

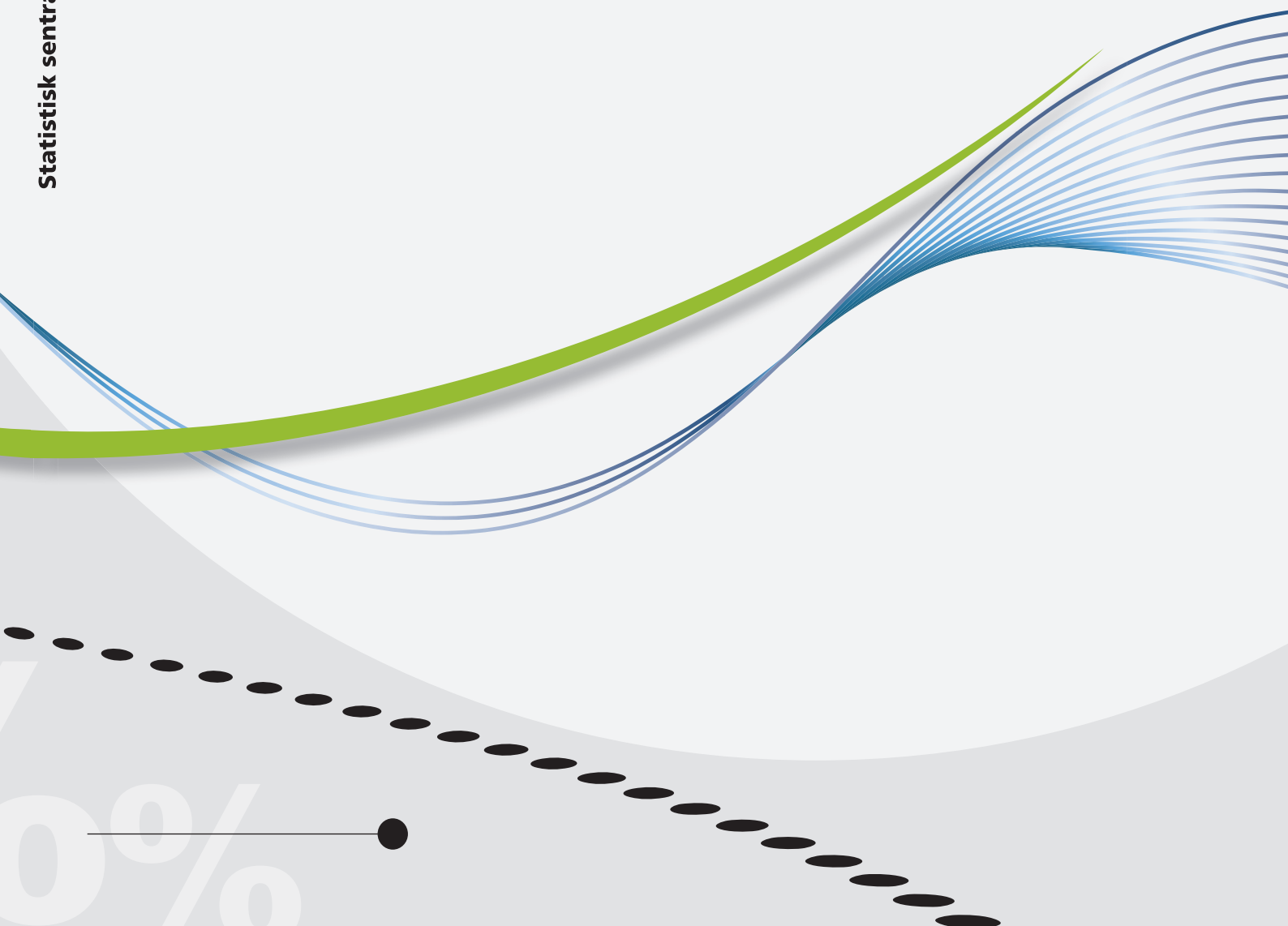
Dinh Quang Pham

Kalendereffekter
Modell og estimering

Statistics Norway



Statistisk sentralbyrå



Dihn Quang Pham

Kalendereffekter

Modell og estimering

© Statistisk sentralbyrå Ved bruk av materiale fra denne publikasjonen skal Statistisk sentralbyrå oppgis som kilde. Publisert juli 2012	Standardtegn i tabeller Tall kan ikke forekomme Oppgave mangler Oppgave mangler foreløpig Tall kan ikke offentliggjøres Null Mindre enn 0,5 av den brukte enheten Mindre enn 0,05 av den brukte enheten Foreløpig tall Brudd i den loddrette serien Brudd i den vannrette serien Desimaltegn	Symbol : - 0 0,0 * — ,
ISBN 978-82-537-8433-5 (trykt) ISBN 978-82-537-8434-2 (elektronisk) ISSN 1891-5906 Emne: 00.90		
Trykk: Statistisk sentralbyrå		

Forord

Dette notatet beskriver modeller og estimering av kalendereffekter med norske data. Resultatene ble presentert for metodegruppen før selve notatet ble utformet. Jeg vil rette en takk til mine kollegaer for gode og konstruktive kommentarer som bidro til bedre modeller og estimeringer. En spesiell takk til Jan Bjørnstad og Øyvind Langsrud. Videre vil jeg takke Terje Skjerpen for gode diskusjoner om Kalmanfilter og Ane Seierstad for retting av manuskriptet.

Statistisk sentralbyrå, 11. juli 2012

Jan Furseth
(fung. adm.dir.)

Sammendrag

Det er mange tidsserier i SSB som er et aggregat av daglige verdier. Effektene som skyldes ulikt antall mandager, tirsdager, ..., søndager fra måned til måned gjør at det blir vanskelig å sammenligne aktiviteten i tidsserien i to tidspunkter. I dette notatet beskriver vi i detalj egenskaper ved kalenderen og gir en forklaring av ukedagseffekter. Vi presenterer flere alternativer for estimering av effektene. Figurer er også brukt for å illustrere analysen. Et konkret eksempel med tall fra VKI (varekonsumprisindeks) blir analysert.

Innhold

1	Innledning	5
2	Egenskaper ved kalenderen	6
3	Forklaring av ukedagseffekter	9
4	Metoder for å sjekke og estimere kalendereffekter	10
4.1	Spektralanalyse for ukedagseffekter	10
4.2	Estimering av ukedagseffekter ved X-11 og X-12-ARIMA	11
4.2.1	Modell basert på irregulærkomponenten	11
4.2.2	Modell basert på regARIMA	12
4.3	En ny korrigering av påske i SSB	13
5	Korrigering ukedageffekter i SSB	13
6	Modell for tidsserie	14
7	Eksempel: Matvarer	15
7.1	Alternativ 1. Korrigeringen med regARIMA og uten hensyn til røde dager og påske i Norge	17
7.2	Alternativ 2. Korrigeringen med regARIMA og med hensyn til røde dager og påske i Norge	17
7.3	Alternativ 3. Korrigeringen ikke med regARIMA modell, men med hensyn til røde dager og påske i Norge	18
7.3.1	Kalmanfilterteknikk	19
7.3.2	Tilstandsmodellene for matvarer	20
8	Oppsummering	21

Tabeller

1	<i>Ulike typer ukedager</i>	7
2	<i>Frekvenser av påske i perioden 1600-2099</i>	8
3	<i>Antall ukedager i mars og april i 2008 og 2009.</i>	8
4	<i>Matvarer og aggregering</i>	15
5	<i>Estimering av kalendereffekter ved alternativ 1.</i>	17
6	<i>Estimering av kalendereffekter ved alternativ 1 og 2.</i>	18
7	<i>Estimering av kalendereffekter ved alternativ 2 og 3</i>	21

Figurer

1	<i>Frekvenser av antall ganger påske kommer</i>	8
2	<i>Ujusterte verdier</i>	16
3	<i>ACF av ujusterte verdier</i>	16
4	<i>Ukedageffekter ved spektralanalysen</i>	16
5	<i>Effektene av mandag-torsdag ved alternativ 2 og 3</i>	23
6	<i>Effektene av fredag og lørdag ved alternativ 2 og 3</i>	24

1 Innledning

Det er mange tidsserier i SSB der verdien Y_t i måned t , er et aggregat av daglige verdier $X_{i,t}$. Da blir $Y_t = \sum_i X_{i,t}$. To eksempler er tidsseriene antall biler registrert første gang i Norge og antall hotellovernattinger foretatt av nordmenn, den observerte verdien i måned t , Y_t , er summen av de daglige observerte verdiene i denne måneden ("flow series"). Det er mange faktorer som påvirker verdien til Y_t . Det er:

- lengden av måneden (28, 29, 30 eller 31 dager).
- antall ulike ukedager i en måned fra år til år (det er 4 mandager, 4 tirsdager, 4 onsdager, 5 torsdager, 5 fredager, 5 lørdager og 4 søndager i juli 2010). Siden vi pleier å handle mest matvarer i slutten av uken blir omsetningen i butikkene i denne måneden høyere enn en måned som har færre torsdager, fredager og lørdager. Det er effekten av ulike antall ukedager i en måned som påvirker omsetningen av butikkene (vi kaller det ukedagseffekter).
- påske eller pinse, som vi kaller bevegelige helligdager. Påske kan ligge i mars eller april og pinse mellom mai og juni. Det varierer fra år til år. Kalenderen gjentar seg etter 28 år men det fins ikke en syklus for påske og pinse.

Variasjoner i kalenderen skaper effekter (vi kaller det kalendereffekter) som påvirker verdien av tidsserien. I sesongjustering må disse effektene fjernes for å spalte opp tidsserien i trend og sesongkomponent.

Det er mange ulemper når vi bruker metoden i X-12-ARIMA for korrigering av kalendereffekter med norske data. Det er:

1. røde dager i Norge blir ikke tatt hensyn til, siden X-12-ARIMA er laget med grunnlag i den amerikanske kalenderen. 1.mai og 17.mai er to helligdager i Norge, men de blir behandlet som arbeidsdager i X-12-ARIMA,
2. påske i USA er bare en søndag mens det er Skjærtorsdag, Langfredag, påskeaften, 1.påskedag og 2.påskedag i Norge. Disse helligdagene varierer i mars og april. I X-12-ARIMA blir de behandlet som arbeidsdager. Folk handler ofte mer matvarer noen dager før påske for å bruke dem i påsken. Når påsken kommer tidlig i april, som for eksempel i 2002 (påskesøndag var 31.mars) handler vi mer matvarer i mars. Omsetningen i mars øker og reduseres i april. Påskeeffekten i 2002 må korrigeres slik at vi kan sammenligne aktiviteten i mars og april. Dette kan vi ikke gjøre med X-12-ARIMA,
3. korrigeringen for kalendereffekter blir utført med en regARIMA modell (regression models with ARIMA errors) i X-12-ARIMA, der man antar at de estimerte verdiene av effektene er konstante over tid. Denne antagelsen ser urimelig ut for lange tidsserier i detaljhandelen. Den beskriver ikke variasjoner i effektene når tiden går, siden mønsteret av handelen i virkeligheten har endret seg over tid. Vi har plottet tidsserien av matvarer med tall fra januar 1979 til september 2010 på figuren. Den viser at handelsmønsteret i tidsserien i de siste årene ikke er det samme som mønsteret i de tidligere årene. En av grunnene er at nå er det mange butikker som har åpent seint på kvelden på hverdager, og det er også mange butikker som har åpent på søndager. Mønsteret i åpningstider har blitt endret. Torsdagen er ikke en viktig handledag lenger sånn som for 10-15 år siden.

Vi fikser ulempene i (1) og (2) ved å selv lage dummyvariable med hensyn til røde dager og påske i Norge. Metoden er beskrevet i notatene 2007/43 og 2008/58. Ved (3) vil vi undersøke hvor store forskjeller det er for lange tidsserier, særlig i detaljhandelen når vi estimerer kalendereffekter med deterministiske modeller slik som i X-12-ARIMA og med stokastiske modeller ved tilstandsmodeller.

Young (1965) viste at kalendervariasjoner påvirker data mer enn sesongvariasjoner for importerte varer. Men de to effektene påvirker data med omtrent like stor effekt for engroshandelen. Young har presentert en enkel modell for å estimere effekter av kalenderen. Metoden er fortsatt brukt i X-11-ARIMA.

Monsell (1983). Dagum, Quenneville og Sutradhar (1992) har laget stokastiske modeller for ukedagseffekter. Monsell med random walk modell av simulerte data og Dagum et al. med differensiert modell av irregulærkomponenten. Harvey (1989) og Bell (2004) har estimert effektene med mer generelle stokastiske modeller.

I dette notatet vil vi beskrive kalendereffekter i detalj, deterministiske og stokastiske modeller for effekter, metoder for korrigerings i SSB med hensyn til røde dager og påske i Norge og et konkret eksempel.

2 Egenskaper ved kalenderen

Den gregorianske kalender er den internasjonalt mest brukte kalenderen. Et år har 12 måneder eller 365 dager. I skuddår blir det 366 dager. En måned har en lengde på 28, 29, 30 eller 31 dager. Hver ukedag (mandag, tirsdag, ..., søndag) forekommer 4 eller 5 ganger i en måned. Merk at kalenderen har en syklus på 28 år. Dermed er kalenderen for 2000 og 2028 helt identisk. Men vi får ikke samme datoer for påske og pinse. 1.påskedag i 2000 er 23.april og i 2028 er det 16.april. I følge den gregorianske kalenderen har påske en syklus på 38000 år eller 456000 måneder (side 176, X-12-ARIMA manualen, v 0.2.10).

Vi skiller mellom to variasjoner av ukedagene: (a) "length-of-month" eller "between month" variasjon. Den beskriver ulike antall dager mellom månedene (28, 29, 30 eller 31 dager). Hvis vi teller lørdag og søndag som helger blir det 23 arbeidsdager for januar 2008 og 20 dager for februar. En nedgang på ca 9% av antall arbeidsdager fra januar til februar. Variasjoner oppstår når vi vil sammenligne antall arbeidsdager i månedene. (b) "within-month" variasjon. Tabell 1 viser antall ukedager etter lengden på måneden 28, 29, 30 eller 31 dager, og hvilken ukedag som kommer først i måneden. For eksempel: en januar som starter med en mandag blir det 5 mandager, 5 tirsdager, 5 onsdager, 4 torsdager, 4 fredager, 4 lørdager og 4 søndager. Februar med 28 dager har samme antall ukedager uansett hvilken ukedag som er første i måneden. For januar blir det tre ekstra ukedager. De kommer 5 ganger. Hvilke ukedager som kommer fem ganger varierer fra år til år.

Vi vil vise "within-month" effekten ved å liste ut kalenderen for for januar 2009 og 2010.

januar 2009							januar 2010						
man.	tir.	ons.	tor.	fre.	lør.	søn.	man.	tir.	ons.	tor.	fre.	lør.	søn.
			1	2	3	4					1	2	3
5	6	7	8	9	10	11	4	5	6	7	8	9	10
12	13	14	15	16	17	18	11	12	13	14	15	16	17
19	20	21	22	23	24	25	18	19	20	21	22	23	24
26	27	28	29	30	31		25	26	27	28	29	30	31

Det blir 5 torsdager, 5 fredager og 5 lørdager i januar 2009 og det er 5 fredager, 5 lørdager og 5 søndager i januar 2010. En torsdag mindre og en søndag ekstra i januar 2010 kan medføre en lavere aktivitet i januar 2010 i forhold til januar 2009.

Påske er den viktigste av de kristne høytidene. Vi i Norge har flere fridager i påsken enn de andre landene. Disse helligidene er Skjærtorsdag, Langfredag, påskeaften, 1.påskedag (søndag) og 2.påskedag (mandag). Påsken i New Zealand starter fom Langfredag og varer tom 2.påskedag, dvs at de har en

helligdag mindre. I Nord Amerika er påsken bare en dag, 1.påskesøndag. Kristi himmelfartsdag og pinse er også bevegelige helligdager. Påske varierer mellom mars og april, mens Kristi himmelfartsdag og pinse varierer mellom mai og juni. 1.påskedag kommer tidligst 22.mars (1818 og 2285) og seinest 25.april (1943 og 2038).

Tabell 1: *Ulike typer ukedager*

Måned med en lengde på 31 dager								
Type	1.dag	#man.	#tir.	#ons.	#tor.	#fre.	#lor.	#son.
1	mandag	5	5	5	4	4	4	4
2	tirsdag	4	5	5	5	4	4	4
3	onsdag	4	4	5	5	5	4	4
4	torsdag	4	4	4	5	5	5	4
5	fredag	4	4	4	4	5	5	5
6	lørdag	5	4	4	4	4	5	5
7	søndag	5	5	4	4	4	4	5

Måned med en lengde på 30 dager								
Type	1.dag	#man.	#tir.	#ons.	#tor.	#fre.	#lor.	#son.
8	mandag	5	5	4	4	4	4	4
9	tirsdag	4	5	5	4	4	4	4
10	onsdag	4	4	5	5	4	4	4
11	torsdag	4	4	4	5	5	4	4
12	fredag	4	4	4	4	5	5	4
13	lørdag	4	4	4	4	4	5	5
14	søndag	5	4	4	4	4	4	5

Februar i skuddår								
Type	1.dag	#man.	#tir.	#ons.	#tor.	#fre.	#lor.	#son.
15	mandag	5	4	4	4	4	4	4
16	tirsdag	4	5	4	4	4	4	4
17	onsdag	4	4	5	4	4	4	4
18	torsdag	4	4	4	5	4	4	4
19	fredag	4	4	4	4	5	4	4
20	lørdag	4	4	4	4	4	5	4
21	søndag	4	4	4	4	4	4	5

Februar i ikke skuddår								
Type		#man.	#tir.	#ons.	#tor.	#fre.	#lor.	#son.
22		4	4	4	4	4	4	4

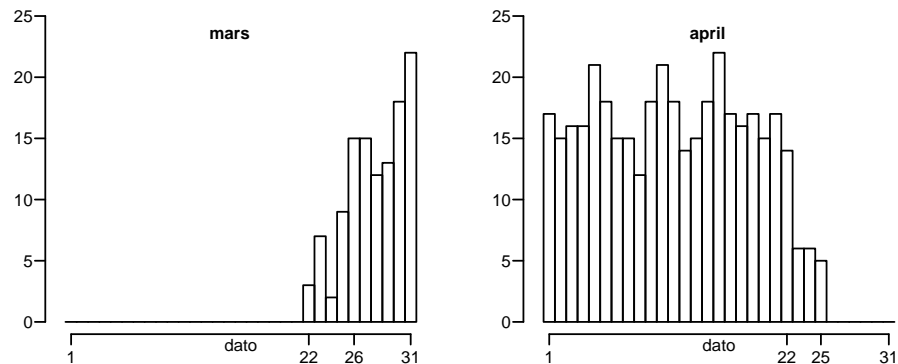
US Census har gitt ut tabeller som viser frekvenser av påske i perioden 1600-2099 av den Gregorianske kalenderen. Dette er listet ut i tabell 2. I årene 1693, 1761 og 1818 er 1.påskedag 22.mars (tidligst dato) og i årene 1666, 1734, 1886, 1943 og 2038 er 1.påskedag 25.april (seinest dato). 1.påskedag 31.mars og 16.april kommer oftest med 22 ganger, mens 1.påskedag 24.mars kommer bare 2 ganger i 1799 og 1940.

Figur 1 viser histogrammene av frekvenser for 1.påskedag i perioden 1600-2099. Den er 116 ganger i mars og 384 ganger i april. Påske kommer oftest mellom 26.mars og 22.april.

Tabell 2: *Frekvenser av påske i perioden 1600-2099*

dato	antall	%	år								
22.mars	3	0.60	1693	1761	1818						
23.mars	7	1.40	1636	1704	1788	1845	1856	1913	2008		
...								
31.mars	22	4.40	1619	1630	1641	1652	1709	1720	1771	1782	1793
			1839	1850	1861	1872	1907	1918	1929	1991	2002
			2013	2024	2086	2097					
1.april	17	3.40	1646	1657	1668	1714	1725	1736	1804	1866	1877
			1888	1923	1934	1945	1956	2018	2029	2040	
...								
25.april	5	1.00	1666	1734	1886	1943	2038				

Figur 1: *Frekvenser av antall ganger påske kommer*



Vi tar et eksempel for å vise at antall arbeidsdager kan variere mye fra måned til måned eller i samme måned fra år til år. Vi antar at mandag-lørdag er arbeidsdager i uken og at påskedagene er Skjærtorsdag, Langfredag, påskeaften, 1.påskedag og 2.påskedag. Tabell 3 viser antall ukedager i mars og april 2008 og mars 2009. 1.påskedag i 2008 er 23.mars og i 2009 er 12.april. Det er 22 arbeidsdager i mars 2008, 26 dager i april 2008 og 26 dager i mars 2009. Dermed vil det være 4 arbeidsdager mindre i mars enn april 2008 (ca 15% lavere) og 4 arbeidsdager mindre i mars 2008 enn mars 2009. Tallene viser at antall arbeidsdager kan variere mye fra måned til måned i samme år eller i samme måned fra år til år. Selv om antall arbeidsdager i april 2008 og mars 2009 er like (26 dager), er ikke antall arbeidsdager for hver ukedag i disse to månedene like. Det er en mandag mindre og en onsdag mer i april 2008 enn i mars 2009. Hvis vektene av ukedager er forskjellige fra hverandre blir det urimelig å sammenligne aktiviteten i disse to månedene uten å korrigere kalendereffekter.

Tabell 3: *Antall ukedager i mars og april i 2008 og 2009.*

måned	#man.	#tir.	#ons.	#tor.	#fre.	#lor.	#son.
mars 2008	5	4	4	4	4	5	5
april 2008	4	5	5	4	4	5	5
mars 2009	5	5	4	4	4	4	5

Vi har korrigert for kalendereffekter i mange tidsserier i SSB, blant annet i varekonsumindeksen (VKI) og produksjonsindeksen. Vi skal forklare kalendereffekter i detalj i neste avsnitt.

3 Forklaring av ukedagseffekter

Vi beskriver ukedagseffekter ved et konkret eksempel. La Y_t være omsetningen for en klesbutikk i måned t . Da er Y_t summen av N_t daglige omsetninger i måneden, der $N_t=28, 29, 30$ eller 31 dager. La t være januar 2010. Måneden starter med en fredag og har en lengde på 31 dager. Ukedagene i januar 2010 beskrives i tabell 1. Vi skriver Y_t slik:

$$\begin{aligned}
 Y_t &= Y_{4,t} + Y_{5,t} + Y_{6,t} + Y_{7,t} + Y_{8,t} + Y_{9,t} + Y_{10,t} + \\
 & Y_{11,t} + Y_{12,t} + Y_{13,t} + Y_{14,t} + Y_{15,t} + Y_{16,t} + Y_{17,t} + \\
 & Y_{18,t} + Y_{19,t} + Y_{20,t} + Y_{21,t} + Y_{22,t} + Y_{23,t} + Y_{24,t} + \\
 & Y_{25,t} + Y_{26,t} + Y_{27,t} + Y_{28,t} + Y_{29,t} + Y_{30,t} + Y_{31,t} \\
 &= \sum_{j=1}^{31} Y_{j,t}
 \end{aligned}$$

der $Y_{1,t}$ er omsetningen for dag 1 (fredag 1.januar), $Y_{2,t}$ er omsetningen for dag 2 (lørdag 2.januar), osv. Siden det er 4 mandager, 4 tirsdager, 4 onsdager, 4 torsdager, 5 fredager, 5 lørdager og 5 søndager kan vi skrive Y_t slik

$$\begin{aligned}
 Y_t &= \frac{4}{4}(Y_{4,t} + Y_{11,t} + Y_{18,t} + Y_{25,t}) + \frac{4}{4}(Y_{5,t} + Y_{12,t} + Y_{19,t} + Y_{26,t}) + \\
 & \frac{4}{4}(Y_{6,t} + Y_{13,t} + Y_{20,t} + Y_{27,t}) + \frac{4}{4}(Y_{7,t} + Y_{14,t} + Y_{21,t} + Y_{28,t}) + \\
 & \frac{5}{5}(Y_{1,t} + Y_{8,t} + Y_{15,t} + Y_{22,t} + Y_{29,t}) + \frac{5}{5}(Y_{2,t} + Y_{9,t} + Y_{16,t} + Y_{23,t} + Y_{30,t}) + \\
 & \frac{5}{5}(Y_{3,t} + Y_{10,t} + Y_{17,t} + Y_{24,t} + Y_{31,t}) \\
 &= 4\bar{Y}_{1,t} + 4\bar{Y}_{2,t} + 4\bar{Y}_{3,t} + 4\bar{Y}_{4,t} + 5\bar{Y}_{5,t} + 5\bar{Y}_{6,t} + 5\bar{Y}_{7,t}
 \end{aligned} \tag{1}$$

der $\bar{Y}_{i,t}$ er gjennomsnitt av 4 eller 5 verdier av ukedag i , i måned t . La i være indeks for 7 dager i uken. $i=1$ betegner mandag, 2 tirsdag, \dots , 7 søndag. Vi kaller $\bar{Y}_{i,t}$ daglig effekt av ukedag i og måned t .

La $N_{i,t}$ være antall dager av ukedag i , i måned t . Ligningen (1) blir

$$Y_t = \sum_{i=1}^7 N_{i,t} \bar{Y}_{i,t} \tag{2}$$

La $\bar{Y}_{0,t} = \frac{1}{7} \sum_{i=1}^7 \bar{Y}_{i,t}$ være gjennomsnitt av 7 daglige effekter. Ligningen (2) blir

$$\sum_{i=1}^7 N_{i,t} \bar{Y}_{i,t} = \sum_{i=1}^6 (N_{i,t} - N_{7,t})(\bar{Y}_{i,t} - \bar{Y}_{0,t}) + N_t \bar{Y}_{0,t} \tag{3}$$

"Trading-day effect" eller ukedagseffekter (på norsk) i måned t er definert som det første leddet av ligningen (3). Vi skriver

$$TD_t = \sum_{i=1}^6 (N_{i,t} - N_{7,t})(\bar{Y}_{i,t} - \bar{Y}_{0,t}) \tag{4}$$

La $\beta_{i,t} = \bar{Y}_{i,t} - \bar{Y}_{0,t}$, $i = 1, \dots, 6$. Denne størrelsen kalles ukedagskoeffisient og beskrives som differansen mellom den daglige effekten av ukedag i og deres gjennomsnitt. I X-12-ARIMA blir $\beta_{i,t}$ estimert ved en regARIMA modell. En positiv verdi av $\beta_{i,t}$ vil vise at effekten av ukedag i er høyere enn gjennomsnittet og en negativ verdi av $\beta_{i,t}$ viser det omvendt. Verdiene $(N_{i,t} - N_{7,t})$, $i = 1, \dots, 6$ er seks regressorer i regresjonsmodellen for å estimere $\beta_{i,t}$.

Vi får

$$\beta_{7,t} = \bar{Y}_{7,t} - \bar{Y}_{0,t} = - \sum_{i=1}^6 \beta_{i,t} \quad (5)$$

Vi ser at koeffisienten for søndag er en negativ sum av de andre ukedagskoeffisientene.

Ligningen (4) viser at størrelsen TD_t er 0 (dvs, det blir ingen ukedagseffekt) når

- $\bar{Y}_{i,t} = \bar{Y}_{0,t}$ for alle i , dvs, det er ingen forskjell mellom de daglige effektene i måned t .
- antall dager av ukedag i i en måned er lik antall søndager, dvs $N_{i,t} = N_{7,t}$, for $i = 1, \dots, 6$.
Dermed blir det ingen ukedagseffekt for februar som har 28 dager.

Differansen $(N_{i,t} - N_{7,t})$ blir 0 for februar. For de andre månedene blir differansen -1 eller 1. Den er -1 når det er 4 søndager og ukedag i har 5 dager. Den blir 1 når det er 5 søndager og ukedag i har 4 dager. I dette tilfellet blir $TD_t \neq 0$. Vi ser at antall ukedager som forekommer 5 ganger i måned skaper ukedageffekter.

Siden $\sum_{i=1}^7 \beta_{i,t} = 0$ blir summen av koeffisienter av ukedager som forekommer 5 ganger lik en negativ sum av av koeffisienter av ukedager som forekommer 4 ganger.

Størrelsen $N_t \bar{Y}_{0,t}$ beskrives som en "length-of-month" effekt. Vi behandler denne effekten på to måter:

- For måneder som ikke er februar blir effekten sugd opp av sesongvariasjoner, siden disse månedene har samme antall dager fra år til år.
- For februar lager man en dummyvariabel med 0 for en februar 28 dager og 1 for en februar med 29 dager.

Dermed betrakter vi "length-of-month"-effekten som en sesongeffekt.

4 Metoder for å sjekke og estimere kalendereffekter

Cleveland og Devlin (1980) viser at ved å bruke spektralanalysen kan man oppdage ukedageffekter i tidsserien. Anta at $x(t)$ er en månedlig tidsserie. La m betegne måned og y år. For hver m og når y varierer får vi en tidsserie $x_m(y)$ som er en delmengde av $x(t)$,

$$x_m(y) = x(12(y-1) + m)$$

Spekteret av $x(t)$ blir beregnet først. Deretter spekteret av $x_m(y)$ for hver m (unntatt februar som oppfører seg anneledes enn de andre månedene). Så anslås gjennomsnittlig spekter over 11 måneder for hver frekvens og det blir lekt etter topper på kalenderfrekvenser slik som vi gjør for $x(t)$.

4.1 Spektralanalyse for ukedageffekter

Spekteret av en tidsserie med sesongvariasjoner har høye spisser i harmoniske frekvenser $\omega_k = 2\pi k/12$, for $k = \pm 1, \dots, \pm 6$. Ved å plote spektraltettheten og sjekke mønsteret i disse frekvensene kan vi si noe om sesongvariasjoner i tidsserien. Vi kan gjøre det på samme måte for kalendereffekter. I spektraltettheten finnes det også frekvenser som er knyttet til variasjoner av kalenderen.

Skuddår kommer hvert fjerde år. Gjennomsnittet av antall dager i løpet av fire år blir 365.25 dager. Lengden for hver måned i gjennomsnitt blir $365.25/12=30.4375$ dager/måned. Siden en uke gjentar seg etter 7 dager blir $30.4375/7=4.348$ syklus/måned. Aliansen for denne frekvensen blir 0.348 syklus/måned.

Frekvensen 0.348 syklus/måned er tilsvarende med $0.348 \times 12 = 4.179$ syklus/år. En alias 0.179 syklus/år. Cleveland og Devlin (1980) har vist at frekvensen 0.432 syklus/måned også er en viktig frekvens i spektraltettheten for å vise ukedagseffekter i tidsserien. Vi skal illustrere anvendelsen av spektralanalysen for ukedagseffekter i et eksempel.

4.2 Estimering av ukedagseffekter ved X-11 og X-12-ARIMA

Estimeringen av ukedagseffekter i X-11-ARIMA er basert på den foreløpige irregulærkomponenten. Dette gjøres også med de omtrent samme modellene i X-12-ARIMA. Men vi kan estimere effektene med en regARIMA modell, der ujusterte data brukes som input.

4.2.1 Modell basert på irregulærkomponenten

Vi ser at når vi erstatter $N_{i,t}$ med $(N_{i,t} - 4)$ i ligningen (3) blir størrelsen $(N_{i,t} - N_{7,t})$ uendret. Dette medfører at det første leddet i ligningen er summen av de ukedagene som forekommer 5 ganger i måneden.

Merk at alle ukedagene (mandag-søndag) har samme gjennomsnittet av antall dager i løpet av 28 år, $(1/28) \sum_{k=1}^{28} N_{i,t+12k}$. Dermed blir summen $\sum_{k=1}^{28} (N_{i,t+12k} - N_{7,t+12k})$ lik null i en periode på 28 år. Dette medfører at det ikke er sesongvariasjoner eller trend i den størrelsen $(N_{i,t+12k} - N_{7,t+12k})$.

N_t , antall dager i måned t , har en syklus på 4 år. Gjennomsnittet $N_t^* = (1/4) \sum_{k=1}^4 N_{t+12k}$ blir N_t for måneder som ikke er februar og 28.25 for februar. Dermed skaper $\bar{y}_{0,t} N_t^*$ trend og sesongvariasjoner i ligningen (3).

I X-11 blir ukedagseffekter estimert ved regresjonsanalyse, med en antagelse om at $\bar{Y}_{i,t} = \bar{Y}_i$, for alle i , dvs at, effektene ikke endrer seg over tid. For å fjerne trend og sesongvariasjoner i ligningen (3) kan man dividere med eller trekke fra størrelsen $\bar{Y}_{0,t} N_t^*$, avhengig av hvilken måte man dekomponerer tidsserien på. Modellen som Young (1965) brukte for å estimere ukedageffekter i X-11 med "ordinary least squares (OLS)"-metoden og ved en multiplikativ dekomponering er

$$N_t^* \hat{I}_t - N_t = \sum_{i=1}^6 \beta_i (N_{i,t} - N_{7,t}) + e_t$$

der $\beta_i = (\bar{Y}_i / \bar{Y}_0) - 1$ for $i = 1, \dots, 6$ og \hat{I}_t er irregulærkomponenten etter å ha estimert trend ved det sentrerte 12-måneder glidende gjennomsnittet og sesongkomponenten ved (3×3) filteret.

Vi viser tilsvarende modeller for ukedagseffekter i X-12-ARIMA, som er basert på irregulærkomponenten I_t , av

- additiv dekomponering. Ved å trekke fra $\bar{Y}_{0,t} N_t^*$ i ligningen (3), får vi en modell for ukedagseffekter slik:

$$\hat{I}_t = \beta_0 (N_t - N_t^*) + \sum_{i=1}^6 \beta_i (N_{i,t} - N_{7,t}) + e_t$$

der $\beta_0 = \bar{Y}_0$ og $\beta_i = \bar{Y}_i - \bar{Y}_0$ for $i = 1, \dots, 6$. Det er sju parametre som skal estimeres.

- pseudo-additiv dekomponering. Ved å dividere med $\bar{Y}_{0,t}N_t^*$ i ligningen (3), får vi en modell slik:

$$\hat{I}_t = 1 + \frac{N_t - N_t^*}{\bar{N}} + \sum_{i=1}^6 \beta_i \frac{N_{i,t} - N_{7,t}}{\bar{N}} + e_t$$

der $\bar{N} = (1/48) \sum_{k=1}^{48} N_{t+k}$.

- log-additiv dekomponering. Ved å dividere med $\bar{Y}_{0,t}N_t^*$ og ta logaritmen deretter i ligningen (3), får vi en modell slik:

$$\hat{I}_t = \log\left(1 + \frac{N_t - N_t^*}{N_t^*} + \sum_{i=1}^6 \beta_i \frac{N_{i,t} - N_{7,t}}{N_t^*}\right) \approx \frac{N_t - N_t^*}{N_t^*} + \sum_{i=1}^6 \beta_i \frac{N_{i,t} - N_{7,t}}{N_t^*}$$

siden $\log(1+x) \approx x$ (Findley, 1996, page 20).

4.2.2 Modell basert på regARIMA

I X-12-ARIMA finnes det fire typer av ukedagseffekter for en "flow" tidsserie. De er TD7, TD6, TD2 og TD1 for regARIMA modell.

- TD7, ukedagseffekter med 7 koeffisienter. Modellen blir

$$\log(Y_t) = \beta_0 LY_t + \sum_{i=1}^6 \beta_i (N_{i,t} - N_{7,t}) + Z_t \quad (6)$$

der Z_t beskrives ved en ARIMA modell og kan inneholde de andre effektene som skyldes av ekstremverdier og bevegelige helligdager. $LY_t = N_t - N_t^*$. Vi får $LY_t = 0.75$ for t er februar i skuddår, $LY_t = -0.25$ for t er februar ikke i skuddår og $LY_t = 0$ ellers. Det er en koeffisient for skuddår og seks koeffisienter for ukedagseffekter i denne modellen.

- TD6, ukedagseffekter med 6 koeffisienter. Findley, et al. (1998) har vist at vi kan velge $\beta_0 = 1/28.25 \approx 0.0354$ isteden for å estimere β_0 som i ligningen (6). Det er dermed en parametre mindre. Vi får en TD6 modell for ukedagseffekter. Det blir

$$\log(Y_t) = \frac{1}{28.25} LY_t + \sum_{i=1}^6 \beta_i (N_{i,t} - N_{7,t}) + Z_t \quad (7)$$

- TD2, ukedagseffekter med 2 koeffisienter. Denne modellen er beskrevet i TRAMO/SEATS. I modellen antar man at mandag-fredag er arbeidsdager og har samme vekt, β_{M-F} , og lørdag, søndag er helger. Ligningen blir

$$\log(Y_t) = \beta_0 LY_t + \beta_{M-F} \left(\sum_{i=1}^5 N_{i,t} - \frac{5}{2} \sum_{i=6}^7 N_{i,t} \right) + Z_t \quad (8)$$

Det er to parametre som skal estimeres: β_0 og β_{M-F} .

- TD1, ukedagseffekter med 1 koeffisient. La $\beta_0 = 1/28.25$. Ligningen (8) blir

$$\log(Y_t) = \frac{1}{28.25} LY_t + \beta_{M-F} \left(\sum_{i=1}^5 N_{i,t} - \frac{5}{2} \sum_{i=6}^7 N_{i,t} \right) + Z_t \quad (9)$$

4.3 En ny korrigering av påske i SSB

I avsnitt 2.5 av notat 2007/43 har vi skrevet om en metode for å lage påskeregressorer for tidsserier i SSB. Vi antar at ujusterte data er påvirket av påskeeffekt i (i) perioden med w_1 dager som ligger før påsken med effekt α^{fpk} , (ii) påskehelligdager med effekt α^{pk} og (iii) perioden med w_2 dager etter påsken med effekt α^{epk} . La

$$\begin{aligned} I_{fpk,t} &= (1/w_1)(\text{antall av } w_1 \text{ dager som faller i måned } t) \\ I_{pk,t} &= (1/5)(\text{antall av 5 helligdager som faller i måned } t) \\ I_{epk,t} &= (1/w_2)(\text{antall av } w_2 \text{ dager som faller i måned } t) \end{aligned} \quad (10)$$

være andelene av dager i måned t i de tre periodene, t er enten mars eller april. Da blir

$$I_{fpk,mars} + I_{fpk,april} = 1, \quad I_{pk,mars} + I_{pk,april} = 1, \quad \text{og } I_{epk,mars} + I_{epk,april} = 1$$

Regressorer blir

- Perioden før påsken,

$$\begin{aligned} E_t^{fpk} &= I_{fpk,t} - \mu_{fpk} \\ &= (1/w_1)(\text{antall av } w_1 \text{ dager som faller i måned } t) - \mu_{fpk} \end{aligned} \quad (11)$$

der $\mu_{fpk} = (1/T) \sum_{t=1}^T I_{fpk,t}$, som er gjennomsnittet av $I_{fpk,t}$.

- Påskeperioden,

$$\begin{aligned} E_t^{pk} &= I_{pk,t} - \mu_{pk} \\ &= (1/5)(\text{antall av helligdager som faller i måned } t) - \mu_{pk} \end{aligned} \quad (12)$$

der $\mu_{pk} = (1/T) \sum_{t=1}^T I_{pk,t}$, som er gjennomsnittet av $I_{pk,t}$.

- Perioden etter påsken,

$$\begin{aligned} E_t^{epk} &= I_{epk,t} - \mu_{epk} \\ &= (1/w_2)(\text{antall av } w_2 \text{ dager som faller i måned } t) - \mu_{epk} \end{aligned} \quad (13)$$

der $\mu_{epk} = (1/T) \sum_{t=1}^T I_{epk,t}$, som er gjennomsnittet av $I_{epk,t}$.

5 Korrigering ukedageffekter i SSB

Proseduren for korrigering av kalendereffekter i X-12-ARIMA er basert på kalenderen i USA. Den fungerer ikke bra for mange tidsserier i SSB, særlig i detaljhandelen og produksjonsindeksen. Grunnen er at det er mange røde dager og bevegelige helligdager i Norge som blir behandlet som arbeidsdager i X-12-ARIMA. Dette medfører at tall i månedene av disse røde dagene ikke blir korrigert. For at X-12-ARIMA skal kunne brukes i landene med en annen kalender enn i USA har US Census laget tilleggsoptionene "user" og "usertype" slik at brukeren selv kan angi forklaringsvariable for røde dager i regARIMA-modellene sine. Vi kan se et lite eksempel i linjene nedenfor:

```
regression{variables=(lpyear ls1987.feb)
                user=(mandag tirsdag onsdag torsdag fredag lordag
                forpaske paske)
                usertype=(td td td td td td holiday holiday)
                file="matvarerdummy.dat"
                format="datevalue"}
```

Her angir vi selv forklaringsvariablene: *lpyear* (skuddsår), *ls1987.feb* (ekstremverdi i februar 1987), *mandag*, *tirsdag*, *onsdag*, *torsdag*, *fredag*, *lordag* som ukedagseffekter og *forpask* (før påske) *pask* (påske) som bevegelige helligdager. Tallene til disse variablene ligger i filen *matvarerdummy.dat*.

Vi korrigerer kalendereffekter i SSB på denne måten og ved modellene (6)-(9). Vi har laget programmer i R for å beregne antall ukedager $N_{i,t}$ med hensyn til røde dager.

6 Modell for tidsserie

Vi antar at tidsserien skrives med følgende modell

$$Y_t = f(X_t; \beta) + Z_t \quad (14)$$

der β en vektor av parametre, X_t en vektor av uavhengige variable som er observert i tidspunkt t og Z_t er et restledd. Hvis Z_t er tilfeldige variasjoner blir ligningen (14) en lineær eller ikke-lineær regresjonsmodell. Ofte er Z_t korrelerte. Dette medfører at tester i regresjonsanalysen er ugyldige, siden testene er basert på antagelsen at restleddene er i.i.d $N(0, \sigma^2)$. Antar at Z_t beskrives på følgende ARIMA modell

$$\phi(B)\delta(B)Z_t = \theta(B)a_t \quad (15)$$

der B er "backshift" operator, $BZ_t = Z_{t-1}$. $\phi(B)$ og $\theta(B)$ er polynomer av B , med $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ og $\theta(B) = \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$. De to polynomene har ikke felles røtter. Restleddene a_t er bare tilfeldige variasjoner.

Ved ligningene (14) og (15) får vi

$$\delta(B)Y_t = \delta(B)f(X_t; \beta) + \frac{\theta(B)}{\phi(B)}a_t \quad (16)$$

der $\delta(B) = (1 - B)^d$. Man kan gjette en verdi for d ved å plote ACF (autokorrelasjonsfunksjon) og PACF (partiell autokorrelasjonsfunksjon) av tidsserien. For en ikke-stasjonær prosess går ACF langsomt ned mot null. Men PACF har høy positiv eller negativ verdi i lag 1. Vi kan transformere en ikke-stasjonær tidsserie til en stasjonær tidsserie ved differanser, $(1 - B)Y_t$ eller $(1 - B)^2 Y_t$.

ACF og PACF kan også brukes for å finne verdi av p og q av polynomene $\phi(B)$ og $\theta(B)$. Når PACF har verdier ($\neq 0$) i lagene $k = 1, 2, \dots, p$ og den er lik null deretter, tilpasser vi dataene med en autoregressiv modell av orden p . Når ACF har verdier ($\neq 0$) i lagene $k = 1, 2, \dots, q$ og deretter er den lik null, tilpasser vi dataene med en MA modell av orden q .

Vi estimerer ϕ og θ ved maksimum likelihood estimator under antagelsen at a_t har en normalfordeling $N(0, \sigma_a^2)$. Vi estimerer $\hat{\sigma}_a^2$ ved $\hat{\sigma}_a^2 = (n - r)^{-1} \sum \hat{a}_t^2$, hvor n er antall observasjoner av tidsserien $\delta(B)Y_t$ og r er antall parametre av ligningen (16). Pierce (1971) har skrevet i en artikkel om metoden for å estimere parametre i en regresjonsmodell der feilleddene er en ARMA modell (autoregressive-moving average errors). Se også Tsay (1984).

Diagnostikk av estimering

Etter estimeringen av parametre foreslår Bell og Hillmer (1983) at vi skal sjekke om

- modellen av $f(X_t; \beta)$ er godt tilpasset. Dette gjøres ved å plote \hat{a}_t mot X_t og de andre forklaringsvariablene.
- \hat{a}_t er en hvit støy prosess. Dette gjøres ved å plote \hat{a}_t mot t .

7 Eksempel: Matvarer

Dataene er hentet fra nasjonalregnskapet, og er fom januar 1979 tom september 2010. Tallene er beregnet fra følgende tidsserier med vekter:

Tabell 4: *Matvarer og aggregering*

Tidsserie	Vekt
Butikkhandel breit vareutvalg	0,91283
Butikkhandel breit vareutvalg, ellers	0,01486
Butikk grønn og frukt	0,00324
Butikk kjøtt og kjøttvarer	0,00257
Butikk Fisk og Skaldyr	0,00379
Bakervarer, konditori	0,02225
Drikkevare	0,00013
Tobakksvare	0,00019
Nærings, nyttels ellers	0,00276
Drivstoff	0,02485
Torghandel med næringsmidler	0,00075
Postordre, internetthandel	0,00099
Utenom butikk, ellers	0,01079

Tidsserien er plottet i figur 2. Det er en klar trend og et tydelig sesongmønster. Vi ser mange lave og høye spisser som kommer med jevne mellomrom. Trenden eller sesongvariasjoner kan sjekkes ved å plotte autokorrelasjonfunksjonen (ACF) av ujusterte verdier, $\hat{\rho}_k$. Vi ser i figur 3 at $\hat{\rho}_k$ går langsomt ned mot null, dette viser trend i tidsserien, og at $\hat{\rho}_k$ er signifikante større enn null i lag 12, 24, . . . , dette viser sesongvariasjoner.

For å sjekke om ukedageffekter plotter vi spekteret av ujusterte verdier. Når det er høye spisser i frekvensene 0.348 eller 0.432 sykluser/måned og 0.179 eller 0.357 sykluser/år vil det vise at det er signifikante effekter av ukedager i tidsserien. Vi ser i figur 4 en høy tynn spiss i frekvensen 0.348. Dette viser ukedageffekter i tidsserien. Vi tolker dette slik at det er noen dager i uken som vi ofte handler matvarer.

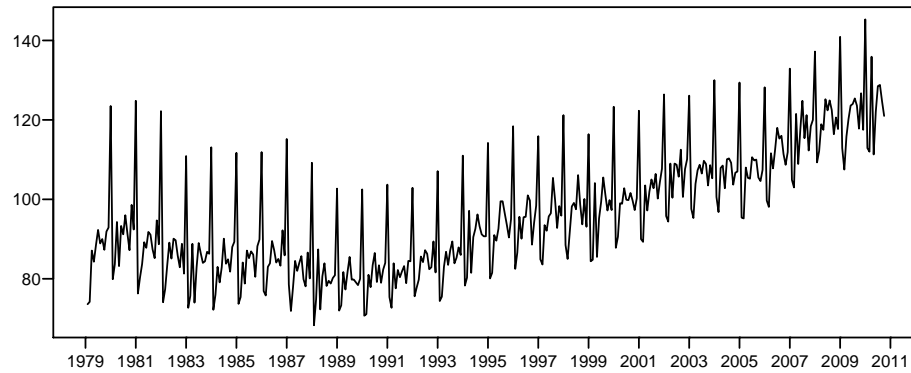
Tidsserien i eksemplet blir kraftig påvirket av sesongvariasjoner og trend. Det blir dermed vanskelig å se høye spisser i frekvenser av ukedager i spekteret siden topper av sesongvariasjoner er mye høyere enn topper av ukedageffekter. Vi bør fjerne trend og sesongvariasjoner fra tidsserien. Den enkleste måten vi kan gjøre det på, er å ta differanser av ujusterte data i lag 1 og 12. Spekteret er plottet i figur 4. Vi får en høy spiss i frekvensen av ukedager (0.348). Det finnes også en test for ukedageffekter i X-12-ARIMA.

Vi estimerer kalendereffekter for matvarer med tre alternativer:

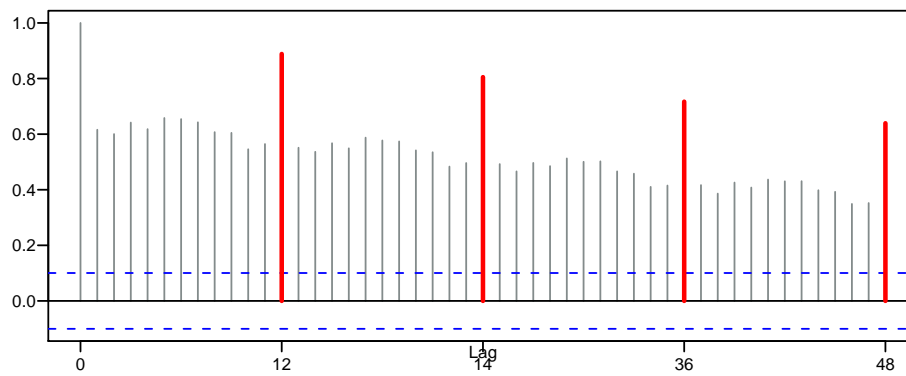
Alt 1. Vi skal korrigere kalendereffekter for norske data med kalenderen i USA. Dvs at de røde dagene i Norge; 1.mai, 17.mai, 24-26 og 31 desember er behandlet som arbeidsdager. Påske er bare en søndag (1.påskedag). 2.påskedag og 2.pinsedag er arbeidsdager. Estimeringen blir utført ved en regARIMA modell, der effektene er konstante over tid.

Alt 2. Vi skal korrigere kalendereffekter for norske data med norsk kalender. De røde dagene; 1.mai, 17.mai, 24-26 og 31 desember er helligdager. Påske er dagene: Skjærtorsdag, Langfredag, påskeaften, 1.påskedag og 2.påskedag. Pinse er dagene: 1.pinsedag og 2.pinsedag. Vi antar også at folk handler mer matvarer enn vanlig i en kort periode før påske. Handelen blir

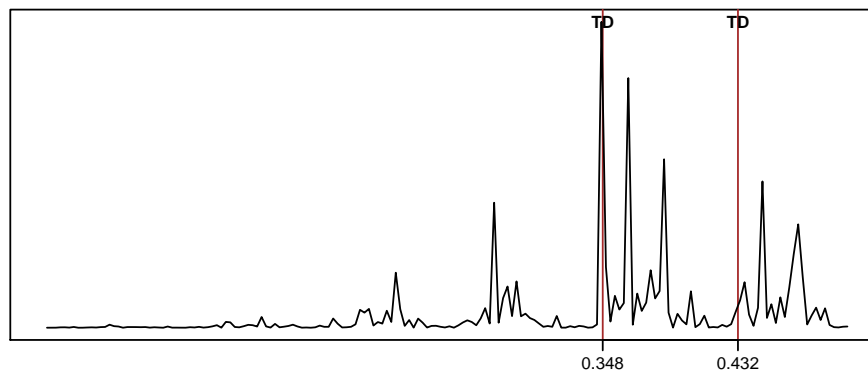
Figur 2: Ujusterte verdier



Figur 3: ACF av ujusterte verdier



Figur 4: Ukedageeffekter ved spektralanalysen



normal igjen etter påske. Vi estimerer effekter ved en regARIMA modell i X-12-ARIMA. De er konstante for hele tidsserien.

Alt 3. Vi korrigerer kalendereffekter på samme måte som i alternativ 2, bortsett fra at effektene beskrives ved stokastiske modeller heller enn deterministiske modeller slik som i alternativ 1 og 2. Her antar vi at effektene kan utvikle seg fra tidspunkt til tidspunkt. regARIMA modellen kan dermed ikke brukes i estimeringen lenger. Effektene skal beskrives ved tilstandsmodeller.

Resultatene av de tre alternativene blir beskrevet i neste avsnitter.

7.1 Alternativ 1. Korrigeringen med regARIMA og uten hensyn til røde dager og påske i Norge

Vi antar at mandag-lørdag er arbeidsdager. Vi har tatt med effektene av ukedager, påske, pinse og ekstremverdier i modellen i første omgang. Men pinsen har en lav t -verdi. Kristi himmelfartsdag varierer fra mai til juni, men ofte i mai. I løpet av 300 år fra 1800 til 2099 faller Kristi himmelfartsdag i juni i de følgende årene: 1848 (1.juni), 1859 (2.juni), 1886 (3.juni), 1905 (1.juni), 1916 (1.juni), 1943 (3.juni), 1954 (3.juni), 2000 (1.juni), 2011 (2.juni), 2038 (3.juni), 2049 (3.juni), 2079 (1.juni) og 2095 (2.juni). Det er bare 13 av 300 år. Vi vil dermed heller behandle Kristi himmelfartsdag som sesongvariasjoner enn en bevegelig helligdag. I estimeringen vil vi ikke ta hensyn til pinse og Kristi himmelfartsdag.

Tabell 5: *Estimering av kalendereffekter ved alternativ 1.*

Parametre	Est.verdi	Std.	t -verdi
mandag	-0.0057	0.0029	-1.95
tirsdag	0.0022	0.0029	0.73
onsdag	0.0024	0.0029	0.81
torsdag	0.0061	0.0029	2.06
fredag	0.0071	0.0029	2.39
lørdag	0.0072	0.0029	2.44
Easter[15]	0.0646	0.0066	9.70
LS1987.feb	-0.0586	0.0159	-3.67

ARIMA(0,1,1)(0,1,1) blir valgt i estimeringen. Ligningen blir

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

Vi får $\hat{\theta} = 0.7619$ med standardavvik 0.0327 og $\hat{\Theta} = 0.6950$ med standardavvik 0.0379. Tabell 5 viser de estimerte verdiene av koeffisientene til ukedager, påske og ekstremverdier. Mandag har en negativ koeffisient. Tirsdag og onsdag har positive koeffisienter, men tallene er ikke signifikante forskjellige fra null på 5%-nivå. Torsdag, fredag og lørdag har positive verdier som er signifikant forskjellige fra null. Vi tolker ukedagseffektene slik at vi handler matvarer mest i slutten av uken og minst på mandag. Dette ser rimelig ut i forhold til virkeligheten.

I tabellen er det en parameter som heter "Easter[15]". Det er en test av påskeeffekt. Vi får en høy positiv t -verdi ($t = 9.70$). Tallet viser at matvarehandelen øker mer enn vanlig i to uker før påskesøndag (dvs uken før påskeuken og uken i påske). Dette ser urimelig ut i forhold til virkeligheten i Norge, siden vi ofte handler mer matvarer enn vanlig bare noen dager før påske og kjøper mye mindre i påske (nesten alle butikker er stengt i påsken i Norge).

Tolkningen av effektene av ukedager og påske gjelder for hele tidsserien.

Det er et brudd i tidsserien. Tallene før og etter februar 1987 ligger på to forskjellige nivå.

7.2 Alternativ 2. Korrigeringen med regARIMA og med hensyn til røde dager og påske i Norge

I dette alternativet vil vi korrigere ujusterte data for kalendereffekter med de røde dagene; 1.januar, 1.mai, 17.mai, 24-26 og 31.desember og påskedagene Skjærtorsdag, Langfredag, påskeaften, 1.påske-

dag og 2.påskedag. Vi har laget programmer for å beregne antall dager av ukedag i i måned t , $N_{i,t}$ og påskeregressorer E_t^{fpk} , E_t^{pk} og E_t^{epk} i ligningene (10). I dette eksemplet antar vi at handelen av matvarer på mandag, tirsdag og onsdag før Skjærtorsdag blir mye høyere enn vanlig (på grunn av at vi handler matvarer for påske). Vi velger $w_1 = 3$. Handelen blir normal igjen etter påske.

Tabell 6: *Estimering av kalendereffekter ved alternativ 1 og 2.*

Parametre	Alternativ 1			Parametre	Alternativ 2		
	Est.verdi	Std.	t -verdi		Est.verdi	Std.	t -verdi
mandag	-0.0057	0.0029	-1.95	mandag	-0.0013	0.0020	-0.63
tirsdag	0.0022	0.0029	0.73	tirsdag	-0.0016	0.0020	-0.75
onsdag	0.0024	0.0029	0.81	onsdag	0.0018	0.0020	0.85
torsdag	0.0061	0.0029	2.06	torsdag	0.0057	0.0020	2.74
fredag	0.0071	0.0029	2.39	fredag	0.0085	0.0020	4.09
lørdag	0.0072	0.0029	2.44	lørdag	0.0056	0.0020	2.77
Easter[15]	0.0646	0.0066	9.70	før påske	0.1159	0.0057	20.07
LS1987.feb	-0.0586	0.0159	-3.67	påske	-0.0909	0.0070	-12.90
				LS1987.feb	-0.0677	0.0149	-4.52

ARIMA(0,1,1)(0,1,1) blir valgt i estimeringen. Vi får $\hat{\theta} = 0.6482$ med standardavvik 0.0390 og $\hat{\theta} = 0.5795$ med standardavvik 0.0418. For å gjøre det lettere å sammenligne resultater ved to alternativer 1 og 2 lister vi dem ut ved siden av hverandre i tabell 6.

Beregningene fra de to alternativene 1 og 2 viser at aktiviteten i handelen øker mot slutten av uken, eller m.a.o at vi pleier å handle matvarer på torsdag, fredag og lørdag, og kjøper mye mindre på de andre dagene. Ved alternativ 1 blir koeffisientene for fredag og lørdag like store og de er litt større enn koeffisienten av torsdag. Ved alternativet 2 er koeffisienten av fredag mye høyere enn torsdag og lørdag.

Vi deler effekten av påske i to perioder: en periode med 3 dager før påske og en periode med påskedagene. Vi får fortsatt to fortegn på koeffisientene 0.1159 og -0.0909 . Tallene beskriver aktivitetene i to perioder. Vi tolker dette slik at folk ofte handler mer matvarer enn vanlig i en kort periode før påske og mindre i påskedagene. Deretter blir handelen normal igjen. Tolkningen ser rimelig ut i forhold til virkeligheten.

Det er et brudd i tidsserien i februar 1987. Tall før og etter februar 1987 ligger på to forskjellige nivåer.

Tolkningene i alternativ 1 og 2 gjelder for hele tidsserien, siden regARIMA-modellen i X-12-ARIMA blir brukt bare for determiniske modeller.

7.3 Alternativ 3. Korrigeringen ikke med regARIMA modell, men med hensyn til røde dager og påske i Norge

Vi gjør det litt annelerdes her enn i de to alternativene 1 og 2. Estimeringen av effektene skal utføres med stokastiske modeller heller enn deterministiske modeller. Vi skal ta hensyn til de røde dagene; 1.januar, 1.mai, 17.mai, 24-26 og 31.desember og påske i Norge slik som vi gjorde i alternativ 2.

Vi skriver en modell for ujusterte data

$$\log(Y_t) = f(X_t; \Xi) + Z_t, \text{ der} \quad (17a)$$

$$f(X_t, \Xi) = \frac{1}{28.25} LY_t + \sum_{i=1}^6 (N_{i,t} - N_{7,t}) \beta_{i,t} + E_t^{fpk} \gamma_{1,t} + E_t^{pk} \gamma_{2,t} + LS_t^{feb87} \lambda_t \quad (17b)$$

$\Xi = (\beta_1, \dots, \beta_6, \gamma_1, \gamma_2, \lambda)$, er henholdvis koeffisientene av ukedagene, påsken og nivåskiftet i februar 1987, Disse parametrene beskrives ved følgende modeller

$$\beta_{i,t} = \beta_{i,t-1} + \epsilon_{i,t}, \quad i = 1, \dots, 6 \quad (18a)$$

$$\gamma_{i,t} = \gamma_{i,t-1} + \zeta_{i,t}, \quad i = 1, 2 \quad (18b)$$

$$\lambda_t = \lambda_{t-1} \quad (18c)$$

Koeffisientene $\beta_{i,t}$, $\gamma_{i,t}$ og λ_t estimeres ved Kalmanfilterteknikken.

7.3.1 Kalmanfilterteknikk

Vi innfører tilstandsmodeller. De består av to ligninger:

$$Y_t = F_t \alpha_t + \nu_t \quad \text{Måleligning} \quad (19a)$$

$$\alpha_t = G_t \alpha_{t-1} + \delta_t \quad \text{Transisjonsligning} \quad (19b)$$

der

Y_t responsvariabel (endogen variabel).

F_t vektor av koeffisienter eller eksogene variable.

α_t vektor av tilstandsvariable.

G_t matrisen som spesifiserer dynamikken i den underliggende prosessen.

ν_t og δ_t , uavhengige restledd. ν_t i.i.d $N(0, \sigma_\nu^2)$ og δ_t i.i.d $N(0, \sigma_\delta^2)$.

Måleligningen knytter sammenhengen mellom de målbare og tilstandsvariablene, mens transisjonsligningen styrer dynamikken i tilstandsvariablene (Lars, Interne notater 87/46).

Eksempel 1. En enkel regresjonsmodell

$$Y_t = x_t \beta + \nu_t. \text{ der } \nu_t \text{ i.i.d } N(0, \sigma_\nu^2).$$

kan skrives som tilstandsmodell slik:

$$\begin{aligned} Y_t &= x_t \beta + \nu_t \\ \beta &= \beta \end{aligned}$$

Vi ser at F_t i ligningen (19) blir x_t , som er vektoren av forklaringsvariable (observasjoner i tidspunkt t). α_t blir β , som er vektoren av regresjonsparametre (tilstandsvariable). G_t blir identitetsmatrisen.

Eksempel 2. En regresjonsmodell med restledd beskrives ved en ARMA(p, q) modell,

$$\begin{aligned} Y_t &= x_t \beta + \nu_t, \text{ der} \\ \nu_t &= \phi_1 \nu_{t-1} + \dots + \phi_p \nu_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \end{aligned} \quad (20)$$

ϵ_t en hvit støy prosess.

Antar at $p = 2$ og $q = 2$. Ligningen (20) skrives som tilstandsmodell slik:

$$Y_t = (x_t \ 1 \ 0 \ 0 \ 0) \alpha_t$$

og

$$\alpha_t = \begin{pmatrix} \beta \\ \nu_t \\ \nu_{t-1} \\ \epsilon_t \\ \epsilon_{t-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \phi_1 & \phi_2 & \theta_1 & \theta_2 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \beta \\ \nu_{t-1} \\ \nu_{t-2} \\ \epsilon_{t-1} \\ \epsilon_{t-2} \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \epsilon_t \quad (21)$$

7.3.2 Tilstandsmodellene for matvarer

Det kreves en ARIMA modell for Z_t i ligningen (17a) for estimering. Vi bruker den samme ARIMA(0,1,1)(0,1,1) i alternativ 2 her, siden alternativ 3 er alternativ 2 bortsett fra at koeffisientene av ukedager og påske blir estimert ved stokastiske modeller heller enn deterministiske modeller. Da blir

$$\log(Y_t) = f(X_t; \Xi) + \frac{(1 - \theta B)(1 - \Theta B^{12})}{\nabla \nabla_{12}} a_t, \text{ der } a_t \text{ i.i.d } N(0, \sigma_a^2) \quad (22)$$

og $f(X_t; \Xi)$ som har beskrevet i ligningen (17b). Ligningen (22) blir

$$\nabla \nabla_{12} \left(\log(Y_t) - \frac{1}{28.25} LY_t \right) = \nabla \nabla_{12} \left(\sum_{i=1}^6 (N_{i,t} - N_{7,t}) \beta_{i,t} + E_t^{fpk} \gamma_{1,t} + E_t^{pk} \gamma_{2,t} + LS_t^{feb87} \lambda_t \right) + (1 - \theta B)(1 - \Theta B^{12}) a_t$$

La $w_t = (1 - \theta B)(1 - \Theta B^{12}) a_t$. Da blir α_t og G_t i ligningene (19),

$$\alpha_t = \left(\beta_{1,t}, \dots, \beta_{6,t}, \gamma_{1,t}, \gamma_{2,t}, \lambda_t, w_t, a_t, \dots, a_{t-12} \right)^T \quad (23)$$

og

$$G_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta & \theta \\ 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad (24)$$

Ligningen (22) er måleligningen og ligningene (23) og (24) er transisjonsligningen for å estimere kalendereffekter av matvarer.

Algoritmen av Kalmanfilteret kan finnes i mange lærebøker: "Time Series Analysis by State Space Methods" av Durbin og Koopman, (2000), eller "Forecasting, Structural Times Series Models and Kalman filter" av Harvey (1989).

Vi får i alternativ 2 en estimert verdi $\hat{\beta}_i^{alt2}$ for koeffisienten av ukedag i , $i = 1, \dots, 6$ og estimerte verdier $\hat{\gamma}_1^{alt2}$ og $\hat{\gamma}_2^{alt2}$ for påske. I tillegg en $\hat{\lambda}^{alt2}$ for ekstremverdien i februar 1987. Disse verdiene er konstante over tid. Når t varierer kan vi betrakte dem som er tidsserier der alle elementer er i tidsserien.

Ved alternativet 3, for hvert tidspunkt t , får vi en estimert verdi $\hat{\beta}_{i,t}^{alt3}$, for ukedag i , $i = 1, \dots, 6$ og $\hat{\gamma}_{i,t}^{alt3}$, $i = 1, 2$, for påske. Tidsserien ($\hat{\beta}_{i,t}^{alt3}$) viser mønsteret av handelen på ukedag i , når t varierer. Vi får også et bilde av påskeeffekten ved å plote tidsserien ($\hat{\gamma}_{i,t}^{alt3}$) mot tid.

Tabell 7: Estimering av kalendereffekter ved alternativ 2 og 3

Parametre	Alternativ 2			Parametre	Alternativ 3		
	Est.verdi	Std.	t -verdi		gj.snitt	snitt	Std.
mandag	-0.0013	0.0020	-0.63	mandag	-0.0017	0.0022	
tirsdag	-0.0016	0.0020	-0.75	tirsdag	-0.0024	0.0032	
onsdag	0.0018	0.0020	0.85	onsdag	0.0022	0.0023	
torsdag	0.0057	0.0020	2.74	torsdag	0.0057	0.0024	
fredag	0.0085	0.0020	4.09	fredag	0.0100	0.0023	
lørdag	0.0056	0.0020	2.77	lørdag	0.0053	0.0040	
før påske	0.1159	0.0057	20.07	før påske	0.1220	0.0048	
påske	-0.0909	0.0070	-12.90	påske	-0.1116	0.0059	
LS1987.feb	-0.0677	0.0149	-4.52	LS1987.feb	-0.0840	0.0004	

Vi sammenligner $\hat{\beta}_i^{alt2}$ med $\hat{\beta}_{i,t}^{alt3}$ og $\hat{\gamma}_i^{alt2}$ med $\hat{\gamma}_{i,t}^{alt3}$ ved å plote dem i figurer (figur 5 og 6). I tabellen vil vi sammenligne $\hat{\beta}_i^{alt2}$ mot $\bar{\beta}_i$ som er gjennomsnittet av $\hat{\beta}_{i,t}^{alt3}$, og $\hat{\gamma}_i^{alt2}$ mot $\bar{\gamma}_i$ som er gjennomsnittet av $\hat{\gamma}_{j,t}^{alt3}$. Vi ser i tabell 7 at $\hat{\beta}_i^{alt2} \approx \bar{\beta}_i$ og $\hat{\gamma}_i^{alt2} \approx \bar{\gamma}_i$. Dette vil si at de koeffisientene som blir estimert ved deterministiske modeller ligger omtrent i midten av koeffisientene ved stokastiske modeller. Vi får de samme tolkningene for ukedagseffekter og påske når vi estimerer dem ved deterministiske eller stokastiske modeller. Vi handler matvarer mest i slutten av uken og vi handler mer matvarer før påske og mye mindre i påsken.

Figuren 5 og 6 viser koeffisientene av ukedagene ved deterministiske og stokastiske modeller. De som blir estimert ved deterministisk modell er plottet med en rett linje. Koeffisientene av mandag og onsdag har lave verdier og varierer rundt null-aksen. Tirsdag har større variasjoner enn mandag og onsdag. Figuren 6 viser at tirsdag har blitt en stor handledag i uken de siste årene. Torsdag, fredag og lørdag er viktige handledager i uken. Mønstrene av deres koeffisienter i figurene viser at vi handler litt mindre på torsdag og mer på lørdag nå enn før. Handelen på fredag er omtrent samme som før.

8 Oppsummering

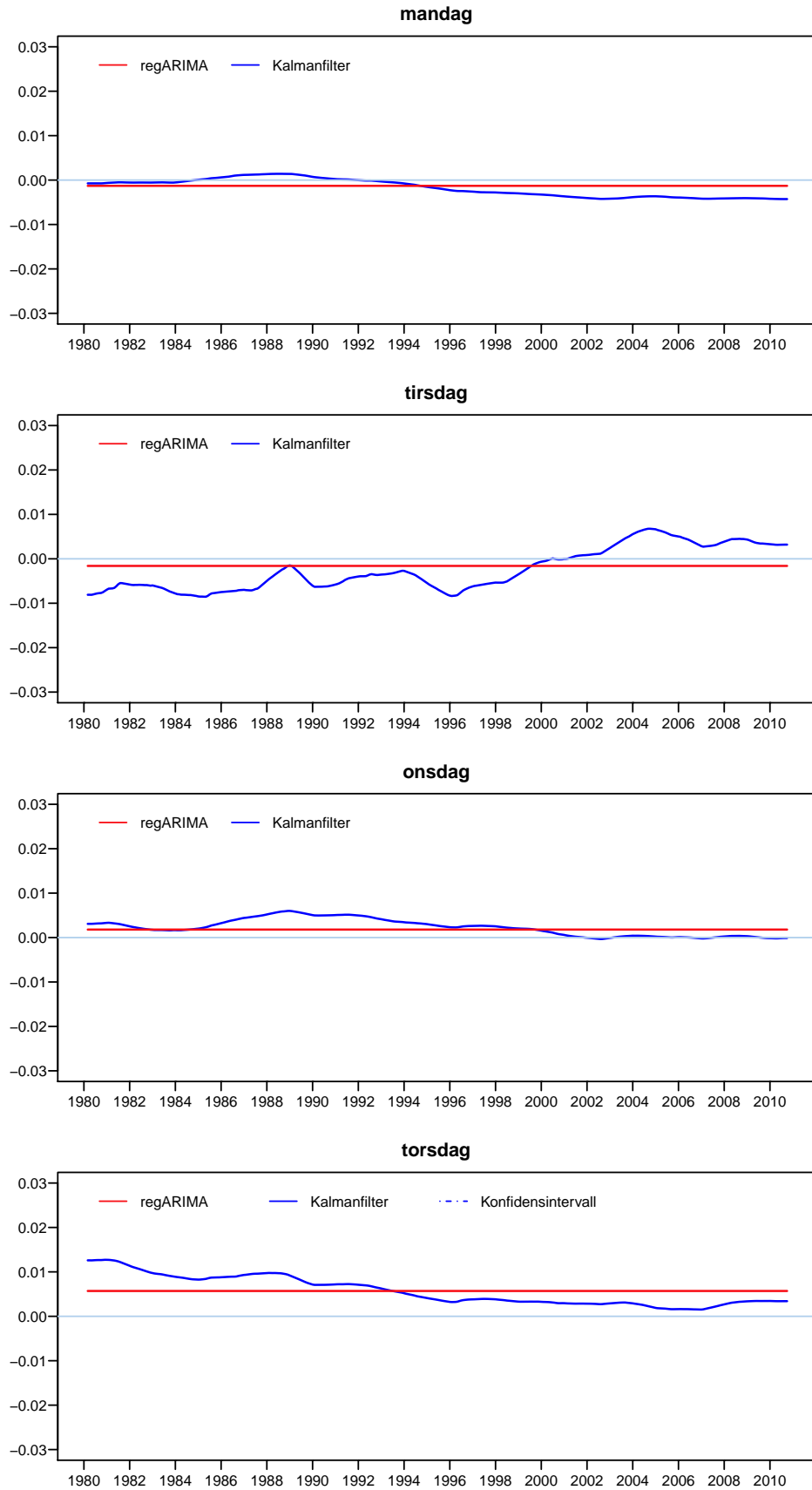
Vi har blitt vant å korrigere kalendereffekter med metoden i X-12-ARIMA for norske data, men for mange tidsserier får vi ikke tilfredsstillende resultater. Et eksempel er matvarer som vi har sett.

Dette skyldes at estimeringen ikke er basert på den norske kalenderen, der 17.mai ikke er en rød dag og påske er bare en søndag i X-12-ARIMA. Vi gjør det bedre ved å lage nye dummyvariable av effekter med hensyn til røde dager og påske i Norge. Påskeeffekten blir behandlet med forskjellige perioder; før påske, påske og etter påske. I estimeringen bruker vi fortsatt regARIMA-modellen i X-12-ARIMA. Resultatene blir det mye bedre i sammenligning med estimeringen som er basert på den amerikanske kalenderen. Vi får en estimert verdi for hver effekt. Verdien gjelder for alle tidspunkter, siden regARIMA er en vanlig regresjonsmodell der koeffisienter er konstante over tid. Så lenge vi bruker X-12-ARIMA vil effekten ha en fast verdi.

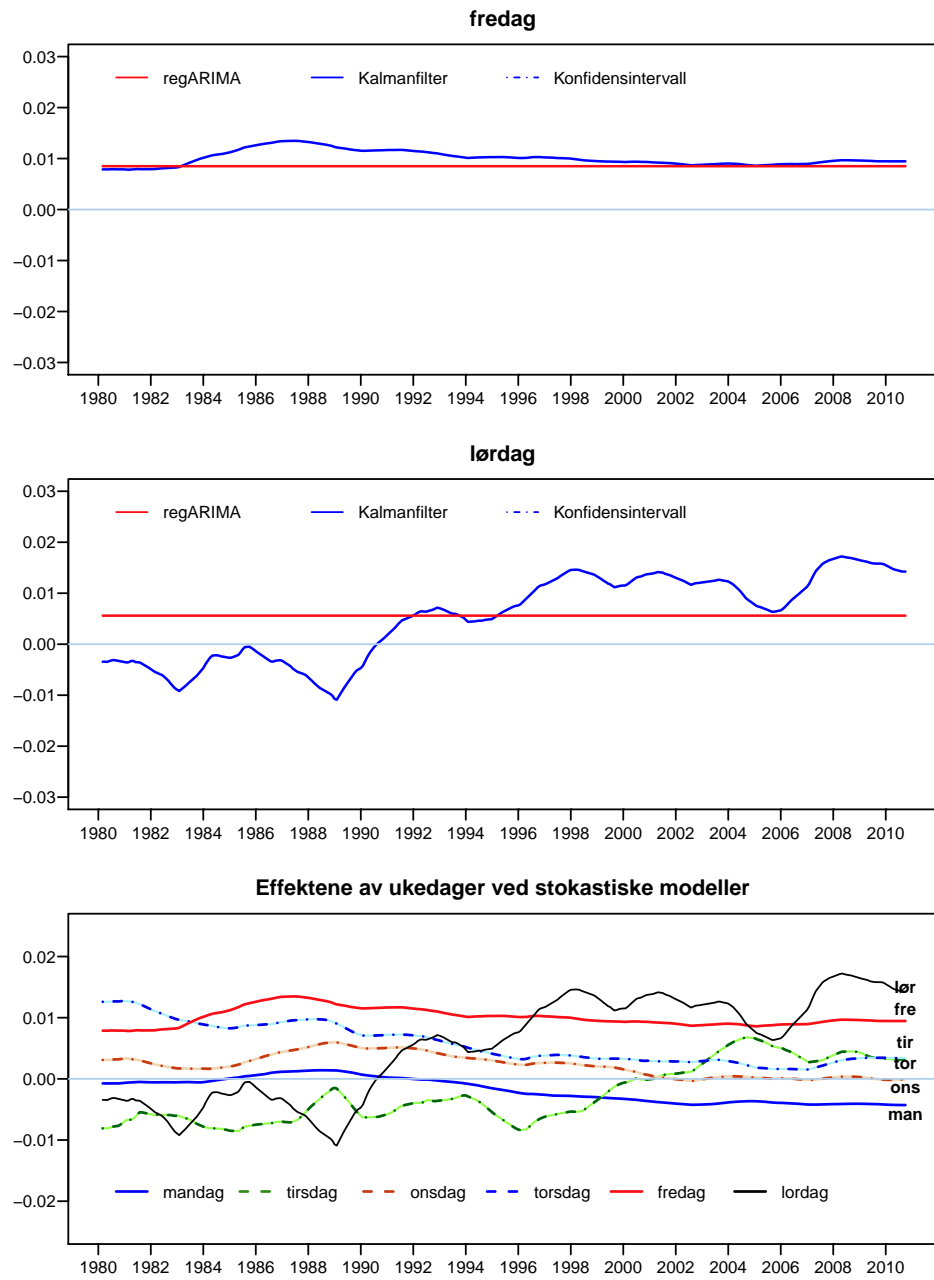
For en lang tidsserie vil det være rimelig å anta at effektene endrer seg over tid. Dette ser fornuftig ut i virkeligheten. For eksempel for matvarer er handlemønsteret i 1979 ikke det samme som i 2010, siden det er mange butikker som holder åpent seint om kvelden alle ukedager. Effektene blir estimert med stokastiske modeller. Vi får estimerte verdier for hvert tidspunkt. Mønstrene deres viser handelen gjennom årene. Vi fant at torsdagen ikke er en viktig handledag i uken lenger de siste årene og det er en liten endring av hendelsmønsteret på fredag gjennom årene.

Kalendereffekter av matvarer blir estimert med regARIMA og Kalmanfilter. Selv om den ene metoden estimerer effekter med deterministiske modeller og stokastiske modeller med den andre metoden får vi ganske like estimerte verdier.

Figur 5: Effektene av mandag-torsdag ved alternativ 2 og 3



Figur 6: Effektene av fredag og lørdag ved alternativ 2 og 3



Referanser

- [1] Bell W. R. og Hillmer S. C. (1983): *Modelling Time Series with Calendar Variation*, Journal of the American Statistical Association, September 1983, Vol. 78.
- [2] Bell, William R. (2004): *On RegComponent Time Series Models and Their Applications*, in State Space and Unobserved Component Models: Theory and Applications, eds. Andrew C. Harvey, Siem Jan Koopman, and Neil Shephard, Cambridge, UK: Cambridge University Press, forthcoming.
- [3] Bureau of the Census: *X-12 ARIMA Reference Manual, Version 0.2.10, July 26, 2002*.
- [4] Cleveland, W. S. and S. J. Devlin (1980): *Calendar Effects in Monthly Time Series: Detection by Spectrum Analysis and Graphical Methods*, Journal of the American Statistical Association, 75, pp. 487-496.
- [5] Dagum, E. B., Quenneville, B. and Sutradhar, B. (1992): *Trading-Day Variations Multiple Regression Models with Random Parameters*, International Statistical Review, 60, 57-73.
- [6] Durbin, J. og Koopman, S. J. (2000): *Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives*, J. R. Statist. Soc. B 62, Part 1, pp 3-56.
- [7] Findley, D. F. Monsell, B. C., Bell, W. R., Otto, M. C. og Chen B. C. (1998): *New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program*. Journal of Business and Economic Statistics, 16, 127-177.
- [8] Harvey A. C. (1989); *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge, U. K.: Cambridge University Press.
- [9] Harvey, A. and Streibel, M. (1998): *Tests for deterministic versus indeterministic cycles*, Journal of Time Series Analysis, 19(5), 505-529.
- [10] Leiv Solheim og Dinh Quang Pham (1997): *Prekorrigering av påskeeffekten for detaljvolumindeksen 1979-1997*. Notater 73/97.
- [11] Monsell, B. C. (1983): *Using the Kalman Smoother to Adjust for Moving Trading Day*, Research Report 83/04, Statistical Research Division, U.S. Census Bureau.
- [12] Pierce, D.A. (1971a), *Least Squares Estimation in the Regression Model With Autoregressive-Moving Average Errors*, Biometrika, 58, 299-312.
- [13] Pham, Q. D. (2007) *Ny metode for påskekorrigerig*, Notater 2007/43.
- [14] Tsay, R. S. (1984): *Regression Models with Time Series Errors*, Journal of the American Statistical Association, Vol. 79, No. 385 (Mar., 1984), pp. 118-124.
- [15] William S. Cleveland og Susan J. Devlin (1982): *Calendar Effects in Monthly Time Series: Modeling and Adjustment*. Journal of the American Statistical Association, Vol. 77, No. 379 (Sep., 1982), pp. 520-528.
- [16] Young, A.H (1965): *Estimating Trading Day Variation in Monthly Economic Time Series*, Technical Paper 12, U.S. Department of Commerce, Bureau of the Census, Washington, D.C.

B Returadresse:
Statistisk sentralbyrå
NO-2225 Kongsvinger

Statistisk sentralbyrå

Oslo:

Postboks 8131 Dep
NO-0033 Oslo
Telefon: 21 09 00 00
Telefaks: 21 09 49 73

Kongsvinger:

NO-2225 Kongsvinger
Telefon: 62 88 50 00
Telefaks: 62 88 50 30

E-post: ssb@ssb.no
Internett: www.ssb.no

ISBN 978-82-537-8433-5 (trykt)
ISBN 978-82-537-8434-2 (elektronisk)
ISSN 1891-5906

ISBN 978-82-537-8433-5



9 788253 784335

