = (1 - B)$  og  $\nabla_s = (1 - B^s)$ .  $\phi(B)$ ,  $\Phi(B^s)$ ,  $\theta(B)$ ,  $\Theta(B^s)$  er polynomer av orden  $p$ ,  $q$ ,  $P_s$  og  $Q_s$ .  $d$  og  $D$  er antall ganger differansen skal tas for  $O_t$  slik at  $\nabla^d\nabla_s^D O_t$  er en stasjonær prosess.

Vi får, fra ligningene 7 og 8

$$\widehat{S}_t = \nu(B)\Psi(B)\epsilon_t = \Xi(B)\epsilon_t \quad (9)$$

hvor  $\Xi(B) = \dots + \xi_{-2}B^2 + \xi_{-1}B + \xi_0 + \xi_1F + \xi_2F^2 + \dots$

La  $\widehat{S}_{t|t+k}$  være sesongkomponenten i tidspunkt  $t$  gitt dataene fram til og med  $t+k$ .  $\widehat{S}_{t|t+k}$  er gitt ved

$$\begin{aligned}
\widehat{S}_{t|t+k} &= E[\xi(B)\epsilon_t | O_{t+k}, O_{t+k-1}, \dots, O_1] \\
&= E[\dots + \xi_{-2}B^2 + \xi_{-1}B + \xi_0 + \xi_1F + \xi_2F^2 + \dots] \epsilon_t \\
&= [\dots + \xi_{-2}B^2 + \xi_{-1}B + \xi_0 + \xi_1F + \xi_2F^2 + \dots + \xi_kF^k] \epsilon_t \\
&= \Xi^k(B)\epsilon_t
\end{aligned} \tag{10}$$

hvor  $\Xi^k(B)$  er  $\Xi(B)$  som er trunkert etter lag  $k$ , siden vi ikke kan framskrive feilledd.

Revisjonen  $r_k$  er gitt ved

$$r_k = \widehat{S}_t - \widehat{S}_{t|t+k} = \sum_{i=k+1}^{\infty} \xi_i \epsilon_{t+i} \tag{11}$$

Vi ser at forskjellige framskrivinger og vektorer til filteret ikke gir samme verdi til revisjonen. Pierce (1960) har vist at  $r_k$  konvergerer. Gitt at dataene er observert fram til og med  $t+k$ . Revisjonen for  $\widehat{S}_t$  når  $O_{t+k+1}$  observeres, er

$$r_k = \widehat{S}_{t|t+k+1} - \widehat{S}_{t|t+k} = \xi_{k+1} \epsilon_{k+1} \tag{12}$$

Vi er ofte interessert i revisjonen for sesongjusterte tall i det siste tidspunktet  $t = N$ ,

$$r_A = A_{N|N+1} - A_{N|N}$$

siden  $O_{N+1}$  ikke er observert, får vi ikke en verdi for  $r_A$ . La  $\widehat{O}_N(1)$  være framskrivningen for  $O_{N+1}$  med variansen  $\widehat{\sigma}_{O_N(1)}^2$ , og  $\widehat{A}_{N|N+1}$  være sesongjustert tall for  $t = N$  med dataene  $O_1, O_2, \dots, O_N, \widehat{O}_N(1)$ . Revisjonen blir

$$\widehat{r}_A = \widehat{A}_{N|N+1} - A_{N|N} \tag{13}$$

Beregningen for sesongjustering i X-12-ARIMA er veldig komplisert med mange prekorrigeringer for effektene av ukedag, påske og intervensjoner. Derfor blir det vanskelig å skrive sesongjusterte tall ved et entydig uttrykk for å kunne estimere variansen til  $\widehat{r}_A$ . Vi vil bruke simuleringer for å gjøre dette, se neste avsnitt.

## 4 Simulering for varians til revisjonen av sesongjusterte tall

Prinsippet er å simulere nye tidsserier for å beregne de simulerte dataene for sesongjusterte tall, trend og sesongkomponenten. Vi utfører følgende trinn:

- 1 Anta at støyleddet  $\epsilon_t$  i den ARIMA modellen som tilpasses, har en normalfordeling,  $N(0, \widehat{\sigma}_\epsilon^2)$ . Vi simulerer framskrivningen for rådata i tidspunkt  $t = N$ ,  $\widehat{O}_N^{(i)}(1)$ , i  $i$ -te replikasjon ved

$$\widehat{O}_N^{(i)}(1) \sim N(\widehat{O}_N(1), \widehat{\sigma}_O^2) \tag{14}$$

2 Vi lager en ny tidsserie med dataene

$$O_1, O_2, \dots, O_N, \widehat{O}_N^{(i)}(1) \quad (15)$$

3 Ved å sesongjustere (15), får vi  $\widehat{A}_{N|N+1}^{(i)}$  som er sesongjusterte tall for  $t = N$ , gitt dataene og framskrivningen  $\widehat{O}_N^{(i)}(1)$ . Revisjonen blir

$$\widehat{r}_{A,i} = \widehat{A}_{N|N+1}^{(i)} - A_{N|N}$$

hvor  $A_{N|N}$  er sesongjusterte tall av tidsserien  $O_1, O_2, \dots, O_N$ .

Vi gjentar trinn 1, 2 og 3, et stort antall ganger  $M$ . Vi får

$$\widehat{A}_{N|N+1}^{(1)}, \widehat{A}_{N|N+1}^{(2)}, \dots, \widehat{A}_{N|N+1}^{(M)} \quad \text{og} \quad \widehat{r}_{A,1}, \widehat{r}_{A,2}, \dots, \widehat{r}_{A,M} \quad (16)$$

Det er enkelt å beregne gjennomsnitt og standardavvik for disse størrelsene. For eksempel, for revisjonen er gjennomsnittet  $\bar{r}_A$  og standardavviket  $\widehat{\sigma}_r$  gitt ved

$$\bar{r}_A = \frac{1}{M} \sum_{i=1}^M \widehat{r}_{A,i}, \quad \text{og} \quad \widehat{\sigma}_r = \sqrt{\frac{1}{M-1} \sum_{i=1}^M (\widehat{r}_{A,i} - \bar{r}_A)^2} \quad (17)$$

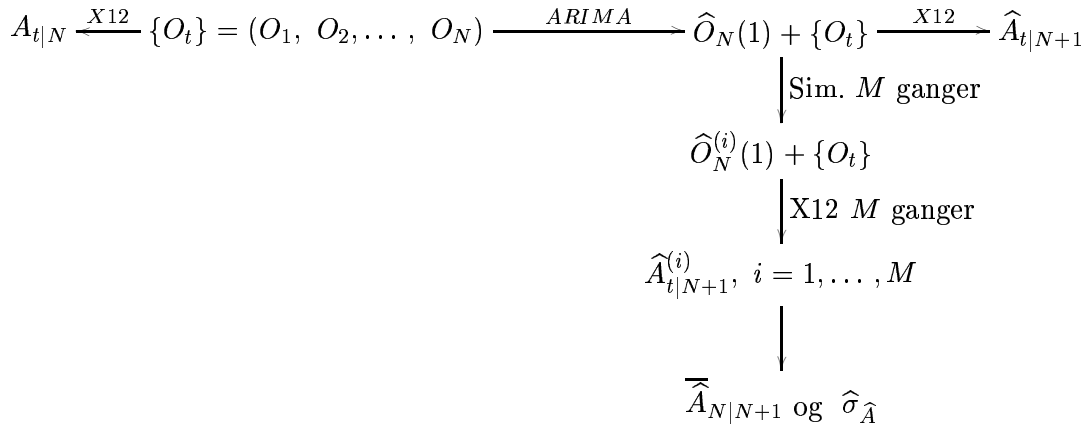
Vi lager et 95% konfidensintervall for  $r_A$  ved å sortere (16) i stigende rekkefølge. Nedre og øvre grenseverdien,  $l95$  og  $u95$ , er gitt ved

$$l95 = 0.5(\widehat{r}_A^{(25)} + \widehat{r}_A^{(26)}) \quad \text{og} \quad u95 = 0.5(\widehat{r}_A^{(975)} + \widehat{r}_A^{(976)})$$

hvor  $\widehat{r}_A^{(1)} \leq \widehat{r}_A^{(2)} \leq \dots \leq \widehat{r}_A^{(M)}$ . Siden  $r_A = A_{N+1|N} - A_{N|N}$ . Et 95% konfidensintervall for  $A_{N|N+1}$ , er

$$[A_{N|N} + l95, A_{N|N} + u95] \quad (18)$$

Vi beregner på samme måte for sesongkomponenten og trend. Metoden beskrives ved diagrammet i figur 1.



Figur 1: Diagrammet for simuleringen



## 5 Eksempler

Vi anvender metoden for tre eksempler: registrerte første gang for personbiler i Norge, detaljvolumindeksen og antall flypassasjerer på utenlands ruter.

### 5.1 Antall førstegangs registrerte personbiler

Dataene er egentlig observert fra januar 1980 til juni 2001 (forkortes med jun.01), men vi vil bruke tallene fram til og med mai 2001 for å sammenligne sesongkomponenten, sesongjusterte tall og trend i mai 2001, når (i)  $O_{jun.01}$  blir erstattet med  $\hat{O}_{jun.01}$  og (ii)  $O_{jun.01}$  observeres.

ARIMA(0 1 1)(0 1 1) modellen gir en god tilpasning for dataene. Framskrivningen  $\hat{O}_{jun.01} = 10531$  med standardavvik 1125 biler. Tallet ligger nær den observerte verdien  $O_{jun.01} = 10321$  biler. For å beregne sesongjusterte i mai 2001, gitt

- dataene fram til og med mai 2001,  $A_{mai|mai.01}$ .
- dataene forlenget med framskrivningen,  $\hat{A}_{mai|jun.01}$ .
- $O_{jun.01}$  observeres,  $A_{mai|jun.01}$ .

sesongjusterer vi de følgende tidsseriene:

$$\begin{array}{llll} O_{jan.80}, O_{feb.80}, \dots, O_{mai.01} & & \text{det gir} & A_{mai|mai.01} = 9555 \\ O_{jan.80}, O_{feb.80}, \dots, O_{mai.01}, \hat{O}_{jun.01} & & & \hat{A}_{mai|jun.01} = 9559 \\ O_{jan.80}, O_{feb.80}, \dots, O_{mai.01}, O_{jun.01} & & & A_{mai|jun.01} = 9532 \end{array}$$

Vi får en god estimering for  $A_{mai|jun.01}$ , siden  $\hat{A}_{mai|jun.01}$  er nær  $A_{mai|jun.01}$ . Vi simulerer  $\hat{O}_{jun.01}^{(i)}$ , i  $i$ -te replikasjoner, med fordelingen

$$\hat{O}_{jun.01}^{(i)} \sim N(10531, 1124)$$

Vi utfører trinn 2 og 3 i avsnitt 4 med 1000 replikasjoner og får

$$\begin{array}{ll} \hat{S}_{mai|jun.01}^{(i)}, i = 1, \dots, 1000 & \text{for sesongkomponenten} \\ \hat{A}_{mai|jun.01}^{(i)}, i = 1, \dots, 1000 & \text{for sesongjusterte tall} \\ \hat{T}_{mai|jun.01}^{(i)}, i = 1, \dots, 1000 & \text{for trend} \end{array}$$

deres gjennomsnitt og standardavvik er gitt i tabell 2.

Symbolet  $t|\hat{O}_{jun.01}$  i tabell 2 står for verdien i tidspunkt  $t$  gitt  $O_1, \dots, O_{mai.01}$  og  $\hat{O}_{jun.01}$ . Vi får gode estimerte verdier for de tre komponentene sammenlignet med de faktiske verdiene. Dessuten er deres 95% konfidensintervallene mye kortere sammenlignet med for rådata ( $[l95, u95] = [8328, 12734]$ ).

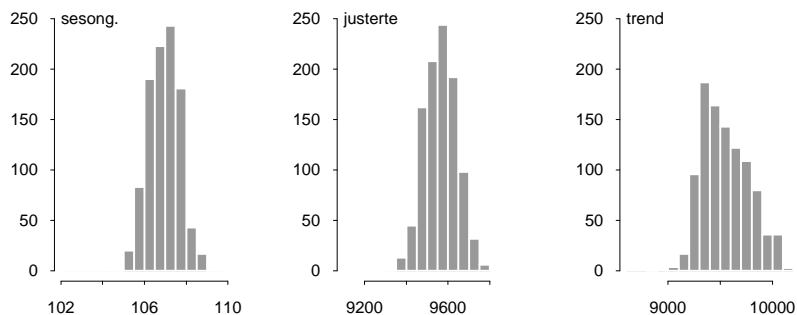
Ved å bruke alle dataene fra januar 1980 fram til juni 2001 i beregningen, får vi

$t$ =juni 2001	$t O_{jun.01}$	$t \hat{O}_{jun.01}$	$\hat{\sigma}$	$l95$	$u95$	
Se. justerte	9613	9592	77	9450	9747	(biler)
Trend	9488	9493	235	9171	10047	(biler)
Se. komp.	106.99	107.22	0.73	105.74	108.57	(i prosent)

Tabell 2: 95% konfidensintervaller for sesongkomponenten, sesongjusterte tall og trend for  $t=mai$  2001

	$t O_{mai01}$	faktiske $t O_{jun01}$	estimerte $t \hat{O}_{jun01}$	$\hat{\sigma}$	$l95$	$u95$	lengden
se. jus. ( $A$ )	9555	9532	9539	133	9339	9818	479
trend ( $T$ )	9615	9553	9615	256	9112	9952	840
se. komp ( $S$ )%	114.68	114.93	114.63	1.42	111.88	117.00	5.12

som er estimerte verdiene for sesongjusterte tall, trend og sesongkomponenten i juni 2001, når tallet i juli er estimert. Vi øker antall replikasjoner fra 1000 til 2000 og finner de samme resultatene. Figur 2 viser histogrammene for sesongkomponenten, sesongjusterte tall og trend for juni 2001, ved 1000 simuleringer. Vi ser at 2.5% og 97.5% kvantiler for sesongkomponenten og sesongjusterte tall ligger nær gjennomsnittet. Dette medfører smale 95% konfidensintervaller. For trenden er fordelingen skjev med 97.5% kvantilen lengre fra gjennomsnittet enn 2.5% kvantilen. Vi får et større 95% konfidensintervall for trenden sammenlignet med sesongjusterte tall.



Figur 2: Histogrammene for sesongkomponenten, sesongjusterte tall og trend for juni 2001, ved 1000 simuleringer

## 5.2 Detaljvolumindeksen, SNN52.IVL.U

Tallene er observert fra januar 1979 til mai 2001. I første omgang bruker vi bare dataene fram til og med april 2001 i beregningen. Grunnen er at vi vil sammenligne de estimerte og faktiske verdiene av sesongkomponenten, sesongjusterte tall og trend, når  $\hat{O}_{mai.01}$  og  $O_{mai.01}$  legges til i tidsserien.

Vi får  $\hat{O}_{mai.01} = 118.06$ , med standardavvik  $\hat{\sigma} = 2.62$ . 95% konfidensintervallet er  $[113.03, 123.31]$ . Vi observerer  $O_{mai.01} = 115.66$ . Det relative avviket mellom den estimerte og faktiske verdien,  $\hat{O}_{mai.01}$  og  $O_{mai.01}$ , er 2.1%. Resultater etter 1000 replikasjoner er vist i tabell 3. Vi ser at den faktiske verdien for sesongjusterte tall i april 2001, når  $O_{mai.01}$  legges til,  $A_{apr|mai.01}$ , er nær den simulerte verdien  $\hat{A}_{apr|mai.01}$

og ligger innenfor 95% konfidensintervallet. Vi finner de samme konklusjonene for sesongkomponenten og trend.

Tabell 3: 95% konfidensintervaller for sesongkomponenten, sesongjusterte tall og trend, for  $t=$ april 2001

	$t O_{apr01}$	faktiske $t O_{mai01}$	estimerte $t \widehat{O}_{mai01}$	$\widehat{\sigma}$	$l95$	$u95$	lengden
se. jus. ( $A$ )	116.90	116.77	117.01	0.25	116.41	117.40	0.99
trend ( $T$ )	117.21	116.64	117.17	0.56	115.69	117.87	2.18
se. komp ( $S$ )%	90.80	90.89	91.72	0.17	90.45	91.15	0.70

Med dataene fram til og med mai 2001, er

$t=$ mai 2001	$t O_{mai01}$	$t \widehat{O}_{jun01}$	$\widehat{\sigma}$	$l95$	$u95$
Se. justerte	116.72	116.67	0.53	115.72	117.66
Trend	116.72	116.71	0.63	115.52	117.86
Se. komp	97.80	97.85	0.43	97.05	98.61

Vi ser at sesongjusterte tall for  $t = mai.01$ , gitt dataene fram til og med mai 2001, er 116.72. Når  $O_{jun.01}$  legges til, blir tallet revidert til 116.67.

### 5.3 Passasjerer på utenlandsruter totalt

Dataene er antall passasjerer som er summert over alle lufthavner på utenlandsruter i Norge, fra januar 1982 til desember 2000.

ARIMA(0 1 1)(0 1 1) modellen gir en god tilpasning til dataene. Vi får  $\widehat{O}_{des.00} = 466396$ , med 95% konfidensintervallet [435488, 499499].  $O_{des.00} = 455519$ . Det relative avviket mellom de to størrelsene er 2.38%.

Ved å bruke dataene fram til og med november 2000, viser tabell 4 følgende resultater

Tabell 4: 95% konfidensintervaller for sesongkomponenten, sesongjusterte tall og trend, for  $t=$ november 2000

	$t O_{nov00}$	$t O_{des00}$	$t \widehat{O}_{des00}$	$\widehat{\sigma}$	$l95$	$u95$	lengden
se. jus. ( $A$ )	577431	576843	577870	1200	574074	577929	3855
trend ( $T$ )	586408	582867	586643	5744	580201	599926	19725
se. komp ( $S$ )%	91.33	91.40	91.26	0.19	91.25	91.86	0.61

Vi får gode estimerte verdier for sesongkomponenten, sesongjusterte tall og trend for november 2000, sammenligning med de faktiske verdiene, når  $O_{des.00}$  tilføyes.

Ved å bruke hele tidsserien fra januar 1982 til desember 2000, er

$t=\text{desember } 2000$	$t O_{des00}$	$t \hat{O}_{jan01}$	$\hat{\sigma}$	$l95$	$u95$
Se. justerte	582022	582300	1421	579952	584764
Trend	585619	585682	4099	579076	591181
Se. komp	78.26	78.23	0.19	77.90	78.55

Tolkningen er at sesongjusterte tall i desember 2001 gitt dataene til og med dette tidspunktet er  $A_{des|des.01} = 582022$ . Det blir estimert med 582300 passasjerer og 95% konfidensintervallet  $[579952, 584764]$ , når tidsserien forlenges med januar 2001.

## 6 Simuleringsstudie

For å vurdere hvor god metoden er, lager vi en sammenligning mellom de estimerte og faktiske verdiene for sesongkomponenten, sesongjusterte tall og trend, når en ny observasjon legges til. Vi velger tidsserien første gangsregistrerte for personbiler for simuleringsstudien siden dataene har store sesongvariasjoner og det er mange ekstremverdier som medfører store usikkerheter for revisjoner. Dermed er det mange verdier som kan ligge utenfor de simulerte 95% konfidensintervallene. Tallene for 27 måneder fra januar 1999 til og med mars 2001, er listet ut i tabell 5. Vi ser at de faktiske verdiene  $S_{N|N+1}$ ,  $A_{N|N+1}$  og  $T_{N|N+1}$  ligger innenfor 95% konfidensintervaller, bortsett fra trenden i juni 1999. Vi får  $T_{jun|jul.99} = 9959$  og  $\hat{T}_{jun|jul.99} = 9516$  med  $[l95, u95] = [8999, 9946]$ . Metoden fungerer bra. Vi får smale konfidensintervaller for sesongkomponenten og sesongjusterte tall. Det er de to viktige komponentene i sesongjustering.

## 7 Oppsummering

Vi har i SSB følgende programmer for sesongjustering:

STAMP. Det krever at brukeren må lage modeller for sesongkomponenten og trend og har kunnskap om Kalmanfilter. Programmet lister ikke ut varianser for estimatene og korrigerer ikke for effektene av ukedagseffekter, påske og intervensjoner. For å kunne gjøre dette må brukeren lage selv dummy variable og kjøre STAMP med disse dummy variablene. Det er en tung jobb.

SEATS/TRAMO, brukes i EU land, som er modellbasert. Programmet gir usikkerheter for sesongjusterte tall og trend, men konfidensintervaller er så store.

X-12-ARIMA, brukes i SSB, som er basert på det glidende gjennomsnittet. Vi får ingen usikkerheter for sesongjusterte tall og trend, men programmet kan korrigere for effektene av ukedager, påske og intervensjoner.

En fordel med metoden i dette notatet er at man ikke trenger å lage modeller for sesongvariasjoner og trend. Metoden passer bra med X-12-ARIMA, siden man kan hente framskrivingen  $\hat{O}_N(1)$  og standardavviket  $\hat{\sigma}(\hat{O}_N(1))$  fra kjøringen for rådata for å lage nye tidsserier og bruker deretter X-12-ARIMA for sesongjustering. Etter et stort antall replikasjoner beregner vi de simulerte 95% konfidensintervaller for sesongjusterte tall og trend. Dessuten er effektene for ukedager, påske og intervensjoner korrigert uten å lage dummy variable. Metoden kan også brukes for SEATS.

De tre eksemplene viser at metoden gir gode resultater. Vi utfører 1000 replikasjoner i hvert tilfelle. Det tar i gjennomsnitt 15 minutter.

Tabell 5: Faktiske og estimerte verdiene for sesongkomponenten, sesongjusterte tall og trend

N	Sesongkomponent					Sesongjusterte tall					Trend				
	$S_{N N+1}$	$\hat{S}_{N N+1}$	$\hat{S}_{N N+1}$	$I95$	$u95$	$A_{N N+1}$	$\hat{A}_{N N+1}$	$\hat{A}_{N N+1}$	$I95$	$u95$	$T_{N N+1}$	$\hat{T}_{N N+1}$	$\hat{T}_{N N+1}$	$I95$	$u95$
1999	januar	98.56	98.30	96.21	100.50	9909	9936	9706	10163	9987	10071	9512	10685		
	februar	91.12	90.62	87.07	93.13	9873	9926	9660	10331	9805	9919	9381	10492		
	mars	103.89	103.89	102.40	105.85	9783	9784	9550	9973	9750	9751	9234	10281		
	april	105.16	104.28	102.79	106.55	9677	9773	9521	9947	9476	9711	9152	10103		
	mai	112.90	112.55	111.54	115.61	9287	9241	8971	9338	9527	9363	8835	9660		
	juni	109.42	110.66	109.33	111.39	9849	9721	9648	9859	9959	9516	8999	9946 (*)		
	juli	121.86	123.18	121.21	123.19	10712	10568	10567	10752	10338	10163	9396	10658		
	august	94.44	94.25	92.78	96.39	10466	10490	10222	10689	10557	10633	9843	11141		
	september	97.33	97.00	95.56	98.53	10735	10777	10574	10968	10616	10754	10095	11216		
	oktober	102.72	103.34	101.75	104.90	10794	10720	10533	10912	10920	10672	10277	11266		
	november	95.91	96.21	94.93	97.49	11287	11245	11066	11427	11171	11011	10565	11663		
	desember	70.14	70.79	69.57	73.27	11536	11418	10993	11647	11594	11283	10663	11874		
2000	januar	97.30	97.97	96.25	98.58	12526	12468	12379	12722	11286	12060	11284	12686		
	februar	88.03	87.45	87.33	88.67	10813	10899	10739	10914	10909	11320	10833	12361		
	mars	104.04	103.55	103.55	105.33	10326	10386	10174	10387	10549	10761	10298	11679		
	april	101.83	102.78	101.61	104.59	10256	10148	9946	10282	10600	10331	9799	10830		
	mai	113.43	114.48	111.27	114.75	10990	10907	10879	11245	10005	10571	9684	11101		
	juni	107.52	108.08	107.02	109.55	9865	9807	9654	9918	10187	9936	9521	10636		
	juli	123.42	123.61	121.06	125.46	10352	10335	10163	10570	10321	10267	9852	10911		
	august	95.59	94.95	93.62	96.53	10367	10447	10251	10619	10235	10466	9982	10932		
	september	97.12	96.72	95.64	98.29	10054	10101	9915	10237	10125	10283	9721	10626		
	oktober	101.01	101.15	99.72	102.74	10113	10098	9921	10262	10151	10100	9720	10564		
	november	95.56	97.04	95.56	99.06	10296	10120	9887	10300	10632	10166	9735	10655		
	desember	72.24	72.75	71.18	72.77	11357	11293	11290	11562	10196	10844	10020	11351		
2001	januar	96.73	96.13	95.74	98.41	10063	10129	9882	10174	10037	10209	9707	11200		
	februar	87.74	87.62	86.56	89.15	9915	9929	9758	10050	9949	9997	9564	10629		
	mars	103.20	103.08	101.54	104.64	9824	9838	9666	10012	9844	9889	9446	10301		

\*  $T_{jun.jul.99}$  ligger utenfor 95% konfidensintervallet

## Referanser

- [1] Bell W. R. og Hillmer S. C. (1983): *Modelling Time Series with Calendar Variation*, Journal of the American Statistical Association, September 1983, Vol. 78
- [2] Bureau of the Census: *X-12 ARIMA Reference Manual, Version 0.2.5, October 1, 1999*
- [3] Dagum Estela Bee (1988): *The X11ARIMA/88 Seasonal Adjustment Method Foundations and User's Manual*, Statistics Canada, Methodology Branch
- [4] Findley D. F., Monsell B. C., Bell W. R., Otto M. C. and Chen B. C.(1996): *New Capabilities and Methods of the X-12 ARIMA Seasonal Adjustment Program*, Journal of Business & Economic Statistics, April 1998, Vol. 16, No.2
- [5] Findley D. F., Monsell B. C., Shulman H. B. and Pugh M. G.(1990): *Sliding-spans Diagnostics for Seasonal and Related Adjustments*, Journal of the American Statistical Association, June 1990, Vol. 85
- [6] Dossé J. and Planas C. (1996): *Revisions in Seasonal Adjustment Methods: an empirical comparison of X-12-ARIMA & SEATS*
- [7] Efron B. (1982): *The Jackknife, the Bootstrap and Other Resampling Plans*. Sam 1982.
- [8] John Higginson (1975): *An F Test for the presence of moving seasonality when using census method II-X-11 variant*, Statistics Canada
- [9] Higginson J. (1975): *An F test for the presence of moving seasonality when using census method II-X-11 variant*, Statistics Canada, Groupe des méthodes de désaisonnalisation et des séries chronologiques
- [10] Lothian J. and Morry M. *A Set of Quality Control Statistics for the X-11-ARIMA Seasonal Adjustment Method*, Statistics Canada, Groupe des méthodes de désaisonnalisation et des séries chronologiques
- [11] Lothian J. and Morry M. (1977): *The problem of aggregation direct and indirect*, Statistics Canada, Groupe des méthodes de désaisonnalisation et des séries chronologiques
- [12] Planas, C. (1997): *Applied Time Series Analysis: Modelling, Forecasting, Unobserved Components Analysis and the Wiener-Kolmogorov Filters*
- [13] Leiv Solheim og Dinh Quang Pham (1997): *Prekorrigering av påskeeffekten for detaljvolumindeksen 1979-1997*. Notater 73/97
- [14] J. Lothian and M. Morry: *A set of Quality Control Statistics for the X-11 ARIMA*, Statistics Canada, Groupe des méthodes de désaisonnalisation et des séries chronologiques
- [15] S-PLUS User'Manual, version 3.2, December 1993

## De sist utgitte publikasjonene i serien Notater

- 2002/39 H. Moafi: Omlegging av folkehøgskolestatistikk. Overgang til elektronisk rapportering. 31s.
- 2002/40 Ø. Kleven: Mediebrukundersøkelsen 2001. Dokumentasjonsrapport. 43s.
- 2002/41 Ø. Kleven: Samordnet levekårsundersøkelse 2000 - panelundersøkelsen. Dokumentasjonsrapport. 129s.
- 2002/42 L. Solheim: Foreløpige tall i FoB2001 Utvalg, vektor, estimering og usikkerhet. 64s.
- 2002/43 A. Andersen, E. Birkeland, J. Epland og M. I. Kirkeberg: Økonomi og levekår for ulike grupper trygdemottakere 2001. Foreløpig rapport. 214s.
- 2002/44 E.E. Eibak og R. Johannessen: Forventningsindikator - konsumprisene. Mai-november 2002. 16s.
- 2002/45 D. Quang Pham: Konkrete problemer med sesongjustering i SSB. 189s.
- 2002/46 A. Akselsen, G. Dahl og B.R. Joneid: FD - Trygd. Dokumentasjonsrapport. 1992-1997. 48s.
- 2002/47 J. Kristiansen: Visualisering av statistikk. Fra tabell til diagram. 40s.
- 2002/48 A. Finstad: Utslippsfaktorer for benzen. 18s.
- 2002/49 T. Bye, K.M. Heide og E. Holmøy: Transportutvikling i langsiktige fremskrivninger for norsk økonomi. Prosjektnotat til ECON. 30s.
- 2002/50 L. Solheim, M.I. Faldmo og D. Sve: Foreløpige tall i Jordbrukstelling 1999. Dokumentasjon av metoder og produksjon. 68s.
- 2002/51 KOSTRA: Arbeidsgrupperapporter 2002. 219s.
- 2002/52 Årsrapport 2001 Kontaktutvalget for helse- og sosialstatistikk. 35s.
- 2002/53 E. Dalheim: En skjembasert komplettering av registeret over befolkningens høyeste utdanning - Opplysninger om opplæring, skolegang og utdannig 1999. 59s.
- 2002/54 O. Klungsøyr: Lineære modeller - Av høyeste rang -. 75s.
- 2002/55 G. Dahl: Innvandrere og trygd. 28s.
- 2002/56 L. Vågane: Samordnet levekårsundersøkelse 2001 - tverrsnittsundersøkelsen. Dokumentasjonsrapport. 121s.
- 2002/57 A-K. H. Grorud: Bedrifts og foretaksregisteret. Regler og rutiner for ajourhold. Oppdatert versjon september 2002. 94s.
- 2002/58 S.I. Pedersen og L. Solheim: Arveavgiftsundersøkelsen. 50s.
- 2002/59 T. Løwe: Boligpreferanser og livsfase. 29s.
- 2002/60 T. Hoel: Cai2000 - Sentraladministrasjon . Systemdokumentasjon. 128s.
- 2002/61 H.C. Hougen, Ø. Kleven og S. Opdahl: Undersøkelse om livsløp og helse blant 55-årige menn 2001. Dokumentasjonsrapport. 68s.
- 2002/62 H.C. Hougen: Den nasjonale reisevaneundersøkelsen 2001. Dokumentasjonsrapport. 114s.
- 2002/63 T.M. Normann: Omnibusundersøkelsen mai/juni 2002. Dokumentasjonsrapport. 53s.
- 2002/64 V.V.H. Bloch: Arealbruksstatistikk for tettsteder. -Områdemodellering. 38s.
- 2002/65 F. Strøm: Arbeidsgiveravgiften - Soneinndelingens utvikling. Endringer i perioden 1998-2002. 31s.
- 2002/66 T. Løwe: Boligkonsum etter alder og kohort. Analyser av boforholdsundersøkelsene 1967-1997. 58s.