



*Dinh Quang Pham*

## **Konkrete problemer med sesongjustering i SSB**

Notater

# Innhold

|                                                                                                                           |           |
|---------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>1 Innledning</b>                                                                                                       | <b>2</b>  |
| <b>2 En metode for å framskrive apriltall, for totale produksjonsindeksen for industrien</b>                              | <b>4</b>  |
| 2.1 Innledning . . . . .                                                                                                  | 4         |
| 2.2 En beskrivelse av metoden for estimering av apriltall . . . . .                                                       | 5         |
| 2.2.1 Månedstidsserie og multiplikativ modell ( <i>log</i> transformasjon)                                                | 5         |
| 2.2.2 Månedstidsserie og additiv modell . . . . .                                                                         | 7         |
| 2.2.3 En alternativ måte å estimere apriltall for en månedstidsserie                                                      | 7         |
| 2.2.4 Kvartalstall . . . . .                                                                                              | 9         |
| 2.3 Oppsummering . . . . .                                                                                                | 11        |
| <b>3 Prekorrigering av påskeeffekten for detaljvolumindeksen 1979-1997</b>                                                | <b>12</b> |
| 3.1 Innledning . . . . .                                                                                                  | 12        |
| 3.2 Symboler . . . . .                                                                                                    | 12        |
| 3.3 Hvorfor må vi korrigere rådata for påskeeffekt . . . . .                                                              | 13        |
| 3.4 Metoden . . . . .                                                                                                     | 13        |
| 3.4.1 Både mars og april er observert i det siste året . . . . .                                                          | 14        |
| 3.4.2 Bare mars er observert i det siste året . . . . .                                                                   | 15        |
| 3.5 Eksempel. Detaljvolumindeksen . . . . .                                                                               | 16        |
| 3.5.1 Bare mars er observert i det siste året . . . . .                                                                   | 16        |
| 3.5.2 Både mars og april er observert i det siste året . . . . .                                                          | 21        |
| 3.5.3 En oppsummering for 5.1 og 5.2 . . . . .                                                                            | 21        |
| 3.5.4 Sammenligning med påskekorrigering og sesongjusterte tall ved X-11 ARIMA . . . . .                                  | 22        |
| 3.6 Oppsummering . . . . .                                                                                                | 23        |
| 3.7 Programmer . . . . .                                                                                                  | 26        |
| 3.8 Anvendelse metoden for de andre seriene i detaljvolumindeksen . . .                                                   | 29        |
| 3.8.1 SNN521.IVL.U: Butikkhandel med bredt vareutvalg. I alt . .                                                          | 29        |
| 3.8.2 SNN5211.IVL.U: Butikkhandel med bredt vareutvalg. Hovedvekt på nærings og nytelsesmidler . . . . .                  | 30        |
| 3.8.3 SNN5212.IVL.U: Butikkhandel med bredt vareutvalg. Bredt vareutvalg ellers . . . . .                                 | 31        |
| 3.8.4 SNN522.IVL.U: Butikkhandel med nærings og nytelsesmidler i spesialforretninger. I alt . . . . .                     | 32        |
| 3.8.5 SNN523.IVL.U: Butikkhandel med apotekvarer, sykepleieartikler, kosmetikk og toalettartikler. I alt . . . . .        | 33        |
| 3.8.6 SNN5231.IVL.U: Butikkhandel med apotekvarer, sykepleieartikler, kosmetikk og toalettartikler. Apotekvarer . . . . . | 34        |

|          |                                                                                                                                                        |           |
|----------|--------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| 3.8.7    | SNN5233.IVL.U: Butikkhandel med apotekvarer, sykepleie-<br>artikler, kosmetikk og toalettartikler. Kosmetikk og toalett-<br>artikler . . . . .         | 35        |
| 3.8.8    | SNN524.IVL.U: Butikkhandel med andre nye varer. I alt . . .                                                                                            | 36        |
| 3.8.9    | SNN5241.IVL.U: Butikkhandel med andre nye varer. Tekstiller<br>og utstyrsvarer . . . . .                                                               | 37        |
| 3.8.10   | SNN5242.IVL.U: Butikkhandel med andre nye varer. Klær . .                                                                                              | 38        |
| 3.8.11   | SNN5243.IVL.U: Butikkhandel med andre nye varer. Skotøy<br>reise effekter av lær og lærvarer. . . . .                                                  | 39        |
| 3.8.12   | SNN5244.IVL.U: Butikkhandel med andre nye varer. Belys-<br>ningsutstyr, kjøkkenutstyr, møbler og innredningsartikler. . .                              | 40        |
| 3.8.13   | SNN5245.IVL.U: Butikkhandel med andre nye varer. Elektris-<br>ke husholdningsapparater, radio, fjernsyn, kassetter og musikk-<br>instrumenter. . . . . | 41        |
| 3.8.14   | SNN5247.IVL.U: Butikkhandel med andre nye varer. Bøker,<br>papir, aviser og blader. . . . .                                                            | 42        |
| 3.8.15   | SNN5248.IVL.U: Butikkhandel med andre nye varer. Butikk-<br>handel ellers. . . . .                                                                     | 43        |
| 3.9      | Figurer . . . . .                                                                                                                                      | 44        |
| <b>4</b> | <b>Påskekorrigeringen for detaljvolumindeksen</b>                                                                                                      | <b>49</b> |
| 4.1      | Innledning . . . . .                                                                                                                                   | 49        |
| 4.2      | Beskrivelse av metoden . . . . .                                                                                                                       | 51        |
| 4.2.1    | Symboler . . . . .                                                                                                                                     | 52        |
| 4.2.2    | Estimeringsmetode . . . . .                                                                                                                            | 52        |
| 4.3      | Eksempel 1. Detaljvolumindeksen . . . . .                                                                                                              | 54        |
| 4.4      | Eksempel 2. Produksjonsindeks for industrien . . . . .                                                                                                 | 60        |
| 4.5      | Konklusjon . . . . .                                                                                                                                   | 62        |
| <b>5</b> | <b>Sesongjustering for ordretilganger i industrinæringen</b>                                                                                           | <b>63</b> |
| 5.1      | Innledning . . . . .                                                                                                                                   | 63        |
| 5.2      | Sesongjustering for tidsseriene $i_{TOT_t}$ , $i_{EKSP_t}$ og $i_{HJEM_t}$ . . . .                                                                     | 64        |
| 5.2.1    | Totalen, $i_{TOT_t}$ . . . . .                                                                                                                         | 64        |
| 5.2.2    | Samlet ordretilgang fra eksportmarkedet, $i_{EKSP_t}$ . . . . .                                                                                        | 68        |
| 5.2.3    | Samlet ordretilgang fra hjemmemarkedet, $i_{HJEM_t}$ . . . . .                                                                                         | 70        |
| 5.3      | Sesongjustering for tidsseriene $v_{TOT_t}$ , $v_{EKSP_t}$ og $v_{HJEM_t}$ . . .                                                                       | 70        |
| 5.3.1    | Totalen, $v_{TOT_t}$ . . . . .                                                                                                                         | 71        |
| 5.3.2    | Samlet ordretilgang fra eksportmarkedet, $v_{EKSP_t}$ . . . . .                                                                                        | 71        |
| 5.3.3    | Samlet ordretilgang fra hjemmemarkedet, $v_{HJEM_t}$ . . . . .                                                                                         | 71        |
| 5.3.4    | Samlet ordretilgang fra hjemmemarkedet, $i_{HJEM_t}$ , med 9-<br>ledd filteret for trenden . . . . .                                                   | 72        |
| 5.4      | Konklusjon. . . . .                                                                                                                                    | 73        |
| <b>6</b> | <b>Estimation of confidence intervals for seasonally adjusted and trend<br/>time series: A case study based on simulation</b>                          | <b>82</b> |
| 6.1      | Introduction . . . . .                                                                                                                                 | 84        |
| 6.2      | The methods . . . . .                                                                                                                                  | 85        |
| 6.2.1    | Method 1. A parametric method based on simulation of data<br>from an ARIMA model . . . . .                                                             | 85        |

|           |                                                                                                                        |            |
|-----------|------------------------------------------------------------------------------------------------------------------------|------------|
| 6.2.2     | Method 2. Non-parametric bootstrapping utilizing the estimated irregular component from X-11-ARIMA . . . . .           | 86         |
| 6.3       | Application. The first time registered new passenger cars in Norway                                                    | 87         |
| 6.3.1     | Application of method 1: The parametric case . . . . .                                                                 | 88         |
| 6.3.2     | Application of method 2. The non-parametric method utilizing the estimated irregular component of X-11-ARIMA . . . . . | 90         |
| 6.3.3     | Application of method 2. The non-parametric method utilizing the estimated irregular component of X-12-ARIMA . . . . . | 91         |
| 6.4       | Concluding remarks . . . . .                                                                                           | 94         |
| <b>7</b>  | <b>Intervensjoner og trend - et eksempel med data for produksjon av kull og petroleumsprodukter.</b>                   | <b>95</b>  |
| 7.1       | Innledning . . . . .                                                                                                   | 95         |
| 7.2       | Resultater . . . . .                                                                                                   | 96         |
| 7.2.1     | Første kjøring . . . . .                                                                                               | 96         |
| 7.2.2     | Andre kjøring . . . . .                                                                                                | 97         |
| 7.2.3     | Tredje kjøring . . . . .                                                                                               | 98         |
| 7.3       | Oppsummering . . . . .                                                                                                 | 101        |
| <b>8</b>  | <b>Faktisk utførte investeringer totalt for industrien</b>                                                             | <b>102</b> |
| 8.1       | Innledning . . . . .                                                                                                   | 102        |
| 8.2       | Nye tidsserier . . . . .                                                                                               | 103        |
| 8.2.1     | Å lage $Y_{1,t}, Y_{2,t}, \dots, Y_{n,t}$ ved å legge til gradvis nye observasjoner                                    | 103        |
| 8.2.2     | Å lage $U_{1,t}, U_{2,t}, \dots, U_{n,t}$ ved å fjerne gradvis gamle observasjoner                                     | 105        |
| 8.2.3     | Tidsserier med fast lengde . . . . .                                                                                   | 106        |
| 8.3       | Resultater fra sesongjustering for tidsserien $\{O_{87.1}, O_{87.2}, \dots, O_{99.4}\}$ .                              | 108        |
| 8.4       | Simuleringsstudie . . . . .                                                                                            | 111        |
| <b>9</b>  | <b>Eksempler på estimering av kalendereffekter og intervensjoner ved default verdier i X-12-ARIMA</b>                  | <b>114</b> |
| 9.1       | Innledning . . . . .                                                                                                   | 114        |
| 9.2       | Faktisk utførte for de totale investeringene totalt for industrien . . .                                               | 115        |
| 9.2.1     | Uten å angi kalendereffekter og intervensjoner på forhånd . .                                                          | 115        |
| 9.2.2     | Angi påskeeffekten på forhånd . . . . .                                                                                | 116        |
| 9.2.3     | Ved å utføre en estimering for intervensjon før påskeeffekt . .                                                        | 116        |
| 9.2.4     | Ved å angi ukedageffekter på forhånd . . . . .                                                                         | 116        |
| 9.3       | Import av tradisjonelle varer IMP.VITSOR.VR.P . . . . .                                                                | 117        |
| 9.3.1     | Uten å angi på forhånd kalendereffekter eller intervensjoner i programmet . . . . .                                    | 117        |
| 9.3.2     | Ved å angi påskeeffekt på forhånd . . . . .                                                                            | 117        |
| 9.3.3     | Ved å angi intervensjon AO1994.jul på forhånd . . . . .                                                                | 118        |
| 9.4       | Konklusjon . . . . .                                                                                                   | 118        |
| <b>10</b> | <b>Korrigerings for streiken i 2000 og Kristi himmelfartsdag i detaljvolum indeksen</b>                                | <b>119</b> |
| 10.1      | Problem . . . . .                                                                                                      | 119        |
| 10.2      | Beskrivelse av metoden . . . . .                                                                                       | 120        |
| 10.3      | Konklusjon . . . . .                                                                                                   | 122        |

|                                                                                                                    |            |
|--------------------------------------------------------------------------------------------------------------------|------------|
| <b>11 Testen for valg av modell ved sesongjustering og sliding spans i X-12-ARIMA</b>                              | <b>124</b> |
| 11.1 Innledning . . . . .                                                                                          | 124        |
| 11.2 Eksempel 1. Antall passasjerer på charter utenlands ruter . . . . .                                           | 125        |
| 11.3 Eksempel 2. Antall registrerte nye biler i Norge . . . . .                                                    | 132        |
| 11.4 Resultater fra sesongjustering . . . . .                                                                      | 132        |
| 11.5 Konklusjoner . . . . .                                                                                        | 140        |
| <b>12 Internasjonalt seminar om sesongjusteringsmetoder: Rapport fra SAM 98 "Seasonal Adjustment Meeting 1998"</b> | <b>141</b> |
| 12.1 Sammendrag . . . . .                                                                                          | 141        |
| 12.1.1 Politikkkvalg for sesongjustering . . . . .                                                                 | 142        |
| 12.1.2 Anbefalinger . . . . .                                                                                      | 143        |
| 12.2 Invitert foredrag . . . . .                                                                                   | 143        |
| 12.3 Avsluttende rundebordsdiskusjon . . . . .                                                                     | 144        |
| 12.4 Teoretiske aspekter ved sesongjustering . . . . .                                                             | 145        |
| 12.5 Sykel- og trenddefinisjoner og sesongjustering . . . . .                                                      | 147        |
| 12.6 Varians ved målinger og varians ved sesongjustering . . . . .                                                 | 147        |
| 12.7 Erfaringer med sesongjustering . . . . .                                                                      | 148        |
| 12.8 Hva er en god sesongjusteringsmetode? . . . . .                                                               | 150        |
| 12.9 Sammenligning av metoder - Metodologi . . . . .                                                               | 153        |
| 12.10 Sammenligning av metoder - Resultater . . . . .                                                              | 154        |
| 12.11 Presentasjon av applikasjoner og programmer . . . . .                                                        | 156        |
| <b>13 Ukedagseffekter og sesongjusterte tall for detaljvolumindeksen: Et nærmere blikk på desember 2000</b>        | <b>158</b> |
| 13.1 Innledning . . . . .                                                                                          | 158        |
| 13.2 Beskrivelse av dataene . . . . .                                                                              | 159        |
| 13.3 Ukedagseffekter . . . . .                                                                                     | 160        |
| 13.3.1 Tester for endringer av ukedagseffektene med 10 år lange tidsserier                                         | 162        |
| 13.3.2 Tester for bevegelser av ukedagseffekter kun for de senere dataene                                          | 165        |
| 13.4 Resultatene for sesongjusteringen . . . . .                                                                   | 168        |
| 13.4.1 Modell for sesongjustering . . . . .                                                                        | 168        |
| 13.4.2 ARIMA modell . . . . .                                                                                      | 168        |
| 13.4.3 Ukedagseffekter . . . . .                                                                                   | 169        |
| 13.4.4 Intervensjoner . . . . .                                                                                    | 169        |
| 13.4.5 Framskrivinger . . . . .                                                                                    | 170        |
| 13.4.6 Tester for sesongmønster og sesongbevegelser . . . . .                                                      | 170        |
| 13.4.7 Kvalitetsmål og sliding spans . . . . .                                                                     | 170        |
| 13.4.8 Figurer . . . . .                                                                                           | 171        |
| 13.4.9 Revisjoner . . . . .                                                                                        | 173        |
| 13.4.10 Sesongjusterte tall i desember 1999 og 2000 . . . . .                                                      | 174        |
| 13.5 Konklusjon . . . . .                                                                                          | 176        |
| <b>14 Stokastiske modeller for ukedagseffekter for detaljvolumindeksen</b>                                         | <b>177</b> |
| 14.1 Innledning . . . . .                                                                                          | 177        |
| 14.2 Beskrivelse av ukedagseffekter . . . . .                                                                      | 177        |
| 14.3 En stokastisk modell for ukedagseffekter . . . . .                                                            | 178        |
| 14.4 Oppsummering . . . . .                                                                                        | 185        |

# Figurer

|      |                                                                                                                                            |    |
|------|--------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1  | <i>Rådata for totale produksjonsindeksen for industrien . . . . .</i>                                                                      | 6  |
| 2.2  | <i>Rådata for matvarekonsummet . . . . .</i>                                                                                               | 11 |
| 3.1  | <i>Et diagram for de tre ukene i forhold til mars og april . . . . .</i>                                                                   | 13 |
| 3.2  | <i>Virkedager og helligdager for uke før, etter og påskeuka . . . . .</i>                                                                  | 14 |
| 3.3  | <i>Virkedager og helligdager for uke før, etter og påskeuka i 1996 . . . . .</i>                                                           | 18 |
| 3.4  | <i>Observerte og korrigerte verdier for mars ved tabell 3.8 . . . . .</i>                                                                  | 24 |
| 3.5  | <i>Observerte og korrigerte verdier for april ved tabell 3.8 . . . . .</i>                                                                 | 24 |
| 3.6  | <i>Observerte og korrigerte verdier for mars ved X-11 ARIMA . . . . .</i>                                                                  | 24 |
| 3.7  | <i>Observerte og korrigerte verdier for april ved X-11 ARIMA . . . . .</i>                                                                 | 25 |
| 3.8  | <i>Korrigerte verdier for mars ved ny metode og X-11 ARIMA . . . . .</i>                                                                   | 25 |
| 3.9  | <i>Korrigerte verdier for april ved ny metode og X-11 ARIMA . . . . .</i>                                                                  | 25 |
| 3.10 | <i>Sesongjusterte tall for mars ved ny metode og X-11 ARIMA . . . . .</i>                                                                  | 26 |
| 3.11 | <i>Sesongjusterte tall for april ved ny metode og X-11 ARIMA . . . . .</i>                                                                 | 26 |
| 3.12 | <i>Rådata og korrigerte data . . . . .</i>                                                                                                 | 45 |
| 3.13 | <i>Rådata og korrigerte data . . . . .</i>                                                                                                 | 46 |
| 3.14 | <i>Rådata og korrigerte data . . . . .</i>                                                                                                 | 47 |
| 3.15 | <i>Rådata og korrigerte data . . . . .</i>                                                                                                 | 48 |
| 4.1  | <i>Et diagram for de tre ukene i forhold til mars og april . . . . .</i>                                                                   | 49 |
| 4.2  | <i>Påskedagene av <math>T_1</math>, <math>T_2</math> og <math>T_3</math> i 1999 . . . . .</i>                                              | 51 |
| 5.1  | <i>Kvalitets mål i X-11-ARIMA . . . . .</i>                                                                                                | 67 |
| 5.2  | <i>Rådata, sesongjusterte tall og trend av total for ordreindustri . . . . .</i>                                                           | 74 |
| 5.3  | <i>Sesong- og irregulærkomponenter av total for ordreindustri . . . . .</i>                                                                | 75 |
| 5.4  | <i>Rådata, sesongjusterte tall og trend av total eksport i ordreindustri . . . . .</i>                                                     | 76 |
| 5.5  | <i>Rådata, sesongjusterte tall og trend av hjemmemarked i total for ordreindustri . . . . .</i>                                            | 77 |
| 5.6  | <i>Rådata, sesongjusterte tall og trend av total for ordreindustri i absoluttall . . . . .</i>                                             | 78 |
| 5.7  | <i>Rådata, sesongjusterte tall og trend av total eksport i ordreindustri, i absoluttall . . . . .</i>                                      | 79 |
| 5.8  | <i>Rådata, sesongjusterte tall og trend av hjemmemarked i total i ordreindustri, i absoluttall . . . . .</i>                               | 80 |
| 5.9  | <i>Rådata, sesongjusterte tall og trend av hjemmemarked i total i ordreindustri, i indekser, med 9-ledd filteret for trenden . . . . .</i> | 81 |
| 6.1  | <i>The original data for car sales . . . . .</i>                                                                                           | 87 |
| 6.2  | <i>The seasonally adjusted values and the estimated 95% confidence intervals. Parametric case . . . . .</i>                                | 89 |
| 6.3  | <i>The trend values and the estimated 95% confidence intervals. Parametric case . . . . .</i>                                              | 89 |

|       |                                                                                                                                   |     |
|-------|-----------------------------------------------------------------------------------------------------------------------------------|-----|
| 6.4   | <i>The estimated irregular component of X-11-ARIMA . . . . .</i>                                                                  | 90  |
| 6.5   | <i>The Box plot of the estimated irregular component from X-11-ARIMA . . . .</i>                                                  | 90  |
| 6.6   | <i>The seasonally adjusted values and the estimated 95% confidence intervals. Non-parametric method with X-11-ARIMA . . . . .</i> | 91  |
| 6.7   | <i>The trend values and the estimated 95% confidence intervals. Non-parametric method with X-11-ARIMA . . . . .</i>               | 91  |
| 6.8   | <i>The original data and adjusted for prior effects . . . . .</i>                                                                 | 92  |
| 6.9   | <i>The estimated irregular component from X-12-ARIMA . . . . .</i>                                                                | 92  |
| 6.10  | <i>The Box plot of the estimated irregular component from X-12-ARIMA . . . .</i>                                                  | 92  |
| 6.11  | <i>The seasonally adjusted values and the estimated 95% confidence intervals. Non-parametric case with X-12-ARIMA . . . . .</i>   | 93  |
| 6.12  | <i>The trend values and the estimated 95% confidence intervals. Non-parametric case with X-12-ARIMA . . . . .</i>                 | 93  |
| 7.1   | <i>Rådata og trend med log transformasjonen . . . . .</i>                                                                         | 97  |
| 7.2   | <i>Rådata og trend i den andre kjøringen . . . . .</i>                                                                            | 98  |
| 7.3   | <i>Rådata og trend i tredje kjøringen . . . . .</i>                                                                               | 100 |
| 7.4   | <i>Sesongkomponenten i tredje kjøringen . . . . .</i>                                                                             | 100 |
| 7.5   | <i>Irregulærkomponenten i tredje kjøringen . . . . .</i>                                                                          | 100 |
| 8.1   | <i>Rådata . . . . .</i>                                                                                                           | 102 |
| 8.2   | <i>Endringene i prosent fra kvartal til kvartal av rådata . . . . .</i>                                                           | 103 |
| 8.3   | <i>Tidsseriene <math>Y_{1,t}, Y_{2,t}, \dots, Y_{n,t}</math> . . . . .</i>                                                        | 104 |
| 8.4   | <i>Tidsseriene <math>U_{1,t}, U_{2,t}, \dots, U_{n,t}</math> . . . . .</i>                                                        | 105 |
| 8.5   | <i>Tidsseriene <math>V_{1,t}, V_{2,t}, \dots, V_{n,t}</math> . . . . .</i>                                                        | 106 |
| 8.6   | <i>Tidsseriene <math>W_{1,t}, W_{2,t}, \dots, W_{n,t}</math> . . . . .</i>                                                        | 107 |
| 8.7   | <i>Histogrammet av revisjoner . . . . .</i>                                                                                       | 110 |
| 8.8   | <i>Rådata, trend og sesongjusterte tall . . . . .</i>                                                                             | 112 |
| 8.9   | <i>Sesongkomponent . . . . .</i>                                                                                                  | 113 |
| 8.10  | <i>Irregulærkomponent . . . . .</i>                                                                                               | 113 |
| 11.1  | <i>Rådata av passasjerer på charter utlandet . . . . .</i>                                                                        | 124 |
| 11.2  | <i>Rådataene for nye personbiler i Norge . . . . .</i>                                                                            | 125 |
| 11.3  | <i>Rådata gjennom årene, over 12 måneder . . . . .</i>                                                                            | 126 |
| 11.4  | <i>Rådata og de korrigererte dataene for intervensjoner . . . . .</i>                                                             | 129 |
| 11.5  | <i>Sesongjusterte tall ved additiv og multiplikativ modell . . . . .</i>                                                          | 130 |
| 11.6  | <i>Avviket i prosent av sesongjusterte tall ved additiv og multiplikativ modell</i>                                               | 131 |
| 11.7  | <i>Histogram av de største endringene i sesongjusterte tall i fire spans . .</i>                                                  | 131 |
| 11.8  | <i>Antall registrerte nye biler i Norge . . . . .</i>                                                                             | 132 |
| 11.9  | <i>Korrigererte dataene for kalenderseffekter og intervensjoner . . . . .</i>                                                     | 134 |
| 11.10 | <i>Korrigererte dataene for kalenderseffekter og intervensjoner . . . . .</i>                                                     | 134 |
| 11.11 | <i>Sesongjusterte tall ved additiv og multiplikativ modell . . . . .</i>                                                          | 138 |
| 11.12 | <i>Relative avvikene i prosent av sesongjusterte tall . . . . .</i>                                                               | 138 |
| 11.13 | <i>Trender ved additiv og multiplikativ modell . . . . .</i>                                                                      | 139 |
| 11.14 | <i>Relative avvikene i prosent av trender . . . . .</i>                                                                           | 139 |
| 13.1  | <i>Rådata av detaljvolumindeksen . . . . .</i>                                                                                    | 160 |
| 13.2  | <i>Box-whisker plot for 12 måneder . . . . .</i>                                                                                  | 160 |
| 13.3  | <i>De nye tidsseriene med lengden 10 år . . . . .</i>                                                                             | 162 |
| 13.4  | <i>Vektene for ukedager fra tabell 13.6 . . . . .</i>                                                                             | 165 |

|       |                                                                                                                                   |     |
|-------|-----------------------------------------------------------------------------------------------------------------------------------|-----|
| 13.5  | <i>De gjennomsnittlige vektene <math>v_1</math> og <math>v_2</math> beregnes fra tabell 13.6</i>                                  | 165 |
| 13.6  | <i>De nye tidsseriene</i>                                                                                                         | 166 |
| 13.7  | <i>Vektene for ukedager i tabell 13.7</i>                                                                                         | 166 |
| 13.8  | <i>De gjennomsnittlige vektene <math>\bar{v}_{1,}</math> og <math>\bar{v}_{2,}</math> beregnes fra tabell 13.7</i>                | 166 |
| 13.9  | <i>Rådata, trend og irregulære komponenten for tidsserien [jan.79 – des.2000]</i>                                                 | 171 |
| 13.10 | <i>Relative avviket i prosent av sesongjusterte tallene</i>                                                                       | 172 |
| 13.11 | <i>Relative avviket i prosent av trendene</i>                                                                                     | 172 |
| 13.12 | <i>Relative avviket i prosent av sesongkomponentene</i>                                                                           | 172 |
| 13.13 | <i>Rådataene og sesongjusterte tall for tidsserien [jan.90 – des.2000]</i>                                                        | 173 |
| 13.14 | <i>Rådataene og trend for tidsserien [jan.90 – des.2000]</i>                                                                      | 173 |
| 13.15 | <i>Sesongkomponenten for tidsserien [jan.90 – des.2000]</i>                                                                       | 173 |
| 13.16 | <i>Sesongkomponenten for 12 måneder</i>                                                                                           | 174 |
| 13.17 | <i>Revisjoner av sesongjusterte tall</i>                                                                                          | 175 |
| 14.1  | <i>De estimerte koeffisientene <math>\hat{\delta}_{i,t}^s</math></i>                                                              | 181 |
| 14.2  | <i>Standardavvikene for <math>\hat{\delta}_{i,t}^s</math></i>                                                                     | 181 |
| 14.3  | <i>Relative avvikene i prosent mellom <math>O_t^d</math> og <math>O_t^s</math></i>                                                | 183 |
| 14.4  | <i>Relative avvikene i prosent mellom sesongjusterte tall med estimeringen for ukedagseffekter fra X-12-ARIMA og Kalmanfilter</i> | 184 |

# Kapittel 1

## Innledning

Vi har i dette notatet samlet konkrete problemer om sesongjustering i SSB. Hvert problem løses i et eget kapittel. Det er fortsatt mange problemer som ikke er løst, for eksempel pinseeffekter, aggregering og publisering av trenden. Det krever mye tid og arbeid. Dessuten kan vi forbedre løsningene fra de problemene vi har forsøkt å løse som ikke er helt tilfredsstillende ved bruk. Kapitlene kan leses i vilkårlig rekkefølge.

- Kapittel 2 presenterer en metode for å framskrive apriltall for produksjonsindeksen for industrien, når den siste observasjonen er mars, siden det viser ved erfaring at påskekorrigeringen i X-12-ARIMA kan fungere dårlig når apriltall ikke er observert.
- Kapittel 3 beskriver en prekorrigering av påskeeffekten for detaljvolumindeksen. Metoden er utviklet av Leiv Solheim og brukes i SSB for å erstatte den rutinen i X-12-ARIMA som ikke er tilpasset for norsk forhold. Vi får mer tilfredsstillende sesongjusterte tall. Utgangspunktet i metoden er å ta hensyn til antall dager som faller i mars av (i) uka før påske, (ii) påskeuka, (iii) uka etter påske. Dette kapitlet er tidligere publisert i notater 97/73.
- Kapittel 4 er en videre utvikling av metoden i kapittel 3. Vi erstatter de tre ukene: før påske, påskeuka og uka etter påske med tre intervallene  $T_1$ ,  $T_2$  og  $T_3$ , der lengdene  $d_1$ ,  $d_2$  og  $d_3$  velges selv av brukeren,  $d_i \geq 0$ . Metoden blir dermed mer fleksibel.
- Kapittel 5 viser beregninger for sesongjusterte tall (betegnes med  $J$ ) for totale verdien ( $TOT$ ), eksporten ( $EKSP$ ) og hjemmemarkedet ( $HJEM$ ), der  $TOT$  er summen av  $EKSP$  og  $HJEM$ . Ved å utføre X-12-ARIMA direkte for disse tidsseriene får vi gode sesongjusterte tall for  $TOT$  og  $EKSP$ , men tallene blir ustabile for  $HJEM$ . Vi har beregnet sesongjusterte tall for  $HJEM$  ved å ta differansen  $J(TOT) - J(EKSP)$ , men resultatet viser den samme konklusjonen.
- Kapittel 6 beskriver en metode utviklet i samarbeid med Terje Skjerpen for å simulere 95% konfidensintervaller for trend og sesongjusterte tall fra X-12-ARIMA. Resultatene ble presentert av Terje Skjerpen på konferansen for sesongjustering SAM-98 i Bucharest 22.-24. oktober 1998.
- Kapittel 7 er et godt eksempel for å vise at man kan få en dårlig trend på grunn av korrigeringen for intervensjoner. I dette tilfellet bør brukeren sjekke tidspunkter og typer for intervensjoner ved hjelp av  $t$ -verdi. Data er produksjon av kull og petroleums produkter og kjerne brensel fra januar 1986 til desember 1999.
- Kapittel 8 er sesongjustering for faktisk de utførte totale investeringene for industri-

en med data fra første kvartal 1975 til fjerde kvartal 1999. Resultatet viser ustabile sesongkomponent og sesongjusterte tall slik at vi ikke kan publisere tallene. Ved å analysere delmengder av dataene får vi en bedre sesongjustering når data før 1987 blir fjernet.

- Kapittel 9 er et eksempel som viser at korrigeringen for kalendereffekter og estimeringen for intervensjoner i samme kjøring med default verdier ikke fungerer for alle dataene. I dette tilfellet bør brukeren angi kalendereffekter på forhånd og estimere intervensjoner eller gjøre det omvendt.

- Kapittel 10 beskriver en metode for å korrigere for streiken i månedskiftet april-mai og Kristi himmelfartsdag i juni 2000 før vi sesongkorrigerer detaljvolumindeksen ved hjelp av dummy variable.

- Kapittel 11 viser testen av en modell for sesongjustering og sliding spans i X-12-ARIMA. Dataene vi bruker er antall passasjerer på charter utenlands ruter og antall registrerte første gang personbiler i Norge.

- Kapittel 12 er et rapport fra SAM-98 "Seasonal Adjustment Meeting 1998" sammen med Anne Berit Dahle, Terje Skjerpen og Joaquin Rodriguez. Seminaret fant sted i Bucharest 22-24 oktober 1998. Dette kapitlet er publisert ved interne dokumenter 99/2.

- Kapittel 13 beskriver ukedagseffekter og sesongjusterte tall for detaljvolumindeksen. Vi har gjennomført en rekke analyser på delmengder av dataene. Det viser at vektene for ukedager har endret seg over tid. I de siste årene fra 1991 er torsdag ikke lenger en så stor handledag.

- Kapittel 14 beskriver stokastiske modeller for ukedagseffekter for detaljvolumindeksen. Vektene  $\delta_{i,t}$  for ukedager kan variere fra tidspunkt til tidspunkt og beskrives ved hjelp av en ARIMA modell. Kalmanfilteret er brukt for estimeringen. Vi får et bilde av variasjoner for  $\delta_{i,t}$ . Det viser at  $\delta_{i,t}$  har endret seg sakte med tiden slik at vi har et godt grunnlag for å anta at de er konstanter, slik som lagt til grunn i X-12-ARIMA.

Vi takker Leiv Solheim og Terje Skjerpen for gode kommentarer.

## Kapittel 2

# En metode for å framskrive apriltall, for totale produksjonsindeksen for industrien

Dinh Quang Pham

### 2.1 Innledning

De fleste tidsseriene i produksjonsindeksen for industrien er ofte påvirket av påske. Påskekorrigeringen i X-11-ARIMA eller i X-12-ARIMA fungerer dårlig når mars er den siste observasjonen. Endringen av sesongjusterte tall fra februar til mars er ofte mye større enn endringen når tallet for april føyes til. For å unngå dette problemet kan vi forlenge den originale tidsserien med en predikert verdi for april. Deretter sesongjusterer vi den nye serien. Vi framskriver en verdi for april i det siste året ved hjelp av en ARIMA modell. Antar at ARIMA (0,1,1)(0,1,1) modellen gir god tilpasning for data  $Y_t$ ,  $t = \text{jan.86}, \dots, \text{mar.99}$ . Da kan vi skrive  $Y_t$  på formen

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + \epsilon_t - \theta\epsilon_{t-1} - \Theta\epsilon_{t-12} + \theta\Theta\epsilon_{t-13} \quad (2.1)$$

Minste kvadraters framskriving for april 1999 er gitt ved

$$\begin{aligned} \hat{Y}_{apr.99} &= E(Y_{apr.98} | Y_{feb.99}, \dots, Y_{jan.86}) \\ &= Y_{mar.99} + Y_{apr.98} - Y_{mar.98} - \hat{\Theta}\epsilon_{apr.98} + \hat{\theta}\hat{\Theta}\epsilon_{mar.98} \end{aligned} \quad (2.2)$$

(2.2) kan gi et estimat som har stor avvik fra den observerte verdien  $Y_{apr.99}$ , siden  $\hat{Y}_{apr.99}$  estimeres uten å ta hensyn til påskeeffekten som skyldes forskyvningen av aktiviteten fra april til mars. Vi bruker den totale produksjonsindeksen for industrien som et eksempel. Data er observert fra januar 1986 til mars 1999 og plottet i figur 2.1. Vi bruker X-12-ARIMA som et verktøy for sesongjustering. I dette programmet er det en rutine som kan estimere antall dager  $d$  før 1. påskedag som er påvirket av påskeeffekten. Vi får  $d = 8$ . Siden påskedagene varierer fra år til år, er en tolkning for denne verdien i 1999 at produksjonen er lavere i en periode fra lørdag den 27.

mars til 1. påskedag den 4. april. Hvis  $Y_{mar.99}$  er summen av dagens aktivitet må denne verdien korrigeres opp for å ha en riktig verdi av mars. I motsetning til 1999, kom påsken i helhet i april i 1998 (1. påskedag var 12 april). Lav produksjon i denne perioden har ikke innvirkning på verdien til mars. Vi har ingen korrigerings for mars 1998 før sesongjustering. Påskeeffekten beskrives av en faktor  $F_t$ . I X-11-ARIMA er  $F_t = 1$  for årene hvor perioden av  $d$  dager før 1. påskedag faller helhet i april og  $F_t \neq 1$  for de andre årene. Mens i X-12-ARIMA blir mars og april korrigert uansett hvor den perioden av  $d$  dager ligger. Vi ser at  $\hat{Y}_{apr.99}$  i ligningen (2.2) beregnes med påskefaktoren av 1998. Dette gir derfor ikke et riktig estimat for  $Y_{apr.99}$ .

I de neste avsnittene vil vi presentere en metode for å estimere  $\hat{Y}_{apr.99}$ . Den er også basert på en ARIMA modell, men beregningen er litt mer komplisert enn ligningen (2.2). I notatet har vi brukt data fra den totale produksjonsindeksen for industrien og matvarekonsummet som eksempler for beregningene. Tallene er observert fra januar 1986 til mars 1999 for den første serien og fra første kvartal 1978 til første kvartal 1999 for den andre.

Vi innfører notasjon som brukes i de neste avsnittene

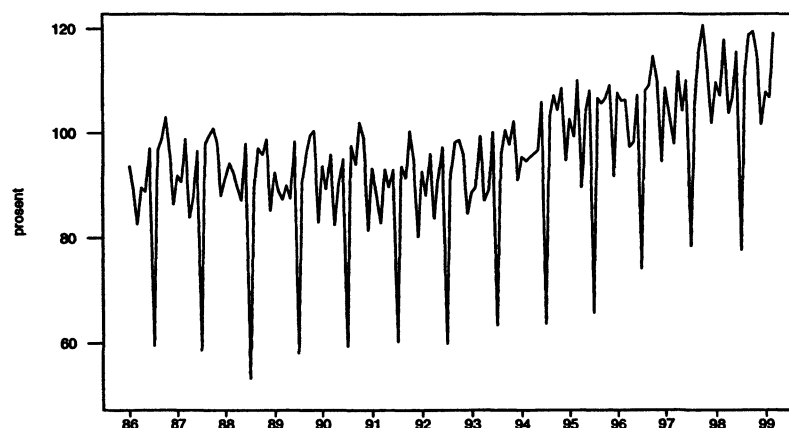
|                           |                                                                             |
|---------------------------|-----------------------------------------------------------------------------|
| $Y_t$                     | den originale (måneds)data                                                  |
| $Y_{mar.99}$              | den observerte verdien for mars i 1999                                      |
| $Y_{apr.99}$              | den observerte verdien for april i 1999                                     |
| $\hat{Y}_{mar.99}$        | den estimerte verdien for mars 1999                                         |
| $\hat{Y}_{apr.99}$        | den estimerte verdien for april 1999                                        |
| $\hat{Y}_{mar.99}^K$      | den korrigerte verdien med påskeeffekten for mars 1999                      |
| $\hat{Y}_{apr.99}^K$      | den korrigerte verdien med påskeeffekten for april 1999                     |
| $Y_t^*$                   | den simulerte serien av $Y_t$ ved hjelp av en ARIMA modell                  |
| $\hat{Y}_{apr.99}^{*(i)}$ | den simulerte verdien av april 1999 i $i$ -te iterasjonen                   |
| $Z_t$                     | den tidsserien med termintall som beregnes ved<br>$Z_t = Y_{2t-1} + Y_{2t}$ |
| $\hat{Z}_t$               | den estimerte verdien av $Z_t$                                              |

## 2.2 En beskrivelse av metoden for estimering av april-tall

Metoden kan anvendes for måneds- og kvartalstidsserier med multiplikativ eller additiv modell. Med den multiplikative modellen mener vi at data blir transformert, for eksempel med  $\log$ , før beregning. For additive modellen er det rådata brukes som input.

### 2.2.1 Månedstidsserie og multiplikativ modell ( $\log$ transformasjon)

For å kunne beregne  $\hat{Y}_{apr.99}$  ved (2.2) må vi standardisere  $Y_{mar.99}$ ,  $Y_{mar.98}$  og  $Y_{apr.98}$  med deres påskefaktorer slik at disse størrelsene ikke er påvirket av påskeeffekten. I utlistingen av X-12-ARIMA er det en tabell som heter A.7 som lister ut påskefaktorene i prosent for mars og april  $F_{mar.}$  og  $F_{apr.}$  for hele serien. Den framskrivingen  $\hat{F}_{apr.99}$



Figur 2.1: Rådata for totale produksjonsindeksen for industrien

finnes i tabell A.7.A. La  $Y_{mar.}^K$  og  $Y_{apr.}^K$  være de korrigerte verdiene for mars og april med påskeeffekten. De beregnes ved

$$Y_{mar.}^K = \frac{Y_{mar.}}{F_{mar.}} \quad \text{og} \quad Y_{apr.}^K = \frac{Y_{apr.}}{F_{apr.}}, \quad \text{for årene før 1999} \quad (2.3)$$

og for april 1999 er  $\hat{Y}_{apr.99}^K = \hat{Y}_{apr.99} / \hat{F}_{apr.99}$ . Vi erstatter  $Y_{mar.}$  og  $Y_{apr.}$  i de originale data med  $Y_{mar.}^K$  og  $Y_{apr.}^K$  og tilpasser med en ARIMA modell. Vi antar at ARIMA (0,1,1)(0,1,1) modellen gir en god tilpasning med data. Den estimerte verdien etter å ha korrigert for påskeeffekten for april 1999 beregnes ved

$$\hat{Y}_{apr.99}^K = Y_{mar.99}^K + Y_{apr.98}^K - Y_{mar.98}^K - \hat{\theta}' \epsilon_{apr.98} + \hat{\theta}' \hat{\theta}' \epsilon_{mar.98} \quad (2.4)$$

Den virkelige estimerte verdien for april 1999 blir

$$\hat{Y}_{apr.99} = \hat{Y}_{apr.99}^K \times \hat{F}_{apr.99} \quad (2.5)$$

#### Ekesempel: Estimering $\hat{Y}_{apr.99}$ for den totale produksjonsindeksen

ARIMA (0,1,1)(0,1,1) modellen gir en god tilpasning til data. Ved å la  $d$  estimeres fra data får vi  $d = 8$ . Vi estimerer  $\hat{Y}_{apr.99}$  med  $d=1$  og 3 også. For eksempel, med  $d = 1$  påstår vi at lørdag den 3. april er påvirket av påskeeffekten. Men for  $d = 3$  er skjærtorsdag, langfredag og lørdag påvirket. Effekten sprer over en lengre periode med  $d = 8$ . Vi får tilsvarende  $F_{mar.}$  og  $F_{apr.}$  for hver verdi av  $d$ . Rådata etter å ha korrigert for påskefaktorene blir lagret i tabell B.1 i utlistingen og brukes senere for å framskrive  $\hat{Y}_{apr.99}^K$ . Tabell 2.1 viser påskefaktorene og de estimerte verdiene for april 1999. Tall i  $[ , ]$  er 95% konfidensintervallene for  $\hat{Y}_{apr.99}^K$ . Den estimerte verdien for april 1999 med  $d = 1$ , er

$$\hat{Y}_{apr.99} = \hat{Y}_{apr.99}^K \times \hat{F}_{apr.99} = 108.18 \times 0.9675 = 104.66$$

Den faktiske verdien for april 1999 er  $Y_{apr.99} = 104.70$ . Vi får svært gode estimater for  $Y_{apr.99}$  ved  $d = 1$  eller  $d = 3$ , mens  $d = 8$  gir et dårlig estimat.

Tabell 2.1: Påskefaktorene og de estimerte verdiene for april 1999

| $d$             | $\hat{F}_{apr.99}$ | $\hat{Y}_{apr.99}^K$    | $\hat{Y}_{apr.99}$ |
|-----------------|--------------------|-------------------------|--------------------|
| 1 (sign. på 5%) | 0.9675             | 108.18 [101.02, 115.84] | 104.66             |
| 3 (sign. på 5%) | 0.9595             | 108.39 [101.58, 115.66] | 104.00             |
| 8 (sign. på 5%) | 1.0249             | 110.18 [108.21, 117.63] | 112.92             |

## 2.2.2 Månedstidsserie og additiv modell

Vi bruker ingen transformasjon for rådata. De påskefaktorene  $F_{mar.}$ ,  $F_{apr.}$  og  $\hat{F}_{apr.99}$  er beregnet i absoluttall og hentet fra tabell A.7 og A.7.A. De korrigerte verdiene for påskeeffekten i mars og april blir

$$Y_{mar.}^K = Y_{mar.} - F_{mar.} \quad \text{og} \quad Y_{apr.}^K = Y_{apr.} - F_{apr.} \quad (2.6)$$

Vi erstatter  $Y_{mar.}$  og  $Y_{apr.}$  i den originale serien med  $Y_{mar.}^K$  og  $Y_{apr.}^K$  fra (2.6) og tilpasser med en ARIMA modell. Deretter beregner vi framskrivingen for april 1999,  $\hat{Y}_{apr.99}^K$ . Den virkelige estimerte verdien for april 1999 blir

$$\hat{Y}_{apr.99} = \hat{Y}_{apr.99}^K + \hat{F}_{apr.99} \quad (2.7)$$

Vi ser at ligningen (2.6) og (2.7) tilsvarende (2.3) og (2.5) for multiplikativ modellen.

**Eksempel: Estimering  $\hat{Y}_{apr.99}$  for den totale produksjonsindeksen**

Vi estimerer  $\hat{Y}_{apr.99}$  med additive modell, dvs ingen transformasjon for rådata er brukt i beregningen. Vi anvender metoden med forskjellige verdier av  $d$ . Resultatene er vist i tabell 2.2.

Tabell 2.2: Påskefaktorene og de estimerte verdiene for april 1999

| $d$             | $\hat{F}_{apr.99}$ | $\hat{Y}_{apr.99}^K$    | $\hat{Y}_{apr.99}$ |
|-----------------|--------------------|-------------------------|--------------------|
| 1 (sign. på 5%) | -3.13              | 107.03 [101.31, 112.75] | 103.90             |
| 3 (sign. på 5%) | -3.89              | 106.98 [101.41, 112.54] | 103.09             |
| 8 (sign. på 5%) | 2.30               | 108.79 [103.00, 114.58] | 111.09             |

Den observerte verdien for april 1999 er  $Y_{apr.99} = 104.70$ . Vi får et relativt godt estimat for  $d=1$  og  $d=3$ , men altfor dårlig med  $d=8$ .

## 2.2.3 En alternativ måte å estimere apriltall for en månedstidsserie

### a. multiplikativ modell

Vi vil estimere  $\hat{Y}_{apr.99}$  ved å fordele den estimerte verdien for både mars og april 1999 med deres påskefaktorene,  $F_{mar.99}$  og  $\hat{F}_{apr.99}$ . Vi lager en termintidsserie,  $Z_t$ , ved å slå sammen

|                        |     |           |
|------------------------|-----|-----------|
| - januar og februar    | for | 1. termin |
| - mars og april        |     | 2. termin |
| - mai og juni          |     | 3. termin |
| - juli og august       |     | 4. termin |
| - september og oktober |     | 5. termin |
| - november og desember |     | 6. termin |

Vi beregner  $Z_t$  ved formen

$$Z_t = Y_{2t-1} + Y_{2t}$$

For eksempel,  $Z_1 = Y_1 + Y_2 = Y_{jan.86} + Y_{feb.86}$ , er den første terminen av 1986.  $\{Z_t\}$  er en tidsserie med 6 observasjoner for hvert år og ikke påvirket av påskeeffekten. La  $\hat{Z}_{2.99}$  være den estimerte verdien for den andre terminen 1999. Vi skriver  $\hat{Z}_{2.99}$  som

$$\hat{Z}_{2.99} = \hat{Y}_{mar.99} + \hat{Y}_{apr.99} \quad (2.8)$$

Det er en ligning med to ukjente variable  $\hat{Y}_{mar.99}$  og  $\hat{Y}_{apr.99}$ . Antar at vi kjenner påskefaktorene  $F_{mar.99}$  og  $\hat{F}_{apr.99}$ , så kan vi finne en annen ligning for  $\hat{Y}_{mar.99}$  og  $\hat{Y}_{apr.99}$ . De korrigerte verdiene for påskeeffekten,  $\hat{Y}_{mar.99}^K$  og  $\hat{Y}_{apr.99}^K$  beregnes ved

$$\hat{Y}_{mar.99}^K = \frac{\hat{Y}_{mar.99}}{F_{mar.99}} \quad \text{og} \quad \hat{Y}_{apr.99}^K = \frac{\hat{Y}_{apr.99}}{\hat{F}_{apr.99}}$$

Siden

$$\hat{Z}_{2.99} = \hat{Y}_{mar.99}^K + \hat{Y}_{apr.99}^K$$

får vi

$$\hat{Z}_{2.99} = \frac{\hat{Y}_{mar.99}}{F_{mar.99}} + \frac{\hat{Y}_{apr.99}}{\hat{F}_{apr.99}} \quad (2.9)$$

Merk at  $F_{mar.99} + \hat{F}_{apr.99} = 2$ . Fra (2.8) og (2.9), får vi

$$\hat{Y}_{mar.99} = \frac{\hat{Z}_{2.99}}{2} \times F_{mar.99} \quad \text{og} \quad \hat{Y}_{apr.99} = \frac{\hat{Z}_{2.99}}{2} \times \hat{F}_{apr.99} \quad (2.10)$$

Vi ser at den observerte verdien av mars 1999 har ikke vært brukt i estimeringen av  $\hat{Y}_{apr.99}$ , men i beregningen av  $F_{mar.99}$  og  $\hat{F}_{apr.99}$  i tabell A.7 og A.7.A.

## b. additiv modell

$\hat{Y}_{mar.99}$  og  $\hat{Y}_{apr.99}$  estimeres med tilsvarende formlene som (2.10), dvs

$$\hat{Y}_{mar.99} = \frac{\hat{Z}_{2.99}}{2} + F_{mar.99} \quad \text{og} \quad \hat{Y}_{apr.99} = \frac{\hat{Z}_{2.99}}{2} + \hat{F}_{apr.99} \quad (2.11)$$

der  $F_{mar.99} + \hat{F}_{apr.99} = 0$  og er i absoluttall.

### Eksempel: Estimering $\hat{Y}_{apr.99}$ for total produksjonsindeksen

#### a. multiplikativ modell

ARIMA (0,1,1)(0,1,1) modellen ble valgt for  $Z_t$ .  $\hat{Z}_{2.99} = 219.60$ . Vi estimerer  $\hat{Y}_{apr.99}$  med tre forskjellige verdier av  $d$ .  $F_{mar.99}$  og  $\hat{F}_{apr.99}$  hentes fra tabell 2.1. Resultatene er vist i tabell 2.3.

Tabell 2.3: Påskefaktorene og estimerte verdiene for mars og april 1999

| $d$             | $\hat{F}_{mar.99}$ | $\hat{F}_{apr.99}$ | $\hat{Y}_{mar.99}$ | $\hat{Y}_{apr.99}$ |
|-----------------|--------------------|--------------------|--------------------|--------------------|
| 1 (sign. på 5%) | 1.0325             | 0.9675             | 113.36             | 106.23             |
| 3 (sign. på 5%) | 1.0405             | 0.9595             | 114.25             | 105.35             |
| 8 (sign. på 5%) | 0.9751             | 1.0249             | 107.07             | 112.53             |

De faktiske verdiene for mars og april 1999 er  $Y_{mar.99} = 119.20$  og  $Y_{apr.99} = 104.70$ . Vi får langt dårligere estimater enn i 2.1.

#### b. additiv modell

$F_{mar.99}$  og  $\hat{F}_{apr.99}$  er hentet fra tabell 2.2. Verdiene av  $\hat{Y}_{mar.99}$  og  $\hat{Y}_{apr.99}$  er vist i tabell 2.4. De ligger langt fra deres faktiske verdier.

Tabell 2.4: Påskefaktorene og estimerte verdiene for mars og april 1999

| $d$             | $\hat{F}_{mar.99}$ | $\hat{F}_{apr.99}$ | $\hat{Y}_{mar.99}$ | $\hat{Y}_{apr.99}$ |
|-----------------|--------------------|--------------------|--------------------|--------------------|
| 1 (sign. på 5%) | 3.13               | -3.13              | 112.93             | 106.67             |
| 3 (sign. på 5%) | 3.89               | -3.89              | 113.69             | 105.91             |
| 8 (sign. på 5%) | -2.30              | 2.30               | 107.50             | 112.10             |

### Kommentarer

Ligningene (2.10) og (2.11) kan brukes for et vilkårlig år  $t$ . Vi ser at når  $F_{mar.t} = \hat{F}_{apr.t} = 1$  for multiplikativ modell eller  $F_{mar.t} = \hat{F}_{apr.t} = 0$  for additiv modell, er  $\hat{Y}_{mar.t} = \hat{Y}_{apr.t} = \hat{Z}_{2.t}/2$ . Dette gir ofte et dårlig estimat enten for  $Y_{mar.t}$  eller  $Y_{apr.t}$ . En annen ulempe er at den observerte verdien av mars i det siste året ikke har vært brukt i beregningen. Dermed kan den estimerte verdien for april gi et større avvik.

#### 2.2.4 Kvartalstall

Data i eksemplet er matvarekonsummet i million kroner, fra 1. kvartal 1978 til 1. kvartal 1999. Vi betegner med  $Q_t$ , hvor  $t = 1kv78, \dots, 1kv99$ . Vi vil estimere en verdi for det andre kvartalet 1999,  $\hat{Q}_{2kv99}$ , med hensyn til påskeeffekten. Vi skriver slik

$$\hat{Q}_{2kv99} = \hat{Y}_{apr.99} + \hat{Y}_{mai.99} + \hat{Y}_{jun.99}$$

Vi bruker samme metoden i seksjon 2.1 og 2.2.

For multiplikativ modell, får vi

$$\hat{Q}_{2kv99}^M = \hat{Q}_{2kv99}^K \times \hat{F}_{2kv99} \quad (2.12)$$

For additiv modell, får vi

$$\hat{Q}_{2kv99}^A = \hat{Q}_{2kv99}^K + \hat{F}_{2kv99} \quad (2.13)$$

hvor  $\hat{Q}_{2kv99}^K$  er den estimerte verdien for 2. kvartal 1999 som er korrigert for påskeeffekten.  $\hat{F}_{2kv99}$  er påskefaktoren for 2. kvartal 1999. Vi ser at for månedstall eller kvartalstall skyldes påskeeffekten flytting av aktiviteten fra april til mars. Dermed er  $F_{1kv99}$  egentlig  $F_{mar.99}$  og  $\hat{F}_{2kv99}$  er  $\hat{F}_{apr.99}$ .

Fra (2.12), har vi

$$\begin{aligned} \hat{Q}_{2kv99}^M &= (\hat{Y}_{apr.99}^K + \hat{Y}_{mai.99} + \hat{Y}_{jun.99}) \times \hat{F}_{2kv99} \\ &= \hat{Y}_{apr.99}^K \times \hat{F}_{2kv99} + \hat{Y}_{mai.99} \times \hat{F}_{2kv99} + \hat{Y}_{jun.99} \times \hat{F}_{2kv99} \\ &\neq \hat{Q}_{2kv99} \end{aligned}$$

Vi ser at mai og juni 1999 blir også korrigert for påskeeffekten. Dette medfører et dårlig estimat for det andre kvartalet 1999.

Fra (2.13), skriver vi

$$\begin{aligned} \hat{Q}_{2kv99}^A &= (\hat{Y}_{apr.99}^K + \hat{Y}_{mai.99} + \hat{Y}_{jun.99}) + \hat{F}_{2kv99} \\ &= (\hat{Y}_{apr.99}^K + \hat{F}_{2kv99}) + \hat{Y}_{mai.99} + \hat{Y}_{jun.99}, \\ &= (\hat{Y}_{apr.99}^K + \hat{F}_{apr.99}) + \hat{Y}_{mai.99} + \hat{Y}_{jun.99}, \quad \text{siden } \hat{F}_{2kv99} = \hat{F}_{apr.99} \\ &= \hat{Y}_{apr.99} + \hat{Y}_{mai.99} + \hat{Y}_{jun.99} \quad \text{ifølge (2.7)} \\ &= \hat{Q}_{2kv99} \end{aligned}$$

Vi ser at påskefaktoren ikke har påvirket mai og juni i beregningen. Vi vil bruke den additive modellen for kvartalstall i estimering apriltall.

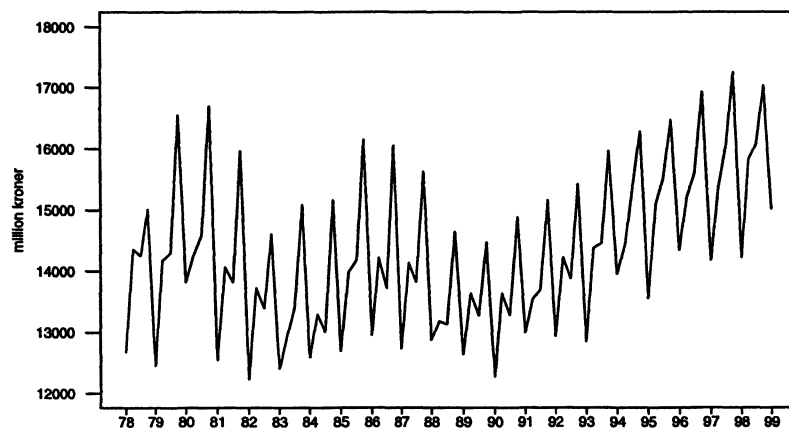
### Eksempel: Estimering $\hat{Q}_{2kv99}$ for matvarekonsummet

Data er plottet i figur 2.2. Additive modellen er brukt for å estimere  $\hat{Q}_{2kv99}$  med tre forskjellige verdier av  $d$ ,  $d=1$ , 8 og 15. Resultatene er vist i tabell 2.5.

Tabell 2.5: Påskefaktorene og de estimerte verdiene for april 1999

| $d$              | $\hat{F}_{2kv99}$ | $\hat{Q}_{2kv99}^K$           | $\hat{Q}_{2kv99}$ |
|------------------|-------------------|-------------------------------|-------------------|
| 1 (ikke på 5%)   | -1.00             | 16009.03 [15162.69, 16855.38] | 16008.03          |
| 8 (sign. på 5%)  | -53.80            | 15970.45 [15151.14, 16789.76] | 15916.65          |
| 15 (sign. på 5%) | -84.54            | 15980.31 [15167.51, 16793.12] | 15895.77          |

Testen viser at  $d=1$  ikke er signifikant på 5% nivået. Vi kjenner ikke den faktiske verdien til 2. kvartal 1999 for å kunne sammenligne med estimerte verdien.



Figur 2.2: Rådata for matvarekonsummet

## 2.3 Oppsummering

Vi bruker samme metode for å estimere apriltall for en månedsserie og kvartalsserie. Den er basert på en ARIMA modell som gir god tilpasning med data. Vi ser at for månedstall vil framskriving på grunnlag av observasjoner fram til mars fungere langt bedre enn framskriving først av 2. termin og deretter fordeling på mars og april. For kvartalstall gir den additive modellen et bedre estimat. Vi fikk et godt estimat for april 1999 for produksjonsindeksen siden data har en regelmessig variasjon. En ulempe med denne metoden er at vi får ikke et konfidensintervall for estimerte verdien, siden X-12-ARIMA ikke lister ut usikkerhetene for de påskefaktorene,  $F_{mar.99}$  og  $\hat{F}_{apr.99}$ . En alternativ måte er å beregne en rekke av  $N$  simulerte verdier  $\hat{Y}_{apr.99}^{(1)}$ ,  $\hat{Y}_{apr.99}^{(2)}$ , ...,  $\hat{Y}_{apr.99}^{(N)}$ . Standardavviket av  $\hat{Y}_{apr.99}$  beregnes da ved

$$Std(\hat{Y}_{apr.99}) = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\hat{Y}_{apr.99}^{(i)} - \hat{Y}_{apr.99}^*)^2}$$

hvor  $\hat{Y}_{apr.99}^* = \frac{1}{N} \sum_{i=1}^N \hat{Y}_{apr.99}^{(i)}$ . For hver iterasjon simulerer vi nye data fra ARIMA modell og korrigerede faktorene  $F_{mar.99}$  og  $\hat{F}_{apr.99}$ .  $\hat{Y}_{apr.99}$  beregnes ved (2.5) eller (2.7). Etter  $N$  simuleringer kan vi beregne standardavviket og konfidensintervallet for estimatet. Vi kan lage et konfidensintervall for  $\hat{Q}_{apr.99}$  på samme måte.

## Kapittel 3

# Prekorrigering av påskeeffekten for detaljvolumindeksen 1979-1997

Leiv Solheim og Dinh Quang Pham

### 3.1 Innledning

I X-11 ARIMA finnes det en rutine for påsekorrigering med opsjoner som er tilpasset for norske forhold. Det har vist seg at for mange serier, spesielt detaljvolumindeksen, fungerer denne metoden dårlig. Vi har derfor laget en ny rutine for påsekorrigeringen for detaljvolumindeksen. Metoden er enkel og basert på å telle antall virke- eller helligdager, i uka før påske, påskeuka og uka etter påske, som faller i mars. Vi bruker multippel regresjonsanalyse for å estimere effektene. I det neste avsnittet vil vi presentere metoden og programmet. Metoden er blitt brukt siden januar 1995.

### 3.2 Symboler

Før vi presenterer metoden, vil vi definere noen symboler:

|                        |                                              |
|------------------------|----------------------------------------------|
| $Y_{M,i}$              | observert verdi for mars i år $i$ .          |
| $Y_{A,i}$              | observert verdi for april i år $i$ .         |
| $\hat{Y}_{M,i}^{korr}$ | prekorrigert påskeverdi for mars i år $i$ .  |
| $\hat{Y}_{A,i}^{korr}$ | prekorrigert påskeverdi for april i år $i$ . |

Når vi vil sammenligne med prekorrigerte verdier for påskeeffekter ved X-11 ARIMA bruker vi

|                                                      |                                                              |
|------------------------------------------------------|--------------------------------------------------------------|
| $\hat{Y}_{M,i}^{Reg,korr}$ , $\hat{Y}_{M,i}^{Reg,k}$ | som prekorrigert verdi for mars i år $i$ ved den nye metoden |
| $\hat{Y}_{M,i}^{X11,korr}$ , $\hat{Y}_{M,i}^{X11,k}$ | som prekorrigert verdi for mars i år $i$ ved X-11 ARIMA      |

Tilsvarende for april er  $\hat{Y}_{A,i}^{Reg,korr}$  og  $\hat{Y}_{A,i}^{X11,korr}$ .

Vi bruker regresjonsanalyse for å estimere  $\hat{Y}_{M,i}^{korr}$  og  $\hat{Y}_{A,i}^{korr}$  under to betingelser:

- Ingen forskyving mellom virke- og helligdager.
- Ingen forskyving ut/inn av de tre ukene.

### 3.3 Hvorfor må vi korrigere rådata for påskeeffekt

Siden påske ikke er en fast uke i året vil sesongmønsteret være avhengig av tidspunktet for påske. Det fører til feil i sesongmønsteret og deretter de sesongjusterte tallene for mars og april dersom en sesongjusterer en serie uten å prekorrigere for påskeeffekten. Vi definerer i notatet

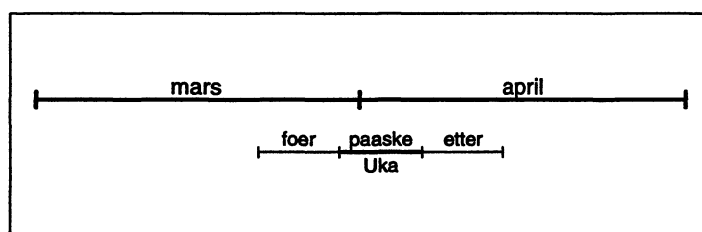
- **Påskeeffekt** effekten som påvirker rådata på grunn av påske.
- **Påskedagene** dagene som rådata blir påvirket av påskeeffekten. I X-11 ARIMA er det en rutine som brukes for å estimere antall dager  $d$  før 1. påskedag som rådata er påvirket av påskeeffekt. Når  $\hat{d} = 9$  betyr det at skjærtorsdag, langfredag er blant de påskedagene, mens når  $\hat{d} = 1$  er bare lørdag, 1. påskedag påskedagene.  $\hat{d}$  varierer fra serie til serie.  $\hat{d} = 1, 2, \dots, 25$ . I vår metode er påskedagene lik 3 uker: en uke før, påskeuka og uka etter påske.

### 3.4 Metoden

Utgangspunktet i metoden er å ta hensyn til

- Antall dager i uka før påske som faller i mars.
- Antall dager i påskeuka som faller i mars.
- Antall dager i uka etter påske som faller i mars.

Mandag er definert som den første dagen i uka. Figur 3.1 viser et eksempel på hvordan påskeuka faller i både mars og april.



Figur 3.1: Et diagram for de tre ukene i forhold til mars og april

Vi må skille mellom to tilfeller:

- 1 Både mars og april er observert det siste året.
- 2 Bare mars er observert det siste året.

### 3.4.1 Både mars og april er observert i det siste året

Dette betyr at vi har en tidsserie som har minst 4 måneder i det siste året. La  $\hat{Y}_{M,i}^{korr}$  og  $\hat{Y}_{A,i}^{korr}$  være estimatorene for  $Y_{M,i}$  og  $Y_{A,i}$  etter å ha korrigert for påskeeffekten. Vi vil korrigere mars og april observasjonene til tall som vi ville forventet å observere dersom de tre ukene i sin helhet faller i april.  $\hat{Y}_{M,i}^{korr}$  og  $\hat{Y}_{A,i}^{korr}$  estimeres på følgende måte:

La  $Y_{M,i}$  og  $Y_{A,i}$  betegne de observerte verdiene i hhv mars og april for år  $i$ .

La  $X_{M,i}$  og  $X_{A,i}$  være differansen mellom hhv mars og april sammenliknet med gjennomsnittet for de to månedene. Vi definerer

$$X_{M,i} = Y_{M,i} - \frac{Y_{M,i} + Y_{A,i}}{2} \quad (3.1)$$

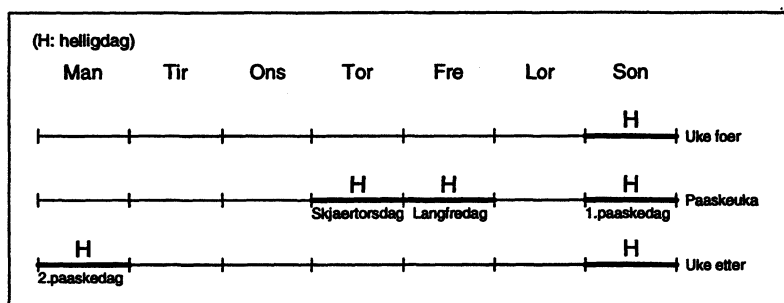
$$X_{A,i} = Y_{A,i} - \frac{Y_{M,i} + Y_{A,i}}{2} \quad (3.2)$$

Merk at  $X_{M,i} + X_{A,i} = 0$ . Kall sesongutslaget for mars  $S$  (dvs sesongutslaget for april blir  $-S$ ) når vi sammenlikner med gjennomsnittet for de to månedene. La

$$\begin{aligned} I_{-1} &= \text{antall virkedager i mars i uka før påske/6} \\ I_0 &= \text{antall virkedager i mars i påskeuka/4} \\ I_1 &= \text{antall virkedager i mars i uka etter påske/5} \end{aligned} \quad (3.3)$$

Vi innfører i tillegg på tilsvarende vis andelen helligdager/søndager i mars i de tilsvarende ukene.

$$\begin{aligned} J_{-1} &= \text{antall helligdager i mars i uka før påske/1} \\ J_0 &= \text{antall helligdager i mars i påskeuka/3} \\ J_1 &= \text{antall helligdager i mars i uka etter påske/2} \end{aligned} \quad (3.4)$$



Figur 3.2: Virkedager og helligdager for uke før, etter og påskeuka

Figur 3.2 illustrerer for virkedagene og helligdagene i uke før, etter og påskeuka. "H" står som helligdager. De blanke feltene er virkedager. Skjærtorsdag, langfredag, 1. påskedag og 2. påskedag er blant helligdagene.

En modell for  $X_{M,i}$  og  $X_{A,i}$  blir da

$$X_{M,i} = S + P_i + \epsilon_{M,i} \quad (3.5)$$

der  $E(\epsilon_{M,i}) = 0$  og

$$P_i = \alpha_{-1}I_{-1,i} + \alpha_0I_{0,i} + \alpha_1I_{1,i} + \beta_{-1}J_{-1,i} + \beta_0J_{0,i} + \beta_1J_{1,i}$$

$$X_{A,i} = -S - P_i + \epsilon_{A,i} \quad (3.6)$$

der  $E(\epsilon_{A,i}) = 0$  og

$$\begin{aligned} -P_i &= \alpha_{-1}(1 - I_{-1,i}) + \alpha_0(1 - I_{0,i}) + \alpha_1(1 - I_{1,i}) + \\ &\quad \beta_{-1}(1 - J_{-1,i}) + \beta_0(1 - J_{0,i}) + \beta_1(1 - J_{1,i}) \end{aligned}$$

Siden

$$0 = X_{M,i} + X_{A,i} = \alpha_{-1} + \alpha_0 + \alpha_1 + \beta_{-1} + \beta_0 + \beta_1 + \epsilon_{M,i} + \epsilon_{A,i}$$

og

$$E(\epsilon_{M,i}) = E(\epsilon_{A,i}) = 0$$

har vi enten

- a.  $\alpha_1 = -\alpha_{-1} - \alpha_0$  og  $\beta_1 = -\beta_{-1} - \beta_0$ , som betyr at det er ingen forskyvning av handelen mellom virkedager og helligdager.
- b.  $\beta_1 = -\alpha_{-1} - \alpha_0 - \alpha_1 - \beta_{-1} - \beta_0$ , som betyr at det er ingen forskyvning av handelen mellom de tre ukene og resten av mars og april.

$S$ ,  $\alpha_{-1}$ ,  $\alpha_0$ ,  $\alpha_1$ ,  $\beta_{-1}$ ,  $\beta_0$ ,  $\beta_1$  må estimeres i begge tilfeller. Deretter predikeres  $\hat{Y}_{M,i}^{korr}$  og  $\hat{Y}_{A,i}^{korr}$  slik

$$\begin{aligned} \hat{Y}_{M,i}^{korr} &= Y_{M,i} - \hat{P}_i \\ &= Y_{M,i} - \hat{\alpha}_{-1}I_{-1,i} - \hat{\alpha}_0I_{0,i} - \hat{\alpha}_1I_{1,i} - \hat{\beta}_{-1}J_{-1,i} - \hat{\beta}_0J_{0,i} - \hat{\beta}_1J_{1,i} \end{aligned} \quad (3.7)$$

$$\begin{aligned} \hat{Y}_{A,i}^{korr} &= Y_{A,i} + \hat{P}_i \\ &= Y_{A,i} - \hat{\alpha}_{-1}(1 - I_{-1,i}) - \hat{\alpha}_0(1 - I_{0,i}) - \hat{\alpha}_1(1 - I_{1,i}) - \\ &\quad \hat{\beta}_{-1}(1 - J_{-1,i}) - \hat{\beta}_0(1 - J_{0,i}) - \hat{\beta}_1(1 - J_{1,i}) \\ &= Y_{A,i} + \hat{\alpha}_{-1}I_{-1,i} + \hat{\alpha}_0I_{0,i} + \hat{\alpha}_1I_{1,i} + \hat{\beta}_{-1}J_{-1,i} + \hat{\beta}_0J_{0,i} + \hat{\beta}_1J_{1,i} \end{aligned} \quad (3.8)$$

### 3.4.2 Bare mars er observert i det siste året

Dvs vi har en tidsserie der den siste måneden er mars. La  $M$  betegne mars i inneværende år.  $\hat{Y}_M^{korr}$  estimeres på følgende måte. Bruk mars og april tallene inntil året før for å beregne  $X_{M,i}$ ,  $X_{A,i}$  og  $\hat{\alpha}_{-1}$ ,  $\hat{\alpha}_0$ ,  $\hat{\alpha}_1$ ,  $\hat{\beta}_{-1}$ ,  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  på samme vis som i 2.1. Den prekorrigerte verdien til mars blir da

$$\hat{Y}_M^{korr} = Y_M - \hat{\alpha}_{-1}I_{-1,M} - \hat{\alpha}_0I_{0,M} - \hat{\alpha}_1I_{1,M} - \hat{\beta}_{-1}J_{-1,M} - \hat{\beta}_0J_{0,M} - \hat{\beta}_1J_{1,M} \quad (3.9)$$

Tabell 3.1: Datoene til 1. påskedag

| ÅR   | MARS | APRIL |
|------|------|-------|
| 1979 |      | 15.   |
| 1980 |      | 6.    |
| 1981 |      | 19.   |
| 1982 |      | 11.   |
| 1983 |      | 3.    |
| 1984 |      | 22.   |
| 1985 |      | 7.    |
| 1986 | 30.  |       |
| 1987 |      | 19.   |
| 1988 |      | 3.    |
| 1989 | 26.  |       |
| 1990 |      | 15.   |
| 1991 | 31.  |       |
| 1992 |      | 19.   |
| 1993 |      | 11.   |
| 1994 |      | 3.    |
| 1995 |      | 16.   |
| 1996 |      | 7.    |
| 1997 | 30.  |       |
| 1998 |      | 12.   |
| 1999 |      | 4.    |
| 2000 |      | 23.   |

### 3.5 Eksempel. Detaljvolumindeksen

I dette eksemplet vil vi presentere beregningene av parametrene  $I_{-1}$ ,  $I_0$ ,  $I_1$ ,  $J_{-1}$ ,  $J_0$  og  $J_1$  samt deres regresjonsestimatorene  $\hat{\alpha}$  og  $\hat{\beta}$ . Vi presenterer også de sesongjusterte tallene fra X-11 ARIMA for rådata etter å ha korrigert for påskeeffekten. Dataene i eksemplet er detaljvolumindeksen for totalen (SNN52.IVL.U) fra januar 1979. Vi antar at

1. Serien er observert til og med mars 1997.
2. Serien er observert til og med april 1997.

Resultater for de to tilfellene er beskrevet nedenfor.

#### 3.5.1 Bare mars er observert i det siste året

Siden vi har tall bare til mars for 1997, vil vi bruke tallene fra mars og april måned i de årene fra 1979 til 1996 for å estimere  $\hat{\alpha}_{-1}$ ,  $\hat{\alpha}_0$ ,  $\hat{\alpha}_1$ ,  $\hat{\beta}_{-1}$ ,  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ . Deretter er  $\hat{Y}_{mars.97}^{korr}$  beregnet ved (3.9).

- Beregning av  $I_{-1,i}$ ,  $I_{0,i}$ ,  $I_{1,i}$ ,  $J_{-1,i}$ ,  $J_{0,i}$  og  $J_{1,i}$ .

For å beregne disse størrelsene må vi kjenne til datoene som 1. påskedag faller på.

Tabell 3.2: Verdiene for  $I_{(.)}$  og  $J_{(.)}$  (\* betyr at mars og april skal korrigeres)

| ÅR   | $I_{-1}$ | $I_0$ | $I_1$ | $J_{-1}$ | $J_0$ | $J_1$ |   |
|------|----------|-------|-------|----------|-------|-------|---|
| 1979 | 0        | 0     | 0     | 0        | 0     | 0     |   |
| 1980 | 1        | 1/4   | 0     | 1        | 0     | 0     | * |
| 1981 | 0        | 0     | 0     | 0        | 0     | 0     |   |
| 1982 | 3/6      | 0     | 0     | 0        | 0     | 0     | * |
| 1983 | 1        | 3/4   | 0     | 1        | 1/3   | 0     | * |
| 1984 | 0        | 0     | 0     | 0        | 0     | 0     |   |
| 1985 | 1        | 0     | 0     | 1        | 0     | 0     | * |
| 1986 | 1        | 1     | 0     | 1        | 1     | 1/2   | * |
| 1987 | 0        | 0     | 0     | 0        | 0     | 0     |   |
| 1988 | 1        | 3/4   | 0     | 1        | 1/3   | 0     | * |
| 1989 | 1        | 1     | 4/5   | 1        | 1     | 1/2   | * |
| 1990 | 0        | 0     | 0     | 0        | 0     | 0     |   |
| 1991 | 1        | 1     | 0     | 1        | 1     | 0     | * |
| 1992 | 0        | 0     | 0     | 0        | 0     | 0     |   |
| 1993 | 1/2      | 0     | 0     | 0        | 0     | 0     | * |
| 1994 | 1        | 3/4   | 0     | 1        | 1/3   | 0     | * |
| 1995 | 0        | 0     | 0     | 0        | 0     | 0     |   |
| 1996 | 1        | 0     | 0     | 1        | 0     | 0     | * |
| 1997 | 1        | 1     | 0     | 1        | 1     | 1/2   | * |
| 1998 | 1/3      | 0     | 0     | 0        | 0     | 0     | * |
| 1999 | 1        | 3/4   | 0     | 1        | 0     | 0     | * |
| 2000 | 0        | 0     | 0     | 0        | 0     | 0     |   |

Tabell 3.1 viser datoene til 1. påskedag fra 1979 til 2000. For eksempel, 1. påskedag for 1996 er 7. april.

På grunnlag av tidspunktet til 1. påskedag kan vi beregne  $I_{-1,i}$ ,  $I_{0,i}$ ,  $I_{1,i}$ ,  $J_{-1,i}$ ,  $J_{0,i}$  og  $J_{1,i}$ . Siden 1. påskedag i 1996 er 7 april, ser vi i følge figur 3.3 at

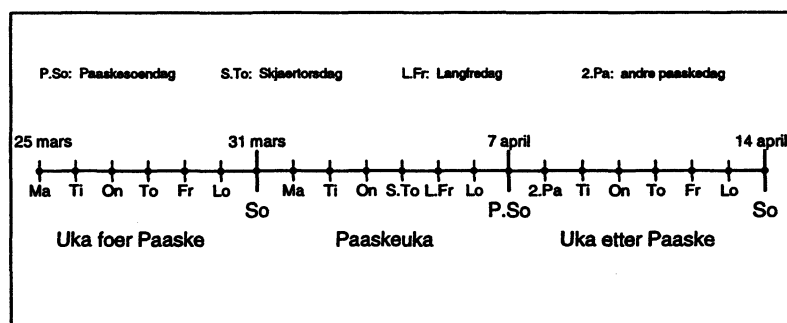
- 1 Uka før påske faller i sin helhet i mars. Dermed ved (3.3) og (3.4) er  $I_{-1} = 1$  og  $J_{-1} = 1$ .
- 2 Påskeuka faller i helhet i april, fra 1. til 7. april som gir  $I_0 = 0$  og  $J_0 = 0$ .
- 3 Uka etter påske faller i april, fra 8. til 14. april. Altså er  $I_1 = 0$  og  $J_1 = 0$ .

Tabell 3.2 viser verdiene for  $I_{(.)}$  og  $J_{(.)}$  fra 1979 til 2000. På grunn av 1. påskedager som faller seint i april blir ikke mars- og april-tallene korrigert i 8 år i perioden 1979-2000 for årene 1979, 1981, 1984, 1987, 1990, 1992, 1995 og 2000.

- Estimering av  $\hat{S}$ ,  $\hat{\alpha}$  og  $\hat{\beta}$

$S$ ,  $\alpha$  og  $\beta$  estimeres ved regresjonsanalyse under forskjellige betingelser for  $\alpha$  og  $\beta$ .

a.  $\alpha_1 = -\alpha_0 - \alpha_{-1}$  og  $\beta_1 = -\beta_0 - \beta_{-1}$



Figur 3.3: Virkedager og helligdager for uke før, etter og påskeuka i 1996

Resultater er vist i tabell 3.3. Vi ser at bare  $\hat{\beta}_{-1}$  er signifikante på 1% nivået.  $\hat{\alpha}_{-1}$  og  $\hat{\beta}_0$  er signifikante på 7% nivået. Vi får  $R^2 = 0,53$ . En tolkning er at handelen på søndager/helligdager i påskeuka går kraftig ned, mens tilsvarende er det stor handel på palmesøndager. Tabell 3.4 viser rådata, predikerte og korrigerte data for mars og april. De korrigerte dataene  $\hat{Y}_{M,i}^{korr}$  og  $\hat{Y}_{A,i}^{korr}$  brukes som utgangspunkt for videre analyse med X-11 ARIMA. Kommandoen "EASTER 6 BUILDUP 0" som beregner påskeeffekt i X-11 ARIMA må ikke brukes. Påskeeffektene ville ellers bli korrigert to ganger og vi ville få feil resultater for sesongjusterte tall i mars og april måned.  $\hat{Y}_{mars,97}^{korr}$  beregnes ved

$$\begin{aligned}\hat{Y}_{mars,97}^{korr} &= Y_{mars,97} - \hat{\alpha}_{-1}I_{-1,97} - \hat{\alpha}_0I_{0,97} - \hat{\alpha}_1I_{1,97} - \hat{\beta}_{-1}J_{-1,97} - \hat{\beta}_0J_{0,97} - \\ &\quad \hat{\beta}_1J_{1,97} \\ &= 92,99 - (-5,96 \times 1 + 3,61 \times 1 + 2,34 \times 0 + 8,64 \times 1 - 7,97 \times 1 - \\ &\quad 0,67 \times 0.5) \\ &= 94,99\end{aligned}$$

Tabell 3.3: Estimering av parametre med betingelse (a) når bare mars 1997 er observert

|                     | df | Estimert | std   | t for $H_0$ | Prob >  t |
|---------------------|----|----------|-------|-------------|-----------|
| $\hat{S}$           | 1  | 0,194    | 0,795 | 0,245       | 0,810     |
| $\hat{\alpha}_{-1}$ | 1  | -5,957   | 2,997 | -1,987      | 0,068     |
| $\hat{\alpha}_0$    | 1  | 3,613    | 3,315 | 1,090       | 0,295     |
| $\hat{\alpha}_1$    | 1  | 2,344    | 3,177 | 0,738       | 0,473     |
| $\hat{\beta}_{-1}$  | 1  | 8,640    | 2,692 | 3,209       | 0,006     |
| $\hat{\beta}_0$     | 1  | -7,969   | 3,902 | -2,042      | 0,062     |
| $\hat{\beta}_1$     | 1  | -0,671   | 4,019 | -0,167      | 0,870     |

$$R^2 = 0,530$$

$$F = 3,702 \text{ (signifikant på 3\% nivået)}$$

Tabell 3.4: Observerte, predikerte og korrigerte tall for mars og april med betingelsen (a) når bare mars 1997 er observert

| År | MARS      |                        |                        | APRIL     |                        |                        |
|----|-----------|------------------------|------------------------|-----------|------------------------|------------------------|
|    | $Y_{M,i}$ | $\hat{Y}_{M,i}^{Pred}$ | $\hat{Y}_{M,i}^{korr}$ | $Y_{A,i}$ | $\hat{Y}_{A,i}^{Pred}$ | $\hat{Y}_{A,i}^{korr}$ |
| 79 | 85,15     | 83,5699                | 85,1500                | 81,60     | 83,1801                | 81,6000                |
| 80 | 84,65     | 86,4359                | 81,0640                | 80,66     | 78,8741                | 84,2460                |
| 81 | 81,06     | 83,2699                | 81,0600                | 85,09     | 82,8801                | 85,0900                |
| 82 | 82,76     | 81,7259                | 85,7390                | 86,26     | 87,2941                | 83,2810                |
| 83 | 85,08     | 83,9512                | 82,3437                | 76,96     | 78,0888                | 79,6963                |
| 84 | 82,44     | 81,5049                | 82,4400                | 80,18     | 81,1151                | 80,1800                |
| 85 | 83,05     | 84,8175                | 80,3674                | 80,83     | 79,0625                | 83,5126                |
| 86 | 81,39     | 84,8061                | 83,3988                | 91,85     | 88,4339                | 89,8412                |
| 87 | 80,39     | 83,8849                | 80,3900                | 86,99     | 83,4951                | 86,9900                |
| 88 | 87,63     | 86,2362                | 84,8937                | 78,98     | 80,3738                | 81,7163                |
| 89 | 80,27     | 80,7416                | 80,4033                | 81,09     | 80,6184                | 80,9567                |
| 90 | 85,11     | 83,2899                | 85,1100                | 81,08     | 82,9001                | 81,0800                |
| 91 | 82,47     | 82,1366                | 84,1433                | 84,76     | 85,0934                | 83,0867                |
| 92 | 82,74     | 84,4299                | 82,7400                | 85,73     | 84,0401                | 85,7300                |
| 93 | 85,15     | 83,0509                | 88,1290                | 86,52     | 88,6191                | 83,5410                |
| 94 | 94,56     | 92,2512                | 91,8237                | 84,08     | 86,3888                | 86,8163                |
| 95 | 91,68     | 89,8099                | 91,6800                | 87,55     | 89,4201                | 87,5500                |
| 96 | 96,14     | 95,8075                | 93,4574                | 89,72     | 90,0525                | 92,4026                |
| 97 | 92,99     | .                      | 94,9900                | .         | .                      | .                      |

Tabell 3.5: Estimering av parametre med betingelsen (b) når bare mars 1997 er observert

|                     | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|---------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$           | 1  | -0,136   | 0,686 | -0,198        | 0,846        |
| $\hat{\alpha}_{-1}$ | 1  | -2,162   | 2,963 | -0,730        | 0,479        |
| $\hat{\alpha}_0$    | 1  | 8,965    | 3,540 | 2,532         | 0,026        |
| $\hat{\alpha}_1$    | 1  | 6,729    | 3,218 | 2,091         | 0,058        |
| $\hat{\beta}_{-1}$  | 1  | 3,903    | 2,974 | 1,313         | 0,213        |
| $\hat{\beta}_0$     | 1  | -12,473  | 3,769 | -3,309        | 0,006        |
| $\hat{\beta}_1$     | 1  | -4,963   | 3,817 | -1,300        | 0,217        |

$$R^2 = 0,690$$

$$F = 5,363 \text{ (signifikant på 1\% nivået)}$$

b.  $\beta_1 = -\alpha_{-1} - \alpha_0 - \alpha_1 - \beta_{-1} - \beta_0$ .

Bruken av regresjonsanalyse med denne betingelsen gir resultatene som er vist i tabell 3.5. Vi ser at  $\hat{\alpha}_0$ ,  $\hat{\alpha}_1$ ,  $\hat{\beta}_0$  er signifikant på 5% nivået med  $R^2 = 0,69$ . Tolkningen fra tabell 3.5 er at en kraftig forskyvning av handel fra helligdagene i påskeuka til virke-

Tabell 3.6: *Observerte, predikerte og korrigerte tall for mars og april med betingelsen (b) når bare mars 1997 er observert*

| År   | MARS      |                        |                        | APRIL     |                        |                        |
|------|-----------|------------------------|------------------------|-----------|------------------------|------------------------|
|      | $Y_{M,i}$ | $\hat{Y}_{M,i}^{Pred}$ | $\hat{Y}_{M,i}^{korr}$ | $Y_{A,i}$ | $\hat{Y}_{A,i}^{Pred}$ | $\hat{Y}_{A,i}^{korr}$ |
| 1979 | 85,15     | 83,238                 | 85,150                 | 81,60     | 83,511                 | 81,600                 |
| 1980 | 84,65     | 86,501                 | 80,667                 | 80,66     | 78,808                 | 84,642                 |
| 1981 | 81,06     | 82,938                 | 81,060                 | 85,09     | 83,211                 | 85,090                 |
| 1982 | 82,76     | 83,292                 | 83,841                 | 86,26     | 85,727                 | 85,178                 |
| 1983 | 85,08     | 85,191                 | 80,772                 | 76,96     | 76,848                 | 81,267                 |
| 1984 | 82,44     | 81,173                 | 82,440                 | 80,18     | 81,446                 | 80,180                 |
| 1985 | 83,05     | 83,544                 | 81,309                 | 80,83     | 80,335                 | 82,570                 |
| 1986 | 81,39     | 82,235                 | 85,638                 | 91,85     | 91,004                 | 87,601                 |
| 1987 | 80,39     | 83,553                 | 80,390                 | 86,99     | 83,826                 | 86,990                 |
| 1988 | 87,63     | 87,476                 | 83,322                 | 78,98     | 79,133                 | 83,287                 |
| 1989 | 80,27     | 81,679                 | 79,134                 | 81,09     | 79,680                 | 82,225                 |
| 1990 | 85,11     | 82,958                 | 85,110                 | 81,08     | 83,231                 | 81,080                 |
| 1991 | 82,47     | 81,712                 | 84,236                 | 84,76     | 85,517                 | 82,993                 |
| 1992 | 82,74     | 84,098                 | 82,740                 | 85,73     | 84,371                 | 85,730                 |
| 1993 | 85,15     | 84,617                 | 86,231                 | 86,52     | 87,052                 | 85,438                 |
| 1994 | 94,56     | 93,491                 | 90,252                 | 84,08     | 85,148                 | 88,387                 |
| 1995 | 91,68     | 89,478                 | 91,680                 | 87,55     | 89,751                 | 87,550                 |
| 1996 | 96,14     | 94,534                 | 94,399                 | 89,72     | 91,325                 | 91,460                 |
| 1997 | 92,99     | .                      | 97,240                 | .         | .                      | .                      |

dagene i påskeuka og uka etter påske. De observerte, predikerte og estimerte verdiene for mars og april måneder er vist i tabell 3.6.  $\hat{Y}_{mars,97}^{korr}$  beregnes ved

$$\begin{aligned}
 \hat{Y}_{mars,97}^{korr} &= Y_{mars,97} - \hat{\alpha}_{-1}I_{-1,97} - \hat{\alpha}_0I_{0,97} - \hat{\alpha}_1I_{1,97} - \hat{\beta}_{-1}J_{-1,97} - \hat{\beta}_0J_{0,97} - \\
 &\quad \hat{\beta}_1J_{1,97} \\
 &= 92,99 - (-2,16 \times 1 + 8,97 \times 1 + 6,73 \times 0 + 3,90 \times 1 - 12,47 \times 1 - \\
 &\quad 4,96 \times 0.5) \\
 &= 97,24
 \end{aligned}$$

### - Konklusjon

Vi har presentert resultater for påskekorrigerings under to sett av betingelser:

$$(a) \alpha_1 = -\alpha_{-1} - \alpha_0 \text{ og } \beta_1 = -\beta_{-1} - \beta_0$$

$$(b) \beta_1 = -\alpha_{-1} - \alpha_0 - \alpha_1 - \beta_{-1} - \beta_0$$

Tabell 3.3 og 3.5 viser at (b) gir en bedre tilpasning til observerte data og også en tolkning som er mer overbevisende.

### 3.5.2 Både mars og april er observert i det siste året

Vi observerer nå til og med april 1997 med  $Y_{april,97} = 100,8$ . Regresjonsanalyse kjøres med den betingelsen  $\beta_1 = -\alpha_{-1} - \alpha_0 - \alpha_1 - \beta_{-1} - \beta_0$  siden vi får en bedre tilpasning til observerte data. Resultater er vist i tabell 3.7. Vi ser at modellen er signifikant på 1% nivået med  $R^2 = 72,89$ .  $\hat{\alpha}_0$ ,  $\hat{\alpha}_1$  og  $\hat{\beta}_0$  er signifikant på under 5% nivået. Tolkningen er at en kraftig forskyvning av handel fra helligdagene i påskeuka til virkedagene i påskeuka og uka etter påske. Observerte, predikerte og korrigerte tall for mars og april fra 1979 til 1997 er skrevet ut i tabell 3.8.  $\hat{Y}_{mars,97}^{korr}$  og  $\hat{Y}_{april,97}^{korr}$  beregnes ved å sette inn i (3.7) og (3.8).

$$\begin{aligned}\hat{Y}_{mars,97}^{korr} &= Y_{mars,97} - \hat{\alpha}_{-1}I_{-1,97} - \hat{\alpha}_0I_{0,97} - \hat{\alpha}_1I_{1,97} - \hat{\beta}_{-1}J_{-1,97} - \hat{\beta}_0J_{0,97} - \\ &\quad \hat{\beta}_1J_{1,97} \\ &= 92,99 - (-2,18 \times 1 + 8,92 \times 1 + 6,36 \times 0 + 3,91 \times 1 - 12,37 \times 1 - \\ &\quad 4,65 \times 0.5) \\ &= 97,03\end{aligned}$$

$$\begin{aligned}\hat{Y}_{april,97}^{korr} &= Y_{april,97} + \hat{\alpha}_{-1}I_{-1,97} + \hat{\alpha}_0I_{0,97} + \hat{\alpha}_1I_{1,97} + \hat{\beta}_{-1}J_{-1,97} + \hat{\beta}_0J_{0,97} + \\ &\quad \hat{\beta}_1J_{1,97} \\ &= 100,8 + (-2,18 \times 1 + 8,92 \times 1 + 6,36 \times 0 + 3,91 \times 1 - 12,37 \times 1 - \\ &\quad 4,65 \times 0.5) \\ &= 96,76\end{aligned}$$

Tabell 3.7: Estimering av parametre med betingelsen (b) når både mars og april 1997 er observert

|                     | df | Estimert | std   | t for $H_0$ | Prob >  t |
|---------------------|----|----------|-------|-------------|-----------|
| $\hat{S}$           | 1  | -0,127   | 0,658 | -0,193      | 0,849     |
| $\hat{\alpha}_{-1}$ | 1  | -2,180   | 2,850 | -0,765      | 0,458     |
| $\hat{\alpha}_0$    | 1  | 8,918    | 3,398 | 2,624       | 0,021     |
| $\hat{\alpha}_1$    | 1  | 6,363    | 2,488 | 2,557       | 0,023     |
| $\hat{\beta}_{-1}$  | 1  | 3,914    | 2,861 | 1,368       | 0,194     |
| $\hat{\beta}_0$     | 1  | -12,369  | 3,589 | -3,446      | 0,004     |
| $\hat{\beta}_1$     | 1  | -4,647   | 3,310 | -1,404      | 0,183     |

$$R^2 = 0,7289$$

$$F = 6,992 \text{ (signifikant på 1\% nivået)}$$

### 3.5.3 En oppsummering for 5.1 og 5.2

Vi ser under betingelsen  $\beta_1 = -\alpha_{-1} - \alpha_0 - \alpha_1 - \beta_{-1} - \beta_0$  at

- I avsnitt 5.1 (tidsserien er observert til og med mars 1997)

Tabell 3.8: Observerte, predikerte og korrigerte tall for mars og april med betingelsen (b) når både mars og april 1997 er observert

| År   | MARS      |                        |                        | APRIL     |                        |                        |
|------|-----------|------------------------|------------------------|-----------|------------------------|------------------------|
|      | $Y_{M,i}$ | $\hat{Y}_{M,i}^{Pred}$ | $\hat{Y}_{M,i}^{korr}$ | $Y_{A,i}$ | $\hat{Y}_{A,i}^{Pred}$ | $\hat{Y}_{A,i}^{korr}$ |
| 1979 | 85,15     | 83,247                 | 85,150                 | 81,60     | 83,502                 | 81,600                 |
| 1980 | 84,65     | 86,492                 | 80,685                 | 80,66     | 78,818                 | 84,624                 |
| 1981 | 81,06     | 82,947                 | 81,060                 | 85,09     | 83,202                 | 85,090                 |
| 1982 | 82,76     | 83,292                 | 83,850                 | 86,26     | 85,728                 | 85,169                 |
| 1983 | 85,08     | 85,193                 | 80,779                 | 76,96     | 76,847                 | 81,260                 |
| 1984 | 82,44     | 81,182                 | 82,440                 | 80,18     | 81,437                 | 80,180                 |
| 1985 | 83,05     | 83,547                 | 81,315                 | 80,83     | 80,333                 | 82,564                 |
| 1986 | 81,39     | 82,452                 | 85,430                 | 91,85     | 90,787                 | 87,810                 |
| 1987 | 80,39     | 83,562                 | 80,390                 | 86,99     | 83,817                 | 86,990                 |
| 1988 | 87,63     | 87,478                 | 83,329                 | 78,98     | 79,132                 | 83,280                 |
| 1989 | 80,27     | 81,603                 | 79,218                 | 81,09     | 79,756                 | 82,141                 |
| 1990 | 85,11     | 82,967                 | 85,110                 | 81,08     | 83,222                 | 81,080                 |
| 1991 | 82,47     | 81,771                 | 84,186                 | 84,76     | 85,459                 | 83,043                 |
| 1992 | 82,74     | 84,107                 | 82,740                 | 85,73     | 84,362                 | 85,730                 |
| 1993 | 85,15     | 84,617                 | 86,240                 | 86,52     | 87,053                 | 85,429                 |
| 1994 | 94,56     | 93,493                 | 90,259                 | 84,08     | 85,147                 | 88,380                 |
| 1995 | 91,68     | 89,487                 | 91,680                 | 87,55     | 89,742                 | 87,550                 |
| 1996 | 96,14     | 94,537                 | 94,405                 | 89,72     | 91,323                 | 91,454                 |
| 1997 | 92,99     | 92,727                 | 97,030                 | 100,80    | 101,062                | 96,760                 |

(i) Dataene fra 1979 til 1996 brukes for å beregne  $X_{M,i}$ ,  $X_{A,i}$ ,  $\hat{\alpha}_{(.)}$  og  $\hat{\beta}_{(.)}$ . Vi får  $F = 5,363$  og  $R^2 = 0,69$ .

(ii)  $\hat{Y}_{mars,97}^{korr}$  beregnes ved ligning (3.9) med  $\hat{\alpha}_{(.)}$  og  $\hat{\beta}_{(.)}$  fra (i) og  $I_{(.)}$ ,  $J_{(.)}$  for 1997. Vi får  $\hat{Y}_{mars,97}^{korr} = 97,24$  med  $Y_{mars,97} = 92,99$  (se tabell 3.6).

- I avsnitt 5.2 (både mars og april 1997 er observert)

(i)  $X_{M,i}$ ,  $X_{A,i}$ ,  $\hat{\alpha}_{(.)}$  og  $\hat{\beta}_{(.)}$  beregnes med dataene fra 1979 til 1997 (dvs en observasjon mer). Vi får  $F = 6,992$  og  $R^2 = 0,73$ .

(ii)  $\hat{Y}_{mars,97}^{korr}$  og  $\hat{Y}_{april,97}^{korr}$  beregnes ved ligningene (3.7) og (3.8). Vi får  $\hat{Y}_{mars,97}^{korr} = 97,03$  med  $Y_{mars,97} = 92,99$  og  $\hat{Y}_{april,97}^{korr} = 96,76$  med  $Y_{april,97} = 100,8$  (se tabell 3.8).

Vi ser i de to tabellene 3.6 og 3.8 at  $\hat{Y}_{mars,i}^{korr}$  og  $\hat{Y}_{april,i}^{korr}$  endrer seg lite når april 1997 legges til.

### 3.5.4 Sammenligning med påskekorrigering og sesongjusterte tall ved X-11 ARIMA

Rådataene fra januar 1979 til 1997 er kjørt ved X-11 ARIMA. Vi får følgende resultater:

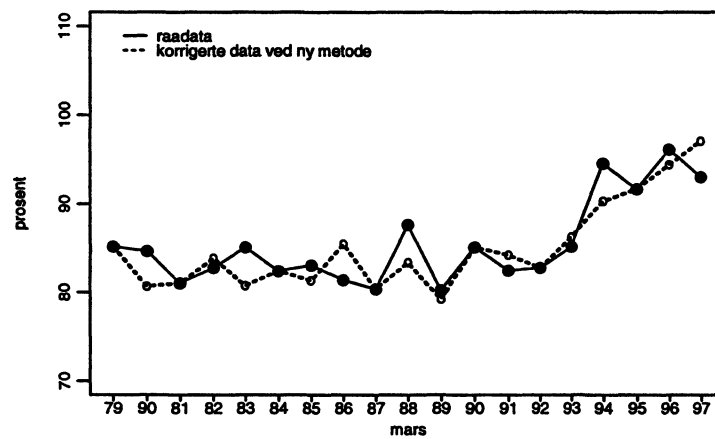
- $\hat{d} = 1$ , hvor  $\hat{d}$  er antall dager før påskesøndag som rådata er påvirket av påskeeffekt. Siden  $\hat{d} = 1$  inneholder påskedagene bare lørdag og 1. påskedag. Med den nye metoden antar vi at påskeeffekten kan variere i tre uker. Det er bare 4 av 19 år som X-11 ARIMA korrigerer for påskeeffektene nemlig for 1986, 1989, 1991 og 1997. Ved den nye metoden korrigeres tallene for 12 år 1980, 1982, 1983, 1985, 1986, 1988, 1989, 1991, 1993, 1994, 1996 og 1997.
- Testen for påskeeffekten ved X-11 ARIMA er signifikant på 1% nivået med  $F = 9,690$ . Vi har også en tilsvarende test. Under betingelsen (a) er  $F = 3,702$  som er signifikant på 3% nivået. Mens under betingelsen (b) er  $F = 5,363$ , som er signifikant på 1% nivået.
- Korrigerte verdiene ved den nye metoden er relative glattere enn ved X-11 ARIMA, særlig i de siste årene. Dette vises i tabell 3.9 og figur 3.8 og 3.9.
- Sesongjusterte tallene for mars og april er også relativt glattere ved den nye metoden enn ved X-11 ARIMA, se figurene 3.10 og 3.11.

Tabell 3.9: Observerte og korrigerte tall for mars og april med betingelsen (b) når både mars og april 1997 er observert, ved den nye metoden og X-11 ARIMA. De blanke feltene i tabellen er observerte verdier

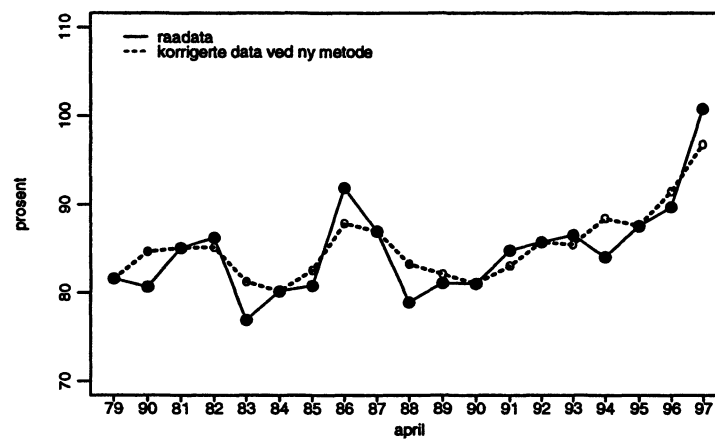
| År   | 1. påskedag | MARS      |                          |                          | APRIL     |                          |                          |
|------|-------------|-----------|--------------------------|--------------------------|-----------|--------------------------|--------------------------|
|      |             | $Y_{M,i}$ | $\hat{Y}_{M,i}^{Reg,k.}$ | $\hat{Y}_{M,i}^{X11,k.}$ | $Y_{A,i}$ | $\hat{Y}_{A,i}^{Reg,k.}$ | $\hat{Y}_{A,i}^{X11,k.}$ |
| 1979 | 15 april    | 85,15     |                          |                          | 81,60     |                          |                          |
| 1980 | 6 april     | 84,65     | 80,68                    |                          | 80,66     | 84,62                    |                          |
| 1981 | 19 april    | 81,06     |                          |                          | 85,09     |                          |                          |
| 1982 | 11 april    | 82,76     | 83,85                    |                          | 86,26     | 85,16                    |                          |
| 1983 | 3 april     | 85,08     | 80,77                    |                          | 76,96     | 81,26                    |                          |
| 1984 | 22 april    | 82,44     |                          |                          | 80,18     |                          |                          |
| 1985 | 7 april     | 83,05     | 81,31                    |                          | 80,83     | 82,56                    |                          |
| 1986 | 30 mars     | 81,39     | 85,43                    | 84,82                    | 91,85     | 87,81                    | 88,28                    |
| 1987 | 19 april    | 80,39     |                          |                          | 86,99     |                          |                          |
| 1988 | 3 april     | 87,63     | 83,32                    |                          | 78,98     | 83,28                    |                          |
| 1989 | 26 mars     | 80,27     | 79,21                    | 83,65                    | 81,09     | 82,14                    | 77,94                    |
| 1990 | 15 april    | 85,11     |                          |                          | 81,08     |                          |                          |
| 1991 | 31 mars     | 82,47     | 84,18                    | 85,94                    | 84,76     | 83,04                    | 81,47                    |
| 1992 | 19 april    | 82,74     |                          |                          | 85,73     |                          |                          |
| 1993 | 11 april    | 85,15     | 86,24                    |                          | 86,52     | 85,42                    |                          |
| 1994 | 3 april     | 94,56     | 90,25                    |                          | 84,08     | 88,38                    |                          |
| 1995 | 16 april    | 91,68     |                          |                          | 87,55     |                          |                          |
| 1996 | 7 april     | 96,14     | 94,40                    |                          | 89,72     | 91,45                    |                          |
| 1997 | 30 mars     | 92,99     | 97,03                    | 96,91                    | 100,80    | 96,76                    | 96,88                    |

### 3.6 Oppsummering

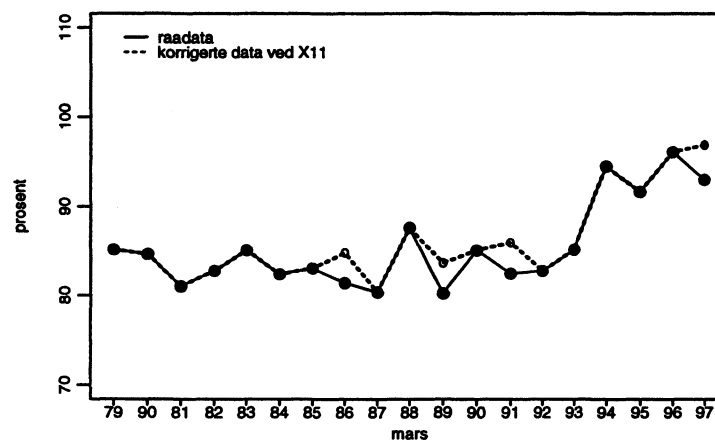
Metoden har vært brukt i Statistisk sentralbyrå for detaljvolumindeksen siden 1995. Den gir gode resultater for detaljvolumindeksen men vi har foreløpig ikke brukt



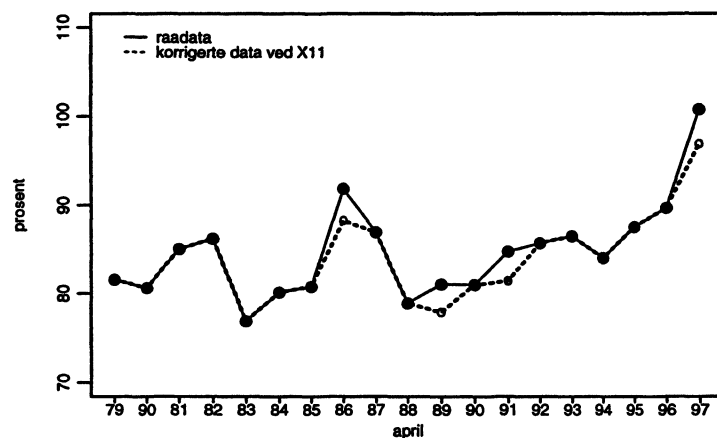
Figur 3.4: Observerte og korrigerende verdier for mars ved tabell 3.8



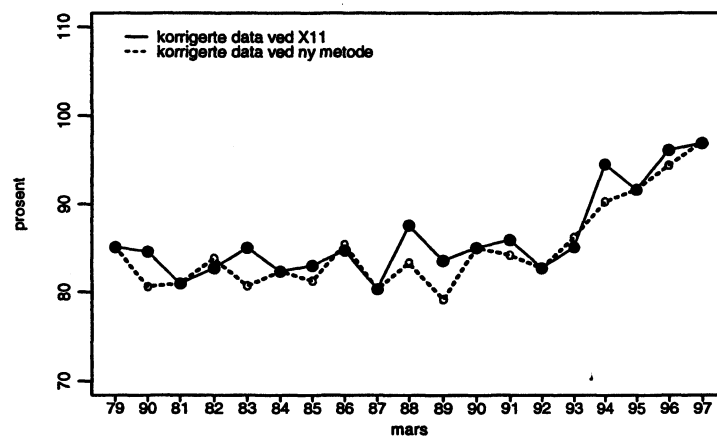
Figur 3.5: Observerte og korrigerende verdier for april ved tabell 3.8



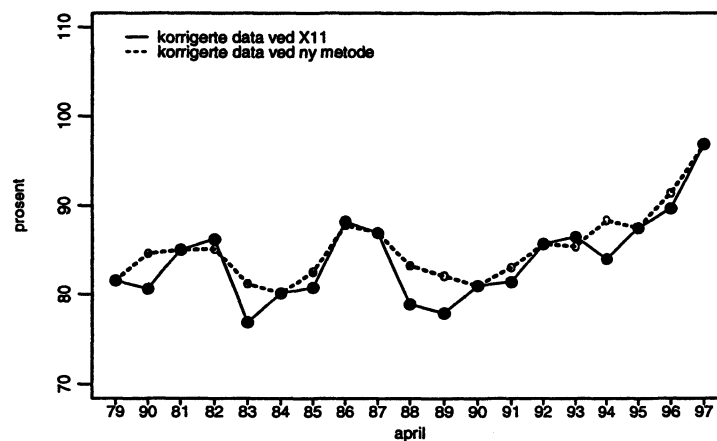
Figur 3.6: Observerte og korrigerende verdier for mars ved X-11 ARIMA



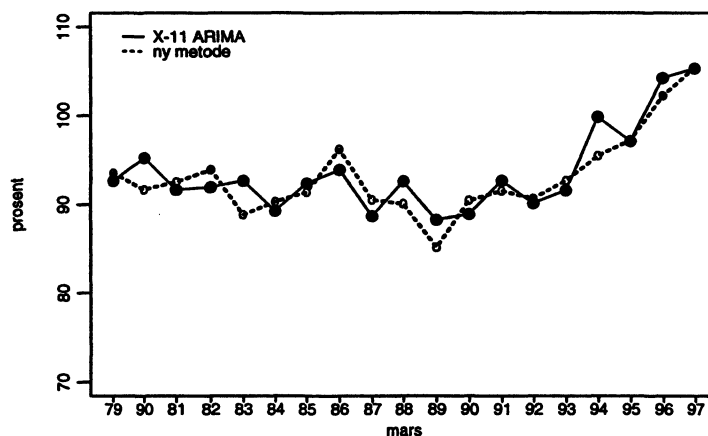
Figur 3.7: *Observerte og korrigerede verdier for april ved X-11 ARIMA*



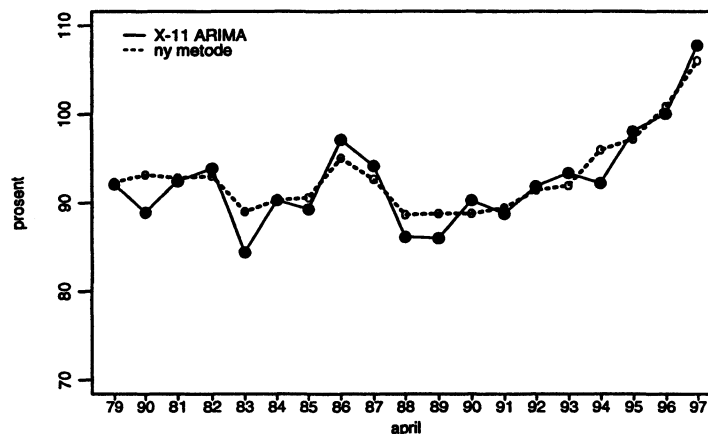
Figur 3.8: *Korrigerede verdier for mars ved ny metode og X-11 ARIMA*



Figur 3.9: *Korrigerede verdier for april ved ny metode og X-11 ARIMA*



Figur 3.10: Sesongjusterte tall for mars ved ny metode og X-11 ARIMA



Figur 3.11: Sesongjusterte tall for april ved ny metode og X-11 ARIMA

metoden på andre serier. Antakelig er ikke betingelsen om at påske kun påvirker uka før, påskeuka og uka etter påske gyldig for andre serier uten videre. Allikevel mener vi at denne type modellen for å prekorrigere tallene er nyttig for andre statistikk områder.

### 3.7 Programmer

Det er to programmer på UNIX som brukes for påskekorrigeringen. De er

- **k\_mars** når den siste observasjonen er mars.
- **k\_april** når den siste observasjonen ikke er mars.

De to programmene kan kjøres med en eller flere serier samtidige. Disse seriene må ligge i en ASCII fil uten tekster, hvor hver serie er en kolonne. Vi starter programmet ved å skrive

## **k\_mars** eller **k\_april**

som er avhengig av tidspunktet av den siste observasjonen. For å gjøre det lettere å forklare programmene vil vi bruke et eksempel.

- **Eksempel:** Rådataene av de to seriene "SNN52.IVL.U" og "SNN521.IVL.U" skal korrigeres for påskeeffekter med den nye metoden før de kjøres med X-11 ARIMA. De to seriene ligger i en ASCII fil med navn "volum.dat". Dataene er observert fra januar 1979 til mars 1997. Siden den siste observasjonen er mars vil vi bruke **k\_mars** for beregningen. Vi taster inn **k\_mars**. Det kommer da følgende spørsmålene:

- Angi det første året i serien (med 4 siffer) < enter >  
1979 < enter >
- Angi antall serier < enter >  
2 < enter >
- Angi innfil < enter >  
volum.dat < enter >
- Angi navn for 1. filen < enter >  
SNN52.IVL.U < enter >
- Angi navn for 2. filen < enter >  
SNN521.IVL.U < enter >

Programmet skal

- (i) beregne  $\hat{Y}_{M,i}^{korr}$  og  $\hat{Y}_{A,i}^{korr}$  for hvert år  $i$ .
- (ii) kjøre X-11 ARIMA med de korrigererte tall fra (i)

Vi får:

1. En fil uten tekster med navn **Tab\_mars** som inneholder 7 kolonner (se tabell 3.10). Den første kolonnen er årstall med ":" som skiller år og måned. De tre neste kolonnene står for prekorrigerte tall for påskeeffekt, prekorrigerte tall for ukedagseffekter og sesongjusterte tall fra X-11 ARIMA for serien SNN52.IVL.U. De tre neste kolonnene tilhører den serien SNN521.IVL.U.
2. To filer med navn **SNN52.IVL.U.lst** og **SNN521.IVL.U.lst** som inneholder  $\hat{\alpha}_{-1}$ ,  $\hat{\alpha}_0$ ,  $\hat{\alpha}_1$ ,  $\hat{\beta}_{-1}$ ,  $\hat{\beta}_0$  og  $\hat{\beta}_1$  og korrigererte tallene for mars og april i alle årene, for hver serie.

Det er en annen måte å kjøre **k\_mars**. Vi lager en fil uten forklarende tekst (for eksempel, filen heter "volum1") med de følgende linjene:

|              |                             |
|--------------|-----------------------------|
| 1979         | (det første året i serien)  |
| 2            | (antall serier)             |
| volum.dat    | (navn på innfil)            |
| SNN52.IVL.U  | (navn til den første filen) |
| SNN521.IVL.U | (navn til den andre filen)  |

Vi skrives

**k\_mars** < **volum1** og < enter >

Tabell 3.10: *Prekorrigerte for påskeeffekt, ukedagseffekter og sesongjusterte tall for SNN52 og SNN521.*

| År      | SNN52   |         |         | SNN521  |         |         |
|---------|---------|---------|---------|---------|---------|---------|
|         | P. Pås. | P. Uke. | Justert | P. Pås. | P. Uke. | Justert |
| 1979:1  | 83,43   | 83,48   | 90,89   | 76,97   | 77,08   | 90,50   |
| 1979:2  | 76,33   | 77,01   | 90,83   | 76,82   | 77,51   | 89,41   |
| 1979:3  | 85,15   | 83,64   | 93,55   | 90,21   | 88,24   | 93,37   |
| 1996:1  | 91,03   | 91,09   | 100,81  | 87,48   | 87,60   | 99,77   |
| 1996:2  | 88,71   | 85,41   | 101,19  | 92,00   | 88,88   | 99,95   |
| 1996:3  | 94,40   | 95,39   | 101,89  | 98,41   | 99,04   | 101,44  |
| 1996:4  | 91,46   | 91,67   | 101,15  | 99,64   | 99,71   | 101,42  |
| 1996:5  | 98,21   | 97,01   | 101,16  | 101,01  | 99,82   | 100,37  |
| 1996:6  | 102,41  | 103,50  | 102,99  | 102,31  | 103,51  | 100,63  |
| 1996:7  | 103,50  | 103,57  | 104,03  | 108,74  | 108,90  | 102,88  |
| 1996:8  | 107,06  | 105,16  | 103,24  | 107,21  | 104,87  | 103,05  |
| 1996:9  | 99,06   | 101,38  | 102,91  | 94,77   | 96,99   | 100,03  |
| 1996:10 | 107,44  | 105,84  | 104,37  | 100,19  | 99,43   | 101,76  |
| 1996:11 | 110,22  | 109,42  | 106,91  | 104,42  | 102,88  | 106,12  |
| 1996:12 | 144,14  | 147,03  | 104,31  | 125,59  | 128,36  | 101,98  |
| 1997:1  | 94,85   | 93,69   | 103,87  | 90,24   | 89,18   | 101,74  |
| 1997:2  | 86,48   | 87,25   | 103,54  | 88,29   | 89,08   | 100,29  |
| 1997:3  | 97,24   | 98,71   | 105,29  | 98,52   | 99,77   | 102,08  |

P. Pås. Prekorrigerte tall for påskeeffekt

P. Uke. Prekorrigerte tall for ukedagseffekter

Justert Sesongjusterte tall

Vi får de samme resultatene. En fordel er at vi kan unngå å svare feil på spørsmålene og starte programmet på nytt. Det tar ca 2 minutter for en kjøring med 16 serier.

### 3.8 Anvendelse metoden for de andre seriene i detaljvolumindeksen

#### 3.8.1 SNN521.IVL.U: Butikkhandel med bredt vareutvalg. I alt

Estimatene for betingelsen (b) når vi observerer til og med mars 1997 og til og med april 1997 er vist i tabellene 3.11 og 3.12.

Tabell 3.11: *Estimering av parametre når bare mars 1997 er observert*

| Estimat             | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|---------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$           | 1  | -0,680   | 0,652 | -1,044        | 0,3172       |
| $\hat{\alpha}_{-1}$ | 1  | -3,668   | 2,816 | -1,303        | 0,2172       |
| $\hat{\alpha}_0$    | 1  | 11,994   | 3,365 | 3,565         | 0,0039       |
| $\hat{\alpha}_1$    | 1  | 3,678    | 3,059 | 1,202         | 0,2524       |
| $\hat{\beta}_{-1}$  | 1  | 6,682    | 2,826 | 2,364         | 0,0358       |
| $\hat{\beta}_0$     | 1  | -11,756  | 3,582 | -3,281        | 0,0066       |
| $\hat{\beta}_1$     | 1  | -6,930   | 3,628 | -1,910        | 0,0803       |

$$R^2 = 0,8381$$

$$F = 12,425 \text{ (signifikant på 1\% nivået)}$$

Tabell 3.12: *Estimering av parametre når både mars og april 1997 er observert*

| Estimat             | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|---------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$           | 1  | -0,654   | 0,634 | -1,031        | 0,3214       |
| $\hat{\alpha}_{-1}$ | 1  | -3,721   | 2,745 | -1,355        | 0,1985       |
| $\hat{\alpha}_0$    | 1  | 11,852   | 3,274 | 3,620         | 0,0031       |
| $\hat{\alpha}_1$    | 1  | 2,571    | 2,397 | 1,073         | 0,3029       |
| $\hat{\beta}_{-1}$  | 1  | 6,716    | 2,756 | 2,437         | 0,0300       |
| $\hat{\beta}_0$     | 1  | -11,443  | 3,458 | -3,309        | 0,0057       |
| $\hat{\beta}_1$     | 1  | -5,975   | 3,189 | -1,873        | 0,0837       |

$$R^2 = 0,8337$$

$$F = 13,033 \text{ (signifikant på 1\% nivået)}$$

### 3.8.2 SNN5211.IVL.U: Butikkhandel med bredt vareutvalg. Hovedvekt på nærings og nytelsesmidler

Estimatene for betingelsen (b) når vi observerer til og med mars 1997 og til og med april 1997 er vist i tabellene 3.13 og 3.14.

Tabell 3.13: *Estimering av parametre når bare mars 1997 er observert*

| Estimat             | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|---------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$           | 1  | -0,581   | 0,669 | -0,869        | 0,4019       |
| $\hat{\alpha}_{-1}$ | 1  | -3,826   | 2,891 | -1,323        | 0,2104       |
| $\hat{\alpha}_0$    | 1  | 12,186   | 3,454 | 3,527         | 0,0042       |
| $\hat{\alpha}_1$    | 1  | 3,442    | 3,141 | 1,096         | 0,2946       |
| $\hat{\beta}_{-1}$  | 1  | 7,101    | 2,902 | 2,447         | 0,0308       |
| $\hat{\beta}_0$     | 1  | -11,637  | 3,678 | -3,163        | 0,0082       |
| $\hat{\beta}_1$     | 1  | -7,267   | 3,725 | -1,951        | 0,0748       |

$$R^2 = 0,8448$$

$$F = 13,061 \text{ (signifikant på 1\% nivået)}$$

Tabell 3.14: *Estimering av parametre når både mars og april 1997 er observert*

| Estimat             | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|---------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$           | 1  | -0,561   | 0,647 | -0,866        | 0,4021       |
| $\hat{\alpha}_{-1}$ | 1  | -3,867   | 2,802 | -1,380        | 0,1908       |
| $\hat{\alpha}_0$    | 1  | 12,073   | 3,341 | 3,613         | 0,0032       |
| $\hat{\alpha}_1$    | 1  | 2,567    | 2,446 | 1,049         | 0,3131       |
| $\hat{\beta}_{-1}$  | 1  | 7,128    | 2,813 | 2,534         | 0,0249       |
| $\hat{\beta}_0$     | 1  | -11,389  | 3,529 | -3,227        | 0,0066       |
| $\hat{\beta}_1$     | 1  | -6,512   | 3,255 | -2,001        | 0,0668       |

$$R^2 = 0,8425$$

$$F = 13,097 \text{ (signifikant på 1\% nivået)}$$

### 3.8.3 SNN5212.IVL.U: Butikkhandel med bredt vareutvalg. Bredt vareutvalg ellers

Estimatene for betingelsen (b) når vi observerer til og med mars 1997 og til og med april 1997 er vist i tabellene 3.15 og 3.16.

Tabell 3.15: *Estimering av parametre når bare mars 1997 er observert*

| Estimat             | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|---------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$           | 1  | -2,008   | 1,050 | -1,912        | 0,0800       |
| $\hat{\alpha}_{-1}$ | 1  | -1,467   | 4,536 | -0,324        | 0,7518       |
| $\hat{\alpha}_0$    | 1  | 11,423   | 5,420 | 2,108         | 0,0568       |
| $\hat{\alpha}_1$    | 1  | 5,781    | 4,927 | 1,173         | 0,2634       |
| $\hat{\beta}_{-1}$  | 1  | 1,206    | 4,552 | 0,265         | 0,7954       |
| $\hat{\beta}_0$     | 1  | -14,748  | 5,771 | -2,556        | 0,0252       |
| $\hat{\beta}_1$     | 1  | -2,196   | 5,843 | -0,376        | 0,7136       |

$$R^2 = 0,4422$$

$F = 1,903$  (signifikant på 17% nivået)

Tabell 3.16: *Estimering av parametre når både mars og april 1997 er observert*

| Estimat             | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|---------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$           | 1  | -1,939   | 1,047 | -1,852        | 0,0869       |
| $\hat{\alpha}_{-1}$ | 1  | -1,605   | 4,530 | -0,354        | 0,7287       |
| $\hat{\alpha}_0$    | 1  | 11,049   | 5,402 | 2,045         | 0,0616       |
| $\hat{\alpha}_1$    | 1  | 2,877    | 3,955 | 0,727         | 0,4798       |
| $\hat{\beta}_{-1}$  | 1  | 1,297    | 4,548 | 0,285         | 0,7799       |
| $\hat{\beta}_0$     | 1  | -13,927  | 5,706 | -2,441        | 0,0297       |
| $\hat{\beta}_1$     | 1  | 0,308    | 5,263 | 0,059         | 0,9541       |

$$R^2 = 0,4001$$

$F = 1,734$  (signifikant på 20% nivået)

### 3.8.4 SNN522.IVL.U: Butikkhandel med nærings og nytelsesmidler i spesialforretninger. I alt

Estimatene for betingelsen (b) når vi observerer til og med mars 1997 og til og med april 1997 er vist i tabellene 3.17 og 3.18.

Tabell 3.17: *Estimering av parametre med betingelsen (b) når bare mars er observert*

|                                                | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|------------------------------------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$                                      | 1  | -0,747   | 0,868 | -0,861        | 0,4061       |
| $\hat{\alpha}_{-1}$                            | 1  | -3,281   | 3,748 | -0,875        | 0,3986       |
| $\hat{\alpha}_0$                               | 1  | 12,710   | 4,478 | 2,838         | 0,0149       |
| $\hat{\alpha}_1$                               | 1  | 2,922    | 4,071 | 0,718         | 0,4866       |
| $\hat{\beta}_{-1}$                             | 1  | 7,748    | 3,761 | 2,060         | 0,0618       |
| $\hat{\beta}_0$                                | 1  | -12,982  | 4,768 | -2,723        | 0,0185       |
| $\hat{\beta}_1$                                | 1  | -7,117   | 4,828 | -1,474        | 0,1662       |
| $R^2 = 0,8041$                                 |    |          |       |               |              |
| $F = 9,8500$ (signifikant på 1 prosent nivået) |    |          |       |               |              |

Tabell 3.18: *Estimering av parametre med betingelsen (b) når både mars og april er observert*

|                                                | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|------------------------------------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$                                      | 1  | -0,672   | 0,889 | -0,756        | 0,4632       |
| $\hat{\alpha}_{-1}$                            | 1  | -3,431   | 3,846 | -0,892        | 0,3886       |
| $\hat{\alpha}_0$                               | 1  | 12,302   | 4,587 | 2,682         | 0,0188       |
| $\hat{\alpha}_1$                               | 1  | -0,244   | 3,358 | -0,073        | 0,9432       |
| $\hat{\beta}_{-1}$                             | 1  | 7,846    | 3,861 | 2,032         | 0,0631       |
| $\hat{\beta}_0$                                | 1  | -12,087  | 4,845 | -2,495        | 0,0268       |
| $\hat{\beta}_1$                                | 1  | -4,386   | 4,468 | -0,982        | 0,3442       |
| $R^2 = 0,7787$                                 |    |          |       |               |              |
| $F = 9,1490$ (signifikant på 1 prosent nivået) |    |          |       |               |              |

### 3.8.5 SNN523.IVL.U: Butikkhandel med apotekvarer, sykepleie- artikler, kosmetikk og toalettartikler. I alt

Estimatene for betingelsen (b) når vi observerer til og med mars 1997 og til og med april 1997 er vist i tabellene 3.19 og 3.20.

Tabell 3.19: *Estimering av parametre med betingelsen (b) når bare mars er observert*

|                                                | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|------------------------------------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$                                      | 1  | 2,965    | 0,817 | 3,631         | 0,0034       |
| $\hat{\alpha}_{-1}$                            | 1  | 0,275    | 3,526 | 0,078         | 0,9392       |
| $\hat{\alpha}_0$                               | 1  | 6,894    | 4,213 | 1,636         | 0,1277       |
| $\hat{\alpha}_1$                               | 1  | 1,918    | 3,830 | 0,501         | 0,6255       |
| $\hat{\beta}_{-1}$                             | 1  | -0,753   | 3,539 | -0,213        | 0,8351       |
| $\hat{\beta}_0$                                | 1  | -15,563  | 4,486 | -3,470        | 0,0046       |
| $\hat{\beta}_1$                                | 1  | 7,229    | 4,542 | 1,592         | 0,1375       |
| $R^2 = 0,6560$                                 |    |          |       |               |              |
| $F = 4,5760$ (signifikant på 2 prosent nivået) |    |          |       |               |              |

Tabell 3.20: *Estimering av parametre med betingelsen (b) når både mars og april er observert*

|                                                | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|------------------------------------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$                                      | 1  | 2,910    | 0,816 | 3,566         | 0,0034       |
| $\hat{\alpha}_{-1}$                            | 1  | 0,385    | 3,529 | 0,109         | 0,9148       |
| $\hat{\alpha}_0$                               | 1  | 7,193    | 4,209 | 1,709         | 0,1112       |
| $\hat{\alpha}_1$                               | 1  | 4,242    | 3,082 | 1,376         | 0,1919       |
| $\hat{\beta}_{-1}$                             | 1  | -0,825   | 3,543 | -0,233        | 0,8194       |
| $\hat{\beta}_0$                                | 1  | -16,220  | 4,446 | -3,649        | 0,0029       |
| $\hat{\beta}_1$                                | 1  | 5,225    | 4,100 | 1,274         | 0,2248       |
| $R^2 = 0,7150$                                 |    |          |       |               |              |
| $F = 6,5240$ (signifikant på 1 prosent nivået) |    |          |       |               |              |

### 3.8.6 SNN5231.IVL.U: Butikkhandel med apotekvarer, sykepleie- artikler, kosmetikk og toalettartikler. Apotekvarer

Estimatene for betingelsen (b) når vi observerer til og med mars 1997 og til og med april 1997 er vist i tabellene 3.21 og 3.22.

Tabell 3.21: *Estimering av parametre med betingelsen (b) når bare mars er observert*

|                                                | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|------------------------------------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$                                      | 1  | 2,979    | 0,756 | 3,942         | 0,0020       |
| $\hat{\alpha}_{-1}$                            | 1  | 0,017    | 3,263 | 0,005         | 0,9960       |
| $\hat{\alpha}_0$                               | 1  | 7,150    | 3,899 | 1,834         | 0,0916       |
| $\hat{\alpha}_1$                               | 1  | 1,926    | 3,545 | 0,543         | 0,5969       |
| $\hat{\beta}_{-1}$                             | 1  | -0,759   | 3,275 | -0,232        | 0,8206       |
| $\hat{\beta}_0$                                | 1  | -15,819  | 4,151 | -3,810        | 0,0025       |
| $\hat{\beta}_1$                                | 1  | 7,486    | 4,204 | 1,781         | 0,1003       |
| $R^2 = 0,6968$                                 |    |          |       |               |              |
| $F = 5,5150$ (signifikant på 1 prosent nivået) |    |          |       |               |              |

Tabell 3.22: *Estimering av parametre med betingelsen (b) når både mars og april er observert*

|                                                | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|------------------------------------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$                                      | 1  | 2,921    | 0,765 | 3,819         | 0,0021       |
| $\hat{\alpha}_{-1}$                            | 1  | 0,134    | 3,308 | 0,040         | 0,9683       |
| $\hat{\alpha}_0$                               | 1  | 7,468    | 3,945 | 1,893         | 0,0808       |
| $\hat{\alpha}_1$                               | 1  | 4,397    | 2,888 | 1,522         | 0,1519       |
| $\hat{\beta}_{-1}$                             | 1  | -0,836   | 3,321 | -0,252        | 0,8052       |
| $\hat{\beta}_0$                                | 1  | -16,517  | 4,167 | -3,964        | 0,0016       |
| $\hat{\beta}_1$                                | 1  | 5,354    | 3,843 | 1,393         | 0,1869       |
| $R^2 = 0,7478$                                 |    |          |       |               |              |
| $F = 7,7100$ (signifikant på 1 prosent nivået) |    |          |       |               |              |

### 3.8.7 SNN5233.IVL.U: Butikkhandel med apotekvarer, sykepleie- artikler, kosmetikk og toalettartikler. Kosmetikk og toalett- artikler

Estimatene for betingelsen (b) når vi observerer til og med mars 1997 og til og med april 1997 er vist i tabellene 3.23 og 3.24.

Tabell 3.23: *Estimering av parametre med betingelsen (b) når bare mars er observert*

|                                                 | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|-------------------------------------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$                                       | 1  | 2,747    | 1,131 | 2,429         | 0,0318       |
| $\hat{\alpha}_{-1}$                             | 1  | 1,756    | 4,884 | 0,360         | 0,7254       |
| $\hat{\alpha}_0$                                | 1  | 5,759    | 5,835 | 0,987         | 0,3431       |
| $\hat{\alpha}_1$                                | 1  | 1,826    | 5,305 | 0,344         | 0,7367       |
| $\hat{\beta}_{-1}$                              | 1  | -0,943   | 4,902 | -0,192        | 0,8506       |
| $\hat{\beta}_0$                                 | 1  | -14,403  | 6,213 | -2,318        | 0,0389       |
| $\hat{\beta}_1$                                 | 1  | 6,006    | 6,291 | 0,955         | 0,3586       |
| $R^2 = 0,4691$                                  |    |          |       |               |              |
| $F = 2,1210$ (signifikant på 14 prosent nivået) |    |          |       |               |              |

Tabell 3.24: *Estimering av parametre med betingelsen (b) når både mars og april er observert*

|                                                | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|------------------------------------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$                                      | 1  | 2,728    | 1,087 | 2,510         | 0,0261       |
| $\hat{\alpha}_{-1}$                            | 1  | 1,793    | 4,702 | 0,381         | 0,7091       |
| $\hat{\alpha}_0$                               | 1  | 5,860    | 5,607 | 1,045         | 0,3150       |
| $\hat{\alpha}_1$                               | 1  | 2,607    | 4,105 | 0,635         | 0,5364       |
| $\hat{\beta}_{-1}$                             | 1  | -0,968   | 4,720 | -0,205        | 0,8407       |
| $\hat{\beta}_0$                                | 1  | -14,624  | 5,922 | -2,469        | 0,0282       |
| $\hat{\beta}_1$                                | 1  | 5,331    | 5,462 | 0,976         | 0,3469       |
| $R^2 = 0,5266$                                 |    |          |       |               |              |
| $F = 2,8920$ (signifikant på 6 prosent nivået) |    |          |       |               |              |

### 3.8.8 SNN524.IVL.U: Butikkhandel med andre nye varer. I alt

Estimatene for betingelsen (b) når vi observerer til og med mars 1997 og til og med april 1997 er vist i tabellene 3.25 og 3.26.

Tabell 3.25: *Estimering av parametre med betingelsen (b) når bare mars er observert*

|                                                | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|------------------------------------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$                                      | 1  | 0,231    | 0,930 | 0,249         | 0,8075       |
| $\hat{\alpha}_{-1}$                            | 1  | -0,343   | 4,014 | -0,085        | 0,9333       |
| $\hat{\alpha}_0$                               | 1  | 4,838    | 4,796 | 1,009         | 0,3330       |
| $\hat{\alpha}_1$                               | 1  | 11,625   | 4,360 | 2,666         | 0,0206       |
| $\hat{\beta}_{-1}$                             | 1  | 0,423    | 4,028 | 0,105         | 0,9181       |
| $\hat{\beta}_0$                                | 1  | -13,092  | 5,106 | -2,564        | 0,0248       |
| $\hat{\beta}_1$                                | 1  | -3,451   | 5,171 | -0,667        | 0,5171       |
| $R^2 = 0,6594$                                 |    |          |       |               |              |
| $F = 4,6460$ (signifikant på 2 prosent nivået) |    |          |       |               |              |

Tabell 3.26: *Estimering av parametre med betingelsen (b) når både mars og april er observert*

|                                                | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|------------------------------------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$                                      | 1  | 0,261    | 0,900 | 0,290         | 0,7764       |
| $\hat{\alpha}_{-1}$                            | 1  | -0,402   | 3,891 | -0,103        | 0,9193       |
| $\hat{\alpha}_0$                               | 1  | 4,678    | 4,640 | 1,008         | 0,3317       |
| $\hat{\alpha}_1$                               | 1  | 10,385   | 3,397 | 3,057         | 0,0092       |
| $\hat{\beta}_{-1}$                             | 1  | 0,462    | 3,906 | 0,118         | 0,9077       |
| $\hat{\beta}_0$                                | 1  | -12,742  | 4,901 | -2,600        | 0,0220       |
| $\hat{\beta}_1$                                | 1  | -2,382   | 4,520 | -0,527        | 0,6071       |
| $R^2 = 0,7155$                                 |    |          |       |               |              |
| $F = 6,5380$ (signifikant på 1 prosent nivået) |    |          |       |               |              |

### 3.8.9 SNN5241.IVL.U: Butikkhandel med andre nye varer. Tekstiler og utstyrsvarer

Estimatene for betingelsen (b) når vi observerer til og med mars 1997 og til og med april 1997 er vist i tabellene 3.27 og 3.28.

Tabell 3.27: *Estimering av parametre med betingelsen (b) når bare mars er observert*

|                                                 | df | Estimert | std    | $t$ for $H_0$ | $Prob >  t $ |
|-------------------------------------------------|----|----------|--------|---------------|--------------|
| $\widehat{S}$                                   | 1  | -3,245   | 2,131  | -1,523        | 0,1537       |
| $\widehat{\alpha}_{-1}$                         | 1  | 4,641    | 9,203  | 0,504         | 0,6232       |
| $\widehat{\alpha}_0$                            | 1  | 5,326    | 10,996 | 0,484         | 0,6368       |
| $\widehat{\alpha}_1$                            | 1  | 14,552   | 9,996  | 1,456         | 0,1711       |
| $\widehat{\beta}_{-1}$                          | 1  | -3,724   | 9,236  | -0,403        | 0,6939       |
| $\widehat{\beta}_0$                             | 1  | -16,395  | 11,707 | -1,400        | 0,1867       |
| $\widehat{\beta}_1$                             | 1  | -4,400   | 11,855 | -0,371        | 0,7170       |
| $R^2 = 0,3945$                                  |    |          |        |               |              |
| $F = 1,5640$ (signifikant på 24 prosent nivået) |    |          |        |               |              |

Tabell 3.28: *Estimering av parametre med betingelsen (b) når både mars og april er observert*

|                                                 | df | Estimert | std    | $t$ for $H_0$ | $Prob >  t $ |
|-------------------------------------------------|----|----------|--------|---------------|--------------|
| $\widehat{S}$                                   | 1  | -3,135   | 2,095  | -1,497        | 0,1584       |
| $\widehat{\alpha}_{-1}$                         | 1  | 4,420    | 9,061  | 0,488         | 0,6338       |
| $\widehat{\alpha}_0$                            | 1  | 4,725    | 10,804 | 0,437         | 0,6690       |
| $\widehat{\alpha}_1$                            | 1  | 9,891    | 7,911  | 1,250         | 0,2332       |
| $\widehat{\beta}_{-1}$                          | 1  | -3,579   | 9,096  | -0,393        | 0,7004       |
| $\widehat{\beta}_0$                             | 1  | -15,078  | 11,412 | -1,321        | 0,2092       |
| $\widehat{\beta}_1$                             | 1  | -0,380   | 10,525 | -0,036        | 0,9718       |
| $R^2 = 0,3888$                                  |    |          |        |               |              |
| $F = 1,6540$ (signifikant på 21 prosent nivået) |    |          |        |               |              |

### 3.8.10 SNN5242.IVL.U: Butikkhandel med andre nye varer. Klær

Estimatene for betingelsen (b) når vi observerer til og med mars 1997 og til og med april 1997 er vist i tabellene 3.29 og 3.30.

Tabell 3.29: *Estimering av parametre med betingelsen (b) når bare mars er observert*

|                                                 | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|-------------------------------------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$                                       | 1  | -4,966   | 1,755 | -2,829        | 0,0152       |
| $\hat{\alpha}_{-1}$                             | 1  | -1,573   | 7,579 | -0,208        | 0,8391       |
| $\hat{\alpha}_0$                                | 1  | 3,561    | 9,056 | 0,393         | 0,7011       |
| $\hat{\alpha}_1$                                | 1  | 17,170   | 8,233 | 2,086         | 0,0590       |
| $\hat{\beta}_{-1}$                              | 1  | 0,670    | 7,607 | 0,088         | 0,9313       |
| $\hat{\beta}_0$                                 | 1  | -13,107  | 9,642 | -1,359        | 0,1990       |
| $\hat{\beta}_1$                                 | 1  | -6,720   | 9,763 | -0,688        | 0,5044       |
| $R^2 = 0,4830$                                  |    |          |       |               |              |
| $F = 2,2430$ (signifikant på 12 prosent nivået) |    |          |       |               |              |

Tabell 3.30: *Estimering av parametre med betingelsen (b) når både mars og april er observert*

|                                                 | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|-------------------------------------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$                                       | 1  | -4,801   | 1,818 | -2,641        | 0,0204       |
| $\hat{\alpha}_{-1}$                             | 1  | -1,903   | 7,864 | -0,242        | 0,8126       |
| $\hat{\alpha}_0$                                | 1  | 2,665    | 9,377 | 0,284         | 0,7807       |
| $\hat{\alpha}_1$                                | 1  | 10,223   | 6,866 | 1,489         | 0,1603       |
| $\hat{\beta}_{-1}$                              | 1  | 0,886    | 7,895 | 0,112         | 0,9123       |
| $\hat{\beta}_0$                                 | 1  | -11,144  | 9,905 | -1,125        | 0,2809       |
| $\hat{\beta}_1$                                 | 1  | -0,728   | 9,135 | -0,080        | 0,9377       |
| $R^2 = 0,4046$                                  |    |          |       |               |              |
| $F = 1,7670$ (signifikant på 19 prosent nivået) |    |          |       |               |              |

### 3.8.11 SNN5243.IVL.U: Butikkhandel med andre nye varer. Skotøy reise effekter av lær og lærvarer.

Estimatene for betingelsen (b) når vi observerer til og med mars 1997 og til og med april 1997 er vist i tabellene 3.31 og 3.32.

Tabell 3.31: *Estimering av parametre med betingelsen (b) når bare mars er observert*

|                                                 | df | Estimert | std    | t for $H_0$ | Prob >  t |
|-------------------------------------------------|----|----------|--------|-------------|-----------|
| $\hat{S}$                                       | 1  | -6,698   | 1,951  | -3,434      | 0,0050    |
| $\hat{\alpha}_{-1}$                             | 1  | -4,738   | 8,424  | -0,562      | 0,5841    |
| $\hat{\alpha}_0$                                | 1  | 1,197    | 10,064 | 0,119       | 0,9073    |
| $\hat{\alpha}_1$                                | 1  | 14,937   | 9,150  | 1,632       | 0,1285    |
| $\hat{\beta}_{-1}$                              | 1  | 1,982    | 8,454  | 0,234       | 0,8186    |
| $\hat{\beta}_0$                                 | 1  | -6,459   | 10,716 | -0,603      | 0,5579    |
| $\hat{\beta}_1$                                 | 1  | -6,919   | 10,851 | -0,638      | 0,5357    |
| $R^2 = 0,3378$                                  |    |          |        |             |           |
| $F = 1,2240$ (signifikant på 36 prosent nivået) |    |          |        |             |           |

Tabell 3.32: *Estimering av parametre med betingelsen (b) når bare mars er observert*

|                                                 | df | Estimert | std    | t for $H_0$ | Prob >  t |
|-------------------------------------------------|----|----------|--------|-------------|-----------|
| $\hat{S}$                                       | 1  | -6.721   | 1.873  | -3.589      | 0.0033    |
| $\hat{\alpha}_{-1}$                             | 1  | -4.692   | 8.100  | -0.579      | 0.5723    |
| $\hat{\alpha}_0$                                | 1  | 1.321    | 9.659  | 0.137       | 0.8933    |
| $\hat{\alpha}_1$                                | 1  | 15.899   | 7.072  | 2.248       | 0.0426    |
| $\hat{\beta}_{-1}$                              | 1  | 1.952    | 8.131  | 0.240       | 0.8140    |
| $\hat{\beta}_0$                                 | 1  | -6.730   | 10.202 | -0.660      | 0.5210    |
| $\hat{\beta}_1$                                 | 1  | -7.749   | 9.409  | -0.824      | 0.4250    |
| $R^2 = 0.4463$                                  |    |          |        |             |           |
| $F = 2.0960$ (signifikant på 14 prosent nivået) |    |          |        |             |           |

### 3.8.12 SNN5244.IVL.U: Butikkhandel med andre nye varer. Belysningsutstyr, kjøkkenutstyr, møbler og innredningsartikler.

Estimatene for betingelsen (b) når vi observerer til og med mars 1997 og til og med april 1997 er vist i tabellene 3.33 og 3.34.

Tabell 3.33: *Estimering av parametre med betingelsen (b) når bare mars er observert*

|                                                | df | Estimert | std   | t for $H_0$ | Prob >  t |
|------------------------------------------------|----|----------|-------|-------------|-----------|
| $\hat{S}$                                      | 1  | 4,196    | 0,963 | 4,357       | 0,0009    |
| $\hat{\alpha}_{-1}$                            | 1  | 0,118    | 4,158 | 0,028       | 0,9779    |
| $\hat{\alpha}_0$                               | 1  | 5,500    | 4,968 | 1,107       | 0,2900    |
| $\hat{\alpha}_1$                               | 1  | 12,067   | 4,517 | 2,671       | 0,0204    |
| $\hat{\beta}_{-1}$                             | 1  | 1,061    | 4,174 | 0,254       | 0,8037    |
| $\hat{\beta}_0$                                | 1  | -13,959  | 5,290 | -2,639      | 0,0216    |
| $\hat{\beta}_1$                                | 1  | -4,786   | 5,357 | -0,893      | 0,3892    |
| $R^2 = 0,6504$                                 |    |          |       |             |           |
| $F = 4,4650$ (signifikant på 2 prosent nivået) |    |          |       |             |           |

Tabell 3.34: *Estimering av parametre med betingelsen (b) når både mars og april er observert*

|                                                | df | Estimert | std   | t for $H_0$ | Prob >  t |
|------------------------------------------------|----|----------|-------|-------------|-----------|
| $\hat{S}$                                      | 1  | 4,222    | 0,930 | 4,542       | 0,0006    |
| $\hat{\alpha}_{-1}$                            | 1  | 0,065    | 4,021 | 0,016       | 0,9873    |
| $\hat{\alpha}_0$                               | 1  | 5,358    | 4,795 | 1,117       | 0,2841    |
| $\hat{\alpha}_1$                               | 1  | 10,965   | 3,511 | 3,123       | 0,0081    |
| $\hat{\beta}_{-1}$                             | 1  | 1,095    | 4,037 | 0,271       | 0,7904    |
| $\hat{\beta}_0$                                | 1  | -13,648  | 5,065 | -2,695      | 0,0184    |
| $\hat{\beta}_1$                                | 1  | -3,835   | 4,671 | -0,821      | 0,4264    |
| $R^2 = 0,7140$                                 |    |          |       |             |           |
| $F = 6,4910$ (signifikant på 1 prosent nivået) |    |          |       |             |           |

### 3.8.13 SNN5245.IVL.U: Butikkhandel med andre nye varer. Elektriske husholdningsapparater, radio, fjernsyn, kassetter og musikkinstrumenter.

Estimatene for betingelsen (b) når vi observerer til og med mars 1997 og til og med april 1997 er vist i tabellene 3.35 og 3.36.

Tabell 3.35: *Estimering av parametre med betingelsen (b) når bare mars er observert*

|                                                | df | Estimert | std   | t for $H_0$ | Prob >  t |
|------------------------------------------------|----|----------|-------|-------------|-----------|
| $\hat{S}$                                      | 1  | 4,163    | 1,107 | 3,759       | 0,0027    |
| $\hat{\alpha}_{-1}$                            | 1  | -0,311   | 4,782 | -0,065      | 0,9492    |
| $\hat{\alpha}_0$                               | 1  | 2,354    | 5,713 | 0,412       | 0,6876    |
| $\hat{\alpha}_1$                               | 1  | 13,031   | 5,194 | 2,509       | 0,0275    |
| $\hat{\beta}_{-1}$                             | 1  | 2,540    | 4,799 | 0,529       | 0,6063    |
| $\hat{\beta}_0$                                | 1  | -11,448  | 6,083 | -1,882      | 0,0843    |
| $\hat{\beta}_1$                                | 1  | -6,166   | 6,160 | -1,001      | 0,3366    |
| $R^2 = 0,5976$                                 |    |          |       |             |           |
| $F = 3,5640$ (signifikant på 4 prosent nivået) |    |          |       |             |           |

Tabell 3.36: *Estimering av parametre med betingelsen (b) når både mars og april er observert*

|                                                | df | Estimert | std   | t for $H_0$ | Prob >  t |
|------------------------------------------------|----|----------|-------|-------------|-----------|
| $\hat{S}$                                      | 1  | 4,280    | 1,168 | 3,664       | 0,0029    |
| $\hat{\alpha}_{-1}$                            | 1  | -0,544   | 5,052 | -0,108      | 0,9159    |
| $\hat{\alpha}_0$                               | 1  | 1,721    | 6,024 | 0,286       | 0,7797    |
| $\hat{\alpha}_1$                               | 1  | 8,116    | 4,411 | 1,840       | 0,0887    |
| $\hat{\beta}_{-1}$                             | 1  | 2,693    | 5,072 | 0,531       | 0,6044    |
| $\hat{\beta}_0$                                | 1  | -10,060  | 6,363 | -1,581      | 0,1379    |
| $\hat{\beta}_1$                                | 1  | -1,926   | 5,869 | -0,328      | 0,7480    |
| $R^2 = 0,5281$                                 |    |          |       |             |           |
| $F = 2,9100$ (signifikant på 6 prosent nivået) |    |          |       |             |           |

### 3.8.14 SNN5247.IVL.U: Butikkhandel med andre nye varer. Bøker, papir, aviser og blader.

Estimatene for betingelsen (b) når vi observerer til og med mars 1997 og til og med april 1997 er vist i tabellene 3.37 og 3.38.

Tabell 3.37: *Estimering av parametre med betingelsen (b) når bare mars er observert*

|                                                | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|------------------------------------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$                                      | 1  | 5,471    | 1,737 | 3,150         | 0,0084       |
| $\hat{\alpha}_{-1}$                            | 1  | 7,393    | 7,500 | 0,986         | 0,3437       |
| $\hat{\alpha}_0$                               | 1  | 8,711    | 8,961 | 0,972         | 0,3502       |
| $\hat{\alpha}_1$                               | 1  | 2,515    | 8,147 | 0,309         | 0,7628       |
| $\hat{\beta}_{-1}$                             | 1  | -8,225   | 7,528 | -1,093        | 0,2960       |
| $\hat{\beta}_0$                                | 1  | -21,534  | 9,542 | -2,257        | 0,0435       |
| $\hat{\beta}_1$                                | 1  | 11,140   | 9,662 | 1,153         | 0,2714       |
| $R^2 = 0,5229$                                 |    |          |       |               |              |
| $F = 2,6300$ (signifikant på 8 prosent nivået) |    |          |       |               |              |

Tabell 3.38: *Estimering av parametre med betingelsen (b) når både mars og april er observert*

|                                                 | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|-------------------------------------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$                                       | 1  | 5,642    | 1,813 | 3,112         | 0,0082       |
| $\hat{\alpha}_{-1}$                             | 1  | 7,050    | 7,842 | 0,899         | 0,3850       |
| $\hat{\alpha}_0$                                | 1  | 7,779    | 9,351 | 0,832         | 0,4205       |
| $\hat{\alpha}_1$                                | 1  | -4,720   | 6,847 | -0,689        | 0,5027       |
| $\hat{\beta}_{-1}$                              | 1  | -7,999   | 7,873 | -1,016        | 0,3281       |
| $\hat{\beta}_0$                                 | 1  | -19,490  | 9,877 | -1,973        | 0,0701       |
| $\hat{\beta}_1$                                 | 1  | 17,381   | 9,110 | 1,908         | 0,0787       |
| $R^2 = 0,4419$                                  |    |          |       |               |              |
| $F = 2,0580$ (signifikant på 14 prosent nivået) |    |          |       |               |              |

### 3.8.15 SNN5248.IVL.U: Butikkhandel med andre nye varer. Butikkhandel ellers.

Estimatene for betingelsen (b) når vi observerer til og med mars 1997 og til og med april 1997 er vist i tabellene 3.39 og 3.40.

Tabell 3.39: *Estimering av parametre med betingelsen (b) når bare mars er observert*

|                                                | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|------------------------------------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$                                      | 1  | 1.511    | 0.691 | 2.185         | 0.0494       |
| $\hat{\alpha}_{-1}$                            | 1  | -2.167   | 2.986 | -0.726        | 0.4819       |
| $\hat{\alpha}_0$                               | 1  | 4.895    | 3.567 | 1.372         | 0.1951       |
| $\hat{\alpha}_1$                               | 1  | 6.783    | 3.243 | 2.091         | 0.0584       |
| $\hat{\beta}_{-1}$                             | 1  | 1.786    | 2.997 | 0.596         | 0.5623       |
| $\hat{\beta}_0$                                | 1  | -11.419  | 3.798 | -3.006        | 0.0109       |
| $\hat{\beta}_1$                                | 1  | 0.122    | 3.846 | 0.032         | 0.9752       |
| $R^2 = 0.6586$                                 |    |          |       |               |              |
| $F = 4.6290$ (signifikant på 2 prosent nivået) |    |          |       |               |              |

Tabell 3.40: *Estimering av parametre med betingelsen (b) når både mars og april er observert*

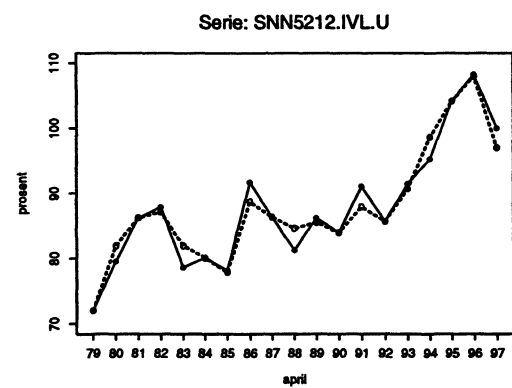
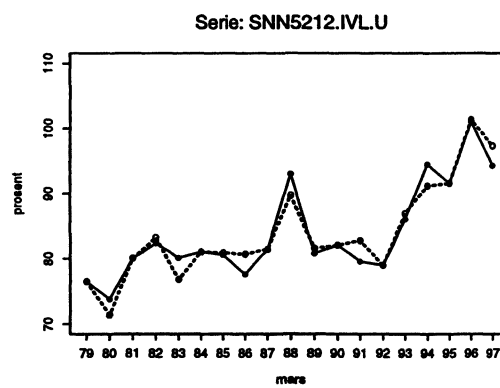
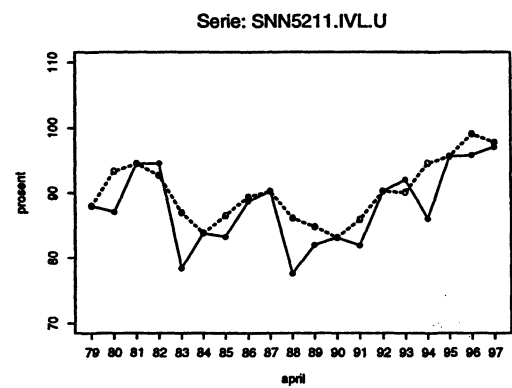
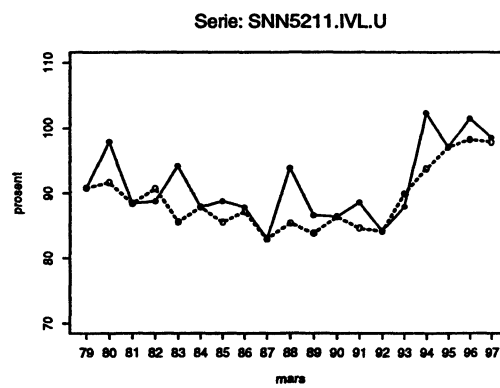
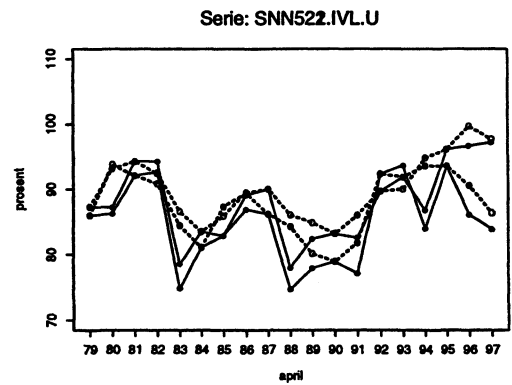
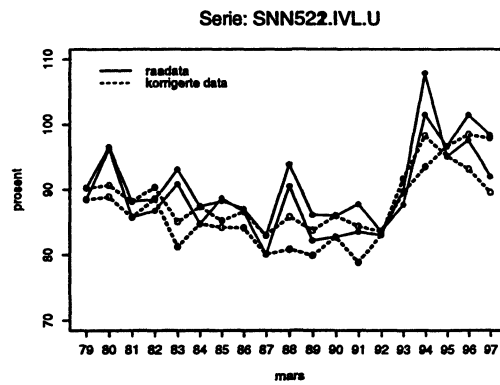
|                                                | df | Estimert | std   | $t$ for $H_0$ | $Prob >  t $ |
|------------------------------------------------|----|----------|-------|---------------|--------------|
| $\hat{S}$                                      | 1  | 1,406    | 0,795 | 1,769         | 0,1004       |
| $\hat{\alpha}_{-1}$                            | 1  | -1,957   | 3,438 | -0,569        | 0,5789       |
| $\hat{\alpha}_0$                               | 1  | 5,466    | 4,100 | 1,333         | 0,2054       |
| $\hat{\alpha}_1$                               | 1  | 11,210   | 3,002 | 3,734         | 0,0025       |
| $\hat{\beta}_{-1}$                             | 1  | 1,648    | 3,452 | 0,477         | 0,6411       |
| $\hat{\beta}_0$                                | 1  | -12,670  | 4,331 | -2,926        | 0,0118       |
| $\hat{\beta}_1$                                | 1  | -3,697   | 3,994 | -0,926        | 0,3715       |
| $R^2 = 0,7588$                                 |    |          |       |               |              |
| $F = 8,1780$ (signifikant på 1 prosent nivået) |    |          |       |               |              |

**- En oppsummering for seriene SNN5221, SNN522, SNN523 og SNN524.-  
IVL.U**

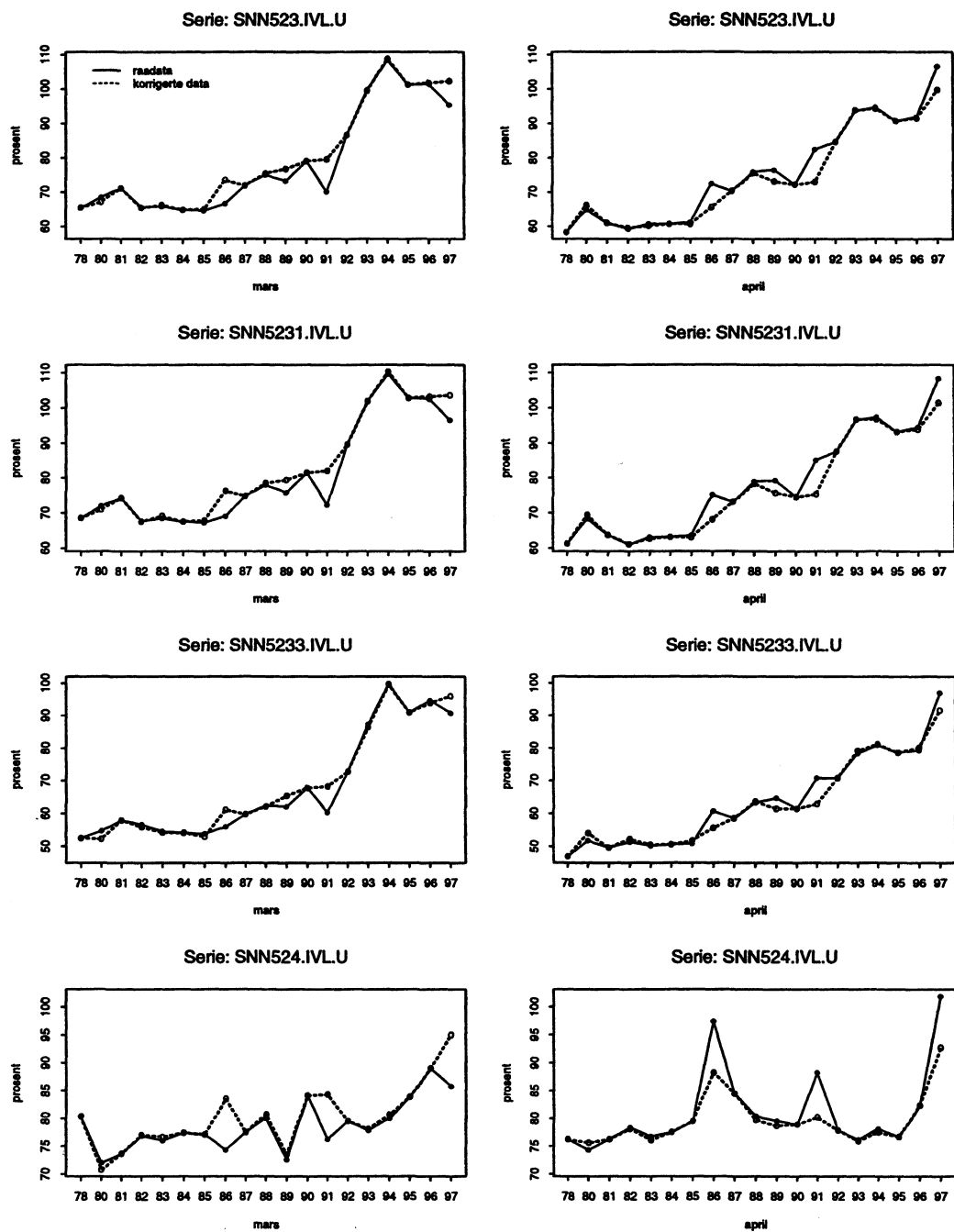
1. For seriene SNN521.IVL.U, SNN522.IVL.U og SNN523.IVL.U er  $\hat{\alpha}_0$ ,  $\hat{\beta}_{-1}$  og  $\hat{\beta}_0$  er blant høyeste og laveste verdiene. Tolkningen er at en kraftig forskyvning av handel fra helligdagene i uke før og påskeuka til virkedagene i påskeuka.
2. For seriene SNN524.IVL.U er  $\hat{\alpha}_0$ ,  $\hat{\alpha}_1$  og  $\hat{\beta}_0$  er blant høyeste og laveste verdiene. Tolkningen er at en kraftig forskyvning av handel fra helligdagene i påskeuka til virkedagene i påskeuka og uka etter.

### **3.9 Figurer**

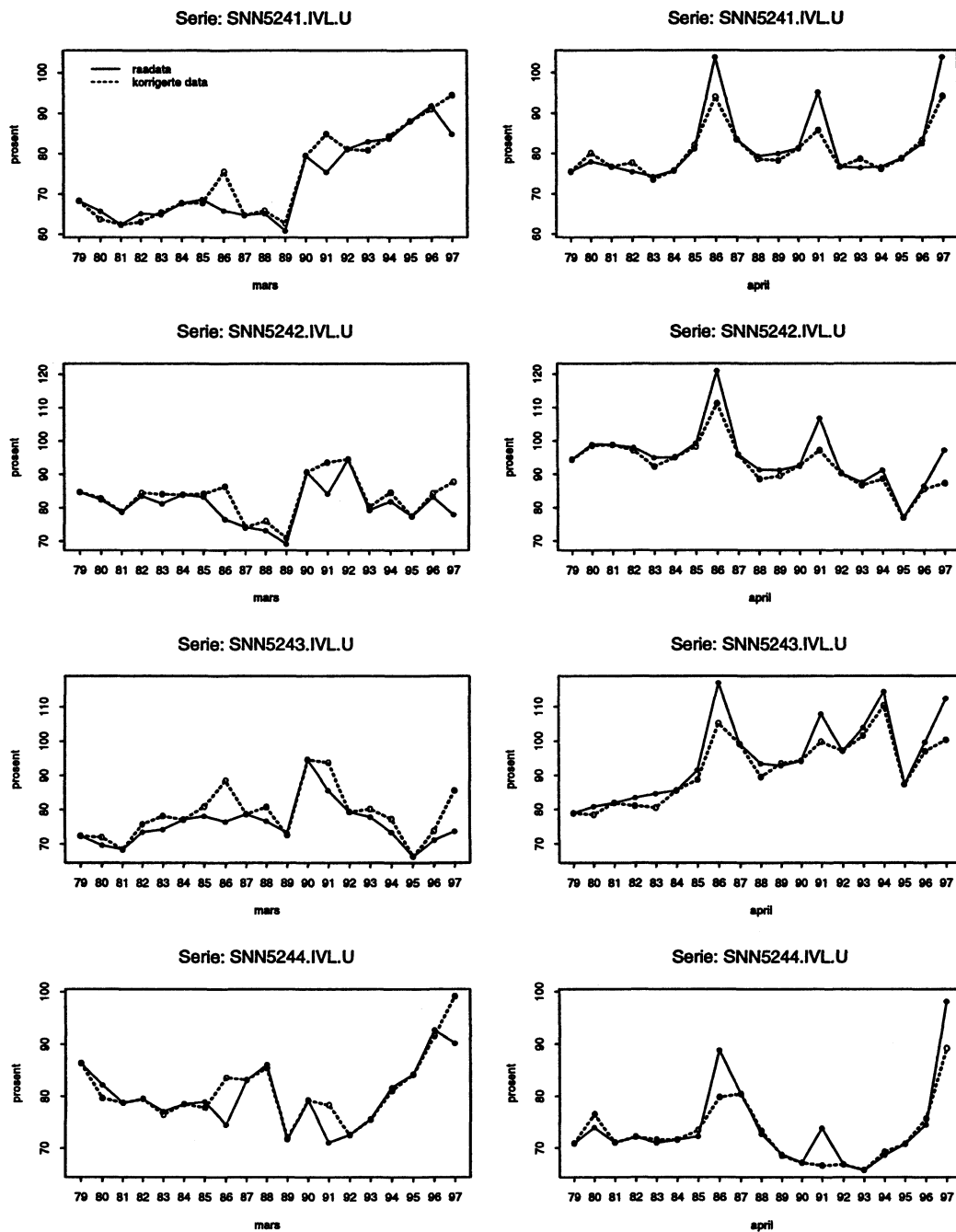
Rådata og korrigerte dataene for seriene SNN521, SNN522, SNN523 og SNN524.IVL.U for mars og april er plottet i figurene 3.12, 3.13, 3.14 og 3.15, med en hel linje for rådata og en brudd linje for korrigerte data.



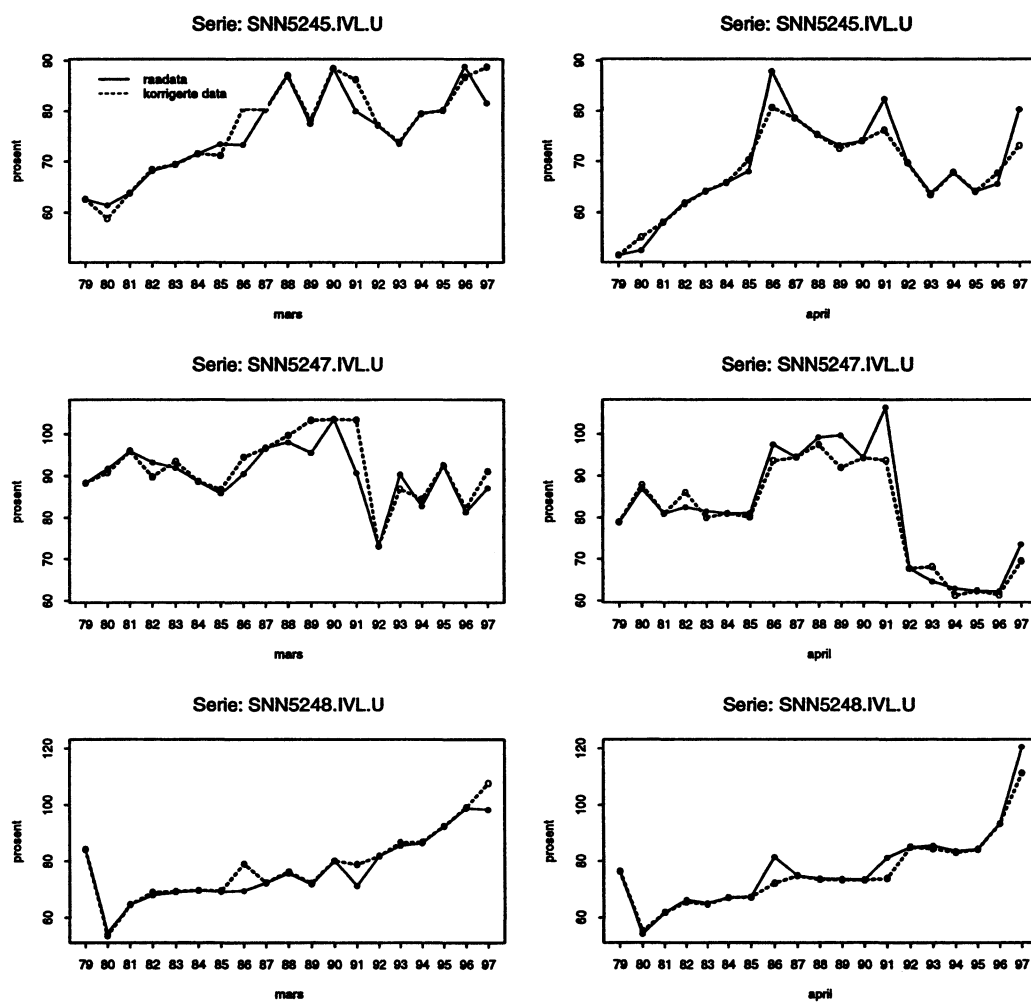
Figur 3.12: Rådata og korrigerende data



Figur 3.13: Rådata og korrigerede data



Figur 3.14: Rådata og korrigerede data



Figur 3.15: Rådata og korrigeret data

## Kapittel 4

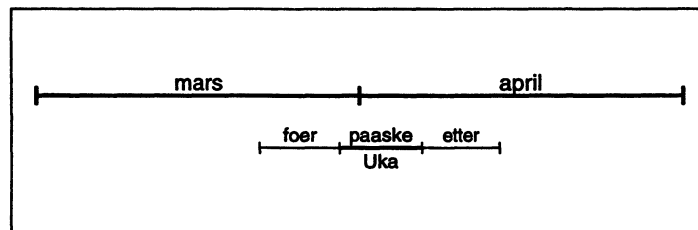
# Påskekorrigeringen for detaljvolumindeksen

Dinh Quang Pham

### 4.1 Innledning

I notat 97/73 presenterte vi en metode av påskekorrigeringen for detaljvolumindeksen. Metoden er tilpasset norske forhold og fungerer godt. Den gir en mer tilfredsstillende sesongjustering av tallene enn ved bruk av rutinene i X-11 og X-12-ARIMA. Metoden er basert på følgende antagelser:

Ingen forskyvning inn/ut av de tre ukene: uken før påskeuken fra mandag til palmesøndag, påskeuken fra mandag etter palmesøndag til 1. påskedag og uken etter påskeuken fra 2. påskedag til søndag.



Figur 4.1: Et diagram for de tre ukene i forhold til mars og april

I framstillingen har vi brukt betegnelsene

- *påskeeffekt*: effekten som påvirker rådata på grunn av påske.
- *påskedagene*: for å beskrive dagene der rådata er påvirket av påskeeffekten. I X-12-ARIMA bruker en  $d$  for disse dagene. For eksempel,  $d = 3$  betyr at skjærtorsdag, langfredag og påskeaften er lik påskedagene.
- *gammel metode*: for å vise til metoden som er beskrevet i notatet 97/73.

Man bruker påskedagene for å lage forklaringsvariable til estimering av påskeeffekten. Ved X-12ARIMA kan man angi  $d$  på forhånd eller den estimeres fra data ved AIC

test. Metode i notat 97/73 er mindre fleksibel enn X-12-ARIMA på dette punktet, siden  $d$  har en fast verdi som er tre uker. Ved å velge  $d$  slik kan metoden ikke tilpasses for tidsserier fra andre statistiske områder. For registrerte nye personbiler som er observert fra januar 1976 til desember 1998, estimeres  $\hat{d} = 1$  ved X-12-ARIMA. Vi vet også at denne serien er lite påvirket av påskeeffekten. Det er kun mars og apriltall i 1978, 1986, 1989, 1991 og 1997 som korrigeres, siden påske faller i mars. Ved å bruke vår metode korrigeres mars og apriltall i 1978, 1980, 1982, 1983, 1985, 1986, 1988, 1991, 1993, 1994, 1997 og 1998. Det blir 14 av totalt 23 år. Vi ser at påsken er seint i april 1993, men rådata blir allikevel justert. Mønsteret av mars og april kan bli forandret på grunn av korrigering. Konsekvensen er at sesongjusterte tall i disse to periodene kan bli dårligere. Vi har utviklet en ny metode ved å anvende fordeler og rette opp svakheter i den gamle metoden. Vi erstatter de tre ukene med tre intervaller  $T_1$ ,  $T_2$  og  $T_3$  (som beskrives mer detaljert i neste avsnitt) der lengdene  $d_1$ ,  $d_2$  og  $d_3$  velges selv av brukeren. For eksempel,  $d_1 = 3$ ,  $d_2 = 4$  og  $d_3 = 1$  for  $t=1976, \dots, 1990$  og  $d_1 = 0$ ,  $d_2 = 7$  og  $d_3 = 0$  for  $t=1991, \dots, 1999$ . Metoden blir derfor mer fleksibel. Vi har ikke laget en modell for å estimere  $d_1$ ,  $d_2$  og  $d_3$  fra data. I X-12-ARIMA estimerer programmet  $d$  ved å sammenligne AIC fra  $d = 1$ ,  $d = 8$  og  $d = 15$ , og plukker den som gir minst AIC. Brukeren velger  $d_i$  ved å bruke sin kunnskap om tidsserien.

I det gamle notatet estimerer vi effektene ikke direkte fra observerte verdiene,  $Y_{M,t}$  og  $Y_{A,t}$ , men fra differansen mellom hhv mars og april sammenliknet med gjennomsnittet for de to månedene,  $X_{M,t}$  og  $X_{A,t}$

$$\begin{aligned} X_{M,t} &= Y_{M,t} - \frac{Y_{M,t} + Y_{A,t}}{2} \\ X_{A,t} &= Y_{A,t} - \frac{Y_{M,t} + Y_{A,t}}{2} \end{aligned} \quad (4.1)$$

Effektene estimeres ved regresjonsanalyse fra modellene

$$\begin{aligned} X_{M,t} &= S + \alpha_{-1}I_{-1,t} + \alpha_0I_{0,t} + \alpha_1I_{1,t} + \beta_{-1}J_{-1,t} + \beta_0J_{0,t} + \beta_1J_{1,t} + \epsilon_{M,t} \\ X_{A,t} &= -S - \alpha_{-1}(1 - I_{-1,t}) + \alpha_0(1 - I_{0,t}) + \alpha_1(1 - I_{1,t}) \\ &\quad + \beta_{-1}(1 - J_{-1,t}) + \beta_0(1 - J_{0,t}) + \beta_1(1 - J_{1,t}) + \epsilon_{A,t} \end{aligned} \quad (4.2)$$

hvor  $S$  er sesongutslaget. Modellen ser kunstig ut siden vi antar at

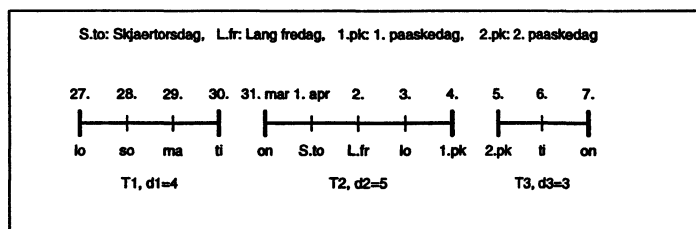
- sesongutslaget for mars og april er like, men med motsatt fortegn.
- andelen av virkedager i mars i uka før påske og andelen av virkedager i april i uka før påske har samme effekten, men med motsatt fortegn. Det gjelder også for  $I_0$  og  $I_1$ .

$\hat{\alpha}_i$  og  $\hat{\beta}_i$  estimeres med betingelsen  $\hat{\alpha}_{-1} + \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\beta}_{-1} + \hat{\beta}_0 + \hat{\beta}_1 = 0$ . Dermed kan (4.2) reduseres til en ligning. Dessuten ser vi bort fra ukedagseffektene fra modellen, men påskeeffekten påvirkes også av ukedagseffekter.

I de neste avsnittene beskriver vi metoden og programmet for å estimere påskeeffekten.

## 4.2 Beskrivelse av metoden

Påskeeffektene estimeres ved regresjonsanalyse. Vi konstruerer forklaringsvariablene ved å telle antall dager i periodene  $T_1$ ,  $T_2$  og  $T_3$  som faller i mars. La  $d_1 = 4$ ,  $d_2 = 5$ ,  $d_3 = 3$  og  $t = 1999$ . Vi vet at 1. påskedag i dette året er 4. april.  $T_2$  inneholder 5 dager som telles fra høyre til venstre side og starter fra 1. påskedag (søndag 4. april). De 4 neste dagene er lørdag (3. april), Langfredag (2. april), Skjærtorsdag (1. april) og onsdag (31. mars).  $T_1$  inneholder 4 dager: tirsdag (30. mars), mandag (29. mars), søndag (28. mars) og lørdag (27. mars).  $T_3$  ligger etter  $T_2$  og starter fra 2. påskedag. Dette intervallet inneholder mandag (5. april), tirsdag (6. april) og onsdag (7. april). I alt er det 12 dager. Figur 4.2 viser  $T_1$ ,  $T_2$  og  $T_3$  i 1999. I vår metode er søndag, skjærtorsdag, langfredag, 1. påskedag og 2. påskedag helligdager.



Figur 4.2: Påskedagene av  $T_1$ ,  $T_2$  og  $T_3$  i 1999

Beregningen av effektene er basert på antall virke- og helligdager av  $T_1$ ,  $T_2$  og  $T_3$  i mars via 6 regresjonsvariablene:

$$\begin{aligned}
 I_{-1,t} &= \frac{\# \text{virkedager av } T_1 \text{ i mars, i år } t}{\# \text{virkedager i } T_1, \text{ i år } t} & J_{-1,t} &= \frac{\# \text{helligdager av } T_1 \text{ i mars, i år } t}{\# \text{helligdager i } T_1, \text{ i år } t} \\
 I_{0,t} &= \frac{\# \text{virkedager av } T_2 \text{ i mars, i år } t}{\# \text{virkedager i } T_2, \text{ i år } t} & J_{0,t} &= \frac{\# \text{helligdager av } T_2 \text{ i mars, i år } t}{\# \text{helligdager i } T_2, \text{ i år } t} \\
 I_{1,t} &= \frac{\# \text{virkedager av } T_3 \text{ i mars, i år } t}{\# \text{virkedager i } T_3, \text{ i år } t} & J_{1,t} &= \frac{\# \text{helligdager av } T_3 \text{ i mars, i år } t}{\# \text{helligdager i } T_3, \text{ i år } t}
 \end{aligned} \tag{4.3}$$

Vi får fra figur 4.2

- Virkedagene i  $T_1$  : lørdag (27/3), mandag (29/3), tirsdag (30/3). 3 dager i mars.
- Virkedagene i  $T_2$  : onsdag (31/3), lørdag (2/4). 1 dag i mars.
- Virkedagene i  $T_3$  : tirsdag (6/4), onsdag (7/4). Ingen dag i mars.
- Helligdagene i  $T_1$  : søndag (28/3). 1 dag i mars
- Helligdagene i  $T_2$  : Skjærtorsdag (1/4), Langfredag (2/4), 1. påskedag (4/4). Ingen dag i mars.
- Helligdagene i  $T_3$  : 2. påskedag (5/4). Ingen dag i mars.

Verdiene av  $I$ . og  $J$ . i 1999 blir:

$$\begin{aligned}
 I_{-1} &= 3/3 = 1, & I_0 &= 1/2 = 0.5, & I_1 &= 0/2 = 0, \\
 J_{-1} &= 1/1 = 1, & J_0 &= 0/3 = 0, & J_1 &= 0/1 = 0.
 \end{aligned}$$

Merk at vi bruker datoen til 1. påskedag i år  $t$  for å beregne  $I_{.,t}$  og  $J_{.,t}$ .

### 4.2.1 Symboler

Vi bruker følgende symboler:

|                         |                                                                         |
|-------------------------|-------------------------------------------------------------------------|
| $Y_{M,t}$               | for observert verdi for mars i år $t$ .                                 |
| $Y_{A,t}$               | for observert verdi for april i år $t$ .                                |
| $\hat{Y}_{M,t}^{K,ny}$  | for prekorrigert påskeverdi for mars i år $t$ ved den nye metoden.      |
| $\hat{Y}_{A,t}^{K,ny}$  | for prekorrigert påskeverdi for april i år $t$ ved den nye metoden.     |
| $\hat{Y}_{M,t}^{K,gam}$ | for prekorrigert påskeverdi for mars i år $t$ ved den gamle metoden.    |
| $\hat{Y}_{A,t}^{K,gam}$ | for prekorrigert påskeverdi for april i år $t$ , ved den gamle metoden. |

### 4.2.2 Estimeringsmetode

Vi vil korrigere mars og april observasjonene til tall som vi ville forventet å observere dersom de tre intervallene  $T_1$ ,  $T_2$  og  $T_3$  i sin helhet faller i april. I den gamle metoden estimerer vi  $\hat{Y}_{M,t}^K$  og  $\hat{Y}_{A,t}^K$  ved regresjonsanalyse for differansen mellom hhv mars og april sammenlignet med gjennomsnittet for de to månedene (se notat 97/73). I dette notatet vil vi estimere  $\hat{Y}_{M,t}^K$  og  $\hat{Y}_{A,t}^K$  fra rådata.

En modell for  $Y_{M,t}$  er

$$Y_{M,t} = C_1 + TD_t + \alpha_{M,-1}I_{-1,t} + \alpha_{M,0}I_{0,t} + \alpha_{M,1}I_{1,t} + \beta_{M,-1}J_{-1,t} + \beta_{M,0}J_{0,t} + \beta_{M,1}J_{1,t} + \epsilon_{M,t}$$

hvor  $C_1$  er konstant og  $TD_t$  er ukedagseffekter. La  $C_{M,t} = C_1 + TD_t$ . Vi får

$$Y_{M,t} = C_{M,t} + \alpha_{M,-1}I_{-1,t} + \alpha_{M,0}I_{0,t} + \alpha_{M,1}I_{1,t} + \beta_{M,-1}J_{-1,t} + \beta_{M,0}J_{0,t} + \beta_{M,1}J_{1,t} + \epsilon_{M,t} \quad (4.4)$$

$C_{M,t}$  er konstant og inneholder ukedagseffekter for mars.  $I_{i,t}$  og  $J_{i,t}$  beregnes ved (4.3).  $\alpha_{M,i}$  og  $\beta_{M,i}$  er regresjonskoeffisienter som knytter til  $I_{i,t}$  og  $J_{i,t}$ ,  $i = -1, 0, 1$ . Siden  $\epsilon_{M,t}$  ofte er korrelerte vil vi beskrive  $\epsilon_{M,t}$  med en ARIMA modell

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D\epsilon_{M,t} = \theta(B)\Theta(B^s)a_{M,t} \quad (4.5)$$

hvor  $\phi(B)$ ,  $\Phi(B^s)$ ,  $\theta(B)$  og  $\Theta(B^s)$  er polynommer av  $B$  og  $a_{M,t}$  er en hvit støy prosess.

Vi estimerer  $\hat{\alpha}_{M,\cdot}$  og  $\hat{\beta}_{M,\cdot}$  ved betingelsen

$$\hat{\alpha}_{M,-1} + \hat{\alpha}_{M,0} + \hat{\alpha}_{M,1} + \hat{\beta}_{M,-1} + \hat{\beta}_{M,0} + \hat{\beta}_{M,1} = 0 \quad (4.6)$$

dvs ingen forskyvning av aktiviteten mellom de tre intervallene  $T_1$ ,  $T_2$  og  $T_3$  og resten av mars og april.

En tilsvarende modell for april blir

$$Y_{A,t} = C_2 + TD_t + \alpha_{A,-1}(1 - I_{-1,t}) + \alpha_{A,0}(1 - I_{0,t}) + \alpha_{A,1}(1 - I_{1,t}) + \beta_{A,-1}(1 - J_{-1,t}) + \beta_{A,0}(1 - J_{0,t}) + \beta_{A,1}(1 - J_{1,t}) + \epsilon_{A,t}$$

La  $C_A = C_2 + TD_t$ . Vi får

$$Y_{A,t} = C_A + \alpha_{A,-1}(1 - I_{-1,t}) + \alpha_{A,0}(1 - I_{0,t}) + \alpha_{A,1}(1 - I_{1,t}) + \beta_{A,-1}(1 - J_{-1,t}) + \beta_{A,0}(1 - J_{0,t}) + \beta_{A,1}(1 - J_{1,t}) + \epsilon_{A,t} \quad (4.7)$$

$C_A$  er konstant og inneholder ukedagseffekter for april.  $\epsilon_{A,t}$  beskrives også ved ARIMA modellen  $\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D\epsilon_{A,t} = \theta(B)\Theta(B^s)a_{A,t}$ , hvor  $a_{A,t}$  er en hvit støy prosess. Vi estimerer  $\hat{\alpha}_{A,i}$  og  $\hat{\beta}_{A,i}$  med betingelsen

$$\hat{\alpha}_{A,-1} + \hat{\alpha}_{A,0} + \hat{\alpha}_{A,1} + \hat{\beta}_{A,-1} + \hat{\beta}_{A,0} + \hat{\beta}_{A,1} = 0$$

De korrigererte påskeverdiene for mars og april i år  $t$ ,  $\hat{Y}_{M,t}^*$  og  $\hat{Y}_{A,t}^*$  beregnes ved

$$\hat{Y}_{M,t}^* = Y_{M,t} - \hat{\alpha}_{M,-1}I_{-1,t} - \hat{\alpha}_{M,0}I_{0,t} - \hat{\alpha}_{M,1}I_{1,t} - \hat{\beta}_{M,-1}J_{-1,t} - \hat{\beta}_{M,0}J_{0,t} - \hat{\beta}_{M,1}J_{1,t}$$

og

$$\begin{aligned}\hat{Y}_{A,t}^* &= Y_{A,t} - \hat{\alpha}_{A,-1}(1 - I_{-1,t}) - \hat{\alpha}_{A,0}(1 - I_{0,t}) - \hat{\alpha}_{A,1}(1 - I_{1,t}) - \hat{\beta}_{A,-1}(1 - J_{-1,t}) \\ &\quad - \hat{\beta}_{A,0}(1 - J_{0,t}) - \hat{\beta}_{A,1}(1 - J_{1,t}) \\ &= Y_{A,t} + \hat{\alpha}_{A,-1}I_{-1,t} + \hat{\alpha}_{A,0}I_{0,t} + \hat{\alpha}_{A,1}I_{1,t} + \hat{\beta}_{A,-1}J_{-1,t} + \hat{\beta}_{A,0}J_{0,t} + \hat{\beta}_{A,1}J_{1,t}\end{aligned}$$

Merk at summen  $\hat{Y}_{M,t}^* + \hat{Y}_{A,t}^*$  kan avvike fra den totalen  $Y_{M,t} + Y_{A,t}$ . La

$$\Delta = 0.5(Y_{M,t} + Y_{A,t} - \hat{Y}_{M,t}^* - \hat{Y}_{A,t}^*)$$

Vi justerer  $\hat{Y}_{M,t}^*$  og  $\hat{Y}_{A,t}^*$  ved

$$\hat{Y}_{M,t}^K = \hat{Y}_{M,t}^* + \Delta \quad (4.8)$$

og

$$\hat{Y}_{A,t}^K = \hat{Y}_{A,t}^* + \Delta \quad (4.9)$$

Vi får

$$\hat{Y}_{M,t}^K + \hat{Y}_{A,t}^K = Y_{M,t} + Y_{A,t}$$

Vi bruker  $\hat{Y}_{M,t}^K$  og  $\hat{Y}_{A,t}^K$  heller enn  $\hat{Y}_{M,t}^*$  og  $\hat{Y}_{A,t}^*$ , som de korrigererte verdiene for påskeeffekten i mars og april. Når vi observerer bare mars men ikke april i det siste året kan vi ikke lenger beregne  $\hat{Y}_{M,t}^K$  og  $\hat{Y}_{A,t}^K$  ved (4.8) og (4.9). I dette tilfellet vil vi framskrive en verdi og bruker den som en observert verdi for april.

Vi ser at i den nye metoden

- $d_1$ ,  $d_2$  og  $d_3$  velges av brukeren.
- Rådata brukes som input i beregningen.
- $\epsilon_{.,t}$  er beskrevet med en ARIMA modell for å unngå korrelerte residualer som kan medføre ugyldige resultater.

Anvendelser av metoden presenteres i de neste avsnittene med data fra detaljvolumindeksen og produksjonsindeks for industrien.

## log transformasjon

Ved å bruke *log* transformasjonen for rådata i estimering. Ligningene (4.4) og (4.7) blir

$$\begin{aligned} Z_{M,t} &= \log(Y_{M,t}) \\ &= C_M + \alpha_{M,-1}I_{-1,t} + \alpha_{M,0}I_{0,t} + \alpha_{M,1}I_{1,t} + \beta_{M,-1}J_{-1,t} + \beta_{M,0}J_{0,t} + \\ &\quad \beta_{M,1}J_{1,t} + \epsilon_{M,t} \end{aligned}$$

$$\begin{aligned} Z_{A,t} &= \log(Y_{A,t}) \\ &= C_A + \alpha_{A,-1}(1 - I_{-1,t}) + \alpha_{A,0}(1 - I_{0,t}) + \alpha_{A,1}(1 - I_{1,t}) + \beta_{A,-1}(1 - J_{-1,t}) \\ &\quad + \beta_{A,0}(1 - J_{0,t}) + \beta_{A,1}(1 - J_{1,t}) + \epsilon_{A,t} \end{aligned}$$

Den korrigerte verdien for påskeeffekten  $\hat{Y}_{M,t}^K$  og  $\hat{Y}_{A,t}^K$  beregnes ved

$$\begin{aligned} \hat{Y}_{M,t}^K &= \exp(\hat{Z}_{M,t}^K) \\ \hat{Y}_{A,t}^K &= \exp(\hat{Z}_{A,t}^K) \end{aligned}$$

## 4.3 Eksempel 1. Detaljvolumindeksen

Tidsserien er volumindeksen for den totalen SNN52.IVL.U. Data er observert fra januar 1979 til april 1999. For å kunne beregne  $I_{.,t}$  og  $J_{.,t}$  må vi bestemme verdiene til  $d_i$ . Vi har laget et program for påskekorrigering og sesongjustering med den nye metoden. Av erfaring vet vi at handelen kan bli påvirket av påsken i en lang periode før 1. påskedag. Derfor vil vi gi  $d_1 = 7$ ,  $d_2 = 7$  og  $d_3 = 0$  i først omgang. Etter kjøring ser vi at avvikene av sesongjusterte tall i 1986 og 1997 er relative store med 5.40 og 5.29 henholdvis. Siden 1. påskedag disse to årene er i mars (30. mars), dermed kan vi sette  $d_1 = 0$ . I andre kjøring velger vi

$$\begin{aligned} d_1 &= 0, d_2 = 7, d_3 = 7 \quad \text{for } t=1986 \text{ og } 1997 \\ d_1 &= 7, d_2 = 7, d_3 = 0 \quad \text{ellers} \end{aligned} \tag{4.10}$$

Vi får da bedre resultater. Verdiene av  $I_{.,t}$  og  $J_{.,t}$  er gitt i tabell 4.1.  $I_{.,t}$  og  $J_{.,t}$  i 1979, 1981, 1984, 1987, 1990, 1992 og 1995 er null. Mars og apriltall i disse årene blir dermed ikke korrigert.

Siden 1. påskedag 1999 er 4. april og  $d_1 = 7$ ,  $d_2 = 7$ ,  $d_3 = 0$  får vi

- $T_1$ : mandag(22. mars), tirsdag, onsdag, torsdag, fredag, lørdag og søndag(28. mars).  
 $I_{-1,1999} = 1$  og  $J_{-1,1999} = 1$ .
- $T_2$ : mandag(29. mars), tirsdag, onsdag(31. mars), Skjærtorsdag(1. april), Langfredag, lørdag og påkesøndag(4. april).  $I_{0,1999} = 3/4$  og  $J_{0,1999} = 0$ .
- $T_3$ : tom mengde.  $I_{1,1999} = 0$  og  $J_{1,1999} = 0$ .

Fra tabell 4.1 ser vi at  $I_{1,t} = 0$  for alle  $t$ . Ligningene (4.4) og (4.7) blir

$$Y_{M,t} = C_M + \alpha_{M,-1}I_{-1,t} + \alpha_{M,0}I_{0,t} + \beta_{M,-1}J_{-1,t} + \beta_{M,0}J_{0,t} + \beta_{M,1}J_{1,t} + \epsilon_{M,t}$$

$$\begin{aligned} Y_{A,t} &= C_A + \alpha_{A,-1}(1 - I_{-1,t}) + \alpha_{A,0}(1 - I_{0,t}) + \beta_{A,-1}(1 - J_{-1,t}) \\ &\quad + \beta_{A,0}(1 - J_{0,t}) + \beta_{A,1}(1 - J_{1,t}) + \epsilon_{A,t} \end{aligned}$$

Tabell 4.1: Verdiene av  $I_{.,t}$  og  $J_{.,t}$ , fra 1979 til 1999

| År   | 1. påskedag | $I_{-1}$ | $I_0$  | $I_1$  | $J_{-1}$ | $J_0$  | $J_1$  |
|------|-------------|----------|--------|--------|----------|--------|--------|
| 1979 | 15-april    | 0.0000   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1980 | 6-april     | 1.0000   | 0.2500 | 0.0000 | 1.0000   | 0.0000 | 0.0000 |
| 1981 | 19-april    | 0.0000   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1982 | 11-april    | 0.5000   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1983 | 3-april     | 1.0000   | 0.7500 | 0.0000 | 1.0000   | 0.3333 | 0.0000 |
| 1984 | 22-april    | 0.0000   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1985 | 7-april     | 1.0000   | 0.0000 | 0.0000 | 1.0000   | 0.0000 | 0.0000 |
| 1986 | 30-mars     | 0.0000   | 1.0000 | 0.0000 | 0.0000   | 1.0000 | 0.5000 |
| 1987 | 19-april    | 0.0000   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1988 | 3-april     | 1.0000   | 0.7500 | 0.0000 | 1.0000   | 0.3333 | 0.0000 |
| 1989 | 26-mars     | 1.0000   | 1.0000 | 0.0000 | 1.0000   | 1.0000 | 0.0000 |
| 1990 | 15-april    | 0.0000   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1991 | 31-mars     | 1.0000   | 1.0000 | 0.0000 | 1.0000   | 1.0000 | 0.0000 |
| 1992 | 19-april    | 0.0000   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1993 | 11-april    | 0.5000   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1994 | 3-april     | 1.0000   | 0.7500 | 0.0000 | 1.0000   | 0.3333 | 0.0000 |
| 1995 | 16-april    | 0.0000   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1996 | 7-april     | 1.0000   | 0.0000 | 0.0000 | 1.0000   | 0.0000 | 0.0000 |
| 1997 | 30-mars     | 0.0000   | 1.0000 | 0.0000 | 0.0000   | 1.0000 | 0.5000 |
| 1998 | 12-april    | 0.3333   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1999 | 4-april     | 1.0000   | 0.7500 | 0.0000 | 1.0000   | 0.0000 | 0.0000 |

Ved regresjonsanalyse får vi resultater som er vist i tabell 4.2.  $\hat{\alpha}_{M,0}$ ,  $\hat{\beta}_{M,0}$ ,  $\hat{\beta}_{A,-1}$  og  $\hat{\beta}_{A,0}$  er signifikante på 5% nivået. Høye verdier for  $R^2$  viser gode modeller.

Tabell 4.2: Estimerte verdiene av  $\alpha_{M,0}$ ,  $\beta_{M,0}$ ,  $\alpha_{A,0}$  og  $\beta_{A,0}$  ved ny metode

| Var.            | Est. verdi | Std.   | t-verdi | Var.            | Est. verdi | Std.   | t-verdi |
|-----------------|------------|--------|---------|-----------------|------------|--------|---------|
| $C_M$           | 83.2855    | 2.1443 | 38.84   | $C_A$           | 80.7369    | 2.8280 | 28.55   |
| $\alpha_{M,-1}$ | 1.3231     | 2.5869 | 0.51    | $\alpha_{A,-1}$ | -4.2649    | 3.0795 | -1.38   |
| $\alpha_{M,0}$  | 12.6261    | 2.7834 | 4.54    | $\alpha_{A,0}$  | 4.2082     | 3.4240 | 1.23    |
| $\beta_{M,-1}$  | -0.6214    | 2.7584 | -0.23   | $\beta_{A,-1}$  | 7.5940     | 3.2958 | 2.30    |
| $\beta_{M,0}$   | -15.7675   | 2.9478 | -5.35   | $\beta_{A,0}$   | -8.2554    | 3.6223 | -2.28   |
| $\beta_{M,1}$   | 2.4397     | 1.2386 | 1.97    | $\beta_{A,1}$   | 0.7181     | 1.4702 | 0.49    |

$$R^2 = 0.89$$

$$R^2 = 0.80$$

### En tolkning for handelen i 1999

Den korrigerte verdien for påskeeffekten i mars 1999 er beregnet fra den observerte verdien  $Y_{M,1999}$  ved

- å trekke fra effekten av virkedager i  $T_1$ :  $-1.32 \times 1$

- å trekke fra effekten av virkedager i  $T_2$ :  $-12.62 \times 0.75$
- å legge til effekten av helligdager i  $T_1$  :  $+0.62 \times 1$

$$\begin{aligned}\hat{Y}_{M,1999}^K &= Y_{M,1999} - 1.32 \times 1 - 12.62 \times 0.75 + 0.62 \times 1 + \Delta \\ &= Y_{M,1999} - 8.32\end{aligned}$$

Tilsvarende den korrigerte verdien  $\hat{Y}_{A,1999}^K$  i april, er beregnet ved

- å trekke fra effekten av virkedager i  $T_2$ :  $-4.20 \times 0.25$
- å trekke fra effekten av helligdager i  $T_3$ :  $-0.71 \times 1$
- å legge til effekten av helligdager i  $T_2$  :  $+8.25 \times 1$

$$\begin{aligned}\hat{Y}_{A,1999}^K &= Y_{A,1999} + 4.26 \times 0 - 4.20 \times 0.25 - 7.59 \times 0 + 8.25 \times 1 - 0.71 \times 1 + \Delta \\ &= Y_{A,1999} - 4.20 \times 0.25 + 8.25 \times 1 - 0.71 \times 1 + \Delta \\ &= Y_{A,1999} + 8.32\end{aligned}$$

Konklusjon:

- Det er en forskyvning av handelen fra april til mars 1999 på grunn av påskeeffekten.
- Handelen har økt på virkedager og gått ned på helligdager.

$\hat{Y}_{M,99}^K$  og  $\hat{Y}_{A,99}^K$  brukes som utgangspunkt for videre analyse med X-12-ARIMA.

Tabell 4.6 viser korrigerte verdier for påskeeffekten og sesongjusterte tall som betegnes med  $J_{.,t}^{ny}$ ,  $J_{.,t}^{gam}$  og  $J_{.,t}^{X12}$  for den nye, gamle metoden og X-12-ARIMA, henholdsvis. Vi får

- omtrent samme verdiene for  $J_{.,t}^{ny}$  og  $J_{.,t}^{gam}$ , bortsett fra 1989.
- X-12-ARIMA gir dårlige sesongjusterte tall i mars og april for detaljvolumindeksen. Endringene er veldig store i 1980, 1983, 1986-1988, 1991, 1993-1995, 1997-1999.

### Stabilitet

Vi kjører opplegget for delmengder av tidsserien:  $[jan.79, apr.96]$ ,  $[jan.79, apr.97]$ ,  $[jan.79, apr.98]$  og  $[jan.79, apr.99]$  med samme verdiene av  $d_1$ ,  $d_2$  og  $d_3$  i (4.10). Sesongjusterte tall av mars og april i 1996, 1997, 1998 og 1999 er listet ut i tabell 4.3.

Vi ser at endringene i samme tidspunkter er veldig små. For eksempel, relative endringen fra den laveste og høyeste verdien i mars 1996 er

$$\frac{102.82 - 102.13}{102.13} \times 100\% = 0.68\%$$

Valget av  $d_1$ ,  $d_2$  og  $d_3$  i (4.10) fungerer for påsekkorrigeringsen av detaljvolumindeksen.

Tabell 4.3: Sesongjusterte tall i mars og april av fire tidsserier

|          | Mars   |        |        |        | April  |        |        |        |
|----------|--------|--------|--------|--------|--------|--------|--------|--------|
|          | 1996   | 1997   | 1998   | 1999   | 1996   | 1997   | 1998   | 1999   |
| ,apr.96] | 102.68 |        |        |        | 101.11 |        |        |        |
| ,apr.97] | 102.13 | 105.12 |        |        | 100.53 | 106.42 |        |        |
| ,apr.98] | 102.52 | 105.66 | 110.66 |        | 100.80 | 106.69 | 111.94 |        |
| ,apr.99] | 102.82 | 106.17 | 111.63 | 115.01 | 101.09 | 107.06 | 112.27 | 116.42 |
| rel.end. | 0.68%  | 0.99%  | 0.88%  | .      | 0.58%  | 0.60%  | 0.92%  | .      |

### Når vi observerer tidsserien bare til mars 1999

Vi ser ofte ved bruk av X-12-ARIMA at endringen av sesongjusterte tall for mars er stor når apriltall kommer i tillegg. Dette kan skyldes at programmet mangler apriltall i beregningen for påskeeffekten. Vi antar nå at vi observerer tidsserien bare til mars 1999. April 1999 blir estimert ved en ARIMA modell, vi får  $\hat{Y}_{apr.99} = 107.96$ , mye høyere enn det faktiske tallet  $Y_{apr.99} = 98.82$ . Vi kjører opplegget med samme  $d_1$ ,  $d_2$  og  $d_3$  fra (4.10). Sesongjusterte tall i mars og april av to tidsserier [*jan.79*, *mar.99*] og [*jan.79*, *apr.99*] er listet ut i tabell 4.4. Vi får omtrent samme tallene bortsett fra april 1999. I notat 97/73 har vi også en metode for å beregne  $\hat{Y}_{M,99}^K$  i dette tilfellet. Mars og apriltallene inntil året før brukes for å estimere  $\hat{\alpha}_i$  og  $\hat{\beta}_i$ . Den korrigerte verdien til mars 1999 blir da

$$\hat{Y}_{M,99}^K = Y_{M,99} - \hat{\alpha}_{-1}I_{-1,99} - \hat{\alpha}_0I_{0,99} - \hat{\alpha}_1I_{1,99} + \hat{\beta}_{-1}J_{-1,99} - \hat{\beta}_0J_{0,99} - \hat{\beta}_1J_{1,99}$$

Vi får  $J_{M,99}^{gam} = 113.44$ . Endringene i prosent i mars 1999 av sesongjusterte tall som er beregnet fra to tidsserier [*jan.79*, *mar.99*] og [*jan.79*, *april.99*], er

- gammel metode :  $(114.97 - 113.44)/113.44 \times 100\% = 1.35\%$
- ny metode :  $(115.01 - 115.57)/115.57 \times 100\% = -0.48\%$

Vi får større endring ved bruk av metoden i notat 97/73.

### Konklusjon

- Metoden fungerer veldig bra og mer fleksibel enn den gamle metoden for detaljvolumindeksen.
- Vi får rimelige og stabile sesongjusterte tall.
- Endringen i mars av sesongjusterte tall i det siste året når apriltall legges til er lavere enn for den gamle metoden.
- Metoden kan anvendes for andre tidsserier.
- Metoden kan beskrive variasjonen av påskeeffekten over tid ved forskjellige valg av verdier av  $d_i$ .

Tabell 4.4: Sesongjusterte tall i mars og april, beregnet fra to tidsserier

| år   | Rådata |        | gam. metode      |        | ny metode        |        |                  |        |
|------|--------|--------|------------------|--------|------------------|--------|------------------|--------|
|      |        |        | [jan.79, mar.99] |        | [jan.79, mar.99] |        | [jan.79, apr.99] |        |
|      | mars   | april  | mars             | april  | mars             | april  | mars             | april  |
| 1979 | 85.15  | 81.60  | 93.64            | 91.97  | 93.65            | 92.02  | 93.60            | 91.95  |
| 1980 | 84.65  | 80.66  | 91.65            | 93.14  | 91.88            | 93.21  | 91.51            | 93.19  |
| 1981 | 81.06  | 85.09  | 92.52            | 92.76  | 92.54            | 92.90  | 92.51            | 92.74  |
| 1982 | 82.76  | 86.26  | 93.19            | 93.30  | 93.17            | 93.44  | 93.29            | 93.14  |
| 1983 | 85.08  | 76.96  | 89.23            | 88.55  | 89.50            | 88.46  | 89.07            | 88.67  |
| 1984 | 82.44  | 80.18  | 90.45            | 89.75  | 90.42            | 89.81  | 90.46            | 89.75  |
| 1985 | 83.05  | 80.83  | 90.98            | 90.64  | 90.94            | 90.70  | 91.03            | 90.58  |
| 1986 | 81.39  | 91.85  | 94.84            | 96.19  | 94.86            | 96.29  | 94.87            | 96.17  |
| 1987 | 80.39  | 86.99  | 90.15            | 92.77  | 90.04            | 92.81  | 90.20            | 92.76  |
| 1988 | 87.63  | 78.98  | 89.76            | 88.79  | 89.92            | 88.62  | 89.64            | 88.91  |
| 1989 | 80.27  | 81.09  | 86.91            | 86.64  | 86.82            | 86.76  | 86.99            | 86.59  |
| 1990 | 85.11  | 81.08  | 89.75            | 89.04  | 89.70            | 89.02  | 89.75            | 89.07  |
| 1991 | 82.47  | 84.76  | 90.63            | 90.02  | 90.63            | 90.03  | 90.69            | 89.97  |
| 1992 | 82.74  | 85.73  | 90.40            | 91.68  | 90.38            | 91.75  | 90.42            | 91.75  |
| 1993 | 85.15  | 86.52  | 92.15            | 92.37  | 92.09            | 92.48  | 92.28            | 92.31  |
| 1994 | 94.56  | 84.08  | 96.25            | 95.80  | 96.46            | 95.72  | 96.13            | 96.01  |
| 1995 | 91.68  | 87.55  | 97.60            | 97.46  | 97.56            | 97.59  | 97.61            | 97.54  |
| 1996 | 95.95  | 89.47  | 102.81           | 101.12 | 102.66           | 100.99 | 102.82           | 101.09 |
| 1997 | 93.28  | 100.34 | 106.27           | 107.12 | 106.13           | 106.66 | 106.17           | 107.06 |
| 1998 | 99.94  | 103.70 | 111.68           | 112.46 | 111.19           | 111.56 | 111.63           | 112.27 |
| 1999 | 113.62 | 107.96 | 113.44           | .      | 115.57           | .      | 115.01           | 116.42 |

Tabell 4.5: Verdiene av  $I_{-,t}$  og  $J_{-,t}$ , fra 1986 til 1999

| År   | 1. påskedag | $I_{-1}$ | $I_0$  | $I_1$  | $J_{-1}$ | $J_0$  | $J_1$  |
|------|-------------|----------|--------|--------|----------|--------|--------|
| 1986 | 30-mars     | 1.0000   | 1.0000 | 0.0000 | 1.0000   | 1.0000 | 0.5000 |
| 1987 | 19-april    | 0.0000   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1988 | 3-april     | 1.0000   | 0.7500 | 0.0000 | 1.0000   | 0.3333 | 0.0000 |
| 1989 | 26-mars     | 0.0000   | 0.0000 | 0.0000 | 0.0000   | 1.0000 | 0.0000 |
| 1990 | 15-april    | 0.0000   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1991 | 31-mars     | 0.0000   | 0.0000 | 0.0000 | 0.0000   | 1.0000 | 0.0000 |
| 1992 | 19-april    | 0.0000   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1993 | 11-april    | 0.5000   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1994 | 3-april     | 1.0000   | 0.7500 | 0.0000 | 1.0000   | 0.3333 | 0.0000 |
| 1995 | 16-april    | 0.0000   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1996 | 7-april     | 0.0000   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1997 | 30-mars     | 0.0000   | 0.0000 | 0.0000 | 0.0000   | 1.0000 | 0.0000 |
| 1998 | 12-april    | 0.0000   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |
| 1999 | 4-april     | 0.7500   | 0.0000 | 0.0000 | 0.0000   | 0.0000 | 0.0000 |

Tabell 4.6: Sesongjusterte tall av mars og april i den gamle og nye metoden

| år   | Korrigerte verdier |           |                        |                        |                         |                         | Sesongjusterte tall |                |                 |                 |                 |                 |
|------|--------------------|-----------|------------------------|------------------------|-------------------------|-------------------------|---------------------|----------------|-----------------|-----------------|-----------------|-----------------|
|      | $Y_{M,t}$          | $Y_{A,t}$ | $\hat{Y}_{M,t}^{K,ny}$ | $\hat{Y}_{A,t}^{K,ny}$ | $\hat{Y}_{M,t}^{K,gam}$ | $\hat{Y}_{A,t}^{K,gam}$ | $J_{M,t}^{ny}$      | $J_{A,t}^{ny}$ | $J_{M,t}^{gam}$ | $J_{A,t}^{gam}$ | $J_{M,t}^{X12}$ | $J_{A,t}^{X12}$ |
| 1979 | 85.15              | 81.60     | 85.15                  | 81.60                  | 85.15                   | 81.60                   | 93.60               | 91.95          | 93.44           | 92.15           | 92.71           | 92.02           |
| 1980 | 84.65              | 80.66     | 80.53                  | 84.77                  | 80.74                   | 84.56                   | 91.51               | 93.19          | 91.58           | 93.33           | 94.91           | 88.93           |
| 1981 | 81.06              | 85.09     | 81.06                  | 85.09                  | 81.06                   | 85.09                   | 92.51               | 92.74          | 92.43           | 92.99           | 91.49           | 92.17           |
| 1982 | 82.76              | 86.26     | 83.49                  | 85.52                  | 84.08                   | 84.93                   | 93.29               | 93.14          | 93.99           | 92.75           | 91.43           | 94.24           |
| 1983 | 85.08              | 76.96     | 80.75                  | 81.28                  | 80.92                   | 81.11                   | 89.07               | 88.67          | 89.03           | 88.79           | 92.29           | 84.74           |
| 1984 | 82.44              | 80.18     | 82.44                  | 80.18                  | 82.44                   | 80.18                   | 90.46               | 89.75          | 89.97           | 89.95           | 89.32           | 90.04           |
| 1985 | 83.05              | 80.83     | 81.03                  | 82.84                  | 81.19                   | 82.68                   | 91.03               | 90.58          | 90.74           | 90.46           | 91.89           | 89.76           |
| 1986 | 81.39              | 91.85     | 84.19                  | 89.04                  | 85.20                   | 88.03                   | 94.87               | 96.17          | 95.25           | 95.02           | 93.66           | 96.56           |
| 1987 | 80.39              | 86.99     | 80.39                  | 86.99                  | 80.39                   | 86.99                   | 90.20               | 92.76          | 89.63           | 92.40           | 88.39           | 93.87           |
| 1988 | 87.63              | 78.98     | 83.30                  | 83.30                  | 83.47                   | 83.13                   | 89.64               | 88.91          | 89.31           | 88.26           | 92.01           | 86.56           |
| 1989 | 80.27              | 81.09     | 81.84                  | 79.51                  | 79.17                   | 82.18                   | 86.99               | 86.59          | 83.99           | 88.87           | 87.52           | 85.73           |
| 1990 | 85.11              | 81.08     | 85.11                  | 81.08                  | 85.11                   | 81.08                   | 89.75               | 89.07          | 89.42           | 88.64           | 88.87           | 90.14           |
| 1991 | 82.47              | 84.76     | 84.04                  | 83.18                  | 83.96                   | 83.26                   | 90.69               | 89.97          | 90.46           | 89.84           | 92.47           | 89.03           |
| 1992 | 82.74              | 85.73     | 82.74                  | 85.73                  | 82.74                   | 85.73                   | 90.42               | 91.75          | 90.15           | 91.75           | 89.75           | 91.78           |
| 1993 | 85.15              | 86.52     | 85.88                  | 85.78                  | 86.47                   | 85.19                   | 92.28               | 92.31          | 92.67           | 91.89           | 90.51           | 93.97           |
| 1994 | 94.56              | 84.08     | 90.23                  | 88.40                  | 90.40                   | 88.23                   | 96.13               | 96.01          | 95.75           | 96.24           | 98.70           | 92.84           |
| 1995 | 91.68              | 87.55     | 91.68                  | 87.55                  | 91.68                   | 87.55                   | 97.61               | 97.54          | 97.19           | 97.87           | 95.37           | 98.34           |
| 1996 | 95.95              | 89.47     | 93.93                  | 91.48                  | 94.09                   | 91.32                   | 102.82              | 101.09         | 102.40          | 101.45          | 102.73          | 100.64          |
| 1997 | 93.28              | 100.34    | 96.08                  | 97.53                  | 97.09                   | 96.52                   | 106.17              | 107.06         | 106.55          | 106.66          | 104.57          | 107.03          |
| 1998 | 99.94              | 103.70    | 100.43                 | 103.20                 | 100.82                  | 102.81                  | 111.63              | 112.27         | 111.47          | 112.41          | 107.94          | 113.88          |
| 1999 | 113.83             | 98.82     | 105.50                 | 107.14                 | 105.82                  | 106.83                  | 115.01              | 116.42         | 114.97          | 116.61          | 120.42          | 109.52          |

indeksen  $M$  står for mars og  $A$  for april,  $t$  for årstall

## 4.4 Eksempel 2. Produksjonsindeks for industrien

Tidsserien er den totale produksjonsindeksen for industrien. Data er observert fra januar 1986 til august 1999 med 164 observasjoner. Ved å estimere påskeeffekten med default verdier fra X-12-ARIMA får vi  $\hat{d} = 8$ . Vi velger  $d_1 = 7$ ,  $d_2 = 7$  og  $d_3 = 7$  i første kjøring. Det gir store endringer av sesongjusterte tall fra mars til april i 1991, 1997 og 1999. Vi vil velge andre verdier for  $d_i$  for disse årene. 1. påskedag faller i mars 1986, 1989, 1991 og 1997. Derfor behøver vi ikke å gi store verdier til  $d_i$  for at mars og april tall blir korrigert av påskeeffekten. For 1999 (1. påskedag 4. april) må vi velge  $d_1$  og  $d_2$  slik at påskedagene ligger i mars. Verdiene av  $d_i$  blir

$$\begin{aligned} d_1 = 7, d_2 = 7, d_3 = 7 & \text{ for } t=1986-1988, 1990 \text{ og } 1992-1995 \\ d_1 = 0, d_2 = 1, d_3 = 0 & \text{ for } t=1989, 1991 \text{ og } 1996-1998 \\ d_1 = 7, d_2 = 0, d_3 = 0 & \text{ for } t=1999 \end{aligned} \quad (4.11)$$

Ved erfaring viser det at av og til må vi gi samme verdi til  $d_i$  for flere år, for eksempel,  $d_1 = 0$ ,  $d_2 = 1$ ,  $d_3 = 0$  for  $t=1996-1998$ , for å få bedre sesongjusterte tall i 1997.  $I_{.,t}$  og  $J_{.,t}$  for hele serien er gitt i tabell 4.5.  $I_{.,t}$  og  $J_{.,t}$  i 1987, 1990, 1992, 1995 og 1998 er lik null, derfor blir mars og april tall ikke korrigert for disse årene. Vi har bare 14 observasjoner tilsvarende med 14 år for å estimere  $\hat{\alpha}_i$  og  $\hat{\beta}_i$ . Data er transformert med  $\log$  i beregningen. Tabell 4.7 viser estimerte verdier for parametre ved regresjonsanalyse.  $\hat{C}_M$ ,  $\hat{\alpha}_{M,0}$ ,  $\hat{\beta}_{M,0}$ ,  $\hat{\beta}_{M,1}$ ,  $\hat{C}_A$ ,  $\hat{\beta}_{A,0}$  og  $\hat{\beta}_{A,1}$  er signifikante på 5% nivået. Høy  $R^2$  viser god tilpasning av modellen. Sesongjusterte tall  $J_{.,t}$  fra den nye metoden og X-12-ARIMA er gitt i tabell 4.8.

Tabell 4.7: Estimerte verdiene av  $\alpha_{M,0}$ ,  $\beta_{M,0}$ ,  $\alpha_{A,0}$  og  $\beta_{A,0}$  ved ny metode

| Var.            | Est. verdi | Std.   | t-verdi |
|-----------------|------------|--------|---------|
| $C_M$           | 4.5688     | 0.0547 | 83.45   |
| $\alpha_{M,-1}$ | -0.0298    | 0.0428 | -0.70   |
| $\alpha_{M,0}$  | -0.6498    | 0.3241 | -2.00   |
| $\beta_{M,-1}$  | 0.4769     | 0.2497 | 1.91    |
| $\beta_{M,0}$   | -0.1319    | 0.0161 | -8.18   |
| $\beta_{M,1}$   | 0.3346     | 0.0771 | 4.34    |

| Var.            | Est. verdi | Std.   | t-verdi |
|-----------------|------------|--------|---------|
| $C_A$           | 4.4029     | 0.0424 | 103.66  |
| $\alpha_{A,-1}$ | 0.0080     | 0.0341 | 0.24    |
| $\alpha_{A,0}$  | -0.0284    | 0.2514 | -0.11   |
| $\beta_{A,-1}$  | -0.0350    | 0.1935 | -0.18   |
| $\beta_{A,0}$   | -0.1030    | 0.0126 | -8.14   |
| $\beta_{A,1}$   | 0.1584     | 0.0602 | 2.63    |

Vi får omtrent samme sesongjusterte tall bortsett fra 1999. Endringen fra mars til april er stor ved X-12-ARIMA. Vi får  $J_{M,t}^{X12} = 112.65$  og  $J_{A,t}^{X12} = 103.12$ , vs  $J_{M,t}^{ny} = 107.16$  og  $J_{A,t}^{ny} = 108.26$ .

Tabell 4.8: Sesongjusterte tall av mars og april i den gamle og nye metoden

| år   | $Y_{M,t}$ | $Y_{A,t}$ | $\hat{Y}_{M,t}^K$ | $\hat{Y}_{A,t}^K$ | $J_{M,t}^{ny}$ | $J_{A,t}^{ny}$ | $J_{M,t}^{X12}$ | $J_{A,t}^{X12}$ |
|------|-----------|-----------|-------------------|-------------------|----------------|----------------|-----------------|-----------------|
| 1986 | 82.60     | 89.60     | 93.43             | 79.20             | 88.29          | 85.83          | 88.96           | 86.23           |
| 1987 | 98.90     | 83.90     | 98.90             | 83.90             | 91.98          | 90.26          | 92.92           | 90.35           |
| 1988 | 92.60     | 89.60     | 100.66            | 82.42             | 91.35          | 90.65          | 91.98           | 90.76           |
| 1989 | 87.40     | 90.00     | 98.29             | 80.02             | 89.52          | 89.64          | 89.80           | 89.89           |
| 1990 | 95.90     | 82.50     | 95.90             | 82.50             | 88.87          | 90.56          | 89.47           | 90.90           |
| 1991 | 82.90     | 93.10     | 93.23             | 82.78             | 89.01          | 88.67          | 89.43           | 88.71           |
| 1992 | 96.10     | 83.70     | 96.10             | 83.70             | 89.98          | 89.11          | 90.28           | 88.63           |
| 1993 | 99.40     | 87.10     | 99.94             | 86.62             | 91.65          | 92.02          | 90.86           | 92.69           |
| 1994 | 95.40     | 96.10     | 103.70            | 88.40             | 94.45          | 95.86          | 94.26           | 95.69           |
| 1995 | 110.10    | 89.60     | 110.10            | 89.60             | 100.38         | 98.86          | 99.76           | 99.19           |
| 1996 | 106.40    | 97.40     | 106.40            | 97.40             | 101.25         | 102.93         | 103.88          | 100.62          |
| 1997 | 98.00     | 111.80    | 110.21            | 99.40             | 104.56         | 104.87         | 104.01          | 105.12          |
| 1998 | 117.80    | 103.90    | 117.80            | 103.90            | 109.61         | 108.54         | 108.58          | 109.45          |
| 1999 | 116.70    | 104.70    | 117.66            | 103.84            | 107.16         | 108.26         | 112.65          | 103.12          |

### Stabilitet

For å vurdere stabiliteten av metoden vil vi kjøre opplegget med samme verdiene til  $d_i$  for delmengder av tidsserien:  $[jan.86, aug.98]$ ,  $[jan.86, mar.99]$  og sammenligne med den originale serien  $[jan.86, aug.99]$ . Sesongjusterte tall er listet i tabell 4.9. Vi ser at endringene av sesongjusterte fra mars til april er små. Det viser en stabilitet og verdiene av  $d_i$  i (4.11) er gode for produksjonsindeks for industrien.

Tabell 4.9: Sesongjusterte tall

| år   | $[jan.86, aug.98]$ |           | $[jan.86, mar.99]$ |           | $[jan.86, aug.99]$ |           |
|------|--------------------|-----------|--------------------|-----------|--------------------|-----------|
|      | $J_{M,t}$          | $J_{A,t}$ | $J_{M,t}$          | $J_{A,t}$ | $J_{M,t}$          | $J_{A,t}$ |
| 1986 | 88.09              | 85.99     | 88.18              | 85.96     | 88.29              | 85.83     |
| 1987 | 92.00              | 90.32     | 91.96              | 90.30     | 91.98              | 90.26     |
| 1988 | 91.46              | 90.58     | 91.43              | 90.68     | 91.35              | 90.65     |
| 1989 | 89.57              | 89.66     | 89.61              | 89.62     | 89.52              | 89.64     |
| 1990 | 88.74              | 90.62     | 88.78              | 90.59     | 88.87              | 90.56     |
| 1991 | 89.03              | 88.65     | 89.03              | 88.69     | 89.01              | 88.67     |
| 1992 | 89.92              | 89.20     | 89.86              | 89.17     | 89.98              | 89.11     |
| 1993 | 91.82              | 91.87     | 91.87              | 91.86     | 91.65              | 92.02     |
| 1994 | 94.51              | 95.85     | 94.46              | 95.93     | 94.45              | 95.86     |
| 1995 | 100.38             | 98.97     | 100.44             | 98.88     | 100.38             | 98.86     |
| 1996 | 101.25             | 103.04    | 101.37             | 102.94    | 101.25             | 102.93    |
| 1997 | 104.37             | 105.23    | 104.63             | 104.96    | 104.56             | 104.87    |
| 1998 | 109.58             | 109.07    | 109.93             | 108.51    | 109.61             | 108.54    |
| 1999 | .                  | .         | 108.12             | 110.38    | 107.16             | 108.26    |

## **4.5 Konklusjon**

Den nye metoden behandler påskeeffekten mer fleksibelt enn den gamle metoden. Den gir gode resultater og kan brukes for tidsserier på flere områder.

## Kapittel 5

# Sesongjustering for ordretilganger i industrinæringen

Dinh Quang Pham

### 5.1 Innledning

Tidsseriene som vi analyserer her er den totale verdien av ordrer i industrien ( $TOT$ ), dvs summen av eksporten ( $EKSP$ ) og hjemmemarkedet ( $HJEM$ ). Dataene er i form av verditall og indekser med 1995 som basisår. Ved å sesongjustere med default verdier på kommandoene i inputfilen får vi dårlige sesongjusteringer. Endringene av sesongjusterte tall fra kvartal til kvartal er mye større enn grenseverdier for en god sesongdekomponering. Dette skyldes store fluktuasjoner i de siste observasjonene. I andre kjøring erstattes default filteret ( $3 \times 5$ ) med et lengre filter ( $3 \times 9$ ) i beregningen av sesongkomponenten. Vi får gode resultater for ( $TOT$ ) og ( $EKSP$ ), men fortsatt ikke for ( $HJEM$ ). Siden verdiene i tidsserien ( $TOT$ ) er summen av ( $EKSP$ ) og ( $HJEM$ ), kan vi beregne de sesongjusterte tallene av ( $HJEM$ ) fra de to andre. I de neste avsnittene vil vi presentere:

- resultater når vi kjører X-12-ARIMA med (i) default og (ii) verdiene som vi velger selv på kommandoene.
- sesongjustering ved å bruke indekser og verdier som input.
- hvilken måte å beregne sesongjusterte tall for ( $HJEM$ ).

#### Symboler.

|             |                                                                      |
|-------------|----------------------------------------------------------------------|
| $v\_TOT_t$  | den totale ordretilgangen i industrien, verdi i tidspunkt $t$        |
| $v\_EKSP_t$ | den totale ordretilgangen for eksport, verdi i tidspunkt $t$         |
| $v\_HJEM_t$ | den totale ordretilgangen for hjemmemarkedet, verdi i tidspunkt $t$  |
| $i\_TOT_t$  | den totale ordretilgangen i industrien, indeks i tidspunkt $t$       |
| $i\_EKSP_t$ | den totale ordretilgangen for eksport, indeks i tidspunkt $t$        |
| $i\_HJEM_t$ | den totale ordretilgangen for hjemmemarkedet, indeks i tidspunkt $t$ |

X-12-ARIMA er brukt for sesongjustering. For å kunne kjøre programmet må vi lage en inputfil som ser slik ut

```
series{title="Total for ordreindustri"
      start=1986.1
      period=4
      file="Total2.dat"
      format="(18x,f9.2)"
      decimals=2}
transform{function=log}
automdl{file="my.mdl"
        identify=all}
forecast{save=(fct)}
outlier{}
```

series, transform, automdl... er kommandoene som X-12-ARIMA leser inn for å behandle data. transform{function=log} betyr at multiplikativ modell skal brukes for sesongdekomponering. Her angir vi ikke kritisk verdi som er  $|t|$ -verdi på forhånd for å identifisere ekstremverdier. Verdien som vi kaller default verdi blir valgt av programmet og gitt i tabell 6.22 i X-12-ARIMA manualen. Den varierer med lengden av tidsserie.

## 5.2 Sesongjustering for tidsseriene $i_{TOT_t}$ , $i_{EKSP_t}$ og $i_{HJEM_t}$ .

### 5.2.1 Totalen, $i_{TOT_t}$ .

#### Første kjøring.

Vi bruker bare default verdier i estimering for ekstremverdier og ukedagseffekter. Vi får følgende resultater:

- ARIMA modell.

ARIMA (0 1 1)(0 1 1) er valgt med

$$\begin{aligned}\hat{\theta} &= 0.3581, & Std(\hat{\theta}) &= 0.1287 \\ \hat{\Theta} &= 0.7884, & Std(\hat{\Theta}) &= 0.0972\end{aligned}$$

- 11 mål for kvaliteten.

$$M7 = 0.365, \quad M8 = 1.003, \quad M10 = 1.394, \quad M11 = 1.307$$

de andre målene er mindre enn 1.

$M7$  er et mål for *grad av bevegelig sesongmønster i forhold til stabilt sesongmønster*. Når  $M7 > 1$  sier vi at sesongmønsteret ikke er identifiserbart og vi vil ikke sesongjustere for denne tidsserien.

$M8$  er et mål for *grad av fluktuasjoner i sesongkomponenten gjennom hele serien*.  $M8 > 1$  viser at fluktuasjonen er så stor at  $S_t$  ikke er stabil lenger.

Dermed får en et dårlig estimat for den sesongjusterte serien. Her ligger  $M8$  ved grenseverdien.

$M10$  er et mål for *grad av fluktuasjoner i  $S_t$  i de siste årene*. Denne testen er den samme som  $M8$ , men bare for de siste årene.

$M11$  er et mål for *grad av lineær bevegelse i  $S_t$  i de siste årene*.  $M11 > 1$  viser at fluktuasjoner i  $S_t$  ikke er tilfeldige.

Disse testene viser store fluktuasjoner i den sesongjusterte serien de siste årene. Selv om verdien av  $Q$  er mindre enn 1, ser sesongdekomponeringen dårlig ut.

#### - Slidingspans.

Vi får:

Prosentandelen av kvartaler hvor sesongkomponenten er definert som upålitelig er 6 av 39 (15.4%). Tallet er for høyt for en god sesongkomponent.

Prosentandelen av endringer fra kvartal til kvartal i den sesongjusterte serien som er upålitelig, er 14 av 38 (36.4%). Tallet er også vurdert som for høyt.

**Konklusjon.** Vi har for dårlig sesongjustering av totalen.

#### Andre kjøring.

Vi ser fra den første kjøringen at  $M8$ ,  $M10$  og  $M11$  er større enn 1. Dette skyldes høye fluktuasjoner av sesongkomponenten i slutten av serien. En måte å dempe variasjoner, er å bruke et lengre filter i beregningen av sesongkomponenten enn default filteret. Vi velger  $(3 \times 9)$  filteret. Vi har også testet med  $(3 \times 15)$  filteret, men det gir dårligere resultater. Inputfilen ser ut slik

```
series{title="Total for ordreindustri"
      start=1986.1
      period=4
      file="Total2.dat"
      format="(18x,f9.2)"
      decimals=2}
transform{function=log}
regression{variables=(td)}
automdl{file="my.mdl"
        identify=all}
forecast{save=(fct)}
outlier{}
estimate{}
x11{seasonalma=s3x9
    save=(d10 d11 d12 d13)}
slidingspans{print=(summary sfspans chngspans sasp ans ychngspans
                    tdspans indpercent indyypercent) fixmdl=no}
```

Vi får følgende resultater:

- ARIMA modell.

ARIMA(0,1,1)(0,1,1) modellen er valgt av programmet med  $\hat{\theta} = 0.3632$  og  $Std(\hat{\theta}) = 0.1289$ ,  $\hat{\Theta} = 0.5782$  og  $Std(\hat{\Theta}) = 0.1198$ . De to parametre er signifikante på 5% nivået.

Gjennomsnittlig absolutt prosent feil for framskriving

siste år: 15.49 forrige år: 6.22 to år tidligere: 3.97

siste tre år: 8.56 som er mindre enn grenseverdien (15%) for å forkaste en ARIMA modell.

- Intervensjoner.

Nivåskiftet i 3. kvartal 1998 som medfører en kraftig endring i trenden fra dette tidspunktet (se figuren 5.2c).

- Framskrivinger.

Vises i tabell 5.1 med 95% konfidensintervaller. Det ser en nedgang i de fremskrevne indeksene for den totale ordretilgangen i industrien.

Tabell 5.1: *Framskrivinger og 95% konfidensintervaller*

| <i>Dato</i> | <i>Nedre</i> | <i>Verdi</i> | <i>Øvre</i> |
|-------------|--------------|--------------|-------------|
| 4.kv 1999   | 110.48       | 129.67       | 152.19      |
| 1.kv 2000   | 109.41       | 130.62       | 155.93      |
| 2.kv 2000   | 104.99       | 128.36       | 156.91      |
| 3.kv 2000   | 99.83        | 125.87       | 158.70      |

- *F*-test for stabilt sesongmønster og sesongbevegelser.

Stabilt sesongmønster er signifikant på 0.1% nivået.

Sesongbevegelser er ikke signifikant på 5% nivået.

- 11 mål for kvalitet.

Det vises i figur 5.1. Vi får

$M1 = 1.336$ ,  $M3 = 1.194$  og  $M4 = 0.967$ . En tolkning av disse målene er at trenden  $T_t$  og den irregulære komponenten  $I_t$  blir vanskelig å skille perfekt fra hverandre i den tredje runden på grunn av store variasjoner i  $I_t$ . Dessuten er verdien for  $M4$  høy. Dette vil si at  $I_t$  kan ikke være uavhengig tilfeldig prosess. Dermed kan alle tester som er basert på  $I_t$  være ugyldige.

$M10 = 1.024$ ,  $M11 = 1.024$ . Dette er målene for grad av fluktusjoner og bevegelse i  $S_t$  i de siste årene. Når  $M10$  og  $M11$  er større enn 1 vil det si at fluktusjonene i de siste årene er så store at  $S_t$  ikke er stabil lenger, og at variasjonene av denne komponenten ikke er tilfeldige.

### F 3. Monitoring and Quality Assessment Statistics

All the measures below are in the range from 0 to 3 with an acceptance region from 0 to 1.

1. The relative contribution of the irregular over one quarter span (from Table F 2.B). M1 = 1.336
2. The relative contribution of the irregular component to the stationary portion of the variance (from Table F 2.F). M2 = 0.535
3. The amount of quarter to quarter change in the irregular component as compared to the amount of quarter to quarter change in the trend-cycle (from Table F2.H). M3 = 1.194
4. The amount of autocorrelation in the irregular as described by the average duration of run (Table F 2.D). M4 = 0.967
5. The number of quarters it takes the change in the trend-cycle to surpass the amount of change in the irregular (from Table F 2.E). M5 = 0.743
6. The amount of year to year change in the irregular as compared to the amount of year to year change in the seasonal (from Table F 2.H). M6 = 0.292
7. The amount of moving seasonality present relative to the amount of stable seasonality (from Table F 2.I). M7 = 0.387
8. The size of the fluctuations in the seasonal component throughout the whole series. M8 = 0.721
9. The average linear movement in the seasonal component throughout the whole series. M9 = 0.389
10. Same as 8, calculated for recent years only. M10 = 1.024
11. Same as 9, calculated for recent years only. M11 = 1.024

\*\*\* ACCEPTED \*\*\* at the level 0.78  
 \*\*\* Check the 4 above measures which failed.  
 \*\*\* Q (without M2) = 0.81 CONDITIONALLY ACCEPTED.

Figur 5.1: Kvalitets mål i X-11-ARIMA

### - Slidingspans.

Blir utført med tre delvis overlappende delmengder

span 1: fra 1. kvartal 1986 til 3. kvartal 1997  
 span 2: fra 1. kvartal 1987 til 3. kvartal 1998  
 span 3: fra 1. kvartal 1988 til 3. kvartal 1999

Testene for sesongmønster og sesongbevegelser for hvert span er vist nedenfor i tabellen

|                       | F-test |        |        |                        |
|-----------------------|--------|--------|--------|------------------------|
|                       | span 1 | span 2 | span 3 |                        |
| Stabilt sesongmønster | 47.86  | 50.74  | 48.16  | signifikant på 5%      |
| Sesongbevegelse       | 0.89   | 0.74   | 0.58   | ikke signifikant på 5% |
| M7                    | 0.32   | 0.30   | 0.30   | tilfredsstillende      |

Andelene av kvartaler i prosent som er definert som ustabile:

For sesongkomponenten : 2 av 47 (4.3%).  
 For sesongjusterte tall : 5 av 47 (10.6%).  
 For endring fra kvartal til kvartal  
 i sesongjusterte tall : 7 av 46 (15.2%).

Tallene ligger langt under grenseverdiene for å forkaste en sesongjustering.

- Figurer.

Rådata og framskrivinger er vist i figuren 5.2a. Sesongjusterte tall, trend er plottet i figuren 5.2b og c. Det er en kraftig nedgang i trenden fra 2. kvartal 1998 (149.40) til 3. kvartal 1998 (116.91) som skyldes nivåskiftet i 3. kvartal 1998. Sesong- og irregulærkomponenten er plottet i figuren 5.3a-c.

**Konklusjon.** Sammenlignet med den første kjøringen får vi en bedre sesongjustering ved å bruke et lengre filter i beregningen for sesongkomponenten.

### 5.2.2 Samlet ordretilgang fra eksportmarkedet, $i_{EKSP_t}$

Dataene er observert fra 1. kvartal 1986 til 3. kvartal 1999 med 55 observasjoner. På grunn av store variasjoner i slutten av tidsserien vil vi bruke  $(3 \times 9)$  filteret i beregningen for sesongkomponent. Vi får i første kjøring  $\hat{\Theta} \approx 1$  (seasonal MA parameter). Dermed vil vi bruke såkalte "fixed seasonal effects" istedenfor det andre leddet  $(P, D, Q)$  i ARIMA modell for å beregne framskrivinger. Det gir bedre tilpasning mellom ARIMA modellen og data og vi får bedre framskrivinger.

Vi får følgende resultater:

- ARIMA modell.

ARIMA(1 1 0) modellen er valgt. Vi får  $\hat{\phi} = -0.0903$  med standardavviket 0.1442.

Gjennomsnittlig absolutt prosent feil for framskriving

siste år: 13.05    forrige år: 8.26    to år tidligere: 7.26

siste tre år: 9.52 som er mindre enn grenseverdien (15%) for å forkaste en ARIMA modell.

- Testen for fixed seasonal effects.

Testen er signifikant med  $\chi^2 = 50.3$  og 3 frihetsgrader.

- Intervensjoner.

Det fins ingen intervensjon.

- Framskrivinger.

Framskrivingene og 95% konfidensintervaller for et år framover er listet ut i tabell 5.2. De viser en nedgang.

-  $F$ -test for stabilt sesongmønster og sesongbevegelser.

Stabilt sesongmønster er signifikant på 0.1% nivået

Sesongbevegelser er signifikant på 5% nivået.

Tabell 5.2: *Framskrivinger og 95% konfidensintervaller*

| <i>Dato</i> | Nedre  | Verdi  | Øvre   |
|-------------|--------|--------|--------|
| 4.kv 1999   | 111.16 | 129.71 | 151.36 |
| 1.kv 2000   | 99.27  | 122.06 | 150.09 |
| 2.kv 2000   | 96.42  | 123.50 | 158.19 |
| 3.kv 2000   | 86.47  | 114.56 | 151.78 |

- 11 mål for kvalitet.

Ingen mål er større enn 1.  $Q = 0.50$ . Basert på disse målene har vi en god sesongdekomponering.

- Slidingspans.

Sesongjusteringen blir utført i tre delvis overlappende delmengder

span 1: fra 1. kvartal 1986 til 3. kvartal 1997

span 2: fra 1. kvartal 1987 til 3. kvartal 1998

span 3: fra 1. kvartal 1988 til 3. kvartal 1999

Testene for sesongmønster og sesongbevegelser for hvert span er vist i tabellen nedenfor

|                       | <i>F</i> -test |        |        |                        |
|-----------------------|----------------|--------|--------|------------------------|
|                       | span 1         | span 2 | span 3 |                        |
| Stabilt sesongmønster | 50.61          | 53.19  | 36.91  | signifikant på 5%      |
| Sesongbevegelse       | 2.36           | 1.87   | 1.59   | ikke signifikant på 5% |
| <i>M7</i>             | 0.37           | 0.34   | 0.40   | tilfredsstillende      |

Andelene i prosent av kvartaler som er definert som upålitelige

For sesongkomponenten : 0 av 47 (0.0%)

For endring fra kvartal til kvartal i sesongjusterte tall : 2 av 46 (4.3%)

- Figurer.

Rådata og framskrivinger med 95% konfidensintervaller er plottet i figuren 5.4a. Figurene 5.4b og c viser sesongjusterte tall og trenden for total eksport i orderindustri.

**Konklusjon.** Resultatet fra slidingspans viser en god sesongjustering for ordretilgangen for eksport.

### 5.2.3 Samlet ordretilgang fra hjemmemarkedet, $i\_HJEM_t$

Resultater fra X-12-ARIMA viser en dårlig sesongjustering med høyt antall kvartaler som har store revisjoner. Vi skal derfor ikke bruke X-12-ARIMA for å beregne sesongjusterte tall, men gjøre dette på følgende måte:

Siden indeksene er beregnet med utgangspunkt i 1995, får vi

$$\begin{aligned}i\_TOT_t &= \frac{v\_TOT_t}{v\_TOT_{95}} \times 100\%, \\i\_EKSP_t &= \frac{v\_EKSP_t}{v\_EKSP_{95}} \times 100\%, \\i\_HJEM_t &= \frac{v\_HJEM_t}{v\_HJEM_{95}} \times 100\%\end{aligned}$$

hvor  $v\_$  og  $i\_$  står for verdi og indeks. Fra ligningen

$$v\_TOT_t = v\_EKSP_t + v\_HJEM_t$$

får vi

$$i\_HJEM_t = \frac{i\_TOT_t \times v\_TOT_{95} - i\_EKSP_t \times v\_EKSP_{95}}{v\_HJEM_{95}}$$

Dermed kan vi beregne trend og sesongjusterte tall for hjemmemarked i total ved

$$\begin{aligned}Ji\_HJEM_t &= \frac{Ji\_TOT_t \times v\_TOT_{95} - Ji\_EKSP_t \times v\_EKSP_{95}}{v\_HJEM_{95}} \\Ti\_HJEM_t &= \frac{Ti\_TOT_t \times v\_TOT_{95} - Ti\_EKSP_t \times v\_EKSP_{95}}{v\_HJEM_{95}}\end{aligned}\quad (5.1)$$

hvor  $J$  står for sesongjusterte tall, og  $T$  står for trend.

- Figurer.

Sesongjusterte tall og trend for hjemmemarkedet er plottet i figurene 5.5a og b. Vi ser en brå endring i sesongjusterte tall og trenden, som skyldes nivåskiftet i 3. kvartal 1998.

- Beregning for revisjoner i sesongjusterte tall.

Slidingspans blir utført i tre delmengder [1.kv 1986, 3.kv 1997], [1.kv 1987, 3.kv 1998] og [1.kv 1988, 3.kv 1999] for  $i\_TOT_t$  og  $i\_EKSP_t$ . Siden vi ikke kjører X-12-ARIMA for  $i\_HJEM_t$ , vil vi estimere sesongjusterte tall for denne tidsserien i disse tre delmengdene ved (5.1). Deretter beregnet vi andelen av kvartaler som klassifiserte som upåtelige for sesongjusterte tall. Vi får 20 av 47 tidspunkter utgjør 42.6%. Tallet er for stort.

**Konklusjon.** Vi får en dårlig sesongjustering for  $i\_HJEM_t$  på grunn av store revisjoner.

## 5.3 Sesongjustering for tidsseriene $v\_TOT_t$ , $v\_EKSP_t$ og $v\_HJEM_t$ .

Vi sesongjusterer med absoluttall her.

### 5.3.1 Totalen, $v\_TOT_t$

Vi får samme konklusjoner på sesongjustering som for  $i\_TOT_t$ . Målene, blant 11 mål for kvalitet, som er større enn 1 er

$$M1 = 1.333, \quad M3 = 1.182, \quad M10 = 1.019 \quad \text{og} \quad M11 = 1.019 \quad \text{med} \quad Q = 0.78$$

Andelene av kvartaler i prosent som er definert som upåtelige er

|                                                              |                    |
|--------------------------------------------------------------|--------------------|
| For sesongkomponenten                                        | : 2 av 47 (4.3%).  |
| For sesongjusterte tall                                      | : 5 av 47 (10.6%). |
| For endring fra kvartal til kvartal<br>i sesongjusterte tall | : 7 av 46 (15.2%). |

- Figurer.

Sesongjusterte tall og trend er plottet i figurene 5.6a-c.

**Konklusjon.** Vi får en god sesongjustering for totalen av ordrer i industrien.

### 5.3.2 Samlet ordretilgang fra eksportmarkedet, $v\_EKSP_t$

Konklusjonen av sesongjustering for  $v\_EKSP_t$ , i absoluttall er den samme som vi kjørte med indeksene. Figurene for sesongjusterte tall og trend er plottet i figurene 5.7a-c.

### 5.3.3 Samlet ordretilgang fra hjemmemarkedet, $v\_HJEM_t$

Siden  $v\_HJEM_t = v\_TOT_t - v\_EKSP_t$  beregner vi sesongjusterte tall og trend av  $v\_HJEM_t$  ved

$$\begin{aligned} Jv\_HJEM_t &= Jv\_TOT_t - Jv\_EKSP_t \\ Tv\_HJEM_t &= Tv\_TOT_t - Tv\_EKSP_t \end{aligned} \quad (5.2)$$

hvor  $J$  står for sesongjusterte tall og  $T$  står for trend.

- Beregning for revisjoner i sesongjusterte tall.

Vi bruker samme måte som for  $i\_HJEM_t$ . Vi får 20 av 47 (42.6%) hvor revisjoner av sesongjusterte tall er større enn 3%.

- Figurer.

Sesongjusterte tall og trend er plottet i figurene 5.8a-c.

**Konklusjon.** Vi får en dårlig sesongjustering av  $v\_HJEM_t$  på grunn av store revisjoner.

### 5.3.4 Samlet ordretilgang fra hjemmemarkedet, $i\_HJEM_t$ , med 9-ledd filteret for trenden

Vi ser i figurene 5.5b og 5.8b at det er store utslag i trenden i slutten av tidsserien. Trenden ser uvanlig ut. For å unngå dette kan vi estimere trenden med et lengre filter enn default filteret. Vi sesongjusterer  $i\_HJEM_t$  med 7-, 9- og 11-ledd filter for trenden. Slidingspans gir minst endringer av sesongjusterte tall fra kvartal til kvartal med 9-ledd filteret. Vi får

- ARIMA modell.

ARIMA(2,1,2) modellen er valgt av programmet med

$$\begin{aligned}\hat{\theta}_1 &= 0.5422 & Std(\hat{\theta}_1) &= 0.1263 \\ \hat{\theta}_2 &= -0.4404 & Std(\hat{\theta}_2) &= 0.1258 \\ \hat{\Theta}_1 &= 1.2196 & Std(\hat{\Theta}_1) &= 0.0796 \\ \hat{\Theta}_2 &= 1.0000 & Std(\hat{\Theta}_2) &= 0.0802\end{aligned}$$

- Slidingspans.

Andelene i prosent av kvartaler som er definert som upålitelige

|                                                           |                    |
|-----------------------------------------------------------|--------------------|
| For sesongkomponenten                                     | : 4 av 47 (8.5%)   |
| For endring fra kvartal til kvartal i sesongjusterte tall | : 15 av 46 (32.6%) |

Selv om tallene er innenfor grensene ligger de ikke langt fra grenseverdiene for en god sesongdekomponering.

- 11 mål for kvalitet.

Vi får  $M1 = 2.397$ ,  $M3 = 1.939$ ,  $M5 = 1.584$ ,  $M7 = 0.548$  og  $Q = 1.14$ .

$M1$  er et mål på relativt bidrag fra  $I_t$  i den totale variasjonen. Når  $M1$  er større enn 1, betyr det at trend  $T_t$  eller sesongkomponent  $S_t$  kan bli forstyrret av variasjoner fra  $I_t$ . Dermed kan hverken  $T_t$  eller  $S_t$  skilles helt perfekt fra  $I_t$ . Vi får lavere kvalitet på sesongjusteringen.

$M3$  er et mål på endring fra kvartal til kvartal i irregulærkomponent  $I_t$  i forhold til endring i  $T_t$ . Når  $M3 > 1$  blir det vanskelig å skille  $I_t$  og  $T_t$  i den tredje runden.

En vil teste om at det fins en  $k$  slik at endringene i  $T_t$  dominerer endringene i  $I_t$  etter  $k$  kvartaler. Testen er ikke signifikant når  $M5 > 1$ .

$Q = 1.14$ .

- Figurer.

Sesongjusterte tall og trenden er plottet i figurene 5.9a-b. Vi får en glattere trend i de siste observasjonene, og trenden tenderer nedover. Figur 5.9c viser trendene som beregnes henholdvis fra X-12-ARIMA med 9-ledd filteret, og ved ligningen (5.1).

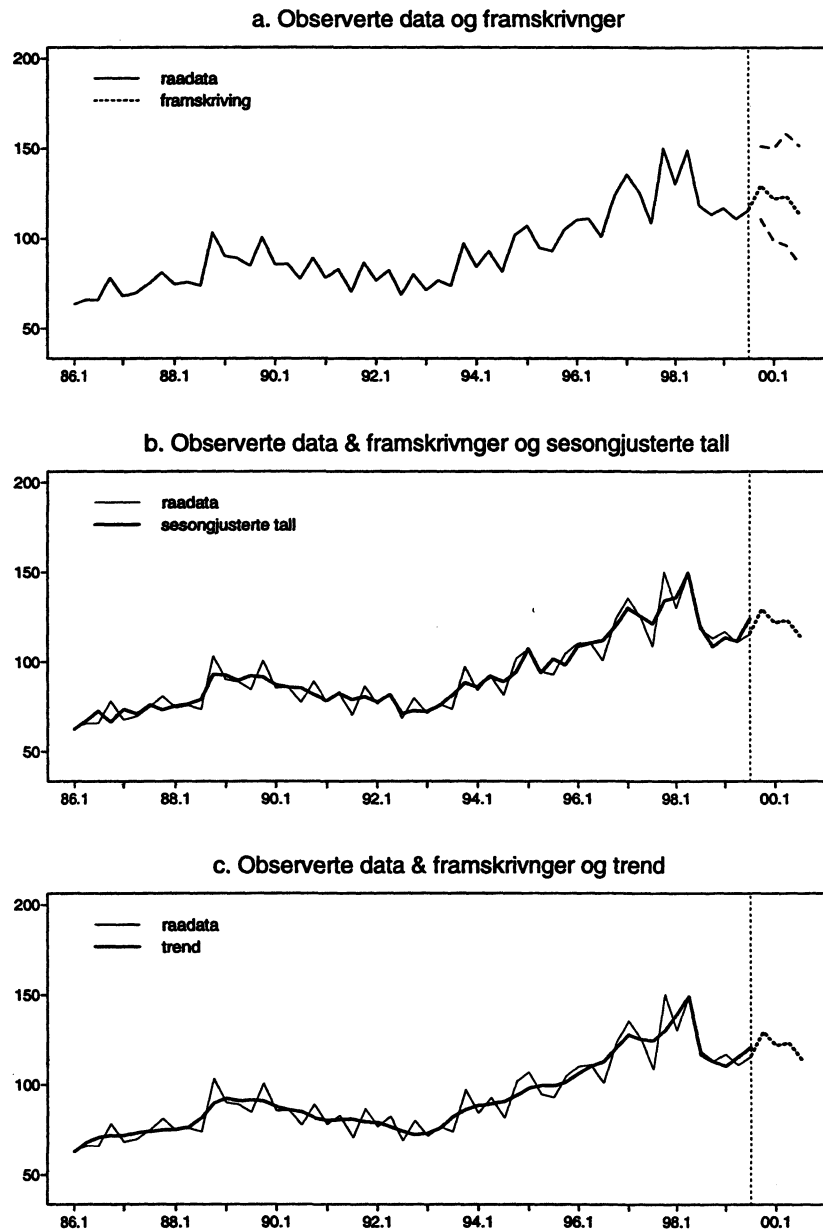
**Konklusjon.** Vi får en bedre sesongjustering for  $i\_HJEM_t$  ved å bruke lengre filtre enn default for sesongkomponenten og trenden i beregningen, men risikerer store endringer av sesongjusterte tall fra kvartal til kvartal når nye observasjoner legges til.

## 5.4 Konklusjon.

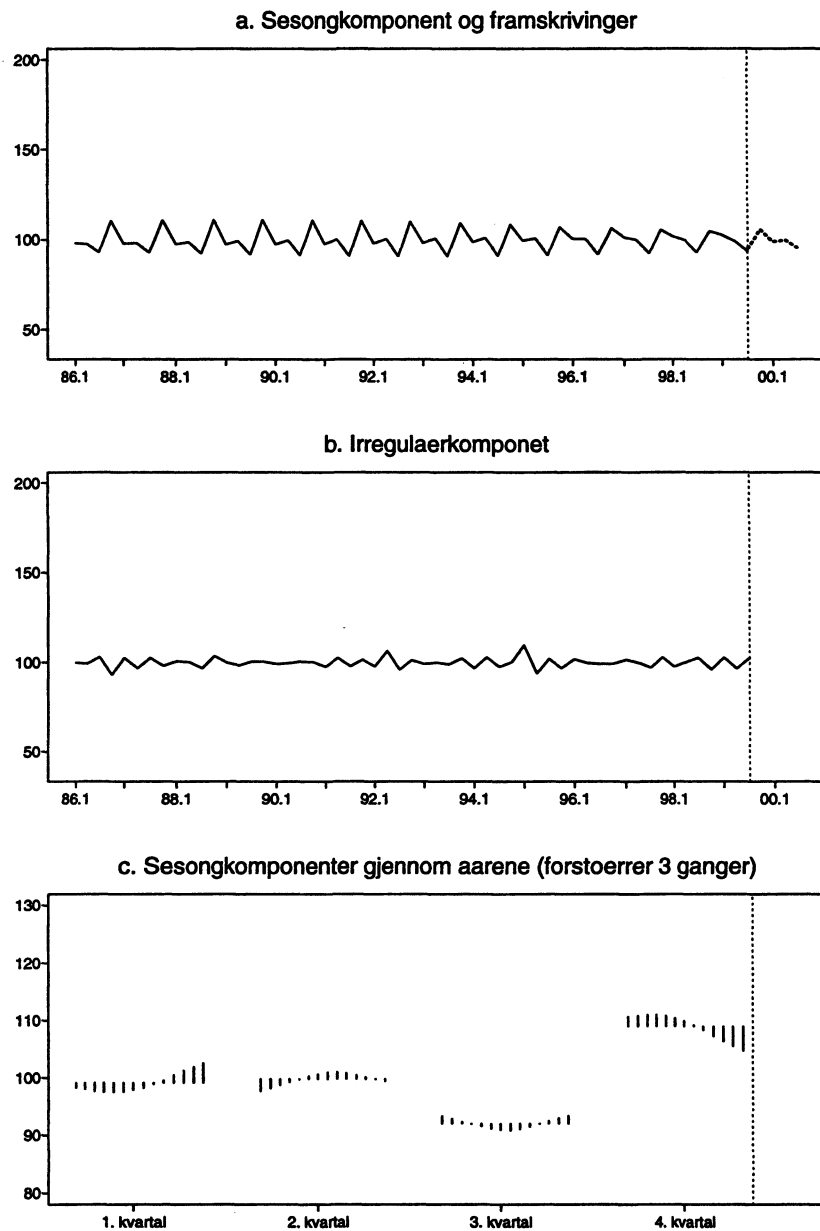
Vi får følgende konklusjoner:

Gode sesongjusteringer for tidsseriene av totalen av ordrer og den totale eksporten av ordrer i industrien, men ikke for hjemmemarkedet på grunn av store revisjoner av sesongjusterte tall. Dermed bør vi bare publisere rådata for denne serien.

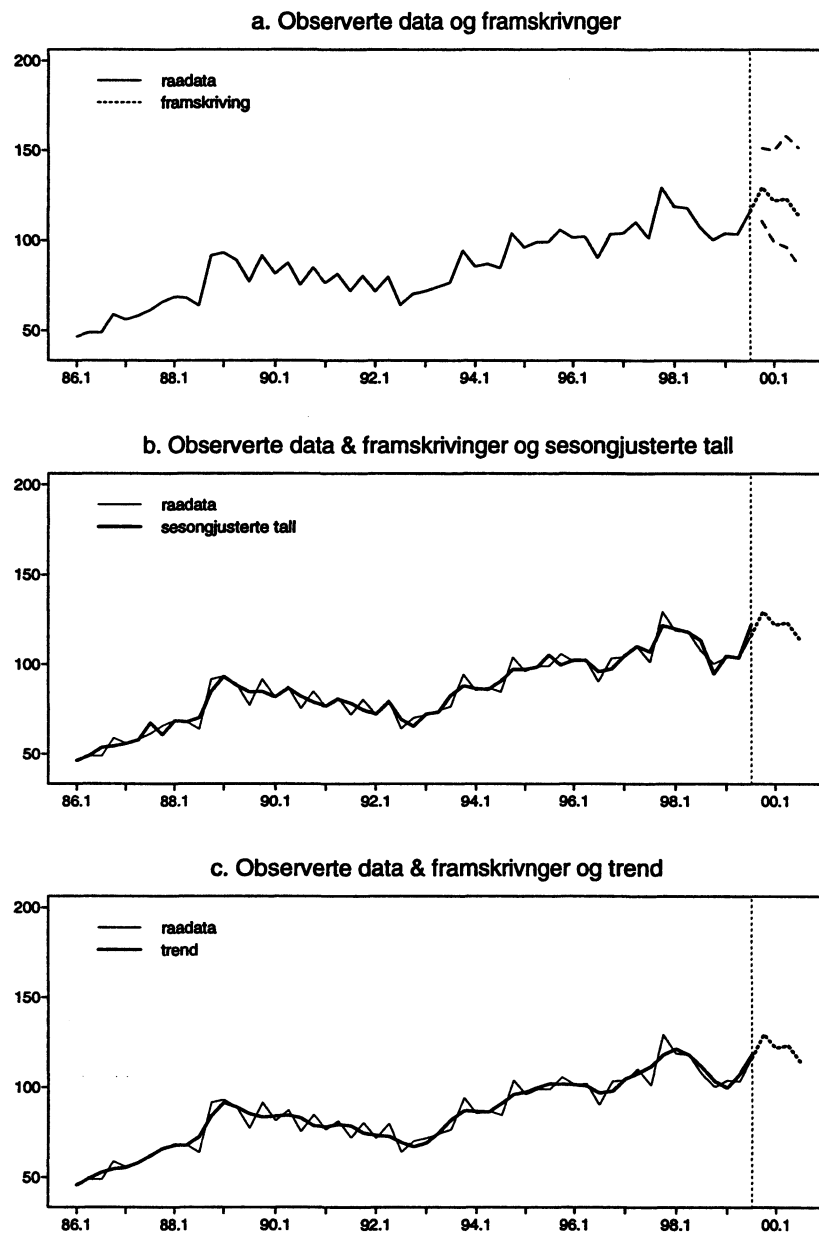
Konklusjonen blir lik, uavhengig av om vi sesongjusterer med indekser eller verditall som input. Sesongjusterte tall beregnet av indekser ved bruk av X-12-ARIMA, og sesongjusterte tall beregnet på verditallene og deretter transformert til indekser med år 1995 som basis, er nesten identiske. Det er derfor ikke nødvendig å sesongjustere to ganger: en for indeks og en på verdi, men kun sesongjustere indekstallene for bruke denne til å korrigere verditallene for sesongjusteringer.



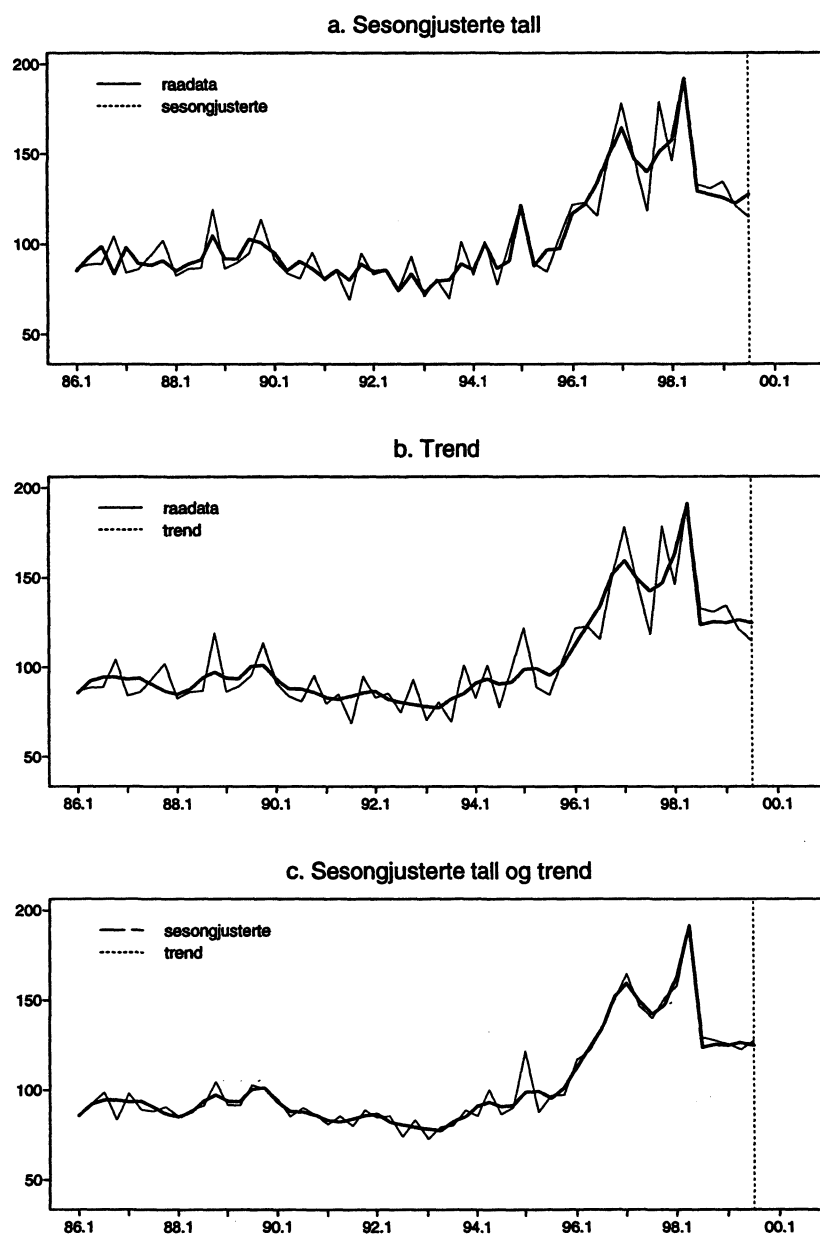
Figur 5.2: *Rådata, sesongjusterte tall og trend av total for ordreindustri*



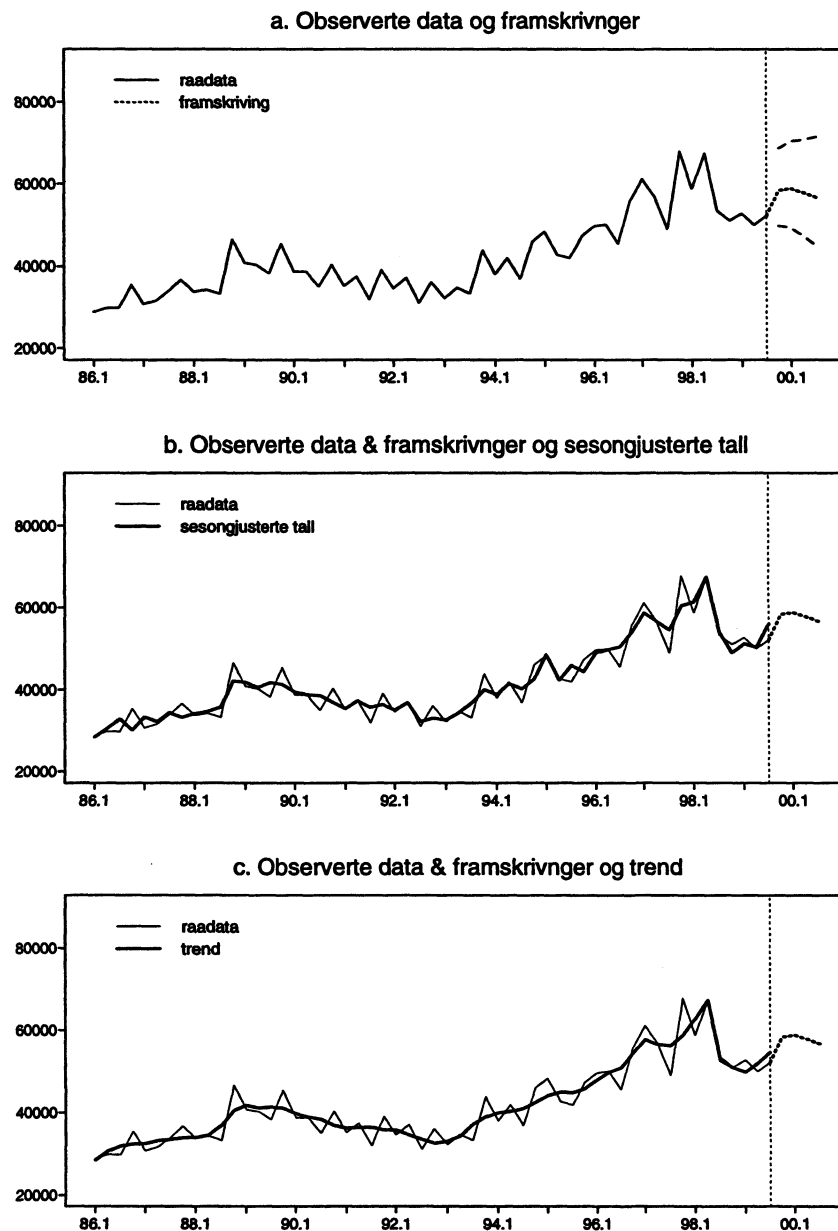
Figur 5.3: *Sesong- og irregulærkomponenter av total for ordreindustri*



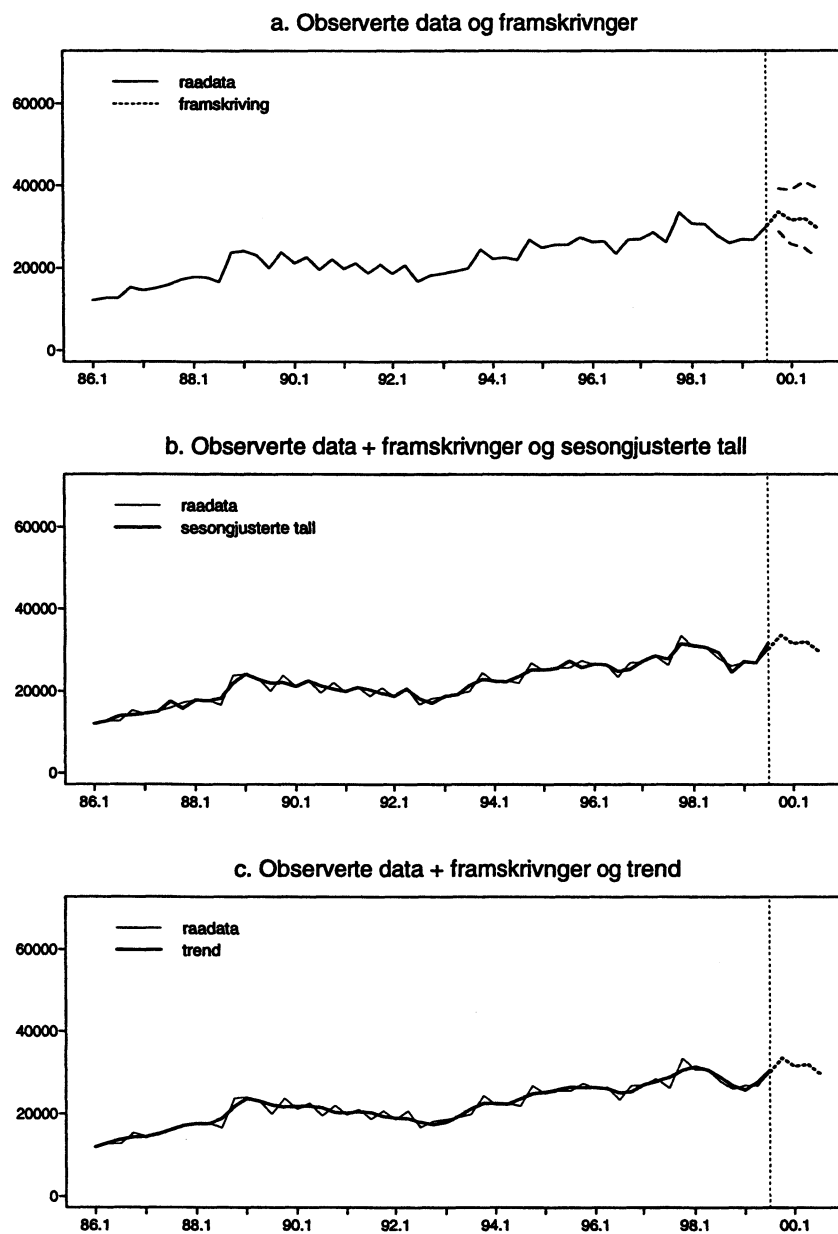
Figur 5.4: Rådata, sesongjusterte tall og trend av total eksport i ordreindustri



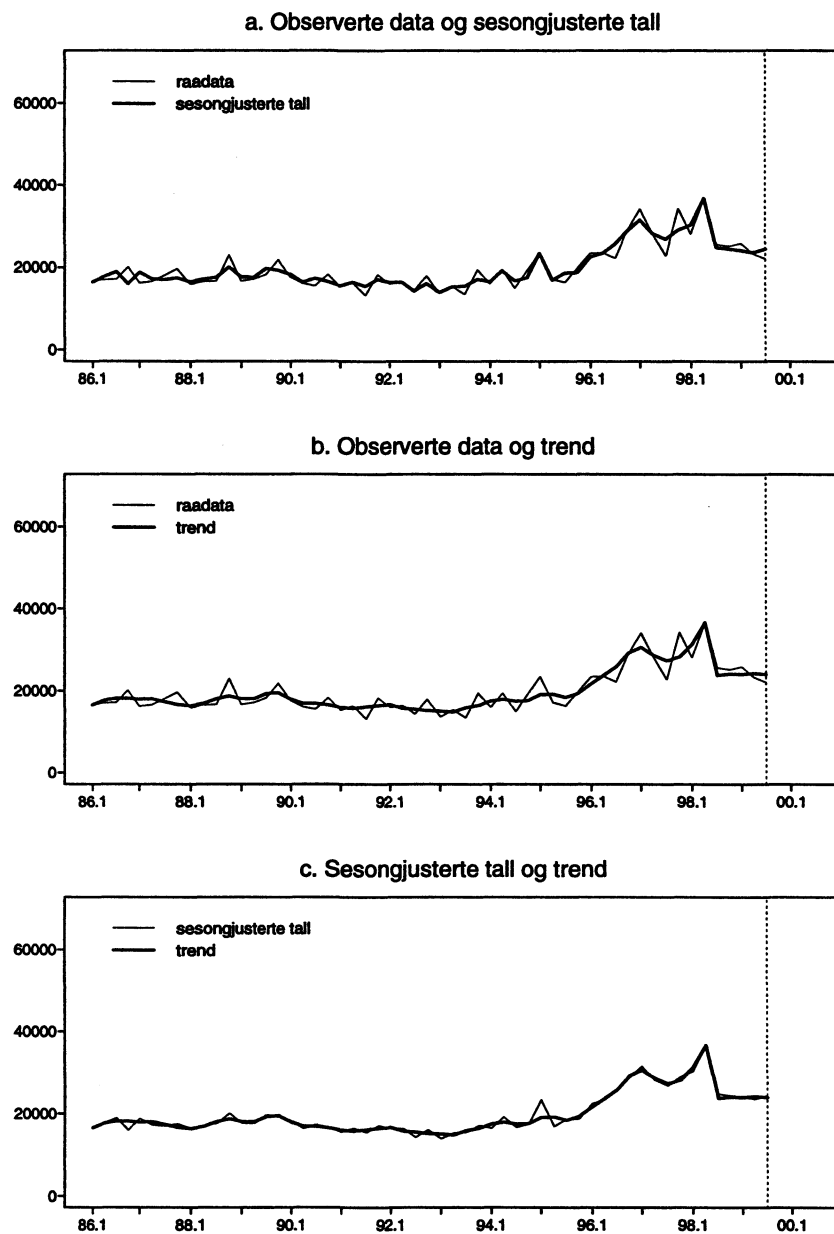
Figur 5.5: Rådata, sesongjusterte tall og trend av hjemmemarked i total for ordreindustri



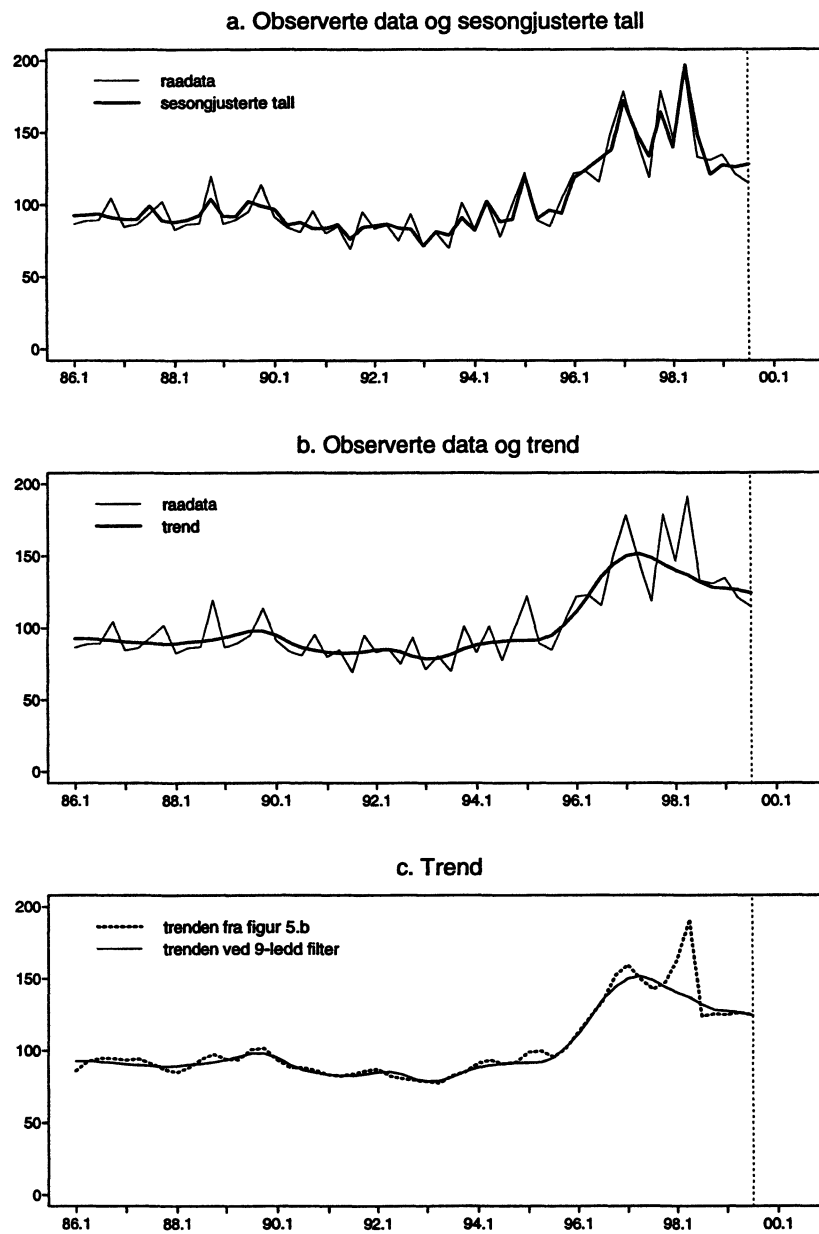
Figur 5.6: Rådata, sesongjusterte tall og trend av total for ordreindustri i absoluttall



Figur 5.7: Rådata, sesongjusterte tall og trend av total eksport i ordreindustri, i absoluttall



Figur 5.8: *Rådata, sesongjusterte tall og trend av hjemmemarked i total i ordreindustri, i absoluttall*



Figur 5.9: Rådata, sesongjusterte tall og trend av hjemmemarked i total i ordreindustri, i indekser, med 9-ledd filteret for trenden

## Kapittel 6

# Estimation of confidence intervals for seasonally adjusted and trend time series: A case study based on simulation

*Paper to be presented at the SAM-98 Seminar, Bucharest, 22-24 October*  
Dinh Quang Pham<sup>1</sup> and Terje Skjerpen<sup>2</sup>

### Abstract

Seasonally adjusted and trend series are often published without providing any measure of uncertainty. In this paper it is shown how simulation techniques can be combined with the seasonal adjustment programs X-11-ARIMA/88 and X-12-ARIMA in order to produce confidence intervals for seasonally adjusted and trend series. Both parametric and non-parametric simulation methods are considered. From the production perspective, a simulation approach to calculation of components uncertainty seems favourable since it allows a high degree of automatisation. As a case study the monthly time series of 'First time registered new passenger cars' (in Norway) is considered. In the parametric case data are resampled on the basis of a (preferred) ARIMA model. From these simulations we obtain  $B$  replications of the observed time series. Each of the replicated time series are used as input series to X-11-ARIMA/88. At each point of time the empirical standard deviations of the seasonally adjusted and trend value are calculated (over the replications). These standard deviations are used when constructing the confidence intervals for the unobserved components. In the non-parametric case the estimated irregular component obtained from applying either X-11-ARIMA/88 or X-12-ARIMA to the original series are utilized in the simulations. The irregulars are replicated by drawing with replacement from the empirical distribution of the estimated irregular component. The seasonally adjusted time series are replicated by combining the simulated irregular component with the trend component obtained from the decomposition of the original series by means of

---

<sup>1</sup>Statistics Norway, Division of Methods and Standards, P.O. Box 8131 Dep., N-0033 Oslo, E-mail: dqp@ssb.no

<sup>2</sup>Statistics Norway, Research Department, P.O. Box 8131 Dep., N-0033 Oslo, E-mail: tes@ssb.no

X-11-ARIMA/88 or X-12-ARIMA. To replicate the trend we apply the 13-term symmetric Henderson filter to the simulated seasonally adjusted time series. Based on the  $B$  non-parametric replications a confidence interval for the seasonally adjusted and trend series can be constructed for each period of time.

**Keywords:** Component uncertainty. Simulations. X-11-ARIMA/88, X-12-ARIMA

**Acknowledgement:** We are grateful to Anne Berit Dahle, Joseph Sexton, Leiv Solheim and Anders Rygh Swensen for comments. Any remaining errors or shortcomings are the responsibility of the authors.

## 6.1 Introduction

Seasonally adjusted and trend series are usually published without any uncertainty measure. However, an other type of uncertainty is often focused on. It is well known that the estimates of the seasonally adjusted values and the trend at the end of the sample are revised as more data become available. This is a property at the end of the sample. In this paper we are however occupied with the entire time series. We want to estimate the uncertainty at each sample point for the seasonal and trend components, given the available data. In the time series literature there has been some attempts to calculate the uncertainty of time series components. For instance, when one uses the state space representation for structural time series models (cf. for instance Burrige and Wallis (1985) and Harvey (1989)), the seasonal component is a state variable and the uncertainty can be obtained as the root mean square error of the smoothed components. Quite recently Pfeiffermann et al (1995) has shown how one can calculate variances of X-11-ARIMA seasonally adjusted estimators by utilizing the sample design for time series data (cf. also Breidt (1992), Pfeiffermann (1994) and Bell and Kramer (1996)). However, when the aim is to decompose a huge amount of time series, it seems natural to utilize simulation techniques because this can be easily automated.

In this paper we utilize different simulation techniques in order to estimate the variance of the estimators of the seasonally adjusted values and the trend. The simulation techniques are used together with the two seasonal adjustment programs X-11-ARIMA/88 (cf. Dagum (1988)) and X-12-ARIMA (cf. Bureau of the Census (1998) and Findley et al (1998)). Hereafter we use the short form X-11-ARIMA for X-11-ARIMA/88. We provide both a parametric and a non-parametric simulation technique. In the parametric case we fit an ARIMA model to the data and we resample data  $B$  times from this model paying regard both to noise and coefficient uncertainty. The  $B$  simulated time series are decomposed by X-11-ARIMA and accordingly we obtain  $B$  simulated values for the seasonal component and for the trend at all time periods. The simulated data can be used for construction of confidence intervals for the true seasonal value and the trend at each period of time. For the parametric case we consider only the X-11-ARIMA program.

The point of departure for the non-parametric case is the estimated irregular component for the original series based on either X-11-ARIMA or X-12-ARIMA. In the simulations we draw from the empirical distribution of the estimated irregular component. When combining the simulated irregular values with the trend obtained by decomposing the original series, replicated simulated seasonally adjusted values are produced. To replicate the trend we utilize the 13-term symmetric Henderson filter. The main gain by using X-12-ARIMA over X-11-ARIMA is that it is more convenient to pay regard to outlier effects in the former since one can utilize the regARIMA facilities. This influences the forecasts and thereby the decomposition at the end of the observed series. We find that the confidence intervals produced by the non-parametric method are much smaller than in the parametric case. Furthermore we find the confidence intervals based on resampling from X-12-ARIMA irregulars smaller than resampling from X-11-ARIMA irregulars.

## 6.2 The methods

The following multiplicative decomposition model is assumed

$$Y_t = T_t \times S_t \times I_t \quad (6.1)$$

In equation (6.1)  $Y_t$  is the observed (original) time series at period  $t$ . It is decomposed into the trend component ( $T_t$ ), the seasonal component ( $S_t$ ) and the irregular component ( $I_t$ ), which all are unobserved. For the estimation of the components we consider X-11-ARIMA and X-12-ARIMA. The seasonally adjusted value is defined as

$$A_t = \frac{Y_t}{S_t} = T_t \times I_t \quad (6.2)$$

The trading day and Easter holiday effects are assumed to be contained within the seasonal factor.

### 6.2.1 Method 1. A parametric method based on simulation of data from an ARIMA model

In this approach one fit an ARIMA to the data and the simulated data are based on the estimated model. One proceeds by the following steps

- Step 1. One searches for an ARIMA model which can be claimed to fit the data well. Let us for expository reasons assume that the so-called "airline" model is adequate. According to this model  $X_t = \log(Y_t)$  is given by

$$(1 - L)(1 - L^{12})X_t = (1 - \theta L)(1 - \Theta L^{12})\epsilon_t, \quad (6.3)$$

where the  $\epsilon_t$ 's are normally i.i.d with mean zero and variance  $\sigma_\epsilon^2$  (i.e white noise process). The parameters  $\theta$ ,  $\Theta$  and  $\sigma_\epsilon^2$  are estimated from the data. The symbol  $L$  denotes the lag operator.

- Step 2. Replicated samples of the data are obtained on the basis of equation (6.3). Let  $i$  denote the  $i$ 'th replication ( $i = 1, \dots, B$ ), such that  $X_t^{(i)}$  denotes the value of  $X$  at period  $t$  in the  $i$ 'th replication.  $X_t^{(i)}$  is computed by

$$X_t^{(i)} = X_{t-1} + X_{t-12} - X_{t-13} + (1 - \theta^{(i)}L)(1 - \Theta^{(i)}L^{12})\epsilon_t^{(i)}, \quad (6.4)$$

where  $\epsilon_t^{(i)}$  is drawn randomly from the univariate distribution  $N(0, \hat{\sigma}_\epsilon^2)$ , where  $\hat{\sigma}_\epsilon^2$  denotes the estimated value of  $\sigma_\epsilon^2$  and where the vector  $(\theta^{(i)} \ \Theta^{(i)})'$  is drawn randomly from the bivariate normal distribution

$$N\left(\begin{pmatrix} \hat{\theta} \\ \hat{\Theta} \end{pmatrix}, \begin{pmatrix} \widehat{var}(\hat{\theta}) & \widehat{cov}(\hat{\theta}, \hat{\Theta}) \\ \widehat{cov}(\hat{\theta}, \hat{\Theta}) & \widehat{var}(\hat{\Theta}) \end{pmatrix}\right)$$

- Step 3. The replicated time series are decomposed by running X-11-ARIMA (after taking anti-log). The estimated trend and seasonally value at time  $t$  corresponding to the  $i$ 'th replication is denoted  $T_t^{(i)}$  and  $A_t^{(i)}$  respectively.

Steps 2 and 3 are repeated a large number of times. At the end we have the replications  $\{T_t^{(1)}, T_t^{(2)}, \dots, T_t^{(B)}\}$  for the trend and the replications  $\{A_t^{(1)}, A_t^{(2)}, \dots, A_t^{(B)}\}$  for the seasonally adjusted value at each period. The means, standard deviations and relative standard deviations of these two replicated series are given by

$$\overline{M}_t^S = \frac{1}{B} \sum_{i=1}^B M_t^{(i)}, \quad (6.5)$$

$$Std(M_t^S) = \sqrt{\frac{1}{B-1} \sum_{i=1}^B (M_t^{(i)} - \overline{M}_t^S)^2} \quad (6.6)$$

and

$$r_t^M = \frac{Std(M_t^S)}{\overline{M}_t^S} \quad (6.7)$$

where  $M$  stands for  $T$  or  $A$ .

A  $(1 - \alpha)100\%$  bootstrapped confidence interval for  $M_t$  is given by

$$\left[ \overline{M}_t^S - z_{\alpha/2} \times Std(M_t^S), \overline{M}_t^S + z_{\alpha/2} \times Std(M_t^S) \right] \quad (6.8)$$

where  $z_{\alpha/2}$  is the  $1 - \alpha/2$  fractile in the standard normal distribution.

### 6.2.2 Method 2. Non-parametric bootstrapping utilizing the estimated irregular component from X-11-ARIMA

Findley and Monsell (1990) presented a simulation method for computing the standard error of the seasonally adjusted values obtained by running a preliminary version of X-12-ARIMA. The main idea is to resample the irregular values non-parametrically by utilizing the empirical distribution. The procedure contains the following steps:

- Step 1. The original data are decomposed multiplicatively by running X-11-ARIMA twice. The first run is carried through in order to produce forecasts. In the second run we use the original series extended with forecasts as input to X-11-ARIMA. From the decomposition of these time series we obtain the estimated irregular component from which we resample. The estimated irregular component is given by  $\{\hat{I}_1, \hat{I}_2, \dots, \hat{I}_N\}$ , where  $N$  denotes the last observation in the sample extended with forecasts. For simplicity one can assume that all these  $N$  values are different.
- Step 2. The irregular component is resampled by drawing with replacement from the empirical distribution of the estimated irregular component. This means that

$$P(I_t^{(i)} = \hat{I}_n) = \frac{1}{N}, \quad t = 1, \dots, N, \quad n = 1, \dots, N, \quad \text{and } i = 1, \dots, B \quad (6.9)$$

The seasonally adjusted and trend values in the  $i$ 'th replication is obtained by the following formulas

$$A_t^{(i)} = \hat{T}_t \times I_t^{(i)} \quad \text{and} \quad T_t^{(i)} = \sum_{j=-H}^H h_j A_t^{(i)}, \quad (6.10)$$

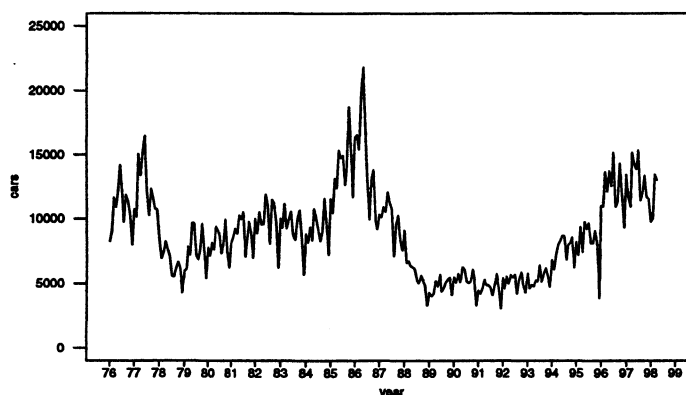
where  $h_j, j = 1, \dots, H$  is a Henderson weight. To select the value of  $H$ , i.e the length of the Henderson filter, we utilize the output from Table D.12 in X-11-ARIMA (cf. Dagum (1988)) which gives information on which filter that has been used when decomposing the input series.

Altogether  $B$  replications are produced and means and standard deviations for the replicated data are again calculated in accordance with equations (6.5) and (6.6). The original series have been extended with forecasts based on the airline model. Because of this extension the symmetric Henderson filter can also be applied at the end of the sample period.

Findley and Monsell (1996) considered a simulation design in which the variance of the irregular component to some extent was assumed to be monthdependent. This implies that the irregular component at a specific month must be replicated by only utilizing the estimated irregular values for the months within a specific group of months. We have also applied this idea to some extent.

### 6.3 Application. The first time registered new passenger cars in Norway

The monthly data starts in January 1976 and the last observation considered is from April 1998. Totally this gives 268 observations. The data have some outliers which can be given specific interpretations. For instance in December 1977 one experienced a very high level since the price was expected to increase next year because of tax changes. A similar effect but with opposite sign was experienced in December 1995 due to the announcement that the car taxes were going to be decreased from the beginning of the subsequent calendar year. The original data are plotted in Figure 6.1.



Figur 6.1: *The original data for car sales*

### 6.3.1 Application of method 1: The parametric case

For an ARIMA model to be adequate it must fulfill the following three criteria (cf. Dagum (1988) p. 32)

- (i) The mean absolute percentage error of the forecasts for the last three years must not be higher than 15%.
- (ii) The chi-squared probability for testing the null hypothesis of randomness of the residuals must be greater than 5%.
- (iii) There must be no evidence of overdifferencing.

None of the five built-in ARIMA models were found adequate by X-11-ARIMA since criterion 1 was not met. However for illustrative purposes we use ARIMA (0,1,1)(0,1,1) after having transformed to natural logarithms. The maximum likelihood estimation results is shown in Table 6.1. The estimators for the parameters  $\theta$  and  $\Theta$  are significant at the 5% level.

Tabell 6.1: *The maximum likelihood estimates of the parameters in the ARIMA model*

| Parameter | Estimate | Std. Error | t-ratio |
|-----------|----------|------------|---------|
| $\theta$  | 0.4318   | 0.0550     | 7.85    |
| $\Theta$  | 0.9148   | 0.0581     | 15.74   |

$$\hat{\sigma}_\epsilon^2 = 0.0164$$

The estimated covariance matrix of  $(\hat{\theta} \ \hat{\Theta})'$ , which we denote by  $\hat{\Sigma}$  is given by

$$\hat{\Sigma} = \begin{pmatrix} 0.0030 & -0.0001 \\ -0.0001 & 0.0034 \end{pmatrix}.$$

Figures 6.2 and 6.3 contain the confidence intervals for the seasonally adjusted and trend values estimated by the parametric bootstrap method.

Tabell 6.2: *The statistics of  $r_t^A$  and  $r_t^T$*

|            | Var.    | Mean  | Std  | Min   | Max   |
|------------|---------|-------|------|-------|-------|
| Method 1   | $r_t^A$ | 20.37 | 4.09 | 10.90 | 47.54 |
| X-11-ARIMA | $r_t^T$ | 5.64  | 0.76 | 4.64  | 11.11 |
| Method 2   | $r_t^A$ | 7.07  | 0.69 | 4.56  | 11.48 |
| X-11-ARIMA | $r_t^T$ | 3.19  | 0.17 | 2.93  | 4.50  |
| Method 2   | $r_t^A$ | 6.54  | 0.79 | 3.99  | 13.57 |
| X-12-ARIMA | $r_t^T$ | 2.95  | 0.16 | 2.60  | 4.07  |

The number of replications, that is  $B$ , has been set to 1000. The wideness of the confidence intervals can also be illustrated by calculating the relative standard deviations over the replications of the seasonally adjusted and trend value at each point

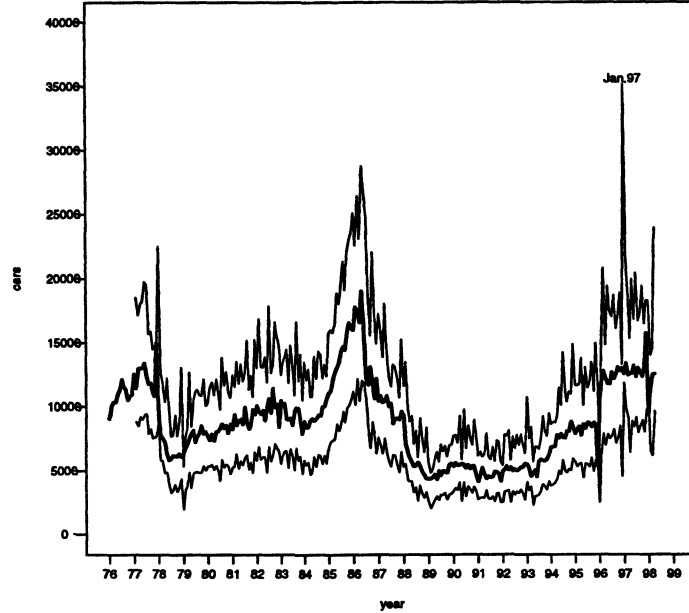


Figure 6.2: *The seasonally adjusted values and the estimated 95% confidence intervals. Parametric case*

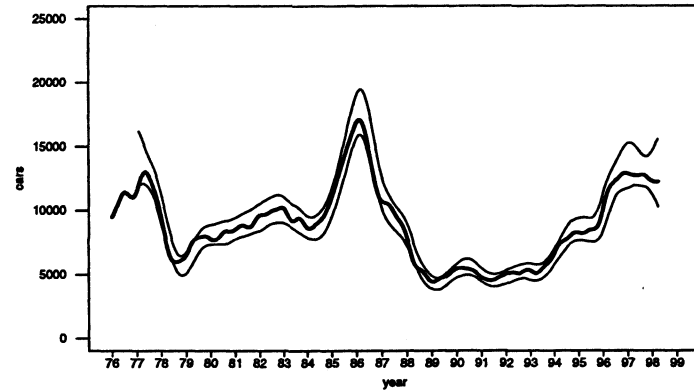


Figure 6.3: *The trend values and the estimated 95% confidence intervals. Parametric case*

of time. Empirical characteristics of these series are given in the first two rows of Table 6.2.

The confidence intervals seem rather wide. The mean of  $r_t^A$  and  $r_t^T$  are 20.37% and 5.64% respectively. For instance for the last observation in the sample, where the seasonally adjusted value from X-11-ARIMA is 12533, the estimated confidence interval for the seasonally adjusted value is [9621, 23956]. The corresponding value for the trend component is 12191 with the estimated confidence interval as [10251, 15527].

### 6.3.2 Application of method 2. The non-parametric method utilizing the estimated irregular component of X-11-ARIMA

Also when using this method the time series are extended with 12 months of forecasts, utilizing the airline model. This able us to use the symmetric Henderson filter for simulation of the trend at the end of the sample. The extended time series starts in January 1976 and ends in April 1999. Altogether there are 280 months with observations. The estimated irregular time series obtained by decomposing the extended time series with X-11-ARIMA is plotted in Figure 6.4. The highest and lowest estimated values of  $I_t$  occur in December 1997 and December 1995, respectively. This can be interpreted as intervention effects. The Box plot for each month is shown in Figure 6.5. We see that the spread of the distribution of March and April are little larger than what is the case for the other months. This may be due to problems with taking



Figure 6.4: *The estimated irregular component of X-11-ARIMA*

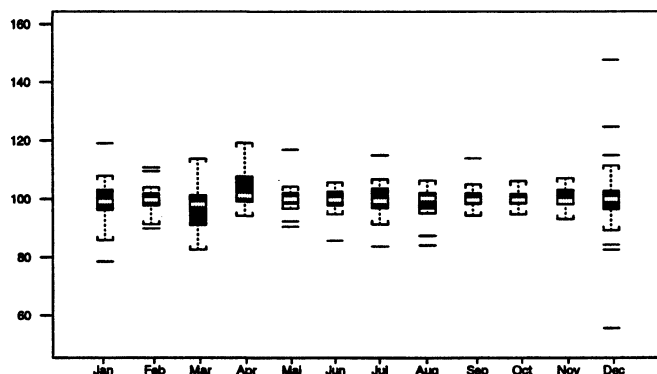


Figure 6.5: *The Box plot of the estimated irregular component from X-11-ARIMA*

properly account of the Easter effect. Most of the outliers occur in December. The presence of outliers in the estimated irregular component represents a problem. A preliminary way of dealing with this problem is to divide the months into two subgroups. The first subgroup consists of the 11 first months of the calendar year, while the second group consists only of December. Thus when we resample a value for the irregular component in period  $t$  and this is in one of the first eleven months we are drawing from the estimated values from these months (cf. Findley and Monsell (1990)). Totally we have 257 such observations. When we are simulating an irregular value for a month in December, we utilize only the estimated values for December. We have 23 such observations. The simulations were carried out both with and without taking account of the heteroscedasticity, but the results were not very different

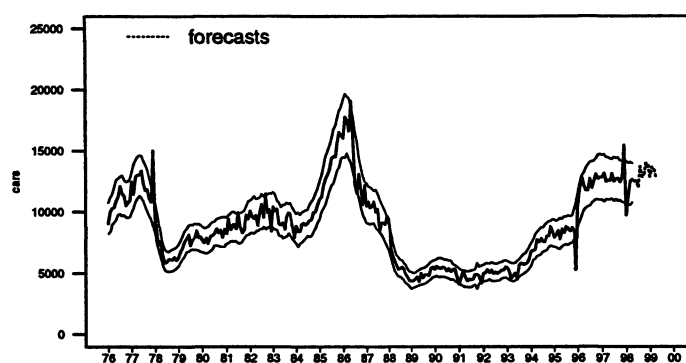


Figure 6.6: *The seasonally adjusted values and the estimated 95% confidence intervals. Non-parametric method with X-11-ARIMA*

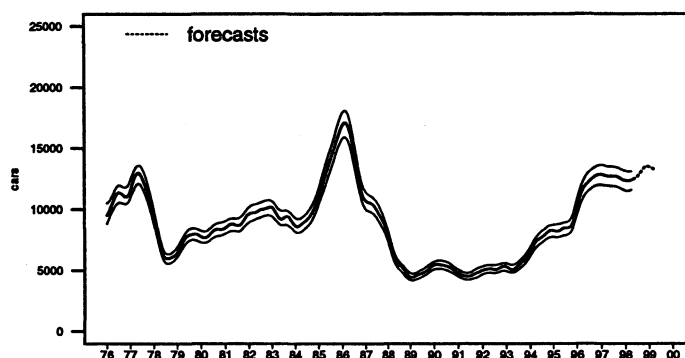


Figure 6.7: *The trend values and the estimated 95% confidence intervals. Non-parametric method with X-11-ARIMA*

and hence we only report the results based on the homoscedastic assumption. Figure 6.6 and 6.7 show the results. The empirical characteristics for the relative standard deviations are given in the third and fourth rows of Table 6.2. One sees that the estimated uncertainty is substantially less when using this method compared to what was the case when using method 1 (the parametric method).

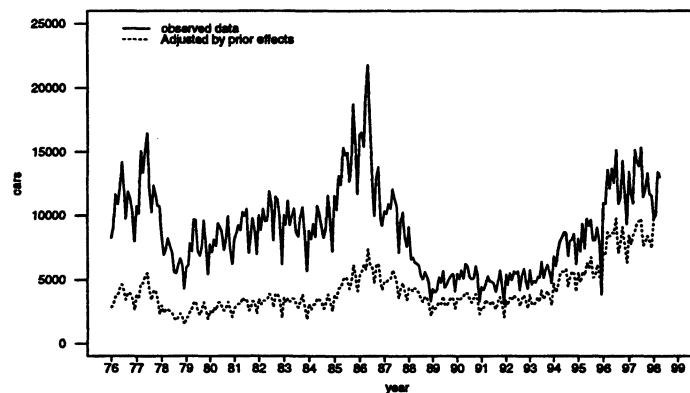
### 6.3.3 Application of method 2. The non-parametric method utilizing the estimated irregular component of X-12-ARIMA

As was the case with X-11-ARIMA none of the five built-in ARIMA models fulfilled the requirements. Thus taking into account prior effects as trading days, Easter and interventions were not sufficient in order to produce an adequate ARIMA model. In spite of this we stick to the most simple model, i.e.  $(0,1,1)(0,1,1)$ .

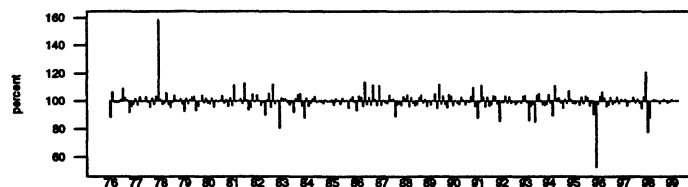
For outlier detection we operate with the default critical value of 3.8. Table 6.3 gives an overview of the intervention effects estimated by X-12-ARIMA. Two additive outliers (AO) and three level shifts (LS) were obtained. Substantial explanation of the two additive outliers was given at the top of section 3. The first two level shifts can be associated with the financial deregulation which took place in the middle of

Tabell 6.3: *The interventions estimated by X-12-ARIMA, with critical value 3.8*

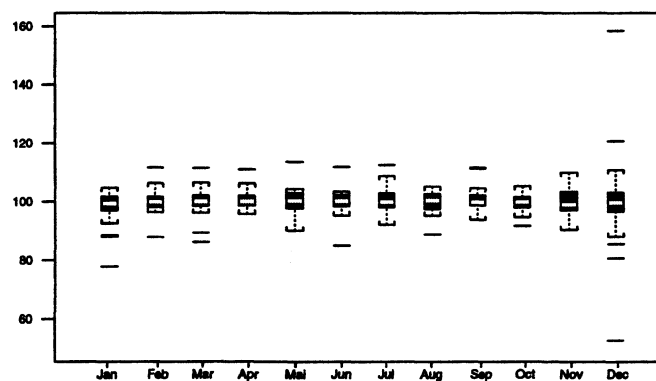
| Month    | Year | Type | Estimate | Std.error | t-value |
|----------|------|------|----------|-----------|---------|
| December | 1977 | AO   | 0.4560   | 0.06820   | 6.69    |
| July     | 1986 | LS   | -0.3217  | 0.08022   | -4.01   |
| February | 1988 | LS   | -0.3400  | 0.08026   | -4.24   |
| December | 1995 | AO   | -0.5989  | 0.06774   | -8.84   |
| January  | 1998 | LS   | -0.4178  | 0.08548   | -4.89   |



Figur 6.8: *The original data and adjusted for prior effects*



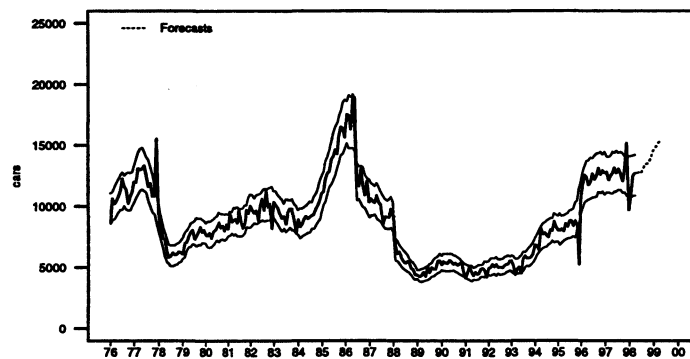
Figur 6.9: *The estimated irregular component from X-12-ARIMA*



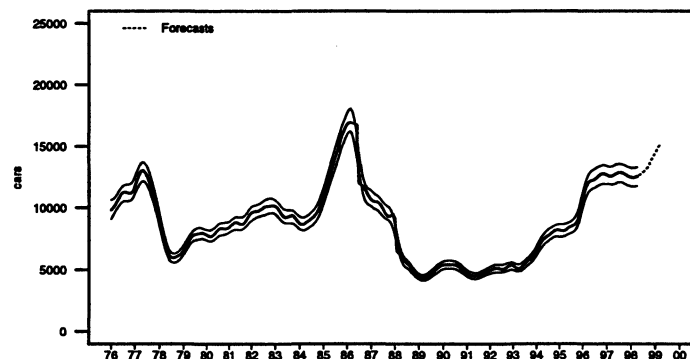
Figur 6.10: *The Box plot of the estimated irregular component from X-12-ARIMA*

Tabell 6.4: *The forecasts computed by X-11-ARIMA and X-12-ARIMA*

| Date     | X-11 ARIMA | X-12 ARIMA |          |       |
|----------|------------|------------|----------|-------|
|          | Forecast   | Lower      | Forecast | Upper |
| 1998.May | 13952      | 11804      | 14029    | 16673 |
| 1998.Jun | 13814      | 11323      | 14042    | 17414 |
| 1998.Jul | 14525      | 11986      | 15377    | 19727 |
| 1998.Aug | 11879      | 8339       | 11032    | 14595 |
| 1998.Sep | 12580      | 9777       | 13306    | 18109 |
| 1998.Oct | 13631      | 10290      | 14362    | 20045 |
| 1998.Nov | 12505      | 8475       | 12108    | 17298 |
| 1998.Dec | 10736      | 7408       | 10823    | 15813 |
| 1999.Jan | 12676      | 9565       | 14295    | 21366 |
| 1999.Feb | 12174      | 8855       | 13492    | 20557 |
| 1999.Mar | 14528      | 10937      | 16991    | 26396 |
| 1999.Apr | 13620      | 9941       | 15725    | 24876 |



Figur 6.11: *The seasonally adjusted values and the estimated 95% confidence intervals. Non-parametric case with X-12-ARIMA*



Figur 6.12: *The trend values and the estimated 95% confidence intervals. Non-parametric case with X-12-ARIMA*

the 1990's. As a result of this institutional change the real cost of borrowing money increased. Besides the price of new cars increased significantly in 1986. The last level shift effect, which is dated January 1998, should be interpreted with great care since the observation period ends in August the same year. Figure 6.8 shows the original series and the original series adjusted by prior effects. Based on the above mentioned regARIMA model we produce 12 months of forecasts. X-12-ARIMA is then applied to the extended time series and the resulting estimated irregular component is depicted in Figure 6.9. The Box plot of the estimated irregular component is given in Figure 6.10.

It is seen that the irregular components in March and April have a distribution that are rather similar to the other months. Thus the effect of the Easter holiday seems to be better captured in X-12-ARIMA. Even if we have taken into account the prior effects there seems to be some remaining outliers in the irregular series. This is likely to be the reason why no adequate ARIMA model is found. The resampling is based on the 280 elements of the estimated value of  $I_t$ . Table 6.4 shows the twelve months of forecasts computed by X-11-ARIMA and X-12-ARIMA.

Since the estimated  $I/C$  ratio is 1.69, the 13-term Henderson filter is used to compute the simulated trend in each replication. The simulated confidence intervals for the seasonally adjusted and the trend values are given in Figure 6.11 and 6.12, respectively. The last two rows of Table 6.2 contain empirical characteristics of the relative standard deviations of the replicated seasonally adjusted and the trend values. It is seen that the mean of both the relative standard deviation of the seasonally adjusted and of the trend values are lower than what was the case when the simulations were based on the estimated irregular component from X-11-ARIMA.

## 6.4 Concluding remarks

Two strategies based on simulations, one parametric and one non-parametric, for producing uncertainty measures for seasonally adjusted and trend time series have been provided. The parametric method, in which one replicates data from an ARIMA model, seems to produce rather large confidence intervals. Moreover it is rather time consuming since each of the  $B$  replicated time series are decomposed by means of X-11-ARIMA. The non-parametric method is more favourable since it is based on less rigid assumptions than the parametric case. However, an important assumption is that one is drawing from an independent sample of irregulars. Thus the estimated irregular component should not show any systematic variation (as for instance autocorrelation). This requirement may in practice not be fulfilled. It should also be emphasized that the non-parametric method (the way we have implemented it) is far less time consuming than the parametric method. The reason for this is that one only has to run X-11-ARIMA or X-12-ARIMA twice. The estimated confidence intervals are somewhat smaller when using X-12-ARIMA instead of X-11-ARIMA. The reason for this is probably connected to different treatments of the Easter holiday in the two programs and to the account of outliers when forecasting data in X-12-ARIMA.

In this paper only one time series has been considered, thus one should be cautious, when drawing conclusions. To gain more experience with the different simulation techniques it seems fruitful to consider a large number of time series with different time series properties (monthly as well as quarterly). Another important topic is to study how the number of replications influences the estimated confidence intervals.

## Kapittel 7

# Intervensjoner og trend - et eksempel med data for produksjon av kull og petroleumsprodukter.

Dinh Quang Pham

### 7.1 Innledning

Vi treffer ofte mange økonomiske tidsserier der eksterne begivenheter, som vi kaller intervensjoner, påvirker dataene kraftig. I X-12-ARIMA er det en rutine som brukes for å prekorrigere disse effektene før sesongjustering. Etter beregningen legger programmet tilbake de estimerte effektene for trenden og sesongjusterte tall slik at de to komponentene har samme nivå som rådata. Derfor får vi store utslag i trenden i tidspunkter der intervensjoner finnes. Dette er et problem ved publisering av endringstall. I estimeringen lar vi ofte programmet selv beregne tidspunkter, typer og effekter samtidig, isteden for å angi dette på forhånd. Dette kan skyldes at vi har mange tidsserier som sesongjusteres hvert publiseringstidspunkt, eller at vi mangler kunnskap om historisk utvikling av tidsserier. Av erfaring ser vi at

- (i) Det ikke alltid finnes en forklaring på hver intervensjon, særlig for tidspunktene som har lave  $t$ -verdier.
- (ii) Det kommer upassende intervensjoner som ødelegger trenden, på grunn av dårlig ARIMA modell brukt i estimeringen.

Vi fikser disse problemene ved å begrense antall intervensjoner etter størrelsen på  $t$ -verdier, eller fjerne de uønskede intervensjonene fra forrige kjøring og utføre en ny estimering.

Vi sesongjusterer ofte data i SSB med *log* transformering (dvs multiplikativ modell) uten å teste om denne modellen gir god tilpasning til dataene. Hvis vi velger en feil modell, kan vi få annerledes intervensjoner som ikke er tilpasset de opprinnelige dataene, fordi mønsteret til dataene blir endret etter transformeringen. Konsekvensen er store utslag i trenden.

Data som vi sesongjusterer i dette kapitlet er produksjon av kull og petroleums

produkter og kjerne brensel. Vi har kjørt data med X-12-ARIMA, med både multiplikativ og additiv modell. Resultater for hver kjøring blir vurdert og rettet opp før neste kjøring. I første omgang bruker vi bare default verdier.

## 7.2 Resultater

Data er produksjon av kull og petroleums produkter og kjerne brensel fra januar 1986 til desember 1999. Rådata er plottet i figur 7.1.

### 7.2.1 Første kjøring

Vi sesongjusterer data med *log* transformasjon, og får

- ARIMA modell

ARIMA(0,1,1)(0,1,1) modellen er valgt av programmet med  $\hat{\theta} = 0.3624$  og  $Std(\hat{\theta}) = 0.0730$ ,  $\hat{\Theta} = 0.7058$  og  $Std(\hat{\Theta}) = 0.0561$ . De to parametre er signifikante på 5% nivået.

Gjennomsnittlig absolutt prosent feil innen framskrivingens

siste år: 13.05, forrige år: 8.87, to år tidligere: 7.41, siste tre år: 9.77

som er mindre enn grenseverdien (15%). Vi får en god ARIMA modell som brukes til å estimere intervensjoner. En dårlig ARIMA modell kan gi ugyldig estimering.

- Intervensjoner

Tidspunkter og typer er gitt i tabell 7.1.

Tabell 7.1: Tidspunkter og *t*-verdier for intervensjoner

| <i>Måned</i> | <i>År</i> | <i>Type</i> | <i>t-verdi</i> |
|--------------|-----------|-------------|----------------|
| September    | 1991      | AO          | -10.12         |
| Oktober      | 1991      | AO          | -4.53          |
| September    | 1995      | AO          | -4.05          |
| Oktober      | 1995      | AO          | -8.12          |
| Mai          | 1999      | LS          | -6.80          |
| Juli         | 1999      | LS          | 5.87           |

Vi får to Level Shift i mai og juli 1999 med høye signifikante *t*-verdier. Trenden og sesongjusterte tall er beregnet fra data som er korrigert for effekter av disse intervensjonene.

- Trend

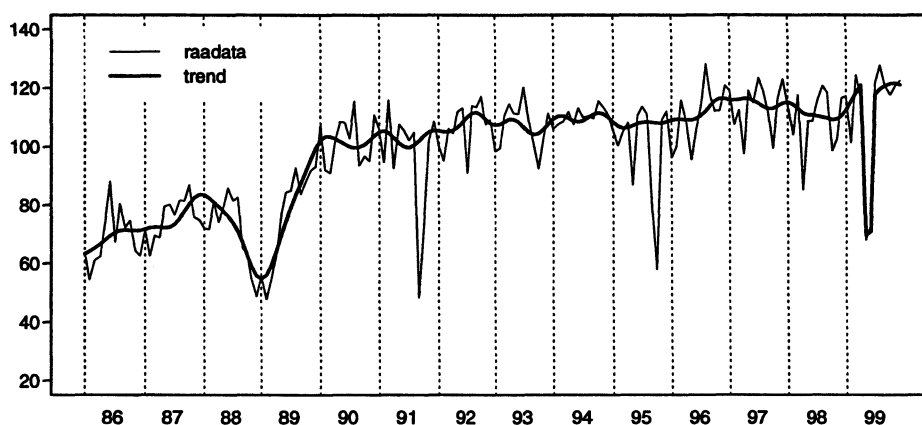
Rådata og trend er plottet i figur 7.1. Vi får et dårlig bilde av trenden i de siste observasjonene. Grunnen er at de korrigerte verdiene for intervensjonene i mai og juli 1999 er lagt tilbake i trenden etter sesongjustering, for å ha samme nivået som rådataene (se side 81 i manualen versjon 0.2.6 "Level shifts and the final Henderson trend").

Endringene i prosent fra måned til måned av trenden for 1999 er vist i tabell 7.2

Tabell 7.2: Endringen i prosent av trenden for 1999

| jan. | feb. | mar. | apr. | mai.   | jun. | jul.  | aug. | sep. | okt. | nov. | des.  |
|------|------|------|------|--------|------|-------|------|------|------|------|-------|
| 1.86 | 2.55 | 2.70 | 2.22 | -42.36 | 1.71 | 66.17 | 1.43 | 0.97 | 0.59 | 0.05 | -0.40 |

Vi ser at tallene i mai og juli 1999 er ganske store i forhold til de andre månedene, med henholdsvis -42.36% og 66.17%. Vi får et dårlig bilde av endringene i trenden i 1999.



Figur 7.1: Rådata og trend med log transformasjonen

### 7.2.2 Andre kjøring

Vi angir på forhånd tidspunkter og typer av intervensjoner som ble estimert i første kjøring av programmet, bortsett fra at Level Shift i mai og juni 1999 er erstattet med Additive Outlier. Data korrigeres med følgende intervensjoner før de sesongjusteres: AO1991.Sep, AO1991.Okt, AO1995.Sep, AO1995.Okt, AO1999.Mai og AO1999.Jul. ARIMA (0 1 1)(0 1 1) er valgt for estimeringen. Vi får

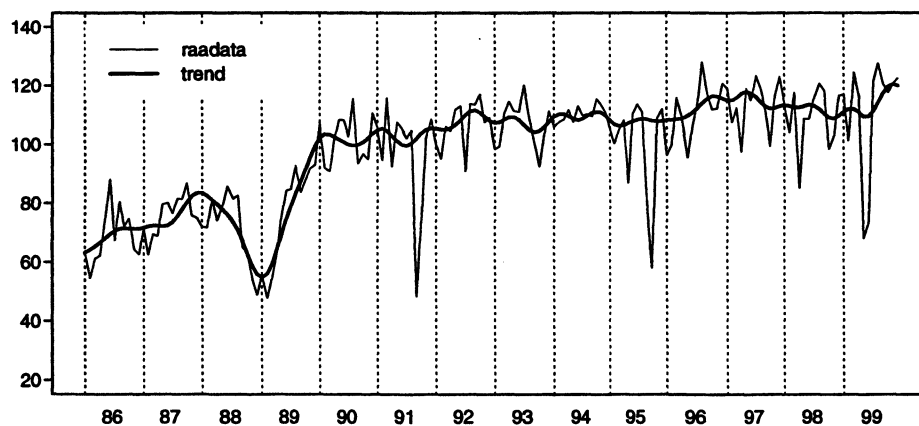
#### - Trend

Rådata og trend er plottet i figur 7.2. Vi får mye mindre utslag for trenden i mai og juli 1999.

Sammenlignet med tallene fra tabel 7.2, er endringene fra april til mai 1999 og fra juni til juli 1999 i andre kjøring mye mindre (se tabell 7.3).

#### - Intervensjoner

Det gir i tabell 7.4. AO1999.Jul er ikke signifikant på 5% nivået.



Figur 7.2: Rådata og trend i den andre kjøringen

Tabell 7.3: Endringen i prosent av trenden for 1999, andre kjøring

| jan. | feb. | mar.  | apr.  | mai.  | jun.  | jul. | aug. | sep. | okt. | nov. | des.  |
|------|------|-------|-------|-------|-------|------|------|------|------|------|-------|
| 1.34 | 0.92 | -0.04 | -1.03 | -1.22 | -0.06 | 1.60 | 2.67 | 2.77 | 2.08 | 0.63 | -0.58 |

Tabell 7.4: Tidspunkter og *t*-verdier for intervensjoner

| <i>Måned</i> | <i>År</i> | <i>Type</i> | <i>t-verdi</i> |
|--------------|-----------|-------------|----------------|
| September    | 1991      | AO          | -9.11          |
| Oktober      | 1991      | AO          | -4.09          |
| September    | 1995      | AO          | -3.70          |
| Oktober      | 1995      | AO          | -7.33          |
| Mai          | 1999      | AO          | -4.55          |
| Juli         | 1999      | AO          | 1.72           |

- Tester for sesongmønster og sesongbevegelser

Sesongmønsteret er signifikant på 1% nivået.

Det er også en signifikans på 5% nivået for sesongbevegelser.

- Kvalitetsmål og slidingspans

Det indikerer dårlig kvalitetssesongjustering.

### 7.2.3 Tredje kjøring

Testen viser at den additive modellen tilpasser dataene bedre enn den multiplikative modellen, som vi har brukt i de to første kjøringene. Vi estimerer effektene av intervensjoner, ukedager og påske med default verdier. Vi får

- Modell for sesongjustering

Additiv modell er valgt. Dvs  $O_t$  skrives slik  $O_t = T_t + S_t + I_t$ , hvor  $O_t$  er observerte data,  $T_t$ ,  $S_t$  og  $I_t$  er henholdsvis trend, sesong- og irregulær komponent.

- ARIMA modell

ARIMA(0,1,1)(0,1,1) modellen er valgt av programmet med  $\hat{\theta} = 0.5485$  og  $Std(\hat{\theta}) = 0.0646$ ,  $\hat{\Theta} = 0.8054$  og  $Std(\hat{\Theta}) = 0.0504$ . De to parametrene er signifikante på 5% nivået.

Gjennomsnittlig absolutt prosent feil innen framskrivingens

siste år: 8.72, forrige år: 9.26, to år tidligere: 6.57, siste tre år: 8.18

som er mindre enn grenseverdien (15%). Dette viser at ARIMA (0 1 1)(0 1 1) er en god modell for dataene. Sammenlignet med tallene fra samme tabell i første kjøring, får vi også mindre gjennomsnittlig absolutt prosent feil i framskrivingene.

- Intervensjoner

Disse vises i tabell 7.5. Vi får Additive Outlier i mai og juni 1999.

Tabell 7.5: Tidspunkter og  $t$ -verdier for intervensjoner

| <i>Måned</i> | <i>År</i> | <i>Type</i> | <i>t-verdi</i> |
|--------------|-----------|-------------|----------------|
| September    | 1991      | AO          | -6.06          |
| Oktober      | 1995      | AO          | -5.10          |
| Mai          | 1999      | AO          | -6.21          |
| Juni         | 1999      | AO          | -6.11          |

- Tester for sesongmønster og sesongbevegelser

Det er signifikant på 5% nivået for sesongmønsteret, men ikke for sesongbevegelser.

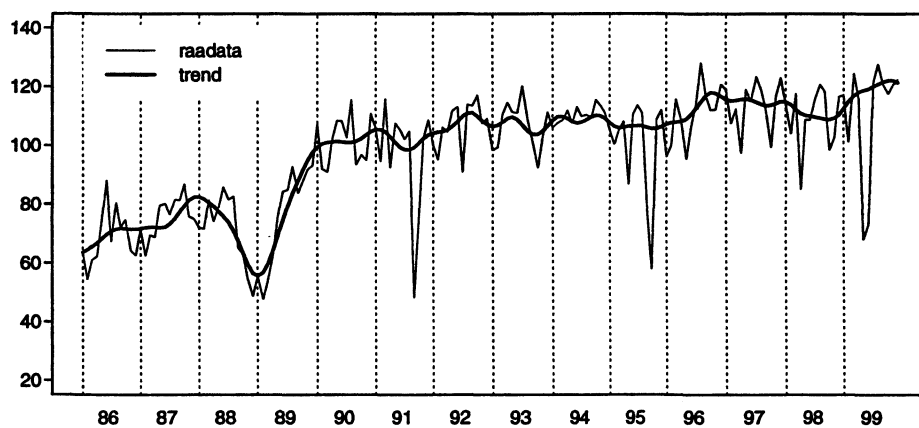
- Kvalitetsmål og slidingspans

11 mål for kvaliteten er gitt i tabellen nedenfor

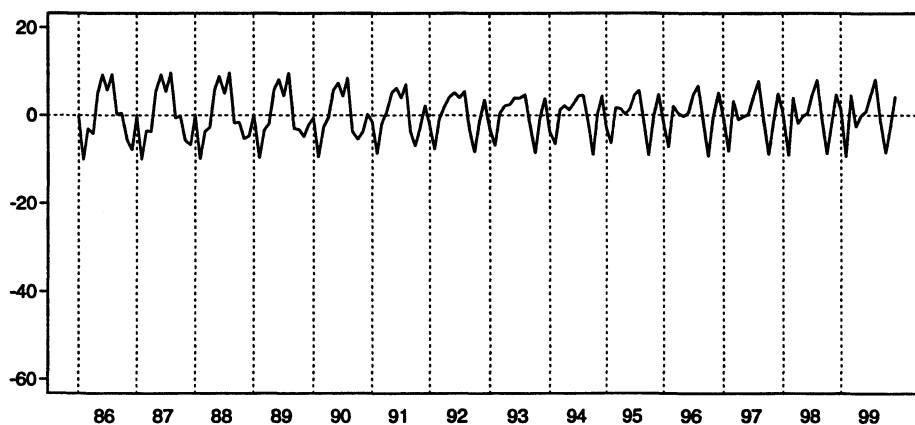
| $M1$       | $M2$  | $M3$  | $M4$  | $M5$  | $M6$  | $M7$  | $M8$  | $M9$  | $M10$ | $M11$ |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1.792      | 1.198 | 0.809 | 0.904 | 1.190 | 0.176 | 0.901 | 1.339 | 0.667 | 1.251 | 1.139 |
| $Q = 1.01$ |       |       |       |       |       |       |       |       |       |       |

Viktige mål som  $M1$ ,  $M2$ ,  $M5$ ,  $M8$  er større enn 1. Vi får også store andeler av måneder der sesongfaktorene er identifisert som upåtelige. Dette indikerer dårlig sesongjustering, og sesongjusterte tall bør derfor ikke publiseres. Trenden og rådata er plottet i figur 7.3.

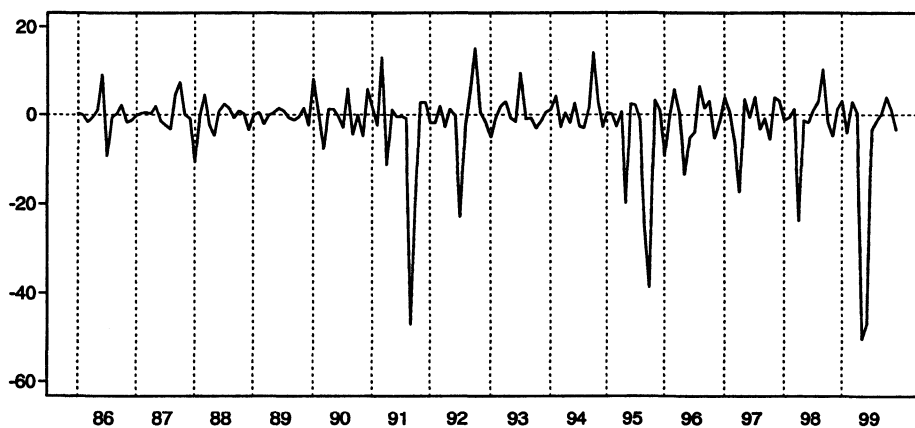
Sesongkomponenten er vist i figur 7.4. Vi ser at mønsteret endres over tid, og derfor ikke kan identifiseres klart. Vi får dårlige sesongjusterte tall. Figur 7.5 viser store variasjoner i den irregulære komponenten.



Figur 7.3: Rådata og trend i tredje kjøringen



Figur 7.4: Sesongkomponenten i tredje kjøringen



Figur 7.5: Irregulærkomponenten i tredje kjøringen

### 7.3 Oppsummering

Additiv modell tilpasser dataene bedre. På grunn av dårlige sesongjusterte tall, bør vi bare publisere rådata og trenden. Programmet som anbefalles for dataene er

```
series{title="produksjon av kull og petro og kjerne brensel"  
      start=1986.1  
      period=12  
      file="kull.dat"  
      decimals=2}  
transform{function=none}  
regression{variables=(ao1991.sep ao1995.oct ao1999.may ao1999.jun)}  
arima{model=(0 1 1)(0 1 1)}  
forecast{}  
estimate{}  
x11{type=trend}
```

Vi kan bruke dette programmet i løpet av et halvt år med korrigeringen for bestemte tidspunkter og typer av intervensjonene. Deretter estimerer vi ARIMA modell, effektene av ukedager, påske og intervensjoner når nye data tilføyes.

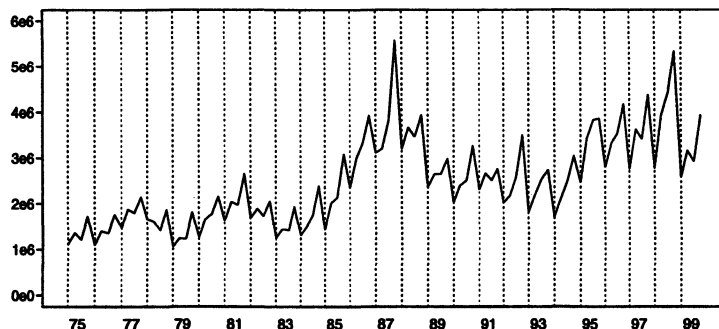
## Kapittel 8

# Faktisk utførte investeringer totalt for industrien

Dinh Quang Pham

### 8.1 Innledning

Data er observert fra første kvartal 1975 til fjerde kvartal 1999. Sammenlignet med den tidsserien for de antatte investeringene, har dataene for de utførte investeringene større utslag. Rådata er plottet i figur 8.1. Den første delen viser en oversikt av variasjoner i rådata i perioden fra 1975 til 1999. Mens den andre delen gir oss et godt bilde av sesongvariasjoner. Vi ser et klart sesongmønster for rådata.



Figur 8.1: *Rådata*

Ved sesongjustering for denne tidsserien får vi

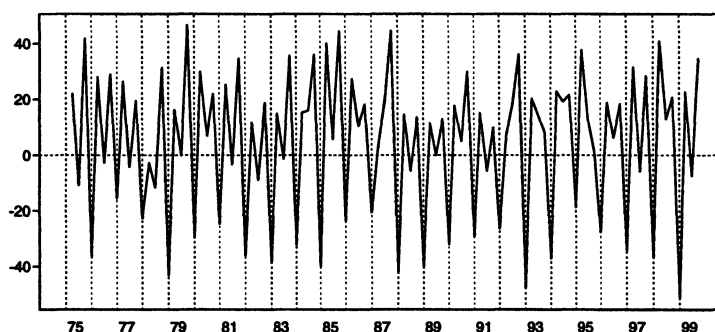
- ustabil sesongkomponent og sesongjusterte tall slik at vi ikke kan publisere tallene. Dvs endringene av sesongjusterte tall er store når tidsserien forlenges med nye observasjoner.
- Ingen av de fem innebygde ARIMA modeller tilfredsstiller kriteriene i programmet. Dette skyldes store avviker mellom data og modellene.

Dagum (1980) har innført ARIMA modell i X-11 metoden for å unngå å beregne med asymmetriske vektorer for de siste tidspunktene som gir lav kvalitet for sesongjusterte tall i slutten av tidsserien. I X-11-ARIMA anvender man ARIMA modell

bare for å framskrive en tidsserie. Kalendereffekter estimeres fra den foreløpige irregulære komponenten ved regresjonsanalyse. Ofte er denne komponenten ikke en hvit støy prosess. Dermed kan resultatene være ugyldige. Findley og Monsell forbedrer svakheter i X-11-ARIMA ved å anvende ARIMA modell for rådata i estimeringer for kalendereffekter og intervensjoner. Dermed er antagelsen om hvit støy prosess av den foreløpige irregulære komponenten ikke nødvendig. Uten ARIMA modell kan rådata ikke prekorrigeres for effektene av ukedager, påske og intervensjoner. I dette tilfellet bruker X-12-ARIMA rådata som input for å beregne trend, sesongkomponent og sesongjusterte tall. Konsekvensen er at vi får et falskt bilde av sesongmønsteret og sesongjusterte tall har store revisjoner i slutten av tidsserien. For å unngå å beregne sesongjusterte tall direkte fra rådata, kan man angi en ARIMA modell på forhånd for programmet. Men ofte blir det ikke bedre.

Figur 8.2 viser endringene i prosent fra kvartal til kvartal av rådata. Tallene varierer fra -51.6% i første kvartal 1999 til 46.8% i fjerde kvartal 1979.

Vi har en lang tidsserie hvor observasjonene går over 25 år. Siden dataene endrer seg raskt fra periode til periode, er det ikke realistisk å anta at effektene av ukedager er de samme for hele serien slik at regresjonsanalyse kan anvendes. Dette gjelder også for påskeeffekt. I estimeringen antar vi at effekten påvirker data i samme periode for hele serien.



Figur 8.2: Endringene i prosent fra kvartal til kvartal av rådata

For de utførte investeringene vil vi bruke en delmengde av de opprinnelige dataene for sesongjustering. Da blir det lettere å finne en ARIMA modell som gir god tilpasning med data. Antagelsene om at kalendereffekter er konstant med tid, kan være oppfylt.

Vi presenterer i neste avsnitt metoder for å lage nye tidsserier og resultater fra sesongjustering.

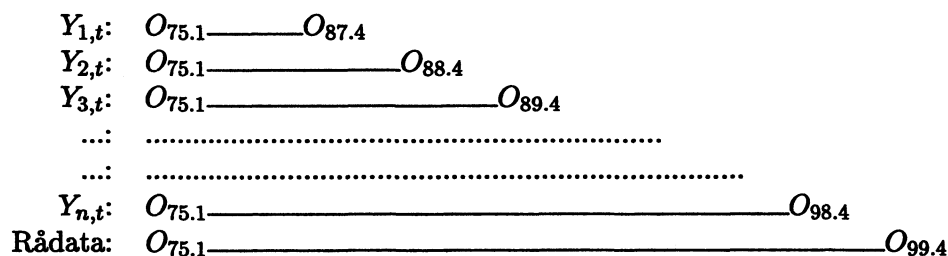
## 8.2 Nye tidsserier

### 8.2.1 Å lage $Y_{1,t}, Y_{2,t}, \dots, Y_{n,t}$ ved å legge til gradvis nye observasjoner

Fra de observerte dataene  $O_{75.1}, O_{75.2}, \dots, O_{99.4}$ , ( $O_{75.1}$  står for verdien i første kvartal 1975), lager vi en rekke av tidsseriene  $Y_{1,t}, Y_{2,t}, \dots, Y_{n,t}$  på følgende måte:

- Den første serien  $Y_{1,t}$  inneholder observasjonene fra 1975 til 1987,  $Y_{1,t} = \{O_{1kv.75}, O_{2kv.75}, \dots, O_{4kv.87}\}$ . Lengden er lang nok til å estimere ARIMA parametre og

Figur 8.3: Tidsseriene  $Y_{1,t}$ ,  $Y_{2,t}$ , ...,  $Y_{n,t}$



kalendereffekter.

- Den neste serien  $Y_{2,t}$  er et år lengre enn  $Y_{1,t}$  ved å legge til observasjonene av 1988. Vi får  $Y_{2,t} = \{O_{1kv.75}, \dots, O_{4kv.87}, O_{1kv.88}, \dots, O_{4kv.88}\}$ .

- Vi gjør på samme måte for  $Y_{3,t}$ ,  $Y_{4,t}$ , ... Tilslutt har vi en rekke av tidsseriene med egenskapen  $Y_{1,t} \subset Y_{2,t} \subset \dots \subset Y_{n,t}$  og som går mot de opprinnelige dataene. Symbolet  $Y_{i,t} \subset Y_{j,t}$  betyr at  $Y_{i,t}$  er en delmengde av  $Y_{j,t}$ . Metoden illustreres i figur 8.3.

Resultater fra X-12-ARIMA for  $Y_{1,t}$ ,  $Y_{2,t}$ , ..., er listet ut i tabell 8.1.

Tabell 8.1: Sesongjustering av  $Y_{1,t}$ ,  $Y_{2,t}$ , ...,  $Y_{n,t}$

| Tids. | ARIMA          | Effekt |       |       | Sliding. i % |       | Inter. |
|-------|----------------|--------|-------|-------|--------------|-------|--------|
|       |                | Uke.   | Påske | Q     | S(%)         | MM(%) |        |
| 75-87 | (0 1 1)(0 1 1) | nei    | nei   | 0.332 | 0.0          | 14.3  | nei    |
| 75-88 | nei            | nei    | nei   | 0.344 | 0.0          | 22.9  | nei    |
| 75-89 | nei            | nei    | nei   | 0.324 | 2.8          | 17.1  | nei    |
| 75-90 | nei            | nei    | nei   | 0.337 | 0.0          | 8.6   | nei    |
| 75-91 | (0 2 2)(0 1 1) | nei    | nei   | 0.285 | 2.8          | 14.3  | nei    |
| 75-92 | nei            | nei    | nei   | 0.276 | 11.1         | 28.6  | nei    |
| 75-93 | nei            | nei    | nei   | 0.341 | 8.3          | 45.7  | nei    |
| 75-94 | nei            | nei    | nei   | 0.314 | 5.6          | 22.9  | nei    |
| 75-95 | nei            | nei    | nei   | 0.277 | 5.6          | 20.0  | nei    |
| 75-96 | (2 1 2)(0 1 1) | nei    | nei   | 0.291 | 8.3          | 25.7  | nei    |
| 75-97 | (2 1 1)(0 1 1) | nei    | nei   | 0.316 | 5.6          | 22.9  | nei    |
| 75-98 | nei            | nei    | nei   | 0.323 | 5.6          | 22.9  | nei    |
| 75-99 | nei            | nei    | nei   | 0.333 | 25.0         | 54.3  | nei    |

Grenseverdiene for  $Q$  er 1,  $S(\%)$  er 15 og  $MM(\%)$  er 35

Vi har her innført to symboler:  $S(\%)$  er prosentandelen av kvartaler hvor sesongkomponenten er definert som upålitelige og  $MM(\%)$  er prosentandelen av kvartaler hvor endringene av sesongjusterte tall identifiseres som upålitelige. De måler stabiliteter for sesongkomponenten og sesongjusterte tall. Grenseverdiene for  $S(\%)$  er 15 og for  $MM(\%)$  er 35.

Vi ser at rådataene ikke er påvirket av ukedagseffekter og påskeeffekt. Det finnes

ingen ARIMA modell som tilfredsstiller kriterier fra X-12-ARIMA for de fleste tids-seriene ellers tilpasses dataene med kompliserte modeller. Lave verdier av  $Q$ ,  $S(\%)$  og  $MM(\%)$  viser stabile sesongjusterte tall, bortsett fra [75-93] og [75-99].

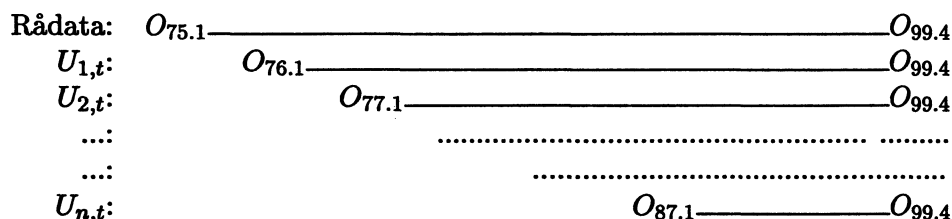
## 8.2.2 Å lage $U_{1,t}$ , $U_{2,t}$ , ..., $U_{n,t}$ ved å fjerne gradvis gamle observasjo- ner

I motsetning til  $Y_{1,t}$ ,  $Y_{2,t}$ , ...,  $Y_{n,t}$ , konstruerer vi de nye tidsseriene  $U_{1,t}$ ,  $U_{2,t}$ , ...,  $U_{n,t}$  på følgende måte:

- Fra de opprinnelige dataene  $O_{75.1}$ ,  $O_{75.2}$ , ...,  $O_{99.4}$  lager vi en ny tidsserie  $U_{1,t}$  som er et år kortere ved å fjerne tallene i 1975. Dvs,  $U_{1,t}$  inneholder observasjonene fra 1976 til 1999,  $U_{1,t} = \{O_{76.1}, O_{76.2}, \dots, O_{99.4}\}$ .
- Fra  $U_{1,t}$  lager vi  $U_{2,t}$  ved å fjerne tallene i 1976. Vi får  $U_{2,t} = \{O_{77.1}, O_{77.2}, \dots, O_{99.4}\}$ .
- Vi gjør på samme måte for  $U_{3,t}$ , osv. Den siste tidsserien  $U_{n,t}$  med observasjonene fra 1987 til 1999,  $U_{n,t} = \{O_{87.1}, O_{87.2}, \dots, O_{99.4}\}$ . For kortere tidsserier kan vi få dårlige estimater for parametrene til ARIMA modellen.

Vi får en rekke av tidsseriene  $U_{1,t}$ ,  $U_{2,t}$ , ...,  $U_{n,t}$  med egenskapen  $U_{1,t} \supset U_{2,t} \supset \dots \supset U_{n,t}$ . Symbolet  $U_{i,t} \supset U_{j,t}$  betyr at  $U_{j,t}$  er en delmengde av  $U_{i,t}$ .  $U_{n,t}$  inneholder kun de senere observasjonene. Metoden illustreres i figur 8.4.

Figur 8.4: Tidsseriene  $U_{1,t}$ ,  $U_{2,t}$ , ...,  $U_{n,t}$



Resultater fra X-12-ARIMA av  $U_{1,t}$ ,  $U_{2,t}$ , ...,  $U_{n,t}$  er vist i tabell 8.2.

Tabell 8.2: Sesongjustering av  $U_{1,t}$ ,  $U_{2,t}$ , ...,  $U_{n,t}$

| Tids. | ARIMA | Effekt |       |       | Sliding. i % |          | Inter. |
|-------|-------|--------|-------|-------|--------------|----------|--------|
|       |       | Uke.   | Påske | $Q$   | $S(\%)$      | $MM(\%)$ |        |
| 75-99 | nei   | nei    | nei   | 0.333 | 25.0         | 54.3     | nei    |
| 76-99 | nei   | nei    | nei   | 0.317 | 25.0         | 54.3     | nei    |
| 77-99 | nei   | nei    | nei   | 0.281 | 25.0         | 54.3     | nei    |
| 78-99 | nei   | nei    | nei   | 0.282 | 25.0         | 54.3     | nei    |
| 79-99 | nei   | nei    | ja    | 0.260 | 27.8         | 48.6     | nei    |
| 80-99 | nei   | nei    | ja    | 0.272 | 25.0         | 48.6     | nei    |

(tabell 8.2 fortsatt)

| Tids. | ARIMA          | Effekt |       | $Q$   | Sliding. i % |          | Inter. |
|-------|----------------|--------|-------|-------|--------------|----------|--------|
|       |                | Uke.   | Påske |       | $S(\%)$      | $MM(\%)$ |        |
| 81-99 | nei            | nei    | ja    | 0.261 | 27.8         | 48.6     | nei    |
| 82-99 | nei            | nei    | ja    | 0.279 | 33.3         | 57.1     | nei    |
| 83-99 | (2 1 2)(0 1 1) | ja     | ja    | 0.231 | 0.0          | 17.1     | nei    |
| 84-99 | (0 1 2)(0 1 1) | ja     | ja    | 0.273 | 2.8          | 28.6     | nei    |
| 85-99 | (0 2 2)(0 1 1) | ja     | ja    | 0.251 | 0.0          | 14.3     | AO3.99 |
| 86-99 | (2 1 0)(0 1 1) | ja     | ja    | 0.272 | 0.0          | 11.4     | AO3.99 |
| 87-99 | (0 1 1)(0 1 1) | ja     | ja    | 0.167 | 0.0          | 5.7      | AO3.99 |

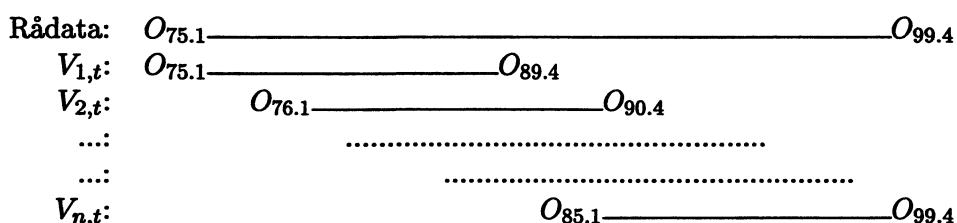
Vi ser at når dataene før 1983 blir fjernet får vi bedre resultater for sesongjustering. Det finnes ARIMA modeller for dataene, kalendereffekter er signifikante på 5% nivå og sesongkomponenten og sesongjusterte tall er ganske stabile slik at tallene kan publiseres. Ved å forlenge tidsserier tilbake blir resultatene dårligere.

### 8.2.3 Tidsserier med fast lengde

Her lager vi tidsseriene med samme antall observasjoner. Vi deler i to tilfeller

a. Å lage  $V_{1,t}$ ,  $V_{2,t}$ , ...,  $V_{n,t}$  med 15 års periodelengden. Den første tidsserien som vi velger,  $V_{1,t}$  inneholder observasjoner fra 1975 til 1989,  $V_{1,t} = \{O_{75.1}, O_{75.2}, \dots, O_{89.4}\}$ . Den neste serien  $V_{2,t}$  har samme lengde som  $V_{1,t}$ , men den starter senere et år. Vi får  $V_{2,t} = \{O_{76.1}, O_{75.2}, \dots, O_{90.4}\}$ . Dvs, når nye observasjoner legges til, blir de tidligste verdiene fjernet med samme antallet. Vi gjør på samme måte for  $V_{3,t}$ , ...,  $V_{n,t}$ . Da får vi en rekke av tidsseriene  $V_{1,t}$ ,  $V_{2,t}$ , ...,  $V_{n,t}$ , hvor  $V_{n,t} = \{O_{85.1}, O_{85.2}, \dots, O_{99.4}\}$ . Metoden illustreres i figur 8.5. Denne teknikken gir oss et bilde om stabiliteten av sesongjusterte tall.

Figur 8.5: Tidsseriene  $V_{1,t}$ ,  $V_{2,t}$ , ...,  $V_{n,t}$



Resultater fra X-12-ARIMA av  $V_{1,t}$ ,  $V_{2,t}$ , ...,  $V_{n,t}$  presenteres i tabell 8.3.

Tabell 8.3: Sesongjustering av  $V_{1,t}$ ,  $V_{2,t}$ , ...,  $V_{n,t}$

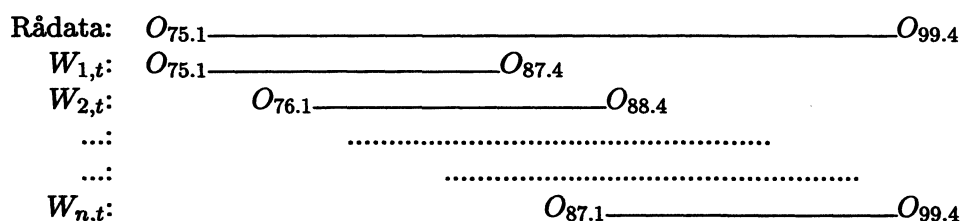
| Tids. | ARIMA          | Effekt |       | $Q$   | Sliding. i % |          | Inter. |
|-------|----------------|--------|-------|-------|--------------|----------|--------|
|       |                | Uke.   | Påske |       | $S(\%)$      | $MM(\%)$ |        |
| 75-89 | nei            | nei    | nei   | 0.324 | 2.8          | 17.1     | nei    |
| 76-90 | nei            | nei    | nei   | 0.270 | 0.0          | 8.6      | nei    |
| 77-91 | (0 1 2)(0 1 1) | nei    | nei   | 0.217 | 0.0          | 5.7      | nei    |

(tabell 8.3 fortsatt)

| Tids. | ARIMA          | Effekt |       | $Q$   | Sliding. i % |          | Inter. |
|-------|----------------|--------|-------|-------|--------------|----------|--------|
|       |                | Uke.   | Påske |       | $S(\%)$      | $MM(\%)$ |        |
| 78-92 | (0 1 2)(0 1 1) | nei    | nei   | 0.192 | 0.0          | 8.6      | nei    |
| 79-93 | (0 1 1)(0 1 1) | nei    | ja    | 0.228 | 0.0          | 31.4     | nei    |
| 80-94 | (0 1 1)(0 1 1) | nei    | ja    | 0.249 | 2.8          | 20.0     | nei    |
| 81-95 | nei            | nei    | ja    | 0.211 | 5.6          | 20.0     | nei    |
| 82-96 | (0 1 1)(0 1 1) | nei    | ja    | 0.209 | 11.1         | 17.1     | nei    |
| 83-97 | (0 1 1)(0 1 1) | ja     | ja    | 0.211 | 5.6          | 11.4     | nei    |
| 84-98 | (0 1 1)(0 1 1) | nei    | ja    | 0.306 | 2.8          | 8.6      | nei    |
| 85-99 | (0 2 2)(0 1 1) | ja     | ja    | 0.251 | 0.0          | 14.3     | AO99.3 |

Tabellene 8.2 og 8.3 viser to rekker av tidsserier:  $\{U_{1,t}, U_{2,t}, \dots, U_{n,t}\}$  og  $\{V_{1,t}, V_{2,t}, \dots, V_{n,t}\}$ . Disse tidsseriene har henholdvis samme startpunkter og sluttpunkter, men lengdene er ikke de samme. Vi ser at ARIMA modeller tilpasser data bedre og de sesongjusterte tallene er mer stabile (lavere verdiene av  $S(\%)$  og  $MM(\%)$ ) når tidsseriene blir kortere. For eksempel, det finnes ingen ARIMA modell for tidsserien [77-99], men for [77-91] er ARIMA (0 1 2)(0 1 1).

b. Å lage  $W_{1,t}, W_{2,t}, \dots, W_{n,t}$  med 13 års periodelengden. Vi lager tidsseriene etter samme prinsippet som vi har brukt i 8.2.3.a. Tidsseriene  $W_{1,t}, W_{2,t}, \dots, W_{n,t}$  er beskrevet i figur 8.6.

Figur 8.6: Tidsseriene  $W_{1,t}, W_{2,t}, \dots, W_{n,t}$ 

Resultatene fra X-12-ARIMA er vist i tabell 8.4.

tabell 8.4 fortsatt

| Tids. | ARIMA          | Effekt |       | $\hat{d}$ | $Q$   | Sliding. i % |          | Inter. |
|-------|----------------|--------|-------|-----------|-------|--------------|----------|--------|
|       |                | Uke.   | Påske |           |       | $S(\%)$      | $MM(\%)$ |        |
| 85-97 | (0 1 1)(0 1 1) | ja     | ja    | 15        | 0.255 | 5.6          | 11.4     | nei    |
| 86-98 | (0 1 1)(0 1 1) | ja     | ja    | 15        | 0.244 | 0.0          | 8.6      | nei    |
| 87-99 | (0 1 1)(0 1 1) | ja     | ja    | 15        | 0.167 | 0.0          | 5.7      | AO3.99 |

$\hat{d}$  er antall dager før påskesøndag som rådata er påvirket av effekten. Vi ser at ARIMA modellene for  $\{V_{1,t}, V_{2,t}, \dots, V_{n,t}\}$  (se tabell 8.3) endrer seg litt med tidsintervaller, men for  $\{W_{1,t}, W_{2,t}, \dots, W_{n,t}\}$  er det samme ARIMA modell for dataene fra 1979 og seinere. Kalendereffekter påvirker dataene i en kort periode før det siste tidspunktet. Det finnes en additiv outlier for de utførte investeringene i tredje kvartal 1999.

Vi har ikke en god forklaring for ukedagseffekter for de utførte investeringene, selv om beregningen viser høy signifikans for effektene. Uten å ta hensyn til ukedags-

Tabell 8.4: Sesongjustering for  $W_{1,t}, W_{2,t}, \dots, W_{n,t}$

| Tids. | ARIMA          | Effekt |       | $\hat{d}$ | $Q$   | Sliding. i % |          | Inter. |
|-------|----------------|--------|-------|-----------|-------|--------------|----------|--------|
|       |                | Uke.   | Påske |           |       | $S(\%)$      | $MM(\%)$ |        |
| 75-87 | (0 1 1)(0 1 1) | nei    | nei   |           | 0.332 | 0.0          | 14.3     | nei    |
| 76-88 | (2 1 2)(0 1 1) | ja     | nei   |           | 0.222 | 0.0          | 11.4     | nei    |
| 77-89 | nei            | nei    | nei   |           | 0.234 | 2.8          | 17.1     | nei    |
| 78-90 | nei            | nei    | nei   |           | 0.232 | 0.0          | 8.6      | nei    |
| 79-91 | (0 1 1)(0 1 1) | nei    | ja    | 15        | 0.186 | 0.0          | 5.7      | nei    |
| 80-92 | (0 1 1)(0 1 1) | nei    | ja    | 15        | 0.276 | 0.0          | 20.0     | nei    |
| 81-93 | (0 1 1)(0 1 1) | nei    | nei   |           | 0.284 | 5.6          | 37.1     | nei    |
| 82-94 | (0 1 1)(0 1 1) | nei    | ja    | 15        | 0.243 | 2.8          | 17.1     | nei    |
| 83-95 | (0 1 1)(0 1 1) | ja     | ja    | 15        | 0.189 | 5.6          | 14.3     | nei    |
| 84-96 | (0 1 1)(0 1 1) | ja     | ja    | 15        | 0.187 | 5.6          | 11.4     | nei    |

effektene i sesongjusteringen får vi høye verdier av  $S(\%)$  og  $MM(\%)$  som viser ustabile sesongjusterte tall. Dette vises i tabell 8.5.

Tabell 8.5:

| Tids. | ARIMA | Påske | $Q$   | Sliding. i % |          | Inter. |
|-------|-------|-------|-------|--------------|----------|--------|
|       |       |       |       | $S(\%)$      | $MM(\%)$ |        |
| 85-99 | nei   | ja    | 0.305 | 33.3         | 57.1     | nei    |
| 87-99 | nei   | ja    | 0.264 | 28.1         | 48.4     | nei    |

**Konklusjon.** Resultatene fra tabellene 8.1-8.5 viser at for de utførte investeringene er det ikke en fordel å ta hele tidsserien fra 1975 til 1999 for sesongjustering. Det er to kandidater:  $V=\{O_{85.1}, O_{85.2}, \dots, O_{99.4}\}$  og  $W=\{O_{87.1}, O_{87.2}, \dots, O_{99.4}\}$ . Vi får bedre sesongjusterte tall for tidsserien  $W$  som inneholder observasjonene fra 1987 til 1999.

### 8.3 Resultater fra sesongjustering for tidsserien $\{O_{87.1}, O_{87.2}, \dots, O_{99.4}\}$

Vi skriver her i mer detaljert form

- Modell for sesongjustering

Dataene er kjørt med multiplikativ modell. Dvs,  $O_t$  skrives på formen

$$O_t = T_t \times S_t \times I_t$$

- ARIMA modell

ARIMA (0 1 1)(0 1 1) er valgt, med

| Par.           | Estimert |        |                  |
|----------------|----------|--------|------------------|
|                | Verdi    | Std    |                  |
| $\hat{\theta}$ | -0.2928  | 0.1301 | sign. på 5% nivå |
| $\hat{\Theta}$ | 0.8031   | 0.0860 | sign. på 5% nivå |

Gjennomsnittlig absolutt prosent feil innen framskrivninger:

siste år: 22.46, forrige år: 9.56, to år tidligere: 5.96

siste tre år: 12.66. Grenseverdien for denne størrelsen er 15%.

#### - Kalendereffekter

Ukedageffekter er signifikante på 5% nivå, med  $\chi^2 = 51.76$  og 6 frihetsgrader.

Påskeeffekt er også signifikant med  $t = -2.77$ ,  $\hat{d} = 15$ .

#### - Intervensjoner

En additiv outlier i tredje kvartal 1999.

#### - Framskrivninger

Framskrivningene av rådata med og uten prekorrigeringen for kalendereffekter og intervensjoner vises i nedenfor tabell

| Dato      | Rådata uten prekorrr. |            |            | med prekorrr. |
|-----------|-----------------------|------------|------------|---------------|
|           | Nedre                 | Fram.      | Øvre       | Fram.         |
| 1.kv 2000 | 2240250.45            | 2650947.80 | 3136936.85 | 2570422.02    |
| 2.kv 2000 | 2335554.66            | 3034602.32 | 3942879.77 | 3103875.88    |
| 3.kv 2000 | 2358943.72            | 3281891.28 | 4565946.33 | 3321410.65    |
| 4.kv 2000 | 2509877.03            | 3708207.16 | 5478674.92 | 3910986.30    |

Merk at X-12-ARIMA lister ikke ut 95% konfidensintervaller for framskrivningene som er beregnet av rådata med prekorrigeringen.

#### - Tester for sesongmønster og sesongbevegelser

$F$ -testene viser at sesongmønsteret er signifikant på 5% nivå,  $F=455.84$ . Men det er ikke signifikant for sesongbevegelser på samme nivå,  $F=0.442$ .

#### - Kvalitetsmål og sliding spans

De 11 målene og  $Q$  verdien er gitt i tabell

| $M1$        | $M2$  | $M3$  | $M4$  | $M5$  | $M6$  | $M7$  | $M8$  | $M9$  | $M10$ | $M11$ |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0.087       | 0.030 | 0.090 | 0.606 | 0.200 | 0.095 | 0.096 | 0.240 | 0.181 | 0.269 | 0.216 |
| $Q = 0.167$ |       |       |       |       |       |       |       |       |       |       |

Basert på disse målene får vi en god sesongjustering.

Sliding spans er utført med fire delvis overlappende delmengder [1989-1996], [1990-1997], [1991-1998] og [1992-1999]. Vi får  $S(\%) = 0$  og  $MM(\%) = 5.7$ . Disse tallene er mye lavere enn deres grenseverdier. Vi får god sesongjustering.

#### - Revisjoner

Revisjoner i prosent for sesongjusterte tall er beregnet fra første kvartal 1995 til tredje kvartal 1999. Tallene er gitt i nedenfor tabell

| år   | 1. kvartal | 2. kvartal | 3. kvartal | 4. kvartal |
|------|------------|------------|------------|------------|
| 1995 | 6.09       | -0.88      | -2.69      | 0.28       |
| 1996 | 1.52       | -0.48      | 1.75       | 5.75       |
| 1997 | 0.94       | -1.05      | 1.98       | -3.19      |
| 1998 | 2.19       | 1.66       | -0.07      | -1.58      |
| 1999 | -3.67      | 0.33       | 0.18       |            |

La  $R_{k,t}$  være revisjonen i kvartal  $k$  og år  $t$ .

Gjennomsnittlige revisjoner i prosent for sesongjusterte tall,

- for hvert kvartal, er:

$$\bar{R}_{1kv} = 2.88, \bar{R}_{2kv} = 0.88, \bar{R}_{3kv} = 1.34 \text{ og } \bar{R}_{4kv} = 2.70.$$

- for hvert år gjennom fire kvartaler, er:

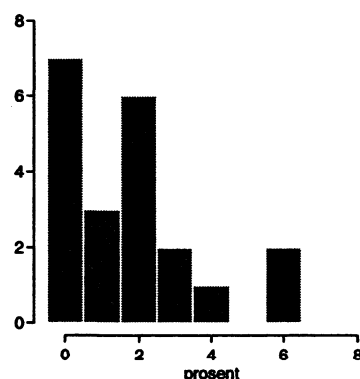
$$\bar{R}_{1995} = 2.49, \bar{R}_{1996} = 2.37, \bar{R}_{1997} = 1.79, \bar{R}_{1998} = 1.37 \text{ og } \bar{R}_{1999} = 1.39.$$

- for hele serien:  $\bar{R}_{.,.} = 1.91$ .

Merk at

$$\bar{R}_{1kv} = \sum_{t=1995}^{1999} R_{1kv,t}, \quad \bar{R}_{1995} = \sum_{k=1}^4 R_{k,1995}, \quad \text{og} \quad \bar{R}_{.,.} = \sum_{k=1}^4 \sum_{t=1995}^{1999} R_{k,t}$$

Histogrammet er gitt ved figur 8.7.



Figur 8.7: Histogrammet av revisjoner

- Figurer

Rådata, trend og sesongjusterte tall er plottet i figur 8.8.

Sesongkomponenten  $S_t$  og irregulærkomponenten  $I_t$  er plottet i figur 8.9 og 8.10. Siden variasjoner av irregulærkomponenten er ganske lave (bortsett fra noen tidspunkter), viser dette at avvikene mellom trenden og sesongjusterte tallene er små (bortsett fra noen tidspunkter).

## 8.4 Simuleringsstudie

For å teste 13 års periodelengden for sesongjustering for de utførte investeringene simulerer vi en ny tidsserie ved å bruke observasjonene  $O_t$  fra 1988 til 1999 og legge til framskrivingene  $\hat{O}_{2000.1}$ ,  $\hat{O}_{2000.2}$ ,  $\hat{O}_{2000.3}$  og  $\hat{O}_{2000.4}$  som er beregnet fra tidsserien  $W_{n,t} = \{O_{87.1}, \dots, O_{99.4}\}$  ved ARIMA (0 1 1)(0 1 1). Dermed inneholder den nye tidsserien observasjonene  $O_{1988.1}, \dots, O_{1999.4}$ ,  $\hat{O}_{2000.1}, \dots, \hat{O}_{2000.4}$ . Lengden er 13 år.

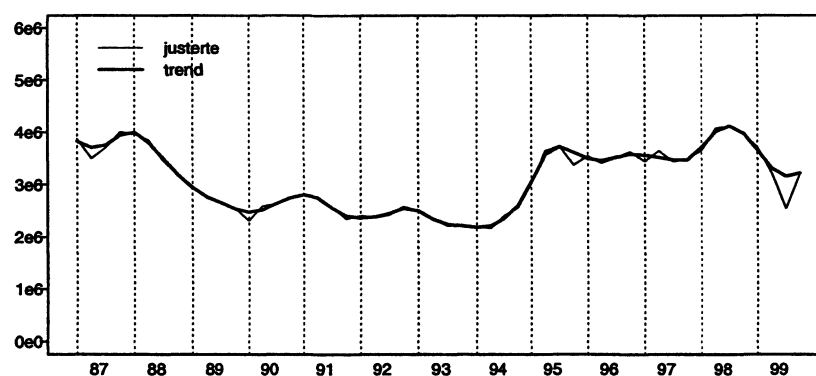
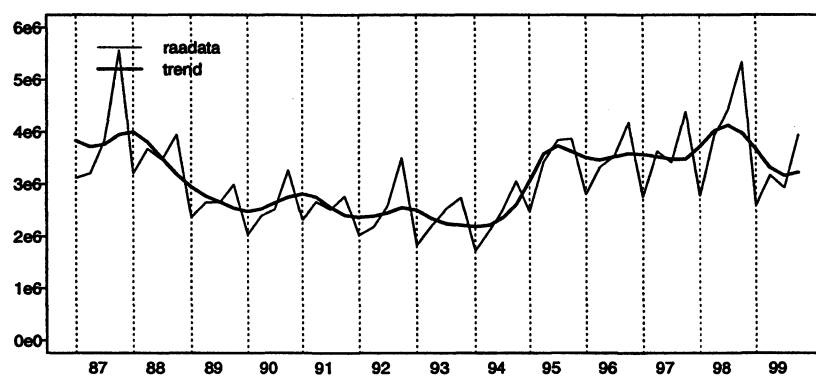
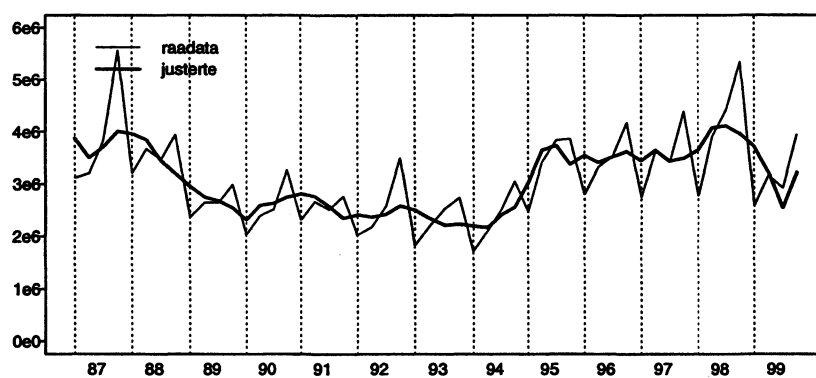
$W_{n,t}$ :  $O_{87.1}$  —————  $O_{99.4}$   
 Sim. :  $O_{88.1}$  —————  $O_{99.4}, \hat{O}_{2000.1}, \hat{O}_{2000.2}, \hat{O}_{2000.3}, \hat{O}_{2000.4}$

Ved å angi på forhånd additive outlier AO1999.3 og estimere kalendereffekter fra dataene, får vi

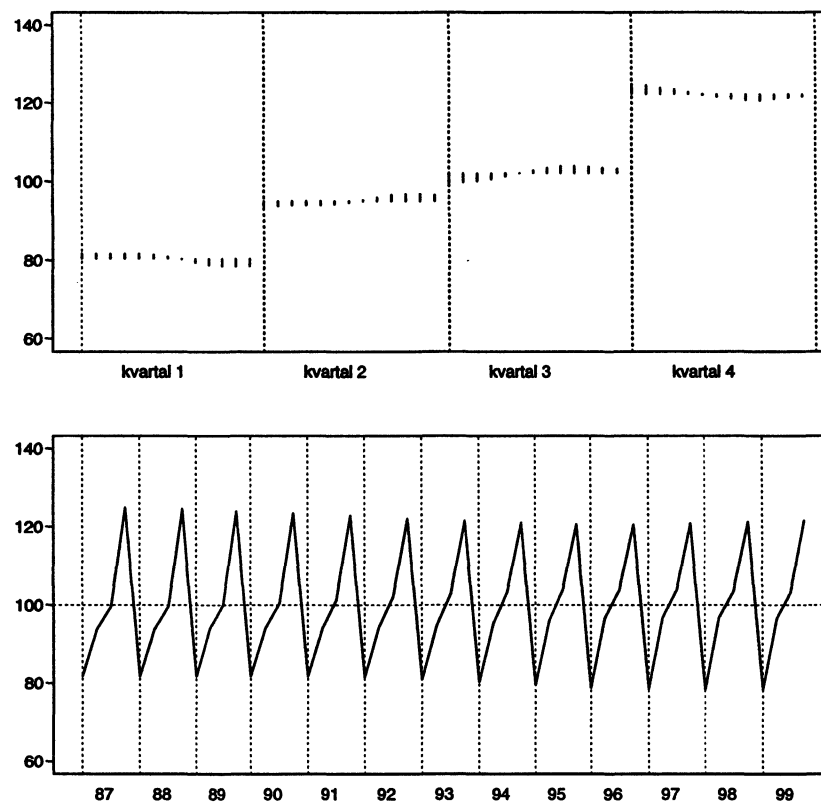
Tabell 8.6: *Sesongjustering for simulerte data*

| Tids.     | ARIMA          | Effekt |       | $Q$   | Sliding. i % |          | Inter. |
|-----------|----------------|--------|-------|-------|--------------|----------|--------|
|           |                | Uke.   | Påske |       | $S(\%)$      | $MM(\%)$ |        |
| 1988-2000 | (0 1 1)(0 1 1) | ja     | ja    | 0.159 | 0.0          | 2.9      | AO99.3 |

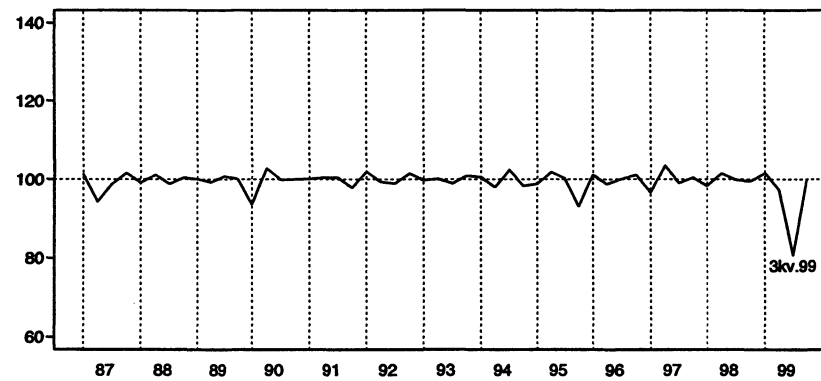
Vi får ganske lave verdiene av  $Q$ ,  $S(\%)$  og  $MM(\%)$  som viser en god sesongjustering. Sammenlignet med tabell 8.4 ser vi at ARIMA modellen ikke endrer seg. Kalenderefektene er fortsatt signifikante på 5% nivå og intervenjonen har forekommet i samme tidspunkt.



Figur 8.8: Rådata, trend og sesongjusterte tall



Figur 8.9: *Sesongkomponent*



Figur 8.10: *Irregulærkomponent*

## Kapittel 9

# Eksempler på estimering av kalendereffekter og intervensjoner ved default verdier i X-12-ARIMA

Dinh Quang Pham

### 9.1 Innledning

En forbedring av X-12-ARIMA er regARIMA modell (regression models with ARIMA errors) slik at en kan prekorrigere rådataene for kalendereffekter og intervensjoner før man utfører en sesongjustering. Formålet er å få bedre kvalitet på sesongjusterte tall. X-12-ARIMA kjøres ved hjelp av en spc file (specification file) som inneholder kommandoer med tilsvarende parametre. For eksempel,

```
series{title="Utførte investeringer"  
        period=12  
        start=1980.1}  
transform{function=auto}  
regression{aictest=(td easter)}  
outlier{}  
osv ...
```

Disse kommandoene beskriver en månedstidsserie (period=12) som starter fra januar 1980 (start=1980.1). Før å sesongjustere, utfører X-12-ARIMA en test om multiplikativ eller additiv modell for dataene (transform{function=auto}). Deretter blir kalendereffekter testet ved å bruke Akaike's (1973, 1974) information criteria, AIC (aictest=(td easter)). Ekstremverdier estimeres med kritisk verdi som velges selv av programmet. Denne verdien kaller vi default verdi (outlier{}). X-12-ARIMA er et fleksibelt program, siden man kan angi verdiene av parametre som vi bestemmer på forhånd, isteden for å bruke default verdier. For eksempel, regression{variables=(td easter[8])}. Dette betyr at kalendereffekter skal estimeres og korrigeres fra rådataene uansett om effektene er signifikante eller ikke.  $d = 8$  er valgt for beregningen. For å ta hensyn til intervensjoner med  $t$ -verdier som er større enn 3.5 skriver vi outlier{critical=3.5}.

Merk at default verdi er verdien som velges selv for parameter av programmet. Denne verdien kan tilpasses for de fleste tidsseriene, men ikke for alle. Når man ikke vil angi en verdi for parameter, blir default verdi brukt i beregningen. For eksempel, kritisk verdien  $C$  for å identifisere en intervensjon i X-12-ARIMA er 3.7974 for en månedstidsserie, hvor lengden  $l$  varierer i intervallet  $73 \leq l \leq 96$ .  $C$  blir 3.8444 for  $97 \leq l \leq 120$ . Vi kan begrense antall intervensjoner ved å sette  $C$  høyere eller lavere enn default verdien. For å korrigere rådata for påske og ukedagseffekter uansett effektene er signifikante eller ikke, kan vi skrive på denne måten `regression{variables=(td easter[8])}`.

Vi har flere tusen av tidsserier, der sesongjusterte tall skal publiseres hver måned. X-12-ARIMA kjøres ofte med default verdier for parametrene i estimeringen for effektene av ukedager, påske og intervensjoner. Vi lar også X-12-ARIMA velge selv filter for sesongkomponenten,  $3 \times 3$ ,  $3 \times 5$  eller  $3 \times 9$ , se Lothian (1984). For de viktige tidsseriene hvor default verdier gir dårlige sesongjusteringer vil vi plukke ut og analysere dem mer nøyaktig. For eksempel, for en ARIMA modell kan vi enten fjerne de parametrene som ikke er signifikant fra modellen og kjøre på nytt, eller velge en annen av de fem innebygde modellene for dataene. Vi kan også forlenge kriterier for å akseptere en ARIMA modell. Transformerings av dataene må vurderes, osv. Det er tilfeller at dataene har regelmessige variasjoner og et klart sesongmønster som tyder på en god tilpasning med ARIMA modellen, men vi får omvendt resultater om ingen ARIMA modell er valgt, påskeeffekt er ikke signifikant selv om historien viser en påvirkning av effekten. Med samme programmet, men påskeeffekten blir angitt på forhånd, får vi ønskede resultater. Dette er en svakhet av X-12-ARIMA.

I dette kapitlet vil vi presenterer problemer når vi estimerer kalendereffekter og intervensjoner i samme kjøring ved to eksempler: (a) Faktisk utførte for de totale investeringene totalt for industrien (b) Import av tradisjonelle varer.

## 9.2 Faktisk utførte for de totale investeringene totalt for industrien

I kapittel 8 har vi vist at det ikke er en fordel å bruke en lang tidsserie for sesongjustering for de utførte investeringene. 13 års periodelengden gir bedre sesongjusterte tall. I simuleringsstudiet, fra de observasjonene  $\{O_{1987.1}, O_{1987.2}, \dots, O_{1999.4}\}$  lager vi en ny tidsserie med samme lengden ved å fjerne de fire første observasjonene  $O_{1987.1}$ ,  $O_{1987.2}$ ,  $O_{1987.3}$  og  $O_{1987.4}$  og legge til framskrivingene  $\hat{O}_{2000.1}$ ,  $\hat{O}_{2000.2}$ ,  $\hat{O}_{2000.3}$  og  $\hat{O}_{2000.4}$ .

Sim.  $O_{87.1}$  —————  $O_{99.4}$   
 $O_{88.1}$  —————  $O_{99.4}, \hat{O}_{2000.1}, \hat{O}_{2000.2}, \hat{O}_{2000.3}, \hat{O}_{2000.4}$

Tidsserien  $\{O_{1988.1}, O_{1988.2}, \dots, O_{2000.4}\}$  skal sesongjusteres på forskjellige måter der det tas hensyn til prekorrigering for kalendereffekter og intervensjoner.

### 9.2.1 Uten å angi kalendereffekter og intervensjoner på forhånd

Kommandoene som utfører regresjonsanalyse ser slik ut

```
regression{aictest=(td easter)}
outlier{}
```

Vi får

- (a) Ingen av de fem ARIMA modellene gir god tilpasning til dataene.
- (b) Påskeeffekten er ikke signifikant på 5% nivå.
- (c) AO1999.3 er ikke en intervensjon lenger.
- (d)  $Q = 0.142$ ,  $S(\%) = 3.1$  og  $MM(\%) = 25.8$ .

Vi ser at  $S(\%)$  og  $MM(\%)$  er høyere sammenlignet med de samme størrelsene for tidsserien fra 1987 til 1999 i tabell 8.4.

### 9.2.2 Angi påskeeffekten på forhånd

I X-12-ARIMA, når kalendereffekter korrigeres ved aictest samtidig med intervensjoner, utfører programmet aictest først, deretter estimeringen for intervensjoner. I denne kjøringen er påskeeffekten ikke signifikant som medfører ingen intervensjon er funnet.

Tabell 8.4 viser at dataene påvirkes av påskeeffekten. Dermed i neste kjøring vil vi angi effekten på forhånd ved kommandoen

```
regression{variables=(easter[15]) aictest=(td)}
outlier{}
```

Vi får

- (a) ARIMA (0 1 1)(0 1 1) er valgt.
- (b) Påskeeffekten er signifikant på 5% nivå.
- (c) AO1999.3 er beregnet som en intervensjon.
- (d)  $Q = 0.159$ ,  $S(\%) = 0.0$  og  $MM(\%) = 2.9$ .

Grenseverdiene for  $Q$ ,  $S(\%)$  og  $MM(\%)$  er 1, 15 og 35, henholdsvis.

### 9.2.3 Ved å utføre en estimering for intervensjon før påskeeffekt

I denne kjøringen vil vi korrigere rådataene for AO1999.3 før å teste om påskeeffekt ved å skrive

```
regression{variables=(ao1999.3) aictest=(easter td)}
outlier{}
```

Vi får

- (a) ARIMA (0 1 1)(0 1 1) er valgt for dataene.
- (b) Kalendereffekter er signifikante på 5% nivå.
- (c) en signifikans for AO1999.3.
- (d)  $Q = 0.127$ ,  $S(\%) = 0.0$  og  $MM(\%) = 8.6$ .

### 9.2.4 Ved å angi ukedagseffekter på forhånd

Vi angir ukedagseffekter på forhånd til programmet og estimere med default verdier for påskeeffekt og intervensjoner deretter.

```
regression{variables=(td) aictest=(td)}
outlier{}
```

Vi får

- (a) Ingen ARIMA modell for dataene.
- (b) Påskeeffekten er ikke signifikant.
- (c) Ingen intervensjoner.
- (d)  $O = 0.142$ ,  $S(\%) = 3.1$  og  $MM(\%) = 25.8$ .

Ved å sammenligne resultater fra 9.2.1, 9.2.2, 9.2.3 og 9.2.4 ser vi at sesongjusteringen er helt annerledes når vi angir påskeeffekt eller intervensjon på forhånd. Resultatene for sesongjusteringen blir også bedre.

### 9.3 Import av tradisjonelle varer IMP.VITSOR.VR.P

Tidsserien er import av varer eksklusive råolje, skip og mobile oljeplattformer. Dataene er observert fra januar 1980 til mars 2000. Verdien i millioner kroner. Vi sesongjusterer tidsserien på følgende måter:

#### 9.3.1 Uten å angi på forhånd kalendereffekter eller intervensjoner i programmet

I denne kjøringen skal programmet utføre en AIC test for kalendereffekter før estimering for intervensjoner. Prosessen skrives ved

```
regression{aictest=(easter td)}  
outlier{}
```

Vi får

- (a) Ingen av de fem innebygde ARIMA modellene blir valgt for dataene.
- (b) Kalendereffekter er signifikant på 5% nivå.
- (c)  $Q = 0.496$ ,  $S(\%) = 5.4$  og  $MM(\%) = 10.0$ .

#### 9.3.2 Ved å angi påskeeffekt på forhånd

Programmet skal ikke utføre en aictest som i 9.3.1. Dataene blir korrigert for påskeeffekten uansett effekten er signifikant eller ikke. Vi skriver i programmet slik

```
regression{variables=(easter[d]) aictest=(td)}  
outlier{}
```

hvor  $d = 1, 8$  og  $15$ . Resultater fra sesongjustering er listet i nedenfor tabell

|                  | $d = 1$                          | $d = 8$                         | $d = 15$                         |
|------------------|----------------------------------|---------------------------------|----------------------------------|
| ARIMA            | (2 1 2)(0 1 1)                   | (0 1 2)(0 1 1)                  | (0 1 2)(0 1 1)                   |
| AIC              | 3681                             | 3729                            | 3729                             |
| Kalendereffekter | sign. 5%                         | sign. 5%                        | sign. 5%                         |
| Intervensjoner   | AO1990.mai<br>AO1994.jul         | AO1994.jul                      | AO1991.apr<br>AO1994.jul         |
| Kvalitetsmål     | $Q = 0.467$                      | $Q = 0.638$                     | $Q = 0.608$                      |
| Sliding spans    | $S(\%) = 0.9$<br>$MM(\%) = 11.8$ | $S(\%) = 2.0$<br>$MM(\%) = 3.4$ | $S(\%) = 5.4$<br>$MM(\%) = 10.0$ |

### 9.3.3 Ved å angi intervensjon AO1994.jul på forhånd

I programmet angir vi på forhånd at AO1994.jul er en additiv outlier. Vi får fra X-12-ARIMA

(a) ARIMA (2 1 2)(0 1 1) er valgt med AIC=3681.

(b) Kalendereffekter er signifikante på 5% nivå og  $\hat{d} = 1$  for påskeeffekten.

(c) AO1990.mai og AO1994.jul er to signifikante intervensjoner av dataene.

(d)  $Q = 0.618$ ,  $S(\%) = 0.9$  og  $MM(\%) = 12.7$ .

Sammenlignet med resultater i avsnitt 9.3.2 for  $\hat{d} = 1$  er det omtrent samme.

## 9.4 Konklusjon

Vi har sett fra to eksempler at svakheten i X-12-ARIMA er at prosedyren for prekorrigering kalendereffekter og intervensjoner i samme kjøring med default verdier i X-12-ARIMA fungerer ikke godt i noen tilfeller. I slike tilfeller anbefaler vi først å finne påskeeffekten og deretter undersøke om det er intervensjon.

## Kapittel 10

# Korrigerering for streiken i 2000 og Kristi himmelfartsdag i detaljvolum indeksen

Dinh Quang Pham

### 10.1 Problem

Ved bruk av det vanlige opplegget for sesongjustering får vi et for høyt sesongjustert tall for mai og for lavt for juni i år 2000 i detaljvolumindeksen. Det er grunn til å anta at dette bl.a. kan skyldes at Kristi himmelfartsdag i år var i juni. Sesongjusterte tall basert på det første opplegget for år 2000 er gitt ved

Tabell 10.1: *Rådata og sesongjusterte tall for detaljvolumindeksen*

|                            | jan.   | feb.   | mar.   | apr.   | mai    | juni.  |
|----------------------------|--------|--------|--------|--------|--------|--------|
| Rådata $O_t$ , i 2000      | 104.74 | 101.92 | 110.25 | 105.90 | 119.90 | 121.10 |
| Endring: $O_t - O_{t-1}$   | -      | -2.82  | 8.33   | -4.35  | 14.00  | 1.20   |
| Se.justerte $J_t$ , i 2000 | 117.46 | 118.20 | 118.41 | 118.76 | 123.17 | 118.56 |
| Endring: $J_t - J_{t-1}$   | -      | 0.74   | 0.21   | 0.35   | 4.41   | -4.61  |
| Gj. snitt for hele serien  | 96.17  | 95.73  | 95.78  | 96.23  | 96.33  | 96.38  |

For hele serien er: min=78.45, max=123.17, gj. snitt=95.83, std. avvik=8.71.

Vi ser at mai i år 2000 har størst verdi for den sesongjusterte serien. Endringene fra april til mai og fra mai til juni er ganske store med 4.41% og -4.61%, henholdsvis.

Et forslag for å justere ned sesongjusterte tall i mai og samtidig øke junitall er å erstatte deres verdier med gjennomsnittet. De nye sesongjusterte tallene for mai og juni 2000 blir

$$\begin{aligned}\hat{J}_{mai.2000} = \hat{J}_{jun.2000} &= \frac{J_{mai.2000} + J_{jun.2000}}{2} \\ &= \frac{123.17 + 118.56}{2} = 120.87\end{aligned}$$

Tabell 10.2 viser sesongjusterte tallene for år 2000 etter forslaget. Vi ser at endringen fra april til mai blir lavere, med 2.11% mot 4.41% i samme tidspunkt før korrigeringen.

Denne metoden bygger imidlertid ikke på en prekorrigering for Kristi himmelfartsdag eller andre faktorer før en utfører en sesongjustering. Andre faktorer kan f. eks, være streiken i månedsskiftet april-mai. Når en legger til nye observasjoner og sesongjusterer den forlengede tidsserien, kan høy verdi i mai påvirke sesongjusterte tall for flere perioder som kommer seinere. Ved å ta gjennomsnittet av de sesongjusterte tallene i mai og juni, løser vi kun problemet for mai og juni 2000, men ikke for de andre månedene.

Tabell 10.2: *Sesongjusterte tall etter forslaget i år 2000*

|                     | jan.   | feb.   | mar.   | apr.   | mai    | juni.  |
|---------------------|--------|--------|--------|--------|--------|--------|
| Sesongjusterte tall | 117.46 | 118.20 | 118.41 | 118.76 | 120.87 | 120.87 |
| Endring             |        | 0.74   | 0.21   | 0.35   | 2.11   | 0.00   |

I neste avsnitt vil vi derfor presentere en metode for prekorrigering for april, mai og juni i 2000 før vi sesongjusterer.

## 10.2 Beskrivelse av metoden

Dataene er observert fra januar 1979 til juni 2000. I alt er det 258 observasjoner. X-12-ARIMA skal brukes for sesongjustering. Siden påsken faller veldig seint i 2000, er det ingen påskekorrigerings. I den nye metoden lager vi to dummyvariable, en for streiken som var i månedsskift april-mai, og en for Kristi himmelfartsdag som var 1. juni. I X-12-ARIMA terminologien kalles dette for "user-supplied variables". De to dummyvariablene ligger i en fil som vi kaller "tillegg.dat". Filen er beskrevet i tabell 10.3.

Tabell 10.3: Dummyvariable for streiken og Kristihimmelfartsdag

|           | Streik | Kristi. |
|-----------|--------|---------|
| jan.1979  | 0      | 0       |
| feb.1979  | 0      | 0       |
| ...       | ...    | ...     |
| des.1999  | 0      | 0       |
| jan.2000  | 0      | 0       |
| feb.2000  | 0      | 0       |
| mar.2000  | 0      | 0       |
| apr.2000  | 1      | 0       |
| mai.2000  | -1     | 1       |
| juni.2000 | 0      | -1      |

Vi ser at summen av de to dummyvariablene i april, mai og juni er lik null. Dermed skal prekorrigeringen for effektene av streiken og Kristi himmelfartsdag ikke påvirke til de andre månedene. Etter å ha kjørt X-12-ARIMA med følgende kommandoer

```

regression{variables=(td)
            user=(streik kristi)
            usertype=holiday
            file="tillegg.dat"}

```

får vi resultater som er vist i tabell 10.4. Vi ser at før sesongjustering ble tallene i

Tabell 10.4: *Rådata, korrigerte dataene, og sesongjusterte tall ved user-supplied variable for streiken og Kristihimmelfartsdag i 2000*

|                            | jan.   | feb.   | mar.   | apr.   | mai    | jun.   |
|----------------------------|--------|--------|--------|--------|--------|--------|
| Rådata $O_t$ , i 2000      | 104.74 | 101.92 | 110.25 | 105.90 | 119.90 | 121.10 |
| Endring: $O_t - O_{t-1}$   | -      | -2.82  | 8.33   | -4.35  | 14.00  | 1.20   |
| Korrigerte dataene $O_t^K$ | 104.74 | 101.92 | 110.25 | 107.01 | 115.54 | 124.37 |
| Endring                    | -      | -2.82  | 8.33   | -3.24  | 8.53   | 8.83   |
| Sesongjusterte tall        | 117.51 | 118.14 | 118.44 | 119.57 | 120.48 | 120.04 |
| Endring                    | -      | 0.63   | 0.33   | 1.13   | 0.91   | -0.44  |

april og juni korrigert opp med 1.11% ( $=107.01-105.90$ ) og 3.27% ( $=124.37-121.10$ ), henholdsvis. Omvendt er mai tallet justert ned med 4.36% ( $=115.54-119.90$ ). Vi får jevne sesongjusterte tall fra april til juni 2000. Endringen fra april til mai er redusert til 0.91% mot 2.11% i det enkle forslaget om å bruke gjennomsnittet. Det er en svak nedgang fra mai til juni med 0.44%.

Kristi himmelfartsdag i år er 1. juni mot normalt i mai. Effekten av denne helligdagen medførte at vi fikk en handledag mer i 2000 enn normalt. Dermed forventet vi at volumet lå omlag  $(1/25) \times 100\% = 4\%$  for høyt. Altså burde mai tallet reduseres med 4% før vi sesongjusterer. Modellen vår beregninger redusjonen til 4.36%. Avviket er bare 0.36%. Merk at 25 er antall ukedager i mai i 2000.

### Resultater fra analyse.

- ARIMA (0 1 1)(0 1 1) med *log* transformasjon er valgt for dataene.
- Det er signifikante for effektene av ukedager, user-supplied variables og intervensjoner. Vi får

|                         |                   |      |              |
|-------------------------|-------------------|------|--------------|
| Ukedagseffekter         | $\chi^2 = 174.69$ | df=6 | p-verdi=0.00 |
| User-supplied variables | $\chi^2 = 6.58$   | df=2 | p-verdi=0.04 |

og

|                | Est. verdi | Std    | t-verdi |
|----------------|------------|--------|---------|
| User-defined   |            |        |         |
| streik         | -0.0104    | 0.0154 | -0.67   |
| Kristi.        | 0.0266     | 0.0155 | 1.72    |
| Intervensjoner |            |        |         |
| LS1987.Feb     | -0.0759    | 0.0144 | -5.29   |
| AO1988.Dec     | -0.0865    | 0.0170 | -5.08   |

Vi ser at det ikke er en signifikant effekt for streiken, men det kan være signifikant for Kristi himmelfartsdag.

### Resultater fra analyse med en dummyvariabel for Kristi himmelfartsdag

Vi får, for ukedagseffekter er  $\chi^2 = 178.48$ , med 6 frihetsgrader og  $p$ -verdi=0.00, en høy signifikans, og

|                | Est. verdi | Std    | t-verdi |                   |
|----------------|------------|--------|---------|-------------------|
| User-defined   |            |        |         |                   |
| Kristi.        | 0.0322     | 0.0130 | 2.47    | signifikant på 5% |
| Intervensjoner |            |        |         |                   |
| LS1987.Feb     | -0.0759    | 0.0144 | -5.29   | signifikant på 5% |
| AO1988.Dec     | -0.0865    | 0.0170 | -5.08   | signifikant på 5% |

Uten å ta hensyn til streiken blir effekten av Kristi himmelfartsdag signifikant på 5% nivået. Korrigerte data og de sesongjusterte tallene er vist i tabell 10.5.

Tabell 10.5: *Rådata, korrigerte dataene, og sesongjusterte tall ved user-supplied variable for Kristihimmelfartsdag i 2000*

|                            | jan.   | feb.   | mar.   | apr.   | mai    | jun.   |
|----------------------------|--------|--------|--------|--------|--------|--------|
| Rådata $O_t$ , i 2000      | 104.74 | 101.92 | 110.25 | 105.90 | 119.90 | 121.10 |
| Endring: $O_t - O_{t-1}$   | -      | -2.82  | 8.33   | -4.35  | 14.00  | 1.20   |
| Korrigerte dataene $O_t^K$ | 104.74 | 101.92 | 110.25 | 105.90 | 116.09 | 125.06 |
| Endring                    | -      | -2.82  | 8.33   | -4.35  | 10.19  | 8.97   |
| Sesongjusterte tall        | 117.52 | 118.23 | 118.43 | 118.85 | 120.83 | 120.46 |
| Endring                    | -      | 0.71   | 0.20   | 0.42   | 1.98   | -0.37  |

Før vi sesongjusterer, ble verdien i mai redusert med  $119.90 - 116.09 = 3.81\%$ . Sammenlignet med tallene i tabell 10.4 er endringen av sesongjusterte tall fra april til mai i tabell 10.5 høyere.

## 10.3 Konklusjon

Det viser seg at bare effekten av Kristi himmelfartsdag er signifikant. Vi kan korrigere dataene i mai og juni for denne effekten ved

(a) å redusere verdien i mai med  $(1/25) \times 100\% = 4\%$  og øke verdien for juni med samme størrelsen.

(b) å bruke user-supplied variables i X-12-ARIMA. Dvs vi lager en fil for dummyvariable og kjører filen sammen med de observerte dataene. Tallet i mai er redusert ca 4% i begge tilfeller: med og uten effekten av streiken.

Vi får omtrent de samme korrigerte og sesongjusterte tallene for april, mai og juni 2000 ved to metoder. Men (a) gjøres manuelt og det blir tungvint og mindre presis hvis det er mange tidspunkter som må korrigeres.

Når mai er siste observasjon så kan vi kun gjøre korreksjonen manuelt. Men når vi har både mai og juni observasjoner bruker vi modellen.

Det er veldig enkelt å bruke user-supplied variables i X-12-ARIMA. Vi får mer informasjon om dummyvariable fra regresjonsanalyse enn å korrigere tallene ved (a).

## Kapittel 11

# Testen for valg av modell ved sesongjustering og sliding spans i X-12-ARIMA

Dinh Quang Pham

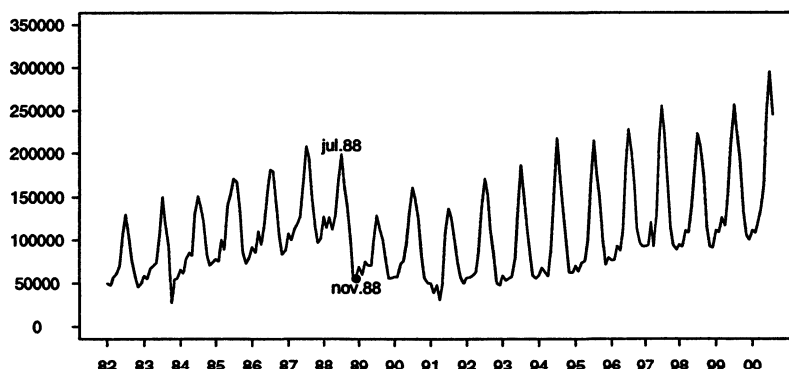
### 11.1 Innledning

Vi spalter opp en tidsserie  $\{O_t\}$  i X-12-ARIMA i trend, sesongjusterte tall ved

$$O_t = T_t + S_t + I_t \quad \text{som vi kaller additiv modell} \quad (11.1)$$

$$O_t = T_t \times S_t \times I_t \quad \text{som vi kaller multiplikativ modell} \quad (11.2)$$

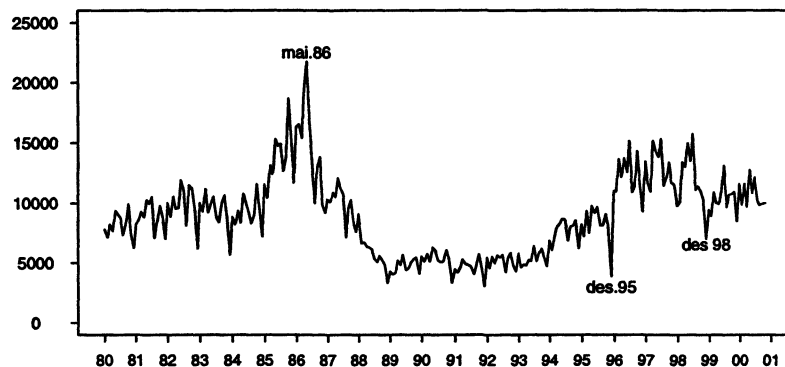
Vi kan lage en modell som er en blanding av (11.1) og (11.2). I praksis velger vi en multiplikativ dekomponering for en tidsserie når sesongvariasjoner er proporsjonale med størrelser av observasjonene. Figur 11.1 viser et eksempel. Det er antall passasjerer som er summert over alle lufthavner på utenlands charter i Norge. Tidsserien er observert fra januar 1982 til august 2000. Vi ser at tidsserien har en regelmessig variasjon, men det blir større i de siste observasjonene.



Figur 11.1: *Rådata av passasjerer på charter utlandet*

For tidsseriene med sesongvariasjoner som endrer seg sakte med tid, vil vi bruke en additiv modell for sesongjustering. Figur 11.2 viser antall registrerte nye personbiler

i Norge fra januar 1980 til november 2000. Testen viser en signifikans på 5% nivå for sesongvariasjoner, men sesongmønsteret i dataene ser ikke tydelig ut. Det blir derfor vanskelig å velge en modell for sesongjustering ved å basere på plottet.



Figur 11.2: Rådataene for nye personbiler i Norge

I X-12-ARIMA er det en test som brukes for å finne en modell for sesongdekomponering. Testen er basert på  $AIC$ -verdiene (Akaike's information criterion) av  $ARIMA(0\ 1\ 1)(0\ 1\ 1)$  tilpasset mot rådataene. En additiv modell velges dersom

$$AICC_{nolog} < AICC_{log} + \Delta_{AICC} \quad (11.3)$$

hvor  $AICC_{nolog}$  er  $F$ -corrected  $AIC$  verdi av  $ARIMA(0\ 1\ 1)(0\ 1\ 1)$  for rådataene uten transformasjon, mens  $AICC_{log}$  er den samme størrelsen med  $log$  transformasjonen.  $\Delta_{AICC} = -2$  for default verdien.

Vi bruker sliding spansanalysen (Findley et al. 1990) for å vurdere hvor god en sesongjustering er. Dersom  $S(\%) < 15$  og  $MM(\%) < 40$  er sesongjustering god. Vi får en mindre kvalitet når  $15 < S(\%) \leq 25$  og  $MM(\%) < 40$ . Ellers er sesongjusteringen dårlig.  $S(\%)$  beskrives som prosentandelen måneder hvor sesongkomponenten er definert som upåtelig.  $MM(\%)$  og  $YY(\%)$  er de målene for sesongjusterte tall.

Vi har kjørt mange tidsserier. Det er tidsserier der den valgte modellen gir mer ustabile sesongjusterte tall enn den modellen som blir forkastet.

I neste avsnitt presenterer vi et eksempel for å vise dette.

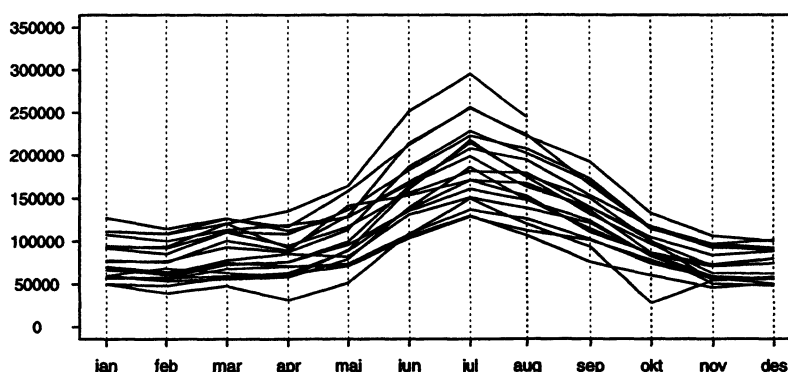
## 11.2 Eksempel 1. Antall passasjerer på charter utenlands ruter

Rådataene er plottet i figur 11.1. Dataene har en regelmessig variasjon i perioden fra januar 1982 fram til oktober 1988. Juli er måneden med de fleste passasjerene. Tabell 11.1 viser rådataene for juli, august, september, oktober og november 1988 samt deres gjennomsnitt i to periodene [1982-1987] og [1989-1999]. Vi ser at tallet i november 1988 er ganske lavt. Endringen fra oktober til november 1988 er stor med -46648 passasjerer, mens gjennomsnittlige endringer fra oktober til november i [1982-1987] er -9154 og [1989-1999] er -25112. Kurven for hele tidsserien ser ut som kan deles i to. Den første inneholder observasjonene fram til oktober 1988 og den andre

starter fra november 1988 med lavere nivå. November 1988 kan være en intervensjon med typen nivåskift. Figur 11.3 viser rådataene som er plottet for 12 måneder. Juni, juli, august og september har store sesongkomponenter. Kurvene viser også betydelige endringer fra år til år i samme måned.

Tabell 11.1: *Rådataene av oktober og november*

| måned          | 1988   | Gjennom- | Gjennom- |
|----------------|--------|----------|----------|
|                |        | snitt i  | snitt i  |
|                |        | 1982-87  | 1989-99  |
| juli           | 198903 | 165338   | 198324   |
| august         | 162789 | 150932   | 172198   |
| september      | 139644 | 119720   | 139245   |
| oktober        | 100491 | 79744    | 97984    |
| november       | 53843  | 70590    | 72872    |
| diff.(aug,jul) | -36114 | -14406   | -26126   |
| diff.(sep,aug) | -23145 | -31212   | -32953   |
| diff.(okt,sep) | -39153 | -39976   | -41261   |
| diff.(nov,okt) | -46648 | -9154    | -25112   |



Figur 11.3: *Rådata gjennom årene, over 12 måneder*

#### - Programmet selv finner en modell for sesongjustering

En additiv modell er valgt, siden  $AICC_{nolog} = 4539$  og  $AICC_{log} = 4623$ . Vi får  $S(\%) = 35.6$ ,  $MM(\%) = 35.9$  og  $YY(\%) = 12.0$ . Disse høye tallene viser en dårlig sesongjustering. Oktober 1983 og mai 1985 er de to intervensjonene med typen AO (additive outlier) i rådataene.

#### - Ved å velge den multiplikative modellen for sesongjustering

Vi får lavere verdier for  $S(\%)$ ,  $MM(\%)$  og  $YY(\%)$ , med  $S(\%) = 21.2$ ,  $MM(\%) = 19.4$  og  $YY(\%) = 4.3$ . Oktober 1983, mai 1985, april 1991 og februar 1994 er intervensjoner med typen AO, mens november 1988 og juni 1991 er to nivåskifter.  $M4 = 1.420$ . Denne høye verdien tyder på at den irregulære komponenten ikke er en hvit støy prosess. For å unngå dette vil vi begrense antall intervensjoner med typen AO i

korrigeringen. Vi tar de tidspunktene som har de høyeste  $t$  verdier som kandidater for intervensjon. Det er oktober 1983 og mai 1985, der  $t$  verdiene er henholdsvis -23.49 og 8.94 som kandidater for intervensjon. Et nivåskift er definert ved  $LS_t^{t_0} = -1$  for  $t < t_0$  og  $LS_t^{t_0} = 0$  for  $t \geq t_0$ , dvs rådataene i tidspunktene etter  $t_0$  blir korrigert. Siden november 1988 og juni 1991 er nivåskifter blir tidspunktene etter november 1988 korrigert to ganger. De to tidspunktene ligger ikke langt fra hverandre. Dermed vil vi beholde LS1991.jun i regARIMA modellen.  $M4$  blir redusert til 0.988. Den irregulære komponenten er nå en hvit støy prosess. Vi får lavere verdier for  $S(\%)$ ,  $MM(\%)$  og  $YY(\%)$  med 14.4, 15.1 og 1.1 henholdsvis. Sesongjusteringen er god.

Vi presenterer i neste avsnitt resultatene i mer detalj.

## 1. ARIMA modell

ARIMA (0 1 1)(0 1 1) er valgt for rådataene og rådataene med *log* transformasjon. Parametrene i ARIMA modellen er estimert med hensyn til effektene av ukedager, påske og intervensjoner. De er listet ut i tabell 11.2. Vi ser at  $\hat{\theta}$  ikke er signifikant på 5% nivå.

Tabell 11.2: De estimerte parametrene av ARIMA(0 1 1)(0 1 1)

| Modell               | Par.           | Estimert |        |                       |
|----------------------|----------------|----------|--------|-----------------------|
|                      |                | Verdi    | Std    |                       |
| Additiv modell       | $\hat{\theta}$ | -0.0905  | 0.0684 | ikke sign. på 5% nivå |
|                      | $\hat{\Theta}$ | 0.4724   | 0.0630 | sign. på 5% nivå      |
| Multiplikativ modell | $\hat{\theta}$ | 0.0378   | 0.0684 | ikke sign. på 5% nivå |
|                      | $\hat{\Theta}$ | 0.4984   | 0.0605 | sign. på 5% nivå      |

Gjennomsnittlig absolutt prosent feil for framskrivninger er gitt i tabell 11.3. Tallene viser at etter å ha transformert med *log* tilpasser dataene bedre med ARIMA modellen i de siste tre årene.

Tabell 11.3: Prosent feil for framskrivninger

|                      | siste år | forrige år | to år tidligere | siste tre år |
|----------------------|----------|------------|-----------------|--------------|
| Additive modell      | 6.13     | 10.65      | 18.35           | 11.71        |
| Multiplikativ modell | 5.15     | 12.22      | 6.34            | 7.90         |

## 2. Kalendereffekter

Effektene er gitt i tabell 11.4. Vi ser at lørdag og søndag har de fleste passasjerene i uken, mens tirsdag (for additiv modell) og onsdag (for multiplikativ modell) har færrest antall passasjerer.

Merk at vi får samme fortegn for de estimerte verdiene for ukedagene. Påskeeffekten er signifikant på 5% nivå.

Tabell 11.4: Ukedageffekter

|         | Additiv modell |        |         | Multiplikativ modell |        |         |
|---------|----------------|--------|---------|----------------------|--------|---------|
|         | effekt         | std    | t-verdi | effekt               | std    | t-verdi |
| mandag  | -668.74        | 675.00 | -0.99   | -0.0070              | 0.0055 | -1.26   |
| tirsdag | -1994.64       | 701.39 | -2.84   | -0.0098              | 0.0058 | -1.68   |
| onsdag  | -980.95        | 674.38 | -1.45   | -0.0133              | 0.0055 | -2.42   |
| torsdag | -718.96        | 674.96 | -1.07   | -0.0075              | 0.0055 | -1.36   |
| fredag  | 421.40         | 716.33 | 0.59    | 0.0085               | 0.0059 | 1.43    |
| lørdag  | 1640.42        | 677.73 | 2.42    | 0.0140               | 0.0055 | 2.53    |
| søndag  | 2301.48        | 697.09 | 3.30    | 0.0151               | 0.0057 | 2.65    |

|       | Additiv modell            | Multiplikativ modell      |
|-------|---------------------------|---------------------------|
| Påske | $\hat{d} = 15$ $t = 4.70$ | $\hat{d} = 15$ $t = 6.17$ |

### 3. Intervensjoner

Tidspunktene, typene og størrelsene av intervensjoner er vist i tabell 11.5, hvor  $f_t$  er effekten av intervensjon i tidspunkt  $t$ .

Tabell 11.5: Intervensjoner for passasjerer på charter utlandet

|                                     | år   | måned   | type | $\hat{f}_t$ | std    | $t$    |
|-------------------------------------|------|---------|------|-------------|--------|--------|
| Additiv modell                      | 1983 | oktober | AO   | -43592      | 4848   | -8.99  |
|                                     | 1985 | mai     | AO   | 37484       | 4721   | 7.94   |
| Multiplikativ modell<br>(log skala) | 1983 | oktober | AO   | -0.9300     | 0.0510 | -18.22 |
|                                     | 1985 | mai     | AO   | 0.3426      | 0.0496 | 6.89   |
|                                     | 1991 | juni    | LS   | 0.2951      | 0.0681 | 4.33   |

For additiv modell.

-  $t$ =oktober 1983. Vi har  $O_t = 27740$  og  $f_t = -43592$ . La  $O_t^k$  være den korrigerte verdien i  $t$ . Vi får  $O_t^k = O_t - f_t = 27740 + 43592 = 71332$ . Tallet viser at hvis det er ingen hendelse i oktober 1983 blir antall passasjerer høyere enn den verdien vi har observert. Dermed er antall passasjerer redusert med  $43592/71332 \times 100\% = 61.11\%$ .

-  $t$ =mai 1985. Vi har  $O_t = 140849$ ,  $f_t = 37484$ . Dermed er  $O_t^k = O_t - f_t = 140849 - 37484 = 103365$ . Vi ser at hendelsen i mai 1985 har økt antall passasjerer fra 103365 til 140849 eller  $37484/103365 \times 100\% = 36.25\%$ .

For multiplikativ modell.

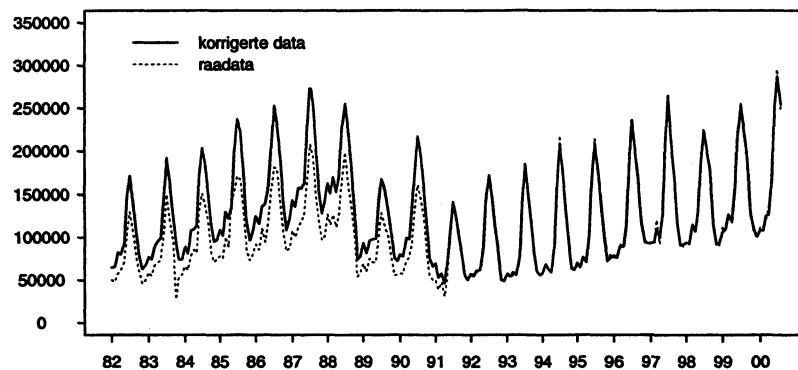
Merk at rådataene blir transformert med  $\log$  i beregningen for intervensjoner.

-  $t$ =oktober 1983. Vi har  $O_t = 27740$ ,  $f_t = -0.93$ . Den korrigerte verdien blir  $O_t^k = \exp(\log(27740) + 0.93) = 70307$ . På grunn av hendelsen i oktober 1983 har antall passasjerer redusert med  $(70307 - 27740)/70307 \times 100\% = 60.54\%$ .

-  $t$ =mai 1985. Vi har  $O_t = 140849$ ,  $f_t = 0.3426$ . Dermed er  $O_t^k = \exp(\log(140849) - 0.3426) = 99991$ . Differansen blir  $140849 - 99991 = 40858$  passasjerer som tilsvarende med  $40858/99991 \times 100\% = 40.86\%$ .

-  $t$ =juni 1991. Siden det er et nivåskift blir observasjonene før juni 1991 korrigert opp (se definisjon av typen LS) med effekten  $f = 0.2951$  i  $\log$  skalaen.

Ved å bruke rådataene  $O_t$  i beregningene blir oktober 1983 og mai 1985 to ekstremverdier i tidsserien. Men da estimeringen ble utført med  $\log(O_t)$  er det flere tidspunkter: oktober 1983, mai 1985 og i intervallet [jan.1982-mai.1991]. Figur 11.4 viser de korrigerte dataene  $O_t^k$  for intervensjoner med multiplikativ modell. Fra januar 1982 til mai 1991 blir  $O_t$  justert opp med en størrelse  $\omega = 0.2951$  (i  $\log$  skala) for nivåskiftet i juni 1991. Dessuten får vi en høyere verdi i oktober 1983 og en lavere verdi i mai 1985. Vi ser at etter korrigeringen er mønsteret for tidspunktene før og etter juni 1991 nesten like. Dermed får vi en bedre tilpasning mellom en ARIMA modell og de korrigerte dataene  $O_t^k$  enn med de opprinnelige dataene  $O_t$  (se tabell 11.3 for prosent feil for framskrivninger).



Figur 11.4: Rådata og de korrigerte dataene for intervensjoner

#### 4. Kvalitetsmål og sliding spans

De 11 kvalitetsmålene er gitt i tabell 11.6.

De to modellene gir omtrent samme verdiene av  $M1$ - $M11$  og  $Q$ . Vi får en lavere verdi for  $M4^{mult}$  enn  $M4^{add}$  som viser en hvit støy prosess for den irregulære komponenten.

Sliding spansanalysen er listet ut i tabell 11.7. Vi får mye lavere verdiene for  $S(\%)$ ,  $MM(\%)$  og  $YY(\%)$  ved den multiplikativ modellen som medfører en bedre sesongjustering.

Tabell 11.6: *De 11 målene*

|                      | <i>M1</i> | <i>M2</i> | <i>M3</i> | <i>M4</i> | <i>M5</i> | <i>M6</i> |
|----------------------|-----------|-----------|-----------|-----------|-----------|-----------|
| additiv modell       | 0.031     | 0.024     | 0.209     | 1.173     | 0.449     | 0.862     |
| multiplikativ modell | 0.029     | 0.026     | 0.113     | 0.988     | 0.403     | 0.826     |

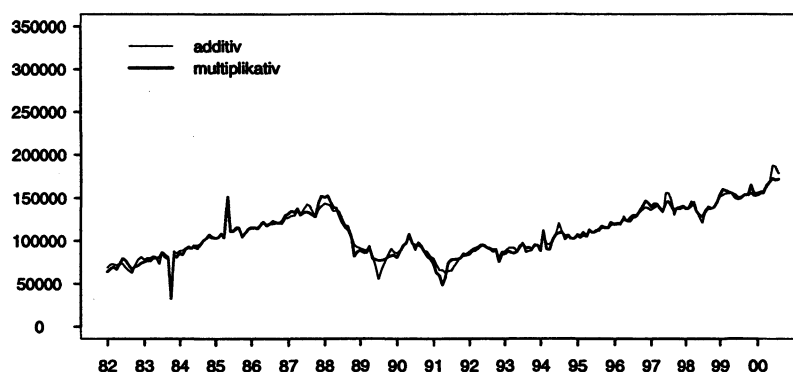
|                      | <i>M7</i> | <i>M8</i> | <i>M9</i> | <i>M10</i> | <i>M11</i> | <i>Q</i> |
|----------------------|-----------|-----------|-----------|------------|------------|----------|
| additiv modell       | 0.339     | 0.425     | 0.306     | 0.448      | 0.439      | 0.353    |
| multiplikativ modell | 0.212     | 0.467     | 0.083     | 0.528      | 0.521      | 0.288    |

Tabell 11.7: *Sliding spans*

|                      | <i>S</i> (%) | <i>MM</i> (%) | <i>YY</i> (%) |
|----------------------|--------------|---------------|---------------|
| additiv modell       | 35.6         | 35.9          | 12.0          |
| multiplikativ modell | 14.4         | 15.1          | 1.1           |

## 5. Figurer

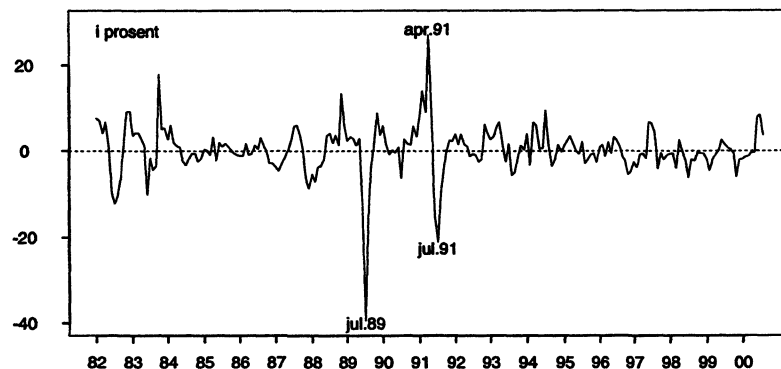
Figur 11.5 og 11.6 viser de sesongjusterte tallene  $A_t$  for de to modellene (11.1) og (11.2) samt deres avvik i prosent  $((A_t^{add} - A_t^{mult}) \times 100\% / A_t^{add})$ .



Figur 11.5: *Sesongjusterte tall ved additiv og multiplikativ modell*

Avvikene i de to kurvene er store i juli 1989 med -39.40, april 1991 med 27.05, juli 1991 med -21.06. Sesongjusterte tall  $A_t$  og rådataene  $O_t$  i disse tidspunktene er vist i tabell 11.8. Det viser at avvikene mellom rådataene og de sesongjusterte tallene er større for den additive modellen.

Sliding spansanalysen er utført i fire delvis overlappende delmengder [jan.1990, aug.1997], [jan.1991, aug.1998], [jan.1992, aug.1999] og [jan.1993, aug.2000]. Programmet skal beregne de største endringene  $A_t^{max}$  av sesongjusterte tall fra



Figur 11.6: Avviket i prosent av sesongjusterte tall ved additiv og multiplikativ modell

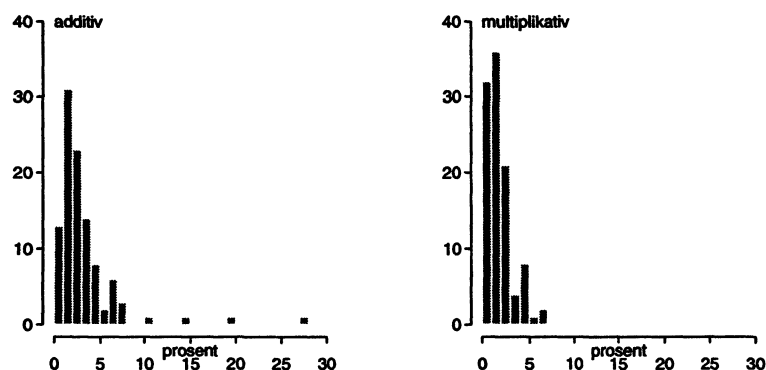
Tabell 11.8:  $O_t$  og  $A_t$  i juli 1989, april 1991 og juli 1991

|             | juli 1989 | april 1991 | juli 1991 |
|-------------|-----------|------------|-----------|
| $O_t$       | 128263    | 31195      | 136606    |
| $A_t$ Add.  | 54943     | 66081      | 64658     |
| $A_t$ Mult. | 76600     | 48205      | 78275     |

de fire intervallene

$$A_t^{max} = \frac{\max_{k \in N_t} A_t(k) - \min_{k \in N_t} A_t(k)}{\min_{k \in N_t} A_t(k)}$$

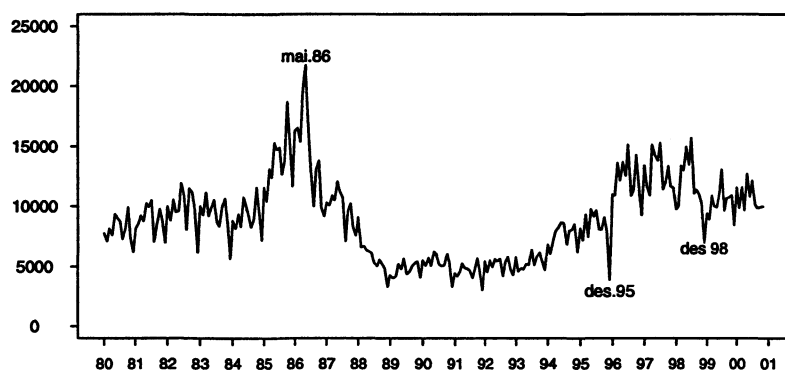
hvor  $t$  går fra januar 1991 til august 1999 (104 tidspunkter). Et sesongjustert tall i måned  $t$  kalles upålitelig hvis  $A_t^{max} > 0.03$ . Det er 37 av 104 tidspunkter (som utgjør 35.6%) der de sesongjusterte tallene er upålitelige for den additive modellen, mens for den multiplikative modellen er det 14.4%. Grenseverdien for denne størrelsen er 15%. Vi får dermed en akseptabel sesongjustering for den multiplikative modellen. Histogrammene av  $A_t^{max}$  er plottet i figur 11.7, de  $x$ -aksen betyr i prosentavvik og  $y$ -aksen er prosentandelen.



Figur 11.7: Histogram av de største endringene i sesongjusterte tall i fire spans

### 11.3 Eksempel 2. Antall registrerte nye biler i Norge

I dette eksemplet vil vi vise at ved å velge en riktig modell for sesongjustering kan sesongjusterte tall bli mer stabile. Månedsserien starter opp i januar 1976 og ender i november 2000. I alt er det 252 observasjoner. Rådataene er plottet i figur 11.8. Det er ekstremverdier som skaper store endringer i rådataene i et tidspunkt eller i en periode. For eksempel, høyere avgifter for nye biler som ble innført fra januar 1978 medførte at antall biler i desember 1977 var mye høyere enn de andre desember månedene, mens vi har en motsatt effekt for desember 1995 på grunn av lave avgifter fra januar 1996. Avgiftsnedgangen var forhåndsvarslet. Salget i desember 1995 ble dermed utsatt. I neste avsnitt vil vi presentere resultater fra sesongjustering med additiv og multiplikativ modell. Vi bruker fortsatt sliding spansanalysen for å vurdere hvor god en sesongjustering er.



Figur 11.8: Antall registrerte nye biler i Norge

### 11.4 Resultater fra sesongjustering

Siden  $AICC_{nolog} = 4088$ ,  $AICC_{log} = 3995$  og  $\Delta_{AICC} = -2$ , har vi  $AICC_{nolog} > AICC_{log} + \Delta_{AICC}$ . Dermed er den multiplikative modellen valgt. Men vi skal beskrive resultater for begge modeller. For hver modell har vi kjørt X-12-ARIMA flere ganger. I første kjøring endrer vi ikke de kritiske verdier som har lagt inn på forhånd i programmet for opsjoner. Utfilen brukes for å justere disse verdiene i de nye kjøringene slik at vi får en bedre sesongjustering.

#### 1. ARIMA modell

Ingen av de fem innebygde ARIMA modellene tilfredstiller kriteriene for å bli valgt ved å bruke den additive modellen, siden  $\chi^2$  for å teste hypotesen om korrelasjonen i residualene er for lav (0.57%). Grenseverdien er 5%. ARIMA(0 1 1)(0 1 1) er brukt for beregningen. De estimerte parametre er listet ut i tabell 11.9.

#### 2. Kalendereffekter

Ukedagseffekter er signifikante på 5% nivå for begge modeller. Effektene er vist i tabellene 11.10 og 11.11. Onsdag og fredag (bare for additiv modell) er de

Tabell 11.9: *De estimerte parametre fra ARIMA modeller*

|       | ARIMA modell     | Est.Par.                | std    | Konklusjon       |
|-------|------------------|-------------------------|--------|------------------|
| Add.  | (0 1 1)(0 1 1)   | $\hat{\theta} = 0.3747$ | 0.0593 | sign. på 5% nivå |
|       | default modell   | $\hat{\Theta} = 0.6479$ | 0.0519 | sign. på 5% nivå |
| Mult. | (0 1 1)(0 1 1)   | $\hat{\theta} = 0.4417$ | 0.0563 | sign. på 5% nivå |
|       | valgt av program | $\hat{\Theta} = 0.8467$ | 0.0438 | sign. på 5% nivå |

to dagene som utpekes til å være de beste salgsdagene i uka, mens lørdag og søndag er de laveste for begge modeller.

Tabell 11.10: *Ukedagseffekter ved additiv modell*

|         | effekt  | std    | t-verdi | Konklusjon             |
|---------|---------|--------|---------|------------------------|
| mandag  | -141.74 | 100.59 | -1.41   |                        |
| tirsdag | 68.51   | 100.27 | 0.68    |                        |
| onsdag  | 325.34  | 98.92  | 3.29    | signifikant på 5% nivå |
| torsdag | -99.07  | 98.96  | -1.00   |                        |
| fredag  | 243.73  | 98.09  | 2.48    | signifikant på 5% nivå |
| lørdag  | -178.04 | 100.00 | -1.78   |                        |
| søndag  | -218.73 | 101.63 | -2.15   | signifikant på 5% nivå |

Tabell 11.11: *Ukedagseffekter ved multiplikativ modell*

|         | effekt  | std    | t-verdi | Konklusjon             |
|---------|---------|--------|---------|------------------------|
| mandag  | 0.0037  | 0.0128 | 0.29    |                        |
| tirsdag | 0.0049  | 0.0127 | 0.39    |                        |
| onsdag  | 0.0256  | 0.0125 | 2.04    | signifikant på 5% nivå |
| torsdag | 0.0003  | 0.0126 | 0.03    |                        |
| fredag  | 0.0164  | 0.0125 | 1.31    |                        |
| lørdag  | -0.0142 | 0.0126 | -1.12   |                        |
| søndag  | -0.0367 | 0.0128 | -2.85   | signifikant på 5% nivå |

Påskeeffekten er signifikant på 5% nivå med  $\hat{d} = 1$  for begge modeller.

## 2. Intervensjoner

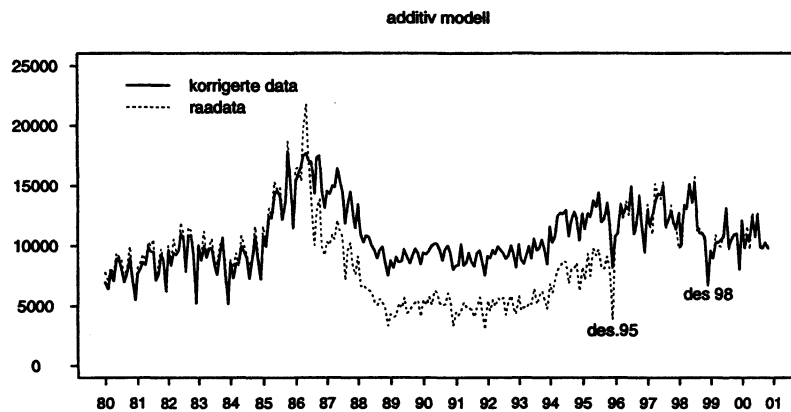
Tabell 11.12 viser typer og effekter for intervensjoner fra begge modeller.

De korrigererte dataene for kalendereffekter og intervensjoner er plottet i figurene 11.9 og 11.10. For additiv modell blir rådataene korrigert opp med 4183 enheter

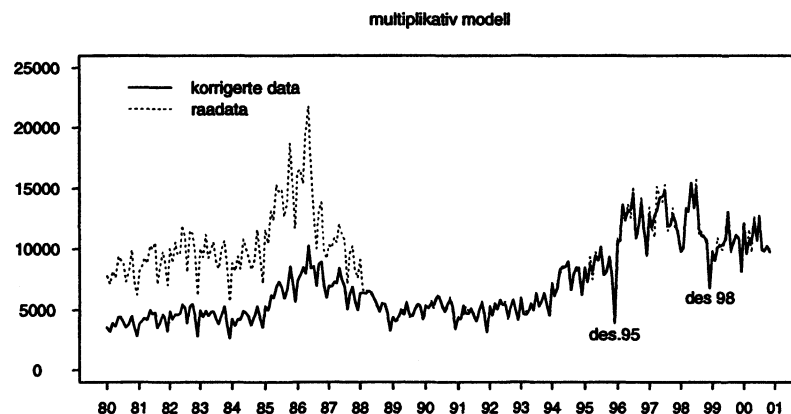
Tabell 11.12: *Intervensjoner ved additiv modell*

|                                     | år   | måned   | type | $f_t$   | std    | $t$   |
|-------------------------------------|------|---------|------|---------|--------|-------|
| Additiv modell                      | 1986 | mai     | AO   | 3614    | 698    | 5.18  |
|                                     | 1986 | juli    | LS   | -4708   | 772    | -6.10 |
|                                     | 1996 | januar  | LS   | 4183    | 758    | 5.52  |
|                                     | 1998 | januar  | AO   | -2920   | 691    | -4.22 |
| Multiplikativ modell<br>(log skala) | 1986 | juli    | LS   | -0.3595 | 0.0883 | -4.07 |
|                                     | 1988 | februar | LS   | -0.3956 | 0.0882 | -4.48 |

for tidspunktene fra januar 1980 til desember 1995 på grunn av nivåskiftet i januar 1996 og deretter ned med 4708 enheter for de tidspunktene fra januar 1980 til juni 1986. Sammenligning med det opprinnelige nivået er de korrigerte dataene høyere med 4183 enheter i perioden [juli 1986, desember 1995] og lavere med  $4708 - 4183 = 525$  enheter i intervallet [januar 1980, juni 1986]. Dessuten er tallene i mai 1986 og januar 1998 også justert.



Figur 11.9: *Korrigerte dataene for kalenderseffekter og intervensjoner*



Figur 11.10: *Korrigerte dataene for kalenderseffekter og intervensjoner*

For multiplikativ modell er rådataene i *log* skalaen korrigert ned to ganger for to nivåskifter i juli 1986 og februar 1988 med 0.3595 og 0.3956, henholdsvis. Dvs, dataene blir justert ned med  $0.3595+0.3956=0.7551$  for tidspunktene fra januar 1980 til juni 1986 og 0.3956 i perioden [juli 1986, januar 1988].

Merk at mønstreret av rådataene kan bli helt annerledes sammenligning med de opprinnelige tallene etter korrigeringen for intervensjoner av typen LS (nivåskift). En ARIMA modell blir ofte tilpasset lettere med de korrigerede dataene. Bortsett fra de spesielle hendelser som påvirker kraftigere rådata får vi samme tidspunkter og typer for intervensjoner fra additiv eller multiplikativ modell i estimeringen. Ellers kan det være ekstreme verdier etter å ha transformert dataene eller det omvendt, de ekstreme verdiene i rådataene forsvinner etter transformeringen. Jo bedre en kjenner tidsserien, desto lettere er det å forklare eller behandle intervensjonene. Vi merker også at korrigeringen for intervensjoner ikke forandrer sesongvariasjonene i tidsserien. For eksempel, devalueringen i mai 1986 førte til at salget av antall nye personbiler gikk kraftig ned fra mai til august før tallet økte i september for deretter å synke ytterligere (se tabell nedenfor). Reduksjonen fra juli til august er -23.70%, mens økningen fra august til september 31.26%. Dette beskrives av nivåskiftet i juli 1986.

| mai.86 | jun.86 | jul.86 | aug.86 | sep.86 |
|--------|--------|--------|--------|--------|
| 21793  | 17241  | 13079  | 9982   | 13103  |

Lavt tall i desember 1995 (3836 biler) har trolig sammenheng med lavere bilavgifter fra januar 1996.

| okt.95 | nov.95 | des.95 | jan.96 | feb.96 |
|--------|--------|--------|--------|--------|
| 9078   | 8006   | 3836   | 11010  | 10920  |

### 3. Framskrivinger

Vi bruker symbolene:

$F_{A,t}$  framskrivningen av rådata med additiv modell og uten korrigering for kalendereffekter

$F_{M,t}$  er den samme størrelsen men med multiplikativ modell

$F_{A,t}^K$  er  $F_{A,t}$  men med korrigering for kalendereffekter for 2001

$F_{M,t}^K$  er den samme størrelsen men med multiplikativ modell

$R_t^F$  relativt avvik i prosent mellom  $F_{A,t}$  og  $F_{M,t}$ .

$$R_t^F = 100\% \times (F_{A,t} - F_{M,t}) / F_{A,t}$$

$\hat{A}_{A,t}$  framskrivningen av sesongjusterte tall  $A_t$  med additiv modell

$\hat{A}_{M,t}$  er den samme størrelsen men med multiplikativ modell

$\hat{T}_{A,t}$  framskrivningen av trend  $T_t$  med additiv modell

$\hat{T}_{M,t}$  er den samme størrelsen men med multiplikativ modell

$R_t^A$  relativt avvik i prosent mellom  $\hat{A}_{A,t}$  og  $\hat{A}_{M,t}$ .

$$R_t^A = 100\% \times (\hat{A}_{A,t} - \hat{A}_{M,t}) / \hat{A}_{A,t}$$

$R_t^T$  relativt avvik i prosent mellom  $\hat{T}_{A,t}$  og  $\hat{T}_{M,t}$ .

$$R_t^T = 100\% \times (\hat{T}_{A,t} - \hat{T}_{M,t}) / \hat{T}_{A,t}$$

Tabell 11.15 viser framskrivingene  $F_{t,t}$  samt deres 95% konfidensintervaller for ett år framover ved additiv og multiplikativ modell. Vi får høye verdier og større konfidensintervaller for  $F_{M,t}$  enn  $F_{A,t}$  i 10 av 12 måneder. Deres relative avvik  $R_t^F$  er store særlig for september, oktober og november 2001, med 8.26, 11.30 og 7.34%.

Framskrivingene for sesongjusterte tall og trend er listet ut i tabell 11.13. Vi får også store avvik mellom dem for de to modellene, særlig i de siste månedene.

Tabell 11.13: *Framskrivinger av sesongjusterte tall og trend*

| år   | måned     | $\hat{A}_{A,t}$ | $\hat{A}_{M,t}$ | $R_t^A$ | $\hat{T}_{A,t}$ | $\hat{T}_{M,t}$ | $R_t^T$ |
|------|-----------|-----------------|-----------------|---------|-----------------|-----------------|---------|
| 2000 | desember  | 10409           | 10414           | -0.04   | 10135           | 10119           | 0.15    |
| 2001 | januar    | 9971            | 9776            | 1.95    | 10032           | 10150           | -1.17   |
|      | februar   | 9693            | 10333           | -6.60   | 9908            | 10200           | -2.94   |
|      | mars      | 9530            | 9928            | -4.17   | 9776            | 10219           | -4.53   |
|      | april     | 10424           | 10947           | -5.01   | 9649            | 10186           | -5.56   |
|      | mai       | 9054            | 9806            | -8.30   | 9563            | 10135           | -5.98   |
|      | juni      | 9550            | 10289           | -7.73   | 9533            | 10108           | -6.03   |
|      | juli      | 9816            | 10030           | -2.18   | 9549            | 10130           | -6.08   |
|      | august    | 9066            | 9585            | -5.72   | 9577            | 10204           | -6.54   |
|      | september | 10021           | 11006           | -9.82   | 9608            | 10305           | -7.25   |
|      | oktober   | 9532            | 10421           | -9.32   | 9636            | 10410           | -8.03   |
|      | november  | 9413            | 10181           | -8.15   | 9667            | 10506           | -8.67   |

De to tabellene 11.15 og 11.13 viser at framskrivingene fra en additiv modell kan være svært forskjellig fra de som beregnes med den multiplikative modellen.

### 3. Tester for sesongmønster og sesongbegelser

De to (additiv og multiplikativ) modellene gir samme konklusjoner om tester. Sesongmønsteret og sesongbevegelsen er signifikante på 1% nivå.

### 4. Kvalitetsmål og sliding spans

Det vises i tabell 11.14. Vi får lavere verdier for  $M8 - M11$  og  $S(\%)$ ,  $MM(\%)$  og  $YY(\%)$  for den multiplikative modellen som viser en mer stabil sesongkomponent. Sliding spansanalysen er utført i fire delvis delmengder [jan.1986-nov.1997], [jan.1987-nov.1998], [jan.1988-nov.1999] og [jan.1989-nov.2000].

Tabell 11.14: *Kvalitetsmål og sliding spans*

|            | $M1$  | $M2$  | $M3$  | $M4$  | $M5$  | $M6$  | $M7$  | $M8$  |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|
| add. mod.  | 0.423 | 0.102 | 0.444 | 0.058 | 0.610 | 0.223 | 0.414 | 0.783 |
| mult. mod. | 0.727 | 0.112 | 0.554 | 0.641 | 0.650 | 0.403 | 0.270 | 0.521 |

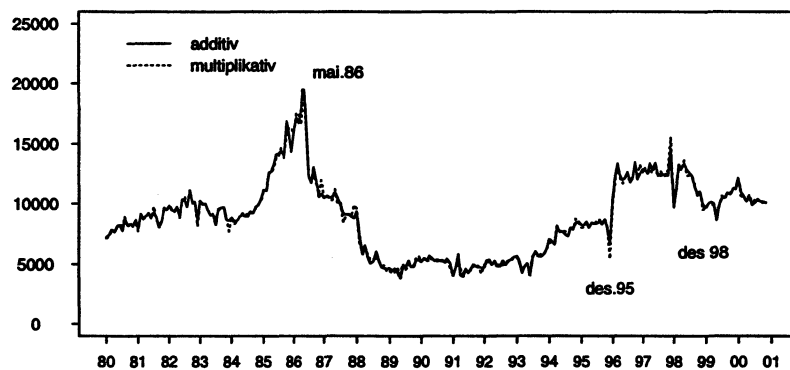
Tabell 11.15: *Framskrivinger og 95% konfidensintervaller*

| år   | måned     | additiv modell |           |       |             | multiplikativ modell |           |       |             | avvik% |
|------|-----------|----------------|-----------|-------|-------------|----------------------|-----------|-------|-------------|--------|
|      |           | Nedre          | $F_{A,t}$ | Øvre  | $F_{A,t}^K$ | Nedre                | $F_{M,t}$ | Øvre  | $F_{M,t}^K$ |        |
| 2000 | desember  | 5180           | 6936      | 8691  | 7089        | 5631                 | 6880      | 8406  | 7122        | 0.81   |
| 2001 | januar    | 8324           | 10396     | 12468 | 10145       | 8333                 | 10473     | 13162 | 10121       | -0.73  |
|      | februar   | 6216           | 8551      | 10885 | 8548        | 7057                 | 9086      | 11699 | 9168        | -5.88  |
|      | mars      | 7768           | 10342     | 12916 | 10021       | 8314                 | 10972     | 14479 | 10574       | -5.74  |
|      | april     | 6420           | 9210      | 12000 | 9924        | 7264                 | 9788      | 13189 | 10470       | -5.90  |
|      | mai       | 8491           | 11482     | 14473 | 11186       | 8674                 | 11918     | 16376 | 11556       | -3.65  |
|      | juni      | 7195           | 10383     | 13571 | 10320       | 7922                 | 11101     | 15555 | 11077       | -6.46  |
|      | juli      | 8434           | 11802     | 15169 | 12091       | 8183                 | 11668     | 16635 | 12000       | 1.14   |
|      | august    | 6008           | 9542      | 13076 | 9071        | 6834                 | 9906      | 14359 | 9496        | -3.67  |
|      | september | 5136           | 8832      | 12528 | 9229        | 6536                 | 9628      | 14183 | 10131       | -8.26  |
|      | oktober   | 6214           | 10070     | 13925 | 9819        | 7587                 | 11354     | 16992 | 10973       | -11.30 |
|      | november  | 5432           | 9437      | 13442 | 9290        | 6706                 | 10185     | 15469 | 10016       | -7.34  |

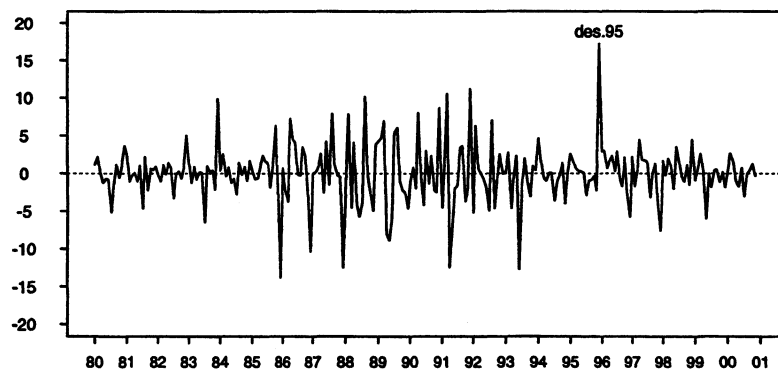
|            | $M9$  | $M10$ | $M11$ | $Q$   | $S(\%)$ | $MM(\%)$ | $YY(\%)$ |
|------------|-------|-------|-------|-------|---------|----------|----------|
| add. mod.  | 0.303 | 0.988 | 0.979 | 0.443 | 9.7     | 18.2     | 1.4      |
| mult. mod. | 0.253 | 0.589 | 0.562 | 0.458 | 4.5     | 13.6     | 0.0      |

#### 4. Figurer

Figurene 11.11 og 11.13 viser sesongjusterte tall  $A_{A,t}$  og  $A_{M,t}$  og trendene  $T_{A,t}$  og  $T_{M,t}$  som beregnes med additiv og multiplikativ modell. Deres relative avviker i prosent er plottet i figurene 11.12 og 11.14.

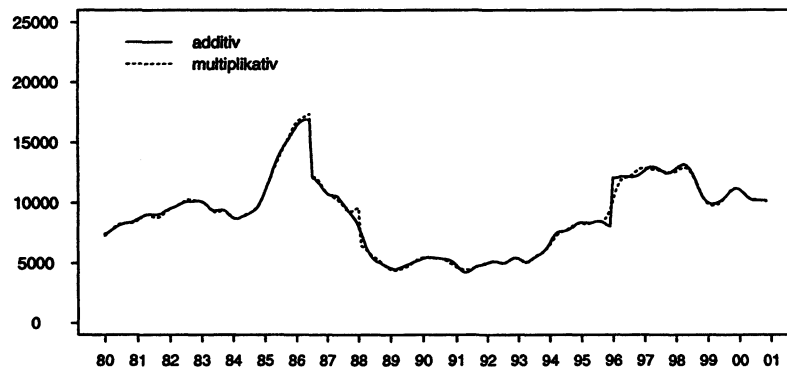


Figur 11.11: Sesongjusterte tall ved additiv og multiplikativ modell



Figur 11.12: Relative avvikene i prosent av sesongjusterte tall

Vi ser at de relative avvikene i prosent  $R_t^A$  av de sesongjusterte tallene  $A_{A,t}$  og  $A_{M,t}$  er store i midten av tidsserien. De varierer rundt intervallet  $[-15\%, 15\%]$ . For trenden er  $R_t^T$  mindre. Tabell 11.16 viser sesongjusterte tall og trender samt deres relative avviker i prosent for år 2000. Forskjellene er små. Man ofte er interessert i de siste observasjonene.



Figur 11.13: *Trender ved additiv og multiplikativ modell*



Figur 11.14: *Relative avvikene i prosent av trender*

Tabell 11.16: *Sesongjusterte tall og trender i 2000 for additiv og multiplikativ modell.*

| måned     | $A_{A,t}$ | $A_{M,t}$ | $R_t^A$ | $T_{A,t}$ | $T_{M,t}$ | $R_t^T$ |
|-----------|-----------|-----------|---------|-----------|-----------|---------|
| januar    | 12167     | 12113     | 0.44    | 11076     | 11078     | -0.01   |
| februar   | 10967     | 10672     | 2.68    | 10908     | 10887     | 0.19    |
| mars      | 10633     | 10454     | 1.68    | 10686     | 10661     | 0.23    |
| april     | 10195     | 10287     | -0.90   | 10468     | 10470     | -0.01   |
| mai       | 10597     | 10781     | -1.73   | 10303     | 10346     | -0.41   |
| juni      | 9958      | 9881      | 0.77    | 10210     | 10288     | -0.76   |
| juli      | 10069     | 10369     | -2.97   | 10172     | 10259     | -0.85   |
| august    | 10394     | 10392     | 0.01    | 10180     | 10222     | -0.41   |
| september | 10185     | 10148     | 0.36    | 10200     | 10171     | 0.28    |
| oktober   | 10203     | 10074     | 1.26    | 10207     | 10129     | 0.76    |
| november  | 10065     | 10099     | -0.33   | 10190     | 10113     | 0.75    |

## 11.5 Konklusjoner

Eksemplet 1 viser at modellen for sesongjustering som er valgt ved testen i X-12-ARIMA, kan gi mer ustabile sesongjusterte tall enn den modellen som er forkastet. I eksemplet 2 viser vi at resultatene fra sesongjustering av en additiv modell kan være ganske forskjellige fra en multiplikativ modell. Dataene får dermed et uklart sesongmønster avhengig av valg av modell, vi bør sesongjustere dataene med både additiv og multiplikativ modell og velge den modellen som gir best resultater. Det tar ikke langt tid å kjøre en tidsserie med X-12-ARIMA. Vi bør studere tall fra tabeller sammen med figurer for å trekke mer sikre konklusjoner.

## Kapittel 12

# Internasjonalt seminar om sesongjusteringsmetoder: Rapport fra SAM 98 "Seasonal Adjustment Meeting 1998"

Anne B. Dahle<sup>1</sup>, Dinh Quang Pham<sup>2</sup>, Joaquin Rodriguez<sup>3</sup> og Terje Skjerpen<sup>4</sup>

<sup>1</sup> Avdeling for økonomisk statistikk, Seksjon for utenrikshandel, energi og industri-statistikk

<sup>2</sup> Avdeling for samordning og utvikling, Seksjon for statistiske metoder og standarder

<sup>3</sup> Avdeling for økonomisk statistikk, Seksjon for økonomiske indikatorer

<sup>4</sup> Avdeling for forskning, Seksjon for makroøkonomi

### 12.1 Sammendrag

Det internasjonale seminaret om sesongjusteringsmetoder ble arrangert av Eurostat 22.-24. oktober 1998. Seminaret fant sted i Bucharest. Det samlet 90 deltakere, heriblant ledende teoretikere, deltakere fra statistikkbyråer og/eller sentralbanker i de fleste EU-land, USA, Canada, Mexico, New Zealand, Tsjekkia, Ungarn, Estland, Russland, Litauen, Macedonia, Israel og Tunis, foruten vertslandet Romania. Konferansen fant sted i verdens nest største bygning, som har et samlet gulvareal på 256 000 m<sup>2</sup> (bare Pentagon er større). Parlamentspalasset er en økonomisk tyngende, håndverksmessig overdådig mastodont av en bygning som laver av marmor, fargestrålende spesialvevde tepper, funkklende krystallysekroner, utsøkt treskurd og - moderne skulpturer og malerier av høy kvalitet.

Foredragene gikk dels i plenum og dels i parallelle sesjoner. Seminaret samlet de fleste av de store teoretikerne innen sesongjustering, heriblant David Findley, Andrew Harvey, Danny Pfeffermann, Agustin Maravall og George Tiao. Sett under ett lå hovedvekten i deres bidrag på vurdering av de teoretiske og praktiske egenskapene ved

hovedmetodene for sesongjustering og strategier for å måle og uttrykke usikkerhet ved de resulterende seriene.

De to hovedmetodene for sesongjustering er X12-ARIMA og TRAMO-SEATS. Hvilken metode som nyttes av enkelte land eller institusjoner avhenger av tradisjoner (opphav og eierskap til metodene) og hva resultatene skal brukes til (brukergruppe). Eurostat justerer nå tidsseriene for korttidsindikatorer vha. sitt egenutviklede TRAMO-SEATS og serier for utenrikshandel med varer med X12-ARIMA som er US Bureau of Census's videreutvikling av Statistics Canadas X11-ARIMA. Office for National Statistics, London, bruker STAMP for å justere et forholdsvis begrenset antall sentrale tidsserier.

### 12.1.1 Politikkvalg for sesongjustering

Vi gir her et sammendrag av de viktigste synspunktene som fremkom under seminaret med hensyn til politikkvalg ved sesongjustering, stort sett med Eurostats systematisering.

#### *Valg av sesongjusteringsmetode*

bør baseres på teoretiske og empiriske kriterier ut fra kvaliteten som kreves av estimatene og hva de skal brukes til (makropolitiske eller andre anvendelser). Makropolitiske anvendelser kan innebære sesongjustering av et begrenset antall serier. Sesongjustering av et stort antall serier krever en metode med muligheter for ulike valg av opsjoner for ulike serier - som f.eks. påvirkes på forskjellig måte av virkedagseffekter - innen ett og samme program.

#### *Skifte av sesongjusteringsmetode*

Metodeansvarlig i Eurostat (Berthold Feldmann) ønsker prinsippavgjørelse på å bruke kun én sesongjusteringsmetode for justeringene som gjøres i Eurostat, mens valg av metode for nasjonal statistikk fortsatt vil være opp til hvert enkelt land å avgjøre. De estimerte seriene vil i større eller mindre grad være avhengig av metoden. Et skifte av metode må derfor være grundig overveid og må kunne begrunnes og forklares overfor brukerne.

#### *Gjennomsiktighet i prosedyrene*

Det ble tilrådd å publisere metadata (valg av metode, valg av opsjoner). Eurostat anbefaler at dette gjøres i en særskilt metodepublikasjon, og at det i tillegg gis informasjon til brukerne ved hver publisering.

#### *Publisering av sesongjusterte eller trend-sykel data*

Med trend-sykel data forstås tall som inneholder både langsiktig, deterministisk trend og kortsiktig konjunkturtrend. Trendsykel-serier kalles også ofte trendserier. Eurostat anbefaler at det publiseres både sesongjusterte data og trend-data i tabeller, men at det i grafer først og fremst publiseres trendserier. Det britiske Office for National Statistics presenterte en interessant måte å gjøre dette på, se seksjonen 12.5.

#### *Revisjoner av komponenter og serier*

Det ble foreslått å binde (de publiserte) seriene etter et visst antall år, selv om de estimerte seriene fortsatt endres litt.

### *Metodisk anvendelse av sesongjusteringsprogrammer*

Avgjørelser om behandlingen av ekstreme observasjoner i en serie, valg av additiv eller multiplikativ modell må være basert på relevante, vitenskapelige tester. Sesongjustering kan ikke fullt ut baseres på default-opsjonene i et standardverktøy: en tilsynelatende god modelltilpasning er ikke nødvendigvis noen garanti for at sesongjusteringen er vellykket og gir gode prognoser.

### *Behandlingen av usikkerhet og publisering av konfidensintervall*

Brukere bør gjøres oppmerksom på usikkerhet ved publiserte tall. Hvordan, og i hvilket omfang var det derimot svært delte meninger om. Eurostat anbefaler at nasjonale statistikkbyråer gir informasjon om standardavvik og konfidensintervall (kun for estimeringsfeilen i selve prosedyren).

### *Effektiv opplæring i sesongjustering*

Det er viktig med effektiv opplæring i sesongjustering, på grunn av risikoen for dårlige resultater ved ukyndig bruk. Det ble også pekt på at det er viktig at metoden som anvendes har tilstrekkelig, brukervennlig diagnostikk.

## **12.1.2 Anbefalinger**

Vår samlede vurdering er at SSB bør prioritere å ta i bruk X12-ARIMA fra U.S. Bureau of Census som sesongjusteringsverktøy. Vi anbefaler at SSB også følger med i utviklingen av Eurostats applikasjon DEMETRA, hvor både X12-ARIMA og TRAMO-SEATS vil være tilgjengelige. DEMETRA foreligger inntil videre kun i PC-versjon.

## **12.2 Invitert foredrag**

Professor George C. Tiao (University of Chicago, USA) var spesielt invitert til å holde en oversiktsforelesning. Tittelen var "An historical review of seasonal adjustment and some perspectives".

Tiao er unektelig en av nestorene innenfor den akademiske forskningen knyttet til dekomponering av tidsserier. Tiaos fremste bidrag var å sammenligne de enkelte angrepsmåtene, som for eksempel den ikke-modellbaserte angrepsvinkel med den modellbaserte. Tiao refererte til Cleveland og Tiao (1976). Der settes det opp en modell i ikke-observerbare komponenter som er laget slik at hvis en sesongjusterer med denne modellen så får en om lag det samme som om en sesongjusterer denne serien i X11. Fra modellen i ikke-observerbare komponenter kan en videre dedusere en ARIMA-modell for inputserien der koeffisientene er kjente. Dette representerer en viktig forbindelseslinje mellom den ikke-modellbaserte og den modellbaserte tilnærmelsene til sesongjustering.

Tiao understreket noen av fordelene, men samtidig noen av begrensningene ved den modellbaserte fremgangsmåte. Fordelene er at en gjør forutsetningene eksplisitt. Dette muliggjør statistisk hypotesetesting. Begrensingen er først og fremst knyttet til spørsmålet om identifikasjon. Dette betyr at det finnes ikke en, men en rekke dekomponeringsmodeller som alle er likeverdige. Det betyr at når en løser modellen slik at de ikke-observerbare komponentene elimineres, sitter en igjen med den samme overordnede modellen. De ikke-observerbare komponentene følger stokastiske prosesser og har egne spesifikke feilledd knyttet til seg. Hvilken varians de ulike feilledd har er

ikke identifiserbart, men må (innenfor visse rammer) postuleres. En spesiell dekomponering er den såkalte kanoniske dekomponeringen se f.eks. Hillmer og Tiao (1982). Den kanoniske dekomponeringen innebærer at en forutsetter at den største mulige variansen er knyttet til den irregulære komponenten. Dette innebærer at trend- og sesongkomponenten utvikler seg langsomt. (En forutsetning om kanonisk dekomponering ligger også til grunn for programpakken SEATS).

Tiao knyttet også noen kommentarer til programmene X12-ARIMA og TRAMO-SEATS og understreket at selv om disse to programmene bygger på forskjellige angrepsmåter gir de i praksis ofte resultater som er meget like. Tiao var også inne på problemstillinger knyttet til prediksjon av ikke-observerbare komponenter som trend og sesong samt til variansen på disse prediksjonsfeilene. Tiao brukte i hele sitt foredrag tre faktiske tidsserier for å få fram sine poenger.

### 12.3 Avsluttende rundebordsdiskusjon

Rundebordsdiskusjonen som avsluttet konferansen hadde som tema "Heuristic and model-based methods in Seasonal adjustment". Deltagere var William Bell (Bureau of Census, USA), David Findley (Bureau of Census, USA), Andrew Harvey (University of Cambridge, UK), Siegfried Heiler (Konstanz University, Tyskland), Agustin Maravall (Central Bank, Madrid) (under innledningen), Danny Pfeffermann (Hebrew University, Jerusalem) og George C. Tiao, samt flere bidragsytere fra salen.

Tiao uttrykte sin tilfredshet med at Eurostat er i ferd med å beslutte seg til å bruke et modellbasert system for sesongjustering og understreket sine hovedpoeng fra det inviterte foredraget.

Harvey fremholdt at det er flere måter å kjøre modellbasert sesongjustering og at det kan være aktuelt å særbehandle visse viktige serier som er nøkkeltall blant økonomiske indikatorer. Videre hevdet han at det er viktig å ha teori i bunnen. Dette gjelder f.eks. revisjoner av estimatene. Å ha en politikk som går ut på at det gjelder å unngå revisjoner er ikke riktig, dersom det er virkelige årsaker til revisjoner.

Pfeffermann var av den oppfatning at i de fleste land er det myndighetene som utvikler sesongjusteringsverktøyer. Myndigheter er generelt sett skeptiske til modellbaserte metoder og vil helst bruke ikke-modellbaserte (ad hoc) metoder. Pfeffermann så tre fordeler ved en helt modellbasert metode: den har modell for residualene, muliggjør multivariat sesongjustering og den kan gi grad av sannsynlighet gitt modellen, dvs. et mål for usikkerhet.

Björn Fischer (tidligere Eurostat, nå Central Bank, Frankfurt) pekte på konflikten mellom optimale, teoretiske løsninger og de praktiske mulighetene for å implementere dem.

Etter Bells mening er det behov for forskning som kobler sammen sesongjusteringen med bruken av resultatene. Han pekte på at noen mulige metoder for å formidle variansen ved de estimerte seriene eller usikkerheten ved resultatene var presentert i foredragene.

Eurostats sesongjusterings- og programmeringsekspert, Gian Luigi Mazzi - som for øvrig er blind - pekte på viktigheten av å gjøre mye mer bruk av informasjonen som ligger i rådata, i forbindelse med sesongjusteringen. Han pekte også på at

det er nødvendig å være oppmerksom på at brukbarheten og nytten av trendserier (trendsykel-serier) avhenger av støynivået i serien og av brukernes nivå (kompetanse)

Maravall fremholdt innledningsvis at det er viktig å vurdere og sammenligne egenskapene ved TRAMO-SEATS og X12-ARIMA. En oppfatning som ble fremført var at sesongjustering må være modellbasert: TRAMO-SEATS er det, mens X12-ARIMA ikke ennå fullt ut modellbasert. På den annen side ble det hevdet at TRAMO-SEATS trenger mer diagnostikk. X12-ARIMA ble hevdet å gi overestimering av sesongkomponenten mot slutten av visse serier, men at et mulig botemiddel kan være å bruke kortere serier i X12-ARIMA. TRAMO-SEATS kan gi svært dårlige prediksjoner, selv med tilsynelatende meget god modelltilpasning. Det var enighet om at det er et stort behov for metadata som beskriver de modellbaserte metodene på en måte som er forståelig for den vanlige bruker.

Om særskilte problemstillinger ved sesongjusteringer av data fra utvalgstillinger ble det påpekt at utvalgsfeil gir kraftig innvirkning på trendestimerer.

Panelet berørte også usikkerheten ved trend ved enden av serie. Fra Eurostats side ble det fremholdt at det er viktig både å vite og si noe om usikkerheten. Office for National Statistics nevnte sin strategi med å publisere trend som stiplet linje, samt tre scenarioer for trend med ytterpunkter "optimistisk" og "pessimistisk".

## 12.4 Teoretiske aspekter ved sesongjustering

Sesjonene I, V A og VIII B hadde som felles tittel "Theoretical aspects in Seasonal Adjustment".

A. Harvey innledet sesjon I med foredraget "Seasonal Adjustment and Structural Time Series Models". Siden strukturelle tidsseriemodeller er kjennetegnet ved at de ulike tidsseriekomponentene modelleres eksplisitt, egner disse modellene seg godt for sesongjustering av tidsserier. Harvey har (i samarbeid med andre) utviklet softwareprogrammet STAMP som blant annet kan utnyttes til sesongjusteringsformål (jf. Harvey (1989) og Koopman mfl. (1995)). I dette programmet utnytter en at de strukturelle tidsseriemodellene kan formuleres som tilstandsmodeller. Tradisjonelt gjennomføres sesongjustering på en univariat basis, dvs. tidsseriene sesongjusteres hver for seg. I STAMP er det også mulig å sesongjustere tidsserier innenfor en ramme der en har en multivariat modell. For eksempel kan en ha den situasjon at to tidsserier er kointegrerte ved sesongfrekvensene. Dette betyr at de to tidsseriene deler en felles stokastisk sesongkomponent. Hvis dette er tilfelle synes det naturlig å sesongjustere innenfor en multivariat ramme.

Begrensningen i STAMP 5.0 er at programmet er menybasert slik at det ikke er mulig å programmere modeller som avviker marginalt fra de som er lagt inn i programmet. Et viktig moment ved sesongjustering er hensynstagen til kalendereffekter (f. eks. påskens plassering i kalenderåret og antall virkedager). Til en viss grad er det mulig å ta hensyn til slike effekter i STAMP 5.0, men i STAMP 6.0 som snart vil bli lansert har det blitt arbeidet en del med å forbedre de rutineene som tar hensyn til dette. Blant annet vil en kunne ha modeller der det tillates at effekten av påske endres over tid. Det bør også nevnes at det eksisterer et mer generelt program der brukeren kan formulere egne tilstandsmodeller. Dette programmet heter SsfPack og er programmert i Ox. I allfall inntil videre kan dette programmet lastes ned gratis fra

Internet (<http://center.kub.nl/stamp/ssfpack.htm>). Det er viktig å være oppmerksom på at disse programmene ikke er aktuelle der en sesongjusterer tidsserier i stor skala, men de er svært anvendelige hvis en vil foreta en mer nitidig dekomponering av enkelte tidsserier som synes spesielt viktige. Koopman foresto en gjennomgang av dette programmet under seminaret.

A. Maravall holdt i sesjon V A foredraget "Signal extraction in economic time series". Dette foredraget kan betraktes som en gjennomgang av det teoretiske grunnlaget som programpakken SEATS-TRAMO bygger på. I likhet med Harveys strukturelle tidsmodeller er dette også en modellbasert prosedyre.

Opplegget bygger på en estimerer en sesong ARIMA-modell (etter korleksjon for en del deterministiske effekter). Når ARIMA-modellen er estimert, kan en gjennomføre en dekomponering ved filtrering for å isolere de ulike tidsseriekomponentene. Ved dekomponering av tidsserier kommer en ikke utenom å gjøre identifiserende forutsetninger. En attraktiv egenskap ved Maravalls metode er at dette kommer eksplisitt fram. Det kan også i denne sammenheng nevnes at G. Tiao i det inviterte innledningsforedraget ("A historical review of seasonal adjustment and some perspectives") var inne på disse identifikasjonsproblemene. Han viste imidlertid ved konkrete eksempler at i mange tilfeller var ikke den sesongjusterte serien veldig følsom overfor de identifiserende forutsetninger. Når dette er tilfelle kunne en snakke om robust sesongjustering.

W. Härdle (Humbolt universitet, Berlin) gikk i sitt foredrag "Local models and non-parametric estimation" i sesjon I inn på hvordan ikke-parametriske metoder kunne utnyttes til sesongjusteringsformål.

I sesjon VIII B ble det presentert fem tema som drøftet forskjellige teoretiske aspekter i tidsseriedekomponering. Det var stor sett deltagere med en solid teoretisk bakgrunn som hadde fullt utbytte av både presentasjoner og diskusjoner rundt de forskjellige tema.

M. Riani (Institute of Statistics, Parma, Italia) presenterte et paper som viser hvordan den teoretiske autokorrelasjonen i sesongkorrigererte serier kan beregnes. Deretter brukte han metoden for å sammenligne dekomponering etter empiriske algoritmer (X11) med den som er modell basert (MB). Blant MB-alternativene valgte han strukturelle modeller i stedet for ARIMA modeller. Dermed mistet vi mulighetene for å koble resultatene til X12-ARIMA og TRAMO-SEATS som begge er basert på ARIMA modeller, ikke strukturmodeller.

R. Depoutot (Eurostat) og C. Planas (EC Joint Research Center, Ispra, Italia) presenterte to papers med titlene "Optimal truncation of linear filters" og "Admissible decomposition". Problemstillingen er hvordan revisjonen i tidligere beregnede sesongjusterte tall kan kobles til estimeringsmetoden. De viste hvordan valget av endelige filtre (som f.eks. X11) kan optimaliseres. Samtidig viste de hvordan uendelige filtre (modellbaserte metoder) kan gi tilsvarende resultater som endelige filtre. Resultatene ble koblet til problemstillingen om hvordan lengden for revisjonsperiode kan kontrolleres.

V. Gomez (Ministry of Economy and Finance, Madrid) presenterte et dokument under tittelen "Infinite versus finite moving average filters". Problemstillingen og konklusjonene var tilsvarende som beskrevet i forrige avsnitt.

R. Lacroix (Banque de France, Paris) drøftet hvordan man kan behandle de grunnleggende problemene som de to mest brukte metoder er konfrontert med: overjustering (X11-familien) og uheldige restriksjoner ved beregning av spektrum (modellbaserte metoder). Behandlingen av problemene baserer seg på nye prosedyrer innenfor spektralanalyse.

## 12.5 Sykel- og trenddefinisjoner og sesongjustering

Sesjonene III og VI A hadde som felles tittel: "Cycle-trend definition and seasonal adjustment".

H. Rudolph (Institute for Employment Research, Tyskland) presenterte problemer med sesongjustering for arbeidsledighet i Tyskland med månedsdata fra 1950 til 1997. På grunn av store endringer i sesongfaktorer kan man ikke finne en additiv eller multiplikativ modell for disse dataene. En metode har blitt utviklet i The German Federal Employment Services for å løse dette problemet.

I X11-ARIMA-metoden har 13-ledds Henderson-filteret vært brukt for å estimere trend. En ulempe med dette filteret er at det skaper mange små bølger i trenden etter en syklus på 9 eller 10 måneder, som man kan forveksle med vendepunkter. Estela Dagum (Statistics Canada) har modifisert filteret slik at antall små bølger som eksisterer i trenden blir redusert, uten at vendepunktene fjernes. En ny modell for trend ble presentert av M. Morry (Statistics Canada). Trenden inngår som en komponent for en modell skrevet på tilstandform og estimeres via Kalman filteret.

T. Jones (Office for National Statistics UK) presenterte en metode for å publisere trend. Trenden er plottet med en hel linje på samme måte som de sesongjusterte tallene, men i de siste observasjonene blir trenden plottet med prikker. Revisjonene for trend er også med i figuren.

## 12.6 Varians ved målinger og varians ved sesongjustering

Sesjon X hadde som tittel "Measurement variance and Seasonal adjustment variance".

Mange tidsserier er karakterisert ved at de er basert på utvalgsundersøkelser. Dette medfører at utvalgsfeil gjør seg gjeldende. Videre kan en ha et samplingsdesign basert på rotasjon slik at en del aktører er med i flere perioder når de først er kommet inn i undersøkelsen. Dette fører til at en får en seriekorrelasjon i utvalgsfeilen. D. Pfeffermann har gjennom flere år arbeidet med estimere variansen til sesongjusterte verdier fra software-programmer som X-11 og X11-ARIMA når en står overfor slike tidsserier, jf. blant annet Pfeffermann (1994) og Pfeffermann mfl. (1995). Variansformlene her er basert på analytiske formler der en benytter seg av at estimering av den sesongjusterte verdien fra programmene nevnt ovenfor tilnærmet kan oppfattes som at en bruker et lineært filter og at en har kjennskap til fordelingen av utvalgsfeilen over tid. På konferansen presenterte Pfeffermann paperet (skrevet sammen med R. Tiller) "Relationships between sampling designs and survey error models with application to variance estimation of seasonally adjusted and trend components". Dette paperet har mange fellestrekk med de som er refert til over, men poenget i

dette arbeidet er å utnytte parametrisk bootstrap for å estimere variansen til den sesongjusterte verdien og trenden (på ethvert tidspunkt).

D. Pfeffermann bidro med nok et paper (sammen med S. Scott, Bureau of Labour, Washington) som i sterk grad baserte seg på de to referansene gitt ovenfor og der metodene var anvendt på arbeidskraftsserier fra U.S. Current Population Survey (CPS). Pfeffermann presenterte metoden som en generell framgangsmåte for estimering av variansen til sesongjusterte serier fra X11-ARIMA. Metoden er basert på X-11 filteret.

W. Bell foresto det tredje bidraget på denne sesjonen med "Bayesian assessment of uncertainty in seasonal adjustment with sampling error present". I motsetning til de to bidragene omtalt over var dette en modell-basert tilnærming til beregning av usikkerhet i sesongjusterte verdier. Som tittelen over indikerer så foregår dette innenfor en bayesiansk ramme og med en eksplisitt modell for utvalgsfeilen.

## 12.7 Erfaringer med sesongjustering

Sesjon VII A hadde tittelen "Experiences in Seasonal adjustments". Denne sesjonen dekket mange ulike forhold, men hadde som fellesnevner at arbeidene var rettet inn mot anvendelser. Samlet var det åtte innlegg. (I seminarprogrammet står sesjonen oppført med sju innlegg, men Millionis' innlegg i sesjon VIII B ble overført hit.)

Det første paperet var ved D. Q. Pham og T. Skjerpen (Statistisk sentralbyrå) og hadde tittelen "Estimation of confidence intervals for seasonally adjusted and trend time series: A case study based on simulations". (Tittelen var noe revidert i forhold til hvordan den er gjengitt i seminarprogrammet). I dette paperet forsøker en å utnytte to simuleringsteknikker for å estimere usikkerheten knyttet til sesongjusterte og glattede tidsserier (dvs. trend). Den parametriske metoden er basert på at en resampler originale data ved hjelp av en enkel ARIMA modell og dekomponerer hver enkelt av disse syntetiske tidsseriene ved hjelp av X11-ARIMA. Den ikke-parametriske metoden er basert på at en resampler den irregulære komponenten som er oppnådd ved å dekomponere tidsserien ved hjelp av X11-ARIMA eller X12-ARIMA. Den sesongjusterte serien repliseres ved å kombinere "nye" irregulære serier med trendserien fra dekomponeringen av originalserien. Trendserien repliseres ved å bruke et Henderson-filter på nyde nye sesongjusterte verdiene. I motsetning til hva som er tilfellet med den parametriske metoden trenger en her kun å kjøre X11-ARIMA eller X12-ARIMA to ganger (en gang for å lage fremskrivninger for originalserien og en gang for å dekomponere den utvidede originalserien). Metoden ble anvendt på (månedsserien) antall første-gangsregistrerte personbiler. I etterkant av sesjonen kom Professor D. Pfeffermann (Hebrew University, Jerusalem, Israel) med noen kritiske merknader der han stilte spørsmål ved de to metodene referert til over. Etter hjemkomst til Norge har det blitt sendt en mail til Pfeffermann der en har anmodet om en skriftlig beskrivelse av innvendingene.

L. Jones og R. Penny (Statistics New Zealand) presenterte paperet "User Driven Seasonal Adjustment". Statistics New Zealand sesongjusterer rundt 1500 serier på regulær basis. Arbeidet foregår desentralisert, men en sentral ekspertgruppe foretar en årlig gjennomgang av prosedyrene. Den sentrale gruppen arbeider dessuten med å ta i bruk nyvinninger innenfor feltet og samarbeider også med sesongjusterings-

eksperter ved Victoria University Wellington. For tre år siden gikk Statistics New Zealand over fra et produksjonssystem der de kjørte X11-ARIMA/80 på en stormaskin over til et system basert på PC X12-ARIMA. Statistics New Zealand har selv laget et system som gjør det enkelt å hente ut informasjon knyttet til dekomponerte tidsserier (jf. MyFAME i SSB). Dette programmet er under stadig utvikling. Bidragsyterne var opptatt av hvordan en skal sesongjustere aggregatserier. Et spørsmål de er opptatt er hvordan en skal behandle et aggregat av serier der kun noen av seriene utviser sesongvariasjon eller der noen av delseriene utviser sesongvariasjon i bare deler av sampeelperioden.

J.-S. Méssonier (Banque de France, Paris) la fram paperet "TRAMO SEATS: An Application To Seasonal Adjustment and Short-Term Forecasting of French Current Account Series". Programmet TRAMO-SEATS benyttes for å sesongjustere 80 tidsserier fra det franske driftsregnskapet overfor utlandet. Méssonier var inne på problemer knyttet til aggregering av sesongjusterte serier og til stabilitet av sesongkomponentene over tid. Det ble foretatt en sammenligning mellom X11-ARIMA og TRAMO-SEATS. Til slutt ble det gitt noen kommentarer knyttet til prediksjoner av tidsserier basert på TRAMO-SEATS.

R. Karaliunas (Nasjonalbanken i Litauen) bidro med paperet "The Daily Forecasting of the Currency Issued". Bakgrunnen for Karaliunas' modellering var knyttet til nasjonalbankens behov for prediksjoner av utviklingen i pengemengden ( $M_0$ ) når de gjennomfører sine markedsoperasjoner. En modell estimert på dagsdata (business days) ble presentert. Til grunn for estimeringen av modellen lå 979 observasjoner som strakte seg fra 1. april 1994 til 31. desember 1997. Karaliunas' modell forklarte 91pengemengden fra dag til dag.

P. Revilla og P. Ray (Instituto Nacional de Estadística, Madrid, Spania) bidro med paperet "Working Day and Seasonal Adjustment in the Spanish Index of Industrial Production". I disse månedsseriene er det tilstede betydelige kalendereffekter. En korrigering for disse kalendereffektene er nødvendig for å fortolke den kortsiktige utviklingen i disse seriene. Flere forhold bidrar til å gi kalendervirkninger. Det at månedene har ulik lengde og ulik sammensetning på de enkelte ukedager er av betydning. En annen kalendereffekt henger sammen med hvordan påsken er plassert mellom mars og april. For Spania kommer i tillegg det inn at en del offentlige helligdager varierer fra et år til et annet og dessuten fra en region til en annen. Disse kalendereffektene tas hensyn til innenfor en ARIMA- modell der disse kalendereffektene er representert ved hjelp av deterministiske variable (jf. hvordan dette tilsvarende behandles innenfor X12-ARIMA og TRAMO-SEATS). De deterministiske effektene håndteres ved hjelp av to variable. Den ene variabelen skal fange opp antall virkedager, dvs. en har eliminert innflytelsen av lørdager, søndager og helligdager. Dessuten er det gjennomført en vekting for å ivareta regional heterogenitet. Den andre variabelen representerer påskeeffekten. I tillegg tas det hensyn til en intervensjonseffekt som følge av en transportstreik i februar 1997. For å virkedags- og sesongkorrigere dekomponeres de estimerte kalendereffektene i en kalendereffekt og i en effekt som ikke er knyttet til sesong.

R. R. Bar-On, L. Burck og E. Salama (Central Bureau of Statistics, Israel) bidro med paperet "The Analysis of Seasonality, Micro-Seasonality and Short-Term Trends in Travel and Tourism". Turisme er i mange land (herunder Israel) en viktig næring som dessuten er i sterk vekst. De tre bidragsyterne har i stor skala betraktet tidsserier

knyttet til reiser og turisme. Betydningen av å ta hensyn til effekter knyttet til religiøse og kulturelle høytider og virkedager ble understreket. På samme måte er det viktig å ta hensyn til effekten av spesielle hendinger (f. eks. effekten av Iraks invasjon i Kuwait i august 1990). Spesielt interessante var figurene som ble presentert.

S. Compton (Office for National Statistics, United Kingdom) holdt foredraget "Training Staff in Seasonal Adjustment Methods". Dette paperet har betydelig relevans for alle institusjoner som sesongjusterer og glatter tidsserier i stor skala (herunder SSB). Det finnes to måter å organisere sesongjusteringsarbeidet på innenfor institusjoner som sesongjusterer tidsserier i stor skala. Den ene måten er et sentralisert opplegg der en ekspertgruppe har dette som sin spesifikke oppgave. Alternativt kan sesongjusteringen foregå desentralisert slik at de samme personer som lager de ujusterte tallene også har ansvaret for sesongjusteringen. Dette kan eventuelt suppleres med en stab av eksperter som kan komme inn å løse konkrete problemer og dessuten ha ansvaret for opplæring (Den siste organisasjonsformen ligger tettest opptil hvordan virksomheten er organisert i SSB). Compton understreket at en ofte er i den situasjon at de som sesongjusterer i praksis ikke er eksperter innenfor analyse av tidsserier. Dette vil kunne medføre at det å ta i bruk nye og mer sofistikerte softwareprogrammer ikke nødvendigvis vil bedre "kvaliteten" på dekomponeringen. Compton uttrykte i samband med dette noe bekymring for de nye dekomponeringsprogrammene X12-ARIMA og TRAMO-SEATS. Han anbefalte i samsvar med dette at det var nødvendig å arbeide ytterligere med brukergrensesnittene for disse to programmene.

A. Milionis presenterte paperet "The Importance of Stationarity in Variance in applied time series modelling. An example analysis". Milionis gjorde seg til talsmann for at det ofte i anvendt arbeid var nødvendig å teste for hvorvidt en har ikke-stasjonaritet i variansen. Hvis en slik ikke-stasjonaritet er til stede og en ikke tar hensyn til den vil en ende opp med en misspesifisert modell. Milionis betrakter et eksempel knyttet til rentemodellering og finner at dersom en tar hensyn til ikke-stasjonaritet i variansen kan en komme fram til en bedre modell med langt færre parametre enn i modellen karakterisert med stasjonaritet i variansen.

## 12.8 Hva er en god sesongjusteringsmetode?

Sesjon IX hadde tittelen "What is a good seasonal adjustment method?". Det var to presentasjoner på denne sesjonen. For begge presentasjonene gjaldt det at en forsøkte å se sesongjusteringen fra en praktisk synsvinkel, der ikke minst brukernes ønsker var satt i høysetet.

Det ene foredraget var ved G. Laroque (INSEE, Frankrike) og hadde tittelen "Seasonal Adjustment: The point of view of economists and policy makers". Laroque var spesielt velegnet til å holde et slikt foredrag siden han både har akademisk erfaring (fra tidligere av) og i sin nåværende rolle har en stilling der han er i løpende kontakt med brukerne av statistikken. Sesongjustering og trendberegning skjer tradisjonelt på en univariat basis. Det vil si at en ser på en og en serie og dekomponerer denne i ulike komponenter. Fra mange økonomers synsvinkel vil det, ifølge Laroque, være mer naturlig å ha et multivariat utgangspunkt der en erkjenner at de ujusterte data har blitt generert ved at sesongmessige sjokk blandes med atferden til økonomiske aktører. En naturlig modellramme vil her være multivariate modeller der en tar hensyn til forekomsten av enhetsrøtter og kointegrasjon ved sesongfrekvensene.

I samband med politikeres behov understreket Laroque sterkt behovet for pålitelige trendestimater og gikk spesielt inn for at en ikke burde revidere sesongjusterte serier og trendserier etter 3 til 4 år.

B. Feldmann og G. L. Mazzi (Eurostat) holdt foredraget "Is an adjustment policy for Eurostat possible", basert på det svært oversiktlige og lesverdige paperet "Eurostat Suggestions concerning Seasonal Adjustment Policy". Bakgrunnen er de standardiseringsbestrebelse som foregår innenfor Eurostat-området når det gjelder sesongjusterte (og glattede) tidsserier. En lang rekke personer utover de to nevnte har også deltatt i dette prosjektet. Paperet gjengir en rekke anbefalinger når det gjelder sesongjustering innenfor Eurostatområdet. I hele notatet understrekes brukernes behov spesielt og at en ved sesongjustering i praksis alltid er nødt til å avveie deres behov mot krav som følger av faglig forsvarlighet.

Eurostat anbefaler modellbasert sesongjustering (implementert ved programmet TRAMO -SEATS), men avviser heller ikke X12-ARIMA. Øvrige programmer ønsker de at medlemslandene etter hvert skal slutte å bruke. Feldmanns synspunkter mht. publisering var at både sesongjusterte serier og trendserier bør gjøres tilgjengelige for brukerne. Figurer bør først og fremst vise trendserier. Sesongjusterte serier kan også gis i figurer dersom det anses forsvarlig ut fra tidsserienes art og relevant for vedkommende publikasjon.

Siden hva som kommer ut av sesongjusteringen i betydelig grad er avhengig av metode, vil et skifte av sesongjusteringsmetode måtte innebære revisjoner. Eurostat anbefaler i overensstemmelse med dette at endringer i sesongjusteringsmetode bare gjennomføres når det foreligger et sterkt faglig belegg for dette.

Eurostat vektlegger at selve sesongjusteringsprosessen bør være så gjennomslutlig som mulig. I samsvar med dette anbefaler de at det publiseres metainformasjon knyttet til sesongjusteringen av tidsseriene. Metainformasjonen bør blant annet inneholde opplysninger knyttet til sesongjusteringsmetode, hva som er forutsatt i samband med aggregering av tidsserier, hvordan en har korrigert for kalendereffekter og hvordan eventuelle spesielle hendelser (som for eksempel streiker) har blitt behandlet.

Eurostat publiserer både sesongjusterte tall fra Eurostat som helhet, men også fra de enkelte medlemsstatene. Et spørsmål er om dette bør gjøres sentralt i Eurostat eller i det enkelte medlemsland. På grunn av en del informasjon som det enkelte land vil ha bedre tilgang til går Eurostat inn for å opprettholde et desentralisert opplegg, men vektlegger samtidig at samarbeid og harmonisering er nødvendig for å øke informasjonsinnholdet i sesongjusterte serier. Ikke minst er dette viktig når en ønsker å sammenligne den kortsiktige utviklingen i de forskjellige landene innenfor Eurostat.

Når det gjelder sesongjusterte serier er det ofte aggregerte tidsserier som blir viet mest oppmerksomhet. Ved serier som er dannet ved å aggregere delserier, f. eks. over næringer eller varegrupper, kan en statistikkprodusent stå overfor valget mellom å sesongjustere hver av delseriene og summere opp til en sesongjustert totalserie (indirekte sesongjustering) eller å sesongjustere direkte på totalserien (direkte sesongjustering). I en del (teoretiske) tilfeller vil metodene være sammenfallende, men i praksis blir det ofte en forskjell. Feldmann gikk inn for den direkte metoden, men hadde forståelse for at den indirekte metoden blir brukt i samband med sesongjustering innenfor nasjonalregnskapssystemer (siden den indirekte metoden fører til at

økosirkidentitetene opprettholdes).

Et annet viktig aggregeringsspørsmål er knyttet til aggregering over tid. Når en sesongjusterer tidsserier vil summen av de sesongjusterte tidsseriene over kalenderåret kunne avvike fra det tilsvarende ujusterte årstallet. En får således et konsistensproblem. En praksis som er i bruk er å gjennomføre en tilleggsjustering som imøtekommer dette konsistenskravet. Det foreligger imidlertid ingen vitenskaplig begrunnelse for å pålegge dette kravet. Blant annet er endringer i sesongmønsteret og forekomsten av kalendereffekter forhold som peker i retning av å utelate konsistensrestriksjonen. Brukerne har ofte en forventning om at de sesongjusterte tallene skal oppfylle konsistenskravet. Feldmann mener at på kort sikt må en gi etter for dette kravet, men på lengre sikt er det viktig å "oppdra" brukerne til å leve med tidsinkonsistensen.

Fra et brukersynspunkt er det vanlig å betrakte revisjoner i sesongjusterte og glattede serier som et problem. Slike revisjoner er spesielt knyttet til den siste delen av tidsserien (som for brukere av korttidsstatistikk også er den mest interessante). Når ny informasjon kommer til, vil en revidere anslaget for sesongjusterte og glattede tiddserier. Disse revisjonene kan ofte for en del tidsserier være av et betydelig omfang. Dersom sesongmønsteret utviser stor stabilitet, vil revisjonene først og fremst gjøre seg gjeldende i trendserien. På den annen side gir revisjonene uttrykk for at en har fått genuin ny informasjon. En står således overfor en avveining der en må balansere ønsket om stabilitet i den informasjon som publiseres, mot på den andre side ønsket om å forbedre anslag for uobserverbare komponenter når ny informasjon om den ujusterte serien tilkommer.

Av betydning for revisjoner er den praksis en har i samband med prediksjoner av tids-serier. Eurostat synes å ta avstand fra løpende sesongjustering (i den engelskspråklige litteraturen omtalt som "concurrent seasonal adjustment"). Begrepet løpende sesongjustering kan illustreres ved å ta utgangspunkt i X11-ARIMA. I dette programmet finnes det 5 innebygde ARIMA modeller som kan brukes til fremskriving av tids-serier. Den vanlige måten å bruke programmet på er å la programmet selv finne frem til en modell. En starter ut med den enkleste modellen og velger denne til prediksjon hvis den oppfyller de på forhånd oppsatte betingelser. Dersom betingelsene ikke er oppfylt går en videre til en noe mer generell modell og gjennomfører en tilsvarende test. I praksis kan det derfor skje at en underveis i kalenderåret bytter fremskrivingsmodell. Dette vil i så fall forårsake revisjoner. Eurostat går følgelig inn for å holde fast på den samme modellen gjennom kalenderåret. Gitt at en holder fast på samme fremskrivingsmodell må en dernest ta stilling til om parameterene i den skal reestimeres hver gang en ny observasjon tilkommer. Eurostat går inn for å fryse parameterene gjennom kalenderåret slik at modellen bare estimeres en gang i året. Disse tiltakene vil redusere betydningen av revisjoner gjennom året.

Betydningen av usikkerhetsmål for beregnede tidsserier ble også understreket. Siden Eurostat fortrinnsvis går inn for en modellbasert sesongjusteringsmetode (jf. igjen at de går inn for TRAMO SEATS) er det kjent hvordan en skal gå frem for å beregne estimeringsusikkerheten til de uobserverbare komponentene. Selv om en også kan ha andre usikkerhetsbidrag, vil det være av interesse å få meddelt den genuine estimeringsusikkerheten.

## 12.9 Sammenligning av metoder - Metodologi

Sesjon II hadde tittelen "Comparison of methods - Methodology".

D. Findley presenterte metodene for diagnostikk basert på "sliding spans" i X12-ARIMA.

Formålet med "sliding spans" metoden (utviklet av D. F. Findley et al. (1990)) er å beregne revisjoner for sesongkomponenten og sesongjustert serie i et bestemt tidspunkt når vi gjentar sesongjusteringsprosedyren på opp til fire forskjellige delvis overlappende delmengder av rådata (såkalte "spans" på engelsk). Hvert "span" er å betrakte som en komplett tidsserie og behandles akkurat som de originale data, det vil si at ukedagseffekter, påskeeffekt, intervensjoner prekorrigeres fra data i intervallet, før sesongkomponent og sesongjustert serie estimeres. For eksempel, sesongjusteringen utføres fire ganger i intervallene [jan.74, des.81], [jan.75, des.82], [jan.76, des.83] og [jan.77, des.84]. Sesongkomponentene og de sesongjusterte tallene for hvert tidspunkt i fire "spans" registreres for å beregne revisjoner.

### *Sliding spans og størrelsen Q i X11-ARIMA.*

For å måle kvaliteten av en sesongjustering har Statistics Canada innført 11 teststørrelser M1-M11 (Lothian og Morry 1978). Disse målestørrelsene har verdi fra 0 til 3, lav verdi indikerer en god kvalitet med 1 som grense for testen. Et sammenveid mål Q er en kombinasjon av de 11 målestørrelsene. Kriteriet for en god dekomponering er at Q er mindre enn 1. I X12-ARIMA er sliding spans metoden en måte å vurdere kvaliteten av sesongjusteringen på. Findley og Monsell har vist at sliding spans metoden i X12-ARIMA gir bedre evaluering av kvaliteten enn Q-størrelsen i X11-ARIMA.

### *Sliding spans og X-11 ukedagseffekter.*

Ukedagseffekter i X11-ARIMA estimeres ved regresjonsanalyse, hvor den foreløpige irregulære komponenten brukes som input. Ofte er elementene i denne serien sterkt korrelerte. Derfor blir også residualer fra regresjonsanalyse korrelerte med hverandre. F-testen som er basert på residualene, er ikke gyldig lenger. Sliding spans metoden har vist at prekorrigering for ukedagseffekter fra X11-ARIMA kan føre til store endringer fra måned til måned og fra år til år for sesongjusterte tall. Spektralsanalyse er et godt verktøy for å oppdage effekter av ukedager.

### *Direkte eller indirekte sesongjustering.*

En tidsserie kan være summen av mange tidsserier i undergrupper. Sesongjusterte tall for den aggregerte serien kan estimeres på to forskjellige måter som vi kaller direkte og indirekte sesongjustering. I den første metoden sesongjusterer man den aggregerte serien, mens ved den andre metoden summerer man for hver delperiode sesongjusterte tall av seriene i undergruppa. Ofte får man ikke samme resultater. Findley foreslår at vi kan bruke sliding spans for å løse dette problemet, ved å velge metoden som gir minst revisjoner i den aggregerte, sesongjusterte serien.

### *Revisjon av sesongjusterte tall*

Det er et fortrinn ved X-12-ARIMA at det i metoden kan beregnes revisjon av sesongjusterte tall for en bestemt periode.

B. Fischer (Central Bank, Frankfurt) og C. Planas' bidrag ga en oppsummering av erfaringer med bruk av univariate ARIMA-modeller med eksogene forklaringsvari-

able når det gjelder å gi en adekvat dekomponering ved sesongjustering av et stort antall økonomiske tidsserier. Fischer og Planas undersøkte mer enn 13 000 av Eurostats månedsserier. Undersøkelsen viste at det er mulig å kvantifisere virkningen av kalendereffekter på forskjellige typer av økonomiske indikatorserier og at det er mulig å få ut detaljert beskrivelse av forekomster og typer av utliggere (ekstremobservasjoner). Videre at det er mulig å vurdere den praktiske relevansen av flere mulige modellspesifikasjoner, samt å få ut detaljert informasjon om hva som er årsaken til manglende modelltilpasning (feilspesifisering), for eksempel ikke-lineære mønstre

B. Fischer (Central Bank, Frankfurt) og C. Planas' bidrag ga en oppsummering av erfaringer med bruk av univariate ARIMA-modeller med eksogene forklaringsvariable når det gjelder å gi en adekvat dekomponering ved sesongjustering av et stort antall økonomiske tidsserier. Fischer og Planas undersøkte mer enn 13 000 av Eurostats månedsserier. Undersøkelsen viste at det er mulig å kvantifisere virkningen av kalendereffekter på forskjellige typer av økonomiske indikatorserier og at det er mulig å få ut detaljert beskrivelse av forekomster og typer av utliggere (ekstremobservasjoner). Videre at det er mulig å vurdere den praktiske relevansen av flere mulige modellspesifikasjoner, samt å få ut detaljert informasjon om hva som er årsaken til manglende modelltilpasning (feilspesifisering), for eksempel ikke-lineære mønstre

Fischer/Planas advarte likevel mot en mekanisk bruk av sesongjusteringsverktøyene, fordi en tilsynelatende god modelltilpasning ikke er noen garanti for kvaliteten av sesongjusteringen. En annen grunn til å advare mot automatisert bruk av et verktøy er at det kreves fagkunnskap på statistikkområdet som råserien representerer for å kunne teste plausibiliteten av transformasjoner, brudd i seriene, regresjonsparametere og årsaken til ekstremobservasjoner.

Bidraget fra R. Depoutot i samarbeid med C. Planas tok i utgangspunktet sikte på å finne en praktisk metode for å sammenligne filtrene i X11 med X12-ARIMA og TRAMO-SEATS filtre ved sesongjustering i stor skala. Dette ble presisert til "Endelige approksimasjoner til lineære filtre og overvåking av revisjoner". Bidraget gir forslag til en fremgangsmåte til å erstatte uendelige med optimale, endelige filtre for å kunne kontrollere lengden på revisjonsperioden. Paperet gir dessuten oversikt over revisjonsanalyse i teori og praksis.

## 12.10 Sammenligning av metoder - Resultater

Sesjonene IV og VII B hadde som felles tittel "Comparison of methods - Results". Her gjengis hovedkonklusjonene i foredragene som gjaldt sammenligninger av TRAMO-SEATS og X12-ARIMA

I sesjon IV ble det presentert fire forskjellige empiriske arbeider for å sammenligne resultater fra TRAMO-SEATS og X12-ARIMA. Følgende momenter ble fokusert for å vurdere kvaliteten på resultatene: brukervennlighet, revisjoner, direkte vs. indirekte justering og behandlingen av ekstreme verdier.

C. Hood, B.C. Monsell og D.F. Findley (U.S. Bureau of the Census) presenterte resultater for 11 serier fra Italia. De var primært interessert i å se egenskapene for sesong- og trend- komponentene fra TRAMO-SEATS og sammenligne dem med de i X12-ARIMA. Konklusjonene kan oppsummeres på følgende måte:

- Hovedresultatene fra de to metodene var ganske like

- X12-ARIMA ga mindre revisjoner i seks av de estimerte seriene, TRAMO-SEATS i tre av seriene
- Fordeler med TRAMO-SEATS var relatert til valg og behandling av ARIMA modell.
- Fordeler med X12-ARIMA var relatert til behandling av data, klarere feilmeldinger, flere opsjoner, mer diagnostikk.

J. Dossé (World Systems, Luxembourg) og C. Planas (Ispira, Italia) presenterte resultater fra et arbeid orientert mot å undersøke revisjonsproblematikken ved bruk av de to ovennevnte metodene. De fokuserte sine resultater på revisjonen i de siste observasjonene i seriene. Flere verktøy for å vurdere omfanget på revisjoner ble presentert. De konkluderte med:

- Ingen av de to metodene gir mindre revisjoner for alle serier.
- Modell basert metode (TRAMO-SEATS) gir generelt mindre revisjoner enn bruk av empiriske filter (X12-ARIMA).
- X12-ARIMA gir større revisjoner
- TRAMO-SEATS gir glattere trendkomponenter
- X12-ARIMA konvergerer raskere mot endelige estimer, men dette skjer på bekostning av større revisjoner.

Arbeidet fra F. Bacchini, A. Pallara og G. Bruno (ISTAT, Rome, Italia) fokuserte på hvordan de to metodene behandler aggregeringsproblematikk. Seriene som ble testet var begrenset til produksjonsindeks-området. Hovedkonklusjoner:

- Valg av kun en type serie gjør at resultatene bør tolkes forsiktig
- Begge metoder gir bedre resultater når man velger direkte i stedet for indirekte justeringer (man justerer direkte på aggregater i stedet for å justere komponentene).
- TRAMO-SEATS viser seg å være mer robust både for å fjerne virkedagseffekter og å være konsekvent mellom direkte eller indirekte justering

S. Öhlen (Statistiska Centralbyrån, Sverige) presenterte et arbeid med å teste hvordan de to metodene fungerer på serier som er preget av forskjellige typer av ekstreme verdier. Han brukte simuleringsrutiner og testobservatorer for å vurdere resultatene. Han konkluderte med:

- Begge metoder er klart bedre enn X11ARIMA.
- X12-ARIMA og TRAMO-SEATS gir like gode resultater.

I sesjon VII B la D. Focarelli og R. Cristadoro (Bank of Italy, Roma) fram resultatene som ble oppnådd når en sammenlignet revisjoner i 11 tidsserier ved hjelp av henholdsvis X12-ARIMA og TRAMO-SEATS. De fant ikke store forskjeller i resultatene, når en hadde gjort like valg knyttet til prekorrigering. De understreker betydningen av den sofistikerte prekorrigeringen i begge.

Det er ønskelig å minimere revisjonsfeil. Det er spesielt ønskelig og viktig å unngå revisjonsfeil i siste delen av en tidsserie, fordi disse er de mest relevante mht. å treffe f.eks. makropolitiske beslutninger.

Det finnes til nå ingen teoretisk metode som kan gjennomføres i praksis når det gjelder å sammenlikne relevans av revisjoner ved ulike sesongjusteringsmetoder. G. Savio og G. Barbieri (Statistical Office, Roma, Italia) sammenligner X12-ARIMA og TRAMO-SEATS ved kjøring på 30 italienske, økonomiske kvartalsserier for 1970-1996.

Sammenligningen ble utført ved at de siste tre år ble fjernet fra hver av seriene. Deretter ble 12 simuleringer foretatt, ved hver simulering ble en observasjon føyd til den trunkerte serien. Det ble beregnet testobservatorer for revisjonene (gjennomsnittlig kvadrerte revisjoner, glatthet og konvergens av revisjonene) under to forskjellige hypoteser: henholdsvis fri modell, frie parametere og fast modell, frie parametere.

Savio/Barbieri fant følgende resultater:

- Ingen av de to programmene er klart bedre enn det andre
- Graden av revisjoner er typisk for hver enkelt serie
- Et ikke ubetydelig antall av seriene viste revisjoner av tidligere verdier ved tillegg av nye observasjoner etter to år
- For de fleste seriene viste TRAMO-SEATS større stabilitet i siste del av seriene, mens stabiliteten ved X12-ARIMA i den første delen av seriene samsvarer godt med praksis hos mange statistikkinstitusjoner som "fryser" verdiene i den første delen av seriene (ikke publiserer revisjonene der)

## 12.11 Presentasjon av applikasjoner og programmer

Dette avsnittet er et referat av våre inntrykk fra sesjonene V B og VI B med fellestitelen "Presentation of packages" og sesjon VIII A som gjaldt direkte demonstrasjoner av oppkoblede programvarepakker med anledning til å teste programmene på data.

I sesjon V B ble det presentert en programvare hvor SAS/GRAPH kan nyttes til å lage figurer for X12-ARIMA. Programvaren er tilgjengelig via Internett i to versjoner, en skjermversjon for å analysere sesongjusteringen ved en enkelt serie og en batch-versjon for å lage diagnostikk-grafer for mange serier på en gang. (Denne applikasjonen er ikke aktuell for SSB, fordi Dinh Quang Pham har laget en egen applikasjon som med langt mindre arbeid kan gi grafer med lik skala for et stort antall serier. Dessuten har vi et nyttig hjelpemiddel i MyFame)

Både TRAMO-SEATS og X12-ARIMA metodene ble opprinnelig utviklet i et DOS-miljø. K. Attal (INSEE, Paris) presenterte sammen med D. Ladiray (Statistics Canada) en applikasjon hvor TRAMO-SEATS og X12-ARIMA kan opereres direkte fra SAS, uten at det er nødvendig å importere data inn og ut fra SAS. Analysefunksjoner i SAS, bl.a. grafikk, kan nyttes direkte.

I sesjon VI B ble det presentert tre programvarer for å behandle bl.a. tidsseriedekomponering:

### *The XploRe environment*

Dette er et omfattende program som i utgangspunktet skulle konkurrere mot SAS og tilsvarende programmer. XploRe er definert som "The interactive statistical computing environment". W. Hrdle (Humbolt University, Berlin), som har utviklet programmet, ga en beskrivelse. XploRe er beregnet for både statistisk analyse, forskning og opplæring. Det består av flere moduler og gir store muligheter for grafisk analyse. Spesielt interessant er at brukerne i stedet for å betale lisens skal betale tidsbruk på tilsvarende måte som dagens telefontjenester.

### *STAMP 6 - SAM (Seasonal Adjustment Module) PROGRAM*

J.S. Koopman (Tilburg University, Nederland) introduserte denne modulen for sesong-korrigerende som er integrert i GiveWin og STAMP-6 miljø. Presentasjonen viste at dette windows-baserte verktøyet er veldig lett å bruke. Programmet genererer mange figurer og analysetabeller. Det gir også store muligheter både for å fremskrive og for å ta hensyn til intervensjoner.

### *DEMETRA*

Eurostat har integrert både TRAMO-SEATS og X12-ARIMA i et enkel grensesnitt kalt DEMETRA. Hovedformålet med dette verktøyet er at det kan anvendes av både erfarne og nye brukere. Det er mulig å kjøre tusenvis av serier i ulike formater som FAME, ASCII eller EXCEL automatisk. En av fire moduler gir muligheter for å sammenligne resultater fra forskjellige modellspesifikasjoner. Samtidig har programmet opsjoner for å sammenligne resultater fra TRAMO-SEATS og X12-ARIMA. DEMETRA er helt windows-basert. Den gir muligheter for å inkludere alle opsjonene som er i disse to programmene.

Hvis det er slik at dette verktøyet fungerer som EUROSTAT påstår, er dette noe som vi i SSB bør vurdere grundig. (I følge Jørgen Ouren finnes det fortsatt ingen UNIX-versjon av DEMETRA.)

I løpet av ca. to timer var det (i sesjon VIII A) muligheter for å se forskjellige demonstrasjoner av programvarene som er beskrevet i de foregående avsnitt. Det var ganske uheldig at pga. tekniske problemer var det ikke mulig å koble seg til Internett. Som en konsekvens av dette var ikke noen demonstrasjon av DEMETRA.

Deltagerne konsentrerte sin oppmerksomhet om TRAMO-SEATS (A. Maravall) og X12-ARIMA (D. Findley, B.C. Monsell). Det var spesielt interessant å se hvordan Findley valgte optimale opsjoner og hvordan han brukte diagnostikken i X12-ARIMA for å fjerne påskeeffekter fra SSBs produksjonindekser.

## Kapittel 13

# Ukedagseffekter og sesongjusterte tall for detaljvolumindeksen: Et nærmere blikk på desember 2000

Dinh Quang Pham

### 13.1 Innledning

Detaljvolumindeksen er kraftig påvirket av ukedagseffekter som skyldes ulikt antall ukedager i hver måned. Siden antall mandager, tirsdager, ... , søndager varierer fra måned til måned vil, f. eks. en klesbutikk selge mer i en måned med 5 lørdager enn en måned med 4 lørdager, selv om de to månedene har samme lengden, for eksempel, 31 dager. Salget kan også øke for en måned med 5 torsdager. Vi prekorrigerer ukedagseffekter fra rådataene for at de sesongjusterte tallene ikke skal påvirkes av antall ukedager som forekommer i en måned. Vi handler mer enn vanlig i ukene før jul. Dermed er tallene i desember mye høyere enn de andre månedene. Dette gjentar seg hvert år og på samme tidspunkt. Tabell 13.1 viser de observerte verdiene  $O_t$  for år 2000. Vi får  $O_{nov.} = 117,44$  og  $O_{des.} = 158,67$ . Endringen i prosent er  $100\% \times (158,67 - 117,44)/117,44 = 35,10\%$ . Tallet kan tolkes som at aktiviteten målt ved salget i desember er 35% høyere enn i november.

Tabell 13.1: *Rådataene for detaljvolumindeksen i 2000*

| jan.   | feb.   | mar.   | apr.   | mai.   | jun.   |
|--------|--------|--------|--------|--------|--------|
| 104.81 | 102.06 | 110.28 | 105.86 | 119.94 | 119.52 |
| jul.   | aug.   | sep.   | okt.   | nov.   | des.   |
| 115.88 | 124.26 | 114.46 | 113.18 | 117.44 | 158.67 |

Tabell 13.2: *Antall ukedager for november og desember 2000*

|           | man. | tir. | ons. | tor. | fre. | lør. | søn. | sum      |
|-----------|------|------|------|------|------|------|------|----------|
| november  | 4    | 4    | 5    | 5    | 4    | 4    | 4    | 30 dager |
| desember  | 4    | 4    | 4    | 4    | 5    | 5    | 5    | 31 dager |
| forskjell | 0    | 0    | -1   | -1   | 1    | 1    | 1    | 1 dag    |

Tabell 13.2 viser antall ukedager for november og desember 2000. Det er flere fredager, lørdager og søndager, mens det er færre onsdager og torsdager i desember enn i november 2000. Hvis effekten for hver ukedag er den samme for hele tidsserien, viser tabell 13.2 at forskjellen av aktiviteten bestemt av ukedagene i disse to månedene er mye mindre enn 35% som er den faktiske endringen fra november til desember 2000. Det er julehandelen som bidrar til en høy verdi i desember. Vi må skille mellom effekten som skyldes ulike antall dager i en måned og den fra sesongvariasjonen som skyldes julehandelen.

Tabell 13.3: *Antall ukedager for desember i 1999 og 2000*

|           | man. | tir. | ons. | tor. | fre. | lør. | søn. |
|-----------|------|------|------|------|------|------|------|
| 1999      | 4    | 4    | 5    | 5    | 5    | 4    | 4    |
| 2000      | 4    | 4    | 4    | 4    | 5    | 5    | 5    |
| forskjell | 0    | 0    | -1   | -1   | 0    | 1    | 1    |

Tabell 13.3 viser antall ukedager i desember 1999 og 2000. Vi ser at det er flere onsdager og torsdager (største vekt), mens det er færre lørdager og søndager (minst vekt) i 1999 enn i 2000. Dessuten var julaften og nyttårsaften i 1999 på fredag og for år 2000 var det søndag. Det er to færre handledager i desember 2000 enn i 1999. Selve metoden for prekorrigering for antall ukedager tar imidlertid ikke hensyn til at juleaften og nyttårsaften faller på feks søndag. Vi får en høyere verdi for desember 1999 enn i 2000,  $O_{des.1999} = 161,88$  og  $O_{des.2000} = 158,55$ , men dette behøver ikke medføre at sesongjusterte tall i 1999 vil bli høyere enn tallet i 2000, siden sesongvariasjoner ofte endrer seg fra år til år. Det blir derfor vanskelig å forutsi om sesongjusterte tall ved å basere seg på rådataene.

I det neste avsnittet presenterer vi i detalj beregningen for ukedagseffekter og deretter resultater fra sesongjusteringen.

## 13.2 Beskrivelse av dataene

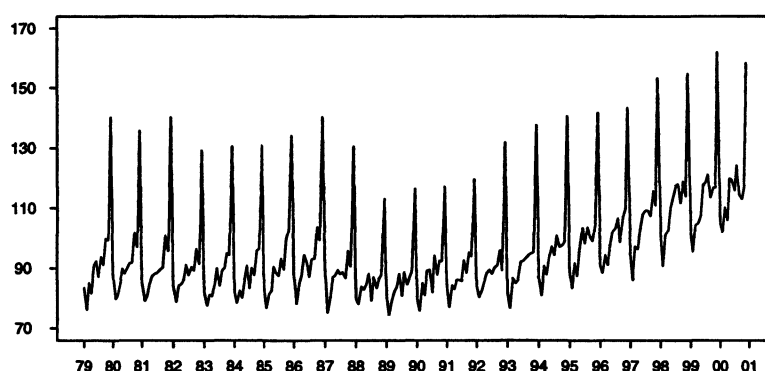
Merk at vi ikke bruker prosedyren i X-12-ARIMA for å korrigere mars og apriltall for påskeeffekten. Tallene blir justert med vår metode før vi sesongjusterer tidsserien. Vi betegner  $O_t$  for rådata i tidspunkt  $t$ .

Tidsserien er detaljhandel, unntatt motorkjøretøyer og motorsykler og reparasjon av husholdningsvarer, og varer til personlig bruk. Tallene er observert fra januar 1979 til desember 2000. I alt er det 22 år med 264 observasjoner. Figur 13.1 viser rådataene. Det er et klart sesongmønster. Figur 13.2 viser Box-whisker plottet for

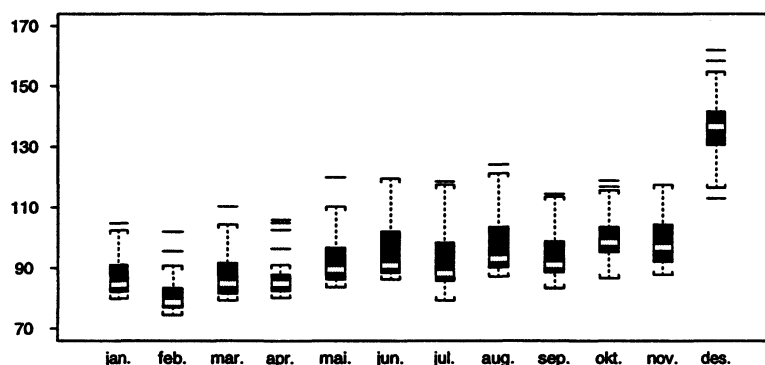
12 måneder. Deres gjennomsnitt og standardavvik er listet ut i tabell 13.4. Juli, august og desember er de månedene som har største spredninger, mens april har fleste ekstremverdier. Fordelingene for juni, juli, august og september blir forskjøvet i retning lave verdier.

### 13.3 Ukedagseffekter

Vi transformerer dataene med logaritmen i beregningen siden vi bruker en multiplikativ modell. Vi lar programmet selv velge en ARIMA modell og estimere intervensjoner. Resultatene er vist i tabell 13.5. Testen viser at dataene er påvirket av ulike antall ukedager i måneder med høy  $p$ -verdi ( $\chi^2 = 183,32$  med 6 frihetsgrader). Vi tolker tabell 13.5 slik for torsdagen og lørdagen er det høy omsetning i detaljhandelen, mens for søndagen er det svært lav (negativ vekt). Positive verdier i tabell 13.5 betyr større omsetning enn i gjennomsnitt, negativt vekt betyr mindre vekt enn i gjennomsnitt. Dette er selvsagt i gjennomsnitt over alle år og måneder vi har data for.



Figur 13.1: Rådata av detaljvolumindeksen



Figur 13.2: Box-whisker plot for 12 måneder

Vi beregner den korrigerte faktoren  $f_t$  for måned  $t$  ved å bruke vektene i tabell 13.5. For eksempel, for desember 1999. Antall mandager, tirsdager, onsdager, torsdager,

Tabell 13.4: Gjennomsnittene og standardavvikene for 12 måneder

|               | jan.  | feb.  | mar.  | apr.  | mai.  | jun.  |
|---------------|-------|-------|-------|-------|-------|-------|
| Gjennomsnitt  | 87.68 | 81.43 | 87.92 | 87.70 | 92.95 | 95.69 |
| Standardavvik | 7.44  | 7.07  | 8.52  | 7.76  | 9.56  | 10.31 |

|               | jul.  | aug.  | sep.  | okt.   | nov.  | des.   |
|---------------|-------|-------|-------|--------|-------|--------|
| Gjennomsnitt  | 93.15 | 97.71 | 94.69 | 100.48 | 99.37 | 136.52 |
| Standardavvik | 12.07 | 11.24 | 9.31  | 8.89   | 9.16  | 13.21  |

Tabell 13.5: Analysen fra regARIMA

|         | estimate | std.avvik | t-verdi | konklusjon       |
|---------|----------|-----------|---------|------------------|
| mandag  | -0.0038  | 0.0025    | -1.51   | ikke             |
| tirsdag | 0.0013   | 0.0024    | 0.54    | ikke             |
| onsdag  | 0.0027   | 0.0025    | 1.10    | ikke             |
| torsdag | 0.0092   | 0.0024    | 3.68    | sign. på 5% nivå |
| fredag  | 0.0026   | 0.0024    | 1.04    | ikke             |
| lørdag  | 0.0055   | 0.0025    | 2.16    | sign. på 5% nivå |
| søndag  | -0.0175  | 0.0025    | -6.86   | sign. på 5% nivå |

fredager, lørdager og søndager i denne måneden, er henholdsvis 4, 4, 5, 5, 5, 4, 4 dager.  $f_{des.99}$  blir

$$\begin{aligned}
 f_{des.99} &= 4 * (-0.0038) + 4 * 0.0013 + 5 * 0.0027 + 5 * 0.0092 + 5 * 0.0026 + \\
 &\quad 4 * 0.0055 + 4 * (-0.0175) \\
 &= 0.0145
 \end{aligned}$$

Vi beregner på samme måte for november og desember 2000.

$$\begin{aligned}
 f_{nov.00} &= 4 * (-0.0038) + 4 * 0.0013 + 5 * 0.0027 + 5 * 0.0092 + 4 * 0.0026 + \\
 &\quad 4 * 0.0055 + 4 * (-0.0175) \\
 &= 0.0119 \\
 f_{des.00} &= 4 * (-0.0038) + 4 * 0.0013 + 4 * 0.0027 + 4 * 0.0092 + 5 * 0.0026 + \\
 &\quad 5 * 0.0055 + 5 * (-0.0175) \\
 &= -0.0094
 \end{aligned}$$

Siden beregningen utføres med  $\log(O_t)$ , får vi de korrigerte verdiene for desember 1999, november og desember 2000 på logaritmisk skala

$$\begin{aligned}
 Y_{des.99}^K &= \log(161.88) - 0.0145 \\
 Y_{nov.00}^K &= \log(117.44) - 0.0119 \\
 Y_{des.00}^K &= \log(158.55) + 0.0094
 \end{aligned}$$

og i nivået av rådataene, er

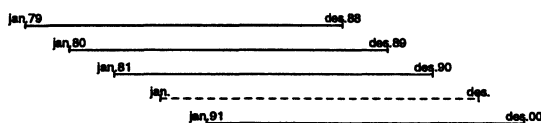
$$\begin{aligned} O_{des.99}^K &= \exp(\log(161.88) - 0.0145) = 159.55 \\ O_{nov.00}^K &= \exp(\log(117.44) - 0.0119) = 116.05 \\ O_{des.00}^K &= \exp(\log(158.55) + 0.0094) = 160.05 \end{aligned}$$

Vi ser at verdien for desember 2000 blir justert opp fra 158,55 til 160,05 etter korrigeringen for ukedagseffekter. Tallet (160,05) er ikke korrigert for at julen og nyttåret falt på en søndag, men inneholder fortsatt sesongvariasjoner. Tallet er høyt fordi folk handler inn til jul.

Vi ser i tabell 13.5 at torsdagen og lørdagen har større vekter enn de andre dagene, men det er ikke helt riktig å tolke dette som at "torsdagen og lørdagen er de to viktigste handledagene i uka", siden utsagnet tyder på at det er riktig uansett hvilken tidsserie som analyseres. Resultatene i tabell 13.5 er beregnet for data fra januar 1979 til desember 2000. En bedre måte å uttrykke dette er å si at ved å basere oss på data fra januar 1979 til desember 2000 er torsdagen og lørdagen to store handledager i uka. Dette beskrives ennå bedre i tabell 13.6.

### 13.3.1 Tester for endringer av ukedagseffektene med 10 år lange tidsserier

Fra rådataene  $O_t$ , hvor  $t$  går fra januar 1979 til desember 2000, lager vi nye tidsserier som er vist i figur 13.3. Den første inneholder dataene fra januar 1979 til desember



Figur 13.3: De nye tidsseriene med lengden 10 år

1988. Lengden er 10 år. Vi forskyver tidsserien et år i retning det siste tidspunktet for å lage den neste, osv. Den siste er fra januar 1991 til desember 2000. Vi lager ikke tidsserier som starter etter 1991, siden det er for få observasjoner for å sesongjustere. Tabell 13.6 viser de estimerte verdiene for vekter. Ved å plote tallene i samme figur (se figur 13.4) får vi et bilde av endringene til vektene for ukedager på grunnlag av 10 år lange tidsseriene. De gjennomsnittlige vektene  $\bar{v}_{1,}$  for (mandagen, tirsdagen, onsdagen) og  $\bar{v}_{2,}$  for (torsdagen, fredagen, lørdagen), for hver tidsserie i tabell 13.6 beregnes. For eksempel, for  $[jan.79 - des.88]$  er

$$\bar{v}_{1,79} = \frac{-0.0011 - 0.0036 + 0.0060}{3} = 0.0004$$

og

$$\bar{v}_{2,79} = \frac{0.0093 + 0.0007 + 0.0064}{3} = 0.0055$$

Disse størrelsene er plottet i figur 13.5. Fra tabell 13.6, figurene 13.4 og 13.5 vi ser at

1. Vektene for ukedager har endret seg med tid.
2. Torsdag er den viktige handledagen i uka fram til 1990.  $t$ -verdien er høyst for dataene i intervallet [*jan.87* – *des.96*]. Deretter gikk den ned for de senere tids-seriene. I de siste årene fra 1991 til 2000 er torsdag ikke en like stor handledag lenger ( $t$  verdien er lav, med 1.20).
3. Lørdag er også en viktig handledag, men i en mindre grad enn torsdag.  $t$ -verdien er høyst for tidsserien fra januar 1987 til desember 1996 og den blir lavere med dataene i de siste årene. Kurvene i figur 13.4 viser at vektene av torsdag og lørdag varierer i samme retning.
4. Fredag har en stor endring i de siste årene. Det viser at handelen blir flyttet fra torsdag og lørdag til fredag, slik at disse tre dagene er nå omtrent like.
5. Mandag, tirsdag og onsdag har lavere vekter. Det vil si disse dagene har mindre innvirkning på handelen i en måned. Handelen mandag-onsdag er omtrent som gjennomsnittet i resten av uka. Det er en svak endring fra onsdag til tirsdag.

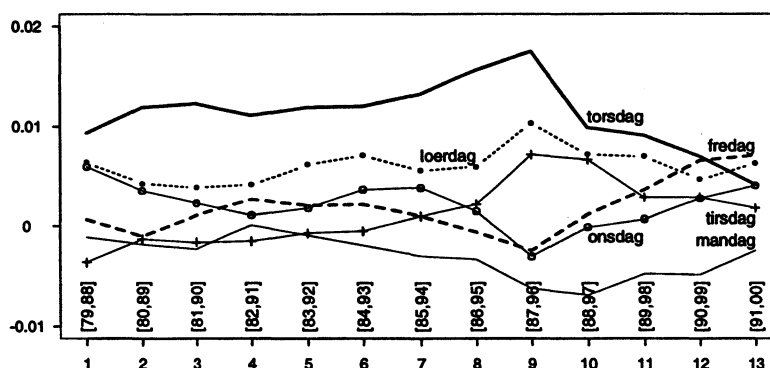
Tabell 13.6: *Ukedagseffekter*

| periode | mandag  |        | tirsdag |        | onsdag  |        | torsdag |        | fredag  |        | lørdag |        | søndag  |        |
|---------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|--------|--------|---------|--------|
|         | estim.  | t-ver. | estim.  | t-ver. | estim.  | t-ver. | estim.  | t-ver. | estim.  | t-ver. | estim. | t-ver. | estim.  | t-ver. |
| [79-88] | -0.0011 | -0.32  | -0.0036 | -1.02  | 0.0060  | 1.68   | 0.0093  | 2.62   | 0.0007  | 0.20   | 0.0064 | 1.74   | -0.0176 | -4.77  |
| [80-89] | -0.0018 | -0.49  | -0.0013 | -0.36  | 0.0036  | 0.96   | 0.0119  | 3.21   | -0.0010 | -0.26  | 0.0043 | 1.12   | -0.0157 | -4.06  |
| [81-90] | -0.0023 | -0.60  | -0.0016 | -0.43  | 0.0023  | 0.62   | 0.0123  | 3.18   | 0.0011  | 0.27   | 0.0039 | 0.99   | -0.0156 | -3.94  |
| [82-91] | 0.0001  | 0.04   | -0.0015 | -0.38  | 0.0011  | 0.29   | 0.0111  | 2.99   | 0.0027  | 0.72   | 0.0042 | 1.12   | -0.0177 | -4.73  |
| [83-92] | -0.0009 | -0.23  | -0.0007 | -0.19  | 0.0018  | 0.45   | 0.0119  | 2.90   | 0.0021  | 0.51   | 0.0062 | 1.55   | -0.0203 | -5.04  |
| [84-93] | -0.0019 | -0.50  | -0.0005 | -0.14  | 0.0037  | 0.96   | 0.0120  | 2.98   | 0.0022  | 0.57   | 0.0071 | 1.83   | -0.0226 | -5.85  |
| [85-94] | -0.0030 | -0.82  | 0.0010  | 0.28   | 0.0039  | 1.01   | 0.0132  | 3.44   | 0.0010  | 0.28   | 0.0056 | 1.46   | -0.0217 | -5.64  |
| [86-95] | -0.0033 | -0.89  | 0.0022  | 1.59   | 0.0015  | 0.39   | 0.0156  | 4.23   | -0.0006 | -0.18  | 0.0060 | 1.61   | -0.0213 | -5.51  |
| [87-96] | -0.0062 | -1.52  | 0.0072  | 1.77   | -0.0030 | -0.71  | 0.0175  | 4.37   | -0.0024 | -0.60  | 0.0103 | 2.55   | -0.0234 | -5.78  |
| [88-97] | -0.0069 | -1.90  | 0.0067  | 1.82   | -0.0001 | -0.03  | 0.0099  | 2.79   | 0.0012  | 0.34   | 0.0072 | 1.99   | -0.0180 | -5.06  |
| [89-98] | -0.0047 | -1.27  | 0.0029  | 0.77   | 0.0007  | 0.19   | 0.0091  | 2.45   | 0.0037  | 0.99   | 0.0070 | 1.94   | -0.0185 | -5.04  |
| [90-99] | -0.0048 | -1.37  | 0.0029  | 0.85   | 0.0028  | 0.83   | 0.0070  | 2.07   | 0.0066  | 1.89   | 0.0047 | 1.37   | -0.0192 | -5.66  |
| [91-00] | -0.0024 | -0.68  | 0.0018  | 0.53   | 0.0041  | 1.19   | 0.0042  | 1.20   | 0.0071  | 2.03   | 0.0063 | 1.79   | -0.0212 | -5.93  |

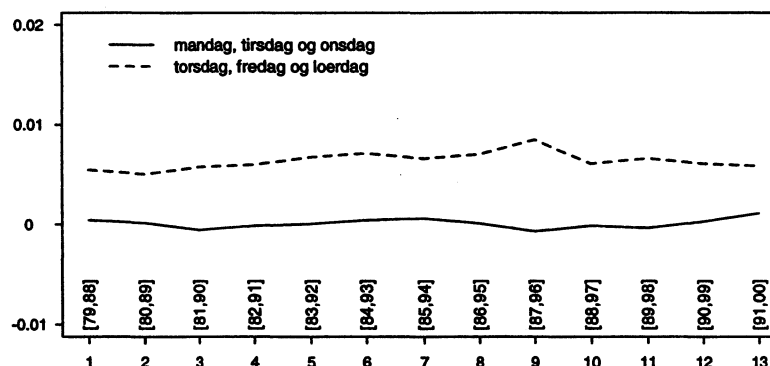
Merk at

- Ingen av 5 ARIMA modeller gir gode tilpasninger for [82-91] og [86-95], ARIMA(0 1 1)(0 1 1) er brukt
- ARIMA(0 1 1)(0 1 1) er valgt av X-12-ARIMA for de andre periodene
- et nivåskift (LS) i februar 1987 og en additiv outlier (AO) i desember 1988

6. Figur 13.5 viser at det er liten endring av vektene mellom de to gruppene (mandag, torsdag og onsdag) og (torsdag, fredag og lørdag). Siden gjennomsnittene  $v_1$  og  $v_2$  ligger nesten på samme nivå, viser dette at handelen i større grad flyttet seg mellom dagene i hver gruppe.



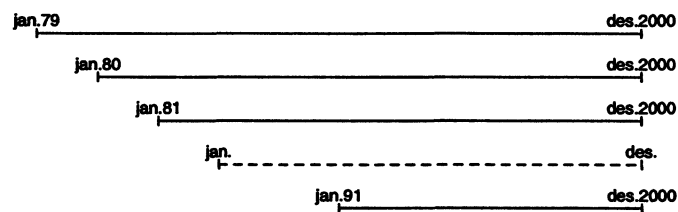
Figur 13.4: Vektene for ukedager fra tabell 13.6



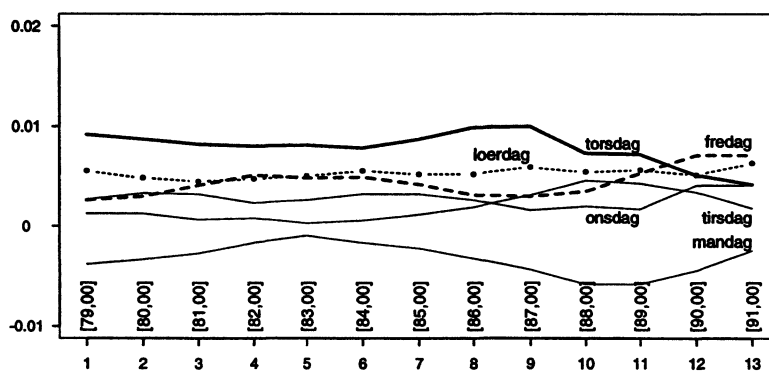
Figur 13.5: De gjennomsnittlige vektene  $v_1$  og  $v_2$  beregnes fra tabell 13.6

### 13.3.2 Tester for bevegelser av ukedagseffekter kun for de senere dataene

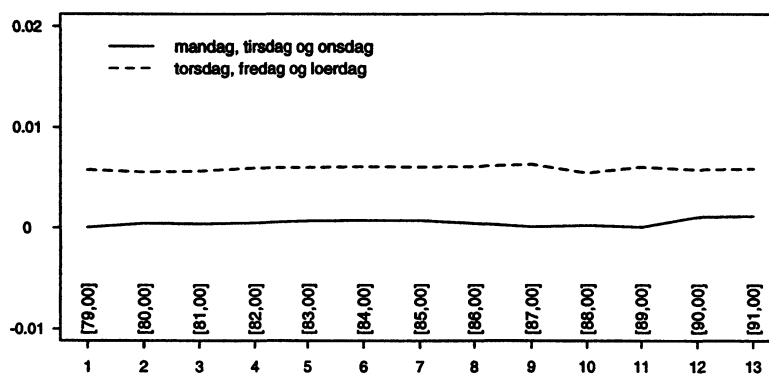
Fra rådataene lager vi nye tidsserier ved å fjerne gradvis de tidligere observasjonene. Den første starter fra januar 1980 til desember 2000, som er et år kortere enn den opprinnelige serien. Deretter blir data i 1980 fjernet for å lage den andre, osv. Den siste serien er fra januar 1991 til desember 2000. Vi får tidsserier som inneholder færre og færre data. Metoden illustreres i figuren 13.6. Vektene for ukedager for hver tidsserie er estimert ved regresjonsanalysen. Resultater er listet ut i tabell 13.7 og plottet i figur 13.7. De gjennomsnittlige vektene  $\bar{v}_1$  for (mandag, tirsdag og onsdag) og  $\bar{v}_2$  for (torsdag, fredag og lørdag) er vist i figur 13.8.



Figur 13.6: *De nye tidsseriene*



Figur 13.7: *Vektene for ukedager i tabell 13.7*



Figur 13.8: *De gjennomsnittlige vektene  $\bar{v}_{1,}$  og  $\bar{v}_{2,}$  beregnes fra tabell 13.7*

Tabell 13.7: Ukedagseffekter

| periode | mandag  |        | tirsdag |        | onsdag |        | torsdag |        | fredag |        | lørdag |        | søndag  |        |
|---------|---------|--------|---------|--------|--------|--------|---------|--------|--------|--------|--------|--------|---------|--------|
|         | estim.  | t-ver. | estim.  | t-ver. | estim. | t-ver. | estim.  | t-ver. | estim. | t-ver. | estim. | t-ver. | estim.  | t-ver. |
| [79-00] | -0.0038 | -1.51  | 0.0013  | 0.54   | 0.0027 | 1.10   | 0.0092  | 3.68   | 0.0026 | 1.04   | 0.0055 | 2.16   | -0.0175 | -6.86  |
| [80-00] | -0.0033 | -1.29  | 0.0013  | 0.52   | 0.0033 | 1.29   | 0.0087  | 3.40   | 0.0030 | 1.17   | 0.0048 | 1.84   | -0.0177 | -6.77  |
| [81-00] | -0.0027 | -1.00  | 0.0007  | 0.27   | 0.0032 | 1.23   | 0.0082  | 3.13   | 0.0041 | 1.55   | 0.0045 | 1.69   | -0.0181 | -6.74  |
| [82-00] | -0.0016 | -0.59  | 0.0008  | 0.29   | 0.0023 | 0.87   | 0.0080  | 3.03   | 0.0051 | 1.91   | 0.0047 | 1.75   | -0.0192 | -7.13  |
| [83-00] | -0.0009 | -0.32  | 0.0003  | 0.10   | 0.0026 | 0.96   | 0.0081  | 2.96   | 0.0048 | 1.74   | 0.0050 | 1.82   | -0.0200 | -7.14  |
| [84-00] | -0.0016 | -0.55  | 0.0006  | 0.23   | 0.0032 | 1.13   | 0.0078  | 2.76   | 0.0049 | 1.74   | 0.0055 | 1.93   | -0.0206 | -7.18  |
| [85-00] | -0.0022 | -0.75  | 0.0012  | 0.42   | 0.0032 | 1.08   | 0.0087  | 2.96   | 0.0042 | 1.45   | 0.0052 | 1.71   | -0.0204 | -6.79  |
| [86-00] | -0.0032 | -1.02  | 0.0019  | 0.62   | 0.0026 | 0.84   | 0.0099  | 3.21   | 0.0031 | 1.02   | 0.0052 | 1.67   | -0.0196 | -6.17  |
| [87-00] | -0.0043 | -1.27  | 0.0031  | 0.92   | 0.0016 | 0.46   | 0.0100  | 3.00   | 0.0030 | 0.89   | 0.0059 | 1.73   | -0.0192 | -5.66  |
| [88-00] | -0.0058 | -1.78  | 0.0046  | 1.43   | 0.0020 | 0.60   | 0.0073  | 2.26   | 0.0035 | 1.11   | 0.0054 | 1.64   | -0.0169 | -5.17  |
| [89-00] | -0.0058 | -1.69  | 0.0043  | 1.29   | 0.0017 | 0.50   | 0.0072  | 2.16   | 0.0053 | 1.60   | 0.0056 | 1.69   | -0.0184 | -5.42  |
| [90-00] | -0.0044 | -1.30  | 0.0034  | 1.04   | 0.0041 | 1.24   | 0.0051  | 1.57   | 0.0071 | 2.15   | 0.0051 | 1.53   | -0.0204 | -6.12  |
| [91-00] | -0.0024 | -0.68  | 0.0018  | 0.53   | 0.0041 | 1.19   | 0.0042  | 1.20   | 0.0071 | 2.03   | 0.0063 | 1.79   | -0.0212 | -5.93  |

Merk at

- ARIMA(0 1 1)(0 1 1) er valgt av X-12-ARIMA for de andre periodene
- et nivåskift (LS) i februar 1987 og en additiv outlier (AO) i desember 1988

Vi ser at

1. For dataene fra 1990 og senere blir vekten til torsdagen mindre. Det tyder på at torsdagen ikke er en like viktig handledag lenger.
2. Vekten for lørdagen beholder nesten samme nivå når tidsserien flytter seg. Lørdagen er den nest viktigste handledagen i uka.
3. Fredagen blir en større handledag mot slutten av perioden.
4. Figur 13.8 viser at det er ingen forskyvning av handelen mellom de to gruppene (mandag, tirsdag, onsdag) og (torsdag, fredag, lørdag), men innen hver av gruppene. Det er særlig en forskyvning av handelen fra torsdag og lørdag til fredag.

Ved å bruke de tidligere dataene fra 1979 i beregningen er torsdag den viktigste handledagen i uka med en vekt 0.0092 og  $t=3.68$  (se tabell 13.5), mens for kun den siste perioden, for eksempel, fra 1990, er det fredag med vekten 0.0071 og  $t=2.03$ . Vekten til torsdagen blir halvert (0.0042) og  $t$ -verdien er ikke signifikant på 5% lenger ( $t=1.20$ ). Vi ser at tolkningen for ukedager er avhengig av de tidsseriene som analyseres. En ulempe ved å sesongjustere en lang tidsserie er at ukedagseffekter eller sesongmønsteret fra de tidligere observasjoner kan skape forvirring i de senere data. For en kort tidsserie er det for få observasjoner i estimeringen.

## 13.4 Resultatene for sesongjusteringen

Vi presenterer resultatene for to tidsserier, den ene er fra januar 1979 til desember 2000 ( $[jan.79 - des.2000]$ ), med 264 observasjoner og den andre er fra januar 1990 til desember 2000 ( $[jan.90 - des.2000]$ ), med 132 observasjoner. Formålet er å studere endringer av sesongjusterte tall når kun de senere årene sesongjusteres. I beregningen lar vi programmet selv finne en ARIMA modell og intervensjoner for dataene. Vi får

### 13.4.1 Modell for sesongjustering

En multiplikativ modell er valgt for sesongjustering (dvs,  $O_t = T_t \times S_t \times I_t$ ). Dataene blir transformert med logaritmisk funksjon ( $\log$ ) i beregningene.

### 13.4.2 ARIMA modell

ARIMA(0 1 1)(0 1 1) er valgt for begge tidsserier. De estimerte verdiene for  $\hat{\theta}$  og  $\hat{\Theta}$  er listet ut i tabell 13.8.

Tabell 13.8: ARIMA modeller

|                | [jan.79 – des.2000] |        |             | [jan.90 – des.2000] |        |             |
|----------------|---------------------|--------|-------------|---------------------|--------|-------------|
|                | est.verdi           | std.   | signifikans | est.verdi           | std.   | signifikans |
| $\hat{\theta}$ | 0.6261              | 0.0487 | på 5%       | 0.7220              | 0.0659 | på 5%       |
| $\hat{\Theta}$ | 0.4767              | 0.0565 | på 5%       | 0.3351              | 0.0887 | på 5%       |

$p$ -verdiene for testen om residualene er ukorrelererte, er 0.18 og 0.81, henholdsvis. Dette viser at den korte tidsserien  $[jan.90 - des.2000]$  er noe bedre tilpasset med ARIMA(0 1 1)(0 1 1).

### 13.4.3 Ukedagseffekter

De estimerte verdiene er listet ut i tabell 13.9.

Tabell 13.9: *Ukedagseffekter*

|         | [jan.79 – des.2000] |        |         | [jan.90 – des.2000] |        |         |
|---------|---------------------|--------|---------|---------------------|--------|---------|
|         | estimert            | std.   | t-verdi | estimert            | std.   | t-verdi |
| mandag  | -0.0038             | 0.0025 | -1.51   | -0.0044             | 0.0033 | -1.30   |
| tirsdag | 0.0013              | 0.0024 | 0.54    | 0.0034              | 0.0032 | 1.04    |
| onsdag  | 0.0027              | 0.0025 | 1.10    | 0.0041              | 0.0032 | 1.24    |
| torsdag | 0.0092              | 0.0024 | 3.68    | 0.0051              | 0.0032 | 1.57    |
| fredag  | 0.0026              | 0.0024 | 1.04    | 0.0071              | 0.0033 | 2.15    |
| lørdag  | 0.0055              | 0.0025 | 2.16    | 0.0051              | 0.0033 | 1.53    |
| søndag  | -0.0175             | 0.0025 | -6.86   | -0.0204             | 0.0033 | -6.12   |

La  $e_{man.}$  være effekten for mandagen i uka. Vi definerer på samme måte for de andre dagene. Summen av ukedagseffekter er gitt i tabell 13.10.

Tabell 13.10:  $\sum_{t=man.}^{ons.} e_t$ ,  $\sum_{t=tor.}^{lør.} e_t$  og  $\sum_{t=man.}^{søn.} e_t$

| perioden            | søndag  | $\sum_{man.}^{ons.}$ | $\sum_{tor.}^{lør.}$ | $\sum_{man.}^{søn.}$ |
|---------------------|---------|----------------------|----------------------|----------------------|
| [jan.79 – des.2000] | -0.0175 | 0.0002               | 0.0173               | 0.0000               |
| [jan.90 – des.2000] | -0.0204 | 0.0031               | 0.0173               | 0.0000               |

Vi ser at

1. Summen av effektene på torsdagen, fredagen og lørdagen er den samme for begge tidsserier. Dette betyr at det er ingen forskyvning ut av de tre dagene. Det er imidlertid forskyvning fra torsdag og lørdag til fredag i de siste årene (fra 1990).
2. Handelen øker for mandag, tirsdag og onsdag de senere årene.

### 13.4.4 Intervensjoner

Tidspunktene, typene og størrelsene av intervensjoner er vist i tabell 13.11.

Tabell 13.11: *Intervensjoner*

| periode             | år                 | måned    | type | est.verdi | std    | t-verdi |
|---------------------|--------------------|----------|------|-----------|--------|---------|
| [jan.79 – des.2000] | 1987               | februar  | LS   | -0.0764   | 0.0148 | -5.14   |
|                     | 1988               | desember | AO   | -0.0862   | 0.0173 | -4.97   |
| [jan.90 – des.2000] | ingen intervensjon |          |      |           |        |         |

Det er et nivåskift i februar 1987 og en additiv outlier i februar 1988 for dataene i perioden [jan.79 – des.2000].

### 13.4.5 Framskrivinger

Framskrivingene av de to tidsseriene [*jan.79 – des.2000*] og [*jan.90 – des.2000*] er listet ut i tabell 13.12. Vi ser at tallene er omtrent de samme, men de 95% konfidensintervallene for framskrivinger som beregnes med dataene i intervallet [*jan.90 – des.2000*], er litt smalene.

Tabell 13.12: *Framskrivinger i 2001*

| måned     | [ <i>jan.79 – des.2000</i> ] |         |        |         | [ <i>jan.90 – des.2000</i> ] |         |        |         |
|-----------|------------------------------|---------|--------|---------|------------------------------|---------|--------|---------|
|           | nedre                        | framsk. | øvre   | lengden | nedre                        | framsk. | øvre   | lengden |
| januar    | 101.32                       | 105.84  | 110.57 | 9.25    | 102.51                       | 106.91  | 111.51 | 9.00    |
| februar   | 93.10                        | 97.50   | 102.11 | 9.01    | 93.85                        | 97.97   | 102.28 | 8.43    |
| mars      | 104.58                       | 109.86  | 115.40 | 10.82   | 105.24                       | 110.11  | 115.20 | 9.96    |
| april     | 98.95                        | 104.21  | 109.75 | 10.80   | 99.79                        | 104.58  | 109.61 | 9.82    |
| mai       | 111.74                       | 117.97  | 124.55 | 12.81   | 113.42                       | 119.05  | 124.97 | 11.55   |
| juni      | 113.17                       | 119.75  | 126.71 | 13.54   | 114.42                       | 120.26  | 126.40 | 11.98   |
| juli      | 109.46                       | 116.10  | 123.14 | 13.68   | 110.52                       | 116.36  | 122.51 | 11.99   |
| august    | 117.98                       | 125.41  | 133.31 | 15.33   | 119.54                       | 126.04  | 132.89 | 13.35   |
| september | 106.26                       | 113.17  | 120.53 | 14.27   | 106.52                       | 112.43  | 118.67 | 12.15   |
| oktober   | 110.81                       | 118.28  | 126.26 | 15.45   | 111.67                       | 118.09  | 124.87 | 13.20   |
| november  | 111.94                       | 119.71  | 128.02 | 16.08   | 112.88                       | 119.48  | 126.48 | 13.60   |
| desember  | 148.40                       | 159.01  | 170.38 | 21.98   | 149.08                       | 158.03  | 167.51 | 18.43   |

### 13.4.6 Tester for sesongmønster og sesongbevegelser

Resultater er vist i tabell 13.13. Testen viser at sesongvariasjoner påvirker dataene. Men det er endringer i sesongmønsteret for de tidligere og senere observasjonene. Dermed kan det være en fordel å sesongjustere kun de senere årene, siden vi får en stabil sesongkomponent.

Tabell 13.13: *Tester for sesongmønster og sesongbevegelser*

|                  | [ <i>jan.79 – des.2000</i> ] |       |    | [ <i>jan.90 – des.2000</i> ] |       |     |
|------------------|------------------------------|-------|----|------------------------------|-------|-----|
|                  | <i>F</i> -verdi              | sign. |    | <i>F</i> -verdi              | sign. |     |
| sesongmønster    | 754.85                       | 0.1%  | ja | 512.59                       | 0.1%  | ja  |
| sesongbevegelser | 2.77                         | 1.0%  | ja | 0.41                         | >5.0% | nei |

### 13.4.7 Kvalitetsmål og sliding spans

De 11 målene *M1 – M11* og *Q* er gitt i tabell 13.14. Vi ser at *M3* og *M5* for tidsserien [*jan.79 – des.2000*] er større enn 1 (grensen for en god test er 1). Dette viser at trenden er glatt slik at endringen fra måned til måned i den irregulære komponenten er større i forhold til endringen i trenden (se figur 13.9). Det blir dermed vanskelig å skille de to komponentene fra hverandre. Kvaliteten til sesongjustering blir mindre.

For tidsserien [jan.90 – des.2000] blir  $M3$  og  $M5$  bedre. Ved å basere oss på sliding spans analysen får vi gode sesongjusteringer for begge tidsserier.

Tabell 13.14: Kvalitetsmål

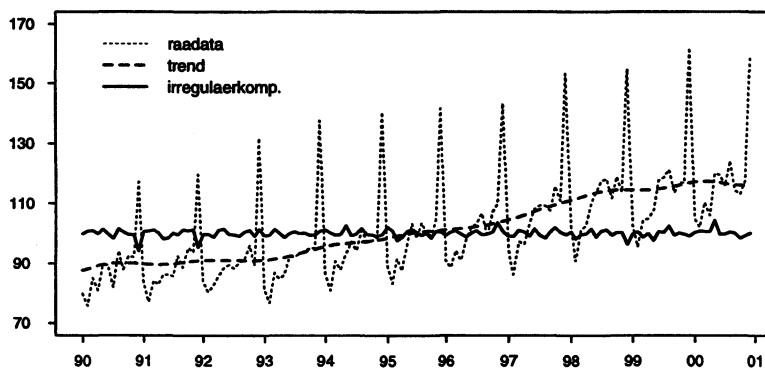
|                     | $M1$  | $M2$  | $M3$  | $M4$  | $M5$  | $M6$  |
|---------------------|-------|-------|-------|-------|-------|-------|
| [jan.79 – des.2000] | 0.072 | 0.050 | 1.306 | 0.322 | 1.244 | 0.219 |
| [jan.90 – des.2000] | 0.063 | 0.048 | 1.095 | 0.215 | 0.867 | 0.356 |

|                     | $M7$  | $M8$  | $M9$  | $M10$ | $M11$ | $Q$   |
|---------------------|-------|-------|-------|-------|-------|-------|
| [jan.79 – des.2000] | 0.101 | 0.208 | 0.126 | 0.235 | 0.220 | 0.388 |
| [jan.90 – des.2000] | 0.090 | 0.226 | 0.175 | 0.259 | 0.241 | 0.333 |

|                     | Sliding spans                                   |
|---------------------|-------------------------------------------------|
| [jan.79 – des.2000] | $S(\%) = 0.0$ , $MM(\%) = 0.0$ , $YY(\%) = 0.0$ |
| [jan.90 – des.2000] | $S(\%) = 0.0$ , $MM(\%) = 0.0$ , $YY(\%) = 0.0$ |



Figur 13.9: Rådata, trend og irregulære komponenten for tidsserien [jan.79 – des.2000]

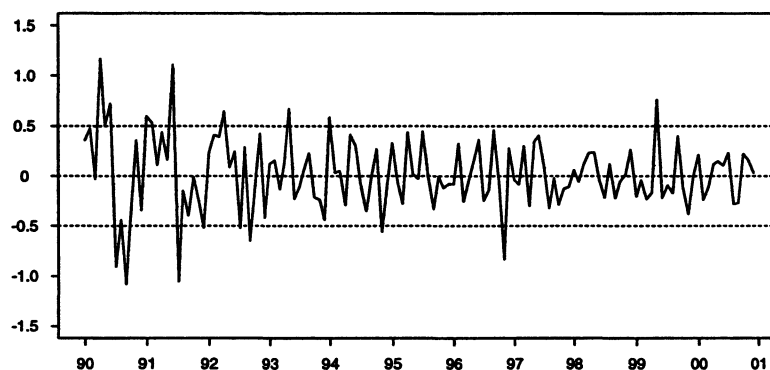
#### 13.4.8 Figurer

La  $A_t^{79}$  og  $A_t^{90}$  være sesongjusterte tall i tidspunkt  $t$  for tidsseriene [jan.79–des.2000] og [jan.90–des.2000], henholdsvis. Vi innfører en størrelse  $R_t^A$  som er relative avviket i prosent i tidspunkt  $t$  for  $A_t^{79}$  og  $A_t^{90}$ . Den beregnes ved

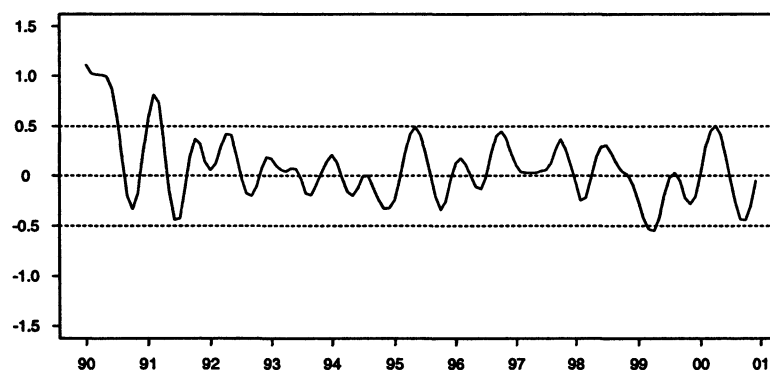
$$R_t^A = 100\% \times \frac{A_t^{90} - A_t^{79}}{A_t^{79}}$$

Figur 13.10 viser disse størrelsene for sesongjusterte tall fra januar 1990 til desember 2000. Vi beregner på samme måte for trend og sesongkomponenten, som betegnes med  $R_t^T$  og  $R_t^S$ . De er plottet i figurene 13.11 og 13.12. Vi ser at  $R_t^A$ ,  $R_t^T$  og  $R_t^S$  er lave. Tallene varierer rundt  $[-0.5, 0.5\%]$  bortsett fra de tidspunktene i 1991 og 1992. Vi får dermed omtrent like store sesongjusterte tall, trender og sesongkomponentene for de to tidsseriene. Vi plottes derfor kun komponentene for tidsserien [jan.90 – des.2000] i figurene 13.13–13.15. Figur 13.16 er sesongkomponentene for årene fra 1990 til 2000

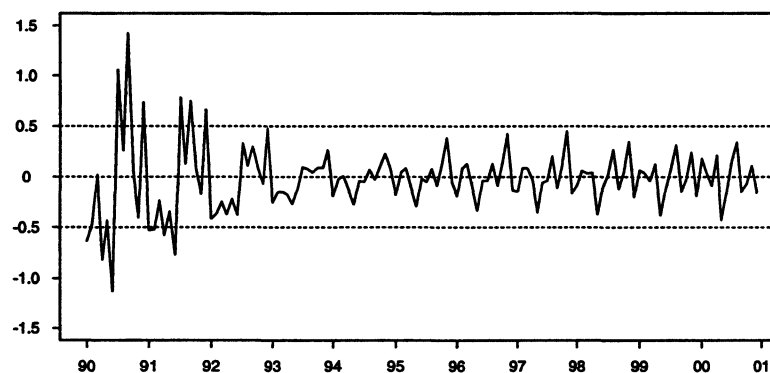
i samme måned mot deres gjennomsnitt. Det gir et godt bilde av sesongvariasjoner fra måned til måned og fra år til år. Vi ser at sesongmønsteret er ganske stabilt. Desember er måneden med det høyeste sesongkomponenten.



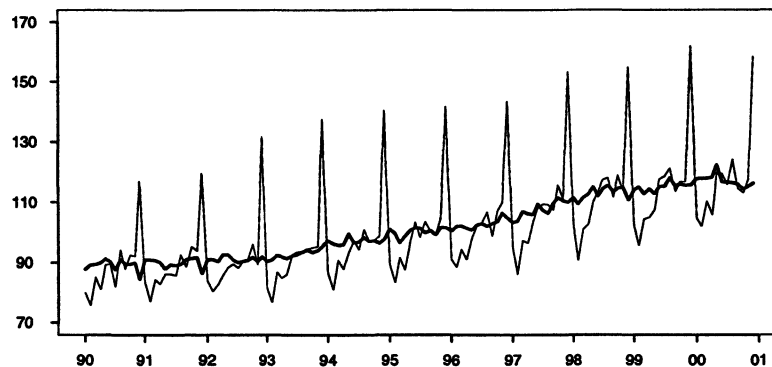
Figur 13.10: *Relative avviket i prosent av sesongjusterte tallene*



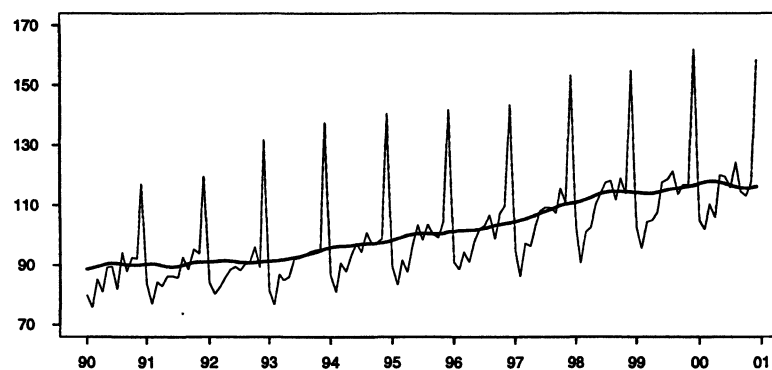
Figur 13.11: *Relative avviket i prosent av trendene*



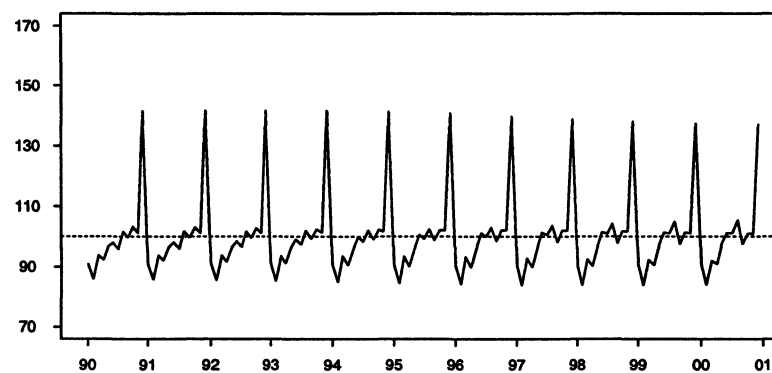
Figur 13.12: *Relative avviket i prosent av sesongkomponentene*



Figur 13.13: *Rådataene og sesongjusterte tall for tidsserien [jan.90 – des.2000]*



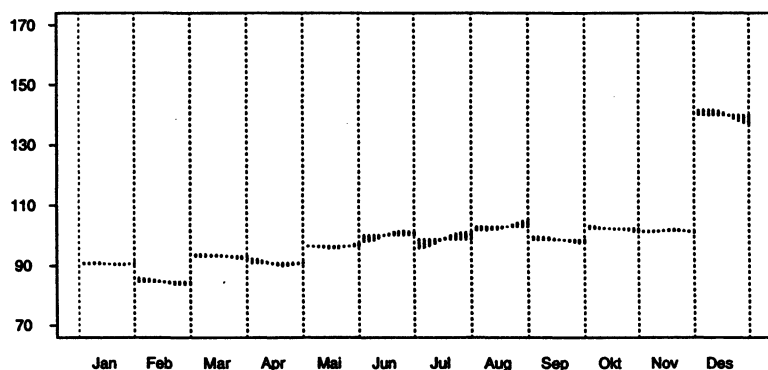
Figur 13.14: *Rådataene og trend for tidsserien [jan.90 – des.2000]*



Figur 13.15: *Sesongkomponenten for tidsserien [jan.90 – des.2000]*

#### 13.4.9 Revisjoner

Formålet er å se endringer i sesongjusterte tall når nye observasjoner legges til. Revisjoner for begge tidsserier er beregnet fra januar 1998 til november 2000. La  $A_{t|N}$  være sesongjusterte tall i tidspunkt  $t$  av tidsserien som er observert til tidspunkt  $N$ ,



Figur 13.16: *Sesongkomponenten for 12 måneder*

$1 \leq t \leq N$ . Dermed er  $A_{t|t}$  sesongjusterte tall i det siste tidspunktet av tidsserien. For eksempel, revisjonen i januar 1998 for tidsserien  $[jan.90 - des.2000]$  beregnes ved

$$\Delta_{jan.98}^{90} = \frac{A_{jan.98|des.00}^{90} - A_{jan.98|jan.98}^{90}}{A_{jan.98|jan.98}^{90}}$$

Vi får ikke revisjoner for tidspunkter  $t$  før januar 1998 siden det er få observasjoner for å estimere sesongjusterte tall  $A_{t|t}^{90}$  med dataene i  $[jan.90 - des.2000]$ .

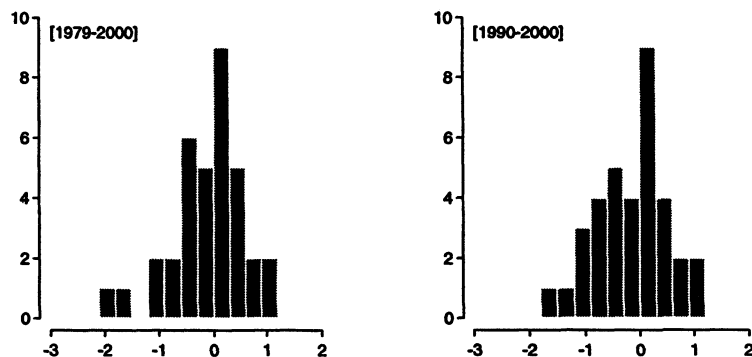
Tabell 13.15: *Revisjoner for sesongjusterte tall i prosent*

| måned     | [jan.79 – des.2000] |       |       | [jan.90 – des.2000] |       |       |
|-----------|---------------------|-------|-------|---------------------|-------|-------|
|           | 1998                | 1999  | 2000  | 1998                | 1999  | 2000  |
| januar    | 0.08                | 0.03  | 0.34  | 0.18                | -0.12 | 0.34  |
| februar   | -0.49               | -0.71 | 0.35  | -0.64               | -0.75 | 0.26  |
| mars      | 1.06                | -0.25 | -0.11 | 1.14                | -0.17 | -0.12 |
| april     | 0.23                | -0.40 | -0.27 | 0.05                | -0.60 | -0.32 |
| mai       | -1.04               | -2.00 | -1.72 | -0.72               | -1.62 | -0.97 |
| juni      | -0.21               | 0.24  | -0.59 | -0.58               | 0.13  | -0.97 |
| juli      | -0.36               | 0.16  | -0.55 | -0.45               | 0.09  | -0.76 |
| august    | -1.17               | 0.16  | -0.74 | -1.32               | 0.12  | -0.98 |
| september | -0.05               | 0.06  | -0.47 | -0.21               | 0.11  | -0.54 |
| oktober   | 0.56                | 1.10  | 0.22  | 0.56                | 1.01  | 0.04  |
| november  | 0.64                | 0.83  | 0.10  | 0.71                | 0.83  | 0.10  |
| desember  | 0.53                | 0.40  |       | 0.54                | 0.46  |       |

Tabell 13.15 og figur 13.17 viser revisjoner for begge tidsserier. Vi ser at revisjoner av sesongjusterte tall i mai for  $[jan.90 - des.2000]$ , er mindre. Ellers får vi omtrent samme størrelser.

#### 13.4.10 Sesongjusterte tall i desember 1999 og 2000

Tabell 13.16 viser tallene for desember 1999 og 2000 for begge tidsserier.  $O_t$ ,  $S_t$ ,  $A_t$ ,  $T_t$  og  $I_t$  står for rådata, sesongkomponenten, sesongjusterte tall, trend og den



Figur 13.17: Revisjoner av sesongjusterte tall

Tabell 13.16: Tallene for desember 1999 og 2000

|       | [jan.79 – des.2000] |        |         |        |        |        |        |
|-------|---------------------|--------|---------|--------|--------|--------|--------|
|       | $O_t$               | $f_t$  | $O_t^K$ | $S_t$  | $A_t$  | $T_t$  | $I_t$  |
| 1999  | 161.88              | 1.0146 | 159.55  | 137.82 | 115.77 | 116.85 | 99.07  |
| 2000  | 158.55              | 0.9906 | 160.06  | 137.48 | 116.42 | 116.25 | 100.15 |
| $R_t$ | -2.05               | -2.37  | 0.32    | -0.25  | 0.56   | -0.51  | 1.09   |

|       | [jan.90 – des.2000] |        |         |        |        |        |        |
|-------|---------------------|--------|---------|--------|--------|--------|--------|
|       | $O_t$               | $f_t$  | $O_t^K$ | $S_t$  | $A_t$  | $T_t$  | $I_t$  |
| 1999  | 161.88              | 1.0165 | 159.26  | 137.56 | 115.77 | 116.61 | 99.29  |
| 2000  | 158.55              | 0.9918 | 159.86  | 137.26 | 116.46 | 116.19 | 100.24 |
| $R_t$ | -2.05               | -2.43  | 0.38    | -0.22  | 0.64   | -0.35  | 0.96   |

irregulære komponenten, henholdsvis.  $f_t$  og  $O_t^K$  er korrigeringsfaktoren og rådata etter korrigeringen for ukedagseffekter.  $R_t$  er det relative avviket i prosent. Vi ser at

1. De komponentene  $S_t$ ,  $A_t$ ,  $T_t$  og  $I_t$  er nesten like for de to tidsseriene.
2. Desember i 1999 og 2000 får omtrent like høye sesongkomponenter  $S_t$  og trender  $T_t$ .
3. Sammenligning med desember 1999, er den observerte verdien for desember 2000 lavere ( $O_{des.99}=161.88$  og  $O_{des.2000}=158.55$ , med et relativt avvik er - 2.05%), mens sesongjusterte tallet blir høyere ( $A_{des.99}^{90} = 115.77$  og  $A_{des.2000}^{90} = 116.46$ , med relativt avviket er 0.64%). Dermed kan vi ikke forvente at det sesongjusterte tallet i desember 2000 skal bli lavere i forhold til desember 1999 på grunn av at det var to handledager mindre i desember 2000, julen og nyttåret falt på en søndag. Vi ser forøvrig at den irregulære komponenten for desember 1999 er nesten en 1% under 100, men såvidt over 100 i desember 2000. Dette kan tyde på at det sesongjusterte tallet i desember 1999 er i høyeste laget sammenliknet med sesongjustert i desember 2000.

4. Verdien for desember 2000 er 158.55. Denne måneden har 4 mandager, 4 tirsdager, 4 onsdager, 4 torsdager, 5 fredager, 5 lørdager og 5 søndager. Vi antar at antall ukedager i desember for 2000 er like med antall ukedager for 1999. Fra tabell 13.16 kan vi beregne den nye verdien  $O_{des.2000}^{ny}$  for desember 2000 ved

$$\log(O_{des.2000}^{ny}) = \log(158.55) - \log(0.9918) + \log(1.0165)$$

Vi får  $O_{des.2000}^{ny} = 162.50$  som er høyere enn den observerte verdien i desember 1999,  $O_{des.99} = 161.88$ . Sesongkomponentene er  $S_{des.99} = 137.56$  og  $S_{des.2000} = 137.26$ . Siden de er nesten like så kan vi anta at sesongkomponentene i desember 1999 og 2000 er de samme for å forklare lettere. Nå kan vi sammenligne sesongjusterte tall i desember 1999 og 2000 ved å bruke de observerte verdiene  $O_{des.99}$  og  $O_{des.2000}^{ny}$ , siden disse to månedene har samme sesongvariasjonen og samme antall ukedager. Siden  $O_{des.2000}^{ny} > O_{des.99}$  medfører dette  $A_{des.2000} > A_{des.99}$ .

## 13.5 Konklusjon

Dataene for detaljvolumindeksen er påvirket kraftig av ukedagseffekter og sesongvariasjoner. Disse effektene må fjernes for å beregne trend og sesongjusterte tall. Dermed er det ikke riktig å sammenligne sesongjusterte tall i to måneder ved å bruke deres observerte verdier. Ved å bruke dataene fra 1990 i sesongjustering får vi et riktigere bilde av ukedagseffekter i de siste årene. Torsdagen og lørdagen er ikke like viktige handledager lenger. Det er en overgang fra disse dagene til fredag. Dessuten blir sesongkomponenten mer stabil. Vi har publisert tallene som er basert på observasjonene fra januar 1979. Når de tidligere dataene før 1990 blir fjernet i beregningene får vi svært like verdier for sesongkomponenten, sesongjusterte tall og trend sammenliknet med de beregnes fra 1979. Selv om handlemønsteret har endret seg i 90-årene sammenliknet med på 80-tallet gir allikevel den lengre tidsserien samme bilde av utviklingen fra 1999 til 2000 som den korte tidsserien som starter i 1990.

## Kapittel 14

# Stokastiske modeller for ukedagseffekter for detaljvolumindeksen

Dinh Quang Pham

### 14.1 Innledning

Modellen for ukedagseffekter som vi bruker i X-11-ARIMA og X-12-ARIMA er utviklet av Young (1965). Den er en deterministisk modell, der vi antar at vekten for hver ukedag er den samme for hele tidsserien. Denne antagelsen ser urimelig ut for tids-serier i detaljvolumindeksen, siden aktiviteten av handelen kan variere fra måned til måned. En random walk modell for ukedagseffekter har beskrevet av Monsell (1983), Dagum og Quenneville (1988), der vekten  $\delta_{i,t}$  for ukedag  $i$ , i tidspunkt  $t$  skrives slik  $\delta_{i,t} = \delta_{i,t-1} + \eta_{i,t}$ , hvor  $\eta_{i,t}$  er støy-ledd.

I dette kapitlet beskriver vi ukedagseffekter med en stokastisk modell og anvender Kalmanfilter for å estimere effektene.

### 14.2 Beskrivelse av ukedagseffekter

La  $O_t$  være en observert verdi i tidspunkt  $t$ . Vi skriver  $O_t$  som summen av daglige verdier  $O_{j,t}$ , der  $j = 1, \dots, N_t$ , og  $N_t$  er antall dager i måneden  $t$ .  $N_t = 28, 29, 30$  eller  $31$ . For eksempel, for en måned som har 30 dager og starter fra mandag, er

$$O_t = \sum_{j=1}^{30} O_{j,t} \quad (14.1)$$

Vi antar at i denne måneden er det 5 mandager, 5 tirsdager, 4 onsdager, 4 torsdager, 4 fredager, 4 lørdager og 4 søndager. Ligningen (14.1) skrives slik

$$\begin{aligned} O_t &= O_{1,t} + O_{8,t} + O_{15,t} + O_{22,t} + O_{29,t} + && \text{for mandager} \\ &O_{2,t} + O_{9,t} + O_{16,t} + O_{23,t} + O_{30,t} + && \text{for tirsdager} \\ &O_{3,t} + O_{10,t} + O_{17,t} + O_{24,t} + && \text{for onsdager} \\ &O_{4,t} + O_{11,t} + O_{18,t} + O_{25,t} + && \text{for torsdager} \end{aligned}$$

$$\begin{aligned}
O_t &= O_{5,t} + O_{12,t} + O_{19,t} + O_{26,t} + && \text{for fredager} \\
&O_{6,t} + O_{13,t} + O_{20,t} + O_{27,t} + && \text{for lørdager} \\
&O_{7,t} + O_{14,t} + O_{21,t} + O_{28,t} + && \text{for søndager}
\end{aligned}$$

eller

$$O_t = 5\bar{O}_{man.} + 5\bar{O}_{tir.} + 4\bar{O}_{ons.} + 4\bar{O}_{tor.} + 4\bar{O}_{fre.} + 4\bar{O}_{lør.} + 4\bar{O}_{søn.} \quad (14.2)$$

der  $\bar{O}_{man.}$  er gjennomsnittet av  $O_{i,t}$  for 5 mandager, osv. Vi innfører et symbol  $i$ , for ukedager:  $i = 1$  står for mandag,  $i = 2$  for tirsdag,  $\dots$ , og  $i = 7$  for søndag. De størrelsene  $\bar{O}_i$ ,  $i = 1, \dots, 7$  kalles *daglige effekter*. La  $N_{1,t}$  være antall mandager,  $\dots$ , og  $N_{7,t}$  antall søndager i måned  $t$ . Ligningen (14.2) skrives slik

$$O_t = \sum_{i=1}^7 N_{i,t} \bar{O}_{i,t} \quad (14.3)$$

La

$$\bar{O}_{0,t} = \frac{1}{7} \sum_{i=1}^7 \bar{O}_{i,t}$$

være gjennomsnittet av daglige effekter. Merk at  $\bar{O}_{0,t}$  og  $\bar{O}_t = (1/N_t) \sum O_t$  er ikke de samme. Ligningen (14.3) blir

$$O_t = \sum_{i=1}^6 (N_{i,t} - N_{7,t})(\bar{O}_{i,t} - \bar{O}_{0,t}) + N_{7,t} \bar{O}_{0,t} \quad (14.4)$$

Det første leddet av (14.4) kalles *ukedagseffekter*. La

$$D_t = \sum_{i=1}^6 (N_{i,t} - N_{7,t})(\bar{O}_{i,t} - \bar{O}_{0,t}) \quad (14.5)$$

Vi ser at  $D_t = 0$  når alle  $\bar{O}_{i,t}$  er like (og lik  $\bar{O}_{0,t}$ ). Det er ingen effekter av ukedager. For en februar måned med lengden 28 dager, er også  $D_t = 0$ , siden  $N_{1,t} = \dots = N_{7,t}$ . Størrelsen  $D_t$  beskriver endringene av antall ukedag  $i$ , i forhold til antall søndager i måned  $t$ .

### 14.3 En stokastisk modell for ukedagseffekter

En enkel modell for ukedagseffekter er

$$w_t = D_t + e_t, \quad t = 1, \dots, T \quad (14.6)$$

hvor  $e_t$  er en hvit støy prosess,  $e_t \sim N(0, \sigma_e^2)$ .  $w_t$  er den originale tidsserien etter å ha fjernet trenden  $T_t$  og den sesongkomponenten  $S_t$ , dvs  $w_t = O_t/(T_t S_t)$  for den multiplikative modellen.

La  $\delta_{i,t} = \bar{O}_{i,t} - \bar{O}_{0,t}$  og  $x_{i,t} = N_{i,t} - N_{7,t}$ , for  $i = 1, \dots, 6$ . Da beskriver  $\delta_{i,t}$  avviket mellom effekten for ukedag  $i$  og gjennomsnittet. Siden

$$\sum_{i=1}^7 \delta_{i,t} = \sum_{i=1}^7 (\bar{O}_{i,t} - \bar{O}_{0,t}) = 0 \quad (14.7)$$

er koeffisienten for søndag,  $\delta_{7,t}$  gitt ved

$$\delta_{7,t} = \bar{O}_{7,t} - \bar{O}_{0,t} = - \sum_{i=1}^6 \delta_{i,t} \quad (14.8)$$

Tabell 14.1: Antall ukedager for januar og mars 2000

|           |  | januar |      |      |      |      |      |      |
|-----------|--|--------|------|------|------|------|------|------|
|           |  | man.   | tir. | ons. | tor. | fre. | lør. | søn. |
| # dager   |  | 5      | 4    | 4    | 4    | 4    | 5    | 5    |
| $x_{i,t}$ |  | 0      | -1   | -1   | -1   | -1   | 0    | 0    |
|           |  | mars   |      |      |      |      |      |      |
|           |  | man.   | tir. | ons. | tor. | fre. | lør. | søn. |
| # dager   |  | 4      | 4    | 5    | 5    | 5    | 4    | 4    |
| $x_{i,t}$ |  | 0      | 0    | 1    | 1    | 1    | 0    | 0    |

Tabell 14.1 viser antall ukedager for januar og mars 2000. Ukedagseffekten  $D_{jan.00}$  blir

$$\begin{aligned} D_{jan.00} &= \sum_{i=1}^6 x_{i,t} \delta_{i,t} = -\delta_{2,t} - \delta_{3,t} - \delta_{4,t} - \delta_{5,t} \\ &= \delta_{1,t} + \delta_{6,t} + \delta_{7,t} \quad \text{siden (14.7)} \end{aligned}$$

For mars 2000, er

$$D_{mar.00} = \delta_{3,t} + \delta_{4,t} + \delta_{5,t}$$

Vi ser at uansett antall søndager i en måned er 4 eller 5 dager, er ukedagseffekten  $D_t$  i denne måneden summen av daglige effektene som forekommer 5 ganger.

I X-12-ARIMA antar vi at  $\delta_{i,t}$ , for  $i = 1, \dots, 6$  ikke endrer seg med tid for hele tidsserien, dvs,  $\delta_{i,t} = \delta_i$ . Men i dette kapitlet antar vi at  $\delta_{i,t}$  følger en "random walk" modell, dvs,

$$\delta_{i,t} = \delta_{i,t-1} + \eta_{i,t} \quad (14.9)$$

der  $\eta_{i,t}$  er en hvit støy prosess,  $\eta_{i,t} \sim N(0, \sigma_\eta^2)$ . For å vise  $\delta_{i,t}$  av en stokastisk modell bruker vi  $\delta_{i,t}^s$  og for en deterministisk modell er  $\delta_{i,t}^d$ .

Modellen for  $\delta_{i,t}^s$  i dette kapitlet er

$$w_t = (x_{1,t} \ x_{2,t} \ \dots \ x_{6,t}) \begin{pmatrix} \delta_{1,t} \\ \delta_{2,t} \\ \vdots \\ \delta_{6,t} \end{pmatrix} + e_t \quad (14.10)$$

$$\begin{pmatrix} \delta_{1,t} \\ \delta_{2,t} \\ \vdots \\ \delta_{6,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} \delta_{1,t-1} \\ \delta_{2,t-1} \\ \vdots \\ \delta_{6,t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \vdots \\ \eta_{6,t} \end{pmatrix} \quad (14.11)$$

for  $t = 1, \dots, T$ . Det er en *tilstand modell* for  $\delta_{i,t}^s$ . Ligningen (14.10) kalles *måleligning* og (14.11) er *transisjonsligning*.  $e_t$  og  $\eta_t$ -ene er uavhengige restledd,  $e_t \sim N(0, \sigma_e^2)$  og  $\eta_{i,t} \sim N(0, \sigma_{\eta_i}^2)$ . Vi estimerer  $\delta_{i,t}^s$  ved Kalmanfilter.

La  $\delta_t^s = (\delta_{1,t}^s, \delta_{2,t}^s, \dots, \delta_{6,t}^s)'$ ,  $\eta_t = (\eta_{1,t}, \eta_{2,t}, \dots, \eta_{6,t})'$  og  $\Phi = I$ , en identitetsmatrise. Ligningen (14.11) skrives

$$\delta_t^s = \Phi \delta_{t-1}^s + \eta_t \quad (14.12)$$

hvor  $Q = E(\eta_t \eta_t')$ . Ved å angi  $\delta_{i,t}^s$  som følger en AR(1) modell, har vi en annen form for matrise  $\Phi$ ,  $\Phi = \text{diag}(\Phi_{1,1}, \Phi_{2,2}, \dots, \Phi_{6,6})$ .

#### a. Estimering $\delta_{i,t}^s$

Vi lager tidsserien  $\{w_t\}$  i ligningen (14.10) ved

$$w_t = \log(O_t) - \log(T_t) - \log(S_t) \quad (14.13)$$

Vi får for hvert tidspunkt  $t$  fra januar 1979 til desember 2000,  $\hat{\delta}_{1,t}^s, \hat{\delta}_{2,t}^s, \dots, \hat{\delta}_{6,t}^s$  for mandag, tirsdag,  $\dots$ , lørdag. Mens i X-12-ARIMA er  $\hat{\delta}_1^d, \hat{\delta}_2^d, \dots, \hat{\delta}_6^d$  for hele tidsserien. Vi beregner gjennomsnittene

$$\bar{\delta}_i^s = \frac{1}{T} \sum_{t=1}^T \hat{\delta}_{i,t}^s, \quad t = 1, \dots, T (= 264 \text{ måneder})$$

for å sammenligne med  $\hat{\delta}_i^d$  fra X-12-ARIMA. De estimerte koeffisientene  $\hat{\delta}_{i,t}^s$  er plottet i figur 14.1.  $\bar{\delta}_i^s$  og  $\hat{\delta}_i^d$  er listet ut i tabell 14.2. Figur 14.2 viser standardavvikene for  $\hat{\delta}_{i,t}^s$ .

Vi ser at

- endringene av  $\hat{\delta}_{i,t}^s$  fra tidspunkt til tidspunkt er svært lave. Deres standardavvik er også ganske lave. Dermed kan vi anta at vektene for ukedager ikke endrer seg med tid for hele tidsserien.
- regARIMA modell i X-12-ARIMA og modellene (14.10) og (14.11) viser at torsdag og lørdag er to viktige handledager i uka. Onsdag er også en stor handledag men med en mindre grad.  $\bar{\delta}_{tor}^s$  er lavere enn  $\hat{\delta}_{tor}^d$ , mens for onsdag er  $\bar{\delta}_{ons}^s > \hat{\delta}_{ons}^d$ . Det ser ut som en forskjvning av vektene mellom onsdag og torsdag. Vektene for mandag og tirsdag er veldig lave.

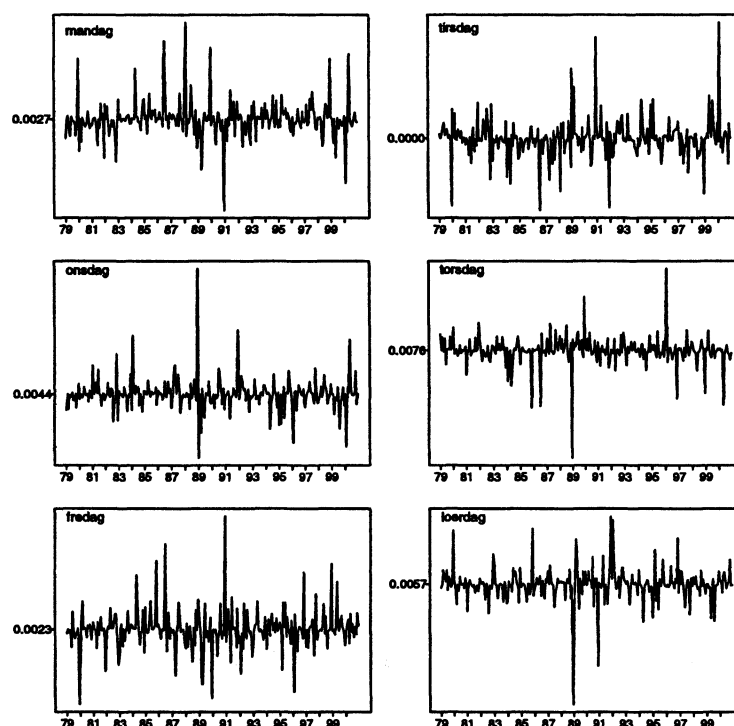
For søndag bruker vi formen (14.8) for å beregne vekten.

Vi får  $\hat{\sigma}_e^2 = 0.0002264$  og  $\hat{\sigma}_{\eta_1}^2 = 5.49 \times 10^{-7}$ ,  $\hat{\sigma}_{\eta_2}^2 = 5.71 \times 10^{-7}$ ,  $\hat{\sigma}_{\eta_3}^2 = 5.73 \times 10^{-7}$ ,  $\hat{\sigma}_{\eta_4}^2 = 5.70 \times 10^{-7}$ ,  $\hat{\sigma}_{\eta_5}^2 = 5.54 \times 10^{-7}$ ,  $\hat{\sigma}_{\eta_6}^2 = 5.82 \times 10^{-7}$ .

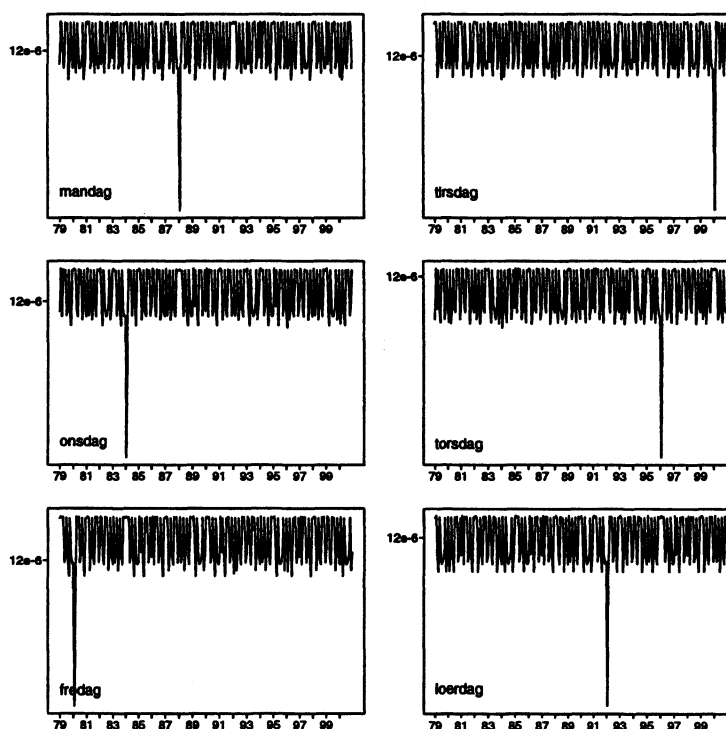
Man estimerer  $\delta_i^d$  i X-12-ARIMA ved en deterministisk modell, såkalt regARIMA modell,

$$O_t = \sum_{i=1}^6 x_{i,t} \delta_i + z_t \quad (14.14)$$

der  $z_t$  følger en ARIMA(p,d,q)(P,D,Q) modell. I dette kapitlet er estimeringen basert på tidsserien som har fjernet trend og den sesongkomponenten. Dessuten antar vi at



Figur 14.1: De estimerte koeffisientene  $\hat{\delta}_{i,t}^s$



Figur 14.2: Standardavvikene for  $\hat{\delta}_{i,t}^s$

restleddene er en hvit støy prosess. Dette er metoden som blir brukt i X-12-ARIMA.

Tabell 14.2: De estimerte verdiene og standardavvikene for koeffisientene ved Kalman-filter og X-12-ARIMA

|         | X-12-ARIMA         |        | Kalman             |                  |
|---------|--------------------|--------|--------------------|------------------|
|         | $\hat{\delta}_i^d$ | std    | $\bar{\delta}_i^s$ | $\overline{std}$ |
| mandag  | -0.0038            | 0.0025 | -0.0027            | 0.000128         |
| tirsdag | 0.0013             | 0.0024 | 0.0006             | 0.000127         |
| onsdag  | 0.0027             | 0.0025 | 0.0044             | 0.000129         |
| torsdag | 0.0092             | 0.0024 | 0.0075             | 0.000129         |
| fredag  | 0.0026             | 0.0024 | 0.0023             | 0.000128         |
| lørdag  | 0.0055             | 0.0025 | 0.0057             | 0.000129         |

En tilstandmodell for ligningen (14.14) er mer komplisert og det er flere parametre som skal estimeres.

La  $f_t^s$  være ukedagseffekten i måned  $t$ . Vi beregner  $f_t$  for desember 2000. Tabell 14.3 viser de koeffisientene  $\hat{\delta}_i^s$  for desember 2000 samt antall mandager, tirsdager, ..., søndager i måneden.

Tabell 14.3: De koeffisientene  $\hat{\delta}_i^s$  for desember 2000

|                    | mandag    | tirsdag  | onsdag   | torsdag  | fredag   | lørdag   | søndag    |
|--------------------|-----------|----------|----------|----------|----------|----------|-----------|
| $\hat{\delta}_i^s$ | -0.002719 | 0.000630 | 0.004417 | 0.007558 | 0.002343 | 0.005712 | -0.017941 |
| $N_{i,t}$          | 4         | 4        | 4        | 4        | 5        | 5        | 5         |

Vi får

$$\begin{aligned}
 f_{des.00}^s &= -0.002719 \times 4 + 0.000630 \times 4 + 0.004417 \times 4 + 0.007558 \times 4 \\
 &\quad + 0.002343 \times 5 + 0.005712 \times 5 - 0.017941 \times 5 \\
 &= -0.0098
 \end{aligned}$$

Korrigerte verdier for ukedagseffekten i denne måneden,  $O'_{des.00}^s$  blir

$$O'_{des.00}^s = \exp(\log(158.55) - (-0.0098)) = 160.11$$

mens ved X-12-ARIMA er  $f_t^d = -0.0094$ , og  $O'_{des.00}^d = 160.05$ . Avvikene mellom disse størrelsene er lave.

For februar beregner  $O_t^s$  på følgende måte

- med 28 dager

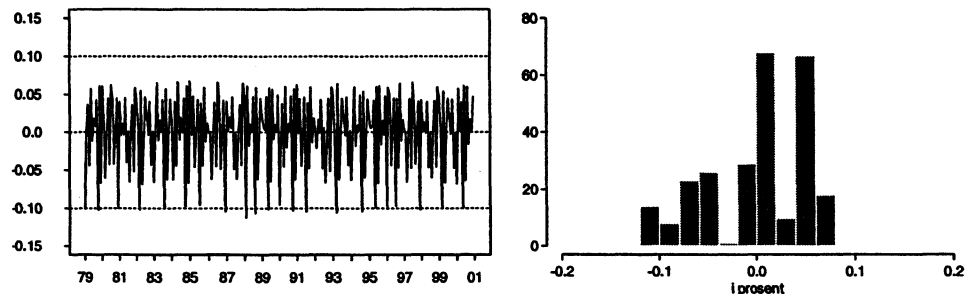
$$O_t^s = O_t \times \frac{28.25}{28}$$

- med 29 dager

$$O_t^s = \frac{28.25}{29} \exp(y_t)$$

der  $y_t = \log(O_t) - \sum_{i=1}^7 x_{i,t} \hat{\delta}_{i,t}^s$ , hvor  $x_{i,t} = N_{i,t} - N_{7,t}$ .

Figur 14.3 og tabell 14.4 (kolonne 2-5) viser de relative avvikene i prosent  $R_t^O$  mellom  $O_t^d$  fra X-12-ARIMA og  $O_t^s$  fra tilstandmodellen. Tallene er lave og ligger innen intervallet  $[-0.1, 0.1]$  prosent.



Figur 14.3: Relative avvikene i prosent mellom  $O_t^d$  og  $O_t^s$

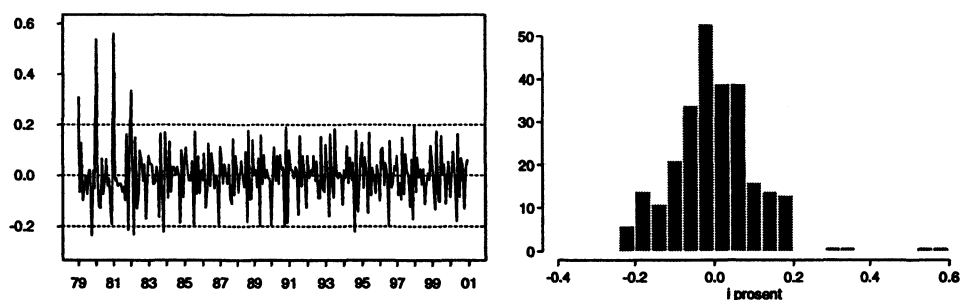
Tabell 14.4: Verdiene for  $O_t^d$  og  $O_t^s$  samt deres sesongjusterte tall,  $A_t^d$  og  $A_t^s$  i desember 2000

| måned     | X-12 Kalman |         |         |             | X-12 Kalman |         |             |
|-----------|-------------|---------|---------|-------------|-------------|---------|-------------|
|           | $O_t$       | $O_t^d$ | $O_t^s$ | $R_t^{O\%}$ | $A_t^d$     | $A_t^s$ | $R_t^{A\%}$ |
| januar    | 104.81      | 106.46  | 106.38  | -0.06       | 117.51      | 117.41  | -0.08       |
| februar   | 102.06      | 99.30   | 99.35   | 0.05        | 118.27      | 118.31  | 0.03        |
| mars      | 110.28      | 108.71  | 108.71  | 0.00        | 118.06      | 118.17  | 0.09        |
| april     | 105.86      | 107.12  | 107.16  | 0.03        | 118.15      | 118.12  | -0.02       |
| mai       | 119.94      | 119.78  | 119.66  | -0.09       | 122.34      | 122.12  | -0.17       |
| juni      | 119.52      | 118.27  | 118.34  | 0.06        | 116.55      | 116.74  | 0.16        |
| juli      | 115.88      | 117.70  | 117.62  | -0.06       | 116.46      | 116.36  | -0.08       |
| august    | 124.26      | 122.63  | 122.70  | 0.05        | 116.86      | 116.94  | 0.06        |
| september | 114.46      | 113.56  | 113.54  | -0.01       | 116.41      | 116.46  | 0.04        |
| oktober   | 113.18      | 115.45  | 115.47  | 0.01        | 114.18      | 114.03  | -0.13       |
| november  | 117.44      | 116.04  | 116.04  | 0.00        | 115.18      | 115.22  | 0.03        |
| desember  | 158.55      | 160.05  | 160.12  | 0.04        | 116.42      | 116.49  | 0.06        |

b. Sammenligning av de to sesongjusterte tallene  $A_t^d$  fra  $O_t^d$  og  $A_t^s$  fra  $O_t^s$ .  $O_t^d$  og  $O_t^s$  brukes som input for sesongjustering. Deres sesongjusterte tall betegnes med  $A_t^d$  og  $A_t^s$ . Relative avviket mellom dem er  $R_t^A$ . Tallene for disse størrelsene er vist i figur 14.3 og kolonnene 6, 7 og 8 i tabell 14.4 for desember 2000. Sammenligning med  $R_t^O$  er  $R_t^A$  litt større.

c. Estimering  $\delta_{i,t}^s$ , for  $t = \text{januar 1990}, \dots, \text{desember 2000}$

Resultater er vist i tabell 14.5. Vi ser at ved å bruke dataene i de siste 10 årene i estimeringen, er vekten for torsdag litt lavere fra 0.0075 (med data for hele tidsserien) til 0.0069. Men det er en økning for fredag fra 0.0023 til 0.0047 og lørdag fra 0.0057 til 0.0077. Dette viser at handelen i de siste årene har flyttet seg fra torsdag til fredag



Figur 14.4: Relative avvikene i prosent mellom sesongjusterte tall med estimeringen for ukedagseffekter fra X-12-ARIMA og Kalmanfilter

og lørdag. Merk at standardavvikene for  $\hat{\delta}_{i,t}^s$  (med data fra [jan.90-des.00]) blir større i gjennomsnitt enn de samme størrelsene med data fra hele tidsserien, siden det er færre observasjoner i beregningen.

Tabell 14.5: De estimerte verdiene for  $\delta_{i,t}^s$

|         | [jan.79 – des.00]  |                  | [jan.90 – des.00]  |                  |
|---------|--------------------|------------------|--------------------|------------------|
|         | $\hat{\delta}_i^s$ | $\overline{std}$ | $\bar{\delta}_i^s$ | $\overline{std}$ |
| mandag  | -0.0027            | 0.000128         | -0.0051            | 0.000176         |
| tirsdag | 0.0006             | 0.000127         | 0.0042             | 0.000171         |
| onsdag  | 0.0044             | 0.000129         | 0.0026             | 0.000173         |
| torsdag | 0.0075             | 0.000129         | 0.0069             | 0.000174         |
| fredag  | 0.0023             | 0.000128         | 0.0047             | 0.000173         |
| lørdag  | 0.0057             | 0.000129         | 0.0077             | 0.000173         |

#### d. En AR(1) modell for $\delta_{i,t}^s$

Ligningen (14.9) er en random walk modell for  $\delta_{i,t}^s$ . Nå vi antar at  $\delta_{i,t}^s$  følger en AR(1) modell. Dvs

$$\delta_{i,t}^s = \phi_{ii}\delta_{i,t-1}^s + \eta_{i,t} \quad (14.15)$$

der  $\phi_{ii} < 1$  og skal estimeres. Ligningen (14.11) blir

$$\begin{pmatrix} \delta_{1,t}^s \\ \delta_{2,t}^s \\ \vdots \\ \delta_{6,t}^s \end{pmatrix} = \begin{pmatrix} \phi_{11} & 0 & \dots & 0 \\ 0 & \phi_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \phi_{66} \end{pmatrix} \begin{pmatrix} \delta_{1,t-1}^s \\ \delta_{2,t-1}^s \\ \vdots \\ \delta_{6,t-1}^s \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \vdots \\ \eta_{6,t} \end{pmatrix} \quad (14.16)$$

Vi får

$$\begin{aligned} \hat{\phi}_{11} &= 0.9793, & \hat{\phi}_{22} &= 0.9643, & \hat{\phi}_{33} &= 0.9615, \\ \hat{\phi}_{44} &= 1.0000, & \hat{\phi}_{55} &= 0.9458, & \hat{\phi}_{66} &= 0.8167 \end{aligned}$$

Analysen viser at torsdag og lørdag har største vekter. Vi ser at bortsett fra  $\hat{\phi}_{66}$  er de andre  $\hat{\phi}_{ii}$  nesten 1 eller lik 1. Dermed kan vi anta at  $\phi_{ii} = 1$  for alle  $i$  som gir en enklere modell.

## 14.4 Oppsummering

I X-12-ARIMA estimerer vi ukedagseffekter ved å bruke en vanlig regresjonsmodell med restledd som følger en  $\text{ARIMA}(p, d, q)(P, D, Q)$ . I analysen må vi anta at vektene  $\delta_i^d$  for ukedager ikke endrer seg med tid, siden modellen er deterministisk. Denne antagelsen kan være urimelig for en lang tidsserie på 22 år, siden handlemønsteret har endret seg over tid. Ved å estimere  $\delta_{i,t}^s$  med en stokastisk modell får vi et godt bilde av vektene som varierer fra måned til måned. I analysen viser det at endringene av  $\hat{\delta}_{i,t}^s$  fra tidspunkt  $t$  til  $t + 1$  er veldig små slik at de betraktes som konstanter. Dermed har vi vist at vi kan bruke en deterministisk modell som beskrives i X-12-ARIMA for å estimere effektene av ukedager for detaljvolumindeksen for perioden 1979-2000.

# Bibliografi

- [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 ARI-MA*, 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

- |         |                                                                                                                                                                              |         |                                                                                                                                                                                                    |
|---------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2002/19 | F. Gundersen og L. Solheim: Regionalisering av FoU-statistikken. 43.s                                                                                                        | 2002/32 | T.M. Normann: Omnibusundersøkelsen februar/mars 2002. Dokumentasjonsrapport. 37s.                                                                                                                  |
| 2002/20 | L. Vågane: Omnibusundersøkelsen november/desember 2001. Dokumentasjonsrapport. 56s.                                                                                          | 2002/33 | S. Reid: Bosettingskriteriene i inntektssystemet til kommunene. Erfaringer med overgang til ny beregningsmåte og nye bosettingskriterier, 2002. 43s.                                               |
| 2002/21 | G. Claus, O. Haugen, P.M. Holt og E. Knutsen: Regnskapsstatistikk. Næringsoppgaver for ikke-finansielle aksjeselskaper, 1999. Dokumentasjon. 34s.                            | 2002/34 | K.E. Engebretsen, P.E. Gjedtjernet, S. Kristoffersen, P.G. Larssen og J.H. Wang: Mottak og tilrettelegging av SLN-data. 49s.                                                                       |
| 2002/22 | M. Takle: Befolkningsstatistikk på rute-nett. Dokumentasjon. 35s                                                                                                             | 2002/35 | D. Rafat: Analyse av sammenheng mellom ektefellers sysselsetting i en familie. 27s.                                                                                                                |
| 2002/23 | D. Roll-Hansen, S. Ferstad, M. Stålnacke, P. Tuhus og R. Nøtnæs: En spørre-skjemametodisk gjennomgang av datainnsamling gjennom Grunnskolens informasjonssystem (GSI). 109s. | 2002/36 | A. Bruvoll og T. Bye: En vurdering av avfallspolitikken bidrag til løsning av miljø - og ressursproblemer. 31s.                                                                                    |
| 2002/24 | T.P. Bøe og I. Håland: Dokumentasjon av arbeidskraftundersøkelsen (AKU). 85s.                                                                                                | 2002/37 | K.I. Bøe: B.R. Joneid: KOSTRA revisjonssystem. Malverk for generelt revisjonssystem - KOSTRA-data. Revidert utgave. 66s.                                                                           |
| 2002/25 | A. Akselsen og T. Sandnes: FD - Trygd: Dokumentasjonsrapport. Stønader til enslig forsørger. 1992-2000. 46s.                                                                 | 2002/38 | N. Arnesen, G. Daugstad, O.E. Hallingstad, E. Skretting Lunde og B.Vold: Kvalitetssikring i KOSTRA. Forslag til dokumentasjonsrutiner med erfaring fra FylkesKOSTRA-helsetjenester, somatikk. 54s. |
| 2002/26 | E. Rønning: Statistisk sentralbyrås tidsbruksundersøkelse 2000/01. Dokumentasjon og resultater fra intervjuet. 125s.                                                         | 2002/39 | H. Moafi: Omlegging av folkehøgskolestatistikk. Overgang til elektronisk rapportering. 31s.                                                                                                        |
| 2002/27 | S. Myro og C. Torp: Stedsfesting av bedrifter i Bedrifts- og foretaksregisteret. Hovedprosjekt. 37s.                                                                         | 2002/40 | Ø. Kleven: Mediebrukundersøkelsen 2001. Dokumentasjonsrapport. 43s.                                                                                                                                |
| 2002/28 | C. Nordseth og T. Sandnes: FD - Trygd: Dokumentasjonsrapport. Foreløpig uførestønad. 1992-2000. 37s.                                                                         | 2002/41 | Ø. Kleven: Samordnet levekårsundersøkelse 2000 - panelundersøkelsen. Dokumentasjonsrapport. 129s.                                                                                                  |
| 2002/29 | S. Derakhshanfar og T. Sandnes: FD - Trygd: Dokumentasjonsrapport. Økonomisk sosialhjelp. 1992-2000. 36s.                                                                    | 2002/42 | L. Solheim: Foreløpige tall i FoB2001 Utvalg, vekter, estimering og usikkerhet. 64s.                                                                                                               |
| 2002/30 | I. Johansen: Undersøking om foreldrebetaling i barnehagar, januar 2002. 42s.                                                                                                 | 2002/43 | A. Andersen, E. Birkeland, J. Epland og M. I. Kirkeberg: Økonomi og levekår for ulike grupper trygdemottakere 2001. Foreløpig rapport. 214s.                                                       |
| 2002/31 | T.M. Køber, H. Moafi, E. Rønning og Ø. Sivertstøl: Bruk av forløpsdatabaser i Statistisk sentralbyrå. 60s.                                                                   |         |                                                                                                                                                                                                    |

Statistisk sentralbyrå

*Oslo:*  
Postboks 8131 Dep.  
0033 Oslo

Telefon: 21 09 00 00  
Telefaks: 21 09 49 73

*Kongsvinger:*  
2225 Kongsvinger

Telefon: 62 88 50 00  
Telefaks: 62 88 50 30

ISSN 0806-3745