

Dinh Quang Pham

**Innføring i tidsserier –
sesongjustering og X-12-ARIMA**

Notater

Innhold

1	Kort om sesongjustering	3
1.1	Hva er en sesongjustering	3
1.2	Grunner som medfører til sesongvariasjoner i tidsserie	5
1.3	Modeller for sesongjustering	6
1.4	Litt historie	6
1.5	Metoder som brukes i Europa	8
1.6	Sesongjustering i SSB	9
1.7	Konfidensintervall	10
2	Sesongjustering for faktisk utførte investeringer, totalt for industrien	11
2.1	Innledning	11
2.2	Nye tidsserier	12
2.2.1	Å lage $Y_{1,t}, Y_{2,t}, \dots, Y_{n,t}$ ved å legge til gradvis nye observasjoner	12
2.2.2	Å lage $U_{1,t}, U_{2,t}, \dots, U_{n,t}$ ved å fjerne gradvis gamle observasjoner	14
2.2.3	Tidsserier med fast lengde	15
2.3	Resultater fra sesongjustering for tidsserien $\{O_{87.1}, O_{87.2}, \dots, O_{99.4}\}$	17
2.4	Simuleringsstudie	20
3	En innføring om begrepene i tidsserieanalyse	23
3.1	Innledning	23
3.2	Om begrepene i tidsserieanalyse	23
3.2.1	Tidsserie	23
3.2.2	Deterministisk funksjon	23
3.2.3	Stokastisk funksjon	24
3.2.4	Autokovariansfunksjon	24
3.2.5	Autokorrelasjonsfunksjon (ACF)	25
3.2.6	Stasjonæritet	26
3.2.7	Hvit støy prosess	26
3.2.8	Autoregressiv prosess av orden 1, AR(1)	26
3.2.9	Glidende gjennomsnitt prosess av orden 1, MA(1)	29
3.2.10	Blandet autoregressiv og glidende gjennomsnitt prosess, ARMA(1,1)	31
3.2.11	Autoregressiv prosess av orden 2, AR(2)	33
3.2.12	Glidende gjennomsnitt prosess av orden 2, MA(2)	35
3.2.13	Blandet autoregressiv og glidende gjennomsnitt prosess, ARMA(2,2)	36

3.2.14	En generell autoregressiv og glidende gjennomsnitt prosess, ARMA(p, q)	37
3.2.15	En generell lineær prosess	38
3.2.16	Ikke-stasjonær prosess. ARIMA(p, d, q)	39
3.2.17	Partiell autokorrelasjonfunksjon (PACF)	41
3.2.18	Anvendelse av autokorrelasjonfunksjon (ACF) og partiell autokorrelasjonfunksjon (PACF) i tidsserieanalyse	44
3.2.19	Estimering av autokorrelasjon (ACF) og partiell autokorrelasjonfunksjon (PACF) fra observasjoner	44
3.2.20	Framskrivning	45
3.2.21	Modell med sesongvariasjoner	49
3.2.22	Framskrivninger for modell med sesongvariasjoner	53
3.2.23	Akaike's AIC og BIC kriterier	55
3.2.24	Anvendelse av autokorrelasjonfunksjoner ved å finne orden til en ARIMA modell	56
3.2.25	Eksempel: Hotellovernattinger for nordmenn	56
3.3	Oppsummering	62
4	Sesongjustering	64
4.1	Innledning	64
4.2	Intervensjoner	64
4.2.1	Additive Outlier i t_0 (Et spesielt utslag for et bestemt tidspunkt t_0)	67
4.2.2	Level shift at t_0 (Nivåskift ved tidspunkt t_0)	68
4.2.3	Innovational Outlier (IO) (Brå endring med avtagende effekt)	68
4.3	Ukedagseffekter	70
4.4	Påskeeffekt	74
4.5	Trend	76
4.5.1	Trend og glidende gjennomsnitt teknikk	76
4.5.2	12 måneders sentrert glidende gjennomsnitt	77
4.5.3	13 måneders glidende gjennomsnitt	78
4.5.4	23 måneders glidende gjennomsnitt	79
4.5.5	En oppsummering for det sentrert 12 mnd-gj., 13 mnd-gj. og 23 mnd-gj.	79
4.5.6	13 måneders Hendersons glidende gjennomsnitt	79
4.5.7	23 måneders Hendersons glidende gjennomsnitt	81
4.5.8	Kommentarer for 13 og 23 måneders Hendersons glidende gjennomsnitt	82
4.5.9	Konklusjoner om trend	83
4.6	Sesongvariasjoner	83
4.6.1	Enkle metoder for å sjekke om sesongvariasjoner i rådataene	84
4.7	Sesongjusterte tall	85
4.8	Irregulærkomponent	85
4.9	Default beregninger i X-12-ARIMA	85
4.10	Diagnostiske prosedyrer for sesongjusteringer i X-12-ARIMA	90
4.11	Eksempel: Detaljvolumindeks for totalen	96

Figurer

1.1	<i>Rådata av detaljvolumindeksen</i>	3
1.2	<i>Sesongfaktorene</i>	4
1.3	<i>Rådata og sesongjusterte tall</i>	4
1.4	<i>Rådata og trend</i>	4
1.5	<i>Diagram av programmene</i>	9
2.1	<i>Rådata</i>	12
2.2	<i>Tidsseriene $Y_{1,t}, Y_{2,t}, \dots, Y_{n,t}$</i>	13
2.3	<i>Tidsseriene $U_{1,t}, U_{2,t}, \dots, U_{n,t}$</i>	14
2.4	<i>Tidsseriene $V_{1,t}, V_{2,t}, \dots, V_{n,t}$</i>	15
2.5	<i>Tidsseriene $W_{1,t}, W_{2,t}, \dots, W_{n,t}$</i>	16
2.6	<i>Histogrammet av revisjoner</i>	20
2.7	<i>Rådata, trend og sesongjusterte tall</i>	21
2.8	<i>Sesongkomponent</i>	22
2.9	<i>Irregulærkomponent</i>	22
3.1	<i>Simulerte data og autokorrelasjonen fra AR(1): $O_t = 0.9O_{t-1} + \epsilon_t$</i>	29
3.2	<i>x-y plott av (O_t, O_{t-1}) og (O_t, O_{t-10}) fra AR(1) prosess: $O_t = 0.9O_{t-1} + \epsilon_t$</i>	30
3.3	<i>Simulerte data og autokorrelasjonen fra MA(1): $O_t = \epsilon_t - 0.5\epsilon_{t-1}$</i>	30
3.4	<i>x-y plott av O_t, O_{t-1} og O_t, O_{t-2} fra MA(1) prosess: $O_t = \epsilon_t - 0.5\epsilon_{t-1}$</i>	31
3.5	<i>Simulerte data og autokorrelasjonen fra ARMA(1,1): $O_t = 0.9O_{t-1} + \epsilon_t - 0.5\epsilon_{t-1}$</i>	32
3.6	<i>x-y plott av (O_t, O_{t-1}) og (O_t, O_{t-2}). ARMA(1,1): $O_t = 0.9O_{t-1} + \epsilon_t - 0.5\epsilon_{t-1}$</i>	33
3.7	<i>Simulerte data og autokorrelasjonen fra AR(2): $O_t = 0.4O_{t-1} - 0.7O_{t-2} + \epsilon_t$</i>	34
3.8	<i>x-y plott av (O_t, O_{t-1}) og (O_t, O_{t-2}). AR(2) prosess: $O_t = 0.4O_{t-1} - 0.7O_{t-2} + \epsilon_t$</i>	34
3.9	<i>Simulerte data og autokorrelasjonen fra MA(2): $O_t = \epsilon_t - 0.65\epsilon_{t-1} - 0.24\epsilon_{t-2}$</i>	36
3.10	<i>x-y plott av O_t og O_{t-1}, O_t og O_{t-2} fra MA(2): $O_t = \epsilon_t - 0.65\epsilon_{t-1} - 0.24\epsilon_{t-2}$</i>	36
3.11	<i>Eksport av tradisjonelle varer (i millioner)</i>	39
3.12	<i>Simulerte dataene fra ARIMA(0,1,1): $O_t = O_{t-1} + \epsilon_t - 0.4\epsilon_{t-1}$</i>	41
3.13	<i>De estimerte ACF og PACF fra de simulerte dataene av AR(2) prosess: $O_t = 0.4O_{t-1} - 0.7O_{t-2} + \epsilon_t$</i>	45
3.14	<i>Detaljvolumindeks</i>	49
3.15	<i>Autokorrelasjonene for ARMA(0,1)(1,0)₁₂</i>	52
3.16	<i>Simulerte data for ARIMA(0,1,1)(0,1,1)₁₂, med $\theta = 0.8$ og $\Theta = 0.6$</i>	54

3.17	<i>Rådata av hotellovernattinger for nordmenn</i>	56
3.18	\widehat{ACF} og \widehat{PACF} av rådata	57
3.19	Y_t	58
3.20	\widehat{ACF} og \widehat{PACF} av Y_t	59
3.21	<i>Residualene</i>	60
3.22	\widehat{ACF} av residualene	61
3.23	<i>Tetthetsplot og normalplot av residualer</i>	61
3.24	<i>Rådata og predikerte verdier (framskrivinger)</i>	62
4.1	<i>Antall nye personbiler</i>	65
4.2	<i>En simulering for additive outlier</i>	67
4.3	<i>En simuleringer for level shift</i>	68
4.4	<i>En simuleringer for innovation outlier</i>	69
4.5	<i>Rådata og prekorrigerte data for nye personbiler</i>	71
4.6	$H(\tau, t)$ for mars og april 1999	76
4.7	<i>Detaljvolumindeks, sentrert 12 mnd glidende gjennomsnitt</i>	78
4.8	<i>Detaljvolumindeks, 13 mnd glidende gjennomsnitt</i>	78
4.9	<i>Detaljvolumindeks, 23 mnd glidende gjennomsnitt</i>	79
4.10	<i>Symmetriske vektorer</i>	80
4.11	<i>Asymmetriske vektorer for $t = N$</i>	81
4.12	<i>Detaljvolumindeks, 13 måneders Hendersons glidende gjennomsnitt</i>	81
4.13	<i>Detaljvolumindeks, 23 måneders Hendersons glidende gjennomsnitt</i>	82
4.14	<i>Sesongvariasjoner ved modell (4.41)</i>	84
4.15	<i>Rådataene av detaljvolumindekse gjennom årene</i>	84
4.16	<i>Spektraltettheten mot frekvenser</i>	85
4.17	<i>Estimerte autokorrelasjoner etter lag k</i>	86
4.18	<i>Diagrammet for X-12-ARIMA</i>	86
4.19	<i>Diagrammet for regARIMA</i>	87
4.20	<i>Prekorrigerte data og framskrivinger</i>	99
4.21	<i>Trenden og sesongjusterte tall</i>	100
4.22	<i>Sesongkomponent</i>	101
4.23	<i>Irregulær komponent</i>	101
4.24	<i>Revisjoner av sesongjusterte tall</i>	102

Tabeller

1.1	Rådata for detaljvolumindeksen for 1999	3
1.2	Sesongkomponent (i prosent), sesongjusterte tall og trend i 1999	5
2.1	Sesongjustering av $Y_{1,t}, Y_{2,t}, \dots, Y_{n,t}$	13
2.2	Sesongjustering av $U_{1,t}, U_{2,t}, \dots, U_{n,t}$	14
2.3	Sesongjustering av $V_{1,t}, V_{2,t}, \dots, V_{n,t}$	16
2.4	Sesongjustering for $W_{1,t}, W_{2,t}, \dots, W_{n,t}$	16
2.5	Resultater fra sesongjustering	17
2.6	Sesongjustering for simulerte data	21
3.1	ρ_k av AR(1) prosess: $O_t = 0.9O_{t-1} + \epsilon_t$	29
3.2	ρ_k av MA(1) prosess: $O_t = \epsilon_t - 0.5\epsilon_{t-1}$	31
3.3	ρ_k av ARMA(1,1) prosess: $O_t = 0.9O_{t-1} + \epsilon_t - 0.5\epsilon_{t-1}$	32
3.4	ρ_k av AR(2) prosess: $O_t = 0.4O_{t-1} - 0.7O_{t-2} + \epsilon_t$	34
3.5	ρ_k av MA(2) prosess: $O_t = \epsilon_t - 0.65\epsilon_{t-1} - 0.24\epsilon_{t-2}$	36
3.6	$\rho_k, \hat{\rho}_k, \phi_{kk}$ og $\hat{\phi}_{kk}$ av AR(2) prosess: $O_t = 0.4O_{t-1} - 0.7O_{t-2} + \epsilon_t$	46
3.7	Estimerte verdier for parametre av ARIMA(1,1,1)(1,1,1)	58
3.8	Estimerte verdier for parametre av ARIMA(1,1,1)(0,1,1)	59
3.9	Estimerte verdier for parametre av ARIMA(0,1,1)(1,1,1)	60
3.10	Estimerte verdier for parametre av ARIMA(0,1,1)(0,1,1)	60
3.11	Framskrivningene og observerte verdiene for år 2000	63
4.1	Intervensjoner av bilsalg	70
4.2	Originale dataene for detaljvolumindeks	72
4.3	$T_{i,t}$ for detaljvolumindeks	73
4.4	Estimerte verdiene for β_i	74
4.5	1. påskedag for 1970-2010	75
4.6	$H(\tau, t)$ i 1999	75
4.7	Detaljvolumindeks for 1999 og 2000	76
4.8	Hendersons vektorer for 13 måneders glidende gjennomsnitt	80
4.9	Hendersons vektorer for 13 måneders glidende gjennomsnitt	80
4.10	11 vektorer for 11 mål	92
4.11	11 modifiserte vektene	93
4.12	Evaluerer en sesongjustering basert på $S(\%), MM(\%)$ og $YY(\%)$	95
4.13	Ukedagseffekter	98
4.14	Framskrivninger fra august 2000 til juli 2001	99
4.15	Kvalitets mål	102

Innledning

Dette notatet beskriver begrepene i tidsserieanalyse som vi bruker i sesongjustering. Det inneholder fire kapitler:

- Kapittel 1: Kort om sesongjustering. Her beskriver vi litt om historien til sesongjustering og årsaker til sesongvariasjon i tidsserier, samt metoden som brukes i SSB.
- Kapittel 2: Sesongjustering for faktisk utførte investeringer, totalt for industrien. Vi presenterer i detalj et konkret problem for sesongjustering i SSB.
- Kapittel 3: En innføring om begrepene i tidsserieanalyse. Brukeren må ha litt kunnskap om tidsserieanalyse for å jobbe med sesongjustering, for eksempel, ARIMA modell, stasjonærhet, autokorrelasjonsfunksjon, beregningen av framskrivinger, osv. Disse begrepene beskrives i dette kapitlet.
- Kapittel 4: Sesongjustering. Vi presenterer metodene for å korrigere kalendereffekter, intervensjoner, samt teknikken for å spalte opp en tidsserie i trend, sesongjusterte tall, sesongkomponenten og den irregulære komponenten. De diagnostiske prosedyrene for sesongjustering i X-12-ARIMA er også beskrevet.

Vi bruker figurer for å forklare og illustrere resultater. Dataene er hentet fra SSB.

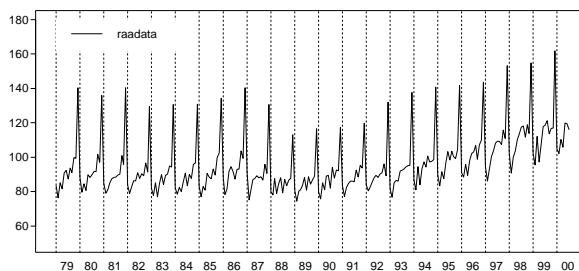
Vi vil takke Leiv Solheim for gode kommentarer.

Kapittel 1

Kort om sesongjustering

1.1 Hva er en sesongjustering

Vi vil beskrive begrepet ved et konkret eksempel istedet for å forklare med ord. Tidsserien er detaljvolumindeksen for totalen (SNN52.IVL.U). Tallene er observert fra januar 1979 til juli 2000. I alt er det 259 observasjoner. Rådataene er plottet i figur 1.1. Merk at en tidsserie¹ er en rekke av observasjoner på etterfølgende tidspunkter over en viss tid.



Figur 1.1: Rådata av detaljvolumindeksen

Vi ser at mønsteret av dataene gjentar seg fra år til år med de laveste verdiene i februar og høyeste i desember. Tallene for 1999 er gitt i tabell 1.1.

Tabell 1.1: Rådata for detaljvolumindeksen for 1999

Jan.	Feb.	Mar.	Apr.	Mai.	Jun.
Jul.	Aug.	Sep.	Okt.	Nov.	Des.
102.41	95.59	111.99	97.25	107.41	117.72
118.63	121.18	113.65	116.84	116.87	161.88

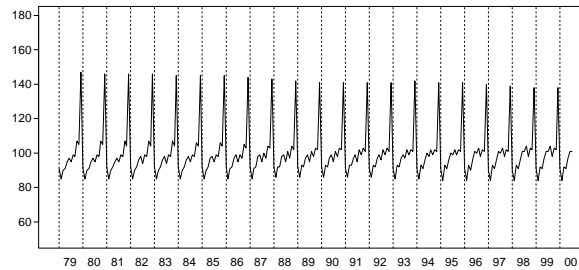
Hvis vi beregner endringen i prosent fra en måned til neste måned ved

$$e_{t,t+1} = \frac{O_{t+1} - O_t}{O_t} \times 100\%,$$

hvor O_t er observasjonen i tidspunkt t , så ser vi store endringer i rådata fra november til desember, 38,51%. Dette skyldes at effekten av sesong har påvirket da-

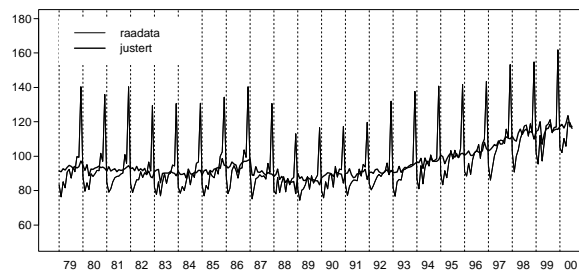
¹se 3.2.1 for mer detaljert

ta. Det blir derfor vanskelig å sammenligne aktiviteten fra måned til måned. Høye verdier i desember betyr ikke at aktiviteten er høyere enn de andre månedene, men julehandelen i desember øker det totale omsetningsvolumet kraftig. Figur 1.2 viser *sesongfaktorene* som beregnes ved X-12-ARIMA. Den tidsserien som dannes av sesongfaktorene kalles *sesongkomponenten*.

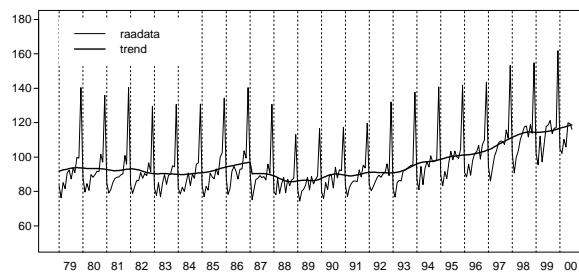


Figur 1.2: *Sesongfaktorene*

Rådata etter å ha justert for sesongeffektene (samt spesielle kalendereffekter og ekstremverdier) kalles *sesongjusterte tall* og vises i figur 1.3. Nå kan vi sammenligne aktiviteten fra måned til måned ved sesongjusterte tallene uten å bli forstyrret av sesongvariasjoner.



Figur 1.3: *Rådata og sesongjusterte tall*



Figur 1.4: *Rådata og trend*

Fluktuasjoner i den sesongjusterte tidsserien er bare tilfeldige og kan ikke forklares ved bestemte årsaker. Senere kaller vi denne tidsserien for den *irregulære komponenten*. Ved å fjerne den irregulære komponenten fra de sesongjusterte tallene får vi en tidsserie som kalles *trenden* (se figur 1.4). Denne viser den langsigtige utviklingen av data. Dersom rådataene har store tilfeldige variasjoner, er

trenden bedre egnet som informasjon enn den sesongjusterte serien. Trenden gir ofte mindre feilaktige konklusjoner for den historiske tallserien. Sesongfaktorene, trend og de sesongjusterte tallene for 1999 er gitt i tabell 1.2.

Tabell 1.2: *Sesongkomponent (i prosent), sesongjusterte tall og trend i 1999*

måned	Se.faktor	Se.just.	Trend
januar	90.55	114.38	114.85
februar	83.64	115.31	114.91
mars	92.59	120.14	114.97
april	91.41	105.93	115.01
mai	95.56	114.27	115.06
juni	101.58	115.60	115.11
juli	100.82	115.84	115.18
august	103.88	118.93	115.28
september	97.67	114.28	115.42
oktober	102.49	115.30	115.62
november	101.58	115.82	115.89
desember	137.88	115.34	116.21

Vi ser at etter å ha justert for sesongeffekter har november og desember omtrent samme nivå, med 115.82 og 115.34, hvor sesongfaktorene for disse månedene er 101.58 og 137.88, henholdsvis.

Hver måned publiserer vi sesongjusterte tall og trend for arbeidsledige etter aldersgrupper. Det vil bli lettere for brukeren å vurdere et tiltak eller se langsiktige variasjoner av arbeidsledigheten ved sesongjusterte og trend enn å bruke rådataene.

1.2 Grunner som medfører til sesongvariasjoner i tidsserie

- **Kalendereffekter.** Faste helligdager som 17. mai, jul (25. desember) og varierende helligdager som påske (1. påske i 1999 var 4. april og i 2000 23. april) eller pinse påvirker mange økonomiske tidsserier. Ukedagseffekter (trading-day effects, på engelsk) som skyldes ulike antall ukedager i samme måned fra år til år, er også en faktor til som bidrar til sesongvariasjoner.

- **Fast periode bestemt av kalenderen.** For eksempel, arbeidsledige blant ungdom øker i feriene. AKU tall viser at arbeidsledigheten er høy i juni, juli og august for aldersgruppene: 16-19 år og 20-24 år. Sysselsetting er totalt lavere i denne perioden på grunn av ferieavvikling.

- **Været.** Endringene av temperatur, nedbør påvirker direkte eller indirekte økonomiske tidsserier som f.eks landsbruksproduksjon, bygg, transport.

- **Forventning av sesongvariasjoner,** som påvirker for planlegging eller investering i en fabrikk. For eksempel, man forventet at salget av fyringsolje ble redusert 20% i forhold til det normale i desember dette året, på grunn av mildvær.

Ofte er sesongvariasjoner en blanding av disse fire eller flere effekter som er avhengig av dataene vi observerer. En sesongkomponent beskrives som regelmessige variasjoner som gjentar seg på samme tidspunkt hvert år, med nesten samme størrelse og kan forklares og estimeres. I mange økonomiske tidsserier er man bare interessert i sesongjusterte tall og trend. Men for planlegging er sesongkomponenten en viktig størrelse. For eksempel, en bedrift må vite om sesongvariasjoner for å gjøre investeringene.

1.3 Modeller for sesongjustering

I sesongjustering beskriver vi observerte data O_t , sesongkomponenten S_t , trenden T_t og den irregulære komponenten I_t ved formlene

$$O_t = S_t + T_t + I_t \quad \text{for additiv modell} \quad (1.1)$$

$$O_t = S_t \times T_t \times I_t \quad \text{for multiplikativ modell} \quad (1.2)$$

Sesongjusterte tall A_t beregnes ved

$$A_t = O_t - S_t \quad \text{for additiv modell} \quad (1.3)$$

$$A_t = O_t/S_t \quad \text{for multiplikativ modell} \quad (1.4)$$

Merk at vi skriver O_t ved (1.1) og (1.2) under antakelsen at det er ingen prekorrigering for kalendereffekter eller ekstremverdier i rådataene. Vi skal beskrive modellene for sesongjustering mer detaljert i neste kapittel. Figurene 1.2-1.4 viser sesongkomponenten S_t , sesongjusterte tall A_t og trenden T_t som beregnes fra de observerte dataene O_t ved (1.2).

1.4 Litt historie

1954. X-0 (**X** står for e**X**perimental) det første elektroniske dataprogram for sesongjustering som ble kalt "Method I" og utviklet av Shiskin. Programmet er basert på metoden som kalles "ratio-to-moving average" av Macaulay på 1920-tallet.

1955. "Method I" ble erstattet av "Method II" med forbedringer. Forskjellige varianter av "Method II" kom etter hvert.

1960. X-3 den første varianten som ble tilgjengelig for offentligheten. De senere versjonene X-4,...,X-8 kom siden.

1961. X-9 og X-10 brukes for serier med store variasjoner.

1965. X-11 fra US Census (Shiskin, Young og Musgrave).

Fordelene med denne versjonen er:

- Forbedring av asymmetriske vekter i endepunktene.
- Metoden mer robust, skyldes korrigeringsrutinen for ekstreme verdier.
- Ukedageffekter blir estimert.
- Flere filtre for sesongkomponent.

Ulemper:

- Fortsatt lav kvalitet for asymmetriske filtre i endepunktene.

1980. X-11-ARIMA fra Statistics Canada (Dagum)

Fordelene med denne versjonen er:

- Modellering av originale data med en tilpasset ARIMA modell. Tre innebygde modeller: $(0\ 1\ 1)(0\ 1\ 1)_s$, $(0\ 2\ 2)(0\ 1\ 1)_s$ og $(2\ 1\ 2)(0\ 1\ 1)_s$, med $s = 12$ for månedstall og $s = 4$ for kvartalstall.
- Mulig å forlenge serien for å kunne bruke symmetrisk filter i endepunktene til trenden. Revisjonene blir derfor mindre.
- Innføring av det sentrerte 24 måneders glidende gjennomsnittet. Filteret gir bedre tilpasning
- for seriene som har syklus mindre enn 3 år eller brå endringer i trenden.

1988. X-11-ARIMA fra Statistics Canada (Dagum)

Viktige forbedringer:

- To ARIMA modeller i tillegg: $(0\ 1\ 2)(0\ 1\ 1)_s$ og $(2\ 1\ 0)(0\ 1\ 1)_s$
- Nye kriterier for tilpasning av ARIMA modell og rådata.
- Effekter av ukedager og påske blir automatisk fjernet fra rådata før en tilpasser en ARIMA modell.
- Korrigering for påskeeffekt (metoden var ikke innført i den tidligere versjonen).
- Økt nøyaktighet for Hendersons filtre.
- Mulig å velge automatisk lengden til sesongfilter.
- Korrigere rådata når de er for små eller for høye.
- Innførte M og Q statistikker for å måle hvor god sesongjusteringen er.

Ulemper med X-11-ARIMA/80 og X-11-ARIMA/88

- Ukedageeffekter og påskeeffekt er estimert indirekte fra et foreløpig anslag av den irregulære komponenten.
- Behandling av ekstreme verdier er ikke god nok.

1994. X-12-ARIMA presenteres av US Census

I forhold til X-11-ARIMA er programmet forbedret med:

- Trend kan estimeres med $(2n + 1)$ -ledd Henderson filteret.
- (3×5) filter er innført for sesongkomponenten.
- Pseudo-additive modell $O_t = T_t \times (S_t + I_t - 1)$, der O_t , T_t , S_t og I_t står for observerte data, trend, sesong- og irregulær komponent.

- regARIMA modell (linear regression model with ARIMA times series errors) brukes for å estimere ukedagseffekter, påskeeffekt og intervensjoner.
- Har bedre framskrivinger.
- Nye diagnostikker for kvaliteten av sesongjustering (spektralanalyse, revisjon, slidingspans).
- Bedre forklaringer med mange eksempler i manualen.

1.5 Metoder som brukes i Europa

- DAINTRIES utviklet i 1979. Det var offisielt program for sesongjustering i European Commission.
- SABL Utviklet i Bell Laboratories.
- The Berlin method BV4 utviklet ved Technische Universitat Berlin og Deutsche Institut fur Wirtschaftsforschung (DIW).
- X-11 familien (X-11, X-11-ARIMA, X-12-ARIMA).
- TRAMO/SEATS utviklet av Maravall og Gomez (1986).
- STAMP utviklet av Koopman, Harvey, Doornik og Shephard.
- SsfPack utviklet av Koopman, Shephard og Doornik.

Figur 1.5 viser diagrammet for programmene. Man deler metodene i to grupper som kalles:

- Implisitt modell med X-11/X-12-ARIMA og SABL. I beregningen trenger man ingen ARIMA modell for trenden og sesongkomponenten.
- Eksplisitt modell, her estimerer man trend og sesongkomponent ved ARIMA modeller. I STAMP er modellene angitt på forhånd. For eksempel,

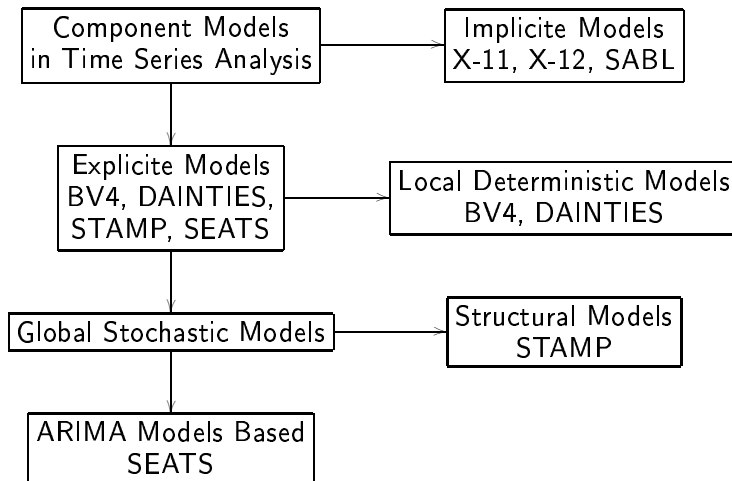
$$\begin{aligned} O_t &= T_t + S_t + I_t \\ T_t &= T_{t-1} + \beta_{t-1} + \eta_t, & \eta_t &\sim NID(0, \sigma_\eta^2), \\ \beta_t &= \beta_{t-1} + \zeta_t, & \zeta_t &\sim NID(0, \sigma_\zeta^2), \end{aligned}$$

T_t , S_t og I_t er trend, sesong- og irregulær komponent, henholdsvis. I_t , η_t og ζ_t are ukorrelerede hverandre. En modell for sesongkomponenten skrives slik

$$S_t = -S_{t-1} - \dots - S_{t-s+1} + \omega_t, \quad \omega_t \sim NID(0, \sigma_\omega^2)$$

mens for TRAMO/SEATS er modellene for T_t og S_t plukket ut fra den ARIMA modellen som bestemmes av rådata.

I denne gruppen finner man BV4, DAINTRIES, STAMP og TRAMO/SEATS. De fire programmene deles i to, den ene er *lokale deterministiske modeller* med BV4 og DAINTRIES og den andre er *globale stokastiske modeller* med STAMP for *strukturelle modeller* og TRAMO/SEATS for *ARIMA baserte modeller*.



Figur 1.5: Diagram av programmene

Forkortelser

TRAMO (Time Series Regression with ARIMA Noise Missing Observations and Outliers)

SEATS (Signal Extraction in ARIMA Time Series)

STAMP (Structural Time Series Alyser, Modeller and Predictor)

SABL (Seasonal Aadjustment at Bell Laboratories)

Begrepene som ARIMA modell, filter, glidende gjennomsnitt, ukedagseffekter, påskeeffekt, M , Q statistikker, Henderson filter, ... skal beskrives i detaljert i de neste kapitlene.

1.6 Sesongjustering i SSB

X-11 og X-11-ARIMA har vært brukt i SSB for sesongjustering siden 1980, i seksjonene:

- Økonomiske indikatorer (240).
- Arbeidsmarked (260).
- Utenrikshandel, energi og industristatistikk (230).
- Samferdsels og reiselivsstatistikk (440).
- Bygg og tjenestestatistikk (460).
- Makroøkonomi (530).

Programmet ble kjørt på stormaskinen. Det finnes også en versjon for PC fra omkring 1987, men den er lite i bruk, siden vi har mye data og overskrider kapasiteten til PC og dessuten er det en tung jobb å overføre tidsseriene fra stormaskin til PC og omvendt. FAME (Forecasting, Analysis and Modeling Environment) for tidsserier blir innført i SSB på UNIX plattform (1992). X-11-ARIMA har en egen versjon som er tilpasset for FAME databasen på UNIX. Fordelen med FAME er at det blir mye lettere for brukeren å hente data, sesongjustere dem og legge tilbake trend og sesongjusterte tall i FAME databasen. Den samme rutinen på stormaskin er svært omfattende. Fra 1999 bruker vi X-12-ARIMA som et offisielt program for sesongjustering i SSB, basert på sammenligning med resultater fra sesongjustering med X-11-ARIMA. Vi har bidratt med mange kommentarer

for å forbedre programmet (X-12-ARIMA). Vi har tusenvis av tidsserier som skal sesongjusteres hver måned. I beregningen lar vi ofte programmet selv estimere kalendereffekter og ekstremverdier samt lengden av filter for sesongvariasjoner. For de tidsseriene som har uregelmessige variasjoner plukker vi dem ut og behandler dem mer nøyaktig. Det kan være å velge en annen ARIMA modell enn en av de fem innebygde modellene for programmet eller justere terskelen for estimering av ekstremverdier eller velge et annet filter for sesongkomponenten. Resultater fra sesongjustering kan bli bedre. Det er et annet program for sesongjustering som heter STAMP og brukes i forskningavdelingen. Det krever større kunnskapen om tidsserier. X12-ARIMA kjøres for månedstall og kvartalstall, men ikke for termintall. Derfor vil vi erstatte med SABL eller TRAMO/SEATS. SABL kan kjøres fra Myfame. Dessuten har vi DEMETRA som er et vindubasert program, der kan man utføre både X-12-ARIMA og TRAMO/SEATS med input data fra Excel og FAME databasen.

Vi har følgende adresser på internett

X-12-ARIMA	ftp.census.gov/pub/ts/x12a
TRAMO/SEATS	http://www.bde.es/servisio/software/econom.htm
DEMETRA	http://forum.europa.eu.int/Public/irc/dsis/eurosam/library
STAMP	http://www.timberlake.co.uk/software/Stamp/Stamp.htm

1.7 Konfidensintervall

Metoden i X-12-ARIMA er basert på det glidende gjennomsnittet med Hendersons filtre for å beregne sesongjusterte tall og trend. Derfor får vi ikke usikkerheter for disse komponentene. Vi publiserer bare rådataene, sesongjusterte tall og trend uten konfidensintervaller. Men for modellbaserte metoder som SEATS, STAMP kan vi beregne usikkerhetene for hver komponent og deretter lage 95% konfidensintervaller. Selv om TRAMO/SEATS er i bruk i mange institusjoner er det ingen land som publiserer tall med konfidensintervaller.

Kapittel 2

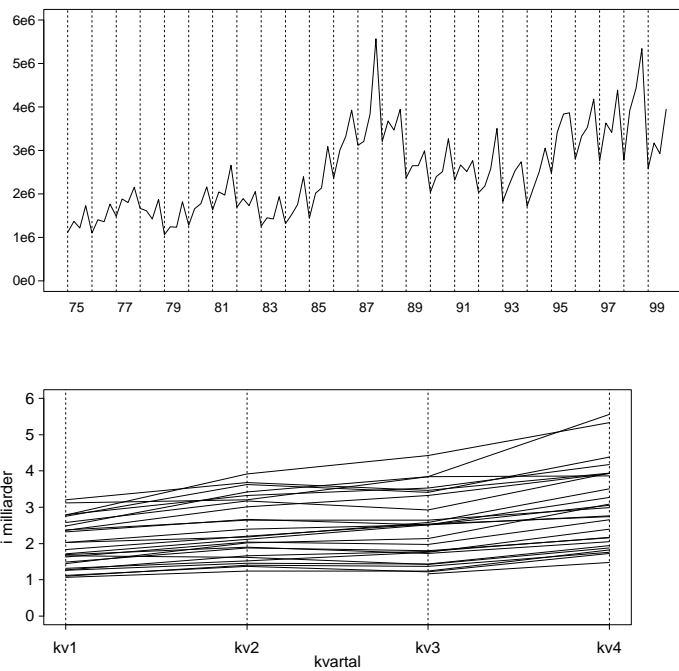
Sesongjustering for faktisk utførte investeringer, totalt for industrien

2.1 Innledning

Vi sesongjusterer tidsserier med X-12-ARIMA fra FAME database. Der har vi laget et standard opplegg som inneholder kommandoer for beregningene. Opplegget er tilpasset for de fleste tidsseriene, men det er noen serier som krever en mer nøyaktig analyse enn å la programmet behandle det selv. De utførte investeringene er et slikt eksempel. Vi får dårlig sesongjustering ved standard opplegget, i den forstand at revisjonene av sesongjusterte tall er store. Tidsserien går fra første kvartal 1975 til fjerde kvartal 1999 og plottet i figur 2.1. Den første delen viser en oversikt av variasjoner i rådataene i perioden fra 1975 til 1999. Mens den andre delen gir oss et godt bilde av sesongvariasjoner. Vi ser at sesongmønsteret er ganske tydelig. Det er et godt tegn for sesongjustering. For tidsserier med et uklart sesongmønster vil vi ikke sesongjustere disse. Grunnen til en dårlig sesongjustering for de utførte investeringene kan skyldes vi har justert en lang tidsserie som går over 25 år, hvor observasjoner endrer seg rask fra periode til periode. Vi har følgende problemer:

- Det er ikke realistisk å anta at effektene av ukedager er de samme for hele serien slik at regresjonsanalyse kan anvendes. Dette gjelder også for påskeeffekt.
- Det blir vanskelig å identifisere en ARIMA modell for dataene. Dermed får vi ikke estimert for ekstremverdier.
- Trend og sesongkomponenten kan ikke skilles effektivt fra den irregulære komponenten. Sesongjusteringen blir dårlig kvalitet.

For å kunne unngå disse problemene må vi finne en metode for å redusere store variasjoner fra periode til periode i rådataene. En enkel måte er å lage en kortere tidsserie ved å kutte ut de første observasjonene fra dataene. Variasjoner er mer regelmessige for en kortere tidsserie. En ARIMA modell blir lettere å identifiseres. I de neste avsnittene presenterer vi metoder for å lage nye tidsserier som har mindre utslag enn de opprinnelige dataene. Resultater fra sesongjustering av disse tidsseriene skal sammenlignes.



Figur 2.1: Rådata

2.2 Nye tidsserier

For å kunne estimere kalendereffekter og intervensjoner lager vi ikke tidsserier hvor lengden er mindre enn 13 år.

2.2.1 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:

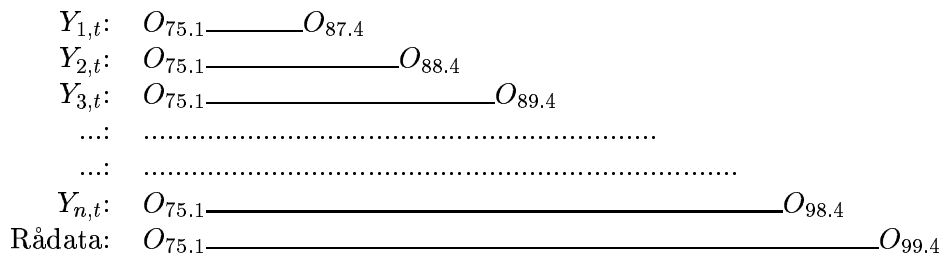
1. Den første serien $Y_{1,t}$ inneholder observasjonene fra 1975 til 1987, dvs,

$$Y_{1,t} = \{O_{1kv.75}, O_{2kv.75}, \dots, O_{4kv.87}\}.$$

Lengden er lang nok til å estimere ARIMA parametre og kalendereffekter.

2. Den neste serien $Y_{2,t}$ er et år lengre enn $Y_{1,t}$, ved å legge til observasjonene for 1988. Vi får $Y_{2,t} = \{O_{1kv.75}, \dots, O_{4kv.87}, O_{1kv.88}, \dots, O_{4kv.88}\}$.
3. Vi gjør på samme måte for $Y_{3,t}, Y_{4,t}, \dots$. 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 2.2.

Resultater fra X-12-ARIMA for $Y_{1,t}, Y_{2,t}, \dots$, er listet ut i tabell 2.1.

Figur 2.2: Tidsseriene $Y_{1,t}$, $Y_{2,t}$, ..., $Y_{n,t}$ Tabell 2.1: Sesongjustering av $Y_{1,t}$, $Y_{2,t}$, ..., $Y_{n,t}$

Tids.	ARIMA	Effekt		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
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

Inter. står for intervensjoner

Vi innfører symbolene:

- Q er veiet gjennomsnitt av 11 kvalitetsmål som presenteres i appendikset. Grenseverdien for Q er 1.
- $S(\%)$ er prosentandelen av kvartaler hvor sesongkomponenten er definert som upålitelige.
- $MM(\%)$ er prosentandelen av kvartaler hvor endringene av sesongjusterte tall identifiseres som upålitelige.

$S(\%)$ og $MM(\%)$ 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 tilfredsstillter kriterier fra X-12-ARIMA for de fleste tidsseriene ellers tilpasses dataene med kompliserte modeller. Lave verdiene av Q , $S(\%)$ og $MM(\%)$ viser stabile sesongjusterte tall, bortsett fra [75-93] og [75-99].

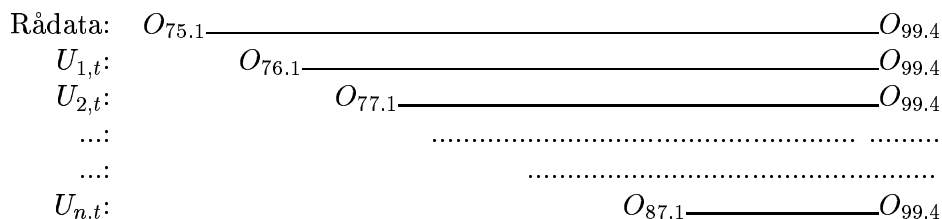
2.2.2 Å lage $U_{1,t}$, $U_{2,t}$, ..., $U_{n,t}$ ved å fjerne gradvis gamle observasjoner

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:

1. 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, $U_{1,t} = \{O_{76.1}, O_{76.2}, \dots, O_{99.4}\}$.
2. Fra $U_{1,t}$ lager vi $U_{2,t}$ ved å fjerne tallene i 1976. Vi får $U_{2,t} = \{O_{77.1}, O_{77.2}, \dots, O_{99.4}\}$.
3. 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 parametre av ARIMA modell.

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 2.3.

Figur 2.3: Tidsseriere $U_{1,t}$, $U_{2,t}$, ..., $U_{n,t}$



Resultater fra X-12-ARIMA av $U_{1,t}$, $U_{2,t}$, ..., $U_{n,t}$ er vist i tabell 2.2.

Tabell 2.2: 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
81-99	nei	nei	ja	0.261	27.8	48.6	nei
82-99	nei	nei	ja	0.279	33.3	57.1	nei

(Tabell 2.2 fortsatt)

Tids.	ARIMA	Effekt			Sliding. i %		
		Uke.	Påske	Q	$S(\%)$	$MM(\%)$	Inter.
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

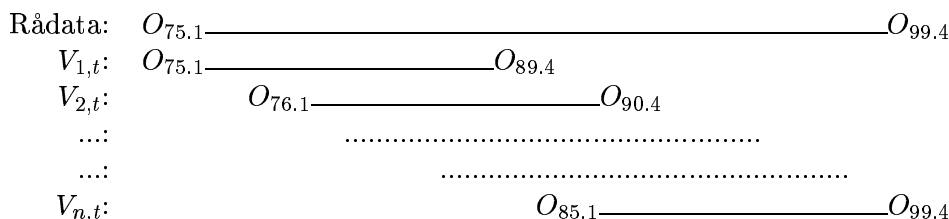
Vi ser at når dataene før 1983 blir fjernet får vi bedre resultater for sesongjustering. Det finnes ARIMA modeller for dataene, kalendereffekter er signifikante på 5% nivå og sesongkomponenten og sesongjusterte tall er ganske stabile slik at tallene kan publiseres. Ved å forlenge tidsserier tilbake blir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

2.2.3 Tidsserier med fast lengde

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

a. Å lage $V_{1,t}, V_{2,t}, \dots, V_{n,t}$ med 15 års periodelengde. 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 et år senere. Vi får $V_{2,t} = \{O_{76.1}, O_{75.2}, \dots, O_{90.4}\}$. Dvs, når nye observasjoner legges til, blir det tilsvarende antallet av de tidligste verdiene fjernet. Vi gjør på samme måte for $V_{3,t}, \dots, V_{n,t}$. Da får vi en rekke av tidsseriene $V_{1,t}, V_{2,t}, \dots, V_{n,t}$, hvor $V_{n,t} = \{O_{85.1}, O_{85.2}, \dots, O_{99.4}\}$. Metoden illustreres i figur 2.4. Denne teknikken gir oss et bilde av stabiliteten til sesongjusterte tall.

Figur 2.4: Tidsseriene $V_{1,t}, V_{2,t}, \dots, V_{n,t}$



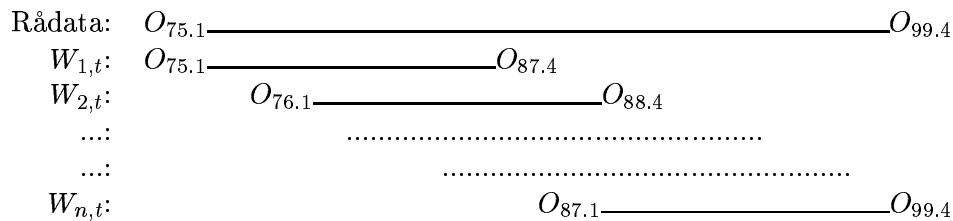
Resultater fra X-12-ARIMA av $V_{1,t}, V_{2,t}, \dots, V_{n,t}$ presenteres i tabell 2.3.

Tabellene 2.2 og 2.3 viser to rekker av tidsseriene $\{U_{1,t}, U_{2,t}, \dots, U_{n,t}\}$ og $\{V_{1,t}, V_{2,t}, \dots, V_{n,t}\}$. Tidsseriene som har samme startpunkter, men lengdene er ikke de samme. Vi ser at ARIMA modeller tilpasser data bedre og sesongjusterte tall er mer stabile (lavere verdiene av $S(\%)$ og $MM(\%)$) når tidsserier 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 periodelengde. Vi lager tidsseriene med samme prinsippet som vi har brukt i 2.2.3.a. Tidsseriene $W_{1,t}, W_{2,t}, \dots, W_{n,t}$ er beskrevet i figur 2.5. Resultater fra X-12-ARIMA er vist i tabell 2.4.

Tabell 2.3: *Sesongjustering av $V_{1,t}, V_{2,t}, \dots, V_{n,t}$*

Tids.	ARIMA	Effekt		Sliding. i %			Inter.
		Uke.	Påske	Q	$S(\%)$	$MM(\%)$	
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
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

Figur 2.5: *Tidsseriene $W_{1,t}, W_{2,t}, \dots, W_{n,t}$* Tabell 2.4: *Sesongjustering for $W_{1,t}, W_{2,t}, \dots, W_{n,t}$*

Tids.	ARIMA	Effekt		\hat{d}	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
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
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

I tabell 2.4 er \hat{d} 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 2.3) 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 ukedagseffektene i sesongjustering får vi høye verdier av $S(\%)$ og $MM(\%)$ som viser ustabile sesongjusterte tall. Dette vises i tabell 2.5.

Tabell 2.5: *Resultater fra sesongjustering*

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 2.1-2.5 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 de beste sesongjusterte tallene for tidsserien W som inneholder observasjonene fra 1987 til 1999.

2.3 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

Testen viser at multiplikativ modell passer bedre med dataene. Dvs, O_t sesongdekomponeres på formen

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

O_t og T_t har samme nivå, mens S_t og I_t beregnes i prosent.

- ARIMA modell¹

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

¹se kapittel 3 for forklaringer

Gjennomsnittlig absolutt prosent feil innen framskrivinger:

siste år: 22.46, forrige år: 9.56, to år tidligere: 5.96
siste tre år: 12.66. Grenseverdien for siste tre år er 15%.

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 (normalfordelte feil). Dermed kan resultater være ugyldige. Findley og Monsell forbedret svakheter i X-11-ARIMA ved å anvende en ARIMA modell for rådata i estimeringen av kalendereffekter og intervensjoner. Dermed er antagelsen om en hvit støy prosess av den foreløpige irregulære komponenten ikke nødvendig. Uten en 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 fra rådata, kan man angi en ARIMA modell på forhånd for programmet. Men ofte blir det ikke bedre.

- Kalendereffekter¹

Ukedageeffekter skyldes ulike antall dager i samme måned fra år til år. Testen viser at effektene har påvirket rådataene på 5% nivå ($\chi^2 = 51.76$ med 6 frihetsgrader).

Påskeeffekt er effekten som påvirker rådataene på grunn av påske. De er helligdager som flytter mellom mars og april, avhengig av året. Testen viser at påskeeffekten er signifikant på 5% nivå ($t = -2.77$).

De er to effekter som bidrar til sesongmønsteret og må fjernes fra rådataene før sesongjustering. Vi har ikke gode forklaringer for ukedageeffekter for de utførte investeringer.

- Intervensjoner¹

Det er spesielle bevingheter som påvirker rådataene, for eksempel streiken i en fabrikk. Effektene kan bidra til endringer i trend og sesongjusterte tall og må fjernes før sesongjustering. Det finnes en additiv outlier i tredje kvartal 1999.

- Framskrivinger¹

beregnes fra den ARIMA modellen som passer for dataene. Hvis tilpassningen er god får vi mindre variasjoner for framskrivingene og omvendt. Tabellen nedenfor viser beregningen med og uten prekorrigeringen for kalendereffekter og intervensjoner.

Dato	Rådata uten prekor.			med prekor.
	Nedre	Fram.	Øvre	Fram.
1.kv 2000	2240250	2650947	3136936	2570422
2.kv 2000	2335554	3034602	3942879	3103875
3.kv 2000	2358943	3281891	4565946	3321410
4.kv 2000	2509877	3708207	5478674	3910986

Vi ser at 95% konfidensintervallene er store. Merk at X-12-ARIMA lister ikke ut konfidensintervaller for de framskrivingene som beregnes av rådata med prekorrigeringen.

- Tester for sesongmønster og sesongbevegelser¹

F -testene viser at sesongmønsteret er signifikant på 5% nivå, $F=455.84$. Men det er ikke signifikant for sesongbevegelser på samme nivå, $F=0.442$. En sesongjustering blir dårlig hvis det er klare sesongbevegelser, siden det da blir vanskelig å identifisere sesongmønsteret.

- Kvalitetsmål og sliding spans¹

brukes for å vurdere hvor god en sesongjustering er. De 11 målene $M1$ - $M11$ og Q (veiet gjennomsnitt av $M1$ - $M11$) er gitt i nedenfor tabell.

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

Lave verdier (mindre enn 1) av M_i og Q viser 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 en god sesongjustering.

- Revisjoner¹

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 revisjon i prosent for sesongjusterte tall,

- for hvert kvartal, er:

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

- for hvert år gjennom fire kvartaler, er:

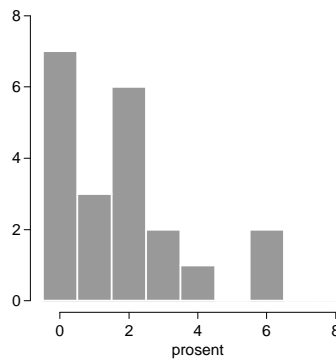
$$\bar{R}_{1995} = 2.49, \quad \bar{R}_{1996} = 2.37, \quad \bar{R}_{1997} = 1.79, \quad \bar{R}_{1998} = 1.37 \text{ og } \bar{R}_{1999} = 1.39.$$

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

Merk at

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

Histogrammet er gitt ved figur 2.6.



Figur 2.6: Histogrammet av revisjoner

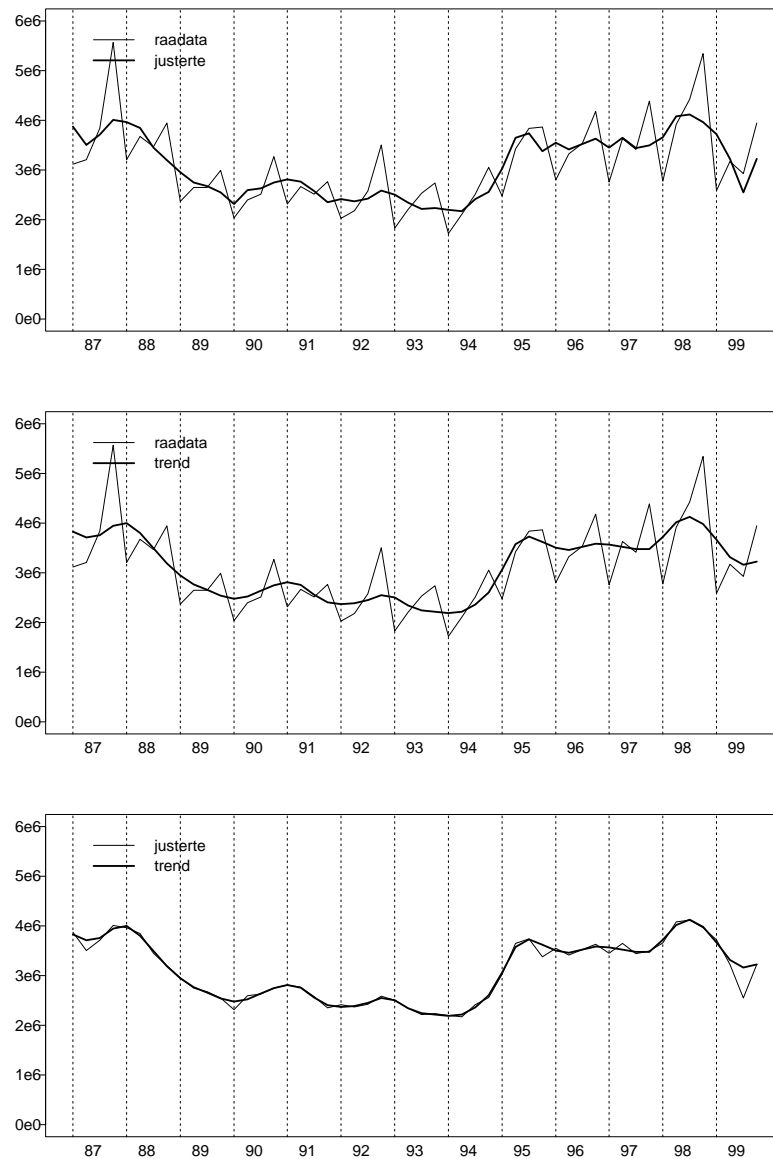
- Figurer

Rådata, trend og sesongjusterte tall er plottet i figur 2.7. Sesongkomponenten S_t og den irregulære komponenten I_t er plottet i figur 2.8 og 2.9. Siden variasjoner av den irregulære komponenten er ganske lave (bortsett fra noen tidspunkter), viser dette at avvikene mellom trenden og de sesongjusterte tallene er små (bortsett fra noen tidspunkter).

2.4 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.

$$\begin{array}{l} W_{n,t}: \quad O_{87.1} \text{-----} O_{99.4} \\ \text{Sim. :} \quad \quad O_{88.1} \text{-----} O_{99.4}, \hat{O}_{2000.1}, \hat{O}_{2000.2}, \hat{O}_{2000.3}, \hat{O}_{2000.4} \end{array}$$



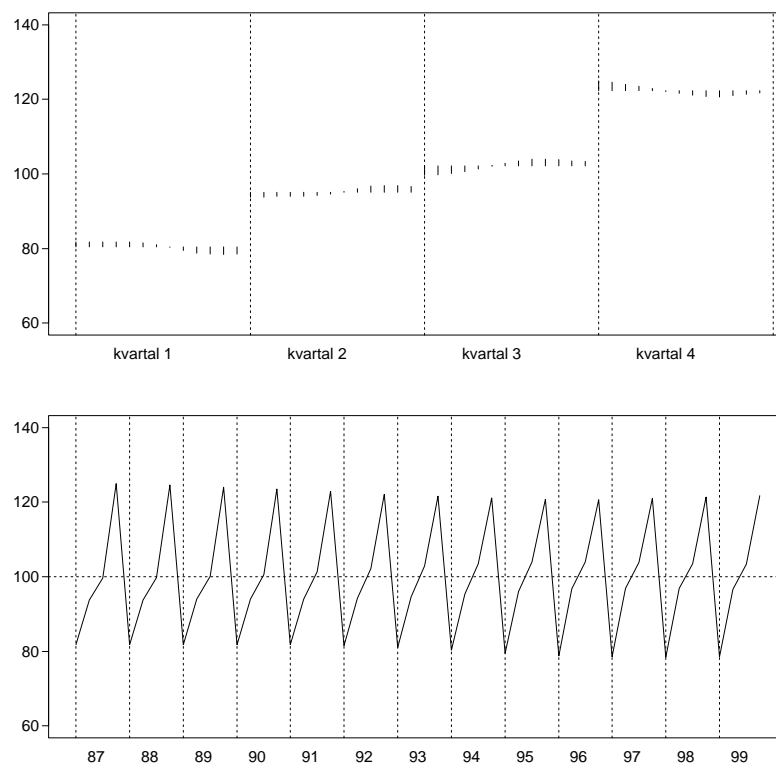
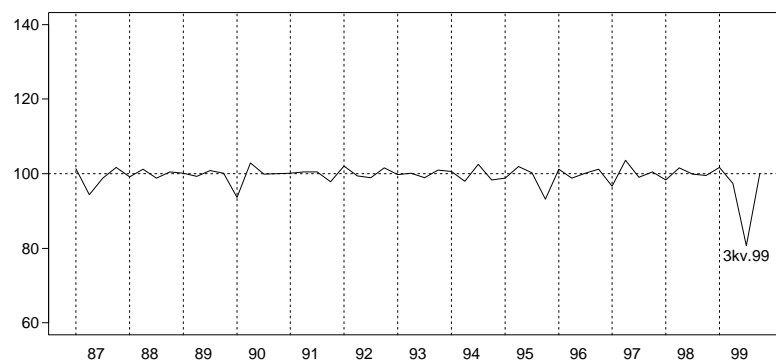
Figur 2.7: *Rådata, trend og sesongjusterte tall*

Ved å angi på forhånd den additive outlieren AO1999.3 og estimere kalenderefekter fra dataene, får vi

Tabell 2.6: *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 verdier av Q , $S(\%)$ og $MM(\%)$ som viser en god sesongjustering. Sammenlignet med tabell 2.4 ser vi at ARIMA modellen ikke endrer

Figur 2.8: *Sesongkomponent*Figur 2.9: *Irregulærkomponent*

seg. Kalendereffektene er fortsatt signifikante på 5% nivå og intervensjonen har forekommet i samme tidspunkt.

Kapittel 3

En innføring om begrepene i tidsserieanalyse

3.1 Innledning

I kapitlene 1 og 2 har vi brukt mange faglige uttrykk som *glidende gjennomsnitt*, *ARIMA modell*, *autokorrelasjon*, *ukedagseffekter* osv... uten nærmere forklaringer. For brukere som ikke kjenner tidsserieanalyse fra før kan dette være vanskelig å forstå. I dette kapitlet vil vi gi en gjennomgang av begrepene som brukes i sesongjustering.

3.2 Om begrepene i tidsserieanalyse

3.2.1 Tidsserie

De utførte investeringene i kapittel 2 er en tidsserie, siden vi observerer tallene for etterfølgende kvartaler over en viss periode. La O_t være en observasjon i tidspunkt t . Vi beskriver de utførte investeringene ved en rekke

$$O_{75.1}, O_{75.2}, O_{75.3}, O_{75.4}, O_{76.1}, \dots, O_{99.4},$$

hvor $O_{75.1}$ er investeringen i første kvartal 1975. Tidsserien er observert fra første kvartal 1975 til fjerde kvartal 1999. I alt er det 100 observasjoner. Ofte plotter man tallene i to dimensjoner for å få et bilde av variasjonene i tidsserien, hvor x -aksen viser tidsforløp og y -aksen viser verdiene av O_t (se figur 2.1). Antall nye personbiler som blir solgt hver måned danner også en tidsserie, hvor verdiene er summen av det daglige salget.

3.2.2 Deterministisk funksjon

La x være en varierende størrelse verdi, en funksjon av x , betegnes med $f(x)$, er en foranderlig størrelses tallverdi som er avhengig av x . En deterministisk funksjon er en funksjon av x , slik at for hver verdi av x , har vi en regel som kan beregne presist verdien av $f(x)$. Ofte er en slik regel gitt ved en matematisk formel, for eksempel, $f(x) = 2x$. La $f(x)$ være arealet til en kvadrat med lengde x . Da er $f(x) = x^2$. For hver verdi av x , har vi en verdi for $f(x)$. Derfor sier vi at $f(x) = x^2$ er en deterministisk funksjon av x .

3.2.3 Stokastisk funksjon

Vi tar et eksempel til, nå er situasjonen den at for hver x kan vi få litt forskjellige verdier til $f(x)$. Lengden l og høyden h av et prosjektil etter utskyting fra en kanon med utskytings hastigheten v_0 og stigningsvinkelen α fra punktet (l_0, h_0) er gitt ved

$$\begin{aligned} l(t) &= (v_0 \cos\alpha)t + l_0 \\ h(t) &= -\frac{1}{2}gt^2 + (v_0 \sin\alpha)t + h_0 \end{aligned}$$

Dersom vi antar at det er ingen usikkerhet for v_0 og α , vil en da, i teorien, i et bestemt tidspunkt t ha samme verdien for $l(t)$ og $h(t)$ under flere utskytinger. Men i praksis, vil verdien for $l(t)$ og $h(t)$ i den første skytingen og den andre være forskjellige. Grunnen er at banen til prosjektilet påvirkes også av for eksempel, luftmotstand, fuktighet, vindhastighet... Disse komponentene forandres seg fra tidspunkt til tidspunkt og det er veldig komplisert å beregne dem, slik at vi vil istedet beskrive dem som usikre. Derfor skriver vi $l(t)$ og $h(t)$ ovenfor på følgende måte

$$\begin{aligned} l(t) &= (v_0 \cos\alpha)t + l_0 + \epsilon_t \\ h(t) &= -\frac{1}{2}gt^2 + (v_0 \sin\alpha)t + h_0 + \epsilon_t \end{aligned}$$

hvor ϵ_t presenter en tilfeldig variasjon. Vi ser at det er umulig å beregne presise verdier til $l(t)$ og $h(t)$, siden vi ikke kjenner ϵ_t . Derfor sier vi at for hver t er $l(t)$ og $h(t)$ to *stokastiske variable* og når t varierer, er $l(t)$ og $h(t)$ *stokastiske prosesser* eller *stokastiske funksjoner*. Kort sagt, en stokastisk prosess er en mengde av tilfeldige variabler, hvor variablene følger en sannsynlighetsfordeling.

3.2.4 Autokovariansfunksjon

Her bruker vi litt statistiske formler for forklaringer. I en tidsserie antar vi at det er en statistisk avhengighet mellom observasjonene på forskjellige tidspunkter. Vi antar også at for hver $t \in T$ er det en sannsynlighetsfordeling $F_t(o)$ for O_t , hvor O_t er en stokastisk variabel. Vi skriver

$$F_t(o) = P(O_t < o)$$

og for s, t

$$F_{s,t}(o_1, o_2) = P(O_s < o_1, O_t < o_2)$$

Antar at $\{O_t\}$, $t = 0, \pm 1, \pm 2, \dots$, er en stokastisk prosess. Autokovariansfunksjonen mellom O_t og O_s er definert som

$$\gamma_{t,s} = Cov(O_t, O_s), \quad \text{for } t, s = 0, \pm 1, \pm 2, \dots \quad (3.1)$$

hvor

$$Cov(O_t, O_s) = E[(O_t - \mu_t)(O_s - \mu_s)] = E(O_t O_s) - \mu_t \mu_s$$

E står for forventning og μ_t er forventningen av O_t , $\mu_t = E(O_t)$.

Vi har følgende egenskaper

$$\begin{aligned} \gamma_{t,t} &= Var(O_t) \\ \gamma_{t,s} &= \gamma_{s,t} \\ |\gamma_{t,s}| &\leq \sqrt{\gamma_{t,t} \gamma_{s,s}} \end{aligned}$$

3.2.5 Autokorrelasjonsfunksjon (ACF)

Autokorrelasjonen mellom O_t og O_s er definert slik

$$\rho_{t,s} = \text{Corr}(O_t, O_s) \quad \text{for} \quad t, s = 0, \pm 1, \pm 2, \dots \quad (3.2)$$

hvor

$$\text{Corr}(O_t, O_s) = \frac{\text{Cov}(O_t, O_s)}{[\text{Var}(O_t)\text{Var}(O_s)]^{1/2}} = \frac{\gamma_{t,s}}{(\gamma_{t,t}\gamma_{s,s})^{1/2}}$$

Vi har det følgende egenskaper

$$\rho_{t,t} = 1, \quad (3.3)$$

$$\rho_{t,s} = \rho_{s,t}, \quad (3.4)$$

$$|\rho_{t,s}| \leq 1 \quad (3.5)$$

Vi ser at $\gamma_{t,s}$ og $\rho_{t,s}$ er to mål. $\gamma_{t,s}$ skal måle kovariansen og $\rho_{t,s}$ skal måle hvor stor korrelasjon og retning (positiv eller negativ) mellom de to stokastiske variablene O_t og O_s . Ofte bruker en heller autokorrelasjonsfunksjonen enn autokovariansfunksjonen, fordi den er lettere å tolke. Når $\rho_{t,s}$ er nær ± 1 indikerer det en sterk avhengighet mellom O_t og O_s , mens når $\rho_{t,s}$ er nær 0 er det liten sammenheng mellom O_t og O_s . Hvis $\rho_{t,s} = 0$ sier vi at O_t og O_s er ukorrelert og hvis O_t og O_s er uavhengige er $\rho_{t,s} = 0$.

Eksempel. La a_1, a_2, \dots , være uavhengige tilfeldige variable med samme fordelingen

$$a_t = \begin{cases} 1 & \text{vinner} \\ -1 & \text{tapper} \end{cases}$$

og $P(a_t = 1) = 0.5$. Vi antar videre at $E(a_t) = 0$ og $\text{Var}(a_t) = \sigma_a^2$ for alle t . Vi definerer en stokastisk prosess $\{Z_t\}$ ved

$$\begin{aligned} Z_1 &= a_1, \\ Z_2 &= a_1 + a_2, \\ \dots &= \dots \\ Z_t &= a_1 + a_2 + \dots + a_t \end{aligned}$$

Forventningen og variansen for Z_t blir da

$$E(Z_t) = E(a_1 + a_2 + \dots + a_t) = E(a_1) + E(a_2) + \dots + E(a_t) = 0$$

fordi $E(a_i) = 0$ for alle i , og

$$\text{Var}(Z_t) = \text{Var}(a_1) + \text{Var}(a_2) + \dots + \text{Var}(a_t) = t\sigma_a^2$$

siden a_1, a_2, \dots er uavhengige. Det viser seg for $1 \leq s \leq t$, autokorrelasjonen mellom Z_s og Z_t er

$$\rho_{s,t} = \sqrt{\frac{s}{t}}$$

Vi får

$$\rho_{1,2} = \sqrt{\frac{1}{2}} = 0.707, \quad \rho_{1,3} = \sqrt{\frac{1}{3}} = 0.577, \quad \rho_{1,4} = \sqrt{\frac{1}{4}} = 0.500, \quad \rho_{1,5} = \sqrt{\frac{1}{5}} = 0.447$$

og

$$\rho_{1,2} = \sqrt{\frac{1}{2}} = 0.707, \quad \rho_{2,3} = \sqrt{\frac{2}{3}} = 0.816, \quad \rho_{3,4} = \sqrt{\frac{3}{4}} = 0.866, \quad \rho_{4,5} = \sqrt{\frac{4}{5}} = 0.894$$

Vi ser at korrelasjonen mellom Z_s og Z_t blir mindre og mindre når s og t ligger langt hverandre. Når s og t er store går $\rho_{s,t}$ mot 1 og $Var(Z_t)$ er en voksende funksjon med t . Dette viser at prosessen har en tendens til å gå langt fra forventningen.

3.2.6 Stasjonæritet

Stasjonæritet beskrives ved at de statistiske egenskaper ved en prosess ikke endrer seg over tid. En svak form av stasjonæritet krever bare at (i) forventningen, (ii) variansen og (iii) autokorrelasjonsfunksjonen er konstant over tid. I praksis, hvis en prosess er stasjonær så kan vi estimere forventningen ved å ta gjennomsnittet av alle observasjonene. Men dette kan vi ikke gjøre for prosesser som ikke er stasjonære, fordi vi har egentlig bare en observasjon for hver tidsintervall. Autokorrelasjonen ρ_k av tidsserien $\{O_t\}$, $t = 1, \dots, n$ estimeres ved

$$r_k = \frac{\sum_{t=k+1}^n (O_t - \bar{O})(O_{t-k} - \bar{O})}{\sum_{t=1}^n (O_t - \bar{O})^2}, \quad \text{for } k = 0, 1, 2, \dots$$

her ser vi at hvis $\{O_t\}$ ikke er stasjonær da blir \bar{O} meningsløs, siden $\{O_t\}$ ikke er samme fordelingen, derfor kan vi ikke beregne $\sum O_t$. Vi sier at når $\{O_t\}$ er en stasjonær prosess, så er det samme kovariansen for de to vilkårlige stokastiske variable som er separert av k tidspunkter. Vi kan skrive

$$\gamma_k = Cov(O_t, O_{t-k}) \quad \text{og} \quad \rho_k = Corr(O_t, O_{t-k})$$

3.2.7 Hvit støy prosess

En prosess $\{\epsilon_t\}$ kalles hvit støy prosess hvis $\epsilon_t, t = 0, \pm 1, \pm 2, \dots$ er uavhengige tilfeldige variable med samme fordelingen. Det er lett å bevise at $\{\epsilon_t\}$ er en stasjonær prosess. Ofte antar vi at $\{\epsilon_t\}$ har forventning 0 og varians σ_ϵ^2 . Det kan vise at

$$\gamma_k = Cov(\epsilon_t, \epsilon_{t+k}) = \begin{cases} 0 & \text{for } k \neq 0, \\ \sigma_\epsilon^2 & \text{for } k = 0 \end{cases}$$

og

$$\rho_k = \begin{cases} 0 & \text{for } k = \pm 1, \pm 2, \dots, \\ 1 & \text{for } k = 0 \end{cases}$$

3.2.8 Autoregressiv prosess av orden 1, AR(1)

$\{O_t\}$ er en autoregressiv prosess med orden 1, hvis O_t skrives på form

$$O_t = \phi O_{t-1} + \epsilon_t \tag{3.6}$$

hvor ϕ er konstant og $\{\epsilon_t\}$ er en hvit støy prosess med forventning μ_ϵ og varians σ_ϵ^2 . Vi antar også at ϵ_t er uavhengig med $O_{t-1}, O_{t-2}, O_{t-3}, \dots$, d.v.s ϵ_t kan ikke forklares ved observasjonene i for tiden. Ligningen (3.6) beskriver "one-step dependence" og tolkes slik: verdien i tidspunkt t er bare avhengig av verdien i

tidspunkt $t - 1$ via en konstant ϕ og en støy ledd ϵ_t . Denne ligningen er en enkel regresjonsligning, hvor O_{t-1} og ϵ_t spiller som en forklaringsvariabel og feilledd. Vi vet at antall arbeidsledige i måned t er avhengig av antall arbeidsledigheter i forrige måned $t - 1$. La O_t være antall arbeidsledige i måned t , så (3.6) er en enkel modell for O_t . I følge (3.6) har vi

$$O_{t-1} = \phi O_{t-2} + \epsilon_{t-1}$$

Herav

$$\begin{aligned} O_t &= \phi(\phi O_{t-1} + \epsilon_{t-1}) + \epsilon_t \\ &= \epsilon_t + \phi \epsilon_{t-1} + \phi^2 O_{t-2} \\ &= \dots \\ &= \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \dots + \phi^{t-1} O_1 \\ &= \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \dots + \phi^{t-1} \epsilon_1 \\ O_t &= \sum_{s=1}^t \phi^{t-s} \epsilon_s \end{aligned} \tag{3.7}$$

siden $O_1 = \phi O_0 + \epsilon_1$ og vi antar at $O_0 = 0$.

Forventningen for O_t er

$$E(O_t) = \mu_\epsilon (1 + \phi + \dots + \phi^{t-1}) = \begin{cases} \mu_\epsilon \left(\frac{1 - \phi^t}{1 - \phi} \right) & \text{for } \phi \neq 1 \\ \mu_\epsilon t & \text{for } \phi = 1 \end{cases}$$

Hvis $\mu_\epsilon \neq 0$ er O_t en funksjon av t , dvs $\{O_t\}$ er ikke en stasjonær prosess.

Hvis $\mu_\epsilon = 0$, da blir $E(O_t) = 0$, en konstant over tid. $\{O_t\}$ er en stasjonær prosess med orden 1.

Antar at $|\phi| < 1$ og t er stor nok, da blir

$$E(Y_t) \approx \frac{\mu_\epsilon}{1 - \phi} \quad \text{uavhengig av } t$$

Vi sier at $\{O_t\}$ er en tilnærmet stasjonær prosess med orden 1.

Variansen til O_t er

$$\begin{aligned} \text{Var}(O_t) &= \text{Var}(\epsilon_t + \phi \epsilon_{t-1} + \dots + \phi^{t-1} \epsilon_1) \\ &= \begin{cases} \sigma_\epsilon^2 \left(\frac{1 - \phi^{2t}}{1 - \phi^2} \right) & \text{for } \phi \neq 1 \\ \sigma_\epsilon^2 t & \text{for } \phi = 1 \end{cases} \end{aligned}$$

for $\phi = 1$, $\text{Var}(O_t)$ er en funksjon av t . For $\phi < 1$ og t er stor nok

$$\text{Var}(O_t) \approx \frac{\sigma_\epsilon^2}{1 - \phi^2}$$

Vi sier at $\{O_t\}$ er en tilnærmet stasjonær prosess med orden 2.

Vi ser at når $|\phi| < 1$ er forventningen og variansen av O_t uavhengige av t . $\{O_t\}$ er en stasjonær prosess. Egentlig er denne konklusjonen ikke gyldig, siden prosessen $\{O_t\}$ er avhengig av startverdi, $O_0 = 0$. Dette medfører at egenskapen stasjonærhet ikke er oppfylt. Vi kan ordne dette ved å anta at $\{O_t\}$ starter fra

et tidspunkt som er ganske langt tilbake etter $t = 0$, dvs $O_{-N} = 0$ og $N \rightarrow \infty$. O_t i (3.7) skrives slik

$$O_t = \sum_{s=-(N-1)}^t \phi^{t-s} \epsilon_s \quad (3.8)$$

La $N \rightarrow \infty$, vi får

$$O_t = \sum_{s=-\infty}^t \phi^{t-s} \epsilon_s = \sum_{s=0}^{\infty} \phi^s \epsilon_{t-s} \quad (3.9)$$

og

$$\text{Var}(O_t) = \text{Var}\left(\sum_{s=0}^{\infty} \phi^s \epsilon_{t-s}\right) = \left(\sum_{s=0}^{\infty} \phi^{2s}\right) \sigma_{\epsilon}^2 = \frac{\sigma_{\epsilon}^2}{1 - \phi^2} < \infty$$

når $|\phi| < 1$. Med denne betingelsen er O_t i (3.9) en stasjonær prosess.

Vi innfører her en notasjon B som er "lagoperator" og definert ved $BO_t = O_{t-1}$. Ligningen (3.6) kan skrives slik

$$(1 - \phi B)O_t = \epsilon_t$$

eller

$$O_t = \frac{1}{1 - \phi B} \epsilon_t \quad (3.10)$$

Vi ser at når $|\phi| < 1$ er $\{O_t\}$ en stasjonær prosess. Denne betingelsen er ekvivalent med absolutt verdi av roten i ligningen $\phi(B) = 1 - \phi B = 0$ er større enn 1.

Vi bruker følgende form for beregningen av kovariansen

$$\text{Cov}(aX + bY + c, Z) = a\text{Cov}(X, Z) + b\text{Cov}(Y, Z) \quad (3.11)$$

hvor X , Y og Z er tilfeldige variablene, med $E(X^2) < \infty$, $E(Y^2) < \infty$ og $E(Z^2) < \infty$. a og b er konstanter.

Vi antar nå at $|\phi| < 1$ for at $\{O_t\}$ er en stasjonær prosess. Vi har

$$\begin{aligned} \gamma_0 = \text{Var}(O_t) &= \text{Cov}(O_t, O_t) = \text{Cov}(\phi O_{t-1} + \epsilon_t, \phi O_{t-1} + \epsilon_t) \\ &= \phi^2 \text{Cov}(O_{t-1}, O_{t-1}) + 2\text{Cov}(O_{t-1}, \epsilon_t) + \text{Cov}(\epsilon_t, \epsilon_t) \\ &= \phi^2 \gamma_0 + \sigma_{\epsilon}^2, \quad \text{siden } \epsilon_t \text{ er uavhengig av } O_{t-1}. \end{aligned}$$

Vi får

$$\gamma_0 = \frac{\sigma_{\epsilon}^2}{1 - \phi^2}$$

For $k \geq 1$

$$\begin{aligned} \gamma_k &= \text{Cov}(O_t, O_{t-k}) = \text{Cov}(\phi O_{t-1} + \epsilon_t, O_{t-k}) \\ &= \phi \text{Cov}(O_{t-1}, O_{t-k}) + \text{Cov}(\epsilon_t, O_{t-k}) \\ &= \phi \gamma_{k-1} = \phi^2 \gamma_{k-2} = \dots \\ &= \phi^k \gamma_0 \end{aligned}$$

Dermed er

$$\rho_k = \phi^k, \quad \text{for } k = 0, 1, 2, \dots$$

Siden $|\phi| < 1$, er ρ_k en avtagende funksjon med k .

- Hvis $\phi > 0$, er ρ_k en positiv avtagende funksjon med k , den går fortere mot null når ϕ er nær 0 og omvendt.

- Hvis $\phi < 0$ skifter ρ_k fortegn og den går fortere mot 0 når $|\phi|$ er nær null.

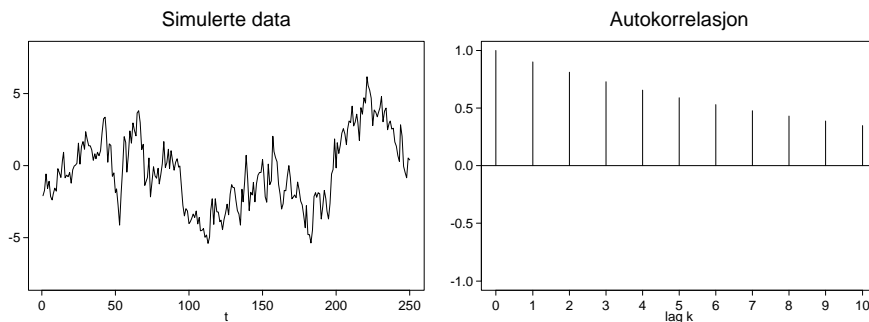
Figur 3.1 viser simulerte dataene fra en AR(1), med ligningen $O_t = 0.9O_{t-1} + \epsilon_t$, hvor ϵ_t er uavhengige og identisk normalfordelte, $N(0, 1)$. Vi ser at autokorrelasjonen er en avtagende funksjon med "lag" k . Den går langsomt mot null siden ϕ er nær 1 ($a = 0.9$).

Tabell 3.1 viser autokorrelasjonen ρ_k for de 10 første lagene, som er plottet i figur 3.1.

Tabell 3.1: ρ_k av AR(1) prosess: $O_t = 0.9O_{t-1} + \epsilon_t$

lag	0	1	2	3	4	5	6	7	8	9	10
ρ_k	1.00	0.90	0.81	0.73	0.66	0.59	0.53	0.48	0.43	0.39	0.35

For å undersøke korrelasjonene mellom O_t og O_{t-1} , O_t og O_{t-10} er de i figur 3.2 plottet mot hverandre. De punktene i plottet til venstre ligger litt tettere langs en rett linje enn i plottet til høyre som antyder en høyere korrelasjon mellom O_t og O_{t-1} enn O_t og O_{t-10} .



Figur 3.1: Simulerte data og autokorrelasjonen fra AR(1): $O_t = 0.9O_{t-1} + \epsilon_t$

3.2.9 Glidende gjennomsnitt prosess av orden 1, MA(1)

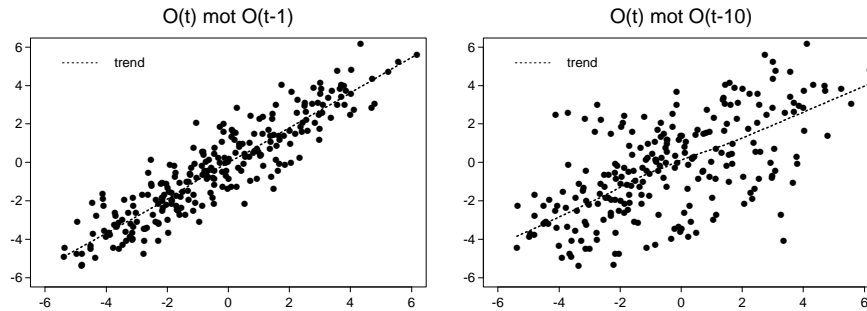
$\{O_t\}$ kalles en glidende gjennomsnitt prosess av orden 1, hvis

$$O_t = \epsilon_t - \theta\epsilon_{t-1} \quad (3.12)$$

hvor ϵ_t er en hvit støy prosess med forventning 0 og varians σ_ϵ^2 .

Forventningen for O_t er

$$E(O_t) = E(\epsilon_t - \theta\epsilon_{t-1}) = 0$$



Figur 3.2: x - y plott av (O_t, O_{t-1}) og (O_t, O_{t-10}) fra $AR(1)$ prosess: $O_t = 0.9O_{t-1} + \epsilon_t$

og variansen for O_t er

$$\begin{aligned} \text{Var}(O_t) &= \text{Var}(\epsilon_t - \theta\epsilon_{t-1}) \\ &= \text{Var}(\epsilon_t) + \theta^2\text{Var}(\epsilon_{t-1}) \\ &= (1 + \theta^2)\sigma_\epsilon^2 \end{aligned}$$

Kovariansen mellom O_t og O_{t-1} er

$$\begin{aligned} \gamma_1 &= \text{Cov}(O_t, O_{t-1}) = \text{Cov}(\epsilon_t - \theta\epsilon_{t-1}, \epsilon_{t-1} - \theta\epsilon_{t-2}) \\ &= \text{Cov}(-\theta\epsilon_{t-1}, \epsilon_{t-1}) = -\theta\sigma_\epsilon^2 \end{aligned}$$

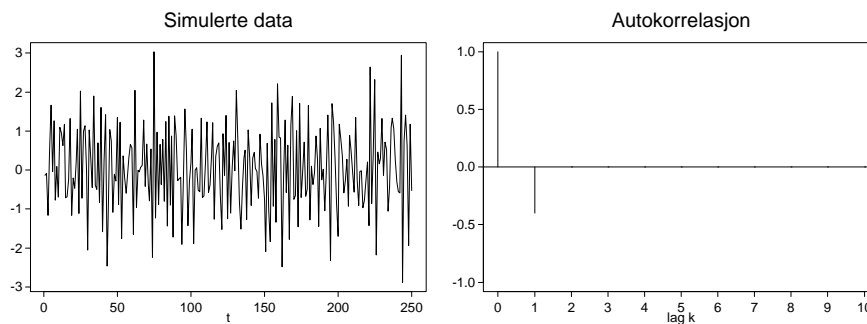
Kovariansen mellom O_t og O_{t-k} , for $k > 1$, er

$$\begin{aligned} \gamma_2 &= \text{Cov}(O_t, O_{t-k}) = \text{Cov}(\epsilon_t - \theta\epsilon_{t-1}, \epsilon_{t-k} - \theta\epsilon_{t-k-1}) \\ &= 0 \end{aligned}$$

Vi får

$$\rho_0 = 1, \quad \rho_1 = \frac{-\theta}{1 + \theta^2}, \quad \text{og} \quad \rho_k = 0 \quad \text{for } k = 2, 3, \dots$$

Herav kan vi si at det er ingen korrelasjon mellom to observasjoner når avstanden i to tidspunkter er større enn 1. $MA(1)$ er alltid en stasjonær prosess.



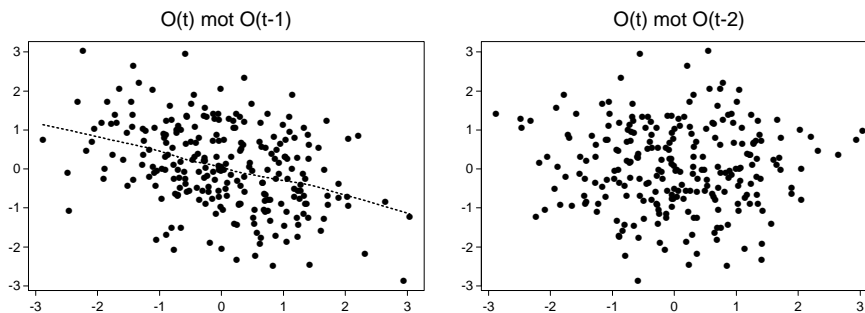
Figur 3.3: *Simulerte data og autokorrelasjonen fra $MA(1)$: $O_t = \epsilon_t - 0.5\epsilon_{t-1}$*

Figur 3.3 viser 250 observasjoner som er simulert fra en $MA(1)$, $O_t = \epsilon_t - 0.5\epsilon_{t-1}$. ρ_k av de 10 første lagene er vist i tabell 3.2.

Figur 3.4 viser korrelasjonene mellom O_t og O_{t-1} , O_t og O_{t-2} . Vi ser at i plottet til venstre ligger punktene litt mer langs en rett linje enn i plottet til høyre som antyder en viss korrelasjon mellom O_t og O_{t-1} , $\rho_1 = -0.4$ og ingen korrelasjon mellom O_t og O_{t-2} , $\rho_2 = 0$.

Tabell 3.2: ρ_k av MA(1) prosess: $O_t = \epsilon_t - 0.5\epsilon_{t-1}$

lag	0	1	2	3	4	5	6	7	8	9	10
ρ_k	1.0	-0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Figur 3.4: x - y plott av O_t, O_{t-1} og O_t, O_{t-2} fra MA(1) prosess: $O_t = \epsilon_t - 0.5\epsilon_{t-1}$

3.2.10 Blandet autoregressiv og glidende gjennomsnitt prosess, ARMA(1,1)

$\{O_t\}$ er en ARMA(1,1) prosess, hvis O_t skrives på form

$$O_t = \phi O_{t-1} + \epsilon_t - \theta \epsilon_{t-1} \quad (3.13)$$

hvor ϕ og θ er konstanter, $\{\epsilon_t\}$ er hvit støy prosess med forventning 0 og varians σ_ϵ^2 . Dessuten er ϵ_t er uavhengige med O_{t-1}, O_{t-2}, \dots . Støyledd i høyre side av ligningen skrives som en glidende gjennomsnitt prosess.

Variansen for O_t er

$$\begin{aligned} \gamma_0 = \text{Var}(O_t) &= \text{Cov}(O_t, O_t) = \text{Cov}(\phi O_{t-1} + \epsilon_t - \theta \epsilon_{t-1}, O_t) \\ &= \phi \text{Cov}(O_{t-1}, O_t) + \text{Cov}(\epsilon_t, O_t) - \theta \text{Cov}(\epsilon_{t-1}, O_t) \end{aligned} \quad (3.14)$$

hvor,

$$\begin{aligned} \text{Cov}(O_{t-1}, O_t) &= \text{Cov}(O_{t-1}, \phi O_{t-1} + \epsilon_t - \theta \epsilon_{t-1}) \\ &= \phi \gamma_0 - \theta \sigma_\epsilon^2 \end{aligned} \quad (3.15)$$

$$\begin{aligned} \text{Cov}(\epsilon_t, O_t) &= \text{Cov}(\epsilon_t, \phi O_{t-1} + \epsilon_t - \theta \epsilon_{t-1}) \\ &= \sigma_\epsilon^2 \end{aligned} \quad (3.16)$$

$$\begin{aligned} \text{Cov}(\epsilon_{t-1}, O_t) &= \text{Cov}(\epsilon_{t-1}, \phi O_{t-1} + \epsilon_t - \theta \epsilon_{t-1}) \\ &= \phi \sigma_\epsilon^2 - \theta \sigma_\epsilon^2 \end{aligned} \quad (3.17)$$

Fra ligningene (3.14)-(3.17) får vi

$$\gamma_0 = \frac{1 - 2\phi\theta + \theta^2}{1 - \phi^2} \sigma_\epsilon^2$$

Generelt har vi, for $k \geq 1$

$$\gamma_k = \frac{(1 - \phi\theta)(\phi - \theta)}{1 - \phi^2} \phi^{k-1} \sigma_\epsilon^2$$

Autokorrelasjonen for ARMA(1,1), $O_t = \phi O_{t-1} + \epsilon_t - \theta \epsilon_{t-1}$ blir

$$\begin{aligned} \rho_0 &= 1 \\ \rho_k &= \frac{(1 - \phi\theta)(\phi - \theta)}{1 - 2\phi\theta + \theta^2} \phi^{k-1}, \quad \text{for } k \geq 1 \end{aligned} \quad (3.18)$$

Vi ser at autokorrelasjonfunksjonen er en avtagende funksjon med lag k .

Ligningen (3.13) skrives på følgende måte

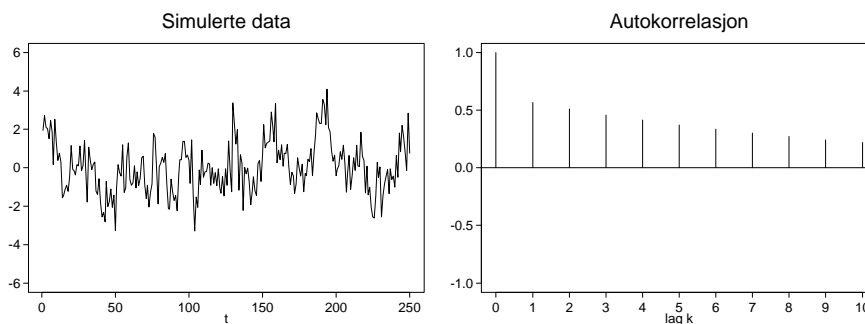
$$(1 - \phi B)O_t = (1 - \theta B)\epsilon_t, \quad \text{hvor } BO_t = O_{t-1}$$

eller

$$\begin{aligned} O_t &= \frac{1 - \theta B}{1 - \phi B} \epsilon_t \\ &= [1 + (\phi - \theta)B + \dots + (\phi - \theta)\phi^{s-1}B^s + \dots] \epsilon_t \\ &= \epsilon_t + (\phi - \theta) \sum_{s=1}^{\infty} \phi^{s-1} \epsilon_{t-s} \end{aligned} \quad (3.19)$$

Dermed er $|\phi| < 1$ en betingelse for at $\{O_t\}$ er en stasjonær prosess.

Figur 3.5 viser 250 simulerte dataene og autokorrelasjonen for ARMA(1,1) modell: $O_t = 0.9O_{t-1} + \epsilon_t - 0.5\epsilon_{t-1}$. Autokorrelasjonen ρ_k i (3.18) er vist i tabell 3.3.

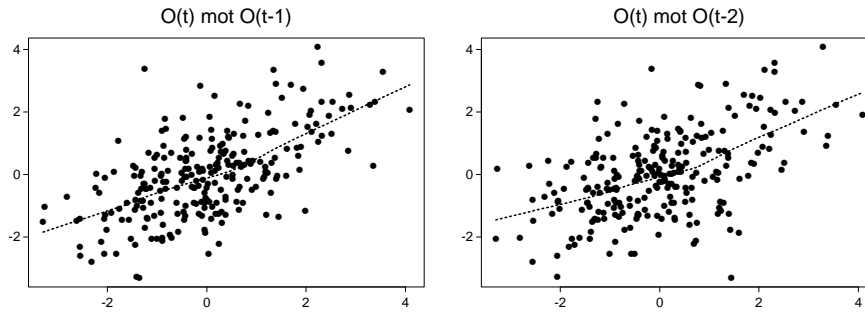


Figur 3.5: *Simulerte data og autokorrelasjonen fra ARMA(1,1): $O_t = 0.9O_{t-1} + \epsilon_t - 0.5\epsilon_{t-1}$*

Tabell 3.3: ρ_k av ARMA(1,1) prosess: $O_t = 0.9O_{t-1} + \epsilon_t - 0.5\epsilon_{t-1}$

lag	0	1	2	3	4	5	6	7	8	9	10
ρ_k	1.00	0.63	0.57	0.51	0.46	0.41	0.37	0.33	0.30	0.27	0.24

Plottene av O_t mot O_{t-1} og O_t mot O_{t-2} er vist i figur 3.6. Vi ser at det er en sterkere korrelasjon i lag 1 enn i lag 2.



Figur 3.6: x - y plott av (O_t, O_{t-1}) og (O_t, O_{t-2}) . $ARMA(1,1)$: $O_t = 0.9O_{t-1} + \epsilon_t - 0.5\epsilon_{t-1}$

3.2.11 Autoregressiv prosess av orden 2, AR(2)

Modellen for O_t blir

$$O_t = \phi_1 O_{t-1} + \phi_2 O_{t-2} + \epsilon_t \quad (3.20)$$

hvor ϕ_1 og ϕ_2 are konstanter. $\{\epsilon_t\}$ er hvit støy prosess. Vi antar at $E(\epsilon_t) = 0$, $Var(\epsilon_t) = \sigma_\epsilon^2$ og ϵ_t er uavhengige med O_{t-1}, O_{t-2}, \dots

Ligningen (3.20) skrives slik

$$(1 - \phi_1 B - \phi_2 B^2)O_t = \epsilon_t$$

hvor $BO_t = O_{t-1}$. Vi får

$$O_t = \frac{1}{1 - \phi_1 B - \phi_2 B^2} \epsilon_t$$

For $\{O_t\}$ er en stasjonær prosess, må røttene av ligningen $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 = 0$ være større enn 1 i absolutte verdier. Dette er ekvivalent med

$$\phi_1 + \phi_2 < 1, \quad \phi_2 - \phi_1 < 1, \quad \text{og} \quad |\phi_2| < 1 \quad (3.21)$$

Antar nå at $\{O_t\}$ er en stasjonær prosess. Autokovariansen γ_k , for $k = 1, 2, \dots$, er gitt ved

$$\begin{aligned} \gamma_k = Cov(O_t, O_{t-k}) &= Cov(\phi_1 O_{t-1} + \phi_2 O_{t-2} + \epsilon_t, O_{t-k}) \\ &= \phi_1 Cov(O_{t-1}, O_{t-k}) + \phi_2 Cov(O_{t-2}, O_{t-k}) + Cov(\epsilon_t, O_{t-k}) \\ &= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} \end{aligned} \quad (3.22)$$

siden ϵ_t er uavhengig med O_{t-k} . Ved å dividere γ_k med variansen til O_t , får vi autokorrelasjonen ρ_k , for $k = 1, 2, \dots$.

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad \text{for } k = 1, 2, \dots \quad (3.23)$$

(3.22) og/eller (3.23) kalles Yule-Walker ligningene.

For $k = 1$, (3.23) blir

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}$$

siden $\rho_0 = 1$ og $\rho_{-1} = \rho_1$.

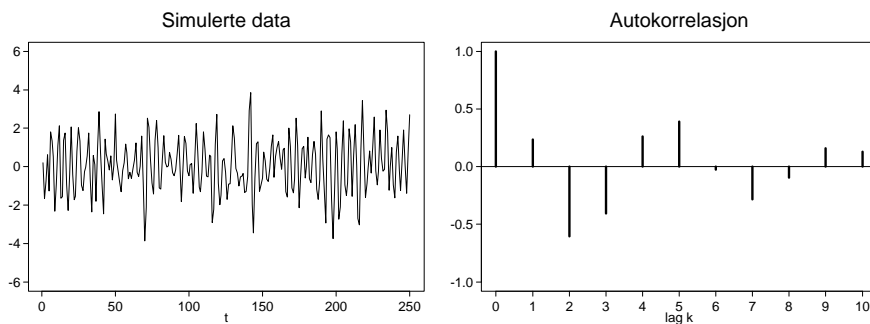
Fra (3.23) og verdien av ρ_1 , får vi, for $k = 2$

$$\begin{aligned}\rho_2 &= \phi_1\rho_1 + \phi_2\rho_0 \\ &= \frac{\phi_2(1 - \phi_2) + \phi_1^2}{1 - \phi_2}\end{aligned}$$

En generell form for ρ_k , $k = 1, 2, \dots$, kan finnes i mange lærebøker, for eksempel, Time Series Analysis av Jonathan D. Cryer. ρ_k går fort ned mot null for store verdier av k . Figur 3.7 viser 250 simulerte dataene og autokorrelasjonen fra AR(2) modell: $O_t = 0.4O_{t-1} - 0.7O_{t-2} + \epsilon_t$, med ϵ_t er uavhengige identisk normalfordelte, $N(0, 1)$. Vi har $\phi_1 = 0.4$ og $\phi_2 = -0.7$. Fra (3.21) følger at $\{O_t\}$ er en stasjonær prosess. Autokorrelasjonen ρ_k for de 10 første lagene er gitt i tabell 3.4.

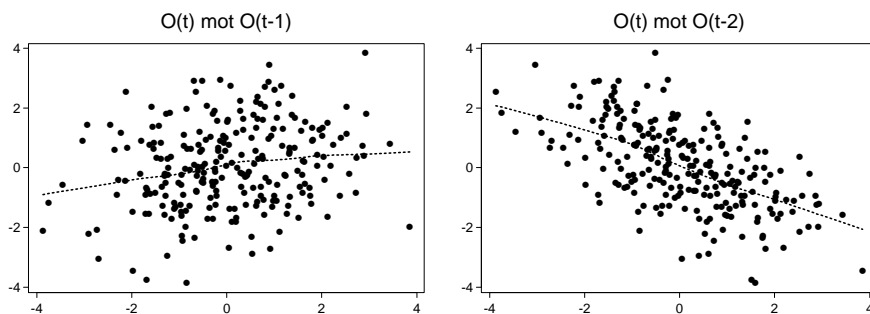
Tabell 3.4: ρ_k av AR(2) prosess: $O_t = 0.4O_{t-1} - 0.7O_{t-2} + \epsilon_t$

lag	0	1	2	3	4	5	6	7	8	9	10
ρ_k	1.00	0.24	-0.61	-0.41	0.26	0.39	-0.03	-0.28	-0.09	0.16	0.13



Figur 3.7: Simulerte data og autokorrelasjonen fra AR(2): $O_t = 0.4O_{t-1} - 0.7O_{t-2} + \epsilon_t$

Vi ser i figur 3.8 at korrelasjonen mellom (O_t, O_{t-2}) er sterkere enn (O_t, O_{t-1}) , siden $|\phi_2| > |\phi_1|$.



Figur 3.8: x - y plott av (O_t, O_{t-1}) og (O_t, O_{t-2}) . AR(2) prosess: $O_t = 0.4O_{t-1} - 0.7O_{t-2} + \epsilon_t$

3.2.12 Glidende gjennomsnitt prosess av orden 2, MA(2)

O_t kalles en glidende gjennomsnitt prosess av orden 2, hvis

$$O_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} \quad (3.24)$$

hvor θ_1 og θ_2 er konstanter og ϵ_t er en hvit støy prosess med forventning 0, og variansen σ_ϵ^2 .

Forventningen for O_t er

$$E(O_t) = E(\epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}) = 0$$

og variansen for O_t er

$$\begin{aligned} \text{Var}(O_t) &= \text{Var}(\epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}) \\ &= (1 + \theta_1^2 + \theta_2^2) \sigma_\epsilon^2 \end{aligned}$$

For $k = 1$, får vi

$$\begin{aligned} \gamma_1 &= \text{Cov}(\epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}, \epsilon_{t-1} - \theta_1 \epsilon_{t-2} - \theta_2 \epsilon_{t-3}) \\ &= \text{Cov}(-\theta_1 \epsilon_{t-1}, \epsilon_{t-1}) + \text{Cov}(-\theta_2 \epsilon_{t-2}, -\theta_1 \epsilon_{t-2}) \\ &= -\theta_1 \text{Cov}(\epsilon_{t-1}, \epsilon_{t-1}) + \theta_1 \theta_2 \text{Cov}(\epsilon_{t-2}, \epsilon_{t-2}) \\ &= -\theta_1 (1 - \theta_2) \sigma_\epsilon^2 \end{aligned}$$

For $k = 2$

$$\begin{aligned} \gamma_2 &= \text{Cov}(\epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}, \epsilon_{t-2} - \theta_1 \epsilon_{t-3} - \theta_2 \epsilon_{t-4}) \\ &= -\theta_2 \text{Cov}(\epsilon_{t-2}, \epsilon_{t-2}) \\ &= -\theta_2 \sigma_\epsilon^2 \end{aligned}$$

og for $k > 2$, er $\gamma_k = 0$.

Autokorrelasjonen ρ_k , blir

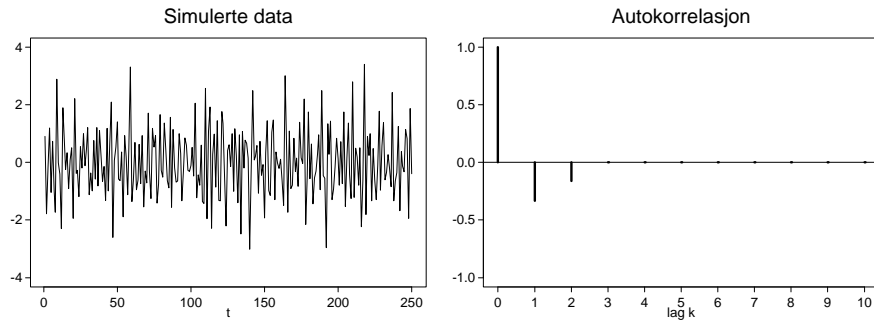
$$\rho_k = \begin{cases} \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}, & k = 1 \\ \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}, & k = 2 \\ 0, & k > 2 \end{cases}$$

Vi ser at ρ_k er forsvunnet etter lag 2. Siden forventningen, variansen og autokovariansen er uavhengige av t , er MA(2) en stasjonær prosess.

Figur 3.9 viser 250 simulerte dataene og autokorrelasjonen ρ_k for MA(2) prosess: $O_t = \epsilon_t - 0.65\epsilon_{t-1} - 0.24\epsilon_{t-2}$, hvor $\epsilon_t \sim N(0, 1)$. Vi får $\rho_0 = 1$, $\rho_1 = -0.333$, $\rho_2 = -0.162$ og $\rho_k = 0$ for $k > 2$.

Autokorrelasjonen ρ_k er gitt i tabell 3.5.

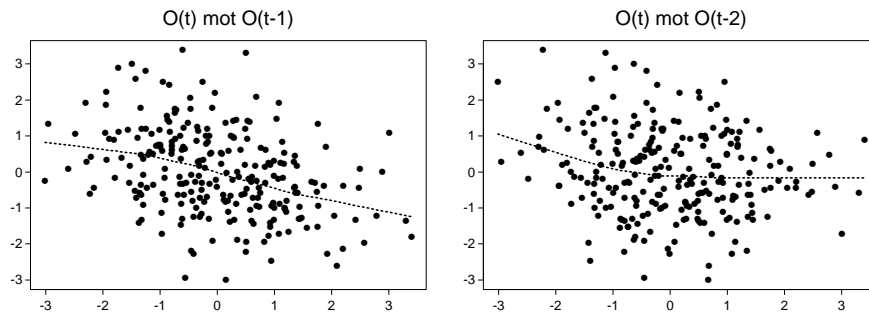
Korrelasjonene mellom (O_t, O_{t-1}) og (O_t, O_{t-2}) er vist i figur 3.10, med $\rho_1 = -0.33$ og $\rho_2 = -0.16$. For $k > 2$ er $\rho_k = 0$.



Figur 3.9: Simulerte data og autokorrelasjonen fra $MA(2)$: $O_t = \epsilon_t - 0.65\epsilon_{t-1} - 0.24\epsilon_{t-2}$

Tabell 3.5: ρ_k av $MA(2)$ prosess: $O_t = \epsilon_t - 0.65\epsilon_{t-1} - 0.24\epsilon_{t-2}$

lag	0	1	2	3	4	5	6	7	8	9	10
ρ_k	1.00	-0.33	-0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00



Figur 3.10: x - y plott av O_t og O_{t-1} , O_t og O_{t-2} fra $MA(2)$: $O_t = \epsilon_t - 0.65\epsilon_{t-1} - 0.24\epsilon_{t-2}$

3.2.13 Blandet autoregressiv og glidende gjennomsnitt prosess, $ARMA(2,2)$

$\{O_t\}$ er en $ARMA(2,2)$ prosess hvis O_t skrives slik

$$O_t = \phi_1 O_{t-1} + \phi_2 O_{t-2} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} \quad (3.25)$$

hvor ϕ_1 , ϕ_2 , θ_1 og θ_2 er konstanter. Vi antar at $\{\epsilon_t\}$ er en hvit støy prosess, med forventning 0 og varians σ_ϵ^2 . Dessuten er ϵ_t uavhengige med O_{t-1} , O_{t-2} , \dots . Prosessen er stasjonær hvis betingelsene (3.21) er oppfylt.

Vi får

$$\begin{aligned} \gamma_k &= Cov(O_{t-k}, O_t) \\ &= Cov(O_{t-k}, \phi_1 O_{t-1} + \phi_2 O_{t-2} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}) \end{aligned}$$

$$\begin{aligned}
\gamma_k &= \phi_1 \text{Cov}(O_{t-k}, O_{t-1}) + \phi_2 \text{Cov}(O_{t-k}, O_{t-2}) + \\
&\quad \text{Cov}(O_{t-k}, \epsilon_t) - \theta_1 \text{Cov}(O_{t-k}, \epsilon_{t-1}) - \theta_2 \text{Cov}(O_{t-k}, \epsilon_{t-2}) \\
&= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \text{Cov}(O_{t-k}, \epsilon_t) - \theta_1 \text{Cov}(O_{t-k}, \epsilon_{t-1}) - \theta_2 \text{Cov}(O_{t-k}, \epsilon_{t-2})
\end{aligned} \tag{3.26}$$

Siden $\text{Cov}(O_{t-k}, \epsilon_{t-i}) = 0$ for $k > i$, får vi

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}, \quad \text{for } k > 2$$

og autokorrelasjonen ρ_k

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad \text{for } k > 2$$

For $k = 1, 2$, beregner vi γ_k ved (3.26), med

$$\begin{aligned}
\text{Cov}(O_{t-k}, \epsilon_t) &= \begin{cases} \sigma_\epsilon^2 & \text{for } k = 0, \\ 0 & \text{for } k = 1, 2 \end{cases} \\
\text{Cov}(O_{t-k}, \epsilon_{t-1}) &= \begin{cases} -\theta_1 \sigma_\epsilon^2 & \text{for } k = 0, \\ \sigma_\epsilon^2 & \text{for } k = 1, \\ 0 & \text{for } k = 2 \end{cases} \\
\text{Cov}(O_{t-k}, \epsilon_{t-2}) &= \begin{cases} -\theta_2 \sigma_\epsilon^2, & \text{for } k = 0, \\ -\theta_1 \sigma_\epsilon^2 & \text{for } k = 1, \\ \sigma_\epsilon^2 & \text{for } k = 2 \end{cases}
\end{aligned}$$

3.2.14 En generell autoregressiv og glidende gjennomsnitt prosess, ARMA(p, q)

En ARMA(p, q) prosess $\{O_t\}$ tilfredsstiller følgende ligningen

$$O_t = \phi_1 O_{t-1} + \phi_2 O_{t-2} + \dots + \phi_p O_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} \tag{3.27}$$

eller med en generell måte

$$\phi(B)O_t = \theta(B)\epsilon_t \tag{3.28}$$

hvor $\{\epsilon_t\}$ er hvit støy prosess med forventningen 0 og variansen σ_ϵ^2 . ϵ_t er uavhengige med O_{t-1}, O_{t-2}, \dots . $\phi(B)$ og $\theta(B)$ er polynomer av B med ordene p og q , henholdsvis. $BO_t = O_{t-1}$.

$$\begin{aligned}
\phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\
\theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q
\end{aligned} \tag{3.29}$$

$\{O_t\}$ blir en stasjonær prosess hvis røttene av $\phi(B) = 0$ ligger utenfor enhets-sirkel.

Antar $\{O_t\}$ er en stasjonær prosess. Ligningen (3.28) kan skrives som en glidende gjennomsnitt prosess ved

$$O_t = \frac{\theta(B)}{\phi(B)} \epsilon_t \tag{3.30}$$

Autokovariansen γ_k beregnes ved

$$\begin{aligned}\gamma_k &= Cov(O_{t-k}, O_t) \\ &= \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \dots + \phi_p\gamma_{k-p} + \\ &\quad Cov(O_{t-k}, \epsilon_t) - \theta_1Cov(O_{t-k}, \epsilon_{t-1}) - \dots - \theta_qCov(O_{t-k}, \epsilon_{t-q})\end{aligned}$$

Siden $Cov(O_{t-k}, \epsilon_{t-i}) = 0$ for $k > i$, får vi

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \dots + \phi_p\gamma_{k-p} \quad \text{for } k \geq (q+1), \quad (3.31)$$

og

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \dots + \phi_p\rho_{k-p} \quad \text{for } k \geq (q+1), \quad (3.32)$$

Vi ser at autokorrelasjonen ρ_k er avhengig av bare parametre $\phi_1, \phi_2, \dots, \phi_p$ av AR delen fra lag $(q+1)$. Men den er en funksjon av både ϕ_i og θ_j for lagene $1, 2, \dots, q$.

3.2.15 En generell lineær prosess

$\{O_t\}$ kalles en generell lineær prosess hvis O_t skrives slik

$$O_t = \sum_{s=0}^{\infty} \psi_s \epsilon_{t-s} \quad (3.33)$$

hvor $\{\epsilon_t\}$ er hvit støy prosess med forventning 0 og varians σ_ϵ^2 . $\{\psi_s\}$ er konstanter med $\sum_{s=0}^{\infty} \psi_s^2 < \infty$. Denne betingelsen er nødt for at O_t har endelig varians, siden $Var(O_t) = (\sum_{s=0}^{\infty} \psi_s^2) \sigma_\epsilon^2$. Autoregressiv (AR), glidende gjennomsnitt (MA) og blandet autoregressiv og glidende gjennomsnitt (ARMA) er spesielle tilfeller av (3.33). For å kunne få verdien til ρ_k skriver vi (3.33) slik

$$O_t = \sum_{s=-\infty}^{\infty} \psi_s \epsilon_{t-s}$$

hvor $\psi_s = 0$ for $s < 0$.

Autokovariansen γ_k er gitt ved

$$\gamma_k = Cov(O_{t-k}, O_t) = \left(\sum_{s=-\infty}^{\infty} \psi_s \psi_{s-k} \right) \sigma_\epsilon^2$$

og autokorrelasjonen ρ_k er

$$\rho_k = \left(\sum_{s=-\infty}^{\infty} \psi_s \psi_{s-k} \right) / \left(\sum_{s=-\infty}^{\infty} \psi_s^2 \right)$$

her må vi også anta at $\sum_{s=-\infty}^{\infty} \psi_s \psi_{s-k} < \infty$ slik at ρ_k kan beregnes.

Eksempel. Antar at $\{O_t\}$ er en ARMA(2,2) stasjonær prosess.

$$O_t = \phi_1 O_{t-1} + \phi_2 O_{t-2} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} \quad (3.34)$$

La $\phi(B) = 1 - \phi_1 B - \phi_2 B^2$ og $\theta(B) = 1 - \theta_1 B - \theta_2 B^2$, hvor $BO_t = O_{t-1}$ og $B\epsilon_t = \epsilon_{t-1}$. Ligningen (3.34) skrives slik

$$(1 - \phi_1 B - \phi_2 B^2) O_t = (1 - \theta_1 B - \theta_2 B^2) \epsilon_t$$

eller

$$\begin{aligned} O_t &= \frac{1 - \theta_1 B - \theta_2 B^2}{1 - \phi_1 B - \phi_2 B^2} \epsilon_t \\ &= (1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \dots) \epsilon_t \\ &= \sum_{s=0}^{\infty} \psi_s \epsilon_{t-s} \end{aligned}$$

hvor

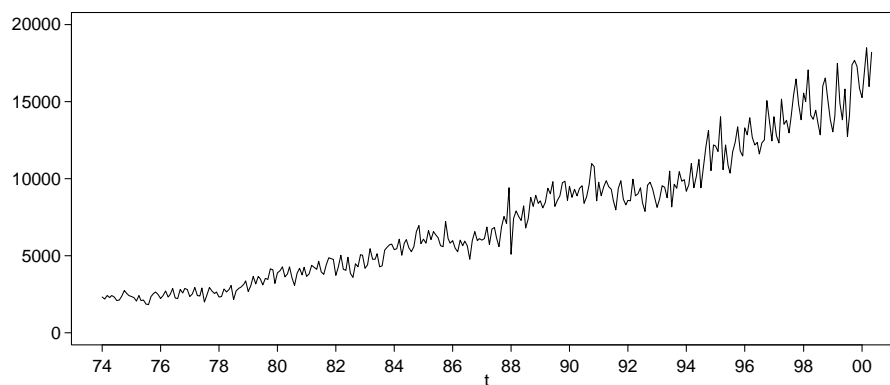
$$\begin{aligned} \psi_0 &= 1, \\ \psi_1 &= \phi_1 - \theta_1, \\ \psi_2 &= \phi_1(\phi_1 - \theta_1) + \phi_2 - \theta_2, \\ \psi_3 &= \phi_1\psi_2 + \phi_2\psi_1, \\ \dots &= \dots \\ \psi_k &= \phi_1\psi_{k-1} + \phi_2\psi_{k-2} \end{aligned} \tag{3.35}$$

Vi ser at O_t i (3.34) skrives nå på form en generell prosess, hvor ψ_i er gitt ved (3.35).

3.2.16 Ikke-stasjonær prosess. ARIMA(p, d, q)

I praksis, kan vi treffe mange tidsserier, særlig i økonomisk statistikk, som ikke tilfredsstillende kraver for en stasjonær prosess. Figur 3.11 viser et eksempel. Dataene er eksport av tradisjonelle varer (i millioner kroner) som er observert fra januar 1974 til mai 2000. Vi ser en klar trend som er en voksende funksjon og variasjoner av rådataene blir større og større med tid. Dette antyder at (i) og (ii) i avsnitt 3.2.6 ikke er tilfredsstillende. Dermed er tidsserien ikke stasjonær.

I dette avsnittet vil vi innføre en ARIMA(p, d, q) (autoregressive integrated moving average) prosess. Dvs, prosessen blir en stasjonær ARMA(p, q) etter å ha utført d ganger differenser. I praksis, vil d være 1 eller 2.



Figur 3.11: Eksport av tradisjonelle varer (i millioner)

Vi innfører notasjonen ∇ som er definert ved $\nabla = 1 - B$, hvor $BO_t = O_{t-1}$. Dvs, $\nabla O_t = O_t - O_{t-1}$. En generell ligning for ARIMA(p, d, q) er

$$\phi(B)\nabla^d O_t = \theta(B)\epsilon_t \tag{3.36}$$

hvor $\{\epsilon_t\}$ er en hvit støy prosess. $\nabla^d = \nabla \nabla \dots \nabla$ (d ganger). $\phi(B)$ og $\theta(B)$ er definert ved (3.29), som er polynomer av B med ordenene p og q , henholdsvis.

$$\begin{aligned}\phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q\end{aligned}$$

Eksempel. La O_t være en ARIMA(0,1,1) prosess. Vi får $p = 0$, $q = 1$ og $d = 1$. La $Z_t = \nabla O_t = (1 - B)O_t = O_t - O_{t-1}$. Z_t er en stasjonær ARMA(0,1) prosess. Ligningen for Z_t blir

$$Z_t = \epsilon_t - \theta \epsilon_{t-1}$$

siden $Z_t = O_t - O_{t-1}$, får vi

$$\begin{aligned}O_t &= O_{t-1} + \epsilon_t - \theta \epsilon_{t-1} \\ &= O_{t-2} + \epsilon_t + (1 - \theta) \epsilon_{t-1} - \theta \epsilon_{t-2} \\ &= \dots \\ &= O_{n_0} + \epsilon_t + (1 - \theta) \epsilon_{t-1} + \dots + (1 - \theta) \epsilon_{n_0+1} - \theta \epsilon_{n_0}\end{aligned}\quad (3.37)$$

for $t > n_0$, hvor n_0 er startpunktet og O_{n_0} er startverdien. Siden $\{O_t\}$ ikke er en stasjonær prosess, kan vi ikke anta at prosessen er startet i tidspunkt $t = -N$, hvor $N \rightarrow \infty$. Fra (3.37) får vi

$$O_{t-k} = O_{n_0} + \epsilon_{t-k} + (1 - \theta) \epsilon_{t-k-1} + \dots + (1 - \theta) \epsilon_{n_0+1} - \theta \epsilon_{n_0}$$

Variansen til O_t blir

$$\text{Var}(O_t) = \sigma_\epsilon^2 + (t - n_0 - 1)(1 - \theta)^2 \sigma_\epsilon^2 = [1 + (t - n_0 - 1)(1 - \theta)^2] \sigma_\epsilon^2$$

Siden O_{n_0} og ϵ_{n_0} er kjente. Variansen til O_{t-k} blir

$$\text{Var}(O_{t-k}) = [1 + (t - k - n_0 - 1)(1 - \theta)^2] \sigma_\epsilon^2$$

og

$$\text{Cov}(O_t, O_{t-k}) = [(1 - \theta) + (t - k - n_0 - 1)(1 - \theta)^2] \sigma_\epsilon^2$$

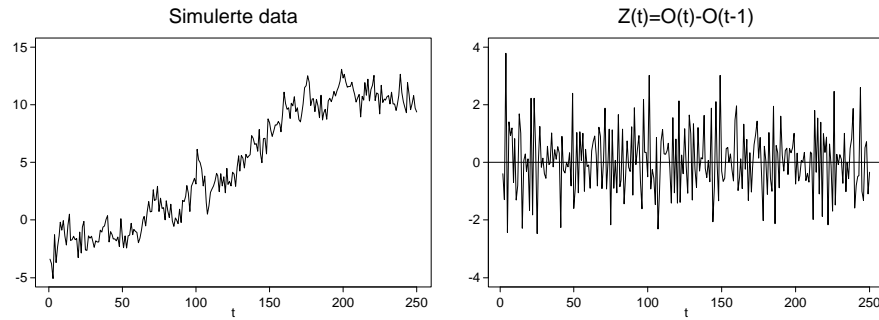
Autokorrelasjonen $\text{Corr}(O_t, O_{t-k})$ blir

$$\text{Corr}(O_t, O_{t-k}) = \frac{(1 - \theta) + (t - k - n_0 - 1)(1 - \theta)^2}{\sqrt{[1 + (t - n_0 - 1)(1 - \theta)^2][1 + (t - k - n_0 - 1)(1 - \theta)^2]}}$$

Vi ser at

1. $\text{Var}(O_t)$ er ikke uavhengig av t . $\text{Var}(O_t) \neq \text{Var}(O_{t-k})$, for $k \neq 0$.
2. $\text{Var}(O_t)$ er en voksende funksjon med t .
3. $\text{Cov}(O_t, O_{t-k})$ og $\text{Corr}(O_t, O_{t-k})$ er ikke bare avhengige av lag k . De avhenger også av t . Hvis n_0 er stor i forhold til k , er $\text{Corr}(O_t, O_{t-k}) \approx 1$. ρ_k går langsomt ned mot null med stor k .

Figur 3.12 viser 250 simulerte dataene fra ARIMA(0,1,1) modell, $O_t = O_{t-1} + \epsilon_t - 0.4\epsilon_{t-1}$, hvor $\epsilon_t \sim N(0, 1)$. Det er en klar trend. La $Z_t = O_t - O_{t-1} = \epsilon_t - 0.4\epsilon_{t-1}$. $\{Z_t\}$ blir en stasjonær prosess. Tidsserien er plottet i høyre side av figuren.



Figur 3.12: Simulerte dataene fra $ARIMA(0,1,1)$: $O_t = O_{t-1} + \epsilon_t - 0.4\epsilon_{t-1}$

3.2.17 Partiell autokorrelasjonsfunksjon (PACF)

La $\{O_t\}$ være en $AR(k)$ stasjonær prosess med følgende ligning

$$O_{t+k} = \phi_{k1}O_{t+k-1} + \phi_{k2}O_{t+k-2} + \cdots + \phi_{kk}O_t + \epsilon_{t+k} \quad (3.38)$$

hvor ϕ_{ki} , $i = 1, \dots, k$ er regresjonskoeffisienter og ϵ_t er hvit støy prosess.

Yule-Walker ligningene skrives

$$\begin{aligned} \rho_1 &= \phi_{k1}\rho_0 + \phi_{k2}\rho_1 + \cdots + \phi_{kk}\rho_{k-1} \\ \rho_2 &= \phi_{k1}\rho_1 + \phi_{k2}\rho_0 + \cdots + \phi_{kk}\rho_{k-2} \\ &\vdots \\ \rho_k &= \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \cdots + \phi_{kk}\rho_0 \end{aligned}$$

Ved å bruke Cramers regel får vi

$$\begin{aligned} \phi_{11} &= \rho_1 \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \end{aligned}$$

$$\begin{aligned}
\phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = \frac{\rho_1 \rho_2^2 - 2\rho_2 \rho_1 + \rho_1^3 - \rho_3 \rho_1^2 + \rho_3}{(\rho_2 - 2\rho_1^2 + 1)(1 - \rho_2)} \\
&\dots \\
\phi_{kk} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_1 & 1 \end{vmatrix}} \quad (3.39)
\end{aligned}$$

ϕ_{kk} kalles partiell autokorrelasjonen mellom O_t og O_{t+k} i lag k .

Den partielle autokorrelasjonen mellom O_t og O_{t+k} beskrives på en annen måte.

La \widehat{O}_{t+k} skrives som en lineær funksjon av $O_{t+1}, O_{t+2}, \dots, O_{t+k-1}$

$$\widehat{O}_{t+k} = \alpha_1 O_{t+k-1} + \alpha_2 O_{t+k-2} + \dots + \alpha_{k-1} O_{t+1}$$

hvor $\alpha_1, \dots, \alpha_{k-1}$ er konstanter som minimerer

$$E(O_{t+k} - \widehat{O}_{t+k})^2 = E(O_{t+k} - \alpha_1 O_{t+k-1} - \alpha_2 O_{t+k-2} - \dots - \alpha_{k-1} O_{t+1})^2$$

Da er \widehat{O}_{t+k} den beste estimatoren for O_{t+k} som minimerer $E(O_{t+k} - \widehat{O}_{t+k})^2$.

La

$$\widehat{O}_t = \beta_1 O_{t+k-1} + \beta_2 O_{t+k-2} + \dots + \beta_{k-1} O_{t+1}$$

hvor $\beta_1, \dots, \beta_{k-1}$ er konstanter som minimerer

$$E(O_t - \widehat{O}_t)^2 = E(O_t - \beta_1 O_{t+k-1} - \beta_2 O_{t+k-2} - \dots - \beta_{k-1} O_{t+1})^2$$

Da er \widehat{O}_t den beste estimatoren for O_t som minimerer $E(O_t - \widehat{O}_t)^2$.

Siden $\{O_t\}$ er en stasjonær prosess, kan det vise at $\alpha_i = \beta_i$, for $i = 1, \dots, k-1$ (se Wei (1990), side 12).

La P_k være den vanlige autokorrelasjonen mellom $(O_t - \widehat{O}_t)$ og $(O_{t+k} - \widehat{O}_{t+k})$

$$P_k = \frac{Cov[(O_t - \widehat{O}_t)(O_{t+k} - \widehat{O}_{t+k})]}{\sqrt{Var(O_t - \widehat{O}_t)} \sqrt{Var(O_{t+k} - \widehat{O}_{t+k})}}$$

Det viser at $P_k = \phi_{kk}$ (se Wei (1990), side 13). En partiell autokorrelasjon mellom O_t og O_{t+k} er dermed en vanlig autokorrelasjonen mellom O_t og O_{t+k} etter å ha fjernet effektene av $O_{t+1}, O_{t+2}, \dots, O_{t+k-1}$.

Eksempel. Vi skal beregne ϕ_{kk} for AR(2) og MA(2) prosesser.

- AR(2) prosess: $O_t = \phi_1 O_{t-1} + \phi_2 O_{t-2} + \epsilon_t$. Vi får (se side 34 for beregninger av ρ_i)

$$\begin{aligned}\rho_1 &= \frac{\phi_1}{1 - \phi_2} \\ \rho_2 &= \frac{\phi_2(1 - \phi_2) + \phi_1^2}{1 - \phi_2} \\ \rho_3 &= \dots\end{aligned}$$

ρ_1 og ρ_2 blir satt inn i (3.39). Vi får

$$\begin{aligned}\phi_{11} &= \rho_1 = \frac{\phi_1}{1 - \phi_2} \\ \phi_{22} &= \frac{\frac{\phi_2(1 - \phi_2) + \phi_1^2}{1 - \phi_2} - \left(\frac{\phi_1}{1 - \phi_2}\right)^2}{1 - \left(\frac{\phi_1}{1 - \phi_2}\right)^2} \\ &= \frac{\phi_2[(1 - \phi_2)^2 - \phi_1^2]}{(1 - \phi_2)^2 - \phi_1^2} = \phi_2 \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \phi_1 + \phi_2\rho_1 \\ \rho_1 & 1 & \phi_1\rho_1 + \phi_2 \\ \rho_2 & \rho_1 & \phi_1\rho_2 + \phi_2\rho_1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} = 0\end{aligned}$$

siden den siste kolonnen i teller er en lineær kombinasjon av de to første kolonnene. Det kan vise at for en AR(p) prosess, er $\phi_{pp} = \phi_p$ og $\phi_{kk} = 0$ for $k > p$. Dvs PACF av en AR(p) prosess er lik null for lagene etter lag p .

- MA(2) prosess: $O_t = \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2}$. Vi får

$$\rho_k = \begin{cases} \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}, & k = 1 \\ \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}, & k = 2 \\ 0, & k > 2 \end{cases}$$

(se side 35 for beregningen av ρ_k). Ved innsetting av ρ_k i (3.39), får vi

$$\begin{aligned}\phi_{11} &= \rho_1 \\ \phi_{22} &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \\ \phi_{33} &= \frac{\rho_1^3 - \rho_1\rho_2(2 - \rho_2)}{1 - \rho_2^2 - 2\rho_1^2(1 - \rho_2)} \\ &\vdots\end{aligned}$$

Vi ser at ρ_k av en MA(2) går fort ned mot null etter lag 2, mens ϕ_{kk} fortsatt varierer rundt null akse for de lagene senere. Anvendelsen av ACF og PACF beskrives i neste avsnitt.

3.2.18 Anvendelse av autokorrelasjonfunksjon (ACF) og partiell autokorrelasjonfunksjon (PACF) i tidsserieanalyse

Man anvender mønstre av ACF og PACF for å kunne identifisere en modell (AR, MA, ARMA eller ARIMA) for en tidsserie. Box og Jenkins (1970) teknikken beskrives ved følgende regler

- *Ikke-stasjonær prosess.* ACF går langsom ned mot null. Men PACF har høy positiv eller negativ verdi i lag 1. Vi kan transformere prosessen til en stasjonær prosess ved differenser, $\nabla O_t, \nabla^2 O_t, \dots, \nabla^d O_t$. Ofte er $d \leq 2$.
- *Autoregressiv prosess.* PACF har verdier ($\neq 0$) i lagene $k = 1, 2, \dots, p$ og den er lik null deretter. I dette tilfellet tilpasser vi dataene med en autoregressiv modell av orden p .
- *Glidende gjennomsnitt prosess.* ACF har verdier ($\neq 0$) i lagene $k = 1, 2, \dots, q$ og deretter er den lik null. Tilpass dataene med en MA modell av orden q .

3.2.19 Estimering av autokorrelasjon (ACF) og partiell autokorrelasjonfunksjon (PACF) fra observasjoner

a. ACF. Fra de observasjonene O_1, O_2, \dots, O_n kan vi estimere autokorrelasjonfunksjonen ρ_k på følgende form

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (O_t - \bar{O})(O_{t+k} - \bar{O})}{\sum_{t=1}^n (O_t - \bar{O})^2}, \quad \text{for } k = 0, 1, 2, \dots \quad (3.40)$$

hvor $\bar{O} = \sum_{t=1}^n O_t/n$. Bartlett (1946) har vist at for en stasjonær normalfordelt prosess, når n er stor, er $\hat{\rho}_k$ tilnærmet normalfordelt med forventning ρ_k og varians som gitt ved

$$\text{Var}(\hat{\rho}_k) \approx \frac{1}{n} \sum_{i=-\infty}^{\infty} (\rho_i^2 + \rho_{i+k}\rho_{i-k} - 4\rho_k\rho_i\rho_{i-k} + 2\rho_k^2\rho_i^2)$$

For prosessen med $\rho_k = 0$ for $k > m$, er

$$\text{Var}(\hat{\rho}_k) \approx \frac{1}{n} (1 + 2\rho_1^2 + 2\rho_2^2 + \dots + 2\rho_m^2)$$

Vi beregner $\text{Var}(\hat{\rho}_k)$ ved å erstatte $\rho_i, i = 1, \dots, m$ med $\hat{\rho}_i, i = 1, \dots, m$. For en hvit støy prosess er $\text{Var}(\hat{\rho}_k) \approx 1/n$.

b. PACF. Vi starter med $\hat{\phi}_{11} = \hat{\rho}_1$. Ved å bruke rekursiv metode fra Durbin (1960) har vi følgende form

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_j} \quad (3.41)$$

for estimering av $\phi_{k+1,k+1}$ fra de observasjonene O_1, O_2, \dots, O_n , hvor

$$\hat{\phi}_{k+1,j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1} \hat{\phi}_{k,k+1-j}, \quad \text{for } j = 1, \dots, k$$

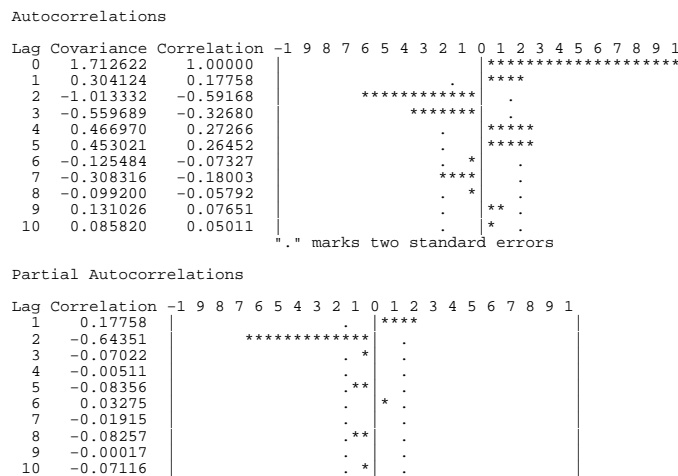
Vi får fra (3.41)

$$\begin{aligned}\hat{\phi}_{11} &= \hat{\rho}_1 \\ \hat{\phi}_{22} &= \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2}, \\ \hat{\phi}_{33} &= \frac{\hat{\rho}_3 - \hat{\phi}_{21}\hat{\rho}_2 - \hat{\phi}_{22}\hat{\rho}_1}{1 - \hat{\phi}_{21}\hat{\rho}_1 - \hat{\phi}_{22} - \hat{\rho}_2}\end{aligned}$$

hvor $\hat{\phi}_{21} = \hat{\phi}_{11} - \hat{\phi}_{22}\hat{\phi}_{11}$.

For en hvit støy prosess, er $Var(\hat{\phi}_{kk}) \approx 1/n$.

Eksempel. Vi har simulert 250 observasjoner fra en AR(2) modell: $O_t = 0.4O_{t-1} - 0.7O_{t-2} + \epsilon_t$. Dataene er plottet i figur 3.7. De estimerte ACF og PACF er beregnet fra SAS med PROC ARIMA. Verdierne og histogrammet er vist i figur 3.13.



Figur 3.13: De estimerte ACF og PACF fra de simulerte dataene av AR(2) prosess: $O_t = 0.4O_{t-1} - 0.7O_{t-2} + \epsilon_t$

Vi har $\phi_1 = 0.4$ og $\phi_2 = -0.7$. Ved å bruke ligningene (3.23), (3.39), (3.40) og (3.41) får vi ρ_k , $\hat{\rho}_k$, ϕ_{kk} og $\hat{\phi}_{kk}$ som listet ut i tabell 3.6. Vi ser i figur 3.13 og tabell 3.6 at $\hat{\phi}_{kk}$ er signifikant større enn null på 5% nivå, i lag 1 og 2. Mens $\hat{\rho}_k$ er fortsatt signifikant større enn null i mange lag. Merk at beregningen av $\hat{\rho}_k$ og $\hat{\phi}_{kk}$ er basert på 250 observasjoner.

3.2.20 Framskrivning

La O_1, O_2, \dots, O_n være en tidsserie som observerer til tidspunkt n . La O_{n+l} være framskrivningen av tidsserien i tidspunkt $n+l$ (hvor l står for l skritt framover som starter fra tidspunkt n). Framskrivningen basert på minste kvadraters metode, for O_{n+l} , som er betegnet med $\hat{O}_n(l)$, er definert slik

$$\hat{O}_n(l) = E(O_{n+l} | O_n, O_{n-1}, \dots) \quad (3.42)$$

dvs, $\hat{O}_n(l)$ beskrives som betingede forventningen av O_{n+l} gitt observasjoner fram til det siste tidspunktet.

Tabell 3.6: ρ_k , $\hat{\rho}_k$, ϕ_{kk} og $\hat{\phi}_{kk}$ av $AR(2)$ prosess: $O_t = 0.4O_{t-1} - 0.7O_{t-2} + \epsilon_t$

lag	ρ_k	$\hat{\rho}_k$	ϕ_{kk}	$\hat{\phi}_{kk}$
1	0.2353	0.1776	0.2353	0.1776
2	-0.6059	-0.5916	-0.7000	-0.6435
3	-0.4071	-0.3268	0.0	-0.0702
4	0.2613	0.2726	0.0	-0.0051
5	0.3895	0.2645	0.0	-0.0835
6	-0.0271	-0.0732	0.0	0.0327
7	-0.2835	-0.1800	0.0	-0.0191
8	-0.0944	-0.0579	0.0	-0.0825
9	0.1607	0.0765	0.0	-0.0001
10	0.1303	0.0501	0.0	-0.0711

Eksempel 1. La $\{O_t\}$ være en stasjonær $AR(1)$ prosess, som vi observerer fram til tidspunkt n . O_t beskrives ved ligningen

$$O_t = \phi O_{t-1} + \epsilon_t$$

hvor ϵ_t er hvit støy prosess, med forventning 0 og varians σ_ϵ^2 .

Ved å bruke (3.42) for å estimere framskrivingen i neste tidspunkt fra tidsserien O_1, O_2, \dots, O_n , får vi

$$\begin{aligned} \hat{O}_n(1) &= E(O_{n+1}|O_n, O_{n-1}, \dots) \\ &= E(\phi O_n + \epsilon_{n+1}|O_n, O_{n-1}, \dots) \\ &= \phi E(O_n|O_n, O_{n-1}, \dots) + E(\epsilon_{n+1}|O_n, O_{n-1}, \dots) \\ &= \phi O_n \end{aligned}$$

siden $E(O_n|O_n, O_{n-1}, \dots) = O_n$ og ϵ_{n+1} er uavhengig med observasjonene som ligger før tidspunkt $n + 1$. Dermed er $E(\epsilon_{n+1}|O_n, O_{n-1}, \dots) = 0$.

I beregningen for 95% konfidensintervall for O_{n+1} , vil vi skrive O_t på form

$$\begin{aligned} O_t &= \frac{1}{1 - \phi B} \epsilon_t, \quad \text{hvor } BO_t = O_{t-1} \\ &= \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \dots \\ &= \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \quad \text{hvor } \psi_0 = 1, \text{ og } \psi_j = \phi^j \end{aligned}$$

Dermed er

$$\begin{aligned} O_{n+1} &= \epsilon_{n+1} + \phi \epsilon_n + \phi^2 \epsilon_{n-1} + \dots + && \text{og} \\ \hat{O}_n(1) &= \phi \epsilon_n + \phi^2 \epsilon_{n-1} + \dots + && \text{siden } \hat{O}_n(1) = \phi O_n \end{aligned}$$

La $e_n(1)$ være differensen mellom O_{n+1} og $\hat{O}_n(1)$. $e_n(1)$ kalles framskrivingsfeil. Vi får

$$\begin{aligned} e_n(1) &= O_{n+1} - \hat{O}_n(1) \\ &= \epsilon_{n+1} \end{aligned} \tag{3.43}$$

siden ϵ_{n+1} er ukorrelerede med ϵ_t , for $t = 1, \dots, n$, vil (3.43) vise at $\widehat{O}_n(1)$ er den beste estimatoren for O_{n+1} . Men hvis ϵ_{n+1} er korrelerte med de tidligere feilleddene, kan vi estimere $\widehat{\epsilon}_{n+1}$ fra dem og forbedre estimeringen for O_{n+1} ved $\widehat{O}_n(1) + \widehat{\epsilon}_{n+1}$.

Variansen for $e_n(1)$ blir

$$\text{Var}(e_n(1)) = \text{Var}(\epsilon_{n+1}) = \sigma_\epsilon^2$$

Antar at $\{O_t\}$ har normalfordeling, så er $e_n(1)$ også en normalfordeling. Vi får

$$P\left[-1.96 \leq \frac{O_{n+1} - \widehat{O}_n(1)}{\sqrt{\text{Var}(e_n(1))}} \leq 1.96\right] = 0.95$$

95% konfidensintervallet for O_{n+1} blir

$$\left[\widehat{O}_n(1) - 1.96\sqrt{\text{Var}(e_n(1))}, \widehat{O}_n(1) + 1.96\sqrt{\text{Var}(e_n(1))}\right]$$

For $l > 1$, den estimerte verdien for O_{n+l} beregnes ved (3.42), er

$$\begin{aligned} \widehat{O}_n(l) &= E(O_{n+l}|O_n, O_{n-1}, \dots) \\ &= \phi E(O_{n+l-1}|O_n, O_{n-1}, \dots) + E(\epsilon_{n+l}|O_n, O_{n-1}, \dots) \\ &= \phi \widehat{O}_n(l-1) = \phi \widehat{O}_n(l-2) = \dots \\ &= \phi^l O_n \end{aligned} \quad (3.44)$$

når $\phi < 1$ og n er stor nok vil $\widehat{O}_n(l)$ gå mot null. Vi skriver O_{n+l} som

$$\begin{aligned} O_{n+l} &= \epsilon_{n+l} + \phi \epsilon_{n+l-1} + \dots + \phi^{l-1} \epsilon_{n+1} + \phi^l \epsilon_n + \phi^{l+1} \epsilon_{n-1} + \dots \quad \text{og fra (3.44)} \\ \widehat{O}_n(l) &= \phi^l \epsilon_n + \phi^{l+1} \epsilon_{n-1} + \dots \end{aligned}$$

Dermed er

$$\begin{aligned} e_n(l) &= O_{n+l} - \widehat{O}_n(l) \\ &= \sum_{j=0}^{l-1} \psi_j \epsilon_{n+l-j} \quad \text{hvor } \psi_0 = 1, \text{ og } \psi_j = \phi^j \end{aligned}$$

Variansen for $e_n(l)$ er

$$\text{Var}(e_n(l)) = \text{Var}\left(\sum_{j=0}^{l-1} \psi_j \epsilon_{n+l-j}\right) = \left(\sum_{j=0}^{l-1} \psi_j^2\right) \sigma_\epsilon^2$$

95% konfidensintervall for O_{n+l} blir

$$\left[\widehat{O}_n(l) - 1.96\sqrt{\text{Var}(e_n(l))}, \widehat{O}_n(l) + 1.96\sqrt{\text{Var}(e_n(l))}\right] \quad (3.45)$$

siden $\text{Var}(e_n(l))$ vokser med l , vil vi få større usikkerheter for framskrivingsberegningen.

Merk at avhengig av modell har vi en bestemt form for ψ_j . For eksempel, for AR(2) modell

$$O_t = \phi_1 O_{t-1} + \phi_2 O_{t-2} + \epsilon_t$$

er ψ_j gitt ved

$$\begin{aligned}\psi_0 &= 1 \\ \psi_1 &= \phi_1 \\ \psi_2 &= \phi_1\psi_1 + \phi_2\psi_0 \\ &\dots \\ \psi_k &= \phi_1\psi_{k-1} + \phi_2\psi_{k-2}\end{aligned}\tag{3.46}$$

Oppdatering av framskrivinger.

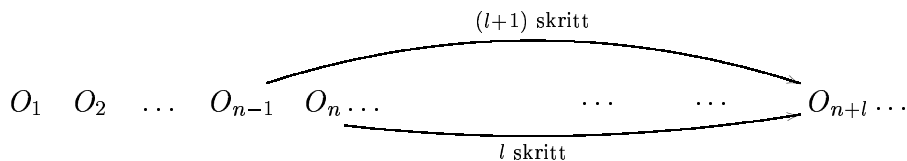
La $\{O_t\}$ være en stasjonær prosess som skrives på form (3.33), dvs

$$O_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

hvor $\{\epsilon_t\}$ er en hvit støy prosess med forventning 0 og varians σ_ϵ^2 . ψ_j er konstanter med $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Vi skal beregne framskrivningen i tidspunkt $n+l$ ved to måter:

a. Ved å bruke dataene fram til den nest siste observasjonen, dvs fra O_1 til O_{n-1} .

Figuren nedenfor illustrerer tidsserien.



Vi får

$$\begin{aligned}O_{n+l} &= \epsilon_{n+l} + \psi_1\epsilon_{n+l-1} + \dots + \psi_l\epsilon_n + \psi_{l+1}\epsilon_{n-1} + \psi_{l+2}\epsilon_{n-2} + \dots \\ \widehat{O}_{(n-1)}(l+1) &= \psi_{l+1}\epsilon_{n-1} + \psi_{l+2}\epsilon_{n-2} + \dots\end{aligned}$$

Dermed er

$$\begin{aligned}e_{(n-1)}(l+1) &= O_{n+l} - \widehat{O}_{(n-1)}(l+1) \\ &= \epsilon_{n+l} + \psi_1\epsilon_{n+l-1} + \dots + \psi_l\epsilon_n\end{aligned}\tag{3.47}$$

b. Ved å bruke dataene helt fram til den siste observasjonen, dvs fra O_1 til O_n .

Vi får

$$\begin{aligned}O_{n+l} &= \epsilon_{n+l} + \psi_1\epsilon_{n+l-1} + \dots + \psi_{l-1}\epsilon_{n+1} + \psi_l\epsilon_n + \psi_{l+1}\epsilon_{n-1} + \dots \\ \widehat{O}_n(l) &= \psi_l\epsilon_n + \psi_{l+1}\epsilon_{n-1} + \dots\end{aligned}$$

Dermed er

$$\begin{aligned}e_n(l) &= O_{n+l} - \widehat{O}_n(l) \\ &= \epsilon_{n+l} + \psi_1\epsilon_{n+l-1} + \dots + \psi_{l-1}\epsilon_{n+1}\end{aligned}\tag{3.48}$$

Fra (3.47) og (3.48), får vi

$$e_{(n-1)}(l+1) = e_n(l) + \psi_l\epsilon_n$$

eller

$$O_{n+l} - \widehat{O}_{(n-1)}(l+1) = O_{n+l} - \widehat{O}_n(l) + \psi_l\epsilon_n$$

Vi får

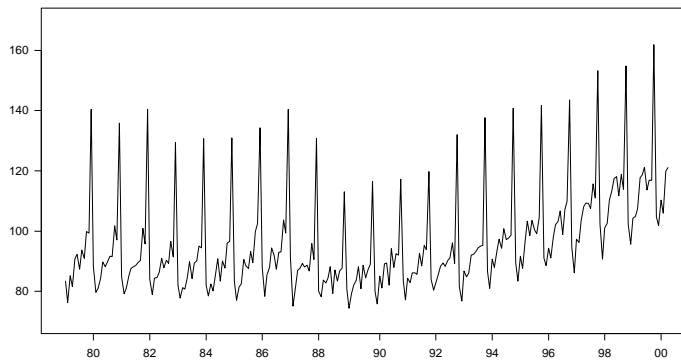
$$\begin{aligned}\widehat{O}_n(l) &= \widehat{O}_{(n-1)}(l+1) + \psi_l [O_n - \widehat{O}_{(n-1)}(1)], & \text{eller} \\ \widehat{O}_{n+1}(l) &= \widehat{O}_n(l+1) + \psi_l [O_{(n+1)} - \widehat{O}_n(1)]\end{aligned}\quad (3.49)$$

Ligningen (3.49) er formen for oppdateringen av framskrivinger. Merk at $\widehat{O}_{(n)}(l+1)$ og $\widehat{O}_{(n+1)}(l)$ er framskrivingen i samme tidspunkt $n+l$. Men den andre beregnes med tidsserien som tilføyes en ny observasjon til.

(3.49) kan tolkes slik: når en ny observasjon blir observert kan man korrigere den framskrivingen som er beregnet ved å legge til størrelsen $\psi_l [O_{(n+1)} - \widehat{O}_n(1)]$ (et skritt framskrivingsfeil).

3.2.21 Modell med sesongvariasjoner

De fleste økonomiske tidsserier inneholder sesongvariasjoner. For eksempel, for



Figur 3.14: *Detaljvolumindeks*

detaljvolumindeksen, i figur 3.14 viser dataene fra januar 1979 til juli 2000 at mønsteret gjentar seg fra år til år i samme tidspunkter, med høye verdier i desember (julehandel) og lave verdier i februar (færrest handledager). Et annet eksempel er antall hotellovernattinger for nordmenn og utlendinger, der ser vi at juli er den måneden som har flest gjesteovernattinger for både nordmenn og utlendinger, mens januar og november er de to månedene som har færrest besøkende. Sesongmønsteret er svært tydelig i dataene.

a. Autoregressiv modell av orden 1, AR(1)s, med sesongvariasjoner

O_t er en autoregressiv prosess av orden 1, med sesong hvis O_t skrives på form

$$O_t = \Phi O_{t-s} + \epsilon_t$$

hvor s er lengden av sesongvariasjoner, $s = 12$ for månedstall og $s = 4$ for kvartalstall, $s = 6$ for terminstall. $|\Phi| < 1$ for $\{O_t\}$ blir en stasjonær prosess. ϵ_t er uavhengig med O_{t-1}, O_{t-2}, \dots . Vi antar at $E(\epsilon_t) = 0$ og $Var(\epsilon_t) = \sigma_\epsilon^2$. La $s = 12$. O_t skrives slik

$$O_t = \Phi O_{t-12} + \epsilon_t \quad (3.50)$$

Autokovariansen for lag k er

$$\begin{aligned}\gamma_k = Cov(O_t, O_{t+k}) &= Cov(O_t, \Phi O_{t+k-12} + \epsilon_{t+k}) \\ &= \Phi Cov(O_t, O_{t+k-12}) \\ &= \Phi \gamma_{k-12}\end{aligned}$$

autokorrelasjonen for lag k blir

$$\rho_k = \Phi \rho_{k-12}$$

For $k = 1$ og $k = 11$ er

$$\begin{aligned}\rho_1 &= \Phi \rho_{11} \\ \rho_{11} &= \Phi \rho_{-1} = \Phi \rho_1\end{aligned}$$

Dermed er $\rho_1 = \rho_{11} = 0$. Vi kan vise med samme måte at $\rho_k = 0$ for $k \neq 12, 24, 36, \dots$

For $k = 12$, er

$$\rho_{12} = \Phi \rho_0 = \Phi$$

og

$$\rho_{12k} = \Phi^k,$$

for $k = 1, 2, \dots$

Man kan vise at for en sesongautoregressiv AR(1) s , forsvinner autokorrelasjonen ρ_k for lag $k \neq s, 2s, 3s, \dots$, og den er en avtagende funksjon med $k = s, 2s, 3s, \dots$

b. Autoregressiv modell av orden P , AR(P) s , med sesongvariasjoner

O_t er en AR(P) s modell hvis O_t skrives på form

$$O_t = \Phi_1 O_{t-s} + \Phi_2 O_{t-2s} + \dots + \Phi_P O_{t-Ps} + \epsilon_t \quad (3.51)$$

hvor s er perioden av sesongvariasjoner. $\{\epsilon_t\}$ er hvit støy prosess med forventning 0 og varians σ_ϵ^2 . ϵ_t er uavhengig med $O_{t-1}, O_{t-2}, \dots, O_t$ i (3.51) kan skrives ved en annen form

$$(1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}) O_t = \epsilon_t$$

La

$$\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$$

$\Phi(B^s)$ er en polynom av B^s med orden P . For $\{O_t\}$ blir en stasjonær prosess, må røttene av $\Phi(B^s) = 0$ være større enn 1 i absoluttverdi. Man kan vise at autokorrelasjonen ρ_k av en AR(Ps) prosess er en avtagende funksjon med lag k , for $k = s, 2s, 3s, \dots$ og forsvunnet for lag $k \neq s, 2s, 3s, \dots$

c. Glidende gjennomsnitt modell av orden Q , MA(Q) s , med sesongvariasjoner

O_t er en MA(Q) s modell hvis

$$O_t = \epsilon_t - \Theta_1 \epsilon_{t-s} - \Theta_2 \epsilon_{t-2s} - \dots - \Theta_Q \epsilon_{t-Qs} \quad (3.52)$$

hvor s er perioden av sesongvariasjoner. $\{\epsilon_t\}$ er hvit støy prosess med forventning 0 og varians σ_ϵ^2 . Man kan vise at autokorrelasjonen ρ_k eksisterer bare i lag $k = s, 2s, \dots$ og beskrive ved

$$\rho_{ks} = \frac{-\Theta_k + \Theta_1\Theta_{k+1} + \dots + \Theta_{Q-k}\Theta_Q}{1 + \Theta_1^2 + \dots + \Theta_Q^2}$$

for $k = 1, 2, \dots, Q$. $\{O_t\}$ er en stasjonær prosess.

d. Multippel ARMA(p, q)(P, Q) s modell med sesongvariasjoner

Modellen inneholder sesong- og ikke sesongvariasjonsledd og beskrives ved ligningen

$$\phi(B)\Phi(B^s)O_t = \theta(B)\Theta(B^s)\epsilon_t \quad (3.53)$$

hvor $\{\epsilon_t\}$ er hvit støy prosess med forventning 0, og varians σ_ϵ^2 . $\phi(B)$, $\Phi(B^s)$, $\theta(B)$ og $\Theta(B^s)$ er polynomer av ordene p , P , q og Q , henholdsvis.

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \Phi(B^s) &= 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps} \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \\ \Theta(B^s) &= 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs} \end{aligned} \quad (3.54)$$

Vi ser at i venstre side av (3.53) er en AR prosess med orden $p + Ps$ og i høyre side er en MA prosess med orden $q + Qs$.

Eksempel. Vi antar at $P = q = 1$ og $p = Q = 0$, $s = 12$. Modellen blir

$$(1 - \Phi B^{12})O_t = (1 - \theta B)\epsilon_t$$

altså

$$O_t = \Phi O_{t-12} + \epsilon_t - \theta \epsilon_{t-1}$$

Autokovariansen γ_k blir

$$\begin{aligned} \gamma_0 &= Cov(O_t, O_t) = Cov(\Phi O_{t-12} + \epsilon_t - \theta \epsilon_{t-1}, \Phi O_{t-12} + \epsilon_t - \theta \epsilon_{t-1}) \\ &= \Phi^2 Cov(O_t, O_t) + Cov(\epsilon_t, \epsilon_t) + \theta^2 Cov(\epsilon_{t-1}, \epsilon_{t-1}) \\ &= \Phi^2 \gamma_0 + \sigma_\epsilon^2 + \theta^2 \sigma_\epsilon^2 \end{aligned}$$

siden ϵ_t er uavhengig med O_{t-12} . Dermed er

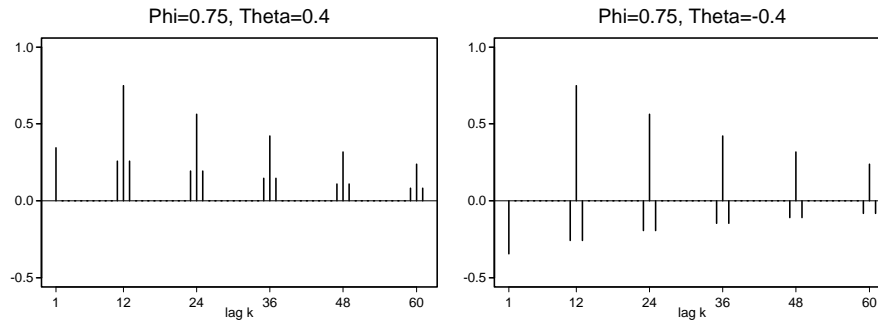
$$\gamma_0 = \frac{1 + \theta^2}{1 - \Phi^2}$$

for $k = 1$ er

$$\begin{aligned} \gamma_1 &= Cov(O_t, O_{t+1}) = Cov(O_t, \Phi O_{t-11} + \epsilon_{t+1} - \theta \epsilon_t) \\ &= \Phi \gamma_{11} - \theta \sigma_\epsilon^2 \end{aligned}$$

og for $k \geq 2$, er

$$\gamma_k = \Phi \gamma_{k-12}$$



Figur 3.15: Autokorrelasjonene for $ARMA(0,1)(1,0)_{12}$

Man kan vise at

$$\begin{aligned} \rho_{12k} &= \Phi^k, & \text{for } k = 1, 2, \dots \\ \rho_{12k-1} &= \rho_{12k+1} = -\frac{\theta}{1+\theta^2}\Phi^k, & \text{for } k = 0, 1, 2, \dots \end{aligned} \quad (3.55)$$

Figur 3.15 viser autokorrelasjonene som er beregnet ved (3.55), med ($\Phi = 0.75$ og $\theta = 0.4$) og ($\Phi = 0.75$ og $\theta = -0.4$).

Vi ser at når $\theta < 0$ blir autokorrelasjonen negative i lagene $(12k-1)$ og $(12k+1)$, for $k = 0, 1, 2, \dots$.

e. Multippel ikke stasjonær $ARIMA(p, d, q)(P, D, Q)_s$ modell med sesongvariasjoner

Vi treffer denne modellen veldig ofte i X-12-ARIMA. Modellen skrives på en generell form

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^D O_t = \theta(B)\Theta(B^s)\epsilon_t \quad (3.56)$$

hvor s er sesongperioden. $\{\epsilon_t\}$ er hvit støy prosess med forventning 0, og varians σ_ϵ^2 . $\phi(B)$, $\Phi(B^s)$, $\theta(B)$ og $\Theta(B^s)$ er polynomer med ordene p , q , P og Q , henholdsvis, og gitt ved (3.54). $\nabla = 1 - B$ og $\nabla_s = 1 - B^s$. d og D er antall ganger differansen skal tas for O_t slik at $\nabla^d\nabla_s^D O_t$ er en stasjonær prosess.

Eksempel. Vi antar at

$$p = P = 0, \quad q = Q = 1, \quad d = D = 1 \quad \text{og} \quad s = 12$$

$ARIMA(p, d, q)(P, D, Q)_s$ blir $ARIMA(0,1,1)(0,1,1)_{12}$ modell. Det er modellen som Box og Jenkins har brukt for å tilpasse internasjonale flyreisedata og kalles "airline model". Ligningen for O_t blir

$$(1 - B)(1 - B^{12})O_t = (1 - \theta B)(1 - \Theta B^{12})\epsilon_t \quad (3.57)$$

hvor $\{\epsilon_t\}$ er en hvit støy prosess. En eksklusiv form for O_t er

$$O_t = O_{t-1} + O_{t-12} - O_{t-13} + \epsilon_t - \theta\epsilon_{t-1} - \Theta\epsilon_{t-12} + \theta\Theta\epsilon_{t-13}$$

Vi ser at verdien i tidspunkt t er avhengig av verdien i et lag tilbake, $t-1$ og verdier i året før $t-12$ og $t-13$. La

$$\begin{aligned} W_t &= (1 - B)(1 - B^{12})O_t \\ &= (1 - \theta B)(1 - \Theta B^{12})\epsilon_t \\ &= \epsilon_t - \theta\epsilon_{t-1} - \Theta\epsilon_{t-12} + \theta\Theta\epsilon_{t-13} \end{aligned}$$

Da er W_t en stasjonær prosess med $E(W_t) = 0$.

Autokovariansen for W_t blir

$$\begin{aligned}\gamma_0 &= Cov(W_t, W_t) = Var(W_t) \\ &= Cov(\epsilon_t - \theta\epsilon_{t-1} - \Theta\epsilon_{t-12} + \theta\Theta\epsilon_{t-13}, \epsilon_t - \theta\epsilon_{t-1} - \Theta\epsilon_{t-12} + \theta\Theta\epsilon_{t-13}) \\ &= (1 + \theta^2 + \Theta^2 + \theta^2\Theta^2)\sigma_\epsilon^2 = (1 + \theta^2)(1 + \Theta^2)\sigma_\epsilon^2\end{aligned}$$

$$\begin{aligned}\gamma_1 &= Cov(W_t, W_{t+1}) \\ &= Cov(\epsilon_t - \theta\epsilon_{t-1} - \Theta\epsilon_{t-12} + \theta\Theta\epsilon_{t-13}, \epsilon_{t+1} - \theta\epsilon_t - \Theta\epsilon_{t-11} + \theta\Theta\epsilon_{t-12}) \\ &= -\theta Cov(\epsilon_t, \epsilon_t) - \theta\Theta^2 Cov(\epsilon_{t-12}, \epsilon_{t-12}) \\ &= -\theta(1 + \Theta^2)\sigma_\epsilon^2\end{aligned}$$

Vi kan vise at

$$\begin{aligned}\gamma_{11} &= \theta\Theta\sigma_\epsilon^2 \\ \gamma_{12} &= -\Theta(1 + \theta^2)\sigma_\epsilon^2 \\ \gamma_{13} &= \theta\Theta\sigma_\epsilon^2 \\ \gamma_j &= 0, \quad \text{ellers}\end{aligned}$$

Dermed er

$$\begin{aligned}\rho_1 &= \frac{-\theta}{1 + \theta^2}, \\ \rho_{11} &= \frac{\theta\Theta}{(1 + \theta^2)(1 + \Theta^2)} = \rho_{13}, \\ \rho_{12} &= \frac{-\Theta}{1 + \Theta^2}, \\ \rho_j &= 0, \quad \text{ellers}\end{aligned}\tag{3.58}$$

Vi simulerte 250 observasjoner fra ARIMA(0,1,1)(0,1,1)₁₂ modell, med $\theta = 0.8$ og $\Theta = 0.6$, hvor $\epsilon_t \sim N(0,1)$. La $\{O_t\}$ være denne tidsserien. Dataene er plottet i venstre side av figur 3.16. Siden kurven er en voksende med t , viser det at forventningen er avhengig av t . $\{O_t\}$ blir dermed ikke en stasjonær prosess. Vi lager en ny tidsserie W_t ved å ta differansen av O_t i lag 1 og 12, $W_t = (1 - B)(1 - B^{12})O_t$. Den plottes i høyre side av figuren. Dataene varierer rundt null-aksen. Det ser ut som $\{W_t\}$ er en stasjonær prosess. Forventningen er ikke avhengig av t lenger.

Autokorrelasjonene for W_t beregnes ved (3.58), med $\theta = 0.8$ og $\Theta = 0.6$, blir

$$\rho_1 = -0.4878, \quad \rho_{11} = \rho_{13} = 0.2152, \quad \rho_{12} = -0.4412,$$

og

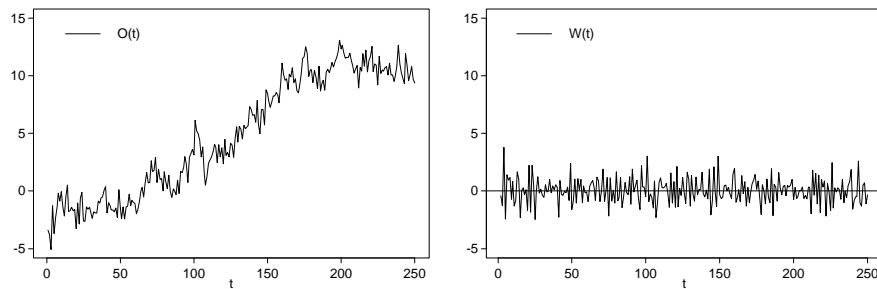
$$\rho_j = 0 \text{ for } j \neq 1, 11, 12 \text{ og } 13.$$

3.2.22 Framskrivinger for modell med sesongvariasjoner

Framskrivningene for en ARIMA(p, d, q)(P, D, Q) beregnes ved (3.42).

Eksempel. ARIMA(0,1,1)(0,1,1)₁₂.

$$O_t = O_{t-1} + O_{t-12} - O_{t-13} + \epsilon_t - \theta\epsilon_{t-1} - \Theta\epsilon_{t-12} + \theta\Theta\epsilon_{t-13}$$



Figur 3.16: Simulerte data for $ARIMA(0, 1, 1)(0, 1, 1)_{12}$, med $\theta = 0.8$ og $\Theta = 0.6$

hvor $\{\epsilon_t\}$ er hvit støy prosess, med forventning 0, og varians σ_ϵ^2 . Dessuten er ϵ_t uavhengig med $O_{t-1}, O_{t-2}, \dots, t = 1, 2, \dots, n$.

Framskrivningene for lag 1, 2 og lag 14 beregnes ved

$$\begin{aligned}\hat{O}_n(1) &= E(O_{n+1}|O_n, O_{n-1}, \dots) \\ &= E(O_n + O_{n-11} - O_{n-12} + \epsilon_{n+1} - \theta\epsilon_n - \Theta\epsilon_{n-11} + \theta\Theta\epsilon_{n-12}|O_n, O_{n-1}, \dots) \\ &= O_n + O_{n-11} - O_{n-12} - \theta\epsilon_n - \Theta\epsilon_{n-11} + \theta\Theta\epsilon_{n-12}\end{aligned}$$

$$\begin{aligned}\hat{O}_n(2) &= E(O_{n+2}|O_n, O_{n-1}, \dots) \\ &= E(O_{n+1} + O_{n-10} - O_{n-11} + \epsilon_{n+2} - \theta\epsilon_{n+1} - \Theta\epsilon_{n-10} + \theta\Theta\epsilon_{n-11}|O_n, O_{n-1}, \dots) \\ &= \hat{O}_n(1) + O_{n-10} - O_{n-11} - \Theta\epsilon_{n-10} + \theta\Theta\epsilon_{n-11}\end{aligned}$$

og for $l = 14$

$$\begin{aligned}\hat{O}_n(14) &= E(O_{n+14}|O_n, O_{n-1}, \dots) \\ &= E(O_{n+13} + O_{n+2} - O_{n+1} + \epsilon_{n+14} - \theta\epsilon_{n+13} - \Theta\epsilon_{n+2} + \theta\Theta\epsilon_{n+1}) \\ &= \hat{O}_{n+13} + \hat{O}_{n+2} - \hat{O}_{n+1}\end{aligned}$$

siden $\epsilon_{n+14}, \epsilon_{n+13}, \epsilon_{n+2}$ og ϵ_{n+1} er uavhengig med O_n, O_{n-1}, \dots . For $l > 13$ er framskrivningene ikke avhengig av støyledene av den glidende gjennomsnitt prosessen.

Vi får

$$\begin{aligned}O_{n+1} &= O_n + O_{n-11} - O_{n-12} + \epsilon_{n+1} - \theta\epsilon_n - \Theta\epsilon_{n-11} + \theta\Theta\epsilon_{n-12} \\ \hat{O}_n(1) &= O_n + O_{n-11} - O_{n-12} - \theta\epsilon_n - \Theta\epsilon_{n-11} + \theta\Theta\epsilon_{n-12}\end{aligned}$$

Framskrivingsfeilen i lag 1, blir

$$\begin{aligned}e_n(1) &= O_{n+1} - \hat{O}_n(1) \\ &= \epsilon_{n+1}\end{aligned}$$

Variansen for $e_n(1)$ blir

$$\text{Var}(e_n(1)) = \sigma_\epsilon^2$$

Vi antar at O_t er normalfordeling, så er $e_n(1)$ også en normalfordeling. 95% konfidensintervallet for O_{n+1} blir

$$\left[\hat{O}_n(1) - 1.96\sqrt{\text{Var}(e_n(1))} \leq O_{n+1} \leq \hat{O}_n(1) + 1.96\sqrt{\text{Var}(e_n(1))} \right]$$

På samme måte kan vi lage et 95% konfidensintervall for O_{n+2}, O_{n+3}, \dots

3.2.23 Akaike's AIC og BIC kriterier

La O_1, O_2, \dots, O_n være en tidsserie av observasjoner. Vi antar at dataene er tilpasset med en ARIMA modell. For å vurdere tilpasningen mellom data og modellen har Akaike (1973, 1974) innført et kriterium som vi kaller AIC (Akaike's information criterion) og dette er definert ved

$$AIC(k) = -2\ln(\text{maksimum likelihood}) + 2k \quad (3.59)$$

hvor k er antall parametre i modellen. Maksimum likelihood funksjon er en funksjon av parametrene i modellen, gitt dataene. For eksempel, for en ARMA(p, q) modell

$$O_t = \phi_1 O_{t-1} + \phi_2 O_{t-2} + \dots + \phi_p O_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q}$$

hvor $\epsilon_t \sim N(0, \sigma_\epsilon^2)$, er log-likelihood funksjon gitt ved

$$\ln L = -\frac{n}{2} \ln 2\pi \sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2} S(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q) \quad (3.60)$$

hvor $S(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ er den betingede kvadratsumfunksjonen.

$$S(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q) = \sum_{t=-\infty}^n [E(\epsilon_t | \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, O_1, \dots, O_n)]^2$$

hvor $E(\epsilon_t | \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, O_1, \dots, O_n)$ er den betingede forventningen av ϵ_t gitt parametre og dataene. Gitt ϕ_i og θ_j , kan vi maksimere $\ln L$ med hensyn til σ_ϵ^2 ved

$$\frac{\partial \ln L}{\partial \sigma_\epsilon^2} = 0$$

Vi får

$$\hat{\sigma}_\epsilon^2 = \frac{S(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)}{n}$$

(3.60) blir

$$\ln L = -\frac{n}{2} \ln(\hat{\sigma}_\epsilon^2) - \frac{n}{2} (\ln(2\pi) + 1) \quad (3.61)$$

siden siste ledd av (3.61) er en konstant, får vi

$$AIC(k) = n \ln(\hat{\sigma}_\epsilon^2) + 2k \quad (3.62)$$

Akaike (1978, 1979) har innført et annet kriterium det såkalte BIC(k) som er gitt ved

$$BIC(k) = n \ln(\hat{\sigma}_\epsilon^2) - (n - k) \ln\left(1 - \frac{k}{n}\right) + k \ln(n) + k \ln\left[\left(\frac{\hat{\sigma}_O^2}{\hat{\sigma}_\epsilon^2} - 1\right) / k\right] \quad (3.63)$$

hvor $\hat{\sigma}_O^2$ er variansen av dataene.

I de fleste statistiske programpakker som SAS, S-Plus eller Minitab blir AIC og BIC listet ut sammen med de estimerte parametre av ARIMA modellen som tilpasses med dataene. Vi velger modellen som gir minst AIC.

3.2.24 Anvendelse av autokorrelasjonsfunksjoner ved å finne orden til en ARIMA modell

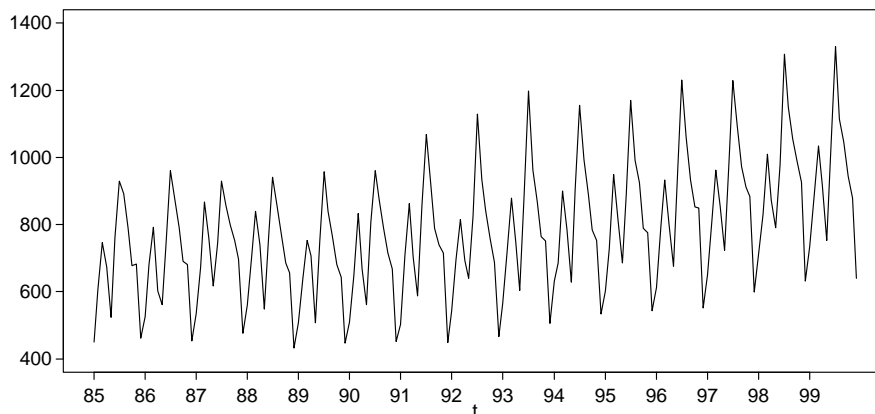
I avsnittene ovenfor har vi beskrevet $AR(p)$, $MA(q)$, $ARMA(p, q)$ og den generelle prosessen med sesongvariasjoner $ARIMA(p, d, q)(P, D, Q)$, samt deres autokorrelasjonsfunksjoner ρ_k . For hver modell har vi et bilde av autokorrelasjonsfunksjon. Når dataene er observert, kan vi estimere autokorrelasjonsfunksjonen $\hat{\rho}_k$ og den partielle autokorrelasjonsfunksjonen $\hat{\phi}_{kk}$ ved (3.40) og (3.41). Ved å plote dem mot lag k og ved å anvende reglene i avsnitt 3.2.18 kan vi gjette verdiene for p , q , P , Q , d og D . Vanligvis tar vi modeller med $p, q, P, Q = 0, 1, 2$, $d = 0, 1, 2$ og $D = 0, 1$. Hvis $\hat{\rho}_k$ er signifikant større enn 0 i lag 12, 24, 36, \dots , vet vi at dataene er påvirket av sesongeffekter. En modell med sesongvariasjoner skal brukes. Når dataene har en god tilpasning for flere modeller, vil vi velge den som gir minst AIC. Framskrivningene skal beregnes med den beste modellen.

3.2.25 Eksempel: Hotellovernattinger for nordmenn

I dette eksemplet vil vi vise hvordan kan man anvende autokorrelasjonsfunksjon og partiell autokorrelasjonsfunksjon $\widehat{\rho}_k$ og $\widehat{\phi}_{kk}$ i praksis for å finne en passende ARIMA modell for dataene.

1. Å plote rådata

Rådata O_t (i 1000) er plottet i figur 3.17.



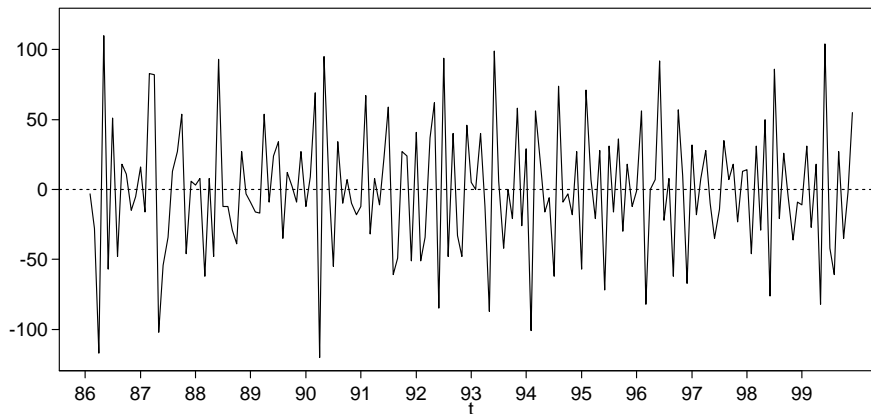
Figur 3.17: Rådata av hotellovernattinger for nordmenn

Verdiene av y -aksen varierer mellom 400 og 1400 med gjennomsnittet er 782 og standard avviket er 183.5.

2. Å sjekke om stasjonærhet og sesongvariasjoner

I figuren ser vi at det er en økning for trenden og variasjonen i rådata blir stadig større med tid. Tidsserien ser ikke ut som den er stasjonær. Det er et klart sesongmønster. Juni, juli og august er månedene som har flest gjesteovernattinger, mens desember og januar er de to månedene som har færrest besøkende. Vi bruker SAS/ETS for å beregne $\hat{\rho}_k$ (ACF) og $\hat{\phi}_{kk}$ (PACF) fra rådataene. Programmet gir verdiene samt et plot for disse størrelsene etter lag k med 95% konfidensintervaller (se figur 3.18). Vi ser at $\hat{\rho}_k$ og $\hat{\phi}_{kk}$ er signifikante større

Merk at Y_t er kortere serie enn O_t , 13 færre observasjoner. Tidsserien går fra februar 1986 til desember 1999. Tallene er plottet i figur 3.19. Det ser ut som det er en stasjonær prosess. $\hat{\rho}_k$ og $\hat{\phi}_{kk}$ av Y_k er vist i figur 3.20. Vi skal anvende regler i avsnitt 3.2.18 for å identifisere verdiene av p , P , q og Q .



Figur 3.19: Y_t

4. Å identifisere p , P , q , Q og estimere ϕ -er, θ -er

Siden $\hat{\phi}_{kk}$ er signifikant større enn null i lag 1 og 12, vil vi velge $p = 1$ og $P = 1$. For å anslå verdiene til q og Q bruker vi $\hat{\rho}_k$. Vi prøver med $q = 1$ og $Q = 1$. Modellen som vi velge i første omgang for Y_t er ARMA(1,1)(1,1). Dermed for O_t blir ARIMA(1,1,1)(1,1,1). Vi får følgende resultater

Tabell 3.7: Estimerte verdier for parametre av ARIMA(1,1,1)(1,1,1)

	Est.verdi	t -verdi	
$\hat{\theta}$	0.8248	12.94	signifikant på 5% nivå
$\hat{\Theta}$	0.7004	6.42	signifikant på 5% nivå
$\hat{\phi}$	0.1504	1.47	ikke sign. på 5% nivå
$\hat{\Phi}$	0.1661	1.21	ikke sign. på 5% nivå
AIC=1646			

Siden $\hat{\phi}$ og $\hat{\Phi}$ er ikke signifikante på 5% nivå vil vi prøve med modellene ARIMA(1,1,1)(0,1,1) og ARIMA(0,1,1)(1,1,1). $\hat{\phi}$ og $\hat{\Phi}$ i tabellene 3.8 og 3.9 er ikke signifikante. Modellen blir redusert til ARIMA(0,1,1)(0,1,1). Resultater er listet i tabell 3.10. ARIMA(0,1,1)(0,1,1) er valgt for dataene. Ligningen for O_t , blir

$$(1 - B)(1 - B^{12})O_t = (1 - \theta B)(1 - \Theta B^{12})\epsilon_t$$

med $\hat{\theta}$ og $\hat{\Theta}$ fra tabell 3.10, får vi

$$O_t = O_{t-1} + O_{t-12} - O_{t-13} + \epsilon_t - 0.7396\epsilon_{t-1} - 0.5747\epsilon_{t-12} + 0.4250\epsilon_{t-13} \quad (3.64)$$

5. Å sjekke om det er en hvit støy prosess for residualer

```

Period(s) of Differencing = 1,12.
Mean of working series = -0.40719
Standard deviation      = 45.34448
Number of observations = 167
NOTE: The first 13 observations were eliminated by differencing.

Autocorrelations

Lag Covariance Correlation -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0 2056.122 1.00000 *****
1 -956.330 -0.46511 *****
2 90.832465 0.04418 *****
3 -177.415 -0.08629 *****
4 -80.623965 -0.03921 *****
5 167.998 0.08171 *****
6 123.089 0.05986 *****
7 -190.653 -0.09272 *****
8 -196.121 -0.09538 *****
9 455.741 0.22165 *****
10 -309.457 -0.15051 *****
11 489.790 0.23821 *****
12 -635.882 -0.30926 *****
13 -37.450319 -0.01821 *****
14 270.230 0.13143 *****
15 50.341162 0.02448 *****
16 -17.492489 -0.00851 *****
17 -43.448338 -0.02113 *****
18 92.280168 0.04488 *****
19 -356.699 -0.17348 *****
20 438.112 0.21308 *****
21 -117.351 -0.05707 *****
22 -296.297 -0.14410 *****
23 401.254 0.19515 *****
24 -372.836 -0.18133 *****
"." marks two standard errors

Partial Autocorrelations

Lag Correlation -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
1 -0.46511 *****
2 -0.21968 *****
3 -0.21908 *****
4 -0.25211 *****
5 -0.12405 *****
6 0.02860 *****
7 -0.05942 *****
8 -0.22726 *****
9 0.09177 *****
10 -0.01911 *****
11 0.22830 *****
12 -0.09235 *****
13 -0.23307 *****
14 -0.06166 *****
15 -0.01194 *****
16 -0.05070 *****
17 -0.02887 *****
18 0.14049 *****
19 -0.11874 *****
20 -0.09269 *****
21 0.13959 *****
22 -0.15088 *****
23 0.13977 *****
24 -0.11718 *****

```

Figur 3.20: \widehat{ACF} og \widehat{PACF} av Y_t Tabell 3.8: Estimerte verdier for parametre av $ARIMA(1,1,1)(0,1,1)$

	Est.verdi	t -verdi	
$\hat{\theta}$	0.8092	12.26	signifikant på 5% nivå
$\hat{\Theta}$	0.5955	8.58	signifikant på 5% nivå
$\hat{\phi}$	0.1303	1.25	ikke sign. på 5% nivå
AIC=1646			

Dersom modellen er riktig og de estimerte parametre er nær de sanne verdiene, er residualene uavhengige identiske normalfordeling.

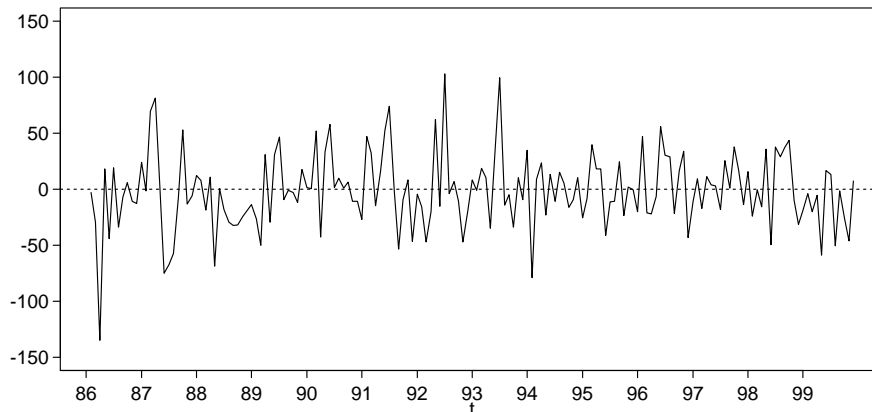
Residualene er definert som differansen mellom observerte dataene og deres predikerte verdier. Tallene er plottet i figur 3.21. For å sjekke om det er en hvit støy prosess for residualer, betrakter vi deres autokorrelasjonfunksjon $\hat{\rho}_{R,k}$. Vi ser i figur 3.22 at det er ingen lag k der $\hat{\rho}_{R,k}$ er sterkt signifikant forskjellig fra null. Dette viser at residualene er en hvit støy prosess.

Tabell 3.9: *Estimerte verdier for parametre av ARIMA(0,1,1)(1,1,1)*

	Est.verdi	<i>t</i> -verdi	
$\hat{\theta}$	0.7447	13.57	signifikant på 5% nivå
$\hat{\Theta}$	0.6700	5.88	signifikant på 5% nivå
$\hat{\phi}$	0.1470	1.04	ikke sign. på 5% nivå
AIC=1646			

Tabell 3.10: *Estimerte verdier for parametre av ARIMA(0,1,1)(0,1,1)*

	Est.verdi	<i>t</i> -verdi	
$\hat{\theta}$	0.7396	13.60	signifikant på 5% nivå
$\hat{\Theta}$	0.5747	8.35	signifikant på 5% nivå
AIC=1645			

Figur 3.21: *Residualene*

I beregningen for konfidensintervaller av framskrivinger antar vi at ϵ_t har normalfordelingen slik at (3.45) kan anvendes. Tettheten av residualer er plottet sammen med den normalfordelingen $N(0, \hat{\sigma}_\epsilon^2)$ i figur 3.23. Vi ser at de to kurvene har store avviker på toppen. Residualene blir derfor ikke perfekt med normalfordelingen. Vi får samme konklusjonen ved normalplott der normalfordelingen plottes med en rett linje. Vi har derfor ikke nøyaktige konfidensintervaller for residualene.

6. Framskrivinger

Predikerte verdier og rådataene er plottet i figur 3.24. Verdiene etter desember 1999 er framskrivinger. Vi ser at modellen har en god tilpasning med dataene.

Framskrivningen for januar 2000 beregnes ved å ta forventningen av (3.64), gitt

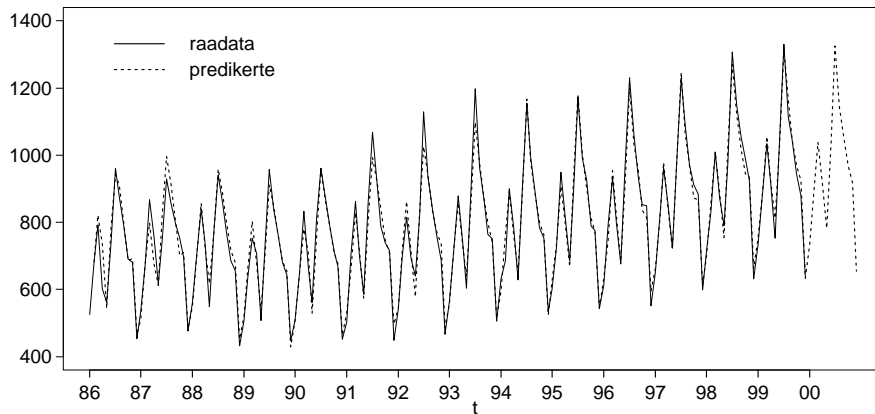
(fortsatt)

år	måned	O_t	\hat{O}_t	$\hat{\epsilon}_t$
1999	mai	753	811.75	-58.74
1999	juni	1043	1026.16	16.84
1999	juli	1331	1317.85	13.15
1999	august	1114	1164.54	-50.53
1999	september	1045	1046.33	-1.32
1999	oktober	944	970.63	-26.63
1999	november	879	924.97	-45.97
1999	desember	640	632.83	7.17

Dermed er

$$\begin{aligned}\hat{O}_{jan.2000} &= 640 + 735 - 632 - 0.7396 \times 7.17 - 0.5747 \times (-19.31) + 0.4250 \times (-31.30) \\ &= 735.49\end{aligned}$$

$$\begin{aligned}\hat{O}_{feb.2000} &= \hat{O}_{jan.2000} + O_{feb.1999} - O_{jan.1999} - \hat{\Theta}\epsilon_{feb.1999} + \hat{\theta}\hat{\Theta}\epsilon_{jan.1999} \\ &= 735.49 + 883 - 735 - (0.5747)(-3.88) + (0.7396)(0.5747)(-19.31) \\ &= 877.50\end{aligned}$$



Figur 3.24: Rådata og predikerte verdier (framskrivinger)

Egentlig har vi tall for hotellovernattinger for nordmenn fra januar 1985 til juni 2000, men i beregningen vil vi bruke dataene fram til desember 1999 og beregne framskrivinger for 12 måneder framover. Tallene skal sammenlignes med de faktiske observerte verdiene for å vurdere hvor god modellen er. Vi ser i tabell 3.11 at de observerte verdiene for januar, februar, . . . , juni 2000 ikke faller langt fra deres framskrivinger og ligger innenfor et 95% konfidensintervaller.

3.3 Oppsummering

Et formål ved tidsserieanalyse er å kunne framskrive en tidsserie fra de observerte dataene. For å gjøre dette må man identifisere en ARIMA modell for tidsserien. Avhengig av dataene har vi et mønster for den estimerte autokorrelasjonfunksjonen (se avsnitt 3.2.19 for beregninger $\hat{\rho}_k$ og $\hat{\phi}_{kk}$). Mønsteret skal

Tabell 3.11: *Framskrivningene og observerte verdiene for år 2000*

	Nedre	Framsk.	Øvre	Obser.	diff.
januar	671	735	799	723	-12
februar	811	877	943	876	-1
mars	971	1039	1107	1075	36
april	845	915	986	858	-57
mai	711	783	855	773	-10
juni	964	1038	1112	1004	-34
juli	1250	1326	1402		
august	1066	1144	1221		
september	974	1054	1133		
oktober	886	968	1049		
november	835	918	1001		
desember	570	655	740		

sammenlignes med den teoretiske autokorrelasjonfunksjonen for å anslå p , d , q , P , D og Q ved regler i avsnitt 3.2.18. Det kan ha mange modeller som gir gode tilpasninger for dataene, men vi vil velge den som har minst AIC (se avsnitt 3.2.23). Vi vil også velge en enklest mulig modell når det er liten forskjell på AIC. Residualene som er differansen mellom observerte og predikerte dataene må sjekkes om det er en hvit støy prosess og den empiriske fordelingen må være rimelig nær normalfordeling, ellers vil beregningen av konfidensintervaller være for unøyaktig.

Kapittel 4

Sesongjustering

4.1 Innledning

Vi antar at en tidsserie $\{O_t\}$ kan skrives på formen

$$O_t = T_t + S_t + I_t \quad (4.1)$$

hvor T_t , S_t og I_t er henholdsvis trend, sesong- og den irregulære komponenten. Disse komponentene skal beskrives i de neste avsnittene. Rådataene til en tidsserie i økonomisk statistikk er ofte påvirket av

(i) kalendereffekter som skyldes ulikt antall arbeidsdager fra måned til måned, eller bevegelige helligdager som påske, pinse og Kristi himmelfartsdag.

(ii) eksterne begivenheter (kalles intervensjoner), f.eks streik, devaluering, kraftige endringer i renta, osv.

(i) medfører at vi trekker feil konklusjon om aktiviteten i januar og mars for en klesbutikk ved å sammenligne bare rådata. De to månedene har til tross for samme lengden (31 dager), ulikt antall mandager, tirsdager, . . . , søndager. Disse dagene har ofte forskjellige vekter eller om en vil forskjellig nivå på handelen de enkelte ukedagene. Påsken kommer ofte i april. I 1986, 1989, 1991, 1997 var påsken i mars. Vi får derfor en forskyvning av handelen fra april til mars. Å sammenligne endringen i prosent av rådata fra mars til april i 1986 med den samme størrelsen i 1987 gir et galt inntrykk, siden påsken i 1987 faller seint i april (19 april). Forskyvningen av påske fra mars i 1986 til april 1986 påvirket handelen.

En streik kan ha rammet aktiviteten i en fabrikk i en periode. Vi får et annet mønster for rådataene i denne perioden.

Vi behandler derfor i dette kapitlet også behovet for prekorrigeringer for disse effektene fra rådata før man utfører en sesongjustering.

4.2 Intervensjoner

Box og Tiao (1965) utviklet intervensjonsanalyse for å studere tidsserier som er påvirket av eksterne begivenheter. La $\{u_t\}$ være en tidsserie. På tidspunktet t observerer vi ikke direkte verdien til u_t , men via O_t som er en sum av u_t og en funksjon $f(t)$, der $f(t)$ beskriver en ekstern begivenhet som vi kaller en

intervensjon ved tidspunkt t .

$$O_t = f(t) + u_t \quad (4.2)$$

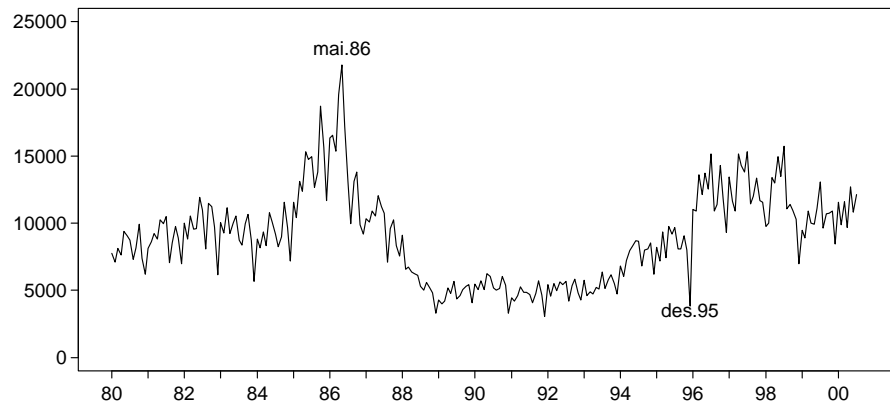
$f(t)$ kalles *overføringsfunksjon* (transfer function) og skrives generelt

$$f(t) = \frac{\omega(B)}{\delta(B)} X_t \quad (4.3)$$

hvor

$$\begin{aligned} \omega(B) &= \omega_0 + \omega_1 B + \omega_2 B^2 + \dots + \omega_h B^h \\ \delta(B) &= 1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r \end{aligned}$$

Eksempel: Registrerte nye biler i Norge. Tallene er observert fra januar 1980 til juli 2000 og plottet i figur 4.1. Vi ser at det er mange uvanlig lave og høye punkter i rådata. Devalueringen i mai 1986 førte til at salget av antall nye personbiler



Figur 4.1: *Antall nye personbiler*

gikk kraftig ned fra mai til august før tallet økte i september for deretter å synke ytterligere (se nedenfor tabell). Reduksjon fra juli til august er -23.70%, mens økningen fra august til september 31.26%.

år	mai	juni	juli	august	september
1986	21793	17241	13079	9982	13103

Lavt tall i desember 1995, $O_{des.1995}=3836$, har trolig sammenheng med lavere bilavgifter fra januar 1996.

okt.95	nov.95	des.95	jan.96	feb.96
9078	8006	3836	11010	10920

Eksempel: Eksportserien. Da Kuwait ble invadert av Irak i begynnelsen av august 1990, økte prisen på råolje kraftig. I slutten av september ble prisen for Nordsjølje notert til over 40 US\$ pr. fat. Prisen gikk litt opp og ned rundt 40 US\$ til midten av oktober. Deretter falt prisen til 30 US\$ pr. fat. Konsekvensen var

at verdien for norsk eksport i september, oktober 1990 og flere måneder senere ble svært høy.

La

$$P_t^{(T)} = \begin{cases} 1, & t = T, \\ 0, & t \neq T. \end{cases} \quad (4.4)$$

og

$$S_t^{(T)} = \begin{cases} 1, & t \geq T, \\ 0, & t < T. \end{cases} \quad (4.5)$$

$P_t^{(T)}$ kalles en *pulsfunksjon* og $S_t^{(T)}$ en *trappfunksjon*.

Merk at

1.

$$P_t^{(T)} = S_t^{(T)} - S_{t-1}^{(T)} = (1 - B)S_t^{(T)} \quad (4.6)$$

2. Overføringsfunksjonen for en intervensjon i tidspunkt T , med effekten ω , men der virkningen blir forsinket med b tidspunkter, beskrives ved

$$\omega B^b P_t^{(T)}, \quad \text{eller} \quad \omega B^b S_t^{(T)} \quad (4.7)$$

som er avhengig av typen intervensjon.

3. Samme som i (2.) men effekten blir gradvis endret for tidspunktene etter T , slik at $f(t)$ er

$$\frac{\omega B^b}{(1 - \delta B)} P_t^{(T)} \quad \text{eller} \quad \frac{\omega B^b}{(1 - \delta B)} S_t^{(T)} \quad (4.8)$$

ved input er en pulsfunksjon eller trappfunksjon, $0 < \delta < 1$. (4.7) er et spesielt tilfelle av (4.8) med $\delta = 0$.

Vi får fra (4.8)

$$\begin{aligned} \frac{\omega B^b}{(1 - \delta B)} P_t^{(T)} &= (1 + \delta B + \delta^2 B^2 + \dots) \omega B^b P_t^{(T)} \\ &= \omega P_t^{(T+b)} + \omega \delta P_t^{(T+b+1)} + \omega \delta^2 P_t^{(T+b+2)} + \dots \end{aligned}$$

hvor $P_t^{(T)}$ er definert i (4.4). Vi ser at i tidspunkt $T + b$ er effekten lik ω . Denne størrelsen blir redusert til $\omega \delta$ for neste tidspunkt $T + b + 1$, osv. Effekten forsvinner etter en kort eller lang periode, avhengig av om verdien av δ ligger nær 0 eller 1.

Fra (4.8),

$$\begin{aligned} \frac{\omega B^b}{(1 - \delta B)} S_t^{(T)} &= (1 + \delta B + \delta^2 B^2 + \dots) \omega B^b S_t^{(T)} \\ &= \omega S_t^{(T+b)} + \omega \delta S_t^{(T+b+1)} + \omega \delta^2 S_t^{(T+b+2)} + \dots \end{aligned}$$

hvor $S_t^{(T)}$ er definert i (4.5). Vi ser at effekten blir øker med verdien $\omega \delta$ for neste tidspunkt. Siden $0 < \delta < 1$, er effekten en konstant etter en periode. Når $\delta = 1$ forsvinner ikke effekten.

Det kan være mange typer for intervensjoner i en tidsserie. $f(t)$ for eksempel, på følgende form

$$f(t) = \frac{\omega_0 B}{1 - \delta B} P_t^{(T)} + \omega_1 B S_t^{(T)}$$

En generell modell for O_t blir

$$O_t = \sum_{j=1}^k \frac{\omega_j(B) B^{b_j}}{\delta_j(B)} I_{jt} + \frac{\theta(B)}{\phi(B)} \epsilon_t \quad (4.9)$$

hvor I_{jt} , $j = 1, \dots, k$ er intervensjonvariable.

4.2.1 Additive Outlier i t_0 (Et spesielt utslag for et bestemt tidspunkt t_0)

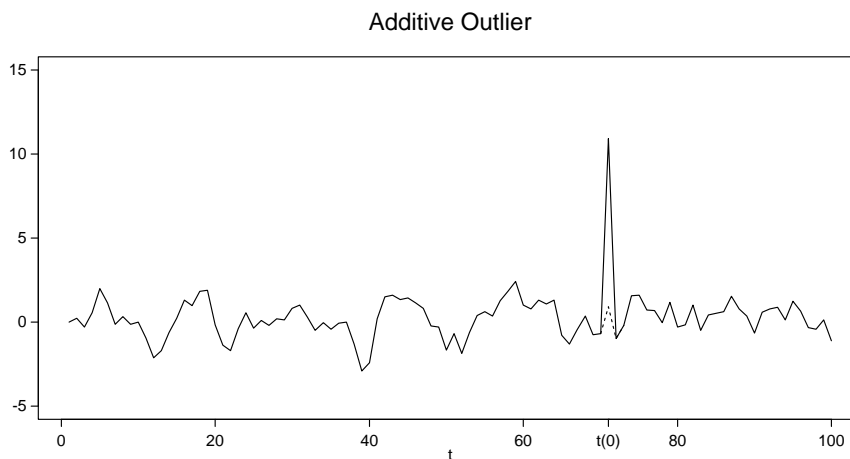
Dataene har en brå endring i t_0 . Effekten påvirker dataene bare i t_0 og forsvinner etter t_0 . $f(t)$ beskrives slik

$$f(t) = \omega_0 P_t^{(t_0)}, \quad P_t^{(t_0)} = \begin{cases} 1 & \text{for } t = t_0 \\ 0 & \text{for } t \neq t_0 \end{cases} \quad (4.10)$$

hvor ω_0 er en konstant. Vi antar at ω_0 er kjent på forhånd. Men denne størrelsen kan også estimeres fra dataene. Ligningen (4.2) blir

$$O_t = \omega_0 P_t^{(t_0)} + u_t$$

ω_0 er effekten i tidspunkt t_0 , ω_0 er vanligvis ukjent og estimeres fra dataene. De to tidsseriene $\{O_t\}$ og $\{u_t\}$ er helt identiske, bortsett fra i tidspunktet $t = t_0$.



Figur 4.2: En simulering for additive outlier

Figur 4.2 viser en simulering for additive outlier fra en AR(1) prosess, $u_t = 0.6u_{t-1} + \epsilon_t$, $t = 1, \dots, 100$, hvor ϵ_t er uavhengige og identisk normalfordelte $N(0, 1)$. $t_0=71$ og $\omega_0 = 10$ er valgt på forhånd. Modellen for O_t blir

$$O_t = 10P_t^{(t_0)} + \frac{1}{1 - 0.6B} \epsilon_t, \quad t = 1, \dots, 100, \text{ og } t_0 = 71$$

4.2.2 Level shift at t_0 (Nivåskift ved tidspunkt t_0)

Modellen for $f(t)$ er

$$f(t) = \omega_0 S_t^{(t_0)}, \quad S_t^{(t_0)} = \begin{cases} 1 & \text{for } t \geq t_0 \\ 0 & \text{for } t < t_0 \end{cases} \quad (4.11)$$

Vi får fra ligningen (4.2)

$$O_t = \omega_0 S_t^{(t_0)} + u_t$$

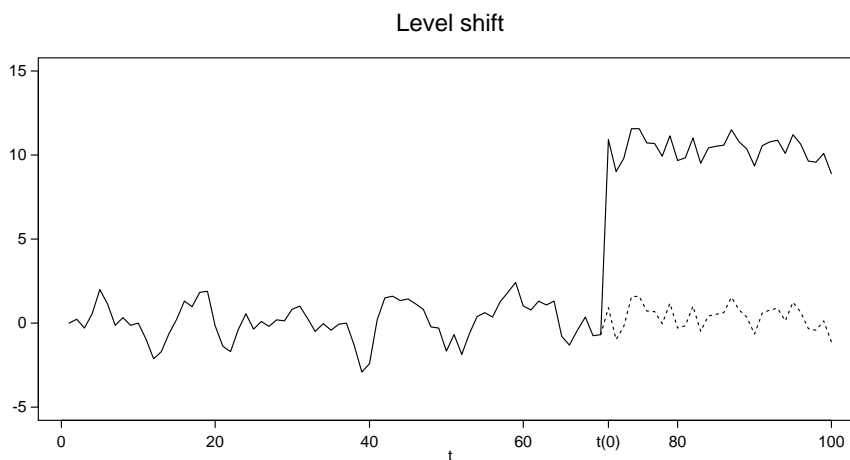
Dataene ser ut som om serien øker eller synker med en størrelse ω_0 fra t_0 og framover. Figur 4.3 viser en simulering fra modellen

$$O_t = 10S_t^{(t_0)} + \frac{1}{1-0.6B}\epsilon_t, \quad t = 1, \dots, 100, \text{ og } t_0 = 71$$

Ved å bruke (4.6) skrives denne ligningen slik

$$O_t = \frac{10}{1-B}P_t^{(t_0)} + \frac{1}{1-0.6B}\epsilon_t, \quad t = 1, \dots, 100, \text{ og } t_0 = 71$$

Etter $t_0 = 71$ er de to tidsseriene forskjøvet med en konstant $\omega_0 = 10$. $\{u_t\}$ plottes med prikker.



Figur 4.3: En simuleringer for level shift

4.2.3 Innovational Outlier (IO) (Brå endring med avtagende effekt)

Modellen for $f(t)$ er

$$f(t) = \frac{\omega_0}{1-\delta B}P_t^{(t_0)} \quad (4.12)$$

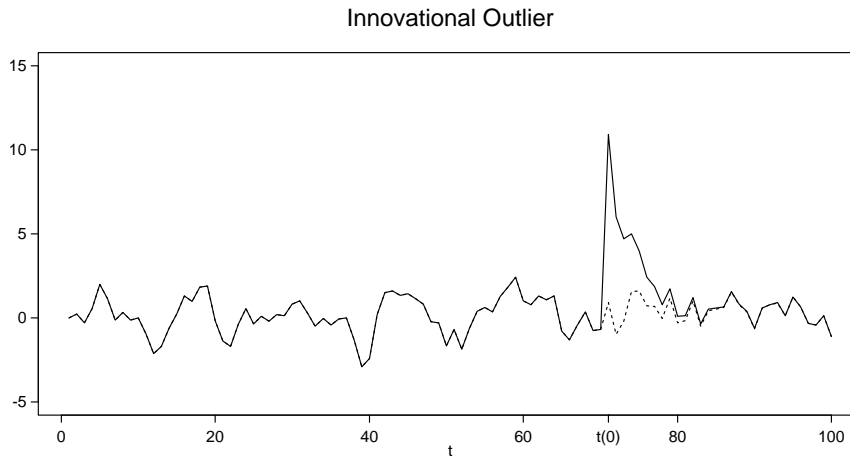
hvor $0 < \delta < 1$. Siden $1/(1-\delta B) = 1 + \delta B + \delta^2 B^2 + \delta^3 B^3 + \dots$, får vi

$$\begin{aligned} O_t &= (1 + \delta B + \delta^2 B^2 + \dots)\omega_0 P_t^{(t_0)} + u_t \\ &= \omega_0 P_t^{(t_0)} + \omega_0 \delta P_t^{(t_0+1)} + \omega_0 \delta^2 P_t^{(t_0+2)} + \dots + \omega_0 \delta^n P_t^{(t_0+n)} + \dots + u_t \end{aligned} \quad (4.13)$$

La Δ_t være differansen mellom O_t og u_t . Vi får

$$\Delta_t = \begin{cases} 0 & t < t_0 \\ \omega_0 & t = t_0 \\ \omega_0 \delta^n & t = t_0 + n, n \geq 0 \end{cases}$$

siden $0 < \delta < 1$, blir størrelsen $\omega_0 \delta^n$ mindre etter hvert når n vokser. For n er stor nok, er avstanden mellom O_t og u_t forsvunnet. De to tidsseriene blir identiske, fra dette tidspunktet.



Figur 4.4: En simulering for innovation outlier

Figur 4.4 viser en simulering for O_t , med modellen $u_t = 0.6u_{t-1} + \epsilon_t$, hvor ϵ_t er uavhengige identiske normalfordel $N(0,1)$. Vi velger $t_0 = 71$, $\omega_0 = 10$ og $\delta = 0.7$. Modellen for O_t blir

$$O_t = \frac{10}{1 - 0.7B} P_t^{(t_0)} + \frac{1}{1 - 0.6B} \epsilon_t$$

I praksis er ω_0 ofte ukjent og estimeres fra dataene. En tidsserie kan påviktes av mange forskjellige typer av intervensjoner. I X-12-ARIMA utfører man en prekorrigering for intervensjoner før sesongjustering. Det er 4 typer av intervensjoner som er beskrevet i X-12-ARIMA. De er

a. Additive outlier at t_0

$$f(t) = \omega_0 AO_t^{(t_0)}, \quad \text{hvor } AO_t^{(t_0)} = \begin{cases} 1 & \text{for } t = t_0, \\ 0 & \text{for } t \neq t_0 \end{cases} \quad (4.14)$$

b. Level shift at t_0

$$f(t) = \omega_0 LS_t^{(t_0)}, \quad \text{hvor } LS_t^{(t_0)} = \begin{cases} -1 & \text{for } t < t_0, \\ 0 & \text{for } t \geq t_0 \end{cases} \quad (4.15)$$

c. Temporary change at t_0 (TC)

$$f(t) = \omega_0 TC_t^{(t_0)}, \quad \text{hvor } TC_t^{(t_0)} = \begin{cases} 0 & \text{for } t < t_0, \\ \alpha^{t-t_0} & \text{for } t \geq t_0, \end{cases} \quad 0 < \alpha < 1 \quad (4.16)$$

d. Ramp, t_0 to t_1

$$f(t) = \omega_0 RP_t^{(t_0, t_1)}, \quad \text{hvor } RP_t^{(t_0, t_1)} = \begin{cases} -1 & \text{for } t \leq t_0, \\ \frac{t - t_0}{t_1 - t_0} - 1 & \text{for } t_0 < t < t_1, \\ 0 & \text{for } t \geq t_1 \end{cases} \quad (4.17)$$

Vi ser at dummyvariabelen $LS_t^{(t_0)}$ i X-12-ARIMA har motsatt fortegn sammenlignet med ligningen (4.11). Fra (4.16) får vi

$$\begin{aligned} f(t) &= \omega_0 TC_t^{(t_0)} = \omega_0 \alpha^{t-t_0} S_t^{(t_0)} = \frac{\omega_0 \alpha^{t-t_0}}{1-B} P_t^{(t_0)} \\ &= (1 + B + B^2 + B^3 + \dots) \omega_0 \alpha^{t-t_0} P_t^{(t_0)} \\ &= \omega_0 \alpha^{t-t_0} P_t^{(t_0)} + \omega_0 \alpha^{t-t_0} P_t^{(t_0+1)} + \omega_0 \alpha^{t-t_0} P_t^{(t_0+2)} + \omega_0 \alpha^{t-t_0} P_t^{(t_0+3)} + \dots \end{aligned} \quad (4.18)$$

Vi ser at (4.12) og (4.18) er på samme form.

Eksempel: Registrerte nye biler i Norge (fortsatt).

X-12-ARIMA er brukt i estimeringen. Typer, tidspunkter og effekter av intervensjoner estimeres selv av programmet fra de observasjonene. Dataene er transformert med logaritmen før vi utfører en estimering med hensyn til ukedagseffekter og påskeeffekt. Resultater er listet i tabell 4.1.

Tabell 4.1: *Intervensjoner av bilsalg*

Dato	Type	$\hat{\omega}_0$	\hat{std}	t -verdi
1986.Jul	LS	-0.3361	0.0771	-4.36
1988.Feb	LS	-0.3575	0.0770	-4.64
1995.Dec	AO	-0.5779	0.0690	-8.36
1997.Dec	AO	0.3311	0.0700	4.72

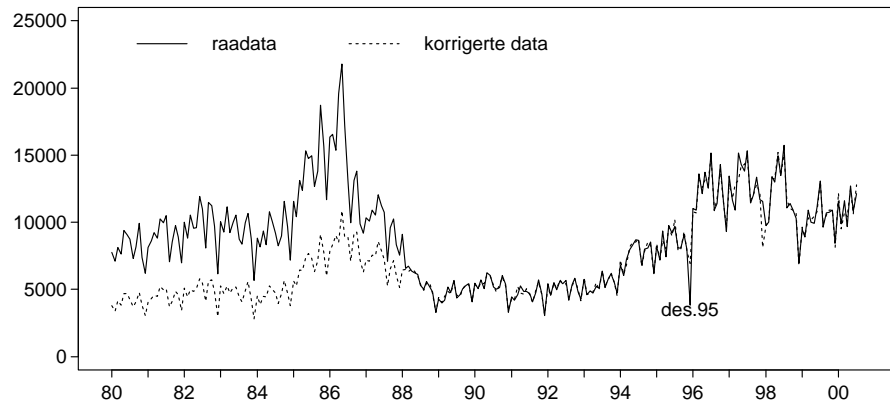
Forklaringene for to tidspunkter juli 1986 og desember 1995 er beskrevet ovenfor. De to andre tidspunktene vet vi ikke årsaken til. De er rett og slett "outliere" i data. En ARIMA(0,1,1)(0,1,1) er brukt i estimeringen. Rådataene og prekorrigerte dataene for ukedagseffekter, påskeeffekt og intervensjoner er plottet i figur 4.5. Vi betegner O_t for observasjonene. Modellen for O_t blir

$$\begin{aligned} \log O_t &= TD_t + E_t - 0.3361 LS_t^{jul.86} - 0.3575 LS_t^{feb.88} \\ &\quad - 0.5779 AO_t^{des.95} + 0.3311 AO_t^{des.97} + u_t \end{aligned}$$

hvor TD_t og E_t står for ukedagseffekter og påskeeffekt. De skal estimeres fra dataene. $\{u_t\}$ er tidsserien som skal spalte opp i trend, sesong- og irregulærkomponent.

4.3 Ukedagseffekter

Det er effekten av ulike antall ukedager som forekommer i en måned. Siden antall mandager, tirsdager, . . . , søndager endrer seg fra måned til måned, slik at en



Figur 4.5: Rådata og prekorrigerte data for nye personbiler

klesbutikk vil selge mer i en måned som har 4 søndager enn i en måned med 5 søndager, salget kan øke mer hvis måneden har 5 tirsdager, mens en kiosk i Vigelands parken vil selge mer i en måned som har 5 søndager. Vi ser at antall ukedager påvirker inntekten til butikken. Dette kan være viktig når vi bruker produksjonen til å vurdere aktiviteten en fabrikk, fordi antall virkedager ikke er det samme for to måneder. På grunn av dette er korrigering av effekten av variasjon i antall ukedager i en gitt måned en viktig del av sesongjusteringen. Tabell 4.2 viser rådataene og antall ukedager for hver måned for detaljvolumindeksen i 1999 og 2000. Merk at ukedager gjentar seg etter 28 år (før 2100 og den 29. februar er utelatt). For eksempel, kalender i år 2000 og 2028 har det samme antall mandager, tirsdager, ..., søndager for hver måned. Bell og Hillmer (1983) har utviklet en enkel metode for å estimere ukedagseffekter av en "flow" tidsserie, dvs tidsserien hvor en observert verdi i en måned er summen av daglige verdiene.

Vi antar at vekten ξ_1 for mandager er det samme i hele serien. Dette gjelder også for de andre ukedagene. Aktiviteten av ukedag i beskrives av ξ_i . Vi betegner $i = 1$ for mandag, $i = 2$ for tirsdag, ... og $i = 7$ for søndag. La $d_{i,t}$ være antall ukedager i , i måned t . Lengden av måned t er

$$TD_t = \sum_{i=1}^7 \xi_i d_{i,t} \quad (4.19)$$

En modell for O_t blir

$$O_t = TD_t + N_t \quad (4.20)$$

Vi antar at N_t følger en ARIMA modell,

$$\phi(B)\delta(B)N_t = \theta(B)\epsilon_t \quad (4.21)$$

hvor ϵ_t er en hvit støy prosess. $\phi(B)$ og $\theta(B)$ er to polynommer av B , som har ikke samme røtter og deres røttene ligger utenfor enhetssirkelen. $\delta(B)$ er en polynom av B for differensiering og skrives som $\delta(B) = (1 - B)^d(1 - B^{12})^D$, slik at $\delta(B)N_t$ er en stasjonær prosess.

Tabell 4.2: *Originale dataene for detaljvolumindeks*

År	måned	O_t	$d_{1,t}$	$d_{2,t}$	$d_{3,t}$	$d_{4,t}$	$d_{5,t}$	$d_{6,t}$	$d_{7,t}$	Σ
1999	januar	102.41	4	4	4	4	5	5	5	31
	februar	95.99	4	4	4	4	4	4	4	28
	mars	111.99	5	5	5	4	4	4	4	31
	april	97.25	4	4	4	5	5	4	4	30
	mai	107.41	5	4	4	4	4	5	5	31
	juni	117.72	4	5	5	4	4	4	4	30
	juli	118.63	4	4	4	5	5	5	4	31
	august	121.18	5	5	4	4	4	4	5	31
	september	113.65	4	4	5	5	4	4	4	30
	oktober	116.84	4	4	4	4	5	5	5	31
	november	116.87	5	5	4	4	4	4	4	30
	desember	161.88	4	4	5	5	5	4	4	31
2000	januar	104.81	5	4	4	4	4	5	5	31
	februar	102.06	4	5	4	4	4	4	4	29
	mars	110.28	4	4	5	5	5	4	4	31
	april	105.86	4	4	4	4	4	5	5	30
	mai	119.94	5	5	5	4	4	4	4	31
	juni	119.52	4	4	4	5	5	4	4	30
	juli	116.16	5	4	4	4	4	5	5	31

$d_{1,t}$: antall mandager i måned t

$d_{2,t}$: antall tirsdager i måned t

...

$d_{7,t}$: antall søndager i måned t

La

$$\bar{\xi} = \frac{1}{7} \sum_{i=1}^7 \xi_{i,t}, \quad T_{i,t} = d_{i,t} - d_{7,t}, \quad \text{og } T_{7,t} = \sum_{i=1}^7 d_{i,t} \quad (4.22)$$

Vi kan skrive (4.19) slik

$$\begin{aligned} TD_t &= \sum_{i=1}^7 (\xi_i - \bar{\xi})(d_{i,t} - d_{7,t}) + \sum_{i=1}^7 (\xi_i - \bar{\xi})d_{7,t} + \bar{\xi} \sum_{i=1}^7 d_{i,t} \\ &= \sum_{i=1}^7 \beta_i T_{i,t} \end{aligned} \quad (4.23)$$

hvor $\beta_i = \xi_i - \bar{\xi}$, (avviket mellom ξ_i og deres gjennomsnittet), $i = 1, \dots, 6$ og $\beta_7 = \bar{\xi}$. Ved å sette TD_t fra (4.23) og modellen til N_t inn (4.20), får vi

$$O_t = \sum_{i=1}^7 \beta_i T_{i,t} + \frac{\theta(B)}{\phi(B)\delta(B)} \epsilon_t \quad (4.24)$$

Fordelen ved å estimere effekter med (4.23) isteden for (4.19) er at det er mindre korrelasjon mellom β_i -ene enn ξ_i -ene i estimeringen. Vekten for søndag, ξ_7 ,

beregnes ved

$$\begin{aligned}\xi_7 - \bar{\xi} &= \sum_{i=1}^7 \xi_i - \bar{\xi} - \sum_{i=1}^6 \xi_i \\ &= 6\bar{\xi} - \sum_{i=1}^6 (\beta_i + \bar{\xi}) = - \sum_{i=1}^6 \beta_i\end{aligned}$$

Dermed er $\xi_7 = \beta_7 - \sum_{i=1}^6 \beta_i$.

Fra $d_{i,t}$ i tabell 4.2 kan vi beregne $T_{i,t}$ ved (4.22). Tallene er listet ut i tabell 4.3. Vi ser at $T_{i,t}$, $i = 1, \dots, 6$ kan være -1, 0 eller 1 og $T_{7,t}$ er lengden av måned t .

Tabell 4.3: $T_{i,t}$ for detaljvolumindeks

År	måned	O_t	$T_{1,t}$	$T_{2,t}$	$T_{3,t}$	$T_{4,t}$	$T_{5,t}$	$T_{6,t}$	$T_{7,t}$
1999	januar	102.41	-1	-1	-1	-1	0	0	31
	februar	95.99	0	0	0	0	0	0	28
	mars	111.99	1	1	1	0	0	0	31
	april	97.25	0	0	0	1	1	0	30
	mai	107.41	0	-1	-1	-1	-1	0	31
	juni	117.72	0	1	1	0	0	0	30
	juli	118.63	0	0	0	1	1	1	31
	august	121.18	0	0	-1	-1	-1	-1	31
	september	113.65	0	0	1	1	0	0	30
	oktober	116.84	-1	-1	-1	-1	0	0	31
	november	116.87	1	1	0	0	0	0	30
	desember	161.88	0	0	1	1	1	0	31
2000	januar	104.81	0	-1	-1	-1	-1	0	31
	februar	102.06	0	1	0	0	0	0	29
	mars	110.28	0	0	1	1	1	0	31
	april	105.86	-1	-1	-1	-1	-1	0	30
	mai	119.94	1	1	1	0	0	0	31
	juni	119.52	0	0	0	1	1	0	30
	juli	116.16	0	-1	-1	-1	-1	0	31

Ved å anta at mandag, tirsdag, \dots , fredag har samme vekt og vekten for hel-
ligdager som lørdag og søndag er like, da blir $T_{i,t}$ i (4.22) slik

$$T_{1,t} = T_{2,t} = \dots = T_{5,t} = (d_{1,t} + d_{2,t} + d_{3,t} + d_{4,t} + d_{5,t}) - \frac{5}{2}(d_{6,t} + d_{7,t}) \quad (4.25)$$

Eksempel: Detaljhandel (SNN52.IVL.U, unntatt med motorkjøretøyer og motor sykler reparasjon av husholdningsvarer og varer til personlig bruk. Volumindeks). Dataene er observert fra januar 1979 til juli 2000 og plottet i figur 3.14. Vi estimerer β_i i (4.24) med X-12-ARIMA. Vi får

1. $\hat{\beta}_i$ er listet ut i tabell 4.4 Vi ser at for onsdag og torsdag er det høy omsetning (positiv β -verdi) i detaljhandelen, mens for søndag er den svært lav (negativ β -verdi).

Tabell 4.4: *Estimerte verdiene for β_i*

	$\hat{\beta}_i$	<i>std</i>	<i>t-verdi</i>
mandag	-0.7449	0.36408	-2.05
tirsdag	0.3158	0.36119	0.87
onsdag	0.7091	0.36311	1.95
torsdag	0.8405	0.36288	2.32
fredag	-0.0639	0.36375	-0.18
lørdag	0.2447	0.36513	0.67
søndag	-1.3013	0.36890	-3.53

- Effekten er signifikant på 5% nivå, med $\chi^2=80$ og 6 frihetsgrader.
- ARIMA(0,1,2)(0,1,1) gir god tilpasning med dataene. Vi får

$$\begin{aligned}\hat{\theta}_1 &= 0.6948, & \text{std}(\hat{\theta}_1) &= 0.0631, \\ \hat{\theta}_2 &= 0.0620, & \text{std}(\hat{\theta}_2) &= 0.0631, \\ \hat{\Theta} &= 0.5199, & \text{std}(\hat{\Theta}) &= 0.0550.\end{aligned}$$

4.4 Påskeeffekt

Påskan påvirker ofte økonomiske data. Siden påskedagene faller i mars eller i april og de ikke faller på samme tidspunkt for hvert år (se tabell 4.5) vil dette påvirke månedstallene for mars og april eller kvartalstallene for første og andre kvartal. Påske påvirker produksjon (på grunn av fri), salg (fære handledager), det kan forskyve salget til perioden før og/eller etter påske. Dersom påske lå fast samme dato, ville påskeeffekten kun være en sesongkomponent. Merk at 1. påskedag varierer fra 22 mars til 25 april.

Bell og Hillmer (1983) beskriver påskeeffekt E_t , i tidspunkt t , ved

$$E_t = \sum_{i=1}^K \bar{\alpha}_i h(i, t) \quad (4.26)$$

K er antall dager før 1. påske hvor dataene påvirkes av påske. $\bar{\alpha}_i$ er effekten for i -te dag før 1. påskedag og

$$h(i, t) = \begin{cases} 1 & \text{hvis } i\text{-te dag før 1. påskedag faller i måned } t \\ & (t \text{ er enten mars eller april}) \\ 0 & \text{ellers} \end{cases} \quad (4.27)$$

Vi ser i ligning (4.26) at effekten for måned t , er summen av daglige effekter. For enkelhets skyld, antar vi at $\bar{\alpha}_1 = \bar{\alpha}_2 = \dots = \bar{\alpha}_\tau = \bar{\alpha}$ og $\bar{\alpha}_{\tau+1} = \dots = \bar{\alpha}_K = 0$. Ligning (4.26) blir

$$E_t = \alpha H(\tau, t) \quad (4.28)$$

hvor $\alpha = \tau \bar{\alpha}$ og $H(\tau, t) = 1/\tau \sum_{i=1}^{\tau} h(i, t)$.

Tabell 4.6 viser verdiene av $H(\tau, t)$ for $\tau = 1, \dots, 10$, og t =mars og april 1999. Merk at 1. påskedag i dette året er 4. april.

Tabell 4.5: 1. påskedag for 1970-2010

år	dato	år	dato
1970	29-mars	1991	31-mars
1971	11-april	1992	19-april
1972	2-april	1993	11-april
1973	22-april	1994	3-april
1974	14-april	1995	16-april
1975	30-mars	1996	7-april
1976	18-april	1997	30-mars
1977	10-april	1998	12-april
1978	26-mars	1999	4-april
1979	15-april	2000	23-april
1980	6-april	2001	15-april
1981	19-april	2002	31-mars
1982	11-april	2003	20-april
1983	3-april	2004	11-april
1984	22-april	2005	27-mars
1985	7-april	2006	16-april
1986	30-mars	2007	8-april
1987	19-april	2008	23-mars
1988	3-april	2009	12-april
1989	26-mars	2010	4-april
1990	15-april		

Tabell 4.6: $H(\tau, t)$ i 1999

τ	0	1	2	3	4	5	6	7	8	9	10
$H(\tau, mars)$	0	0	0	0	1/4	2/5	3/6	4/7	5/8	6/9	7/10
$H(\tau, april)$	1	1	1	1	3/4	3/5	3/6	3/7	3/8	3/9	3/10

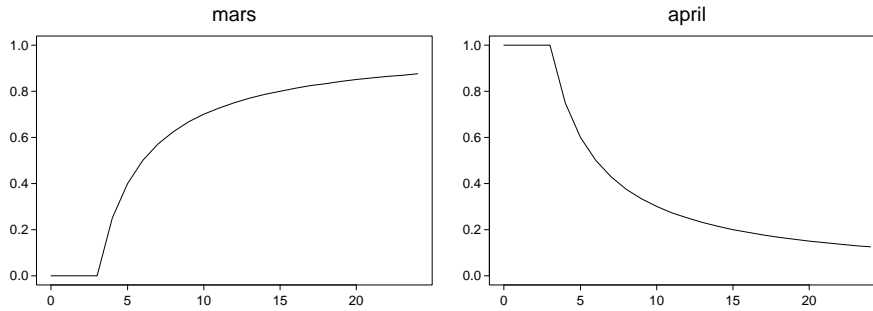
For eksempel, for $\tau = 5$ er 3. april, 2. april, 1. april, 31 mars og 30 mars, de fem dagene før 1. påske (4. april). Blant dem er det to dager som faller i mars og tre dager i april. Dermed er

$$\begin{aligned}
 H(5, mars) &= 2/5, \text{ som er andel av dagene før 1. påske som faller i mars,} \\
 H(5, april) &= 1 - H(5, mars) = 3/5, \text{ som er andel av dagene før 1. påske} \\
 &\text{som faller i april}
 \end{aligned}$$

Figur 4.6 viser $H(\tau, t)$ for mars og april 1999. En modell for en tidsserie med ukedagsvariasjoner og påskevariasjon, er

$$\begin{aligned}
 O_t &= TD_t + E_t + N_t \\
 &= \sum_{i=1}^7 \beta_i T_{it} + \alpha H(\tau, t) + \frac{\theta(B)}{\phi(B)\delta(B)} \epsilon_t
 \end{aligned} \tag{4.29}$$

se (4.23), (4.28) og (4.21) for beskrivelsene av TD_t , E_t og N_t .

Figur 4.6: $H(\tau, t)$ for mars og april 1999

Vi har utviklet en alternativ metode for korrogering av påskeeffekten i SSB (notat 97/73). Metoden fungerer bedre for detaljhandel enn den som brukes i X-12-ARIMA.

4.5 Trend

Det er komponenten som beskriver en langtidsutvikling av tidsserien. Vi betegner trenden i tidspunkt t med T_t . Det å bestemme langtidsutviklingen i en tidsserie kan også omfattes som å glatte tidsserier. Det finnes en rekke metoder for glatting av tidsserier, en av disse er såkalte glidende gjennomsnitt.

La i dette avsnittet O_1, \dots, O_N være en tidsserie med N observasjoner. Vi betegner tidsserien med $\{O_t\}$, $t = 1, \dots, N$.

4.5.1 Trend og glidende gjennomsnitt teknikk

Vi illustrerer denne teknikken ved et enkelt eksempel. Dataene er detaljvolumindeksen fra januar 1979 til juli 2000. Tabell 4.7 viser tallene for 1999 og 2000.

Tabell 4.7: *Detaljvolumindeks for 1999 og 2000*

år	jan.	feb.	mar.	apr.	mai.	jun.
	jul.	aug.	sep.	okt.	nov.	des.
1999	102.41	95.59	111.99	97.25	107.41	117.72
	118.63	121.18	113.65	116.84	116.87	161.88
2000	104.81	102.06	110.28	105.86	119.94	119.52
	116.16					

Trenden som beregnes ved 3 måneders glidende gjennomsnitt, er gitt ved

$$\hat{T}_t = \frac{1}{3}O_{t-1} + \frac{1}{3}O_t + \frac{1}{3}O_{t+1} \quad (4.30)$$

Vi får fra tabell 4.7

$$\begin{aligned}\widehat{T}_{feb.99} &= \frac{1}{3}O_{jan.99} + \frac{1}{3}O_{feb.99} + \frac{1}{3}O_{mar.99} \\ &= (102.41 + 95.59 + 111.99)/3 = 103.33 \\ \widehat{T}_{mar.99} &= \frac{1}{3}O_{feb.99} + \frac{1}{3}O_{mar.99} + \frac{1}{3}O_{apr.99} \\ &= (95.59 + 111.99 + 97.25)/3 = 101.61 \\ &\dots \\ \widehat{T}_{jun.00} &= \frac{1}{3}O_{mai.00} + \frac{1}{3}O_{jun.00} + \frac{1}{3}O_{jul.00} \\ &= (119.94 + 119.52 + 116.16)/3 = 118.54\end{aligned}$$

Vi får ikke verdi for trendestimat \widehat{T}_t i juli 2000, siden $O_{aug.00}$ ikke er observert.

En generell form for trendestimat, er

$$\widehat{T}_t = \sum_{j=-n}^n r_j O_{t+j} \quad (4.31)$$

hvor r_{-n}, \dots, r_n er gitt konstanter. Vanligvis, er $\sum_{j=-n}^n r_j = 1$. I ligning (4.30) er $r_j = 1/3$. \widehat{T}_t tolkes som et veiet gjennomsnitt av de $2n+1$ verdiene O_{t-n}, \dots, O_{t+n} . I praksis velger vi n slik at $11 \leq 2n+1 \leq 21$.

4.5.2 12 måneders sentrert glidende gjennomsnitt

Trenden skrives på formen

$$\widehat{T}_t^{(12)} = \frac{1}{24}O_{t-6} + \frac{1}{12}O_{t-5} + \dots + \frac{1}{12}O_t + \dots + \frac{1}{12}O_{t+5} + \frac{1}{24}O_{t+6} \quad (4.32)$$

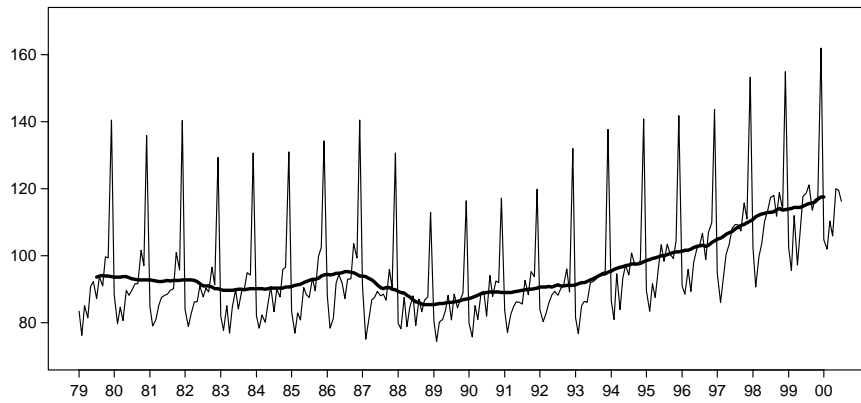
hvor $7 \leq t \leq N-6$, N er antall observasjoner. Vi ser at $\widehat{T}_t^{(12)}$ er summen av 13 observasjoner som ligger etter hverandre på begge sider av t . Vekten av den første og siste er $1/24$ ($r_{-6} = r_6 = 1/24$), mens for de observasjonene som ligger i midten er vekten lik $1/12$ ($r_{-5} = \dots = r_0 = \dots = r_5 = 1/12$) og summen av disse vektene er lik 1 ($\sum_{j=-6}^6 r_j = 1$). Trendestimatet i tidspunkt $t+1$ blir

$$\widehat{T}_{t+1}^{(12)} = \frac{1}{24}O_{t-5} + \frac{1}{12}O_{t-4} + \dots + \frac{1}{12}O_{t+1} + \dots + \frac{1}{12}O_{t+6} + \frac{1}{24}O_{t+7} \quad (4.33)$$

Vektene $(\frac{1}{24}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{24})$ blir flyttet fram en observasjon.

Figur 4.7 viser trenden som beregnes ved det sentrert 12 måneders glidende gjennomsnittet for detaljvolumindeksen. Vi får ikke verdiene for trenden etter januar 2000, siden dataene er observert til juli 2000.

Det sentrert 12 måneders glidende gjennomsnittet brukes for å beregne trend i første runde i X-12-ARIMA.



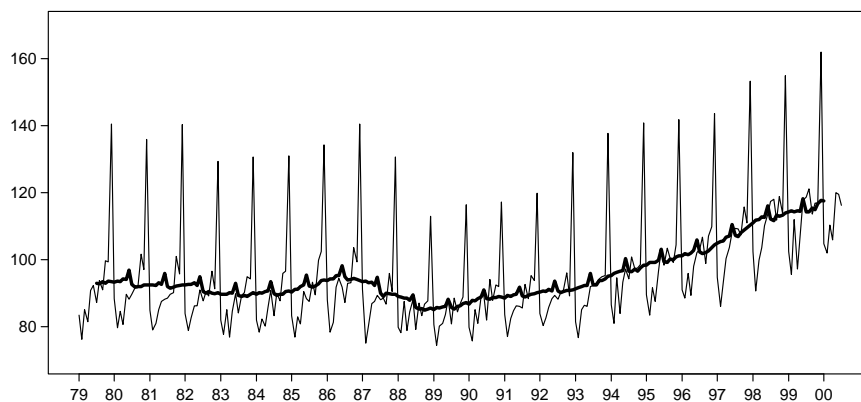
Figur 4.7: *Detaljvolumindeks, sentrert 12 mnd glidende gjennomsnitt*

4.5.3 13 måneders glidende gjennomsnitt

Trenden beregnes på formen

$$\begin{aligned}\hat{T}_t^{(13)} &= \frac{1}{13}O_{t-6} + \frac{1}{13}O_{t-5} + \cdots + \frac{1}{13}O_t + \cdots + \frac{1}{13}O_{t+5} + \frac{1}{13}O_{t+6} \\ &= \frac{1}{13} \sum_{j=-6}^{j=6} O_{t+j}\end{aligned}\quad (4.34)$$

hvor $7 \leq t \leq N - 6$. Vi har like vekter for alle observasjonene.



Figur 4.8: *Detaljvolumindeks, 13 mnd glidende gjennomsnitt*

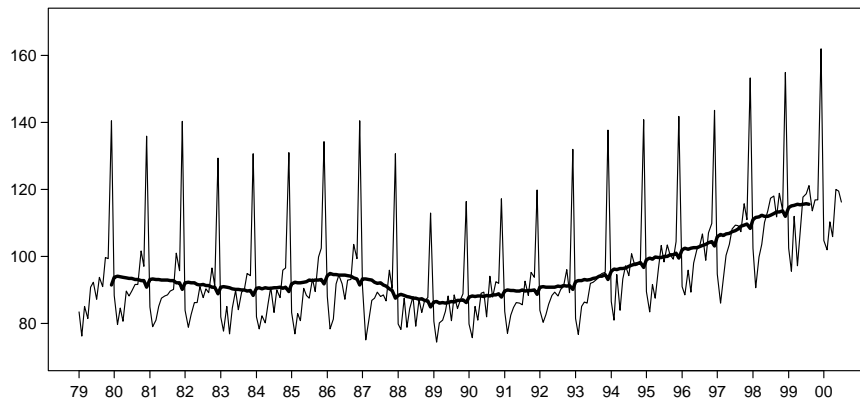
Figur 4.8 viser trend som beregnes med 13 måneders glidende gjennomsnitt for detaljvolumindeksen. Merk at vi ikke har verdiene for trenden i de 6 første og siste tidspunktene av tidsserien.

4.5.4 23 måneders glidende gjennomsnitt

Trendestimatet beregnes på formen

$$\begin{aligned}\widehat{T}_t^{(23)} &= \frac{1}{23}O_{t-11} + \frac{1}{23}O_{t-10} + \cdots + \frac{1}{23}O_t + \cdots + \frac{1}{23}O_{t+10} + \frac{1}{23}O_{t+11} \\ &= \frac{1}{23} \sum_{j=-11}^{j=11} O_{t+j}\end{aligned}\quad (4.35)$$

hvor $12 \leq t \leq N - 11$. Alle tidspunktene beregnes med samme vekt. Figur 4.9 viser trenden som beregnes med 23 måneders glidende gjennomsnitt. Vi får ikke verdiene for $\widehat{T}_t^{(23)}$ før desember 1979 og etter august 1999.



Figur 4.9: *Detaljvolumindeks, 23 mnd glidende gjennomsnitt*

4.5.5 En oppsummering for det sentrert 12 mnd-gj., 13 mnd-gj. og 23 mnd-gj.

1. $\widehat{T}_t^{(23)}$ er estimert med lavere vekter og med lengre periodefilter enn $\widehat{T}_t^{(12)}$ og $\widehat{T}_t^{(13)}$.
2. $\widehat{T}_t^{(12)}$, $\widehat{T}_t^{(13)}$ og $\widehat{T}_t^{(23)}$ er glattere enn den opprinnelige tidsserien og endrer seg senere over tid.
3. Sesongvariasjoner blir fjernet fra $\widehat{T}_t^{(12)}$, men de finnes fortsatt i $\widehat{T}_t^{(13)}$ og $\widehat{T}_t^{(23)}$.
4. Trendestimatet gir ofte ikke et riktig bilde av bevegelsen i tidsserien. Høye punkter i trenden presenterer ikke høye punkter i rådataene og omvendt.

4.5.6 13 måneders Hendersons glidende gjennomsnitt

Ved å sette de to tabellene 4.8 og 4.9 ved siden av, får vi en komplett tabell for vektene.

Forklaringene for disse tabellene beskrives følgende

Tabell 4.8: *Hendersons vektorer for 13 måneders glidende gjennomsnitt*

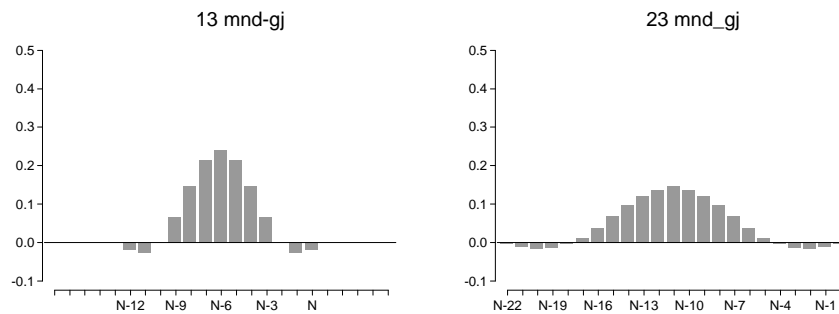
t	$N-12$	$N-11$	$N-10$	$N-9$	$N-8$	$N-7$	$N-6$
N	0	0	0	0	0	0	-0.092
$N-1$	0	0	0	0	0	-0.043	-0.038
$N-2$	0	0	0	0	-0.016	-0.025	0.003
$N-3$	0	0	0	-0.009	-0.022	0.004	0.066
$N-4$	0	0	-0.011	-0.022	0.003	0.067	0.145
$N-5$	0	-0.017	-0.025	0.001	0.066	0.147	0.213
$N-6$	-0.019	-0.028	0	0.066	0.147	0.214	0.240

Tabell 4.9: *Hendersons vektorer for 13 måneders glidende gjennomsnitt*

t	$N-5$	$N-4$	$N-3$	$N-2$	$N-1$	N
N	-0.058	0.012	0.120	0.244	0.353	0.421
$N-1$	0.002	0.080	0.174	0.254	0.292	0.279
$N-2$	0.068	0.149	0.216	0.241	0.216	0.148
$N-3$	0.145	0.208	0.230	0.201	0.131	0.046
$N-4$	0.210	0.235	0.205	0.136	0.050	-0.018
$N-5$	0.238	0.212	0.144	0.061	-0.006	-0.034
$N-6$	0.214	0.147	0.066	0	-0.028	-0.019

1. For $7 \leq t \leq N-6$ er symmetriske vektorer brukt for å beregne \hat{T}_t^{13H} . Vektene beskrives i den siste linjen av tabellene og plottes i figur 4.10. For eksempel,

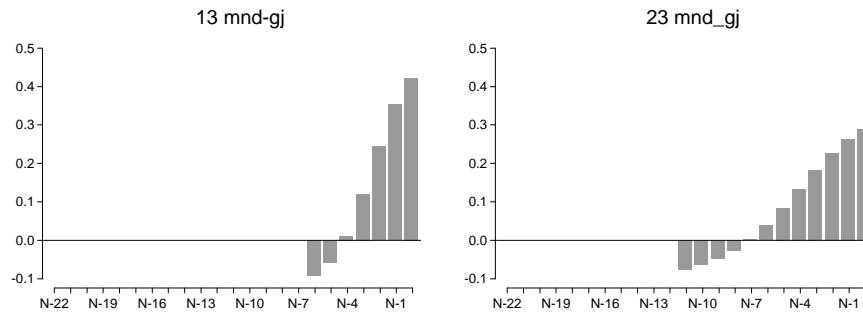
$$\begin{aligned}
 \hat{T}_{N-6}^{13H} = & -0.019 O_{N-12} - 0.028 O_{N-11} + 0 O_{N-10} + 0.066 O_{N-9} \\
 & + 0.147 O_{N-8} + 0.214 O_{N-7} + 0.240 O_{N-6} \\
 & + 0.214 O_{N-5} + 0.147 O_{N-4} + 0.066 O_{N-3} \\
 & + 0 O_{N-2} - 0.028 O_{N-1} - 0.019 O_N
 \end{aligned} \tag{4.36}$$

Figur 4.10: *Symmetriske vektorer*

2. For $N-5 \leq t \leq N$ er asymmetriske vektorer brukt. Figur 4.11 viser vektene for $t = N$. De 6 første linjene av tabellene 4.8 og 4.9 er vektene for $t =$

$N, \dots, N - 5$. Vi får

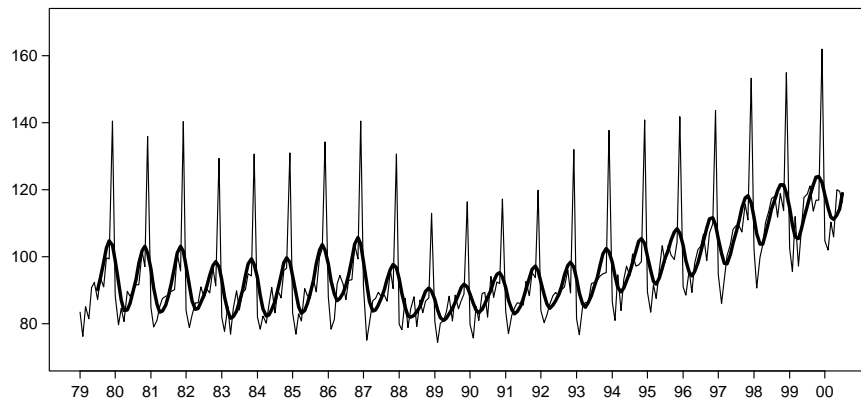
$$\begin{aligned} \hat{T}_N^{13H} &= 0 O_{N-12} + 0 O_{N-11} + 0 O_{N-10} + 0 O_{N-9} \\ &\quad + 0 O_{N-8} + 0 O_{N-7} - 0.092 O_{N-6} \\ &\quad - 0.058 O_{N-5} + 0.012 O_{N-4} + 0.120 O_{N-3} \\ &\quad + 0.244 O_{N-2} + 0.353 O_{N-1} + 0.421 O_N \end{aligned} \quad (4.37)$$



Figur 4.11: *Asymmetriske vektorer for $t = N$*

Vi ser at for det 13 måneders Hendersons glidende gjennomsnittet er symmetriske vektorer brukt i midten og asymmetriske vektorer brukt for observasjoner i slutten av serien. Dermed er trenden beregnet til det siste tidspunktet.

Figur 4.12 viser trenden som beregnes med 13 måneders Hendersons glidende gjennomsnittet for detaljvolumindeksen.



Figur 4.12: *Detaljvolumindeks, 13 måneders Hendersons glidende gjennomsnitt*

4.5.7 23 måneders Hendersons glidende gjennomsnitt

Vektene for det 23 måneders Hendersons glidende gjennomsnittet er gitt i X-11-ARIMA manualen (side 63). Det er en lang tabell med symmetriske vektorer som brukes i midten ($12 \leq t \leq N - 11$), og asymmetriske vektorer for de observasjonene

i slutten ($N - 10 \leq t \leq N$) av serien. For eksempel, trendestimatet i $t = N - 11$ beregnes ved

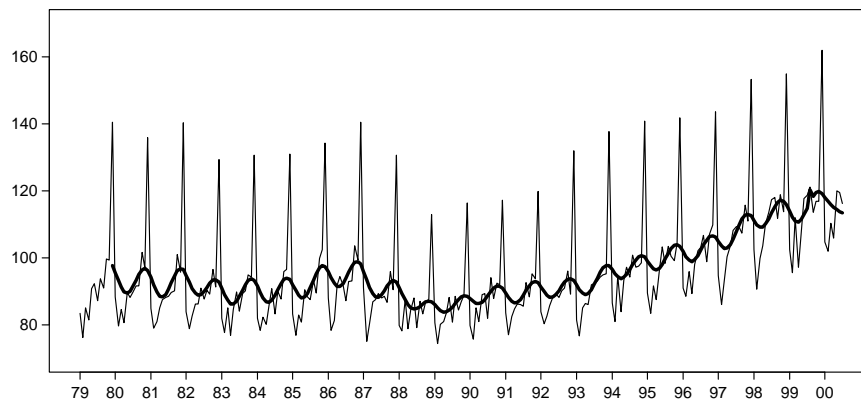
$$\begin{aligned} \hat{T}_{N-11}^{23H} = & -0.004 O_{N-22} - 0.011 O_{N-21} - 0.016 O_{N-20} - 0.015 O_{N-19} \\ & -0.005 O_{N-18} + 0.013 O_{N-17} + 0.039 O_{N-16} + 0.068 O_{N-15} \\ & + 0.097 O_{N-14} + 0.122 O_{N-13} + 0.138 O_{N-12} + 0.148 O_{N-11} \\ & + 0.138 O_{N-10} + 0.122 O_{N-9} + 0.097 O_{N-8} + 0.068 O_{N-7} \\ & + 0.039 O_{N-6} + 0.013 O_{N-5} - 0.005 O_{N-4} - 0.015 O_{N-3} \\ & - 0.016 O_{N-2} - 0.011 O_{N-1} - 0.004 O_N \end{aligned} \quad (4.38)$$

og for $t = N$ er

$$\begin{aligned} \hat{T}_t^{23H} = & 0.288 O_N + 0.263 O_{N-1} + 0.227 O_{N-2} + 0.182 O_{N-3} + \\ & 0.133 O_{N-4} + 0.084 O_{N-5} + 0.039 O_{N-6} + 0.002 O_{N-7} - \\ & 0.028 O_{N-8} - 0.049 O_{N-9} - 0.064 O_{N-10} - 0.077 O_{N-11} \end{aligned} \quad (4.39)$$

Symmetriske og asymmetriske ($t = N$) vektene plottes i figur 4.10 og 4.11.

Trendestimatet \hat{T}_t^{23H} for detaljvolumindeksen er plottet i figur 4.13.



Figur 4.13: *Detaljvolumindeks, 23 måneders Hendersons glidende gjennomsnitt*

4.5.8 Kommentarer for 13 og 23 måneders Hendersons glidende gjennomsnitt

1. \hat{T}_t^{23H} har lavere vekt og lengre periode enn \hat{T}_t^{13H} .
2. De symmetriske vektene brukes i midten av serien og asymmetriske vektene brukes i endepunkter.
3. Trenden gir et godt bilde av bevegelse i serien. Filteret forandrer ikke tidspunktene for høye eller lave verdier.
4. Bruk Hendersons 23 måneders glidende gjennomsnitt for dataene som har store utslag.

4.5.9 Konklusjoner om trend

1. Jo lenger filter en benytter, desto glattere blir trendestimatet, ”overglating” kan medføre at en mister noe av bevegelsene i trendkomponenten.
2. Trenden endrer seg bare for de siste 11 eller 6 måneder når nye observasjoner tilføyes.
3. Hendersons filteret gir betydelige revisjoner i trendkomponenten i slutten av serien.
4. Trenden gir et godt bilde av variasjoner i serien.
5. Hendersons filteret fjerner ikke sesongvariasjoner i trendestimatet.
6. Vi kan tolke i feil retning fra de originale dataene på grunn av ekstreme observasjoner.

4.6 Sesongvariasjoner

Det er variasjoner som gjentar seg fra år til år i samme tidspunkter eller i samme periode som vi kan måle og fjerne fra de observerte dataene for å få en bedre trend. Det er ofte fire årsaker som medfører sesongvariasjoner i en tidsserie

- Kalendereffekter slik som påske, jul eller 17. mai. Disse helligdagene påvirker varehandel, produksjon, transport, import, eksport. For eksempel, folk handler mer varer i ukene før julen. Dette skjer hvert år og på samme tidspunktet.

- Vær endring er en annen grunn som påvirker landbruks produksjon, transport, konstruksjon. For eksempel, færre bygg blir bygd om vinteren enn andre årstider.

- I perioden mellom to semestre øker arbeidsledigheten på grunn av at studenter søker arbeid.

- Forventer en at salget av sukkeføy skal øke rett før jul vil medføre høy produksjon av sukkeføy i f.eks november.

Modell for sesongvariasjoner.

La $\{S_t\}$ være en månedlig prosess med sesongvariasjoner. Vi antar at S_t er en periodisk funksjon med perioden 12 måneder, dvs $S_t = S_{t-12}$. Modellene for S_t er

$$S_t = \sum_{j=1}^6 a_j \cos(\omega j t + \Theta_j), \quad \omega = 2\pi/6 \quad (4.40)$$

eller

$$S_t = \sum_{j=1}^{12} b_j I_j, \quad I_j = \begin{cases} 1, & \text{for måned } j \\ 0 & \text{for de andre månedene} \end{cases} \quad (4.41)$$

S_t i (4.40) og (4.41) er deterministiske funksjoner.

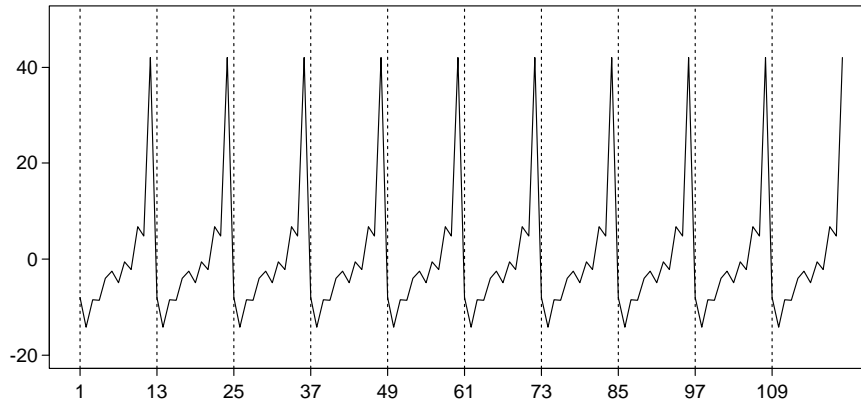
Dessuten kan S_t beskrives ved en hjelp av en ARIMA modell

$$\phi(B)S_t = \theta(B)\epsilon_t \quad (4.42)$$

hvor ϵ_t er en hvit støy prosess. $\phi(B)$ og $\theta(B)$ er polynommer av B . For eksempel,

$$S_t = \phi S_{t-12} + \epsilon_t - \theta \epsilon_{t-1}$$

hvor $\phi < 1$ og ϵ_t er hvit støy prosess.

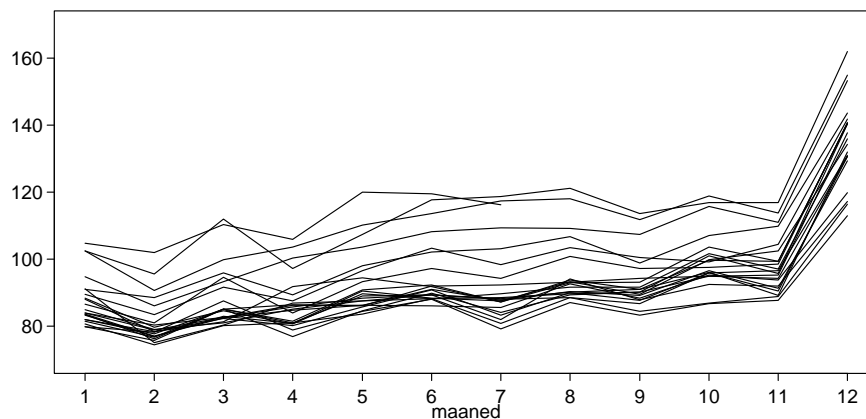


Figur 4.14: *Sesongvariasjoner ved modell (4.41)*

Figur 4.14 viser simulerte dataene av S_t ved modell (4.41), hvor $a_1 = -8.06$, $a_2 = -14.18$, $a_3 = -8.52$, $a_4 = -8.56$, $a_5 = -4.03$, $a_6 = -2.47$, $a_7 = -4.96$, $a_8 = -0.61$, $a_9 = -2.20$, $a_{10} = 6.77$, $a_{11} = 4.81$ og $a_{12} = 42.01$. Vi har $\sum_{j=1}^{12} a_j = 0$.

4.6.1 Enkle metoder for å sjekke om sesongvariasjoner i rådataene

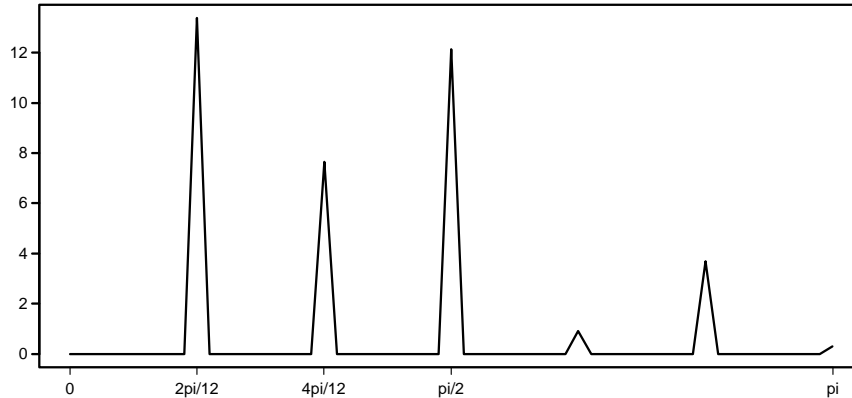
1. Å plote rådataene for hvert år over hverandre. Det skal vise endringer fra år til år i et bestemt tidspunkt. Figur 4.15 viser dataene av detaljvolumindeksen plottet på denne måten. Det viser et klart sesongmønster i tidsserien.



Figur 4.15: *Rådataene av detaljvolumindekse gjennom årene*

2. Å plote spektraltettheten mot perioden

- En prosess med tilfeldig variasjon har en flat spektraltetthet
- Spektraltettheten har tynne og høye frekvenser $j\omega_s$, $j = 1, 2, \dots$ og $\omega_s = 2\pi/12$, for en tidsserie med sesongvariasjoner.
- En tidsserie hvor trenden er dominerende, er en høy spektraltetthet nær null. Figur 4.16 viser spektraltettheten av simulerte dataene ved



Figur 4.16: *Spektraltettheten mot frekvenser*

(4.41).

3. Ved å plote de estimerte autokorrelasjonene \widehat{ACF} etter lag k . Da får vi også et godt bilde av sesongvariasjonene. Figur 4.17 viser \widehat{ACF} av detaljvolumindeksen. Vi ser at $\hat{\rho}_k$ er signifikant større enn null i lagene 12, 24, 36. Dette viser sesongvariasjoner i dataene.

4.7 Sesongjusterte tall

Det er en tidsserie av rådataene som ikke inneholder sesongvariasjoner. Den beskrives av trenden og tilfeldige variasjoner.

4.8 Irregulærkomponent

Det er komponenten som beskriver variasjoner som ikke gjentar seg fra år til år og ikke forklares. Det er restleddene av rådataene etter å ha fjernet trend og sesongvariasjoner.

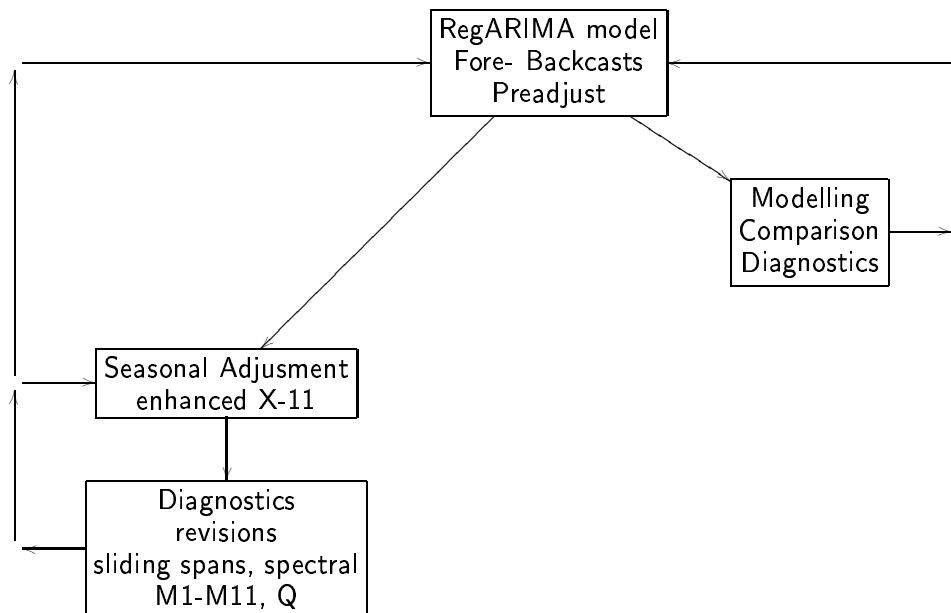
4.9 Default beregninger i X-12-ARIMA

Figur 4.18 viser diagrammet for programmene. Dataene blir først prekorrigert for effektene av ukedager, påske og intervensjoner. Deretter forlenges tidsserien i begge to sider for å kunne anvende synmmetriske vekter for hele serien i estimeringen for trend og sesongjusterte tall.

Autocorrelations

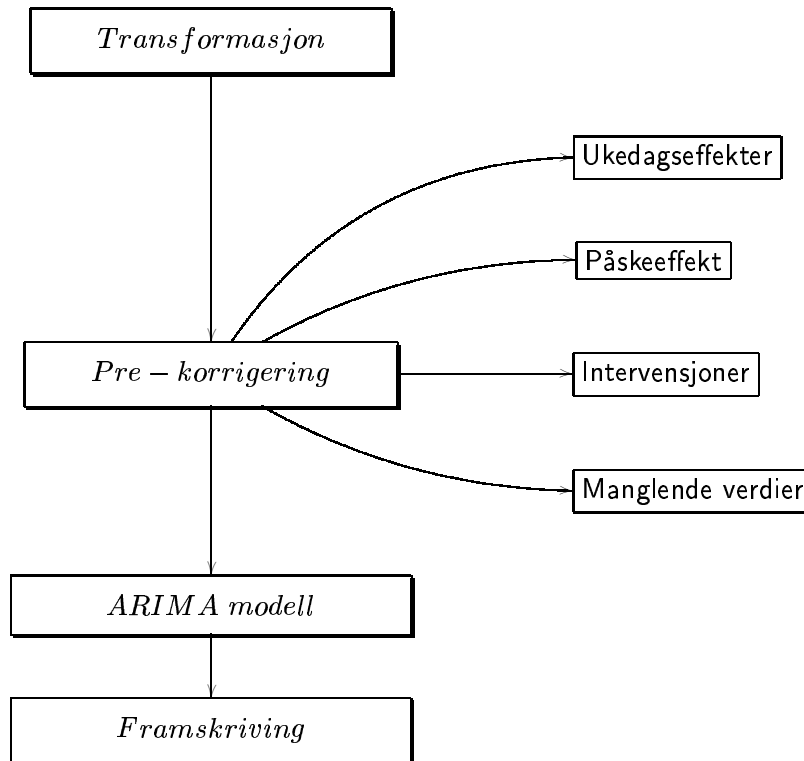
Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1
0	250.037	1.00000												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
1	84.010986	0.33599												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
2	56.423531	0.22566												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
3	46.094109	0.18435												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
4	39.865038	0.15944												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
5	48.048473	0.19217												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
6	60.989006	0.24392												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
7	46.163633	0.18463												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
8	35.281858	0.14111												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
9	37.344368	0.14936												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
10	42.999484	0.17197												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
11	70.603831	0.28237												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
12	218.912	0.87552												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
13	66.950202	0.26776												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
14	43.255673	0.17300												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
15	31.244509	0.12496												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
16	26.238493	0.10494												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
17	33.128568	0.13249												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
18	45.550529	0.18218												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
19	32.843952	0.13136												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
20	24.214847	0.09684												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
21	21.782509	0.08712												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
22	28.926098	0.11569												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
23	54.243849	0.21694												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
24	190.528	0.76200												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
25	50.144486	0.20055												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
26	26.042737	0.10416												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
27	14.712829	0.05884												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
28	11.049199	0.04419												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
29	16.320659	0.06527												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
30	29.714456	0.11884												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
31	19.197049	0.07678												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
32	10.092209	0.04036												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
33	7.325522	0.02930												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
34	14.337057	0.05734												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
35	35.592716	0.14235												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
36	162.918	0.65157												*****	*****	*****	*****	*****	*****	*****	*****	*****	*****

"." marks two standard errors

Figur 4.17: Estimerte autokorrelasjoner etter lag k 

Figur 4.18: Diagrammet for X-12-ARIMA

En forbedring for X-12-ARIMA er at en regARIMA modell (linear regression model with ARIMA time series errors) er brukt for å estimere effektene av uke-dager, påske og intervensjoner, samt framskrivningberegninger. Figur 4.19 viser diagrammet for regARIMA i X-12-ARIMA.



Figur 4.19: Diagrammet for regARIMA

La O_t skrives ved en regresjonsmodell

$$O_t = \sum_i \beta_i X_{i,t} + z_t \quad (4.43)$$

hvor β_i er regresjonsparametre, $X_{i,t}$ regresjonsvariable (ukedageffekter, påskeeffekt og intervensjoner). Fra (4.43) får vi

$$z_t = O_t - \sum_i \beta_i X_{i,t}$$

z_t -ene er ofte korrelerte for de fleste tidsserier. Ved å anta at z_t er ukorrelerte i beregninger kan medføre ugyldige resultater. La $z_t \sim \text{ARIMA}(p, d, q)(P, D, Q)$. Vi får

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D z_t = \theta(B)\Theta(B^s)\epsilon_t$$

O_t i (4.43) blir

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D(O_t - \sum_i \beta_i X_{i,t}) = \theta(B)\Theta(B^s)\epsilon_t \quad (4.44)$$

hvor $BO_t = O_{t-1}$, s angir antall observasjoner per år, ϵ_t er hvit støy prosess. $\phi(B)$, $\Phi(B^s)$, $\theta(B)$ og $\Theta(B^s)$ er polynommer av B . Ligningen (4.44) brukes i X-12-ARIMA for å estimere effektene av ukedager, påske og intervensjoner, samt å beregne framskrivninger.

X-11-ARIMA og X-12-ARIMA har brukt forskjellige rutiner for å prekorrigere effektene av ukedager og påske, mens metoden for å estimere trend T_t , sesongjusterte tall A_t , sesongkomponent S_t og irregulærkomponent I_t er det samme. For å beskrive metoden enklere, antar vi at

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

De tre modellene for sesongjustering er

$$\begin{array}{ll}
 \text{Multiplikativ modell (M)} & O_t = T_t \times S_t \times I_t \\
 \text{Additiv modell (A)} & O_t = T_t + S_t + I_t \\
 \text{Pseudo-Additiv modell (PA)} & O_t = T_t \times (S_t + I_t - 1)
 \end{array} \quad (4.45)$$

- Multiplikativ modell brukes når sesongen er en fast andel av nivået.
- Additiv modell brukes når sesongen er en fast verdi.
- Pseudo-Additiv er en ny modell i X-12-ARIMA. Den brukes for ikke negative tidsserier hvor observasjonene har ganske lave verdier i samme måneder hvert år.

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

- Trinn 1. Initiale estimeringer

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

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

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

(ii) SI_t ratene

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

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

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

(iv) Sesongfaktor

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

(v) Sesongjusterte

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

- Trinn 2 Estimerer sesongfaktor og sesongjusterte tall

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

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

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

(ii) SI -ratene

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

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

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

(iv) Sesongfaktor

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

(v) Sesongjusterte

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

- Trinn 3. Estimerer endelig Henderson trend og endelig irregulær komponent

(i) Endelig trend

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

(ii) Endelig irregulær komponent

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

Vi har

Sesongdekomponering

$$\begin{aligned}
 \text{(M)} \quad O_t &= T_t^{(3)} \times S_t^{(2)} \times I_t^{(3)} \\
 \text{(A)} \quad O_t &= T_t^{(3)} + S_t^{(2)} + I_t^{(3)} \\
 \text{(PA)} \quad O_t &= T_t^{(2)} \times (S_t^{(2)} - 1) + T_t^{(3)} \times I_t^{(3)}
 \end{aligned} \tag{4.46}$$

Hendersons vektene for kvartals og månedstall er listet ut i tabell III side 63 i X-11-ARIMA manualen. Merk at i trinn 1 i beregningen, er trenden beregnet med det sentrerte 12 måneders glidende gjennomsnittet. Vi får en foreløpig trend som ikke inneholder sesongvariasjoner. Trenden beregnes til de siste observasjonene siden tidsserien er forlenget med framskrivninger. I trinn 2 og 3 er trenden beregnet fra de sesongjusterte tall med Hendersons vektorer.

4.10 Diagnostiske prosedyrer for sesongjusteringer i X-12-ARIMA

I X-11-ARIMA har vi de 11 kvalitets målene M_1, \dots, M_{11} for trend T_t , sesongkomponent S_t og den irregulære komponenten I_t . De diagnostiske prosedyrene er forbedret i X-12-ARIMA med "sliding spans" og revisjon for å vurdere stabiliteten av sesongkomponenten og hvor stor endringene av sesongjusterte tall er når vi får nye observasjoner i tidsserien.

4.10.1 Kvalitets mål, $M_1 - M_{11}$ og Q

X-11 ARIMA programmet lister ut i tabell F.2 de 11 målene $M_1 - M_{11}$, hvor deres verdier varierer fra 0 til 3. For verdier under 1 er kvaliteten tilfredsstillende. Kvaliteten blir dårligere når M_i går mot 3. Formlene for $M_1 - M_{11}$ er gitt i notatet av J. Lothian og M. Morry (1978.c). La T_t , S_t og I_t være trenden, sesongkomponenten og irregulærkomponenten i en tidsserie. A_t er den sesongjusterte serien. De 11 målene beskrives i det følgende.

- M_1 : *Relativt bidrag fra I_t i den totale variasjonen.* La

$$\bar{I}(k)^2 = \frac{1}{N-k} \sum_{t=k+1}^N (I_t - I_{t-k})^2$$

$\bar{I}(k)^2$ er den totale variasjonen i lag k fra I_t . For de originale dataene er beregnet ved

$$\bar{O}'(k)^2 = \bar{T}(k)^2 + \bar{S}(k)^2 + \bar{I}(k)^2$$

der $\bar{T}(k)^2$ og $\bar{S}(k)^2$ beregnes på samme måte som $\bar{I}(k)^2$, fra tabellene D12 og D10. Det relative bidraget fra I_t i forhold til originale dataene er

$$R_{I(k)} = \frac{\bar{I}(k)^2}{\bar{O}'(k)^2} \times 100\%$$

Ved erfaringer viser man at $k = 3$ er best for å sammenligne. Grenseverdien for $R_{I(k)}$ er 10%. La

$$M1 = \frac{R_{I(3)}}{10}$$

så grenseverdien for $M1$ er 1. Når $M1 > 1$ viser det at $\bar{I}(k)^2$ er høy i forhold til $\bar{O}'(k)^2$, slik at trend eller sesongkomponent kan bli forstyrret av variasjoner fra I_t . Dermed kan hverken T_t eller S_t skilles helt perfekt fra I_t og sesongjusteringen er av lavere kvalitet.

- $M2$: *Relativt bidrag fra I_t til den stasjonære delen.* Bidraget av I_t sammenlignes med den originale serien etter å ha fjernet trendkomponenten fra den første runden. En størrelse $M2$ er innført for denne testen. Hvis $M2 > 1$ vil det vise at bidraget fra I_t er høyt.
- $M3$: *Endring fra måned til måned i I_t i forhold til endring i T_t .* Formålet i sesongdekomponeringen er å fjerne sesongkomponenten fra den originale serien for å ha et godt bilde av sesongjusterte serien. Ikke bare det, trenden og den irregulære komponenten må også være helt klar. Hvis endringen fra måned til måned i I_t er større enn endringen i trenden, blir det vanskelig å skille de to komponentene I_t og T_t i den tredje runden. Kvaliteten til dekomponeringen blir mindre. En størrelse $M3$ er definert for denne testen. Hvis $M3 > 1$ er variasjonen til I_t for høy.
- $M4$: *Mål på autokorrelasjon i I_t .* I sesongjustering ønsker man at I_t er en uavhengig og tilfeldig prosess. Man antar at I_t skrives på formen $I_t = \rho I_{t-1} + \epsilon_t$, (ϵ_t er hvit støy prosess). Testen er $H_0 : \rho = 0$ mot $H_1 : \rho \neq 0$. Et mål $M4$ er innført for denne testen. Når $M4$ er større enn 1, vil det vise en signifikans for $\rho \neq 0$. Da er I_t ikke en uavhengig tilfeldig prosess. Dette medfører at alle tester som er basert på I_t kan være ugyldige.
- $M5$: *Antallet måneder for at endringen i T_t skal bli større enn endringen i I_t .* En vil teste at det fins en k slik at etter k måneder skal endringene i T_t dominere endringene i I_t . Testen er ikke signifikant når $M5 > 1$.
- $M6$: *Endring fra år til år i I_t i forhold til endringen i S_t .* For å skille de to komponentene I_t og S_t bruker man (3×5) filtret for SI komponenten i den andre runden. Ved erfaringer viser det at når forholdet \bar{I}/\bar{S} er for lavt ($< 1,5$), er (3×5) filtret ikke tilpasset, siden filtret ikke er fleksibelt nok til å følge sesongvariasjonen. Omvendt når \bar{I}/\bar{S} er for høyt ($> 6,5$), da er det (3×5) filtret for fleksibelt slik at sesongkomponenten blander seg i den irregulære komponenten. En størrelse $M6$ er innført for å gi opplysninger til filteret. Hvis $M6 > 1$ betyr det at (3×5) filtret ikke er tilpasset. Et alternativ er å bruke (3×1) filtret når forholdet $\bar{I}/\bar{S} < 1,5$ eller stabil opsjonen når $\bar{I}/\bar{S} > 6,5$. Forholdet \bar{I}/\bar{S} er listet ut i tabell F2.H.
- $M7$: *Grad av bevegelig sesongmønster i forhold til stabilt sesongmønster.* Formålet er å måle hvor stor andelen av den stabile delen er i forhold til bevegelige delen i sesongkomponenten. En lager et mål $M7$ ved å basere seg på to F -tester, en for stabiliteten F_S og en for bevegelsen F_M . Lave verdier i F_S og høye verdier i F_M vil medføre til at sesongmønsteret ikke er

identifiserbart og verdien til $M7$ øker. Grensen for $M7$ er 1. Når $M7 > 1$ sier vi at sesongmønsteret ikke er identifiserbart.

Blant 11 tester for kvaliteten er denne testen viktigst. Årsakene til at $M7 > 1$ er:

- (i) Ingen sesong i det hele tatt i den originale serien.
- (ii) Variasjonen i sesongkomponenten er så stor at sesongmønsteret ikke kan identifiseres.
- (iii) Additiv modell er brukt når serien tilpasses bedre med multiplikativ modell.

(iii) kan rettes ved å endre modell for den originale serien. Verdien til $M7$ kan da bli mindre.

Ofte vil man ikke sesongjustere en serie når verdien til $M7$ er større enn 1.

- $M8$: *Grad av fluktuasjoner i sesongkomponenten gjennom hele serien.* En størrelse $M8$ er innført for måle fluktuasjoner i den sesongkomponenten S_t . $M8 > 1$ viser at fluktuasjonene er så store at S_t ikke er stabil lenger. Dermed får en et dårlig estimat for den sesongjusterte serien.
- $M9$: *Grad av lineær bevegelse i S_t i hele serien.* $M9 > 1$ viser at fluktuasjoner i S_t ikke er tilfeldige.
- $M10$: *Grad av fluktuasjoner i S_t i de siste årene.* Denne testen er samme som $M8$ men bare for de siste årene.
- $M11$: *Grad av lineær bevegelse i S_t i de siste årene.* Denne testen er samme som $M9$ men bare for de siste årene.

De to målene $M10$ og $M11$ gir informasjon om fluktuasjoner og bevegelser i sesongkomponenten for de siste årene. Testen for grad av fluktuasjoner ($M8$ og $M10$) er mindre viktig enn testen for grad av bevegelser ($M9$ og $M11$) i sesongkomponenten. Til slutt har vi en størrelse Q som er definert slik

$$Q = \frac{\sum_{i=1}^{11} w_i M_i}{\sum_{i=1}^{11} w_i}$$

hvor w_i er vekten for M_i . Målene som er relativt viktige, har større vekt. De 11 w_i -ene er gitt i tabell 4.10

Tabell 4.10: 11 vekter for 11 mål.

	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}
Verdi	13	13	10	5	11	10	16	7	7	4	4

Vi ser at $M1$, $M2$ og $M7$ har størst vekter fordi de er relativt sett viktigere enn de andre. Det er noen punkter som vi synes er viktige for brukeren for å kunne vurdere kvaliteten til den sesongjusterte serien.

- De 11 målene gjelder for de fleste seriene, men ikke for alle. De er bare empiriske mål.
- Mål som har større vekt er relativt sett viktigere.
- Hvis en serie der alle $M_i > 1$ vil vi ikke sesongjustere denne serien.
- Ingen mål kan alene bestemme kvaliteten av sesongdekomponeringen.
- Når $M_7 > 1$ skal en være svært forsiktig med å sesongjustere serien.
- Hvis brukeren bruker ett annet filter enn (3×5) filtret for estimering av sesongfaktorene, er $w_6 = 0$.
- For en serie som er mindre enn 6 år eller stabil opsjon er brukt for estimering av sesongfaktorer, får vektene w_8 , w_9 , w_{10} og w_{11} verdiene null. De 11 modifiserte vektene for de 11 målene er vist i tabell 4.11.

Tabell 4.11: 11 modifiserte vektene

	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}
Verdi	17	17	10	5	11	10	30	0	0	0	0

4.10.2 Sliding spans.

Formålet med sliding spans metoden (utviklet av David F. Findley, Brian C. Monsell, Holly B. Shulman og Marian G. Pugh, 1990), er å vurdere stabiliteter for sesongkomponenten S_t og sesongjusterte tall A_t i et bestemt tidspunkt når vi gjentar sesongjusteringsprosedyren på opp til fire forskjellige delvis overlappende delmengder av rådata (såkalt "span" på engelsk). Hver span er å betrakte som en komplett tidsserie og behandles akkurat som den originale tidsserien, dvs ukedagseffekter, påskeeffekt og intervensjoner er prekorrigert også fra dataene, før vi estimerer sesongkomponent S_t og sesongjusterte tall A_t . For eksempel, dataene av detaljvolumindeksen er observert fra januar 1979 til juli 2000, med 259 observasjoner. Sesongjusteringen utføres fire ganger i intervallene, [*jan.1989, jul.1997*], [*jan.1990, jul.1998*], [*jan.1991, jul.1999*] og [*jan.92, jul.2000*]. Sesongkomponentene og sesongjusterte tallene for hvert tidspunkt i fire span er registrert for å beregne revisjoner.

span 1	jan.89	_____	jan.95	_____	jul.97			
span 2		jan.90	_____	jan.95	_____	jul.98		
span 3			jan.91	_____	jan.95	_____	jul.99	
span 4				jan.92	_____	jan.95	_____	jul.00

Vi må definere noen symboler. La

- $S_t(k)$ være sesongfaktor estimert fra span k for måned t .
- $A_t(k)$ være sesongjusterte tall estimert fra span k for måned t .

- $MM_t(k)$ være endring i prosent fra måned til måned av sesongjusterte tall, fra span k for måned t . $MM_t(k)$ er definert ved

$$MM_t(k) = \frac{A_t(k) - A_{t-1}(k)}{A_{t-1}(k)}$$

- $YY_t(k)$ være endring i prosent fra år til år av sesongjusterte tall, fra span k for måned t . $YY_t(k)$ er definert slik

$$YY_t(k) = \frac{A_t(k) - A_{t-12}(k)}{A_{t-12}(k)}$$

- $N_t = \{k: \text{måned } t \text{ er i den } k\text{-te span}\}$.
- $N_{12_t} = \{k: \text{måned } t \text{ og } t - 1 \text{ er i den } k\text{-te span}\}$.

En sesongfaktor S_t i måned t kalles *upålitelig* hvis

$$S_t^{max} = \frac{\max_{k \in N_t} S_t(k) - \min_{k \in N_t} S_t(k)}{\min_{k \in N_t} S_t(k)} > 0,03 \quad (4.47)$$

Denne relasjonen er en test for at den største differansen i prosent for sesongfaktorene i måned t , S_t^{max} , er større enn 3%.

Vi sier at endringen i prosent fra måned til måned i den sesongjusterte serien er *upålitelig* hvis

$$MM_t^{max} = \max_{k \in N_{12_t}} MM_t(k) - \min_{k \in N_{12_t}} MM_t(k) > 0,03 \quad (4.48)$$

Denne relasjonen er en test for at den største differansen i prosent for sesongjusterte tallene i måned t , MM_t^{max} , er større enn 3%.

La $N_{12_t} = \{k: \text{måned } t \text{ og } t - 12 \text{ er i span } k\}$ og

$$YY_t(k) = \frac{A_t(k) - A_{t-12}(k)}{A_{t-12}(k)}$$

$YY_t(k)$ er årlig endringen i prosent av sesongjusterte tall for måned t . $YY_t(k)$ kalles *upålitelig* hvis

$$YY_t^{max} = \max_{k \in N_{12_t}} YY_t(k) - \min_{k \in N_{12_t}} YY_t(k) > 0,03 \quad (4.49)$$

La

- $S(\%)$ være prosentandelen måneder hvor sesongkomponenten er definert som *upålitelig* (dvs $S_t^{max} > 0,03$).
- $MM(\%)$ være prosentandel måneder hvor $MM_t^{max} > 0,03$.
- $YY(\%)$ være prosentandel måneder hvor $YY_t^{max} > 0,03$.

Grenseverdiene for $S\%$, $MM\%$ er gitt i tabell 4.12

4.10.3 Revisjon

Tabell 4.12: *Evaluering en sesongjustering basert på $S(\%)$, $MM(\%)$ og $YY(\%)$*

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

Vi antar at O_t skrives som en additive dekomponering

$$O_t = S_t + T_t + I_t$$

hvor S_t er sesongkomponent. La $\nu(B)$ være filter for å estimere sesongkomponenten i den siste runden. Vi får

$$\widehat{S}_t = \nu(B)O_t \quad (4.50)$$

$\nu(B)$ er symmetrisk og skrives slik

$$\nu(B) = \cdots + \nu_{-2}B^2 + \nu_{-1}B + \nu_0 + \nu_1F + \nu_2F^2 + \cdots \quad (4.51)$$

hvor $BO_t = O_{t-1}$ og $FO_t = O_{t+1}$. For X-11 eller X-12-ARIMA er $\nu(B)$ det 3×3 , 3×5 eller 3×9 filteret, mens $\nu(B)$ er Wiener-Kolmogorov filteret i SEATS.

Vi antar at O_t beskriver som en rekke av ϵ_t ,

$$O_t = \epsilon_t + \psi_1\epsilon_{t-1} + \psi_2\epsilon_{t-2} + \cdots = \psi(B)\epsilon_t \quad (4.52)$$

hvor ϵ_t er hvit støy prosess. Fra (4.50) og (4.52) får vi

$$\widehat{S}_t = \nu(B)\psi(B)\epsilon_t = \xi(B)\epsilon_t \quad (4.53)$$

hvor $\xi(B) = \cdots + \xi_{-2}B^2 + \xi_{-1}B + \xi_0 + \xi_1F + \xi_2F^2 + \cdots$. La $\widehat{S}_{t|t+k}$ være estimerte verdien av S_t gitt data til og med tidspunkt $t+k$.

$$\begin{aligned} \widehat{S}_{t|t+k} &= E_{t+k}[\xi(B)\epsilon_t] \\ &= E_{t+k} \left[\left(\cdots + \xi_{-1}B + \xi_0 + \xi_1F + \cdots + \xi_kF^k + \cdots + \xi_{k+l}F^{k+l} + \cdots \right) \epsilon_t \right] \\ &= \left(\cdots + \xi_{-1}B + \xi_0 + \xi_1F + \cdots + \xi_kF^k \right) \epsilon_t \\ &= \xi^k(B)\epsilon_t \end{aligned} \quad (4.54)$$

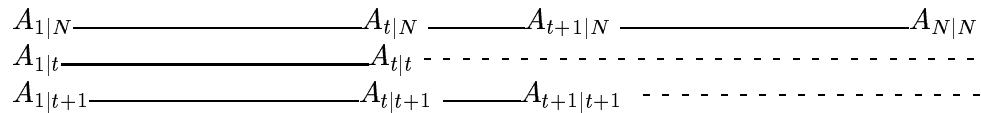
hvor $\xi^k(B)$ er et tilsvarende filter med $\xi(B)$, men det blir "trunkert" etter F^k , dvs $\xi_{k+l} = 0$ for $l > 0$. Revisjonen for \widehat{S}_t gitt data til og med tidspunkt $t+k$ er gitt ved

$$R_k = \widehat{S}_t - \widehat{S}_{t|t+k} = \sum_{i=k+1}^{\infty} \xi_i \epsilon_{t+i} \quad (4.55)$$

Grunnene til revisjonen når nye observasjoner tilføyes, er

- i. Serien forlenges med nye framskrivinger.
- ii. På grunn av (i) kan $\nu(B)$ i (4.51) og $\psi(B)$ i (4.52) få et mønster som er annerledes enn det tidligere mønsteret. Dermed blir modellen for $\xi(B)$ i (4.53) endret. Vi har en revisjon for sesongkomponenten \widehat{S}_t i tidspunkt t .

Beregningen for revisjonen av sesongjusterte tallene A_t kan beskrives ved figur



hvor

$A_{t|u}$ sesongjusterte tall av serien O_t , hvor $1 \leq t \leq u \leq N$.

$A_{t|t}$ siste justering (concurrent adjustment)

$A_{t|N}$ most recent adjustment

Prinsippet for å beregne revisjoner for sesongjusterte tall i X-12-ARIMA fra tidspunkt t er at man sesongjusterer bare en del av den originale serien med observasjonene til tidspunkt t . Sesongjusterte verdien $A_{t|t}$ beregnes og sammenlignes med verdien i tidspunkt t for hele serien $A_{t|N}$. Merk at t må velges slik at lengden av en delmengde av observasjoner er langt nok for at effektene av ukedager, påske og intervensjoner kan korrigeres liksom hele serien.

La

$$R_{t|N}^A = 100 \times \frac{A_{t|N} - A_{t|t}}{A_{t|t}} \quad (4.56)$$

For bestemte tidspunkter N_0 og N_1 er $R_{t|N}^A$ *revisjonshistorier* av sesongjusterte serien i intervallet $[N_0, N_1]$.

Prosentvis endring fra periode til periode

$$\Delta A_{t|u} = 100 \times \frac{A_{t|u} - A_{t-1|u}}{A_{t-1|u}} \quad (4.57)$$

4.11 Eksempel: Detaljvolumindeks for totalen

Tallene er observert fra januar 1979 til juli 2000, i alt 259 observasjoner. I estimate- ringen av trend T_t , sesongkomponent S_t , sesongjusterte tall A_t , og den irregulære komponenten I_t lar vi programmet selv velge

- den multiplikative ($O_t = T_t \times S_t \times I_t$) eller additve modellen ($O_t = T_t + S_t + I_t$).
- lengden av Hendersons filter for trenden og sesongkomponenten.

- en av de fem innebygde ARIMA modellene:

(0,1,1)(0,1,1)

$$O_t = O_{t-1} + O_{t-12} - O_{t-13} + \epsilon_t - \theta\epsilon_{t-1} - \Theta\epsilon_{t-12} + \theta\Theta\epsilon_{t-13}$$

(0,1,2)(0,1,1)

$$O_t = O_{t-1} + O_{t-12} - O_{t-13} + \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \Theta\epsilon_{t-12} + \theta_1\Theta\epsilon_{t-13} + \theta_2\Theta\epsilon_{t-14}$$

(2,1,0)(0,1,1)

$$O_t = O_{t-1} + \phi_1 O_{t-1} - \phi_1 O_{t-2} + \phi_2 O_{t-2} - \phi_2 O_{t-3} + O_{t-12} - O_{t-13} - \phi_1 O_{t-13} + \phi_1 O_{t-14} - \phi_2 O_{t-14} + \phi_2 O_{t-15} + \epsilon_t - \Theta\epsilon_{t-12}$$

(0,2,2)(0,1,1)

$$O_t = 2O_{t-1} - O_{t-2} + O_{t-12} - 2O_{t-13} + O_{t-14} + \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \Theta\epsilon_{t-12} + \theta_1\Theta\epsilon_{t-13} + \theta_2\Theta\epsilon_{t-14}$$

(2,1,2)(0,1,1)

$$O_t = O_{t-1} + \phi_1 O_{t-1} - \phi_1 O_{t-2} - \phi_2 O_{t-3} + O_{t-12} - O_{t-13} - \phi_1 O_{t-13} + \phi_1 O_{t-14} - \phi_2 O_{t-14} + \phi_2 O_{t-15} + \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \Theta\epsilon_{t-12} + \theta_1\Theta\epsilon_{t-13} + \theta_2\Theta\epsilon_{t-14}$$

Tre kriterier for å velge en ARIMA modell er:

- (i) Gjennomsnittlig absolutt prognosefeil i prosent i de siste 3 årene er mindre eller lik 15%.
- (ii) χ^2 for testing om residualene er ukorrelerte, må være større enn 5%.
- (iii) Det er ingen tegn til "overdifferencing" ($\sum_i \theta_i > 0.9$ eller $\sum_i \Theta_i > 0.9$).

Vi lar også programmet selv prekorrigere for ukedagseffekter, påskeeffekt og intervensjoner når testen viser signifikante effekter.

Vi får følgende resultater:

- Modell for sesongjustering

Testen viser at den multiplikative modellen passer best med data. Dvs, alle beregningene i programmet er utført med *log* transformasjon, isteden for med rådata. Etter beregningen blir trenden, sesongkomponenten og sesongjusterte tall transformert med *exp(x)* tilbake til den opprinnelige skalaen.

- ARIMA modell

ARIMA (0,1,1)(0,1,1) er valgt. Vi får

	Est.verdi	Std.	t -verdi	
$\hat{\theta}$	0.7225	0.0433	16.67	signifikant på 5%
$\hat{\Theta}$	0.6231	0.0509	12.24	signifikant på 5%

I modellen er det to parametre som estimeres. De er θ for ikke-sesongvariasjoner og Θ for sesongvariasjoner. Begge parametre er signifikante på 5% nivå.

Gjennomsnittlig absolutt prosent feil for framskrivningen er

siste år: 2.09, forrige år: 3.46, to år tidligere: 1.91
siste tre år: 2.49

Vi får ganske god tilpasning til data med modellen

- Ukedagseffekter og påskeeffekt

De to effektene er signifikante på 5% nivå.

Ukedager $\chi^2 = 124.38$ med 6 frihetsgrader
Påske $t = -5.45$

Tabell 4.13: *Ukedagseffekter*

	Est.verdi	Std.	t -verdi
mandag	-0.0055	0.0033	-1.66
tirsdag	-0.0007	0.0033	-0.20
onsdag	0.0052	0.0033	1.59
torsdag	0.0130	0.0033	3.99
fredag	-0.0019	0.0033	-0.59
lørdag	0.0037	0.0033	1.14
søndag	-0.0139	0.0033	-4.19

Tolkningen av denne tabellen er at aktiviteten er høyest på torsdagen og lavest på søndagen.

- Intervensjoner

Februar 1987 og desember 1988 er to intervensjoner. Størrelser og typer effekter er vist i tabellen nedenfor

Tidspunkt	Type	Est.verdi	Std.	t -verdi
Februar 1987	LS	-0.0749	0.0159	-4.69
Desember 1988	AO	-0.0890	0.0215	-4.13

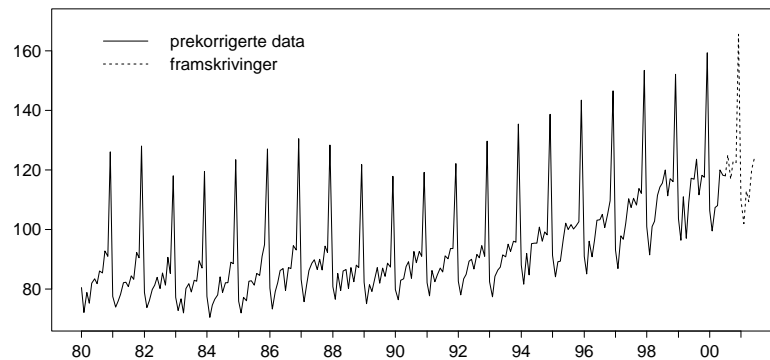
I X-12-ARIMA lister man ut data i tabellene. For eksempel, A 1 inneholder rådata og faktorene for prekorrigering av ukedagseffekter, påskeeffekt og intervensjoner ligger i A6, A7 og A8 henholdsvis. I følge tabell A8 er tidspunktene etter februar 1987 korrigeret ned med faktor 1.0778 og desember 1988 korrigeret opp med faktor 0.9148 på grunn av ekstreme verdier.

- Framskrivinger

Figur 4.14 viser framskrivinger $\hat{O}_t^{(1)}$ for rådata og de framskrivingene $\hat{O}_t^{(2)}$ som er beregnet fra rådata etter å ha korrigert for ukedagseffekter, påskeeffekt og intervensjoner. Vi får bare et 95% konfidensintervall for $\hat{O}_t^{(1)}$, hvor $l95$ og $u95$ er nedre og øvre grensen. De prekorrigerte data og framskrivingene $\hat{O}_t^{(2)}$ er plottet i figur 4.20.

Tabell 4.14: *Framskrivinger fra august 2000 til juli 2001*

år	måned	$l95$	$\hat{O}_t^{(1)}$	$u95$	$\hat{O}_t^{(2)}$
2000	august	121.04	127.30	133.88	125.08
	september	111.21	117.17	123.46	116.96
	oktober	114.13	120.46	127.13	122.89
	november	118.00	124.77	131.93	122.52
	desember	154.35	163.46	173.12	165.45
2001	januar	103.45	109.74	116.42	109.84
	februar	95.22	101.13	107.42	102.04
	mars	108.79	115.78	123.23	112.95
	april	99.64	106.19	113.16	109.35
	mai	114.50	122.24	130.50	120.12
	juni	115.81	123.79	132.32	123.56
	juli	112.70	120.66	129.17	123.10

Figur 4.20: *Prekorrigerte data og framskrivinger*

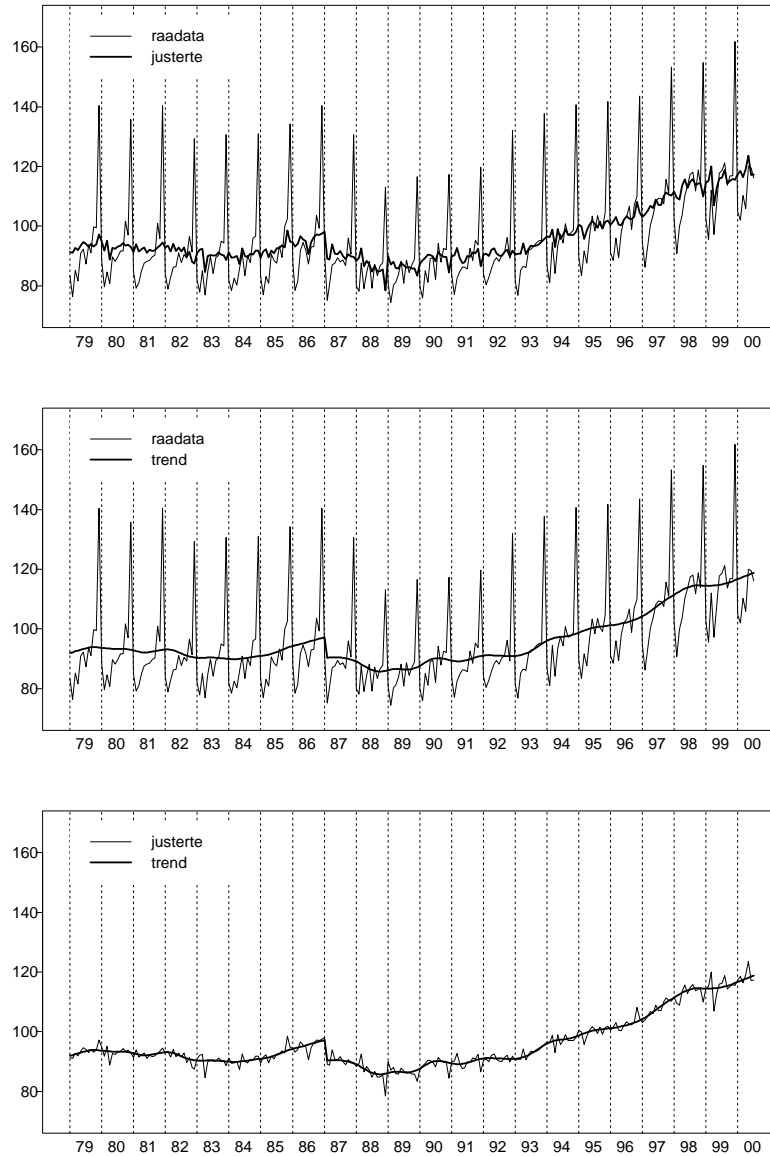
- Testen for sesongmønsteret og sesongbevegelsen

Sesongmønsteret er signifikant på 5% nivå, med F -verdi=657.74. Figur 4.15 viser sesongvariasjoner gjennom årene.

Sesongbevegelsen er også signifikant på 5% nivå, med F -verdi=2.60

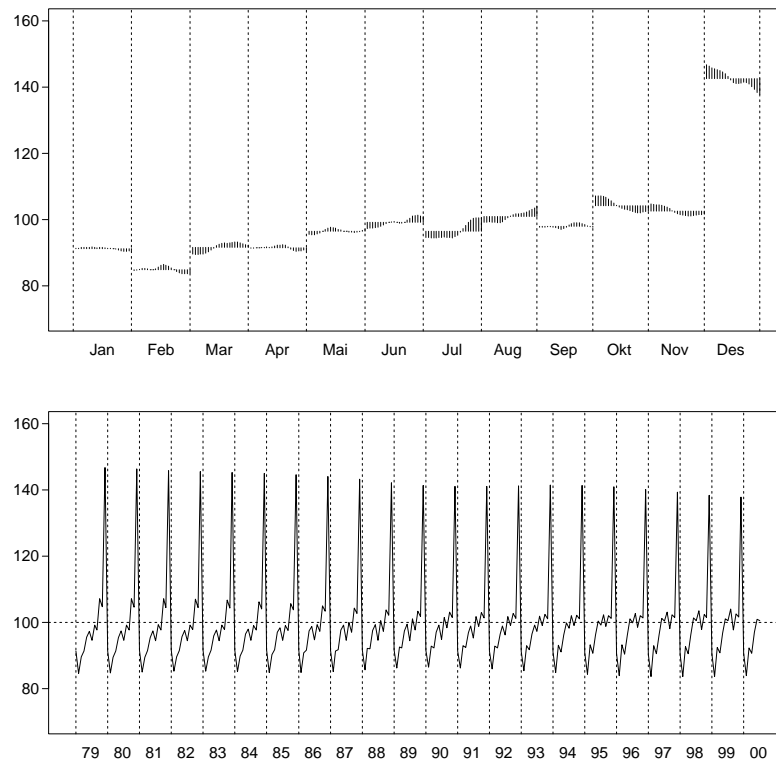
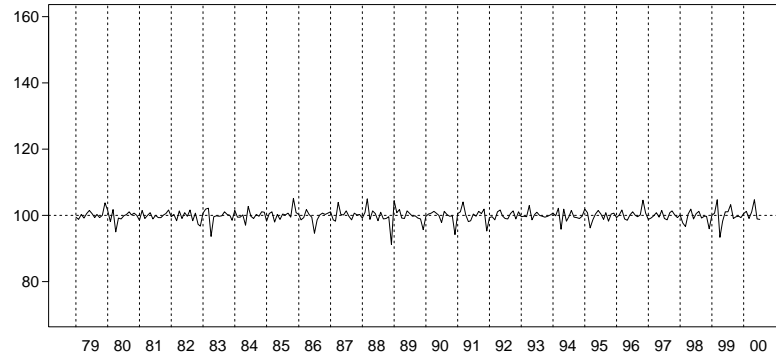
- Figurer

Trenden og sesongjusterte tall plottes i figur 4.21. Vi ser at trenden er glatt og det er et nivåskift i februar 1987. Sammenlignet med rådata er endringene av sesongjusterte tall mye lavere. Dette viser høye sesongvariasjoner.



Figur 4.21: *Trenden og sesongjusterte tall*

Sesongkomponenten er vist i to bilder i figur 4.22. I det første bildet sorterer vi sesongfaktorene etter årene for hver måned, og plotter dem mot deres gjennomsnittet. En fordel er at det blir lettere å se endringene av sesongvariasjoner fra år til år i samme måned. Dessuten er det en enkel metode for å tolke størrelsen og formen på sesongfaktorene for 12 måneder. I det andre bildet plotter vi sesongfaktorene S_t , med t =jan.79, feb.79, ..., jul.2000.

Figur 4.22: *Sesongkomponent*Figur 4.23: *Irregulær komponent*

Vi har den irregulære komponenten i figur 4.23. Testen viser at det er en hvit støy prosess. Dvs, vi har en gevinst av å korrigere dataene for effektene av ukedager, påske og intervensjoner samt å fjerne sesongvariasjoner.

- Kvalitetsmålene $M1 - M11$ og Q

Verdiene av de 11 målene $M1 - M11$ og Q er gitt i tabell 4.15. Vi får høye verdier av $M3$ og $M5$ (større enn 1) på grunn av at trenden er glattere enn den irregulære komponenten. De andre målene og Q viser en god sesongjustering.

Tabell 4.15: *Kvalitets mål*

$M1$	$M2$	$M3$	$M4$	$M5$	$M6$
0.089	0.056	1.510	0.536	1.620	0.200
$M7$	$M8$	$M9$	$M10$	$M11$	Q
0.106	0.184	0.110	0.190	0.175	0.460

- Sliding spans

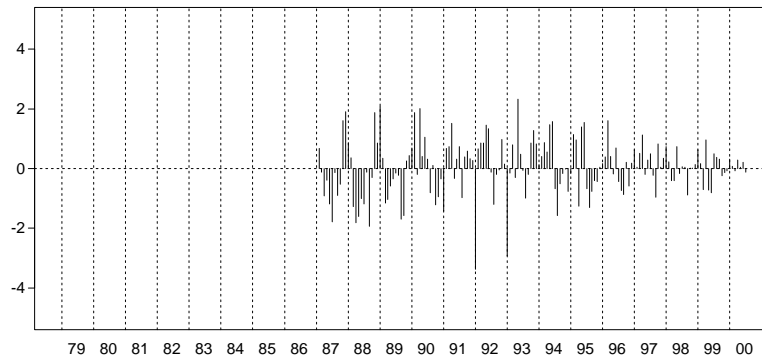
Sliding spans ble utført med fire delvis overlappende delmengder: [januar 1989, juli 1997], [januar 1990, juli 1998], [januar 1991, juli 1999], og [januar 1992, juli 2000]. Dataene i disse intervallene sesongjusteres på samme måte som den opprinnelige tidsserien. Sesongjusterte tall for hver span blir brukt for å beregne $S(\%)$, $MM(\%)$ og $YY(\%)$. Vi får

$$S(\%) = 0, \quad MM(\%) = 0, \quad \text{og} \quad YY(\%) = 0$$

Grenseverdiene for disse størrelsene er gitt i tabell 4.12. Vi får en god sesongjustering for detaljvolumindeksen.

- Revisjoner for sesongjusterte tall

La $A_{jan.87|jan.87}$ være sesongjusterte tall i januar 1987, som er beregnet med rådata fra januar 1979 til januar 1987, og $A_{jan.87|jul.2000}$ være verdien i januar 1987 for den tidsserien fra januar 1979 til juli 2000.

Figur 4.24: *Revisjoner av sesongjusterte tall*

Revisjonen for sesongjusterte tall i januar 1987 er gitt ved

$$R_{jan.87} = \frac{A_{jan.87|jul.2000} - A_{jan.87|jan.87}}{A_{jan.87|jan.87}} \times 100\%$$

Vi beregner på samme måte for februar 1987, \dots , juni 2000. Tallene varierer i mellom -2% og 2% og plottet i figur 4.24.

Register

- 12 mnd. gj, 77
- 13 mnd. Hendersons gj., 79
- 13 mnd. gj., 78
- 23 mnd. Hendersons gj., 81
- 23 mnd. gj., 79

- Additive Outlier (AO), 67
- AIC, 55
- AR(P) modell med sesongvariasjoner, 50
- AR(1) modell med sesongvariasjoner, 49
- ARIMA(p, d, q)(P, D, Q)_s modell med sesongvariasjoner, 52
- ARMA(p, q), 37
- ARMA(p, q)(P, Q) modell med sesongvariasjoner, 51
- Autokorrelasjonsfunksjon (ACF), 25
- Autokovariansfunksjon, 24
- Autoregressiv prosess av orden 1, AR(1), 26
- Autoregressiv prosess av orden 2, AR(2), 33

- BIC, 55
- Blandet autoregressiv og glidende gjennomsnitt prosess, ARMA(1,1), 31
- Blandet autoregressiv og glidende gjennomsnitt prosess, ARMA(2,2), 36

- Deterministisk funksjon, 23

- En generell lineær prosess, 38

- Framskrivning, 45

- Glidende gjennomsnitt prosess av orden 1, MA(1), 29
- Glidende gjennomsnitt prosess av orden 2, MA(2), 34

- Hvit støy prosess, 26

- Ikke-stasjonær prosess. ARIMA(p, d, q), 39
- Innovational outlier (IO), 68
- Intervensjon, 64
- Irregulærkomponent, 85

- Kvalitets mål M_1, \dots, M_{11} og Q , 90

- Lag, 28
- Level shift (LS), 68

- MA(Q) modell med sesongvariasjoner, 50

- Påskeeffekt, 74
- Partiell autokorrelasjonsfunksjon (PACF), 41
- Pulsfunksjon, 66

- Ramp, t_0 to t_1 , 70
- Revisjon, 95

- Sesongjusterte tall, 85
- Sesongvariasjon, 49
- Sliding spans, 93
- Stasjonæritet, 26
- Stokastisk funksjon, 24

- Temporary change (TC), 69
- Tidsserie, 23
- Trappfunksjon, 66
- Trend, 76

- Ukedageeffekter, 70

- X-12-ARIMA, 85

Bibliografi

- [1] Alan Pankratz (1991), "Forecasting with Dynamic Regression Models", Wiley Interscience
- [2] Bell W. R. and Hillmer S. C. (1983). "Modelling Time Series With Calendar Variation". *Journal of the American Statistical Association*, 78, 526-534
- [3] Cleveland W. S. and Susan J. D. (1980), "Calendar Effects in Monthly Times Series: Detection by Spectrum Analysis and Graphical Methods", *Journal of the American Statistical Association*, 75, 487-495
- [4] Findley D. F., Brian C. Monsell, William R. Bell, Mark C. Otto and Bor-Chung Chen (1998). "New Capabilities and Methods of the X-12 ARIMA Seasonal Adjustment Program", *Journal of Business & Economic Statistics*, 16, 127-177
- [5] Dagum Estela Bee, Benoit Quenneville and Brajendra Sutradhar (1992). "Trading-day Variations Multiple Regression Models with Random Parameters", *International Statistical Review*, 60, 57-73
- [6] Dagum Estela Bee (1988). "The X11ARIMA/88 Seasonal Adjustment Method Foundations and User's Manual"
- [7] John Higginson (1975). "An F Test for the presence of moving seasonality when using census method II-X-11 variant"
- [8] Lars A. Loe (1987). "Framskrivning av tidsseriedata i kvartalsvis nasjonalregnskap", Notater 87/1
- [9] Leiv Solheim og Dinh Quang Pham (1997). "Prekorrigering av påskeeffekten for detaljvolumindeksen 1979-1997", Notater 73/97
- [10] Lothian J. and M. Morry. "A set of Quality Control Statistics for the X-11 ARIMA"
- [11] Bureau of the Census. "X-12 ARIMA Reference Manual, Version 0.2.5, October 1, 1999"
- [12] S-PLUS User'Manual, version 3.2, December 1993

De sist utgitte publikasjonene i serien Notater

- 2000/53 G. Dahl og B.R. Joneid: FD - Trygd: Dokumentasjonsrapport: Inntekt og formue, 1992-1997. 36s.
- 2000/54 K. Flugsrud og G. Haakonsen: Utslipp av klimagasser i norske kommuner: En gjennomgang av datakvaliteten i utslippsregnskapet. 44s.
- 2000/56 T. Hægeland: "Ny" vekstteori: Et nytt forskningsprogram eller naturlig progresjon? Utviklingen innenfor økonomisk vekstteori vurdert i forhold til Imre Lakatos' vitenskapsfilosofi. 19s.
- 2000/57 K.-G. Lindquist: SAS-programmer for korrigering av data fra industristatistikken og beregning av variable for analyseformål. 53s.
- 2000/58 A. Akselsen: FD - Trygd: Dokumentasjon av uttak til Sandmanutvalget (SHD). 28s.
- 2000/59 J. Johansen og Ø. Sivertstøl: FD - Trygd: Dokumentasjonsrapport: Fødsels- og sykepenger, 1992-1998. 109s.
- 2000/60 K-G. Lindquist: Database for energintensive næringer: Tall fra industristatistikken: Oppdatert versjon av notat 97/30. 17s.
- 2000/61 O. Haugen: Utrekning av vekter til inntekts- og formuesundersøkingane 1998. 24s.
- 2000/62 Ø. Kleven: Ferieundersøkelsen 1999 Panel: Dokumentasjonsrapport. 50s.
- 2000/63 P.G. Larssen: Overvakingssystemet for bedrifter i BoF. 29s.
- 2000/64 R. N. Johnsen: Undersøking om foreldrebetaling i barnehagar, august 2000. 36s.
- 2000/65 A. Thomassen: Byggekostnadsindeks for rørleggerarbeid i kontor- og forretningsbygg. 14s.
- 2000/67 A.G. Hustoft og G. Olsen: Metadata for statistikk om personer og husholdninger : Forprosjektrapport. 34s.
- 2000/68 A. Bruvoll, K. Flugsrud og H. Medin: Dekomponering av endringer i utslipp til luft i Norge - dokumentasjon av data. 19s.
- 2000/69 M. Vik Dysterud og E. Engelian: Tettstedsavgrønsing: Teknisk dokumentasjon 2000. 53s.
- 2000/70 A. Akselsen, G. Dahl, J. Lajord og Ø. Sivertstøl: FD - Trygd: Variabelliste. 48s.
- 2000/71 B.O. Lagerstrøm: Kompetanse i grunnskolen , del 2: Dokumentasjonsrapport. 19s.
- 2000/72 B.O. Lagerstrøm: Kompetanse i grunnskolen: Hovedresultater 1999/2000 170s.
- 2000/73 J.H. Wang: Kvartalsvis investeringsstatistikk. 57s.
- 2000/74 P.O. Lande og T. Hoel: Dødsårsaksregisteret: Systemdokumentasjon. 90s.
- 2000/75 A.G. Pedersen, P.O. Lande og T. Hoel: Dødsårsaksregisteret: Brukerdokumentasjon. 99s.
- 2000/76 A.G. Hustoft, B. Vannebo: En undersøkelse av frafallet i utvalgsundersøkelser i perioden 1997-2000. 56s.
- 2000/77 P.O. Lande og J. Kittelsen: Forbruksundersøkinga 2000. Innlasting/Innsjekking: Brukerdokumentasjon. 17s.
- 2000/78 J. Fosen, A.K. Johnsen og G. Røyne: Frafall blant innvandrere. En undersøkelse av frafall i Utdanningsundersøkelsen 1999 og i valgundersøkelser blant innvandrere. 53s.
- 2000/79 J. Kittelsen og P.O. Lande: OPPSLAG - Forbruksundersøkelsen. Brukerdokumentasjon. 39s.
- 2000/80 J. Kittelsen og P. O. Lande: Forbruksundersøkinga 2000. Systemdokumentasjon . 156s.
- 2000/81 J. T. Lind: Testing av stokastiske individuelle effekter i paneldatamodeller. 16s.