

**Om muligheten for å predikere norsk inflasjon
ved hjelp av ARIMA-modeller**

av

Kjell-Arild Rein



Hovedfagsoppgave i samfunnsøkonomi

Våren 2001

Institutt for økonomi

Universitetet i Bergen

1. INNLEDNING.....	1
2. LITTERATUR.....	3
2.1 Terminologi	3
2.2 Makroøkonomiske modeller	4
2.3 Tidligere studier av inflasjon og prognoser i Norge	7
3. METODE.....	9
3.1 Stokastiske tidsserier	9
3.1.1 Autokorrelasjonsfunksjonen av en <i>hvit-støy</i> -prosess	9
3.1.2 Stasjonære og ikke-stasjonære tidsserier	10
3.2 ARIMA-modellen	11
3.2.1 Glidende gjennomsnittsmoeller	11
3.2.2 Egenskaper ved glidende gjennomsnittsmoeller	12
3.2.3 Autoregressive moeller	15
3.2.4 Egenskaper ved autoregressive moeller	16
3.2.5 Blandede autoregressive – glidende gjennomsnittsmoeller	21
3.2.6 Egenskaper ved ARMA(p,q)-moeller	22
3.2.7 ARIMA-modeller	24
3.2.8 ARIMA-modeller med sesong	25
3.2.9 Intervensjonsanalyse	27
4. BESKRIVELSE AV DATAMATERIALE.....	29
5. ARIMA - prognoser i praksis.....	32
5.1 Innsamling og undersøkelse av data	33
5.2 Test av stasjonaritetsegenskaper	36
5.3 Modellidentifisering og -estimering	39
5.3.1 Box-Jenkins-prosedyren	39
5.3.2 En ren AR(p)-prosess	39
5.3.3 En ren MA(q)-prosess	40
5.3.4 Blandede ARMA-prosesser	40

5.4 Diagnostisk kontroll av modellen	44
5.5 Prognoser og evaluering av prognoser	46
5.5.1 Beregning av prognoser basert på ARIMA-modeller	48
5.5.2 Prognoseevaluering	49
5.5.3 Prognoser for året 2001	54
6. AVSLUTTNING.....	56
REFERANSER.....	58
APPENDIKS.....	61
Appendiks 1: De estimerte modellene	61
Appendiks 2: Evaluering av prognoseegenskaper	69
Appendiks 3: Prognose for 2001	77

1. INNLEDNING

Hovedmålet med denne oppgaven er å undersøke muligheten for å modellere norsk inflasjon ved hjelp av ARIMA-modeller. Med inflasjon menes en kontinuerlig økning i det generelle prisnivået. Prisenivået vil i denne oppgaven bli representert ved konsumprisindeksen (KPI).

29. mars 2001 innførte Regjeringen formelt inflasjonsmål som mandat for den norske pengepolitikken, med et operativt inflasjonsmål på 2,5 prosent. Norges Bank har imidlertid tolket sitt tidligere mandat om valutakursstabilitet som et inflasjonsmål siden Svein Gjerdrem overtok stillingen som sentralbanksjef i januar 1999.

De siste årene har inflasjon som fenomen gradvis blitt tillagt større vekt både i økonomiske og politiske kretser. Den underliggende årsaken til dette er en bred enighet om at inflasjon er skadelig for et lands økonomi. Denne utviklingen har også vært lett å spore i Norge. I diskusjonen omkring statsbudsjettet og rentepolitikken er inflasjonsutviklingen ett av de argumenter som oftest trekkes frem, og den siste tidens renteøkninger er alle blitt begrunnet med et ønske om å dempe inflasjonspresset i økonomien. Det synes derfor som om det er en bred enighet om at man må bruke penge- og finanspolitikken til å stabilisere inflasjonen på et lavt nivå. Denne diskusjonen har medført at Norge nå slår følge med en rekke andre land¹ og innfører inflasjonsmål som mandat for pengepolitikken.

For å kunne påvirke inflasjonsutviklingen på en effektiv måte er det viktig at man kan lage gode prognoser for fremtidig inflasjon, slik at man skal kunne iverksette de riktige tiltakene til riktig tid. Det er derfor viktig at man prøver ut flere måter å modellere inflasjon på. Erfaringer fra andre land viser at man med gode resultater kan gjøre dette ved hjelp av ARIMA-tidsseriemodeller. Jeg vil derfor i denne oppgaven vurdere mulighetene for å gjøre dette i Norge.

Inndelingen av oppgaven er som følger: I kapittel 2 vil jeg kort gjøre rede for den faglige diskusjonen omkring prognoser og prognosemodeller. Det vil her bli lagt vekt på hvorvidt modellene først og fremst bør være basert på økonomisk teori eller statistisk analyse.

¹ Australia, Canada, Israel, Mexico, New Zealand, Sverige og Storbritannia.

Kapittelet avsluttes med en gjennomgang av de modellene man har brukt for å predikere inflasjon i Norge. I kapittel 3 blir det gjort rede for ARIMA-tidsseriemodeller, og i kapittel 4 vil jeg gjøre nærmere rede for konsumprisindeksen. I kapittel 5 utføres en praktisk analyse av konsumprisindeksen ved hjelp av ARIMA-modeller. De økonometriske testene er utført i EDB programmet Eviews.

2. LITTERATUR

2.1 Terminologi

Statistiske modeller for å beskrive variasjon i en enkelt variabel deles vanligvis inn i to hovedtyper. Disse to typene av modeller kommer til å stå sentralt i diskusjonen som følger.

Den ene typen er *flerdimensjonale (reduisert form) modeller* hvor variasjon i en variabel forklares ved endringer i en eller flere andre variabler. Det vil si at endringer i den *avhengige variabelen* forklares ved hjelp av en eller flere *uavhengige variabler*. Et klassisk eksempel er å forklare husholdningens forbruk i periode t (c_t) som en lineær funksjon av inntekt i samme periode (y_t):

$$c_t = \alpha + \beta y_t + \varepsilon_t$$

Forbruket i periode t er avhengig av inntekt i samme periode, her spesifisert ved en lineær funksjon, hvor α og β er *parametre*. Modellbyggerens oppgave er, på basis av empirisk data for c_t og y_t , å estimere parametrene. Sammenhengen mellom modellens variabler og parametre kan i praksis ikke forventes å være helt presis, noe som uttrykkes ved *restleddet* ε_t . Dette tas med for å fange opp relasjoner som ikke eksplisitt fremgår av den spesifiserte modellen.

Den andre typen er *endimensjonale (univariate) modeller* hvor verdien av en variabel i en periode forklares ved observasjoner av samme variabel i en eller flere foregående perioder, det vil si tidsforsinkede verdier av den samme variabelen. Eksempelvis kan man modellere husholdningens forbruk i periode t som funksjon av foregående periode $t-1$, ved funksjonen:

$$c_t = \alpha + \beta c_{t-1} + \varepsilon_t$$

Modellen har samme lineære struktur som i modellen på redusert form ovenfor, men inntekt erstattes av forbruk i foregående periode (tidsforsinket en periode). Denne modellen forsøker imidlertid ikke å gi noen uttømmende forklaring på hvorfor husholdningens forbruk utvikler

seg som den gjør. Relasjonen er mer et uttrykk for et postulat av typen ”forbrukerne gjør noe i dag fordi man også gjorde det i går”.

2.2 Makroøkonomiske modeller

Makroøkonomiske prognoser har lenge vært en viktig del av økonomisk forskning, men det var først på 60-tallet at dette ble en industri. Modellene man brukte på denne tiden var store makroøkonomiske modeller som tok sikte på å beskrive en nasjons økonomi. Disse besto av et system av ligninger som skulle beskrive de enkelte økonomiske sammenhengene, og en slik modell kunne ofte komme opp i flere hundre likninger. Felles for disse modellene var at de i stor grad bygget på økonomisk teori og i mindre grad empiriske analyse. Her er den norske tradisjonen et godt eksempel. Med utgangspunkt i arbeidet til Ragnar Frisch laget Statistisk sentralbyrå modellene MODIS og MODAG, som begge var planleggingsmodeller. De ble hovedsakelig brukt av Finansdepartementet i budsjettprosessen, men økonometrisk spesifisering, testing og evaluering spilte bare en mindre rolle når de ble laget. Først ved introduksjonen av KVARTS-modellen på slutten av 1980-tallet begynte dette å endre seg².

Ut over 70-tallet vokste det fram en økende skepsis mot denne type modeller, da det hadde vist seg at disse modellene ikke var særlig egnet til å lage økonomiske prognoser med³. Som en reaksjon på modellene man hadde benyttet på 60-tallet, begynte man å utvikle nye modeller som baserte seg på rene statistiske prinsipper. Disse bygget ofte på prinsippene for univariate modeller, og ARIMA-modellen ble et mye brukte rammeverket for denne type modeller. Disse modellene økte raskt i popularitet, da de var enkle å bruke og ofte hadde bedre prognoseegenskaper enn modellene fra 60-tallet.

Disse to tradisjonene sto i skarp kontrast til hverandre, den ene var først og fremst basert på økonomisk teori, mens den andre var basert på rene statistiske prinsipper. Denne kontrasten førte med seg en lang diskusjon om hvilke type modeller man skulle bruke for å produsere gode prognoser. Resultatet av denne diskusjonen ble at man utover på 1980-tallet, i større grad integrerte statistisk metode inn i makroøkonomisk analyse og modellbygging.

² Se Jansen (2000).

³ Se Granger (1986).

Det viktigste enkeltbidraget statistikken gav økonometrien på 1980-tallet hadde sitt utgangspunkt i studier av ikke-stasjonaritet i tidsserier. Statistikerne Clive Granger viste i en rekke artikler viktigheten av at økonometriske likninger var balansert, det vil si at en stasjonær variabel ikke kan bli forklart av en ikke-stasjonær variabel og omvendt⁴. Som en følge av dette ble teorien om kointegrasjon, det at den lineære kombinasjonen av to eller flere ikke-stasjonære variabler kan bli stasjonær, utviklet. På dette grunnlag har også statistikerne Søren Johansen, innen rammeverket av en generell vektor-autoregressiv modell (VAR), utviklet den mest brukte metoden for testing av kointegrasjon i en multivariat setting. I tillegg har verdien av å evaluere tidligere prognoser, som en del av den kontinuerlige modelleringsprosessen, blitt gitt en økende oppmerksomhet. Den såkalte LSE-metoden legger spesielt vekt på viktigheten av å teste og evaluere økonometriske modeller⁵. Denne utviklingen har medført at de makroøkonomiske modellene som brukes i dag i større grad er basert både på økonomisk teori og empirisk analyse.

Et åpent spørsmål er om utviklingen man har sett innenfor makroøkonomisk modellering på 80- og 90-tallet har gjort de rene statistiske modellene som ble utviklet på 70- og 80-tallet overflødige. Det er lite i den økonomiske litteraturen som tyder på det. ARIMA-modellene har fremdeles flere fordeler fremfor modeller på redusert form. Om man ønsker å lage prognoser for inflasjon med en modell på redusert form, er man avhengig av å ha gode prognoser for de forklarende variablene. Dette øker risikoen for at noe kan gå galt i prosessen. Det er dessuten ikke sikkert at det publiseres empirisk data for de forklarende variablene med samme frekvens man kunne ønske seg. Det ville for eksempel være vanskelig å lage prognoser på månedlig basis om man måtte basere modellen på kvartalvise datasett. Med en ARIMA-modell derimot er analysen uavhengig av andre variabler. Det har også vist seg at ARIMA-modeller er mer treffsikre på kort sikt og i perioder der utviklingen i tidsserien er relativt stabil⁶. ARIMA-modellene er dessuten adskillig mindre kostnadskrevenende å bruke.

Det er selvsagt også ulemper forbundet med bruken av disse modellene. Da modellene ikke har noe grunnlag i økonomisk teori, er det ikke mulig å bruke disse modellene om man ønsker å simulere konsekvensene av en endring i økonomien ved for eksempel en endring i den økonomiske politikken. Det er heller ikke mulig å bruke disse modellene til å predikere et

⁴ Se for eksempel Granger (1990).

⁵ Denne utviklingen er også lett å spore i Norge, se for eksempel kapittelet om etterprøving av Norges Banks anslag for 1999 i Inflasjonrapport 4/2000. For en nærmere diskusjon av LSE-metoden, se Hendry (1993, 1995).

⁶ Se for eksempel Mitra og Rashid (1996).

vendepunkt i tidsserien, for eksempel en endring inflasjonsutviklingen som følge av en endret finans- eller pengepolitikk. Det vil derfor være naturlig å bruke en ARIMA-modell om man skal lage prognoser for en tidsserie som er relativt stabil. Man bør derimot bruke en modell på redusert form om man ønsker å predikere en tidsserie som ikke er fullt så stabil, eller om man ønsker å predikere konsekvensene av endringer i eksogene variabler.

Det er også vanlig å bruke ARIMA-modeller som målestokk om man ønsker å vurdere prognoseegenskapene til modeller på redusert form. Den underliggende tanken her er at modellen ikke er god nok hvis den ikke predikerer minst like bra som en ren statistisk modell. Et godt eksempel på dette finner man i Cecchetti *et al.* (2000). Her evaluerer man verdien av å bruke ledende indikatorer for å predikere inflasjon i USA. Resultatet i denne studien er heller nedslående. De finner at det er få av de mest vanlig brukte indikatorene som gjør det bedre enn enkle autoregressive modeller.

I tillegg til dette er det mange i det økonomiske miljøet som stiller seg helt avvisende til real-økonomiske variabler som indikatorer på inflasjon. Freeman (1998) brukte kointegrasjon og Grangers kausalitetstest for tradisjonelle og nye metoder for beregning av underliggende inflasjon i USA. Han fant at begge metodene har de ønskede egenskapene for beregning av underliggende inflasjon, men at ingen av disse var særlig effektive som prognosemodeller. Cecchetti (1995) fant at det var ekstremt vanskelig å predikere inflasjon på denne måten selv med en kort tidshorison. Dette var grunnet i to forhold: han fant at det er svært få av de mest vanlig brukte indikatorene som kan predikere inflasjon på en tilfredsstillende måte, og at hverken inflasjonen eller de brukte indikatorene er stabile over tid.

Et problem med studiene jeg har nevnt ovenfor, er at de er gjort for to store økonomier. Det er rimelig å tro at inflasjonen i store økonomier som USA og Canada i større grad kan forklares ved rene autoregressive prosesser enn det som er tilfelle i små, åpne økonomier som i Norge. Dette fordi inflasjonen i land som USA og Canada stort sett blir bestemt innenlands, mens den i små, åpne økonomier som Norge i stor grad blir påvirket av importert inflasjon. Det kan derfor stilles spørsmål om univariate prognosemetoder vil ha samme validitet i Norge som de har i større og mer lukkede økonomier. Det er derfor hensiktsmessig å se på hvilket arbeid som er gjort på dette området i land som ligner Norge, hvorav Irland er et godt eksempel. Landet har en liten, åpen økonomi og har, som Norge, lenge hatt valutastabilitet mot Europa som høyeste mål for pengepolitikken. *Central Bank of Ireland* har gjort et omfattende arbeid

med å prøve ut forskjellige metoder for å modellere irsk inflasjon, summert opp av Quinn, Kenny og Meyer (1999). De fant at modeller på redusert form basert på kjøpekraftsparende, lønnsdannelse og aggregert pengemengde ikke gav gode prognoser for inflasjon. Dette fordi forholdet mellom inflasjon og indikatorene ofte ikke var stabilt over tid. De fant også at ARIMA-modeller predikerte den irske inflasjonen best på kort sikt, det vil si opp til et år.

2.3 Tidligere studier av inflasjon og prognoser i Norge.

I Norge er det først og fremst Norges Bank som fører pengepolitikken, og deres målsetning er å stabilisere prisstigningen. Derfor er mye av arbeidet med å lage modeller for norsk inflasjon gjort i Norges Bank. Resultatet av dette arbeidet har siden 1993 blitt publisert i en egen *Inflasjonsrapport* som nå kommer ut hvert kvartal. Her gir Norges Bank en beskrivelse av hvordan prisutviklingen har vært den siste tiden og hvordan de tror den vil utvikle seg de nærmeste to årene. Norges Bank baserer disse anslagene på RIMINI-modellen. Dette er en makroøkonomisk modell der man prøver å modellere den norske økonomien som helhet. Kjernemodellen består av et 30-talls stokastiske ligninger med ca. 100 eksogene variabler som må predikeres⁷. Dette er en modell på relativt aggregert nivå som er tydelig inspirert av den utviklingen man har hatt innenfor makroøkonomisk modellering de siste tyve årene. De enkelte ligningene er stort sett beregnet ved kointegrasjonsanalyse som beskrevet ovenfor, der anslagene etterprøves mot de reelle verdiene en gang i året.

Delmodellen som særlig omhandler prognoser av inflasjon er gitt i Bårdsen, Jansen og Nymoene (1999). De har laget en dynamisk modell for underliggende inflasjon der det spesielt fokuseres på sammenhengen mellom inflasjon og lønns- og kostnadsvekst. Modellen tar utgangspunkt i de motstridende lønnskravene fra fagforeninger og bedrifter for en liten åpen økonomi, men tar også høyde for eksogene variabler som sjokk og pengepolitiske instrumenter, som Norges Banks utlånsrente. Dermed gir modellen ikke bare et grunnlag for prognoser, den gir også retningslinjer for hvordan banken skal bruke rentepolitikken til å nå de ønskede mål for inflasjon.

I tillegg gir både Finansdepartementet og Statistisk sentralbyrå (SSB) ut egne prognoser for inflasjonsutviklingen, gjennom henholdsvis Nasjonalbudsjettet og Økonomisk Analyse.

⁷ Se Jansen (2000).

Finansdepartementet baserer seg her på den makroøkonomiske modellen MODAG, mens SSB baserer seg på modellen KVARTS.

Isachsen, Stoknes og Bjønnes (1999) rangerte syv prognosemakere ved å sammenligne prognosene de hadde laget for fire makroøkonomiske variabler i perioden 1988-96 med de faktiske verdiene som ble observert i denne perioden. De finner at modellene gjennomsnittlig gir gode prognoser. Når det gjelder inflasjon blir SSB rangert som nummer to og Norges Bank som nummer tre, men begge må se seg slått av Bankforeningen⁸. Finansdepartementet havner på en sjuende plass som den dårligste i sammenlikningen.

Det er også gjort et par forsøk på å modellere inflasjon utenfor de store institusjonene. Et eksempel er modellen utviklet av Kvilekval, Vaage og Vårdal (1998). Denne modellen fokuserer først og fremst på sammenhengen mellom pengemengdeveksten og inflasjon, og man bruker her en feiljusteringsmodell for å identifisere kortsiktige- og langsiktige sammenhenger. Den eneste forklaringsvariabelen denne modellen har til felles med modellen man bruker i Norges Bank er at de begge bruker importert inflasjon som en eksogen variabel. Likevel har de mye til felles i den grunnleggende tankegangen, da begge er modeller på redusert form som vektlegger realøkonomiske variabler, lønnsdannelse og pengemengdevekst når de skal forklare inflasjonsutviklingen.

Jeg kjenner ikke til, at det til dags dato, er gjort noen forsøk på å modellere inflasjon i Norge ved hjelp av ARIMA-modeller⁹. Diskusjonen ovenfor skulle tyde på at en slik modell vil kunne gi gode resultater, siden inflasjonen i Norge har vært lav og stabil siden slutten av 80-tallet.

⁸ Bankforeningen basere sine anslag på anslagene fra Finansdepartementet, men man forsøker å luke ut mulig politisk innflytelse på analysene.

⁹ Man har riktignok brukt X-11 ARIMA modeller for å sesongjustere tidsserier i SSB.

3. METODE

I dette kapitlet vil jeg gjøre rede for ARIMA-modellens oppbygning og virkemåte. Disse modellene er konstruert for å kunne modellere den stokastiske utviklingen i en tidsserie. For å forstå den grunnleggende tankegangen bak disse modellene, vil jeg begynne med å gjennomgå de statistiske konsepter som ligger til grunn. Fremstillingen er i stor grad basert på Pindyck og Rubinfeld (1998).

3.1 Stokastiske tidsserier

Tidsseriemodellen som er beskrevet i dette kapitlet er basert på en antakelse om at tidsserien som skal modelleres er generert ved en stokastisk prosess. Det vil si at vi antar at hver enkelt verdi y_1, y_2, \dots, y_T i tidsserien er valgt tilfeldig fra en sannsynlighetsfordeling. Ved å modellere denne prosessen ønsker vi å beskrive typiske trekk som kan hjelpe oss å si noe om sannsynligheten knyttet til de alternative fremtidige verdiene av tidsserien. Et enkelt eksempel på en stokastisk tidsserie er en *random walk*-prosess:

$$(3.1) \quad y_t = y_{t-1} + \varepsilon_t$$

der endringen i tidsserien bestemmes av et tilfeldig valgt restledd. Vi antar at restleddet oppfylle kravene $E[\varepsilon_t] = 0$, $E[\varepsilon_t^2] = \sigma_\varepsilon^2$ og $Cov[\varepsilon_t, \varepsilon_s] = 0$ for $s \neq t$. Når restleddet oppfyller disse kravene, betegnes det som generert av en *hvit støy*-prosess, et begrep som kommer til å stå sentralt i denne fremstillingen.

3.1.1 Autokorrelasjonsfunksjon av en *hvit-støy*-prosess

Autokorrelasjonsfunksjonen (AKF) forteller hvor mye samvarians det er mellom nærliggende observasjoner i en tidsserie y_t og utgjør et svært nyttig hjelpemiddel når man skal modellere en tidsserie. AKF for k tidsforsinkelser er gitt ved

$$(3.2) \quad \rho_k = \frac{E[(y_t - \mu_y)(y_{t+k} - \mu_y)]}{\sqrt{E[(y_t - \mu_y)^2]E[(y_{t+k} - \mu_y)^2]}} = \frac{Cov(y_t, y_{t+k})}{\sigma_{y_t} \sigma_{y_{t+k}}}$$

Telleren er her lik kovariansen mellom y_t og y_{t+k} , γ_k . For en stasjonær prosess er variansen i nevneren, på tidspunkt t , den samme som på tidspunkt $t+k$. Dette gir at nevneren er lik variansen av den stokastiske prosessen, γ_0 .

Anta at vi har en *hvit-støy*-prosess der $y_t = \varepsilon_t$. Dette vil gi

$$(3.3) \quad \rho_0 = \frac{E[(\varepsilon_t - E(\varepsilon_t))(\varepsilon_t - E(\varepsilon_t))]}{\sqrt{\text{var}(\varepsilon_t)\text{var}(\varepsilon_t)}} = \frac{\text{var}(\varepsilon_t)}{\text{var}(\varepsilon_t)} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2} = 1$$

$$(3.4) \quad \rho_1 = \frac{E[(\varepsilon_t - E(\varepsilon_t))(\varepsilon_{t-1} - E(\varepsilon_{t-1}))]}{\sqrt{\text{var}(\varepsilon_t)\text{var}(\varepsilon_{t-1})}} = \frac{0}{\sigma_\varepsilon^2} = 0$$

og tilsvarende vil det være for alle $i > 0$, det vil si $\rho_k = 0 \quad \forall i > 0$. Det vil igjen si at en *hvit-støy*-prosess ikke vil gi noen utslag i et autokorrellogram. Dette er et vesentlig funn da hensikten med univariate tidsseriemodeller er å trekke ut den systematiske informasjonen av tidsserien slik at det bare er *hvit støy* igjen i restleddet.

3.1.2 Stasjonære og ikke-stasjonære tidsserier

Før man begynner å modellere en tidsserie, er det viktig å vite om man kan anta at den underliggende stokastiske prosessen som genererer serien ikke vil endre seg med tiden. Hvis de karakteristiske egenskapene til den stokastiske prosessen endrer seg med tiden, det vil si hvis tidsserien er *ikke-stasjonær*, må man ta høyde for dette når man skal modellere tidsserien. Et typisk eksempel på en *ikke-stasjonær* tidsserie, er en tidsserie som stiger jevnt med tiden, som for eksempel BNP.

For at en tidsserie skal kunne defineres som stasjonær må gjennomsnittet av tidsserien, definert ved $\mu_y = E(y_t)$, være konstant over tid:

$$(3.5) \quad E(y_t) = E(y_{t+m})$$

for enhver t og m .

I tillegg må variansen, definert ved $\sigma_y^2 = E[(y_t - \mu_y)^2]$, være konstant over tid:

$$(3.6) \quad E[(y_t - \mu_y)^2] = E[(y_{t+m} - \mu_y)^2],$$

for enhver t og m . Til slutt må kovariansen for enhver tidsforsinkelse k , definert ved $\gamma_k = Cov(y_t, y_{t+k}) = E[(y_t - \mu_y)(y_{t+k} - \mu_y)]$ være konstant over tid:

$$(3.7) \quad Cov(y_t, y_{t+k}) = Cov(y_{t+m}, y_{t+m+k}).$$

Formell testing av tidsseriens stasjonaritetsegenskaper vil bli gitt i kapittel 5.2

3.2 ARIMA-modellen

Autoregressive Integrated Moving Average (ARIMA) modeller er den vanligste typen av univariate modeller. Modellen beskriver verdien av en tidsserie i periode t uttrykt som en funksjon av tidligere verdier av den samme tidsserien (AR: autoregressiv), pluss en kombinasjon av løpende og tidligere verdier av restleddet (MA: moving average, glidende gjennomsnitt). Den integrerte komponenten (I) refererer til det antall ganger serien må differensieres for at den skal bli stasjonær. I en ARIMA(p,d,q) betegner p antall AR(p)-ledd og q antall MA(q)-ledd, mens d betegner antall ganger tidsserien er blitt differensiert.

3.2.1 Glidende gjennomsnittsmoeller

En glidende gjennomsnittsprosess refererer til en type modeller hvor den avhengige variabelen kan uttrykkes som et veid gjennomsnitt av restleddets verdier i denne og foregående perioder. En MA(q)-prosess skrives som

$$(3.8) \quad y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

der μ er et konstantledd og parametrene $\theta_1, \dots, \theta_q$ kan være både positive og negative. En forutsetning er at de stotastiske restleddene er uavhengig fordelt over tid, det vil si at de er generert som en *hvit-støy*-prosess.

Ved bruk av en tidsforsinkelsesoperator B kan prosessen skrives som

$$(3.9) \quad y_t = \mu + \theta(B)\varepsilon_t$$

der $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$.

3.2.2 Egenskaper ved glidende gjennomsnittsmoeller

Gjennomsnittet av en glidende gjennomsnittsprcess er uavhengig av tiden, da $E(y_t) = \mu$. Hver ε_t er antatt å være generert fra den samme hvit-støy-prcessen, det vil si at $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ og $E(\varepsilon_t \varepsilon_{t-k}) = 0$ for alle t og $k \neq 0$. Variansen γ_0 av en glidende gjennomsnittsprcess av grad q er gitt ved

$$(3.10) \quad \begin{aligned} \text{Var}(y_t) = \gamma_0 &= E[(y_t - \mu)^2] \\ &= E(\varepsilon_t^2 + \theta_1^2 \varepsilon_{t-1}^2 + \dots + \theta_q^2 \varepsilon_{t-q}^2 - 2\theta_1 \varepsilon_t \varepsilon_{t-1} - \dots) \\ &= \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2 + \dots + \theta_q^2 \sigma_\varepsilon^2 \\ &= \sigma_\varepsilon^2 (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \end{aligned}$$

Jeg vil nå se på noen enkle glidende gjennomsnittsprcesser og kalkulere gjennomsnittet, variansen, kovariansen og autokorrelasjonsfunksjonen for disse. Dette er statistiske mål som vil være til hjelp når jeg skal identifisere modellene i kapittel 4. En enkel MA(1)-prcess er gitt ved

$$(3.11) \quad y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Denne prcessen har gjennomsnittet μ og variansen $\gamma_0 = \sigma_\varepsilon^2 (1 + \theta_1^2)$. Kovariansen for en én-periode tidsforsinkelse, γ_1 , er da gitt ved

$$(3.12) \quad \gamma_1 = E[(y_t - \mu)(y_{t-1} - \mu)] = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1})(\varepsilon_{t-1} - \theta_1 \varepsilon_{t-2})] = -\theta_1 \sigma_\varepsilon^2$$

Generelt kan man bestemme kovariansen for $k > 1$ som

$$(3.13) \quad \gamma_k = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1})(\varepsilon_{t-k} - \theta_1 \varepsilon_{t-k-1})] = 0$$

Dette gir at en MA(1)-prosess har en kovarians lik null når tidsforsinkelsen blir større enn én periode. Det vil si at y_t er korrelert med y_{t-1} og y_{t+1} , men ingen av de andre verdier av tidsserien. Man sier at en MA(1)-prosessen har en *hukommelse* på én periode.

Autokorrelasjonsfunksjonen for en MA(1)-prosess er altså:

$$(3.14) \quad \rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \frac{-\theta_1}{1+\theta_1^2} & k = 1 \\ 0 & k > 1 \end{cases}$$

Om man ser på en MA(2)-prosess, er denne gitt ved:

$$(3.15) \quad y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

Denne prosessen har gjennomsnittet μ , variansen $\gamma_0 = \sigma_\varepsilon^2(1 + \theta_1^2 + \theta_2^2)$ og kovariansene

$$(3.16) \quad \begin{aligned} \gamma_1 &= E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2})(\varepsilon_{t-1} - \theta_1 \varepsilon_{t-2} - \theta_2 \varepsilon_{t-3})] \\ &= -\theta_1 \sigma_\varepsilon^2 + \theta_2 \theta_1 \sigma_\varepsilon^2 = -\theta_1(1 - \theta_2) \sigma_\varepsilon^2 \end{aligned}$$

$$(3.17) \quad \begin{aligned} \gamma_2 &= E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2})(\varepsilon_{t-2} - \theta_1 \varepsilon_{t-3} - \theta_2 \varepsilon_{t-4})] \\ &= -\theta_2 \sigma_\varepsilon^2 \end{aligned}$$

$$(3.18) \quad \gamma_k = 0 \quad \text{for } k > 2.$$

Da er autokorrelasjonsfunksjonen gitt ved

$$(3.19) \quad \rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}$$

$$(3.20) \quad \rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}$$

$$(3.21) \quad \rho_k = 0 \quad \text{for } k > 2.$$

Dette gir at en MA(2)-prosess har en *hukommelse* på nøyaktig to periode, slik at verdien av y_t bare er påvirket av det som har skjedd i inneværende periode, forrige periode og perioden før der igjen.

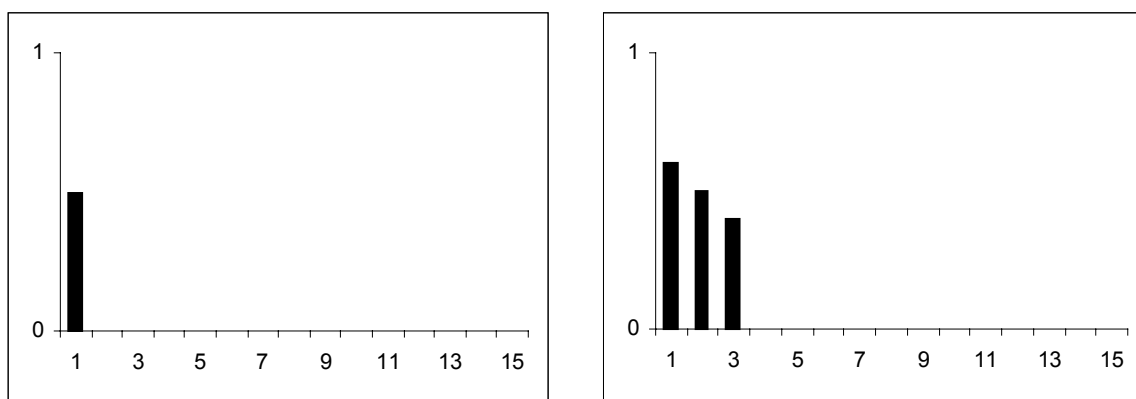
Det kan vises at autokorrelasjonsfunksjonen for en glidende gjennomsnittsprosess av grad q er gitt ved

$$(3.22) \quad \rho_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \dots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & k = 1, \dots, q \\ 0 & k > q \end{cases}$$

Dette gir at korrelogrammet til MA(q)-prosess vil gi et utslag ulikt null opp til og med tidsforsinkelse q og synker så til null.

Autokorrelasjonsfunksjonen for en tidsserie y_t kan representeres grafisk ved et korrelogram, se figur 3.1. Dette er et nyttig hjelpemiddel om man ønsker å identifisere en glidende gjennomsnittsprosess i en tidsserie. For en MA(1)-prosess vil man i et korrelogram observere en verdi ulik null for den første tidsforsinkelsen, mens de påfølgende tidsforsinkelsene vil gi utslag lik null. I figur 3.1.1 er det gitt et eksempel på hvordan et slikt korrelogram kan se ut, mens jeg i figur 3.1.2 har vist et tenkt eksempel for en MA(3)-prosess.

Figur 3.1.1 og 3.1.1 – Eksempel: Korrellogram for en MA(1)- og en MA(3)-prosess.



3.2.3 Autoregressive modeller

I en autoregressiv modell av grad p blir en observasjon i periode t generert som et veid gjennomsnitt av tidligere observasjoner p perioder tilbake i tid. En ren AR(p)-prosess skrives som:

$$(3.23) \quad y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \delta + \varepsilon_t$$

der ϕ_p er parametrene som angir hvor mye vekt man skal legge på observasjonen i periode $t-p$, δ er et konstantledd og ε_t er restledd for periode t . Restleddet antas å være *hvit støy*. Betegnelsen autoregressiv refererer til at modellens parametre kan estimeres ved en regresjonsanalyse, hvor den uavhengige variabelen er verdien av den avhengige variabelen i foregående perioder.

Ved bruk av en tidsforsinkelsesoperator B kan prosessen skrives

$$(3.24) \quad \phi(B)y_t = \delta + \varepsilon_t$$

$$\text{der } \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p.$$

3.2.4 Egenskaper ved autoregressive modeller

Om den autoregressive prosessen er stasjonær, vil gjennomsnittet μ være konstant. Det vil si at $E(y_t) = E(y_{t-1}) = E(y_{t-2}) = \dots = \mu$. Dette gir at gjennomsnittet μ er gitt ved $\mu = \phi_1\mu + \phi_2\mu + \dots + \phi_p\mu + \delta$ eller

$$(3.25) \quad \mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \dots - \phi_p}$$

For at prosessen skal være stasjonær må $\phi_1 + \phi_2 + \dots + \phi_p < 1$, som er en nødvendig, men ikke tilstrekkelig betingelse for stasjonaritet.

For å illustrere egenskapene til de autoregressive modellene vil jeg her se nærmere på noen enkle autoregressive prosesser. En AR(1)-prosess er gitt ved

$$(3.26) \quad y_t = \phi_1 y_{t-1} + \delta + \varepsilon_t$$

Denne prosessen har et gjennomsnitt

$$(3.27) \quad \mu = \frac{\delta}{1 - \phi_1}$$

og er stasjonær dersom $|\phi_1| < 1$. For å beregne variansen γ_0 antar jeg at prosessen er stasjonær, slik at den har en konstant varians, og at $\delta = 0$. Dette gir

$$(3.28) \quad \gamma_0 = E[(\phi_1 y_{t-1} + \varepsilon_t)^2] = E(\phi_1^2 y_{t-1}^2 + \varepsilon_t^2 + 2\phi_1 y_{t-1} \varepsilon_t) = \phi_1^2 \gamma_0 + \sigma_\varepsilon^2$$

som igjen gir

$$(3.29) \quad \gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}$$

Kovariansen ved én og to tidsforsinkelser er da gitt ved

$$(3.30) \quad \gamma_1 = E[y_{t-1}(\phi_1 y_{t-1} + \varepsilon_t)] = \phi_1 \gamma_0 = \frac{\phi_1 \sigma_\varepsilon^2}{1 - \phi_1^2}$$

$$(3.31) \quad \gamma_2 = E[y_{t-2}(\phi_1^2 y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t)] = \phi_1^2 \gamma_0 = \frac{\phi_1^2 \sigma_\varepsilon^2}{1 - \phi_1^2}$$

På samme måten er kovariansen til en k -perioder tidsforsinkelse gitt ved

$$(3.32) \quad \gamma_k = \phi_1^k \gamma_0 = \frac{\phi_1^k \sigma_\varepsilon^2}{1 - \phi_1^2}$$

Dette gjør at autokorrelasjonsfunksjonen for en stasjonær AR(1)-modell blir spesielt enkel. Den begynner i $\rho_0 = 1$ og synker så geometrisk med verdien av k :

$$(3.33) \quad \rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^k$$

Legg merke til at denne prosessen har *uendelig hukommelse*. Det vil si at verdien i inneværende periode er avhengig av alle tidligere verdier, selv om betydningen av avhengigheten er avtagende med tiden.

En AR(2)-prosess er gitt ved

$$(3.34) \quad y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \delta + \varepsilon_t$$

Gjennomsnittet av denne prosessen blir da

$$(3.35) \quad \mu = \frac{\delta}{1 - \phi_1 - \phi_2}$$

og en nødvendig betingelse for stasjonaritet er at $\phi_1 + \phi_2 < 1$.

Variansen og kovariansen er da gitt ved

$$(3.36) \quad \gamma_0 = E[y_t(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t)] = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_\varepsilon^2$$

$$(3.37) \quad \gamma_1 = E[y_{t-1}(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t)] = \phi_1 \gamma_0 + \phi_2 \gamma_1$$

$$(3.38) \quad \gamma_2 = E[y_{t-2}(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t)] = \phi_1 \gamma_1 + \phi_2 \gamma_0$$

og generelt, for $k \geq 2$

$$(3.39) \quad \gamma_k = E[y_{t-k}(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t)] = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}$$

Man kan løse ligningene (3.36), (3.37) og (3.38) simultant for å få γ_0 gitt ved ϕ_1 , ϕ_2 og σ_ε^2 .

Ligning (3.37) kan omskrives som

$$(3.40) \quad \gamma_1 = \frac{\phi_1 \gamma_0}{1 - \phi_2}$$

Om man substituerer ligning (3.38) inn i ligning (3.36) får vi

$$(3.41) \quad \gamma_0 = \phi_1 \gamma_1 + \phi_2 \phi_1 \gamma_1 + \phi_2^2 \gamma_0 + \sigma_\varepsilon^2$$

Man kan da bruke ligning (3.40) til å eliminere γ_1 slik at man får

$$(3.42) \quad \gamma_0 = \frac{\phi_1^2 \gamma_0}{1 - \phi_2} + \frac{\phi_2 \phi_1^2 \gamma_0}{1 - \phi_2} + \phi_2^2 \gamma_0 + \sigma_\varepsilon^2$$

som etter litt omorganisering gir

$$(3.43) \quad \gamma_0 = \frac{(1 - \phi_2) \sigma_\varepsilon^2}{(1 + \phi_2)[(1 - \phi_2)^2 - \phi_1^2]}$$

Disse ligningene kan også brukes til å utlede autokorrelasjonsfunksjonen ρ_k . Fra ligningene (3.38) og (3.40) får man at

$$(3.44) \quad \rho_1 = \frac{\phi_1}{1 - \phi_2}$$

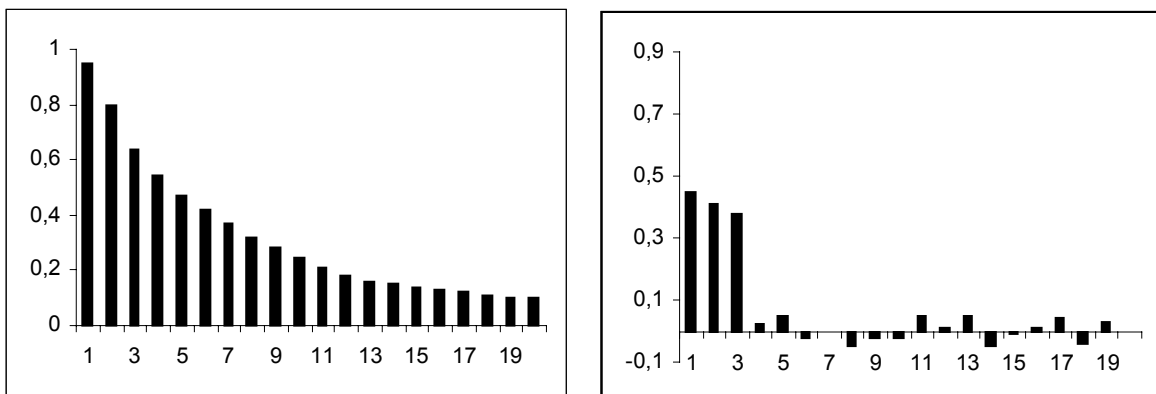
$$(3.45) \quad \rho_2 = \phi_2 + \frac{\phi_1^2}{1 - \phi_2}$$

Og fra ligning (3.39) kan man se at for $k \geq 2$, så er

$$(3.46) \quad \rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$$

som man kan bruke til å beregne autokorrelasjonsfunksjonen for $k > 2$. Ligning (3.44) og (3.45) kalles *Yule-Walker-ligninger*. Dette gir at autokorrelasjonsfunksjonen av en autoregressiv modell synker gradvis med antall tidsforsinkelser. Et typisk eksempel på et slikt korrellogram er gitt i figur 3.2.1. Det er tydelig at autokorrelasjonsfunksjonen ikke kommer til å være til mye hjelp om man ønsker å identifisere graden p av AR(p)-prosessen. For å gjøre denne identifiseringsprosessen enklere bruker man den partielle autokorrelasjonsfunksjonen.

Fig 3.2.1 og 3.2.2 – Eksempel: AKF og PAKF for en typisk AR(3) prosess.



Den *partielle autokorrelasjonsfunksjonen* (PAKF) mellom y_t og y_{t-k} er korrelasjonen mellom y_{t-k} og y_t minus den delen som er forklart av mellomliggende tidsforsinkelser. Det vil si at man får se hvor stort utslag hver enkelt tidsforsinkelse gjør. For å forstå hvordan den partielle autokorrelasjonsfunksjonen virker må man ta utgangspunkt i kovariansen og autokorrelasjonsfunksjonen for en autoregressiv prosess av grad p . Kovariansen er her gitt ved

$$(3.47) \quad \gamma_k = E[y_{t-k} (\phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t)]$$

Om man nå setter $k = 0, 1, \dots, p$ får vi $p + 1$ ligninger som man kan løse for $\gamma_0, \gamma_1, \dots, \gamma_p$:

$$(3.48) \quad \begin{aligned} \gamma_0 &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma_\varepsilon^2 \\ \gamma_1 &= \phi_1 \gamma_0 + \phi_2 \gamma_1 + \dots + \phi_p \gamma_{p-1} \\ \dots & \dots \dots \dots \dots \dots \dots \dots \dots \\ \gamma_p &= \phi_1 \gamma_{p-1} + \phi_2 \gamma_{p-2} + \dots + \phi_p \gamma_0 \end{aligned}$$

For tidsforsinkelser k større enn p er kovariansen gitt ved

$$(3.49) \quad \gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p}$$

Ved å dele ligningene i (3.48) med γ_0 får vi p ligninger som til sammen gir de p første verdiene av autokorrelasjonsfunksjonen:

$$(3.50) \quad \begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \dots & \dots \dots \dots \dots \dots \dots \dots \dots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p \end{aligned}$$

For tidsforsinkelser k større enn p , med bakgrunn i ligning (3.49), er autokorrelasjonsfunksjonen gitt ved

$$(3.51) \quad \rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}$$

Ligningene i (3.50) er *Yule-Walker*-ligninger tilsvarende ligning (3.44) og (3.45) for AKF, det vil si at om man kjenner til en av verdiene for $\rho_1, \rho_2, \dots, \rho_p$ så kan man bruke disse til å beregne verdiene av $\phi_1, \phi_2, \dots, \phi_p$.

Men for å løse *Yule-Walker*-ligningene i (3.50) må man vite hvor stor p , graden av den autoregressive prosessen, er. Om man ikke vet dette løser man *Yule-Walker*-ligningen en etter en for etterfølgende verdier av p . Det vil si at man begynner med hypotesen om at $p = 1$. Da gir ligning (3.50) resultatet $\rho_1 = \phi_1$, eller om vi bruker autokorrelasjonen av datautvalget, $\hat{\rho}_1 = \hat{\phi}_1$. Hvis den kalkulererte verdien av $\hat{\phi}_1$ er signifikant forskjellig fra null, vet vi at den autoregressive prosessen minst er av grad 1. Jeg kommer videre til å betegne verdien av $\hat{\phi}_1$ som a_1 .

Om man så antar at $p=2$, vil *Yule-Walker*-ligningen i (3.50) gi oss et nytt sett av estimerte verdier for $\hat{\phi}_1$ og $\hat{\phi}_2$. Hvis den estimerte verdien av $\hat{\phi}_2$ er signifikant forskjellig fra null, vet vi at den autoregressive prosessen minst er av grad 2. Jeg kommer videre til å betegne verdien av $\hat{\phi}_2$ som a_2 . Om man repeterer denne prosessen for de etterfølgende verdiene av p vil man få en serie verdier a_1, a_2, a_3, \dots som gir den partielle autokorrelasjonsfunksjonen.

Vi har dermed et nyttig redskap for å identifisere graden p av en autoregressiv prosess. Om den faktiske graden av den autoregressive prosessen er lik p , vil vi observere at den partielle autokorrelasjonsfunksjonen er tilnærmet lik null for tidsforsinkelser større enn p . I figur 3.2.2 har jeg vist hvordan den partielle autokorrelasjonsfunksjonen typisk vil se ut for en AR(3)-prosess.

3.2.5 Blandede autoregressive – glidende gjennomsnittsmoeller.

Mange stasjonære stokastiske tidsserier lar seg ikke modellere som en ren AR(p)- eller MA(q)-prosess, da serien inneholder elementer av begge disse prosessene. For å kunne modellere en slik tidsserie setter man de to foregående modellene sammen til én modell.

En ARMA(p,q)-modell skrives som

$$(3.52) \quad y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

eller

$$(3.53) \quad \phi(B)y_t = \delta + \theta(B)\varepsilon_t$$

hvor $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$

og $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$

3.2.6 Egenskaper ved ARMA(p,q)-modellen.

ARMA-modeller er adskillig vanskeligere å identifisere enn rene AR(p)- eller MA(q)-modeller, først og fremst fordi de gir motsatt utslag i AKF- og PAKF-korrellogrammene. Som vi har sett ovenfor, vil utslaget av en ren AR(p)-modell synke langsomt i et AKF-korrellogram, mens utslaget i PAKF-korrellogrammet vil synke til null for tidsforsinkelser større enn p . Mønsteret for en ren MA(q)-prosess vil typisk være motsatt. Et korrellogram kan likevel gi en god pekepinn på hvilke modeller det er verd å undersøke videre.

Anta at prosessen er stasjonær slik at gjennomsnittet er konstant over tid. Da er gjennomsnittet gitt ved $\mu = \phi_1 \mu + \dots + \phi_p \mu + \delta$ eller

$$(3.54) \quad \mu = \frac{\delta}{1 - \phi_1 - \dots - \phi_p}$$

Dette gir at $\phi_1 + \phi_2 + \dots + \phi_p < 1$ er en nødvendig betingelse for stasjonaritet.

For å vise egenskapene til en blandet autoregressive – glidende gjennomsnittsmodell tar jeg utgangspunkt i en enkel ARMA(1,1):

$$(3.55) \quad y_t = \phi_1 y_{t-1} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Variansen og kovariansen til denne prosessen, når man antar at $\delta = 0$, er gitt ved

$$(3.56) \quad \begin{aligned} \gamma_0 &= E[y_t(\phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})] = E[(\phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})^2] \\ &= \phi_1^2 \gamma_0 - 2\phi_1 \theta_1 E[y_{t-1} \varepsilon_{t-1}] + \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2 \end{aligned}$$

Siden $E(y_{t-1} \varepsilon_{t-1}) = \sigma_\varepsilon^2$, får vi

$$(3.57) \quad \gamma_0(1 - \phi_1^2) = \sigma_\varepsilon^2(1 + \theta_1^2 - 2\phi_1 \theta_1)$$

slik at variansen er gitt ved

$$(3.58) \quad \gamma_0 = \frac{1 + \theta_1^2 - 2\phi_1 \theta_1}{1 - \phi_1^2} \sigma_\varepsilon^2$$

Ved å gjenta denne prosessen kan vi fastslå kovariansen $\gamma_1, \gamma_2, \dots$ gitt ved

$$(3.59) \quad \gamma_1 = E[y_{t-1}(\phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})] = \phi_1 \gamma_0 - \theta_1 \sigma_\varepsilon^2 = \frac{(1 - \phi_1 \theta_1)(\phi_1 - \theta_1)}{1 - \phi_1^2} \sigma_\varepsilon^2$$

$$(3.60) \quad \gamma_2 = E[y_{t-2}(\phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})] = \phi_1 \gamma_1$$

og på samme måte $\gamma_k = \phi_1 \gamma_{k-1} \quad k \geq 2$.

Autokorrelasjonsfunksjonen er da gitt ved

$$(3.61) \quad \rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1 - \phi_1 \theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1 \theta_1}$$

og for tidsforsinkelser større enn 1,

$$(3.62) \quad \rho_k = \phi_1 \rho_{k-1} \quad k \geq 2$$

Dette gir at autokorrelasjonsfunksjonen begynner i startverdien ρ_1 (som er en funksjon av både ϕ_1 og θ_1) og synker så geometrisk etter denne startverdien. Dette reflekterer det faktum at den glidende gjennomsnittsdelen av prosessen bare har en hukommelse på én periode.

3.2.7 ARIMA-modeller

Frem til nå har jeg i denne gjennomgangen antatt at tidsserien er stasjonær. Om tidsserien er ikke-stasjonær kan den som oftest transformeres slik at den blir det. Dette gjøres ved å *differensiere* tidsserien en eller flere ganger. Den transformerte stasjonære serien er gitt ved

$$(3.63) \quad w_t = \Delta^d y_t$$

der d angir hvor mange ganger tidsserien må differensieres for å oppnå stasjonaritet. Differensiering av grad én og to er da gitt ved

$$(3.64) \quad \Delta y_t = y_t - y_{t-1}$$

$$(3.65) \quad \Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$$

Hvor mange ganger den må differensieres kan avgjøres ved hjelp av en ADF-test, se kapittel 5.2. Legg merke til at om man har en differensiert tidsserie w_d , kan man gå tilbake til den opprinnelige tidsserien ved å summere w_t d ganger:

$$(3.66) \quad y_t = \sum^d w_t$$

Det vil si at om den differensierte serien er gitt ved $w_t = \Delta y_t$, så kan man beregne y_t ved

$$(3.67) \quad y_t = \sum w_t = y_0 + w_1 + w_2 + \dots + w_t$$

Etter at man har differensiert tidsserien y_t for å produsere den stasjonære tidsserien w_t , kan man modellere w_t som en vanlig ARMA-prosess. Om $w_t = \Delta^d y_t$ og w_t kan modelleres som

en ARMA(p,q)-prosess, sier man at y_t er modellert som en *integrert autoregressiv – glidende gjennomsnittsprosess av grad (p,d,q)*, forkortet som ARIMA(p,d,q). En ARIMA(p,d,q)-modell skrives som

$$(3.68) \quad \phi(B)\Delta^d y_t = \delta + \theta(B)\varepsilon_t$$

hvor $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$

og $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$

Gjennomsnittet av $w_t = \Delta^d y_t$ er gitt ved

$$(3.69) \quad \mu_w = \frac{\delta}{1 - \phi_1 - \phi_2 - \dots - \phi_p} P$$

Dette gir at om δ ikke er lik null, så inneholder den integrerte tidsserien y_t en deterministisk trend. En ARIMA(p,d,q)-modell på generell form er derfor gitt ved

$$(3.70) \quad \phi(B)\Delta^d y_t = \theta(B)\varepsilon_t$$

3.2.8 ARIMA-modeller med sesong.

Vi ser ofte at tidsserier inneholder en sesongkomponent, der gjennomsnittet eller variansen følger et gjentakende mønster. Disse vil ofte ikke bli fanget opp av en vanlig ARMA(p,q)-modell, fordi modellen bare tar hensyn til observasjoner som ligger nær periode t i tid. Dette kan man lett se for en AR(1)-modell, der man bare ser på korrelasjonen mellom y_t og y_{t-1} .

Box og Jenkins (1994) anbefaler å løse dette problemet ved å inkludere sesongautoregressive (SAR) og sesong-glidende gjennomsnitt (SMA) ledd i ARMA-modellen. Disse leddene fungerer på samme måte som AR- og MA-leddene nevnt ovenfor, men istedenfor å se på sammenhengen mellom nærliggende observasjoner, ser man på sammenhengen mellom de observasjonene som utgjør sesongkomponenten. Det vil si at om man har en årlig sesong-

komponent i tidsserien, så gir SAR(P) korrelasjonen mellom y_t og verdien av tidsserien med et års tidsforsinkelse, y_{t-s} , der $s = 4$ for kvartalsdata og $s = 12$ for månedlige data.

Som jeg har vist tidligere, er en AR(1)-prosess uten sesong eller konstant gitt ved

$$(3.71) \quad y_t = \phi_1 y_{t-1} + \varepsilon_t$$

som også kan skrives med en tidsforsinkelsesoperator B , som

$$(3.72) \quad (1 - \phi_1 B)y_t = \varepsilon_t.$$

Om man inkluderer et SAR(1)-ledd for en årlig sesongkomponent, med månedlige data, kan modellen skrives som

$$(3.73) \quad (1 - \phi_1 B)(1 - \Phi B^{12})y_t = \varepsilon_t$$

eller om man skriver den ut

$$(3.74) \quad y_t = \phi_1 y_{t-1} + \Phi y_{t-12} - \Phi \phi_1 y_{t-13} + \varepsilon_t$$

På tilsvarende måte kan en MA(1)-prosess, uten sesong eller konstant, skrives som

$$(3.75) \quad y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

eller med en tidsforsinkelsesoperator B , som

$$(3.76) \quad y_t = (1 + \theta_1 B)\varepsilon_t$$

Om man inkluderer et SMA(1)-ledd for en årlig sesongkomponent, med månedlige data, kan modellen skrives som

$$(3.77) \quad y_t = (1 - \theta_1 B)(1 - \Theta B^{12})\varepsilon_t$$

eller om man skriver den ut

$$(3.78) \quad y_t = \varepsilon_t - \theta\varepsilon_{t-1} - \Theta\varepsilon_{t-12} + \Theta\theta\varepsilon_{t-13}$$

En generell sesong ARMA(p,q)x(P,Q)-modell kan derfor skrives som

$$(3.79) \quad \phi(B)\Phi(B)y_t = \theta(B)\Theta(B)\varepsilon_t$$

$$\text{hvor} \quad \Phi(B) = 1 - \Phi_{1s}B^{1s} - \Phi_{2s}B^{2s} - \dots - \Phi_{Ps}B^{Ps}$$

$$\text{og} \quad \Theta(B) = 1 - \Theta_{1s}B^{1s} - \Theta_{2s}B^{2s} - \dots - \Theta_{Qs}B^{Qs}$$

der s er lik den sesongbetonte rekkevidden, det vil si at for kvartalsdata er $s = 4$ og for månedlige data er $s = 12$. P og Q refererer her seg til graden av sesongleddene.

Der er også mulig at det er et sesongmønster som gjør at tidsserien er ikke-stasjonær i utgangspunktet. Man kan da sesongdifferensiere tidsserien for å filtrere ut denne sesongkomponenter. Antall ganger man sesongdifferensiere tidsserien betegnes ved D. En generell sesong ARIMA(p,d,q)x(P,D,Q)-modell kan derfor skrives som

$$(3.74) \quad \phi(B)\Phi(B)\nabla^d\nabla_s^D y_t = \theta(B)\Theta(B)\varepsilon_t$$

hvor den transformerte tidsserien $w_t = \nabla^d\nabla_s^D y_t$ er en stasjonær tidsserie, og $\nabla^d = (1-B)^d$ representerer antallet ordinære differensieringer, mens $\nabla_s^D = (1-B^s)^D$ representerer antallet sesongdifferensieringer som er nødvendig for å gjøre y_t stasjonær.

3.2.9 Intervensjonsanalyse

En tidsserie kan ofte være påvirket av spesielle hendelser slik som politiske endringer, økonomiske sjokk eller lignende. Slike hendelser vil her bli referert til som intervensjoner. Om tidsserien er påvirket av slike hendelser må det tas med i modelleringsprosessen. Det kan forekomme to typer av intervensjoner: engangshendelse som bare påvirker tidsserien i én periode, og hendelser som medfører et strukturelt skift i tidsserien. For å korrigere for disse kan man benytte dummyvariabler.

For å korrigere for avvikende observasjoner i periode T , kan man benytte dummyvariabler gitt ved

$$(3.75) \quad du_{p,t} = \begin{cases} 0 & t \neq T \\ 1 & t = T \end{cases}$$

For å korrigere for et strukturelt skift i tidsserien i periode T , kan man benytte dummyvariabler gitt ved

$$(3.76) \quad du_{s,t} = \begin{cases} 0 & t < T \\ 1 & t \geq T \end{cases}$$

Der finnes flere formelle tester for å avgjøre om en observasjon er avvikende eller om det er et strukturelt brudd i tidsserien. Men i denne oppgaven vil jeg begrense meg til å bruke grafen av residualen til de estimerte modellene til å identifisere eventuelle avvikende observasjoner.

4. BESKRIVELSE AV DATAMATERIALE

Som nevnt i innledningen mener man med inflasjon en kontinuerlig økning i det generelle prisnivået, her målt ved konsumprisindeksen (KPI). Det kan reises et par innvendinger mot dette. Konsumprisindeksen dekker et bredt spekter av priser, men langt fra alle prisene i økonomien. Det er bare private husholdningers forbruk som inngår i vektgrunnlaget, og prisene hentes dermed inn fra de utsalgsstedene husholdningene benytter seg av. Andre sektorer av økonomien fanges dermed ikke direkte opp. Det kan imidlertid argumenteres for at bevegelser i tidligere produksjons- og omsetningsledd i noe grad vil gjenspeile seg i de prisene konsumentene møter. Konsumprisindeksen vil da indirekte fange opp dette.

Et annet problem med KPI er at den kan bli betydelig påvirket av enkeltpriser som endrer seg kraftig. Anta for eksempel at en vare som utgjør 5% av vektgrunnlaget i indeksen stiger med 200%, mens prisen på alle andre varer forblir uendret. KPI vil da stige med 10% uten at det generelle prisnivået i økonomien kan sies å ha godt opp. Hvordan de enkelte delprissettene som utgjør KPI skal vektes, har ofte vært gjenstand for kritikk. Statistisk sentralbyrå (SSB) prøver å løse dette problemet ved å beregne vektene i konsumprisindeksen på grunnlag av resultatene av SSBs årlige Forbrukerundersøkelser. Her kartlegger man hvor stor del varene i de enkelte delindeksens utgjør av husstandenes budsjett. Vektene beregnes som et gjennomsnitt av budsjettandelene i de siste tre årene (3-årig glidende gjennomsnitt).

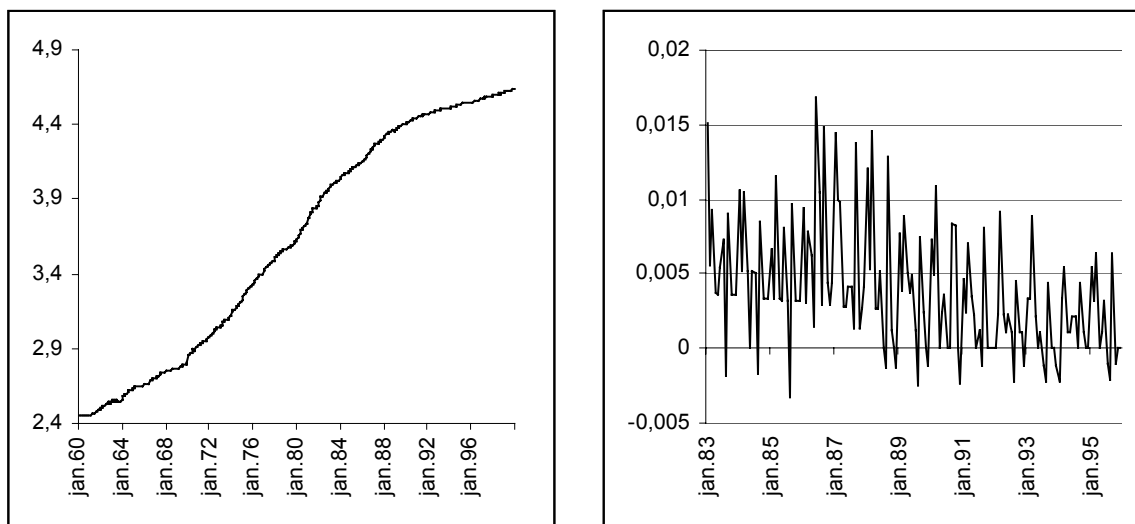
Et tredje problem med konsumprisindeksen er at rapporterte prisendringer kan skyldes endringer i kvaliteten av konsumvarene fremfor en ren prisendring. SSB prøver å rette på dette ved å velge ut representantvarer og ved å hente inn detaljerte beskrivelser av hver vare og tjeneste.

Til tross for disse problemene er konsumprisindeksen det mest vanlig brukte målet på generell prisstigning. Dette gir god mening sett på bakgrunn av at formålet med all økonomisk aktivitet i siste instans er konsum og at konsumprisindeksen tar for seg de prisene konsumentene står ovenfor. Da konsumprisindeksen er den prisindeksen som er mest vanlig brukt i tilsvarende studier både i Norge og i utlandet, vil valget av denne indeksen gjøre det lettere å sammenligne dette studiet med tilsvarende undersøkelser.

Jeg har valgt å bruke månedlige data for konsumprisindeksen som beregningsgrunnlag. Dette har bakgrunn i et ønske om å kunne predikere inflasjon på kort sikt. Men valget av månedlige data kan også utgjøre et problem. Konsumprisindeksen er basert på innrapporterte priser på et utvalg av varer og tjenester fra et utvalg av tilbydere. Disse prisene er stort sett innrapportert månedlig. Men den største delindeksen, utgifter til bolig, som rapporteres gjennom *Husleieundersøkelsen*, innrapporteres kvartalsvis¹⁰. Denne delindeksen utgjør ca. 1/5 av den totale konsumprisindeksen.

Det er nødvendig med en viss størrelse på datasettet for å kunne gi det nødvendige grunnlaget for estimering av en ARIMA-modell. Om det er mulig, bør man bruke minst 50 og helst 100 eller flere observasjoner (Box, Jenkins og Reinsel, 1994, s. 15). For inflasjon er det imidlertid ikke noe problem å finne lange nok tidsserier. Konsumprisindeksen er rapportert tilbake til 1959¹¹. Men selv om lange tidsserier er tilgjengelig, kan tidsserien inneholde strukturelle brudd, der den underliggende prosessen i tidsserien endret seg slik at den i ettertid følger et annet utviklingsmønster. Dette kan gjøre det nødvendig å avgrense tidsserien til en periode uten brudd. Om dette ikke er mulig eller ønskelig, kan man også korrigere tidsserien for strukturelle brudd ved hjelp av intervensjonsanalyse.

Fig. 4.1.1 og 4.1.2 - Den naturlige logaritmen til KPI og den førstedifferansen av denne.



¹⁰ Man begynte i januar 2000 å innrapportere husleieundersøkelsen månedlig, se Økonomisk Analyse 9/2000. Dette får imidlertid liten betydning for analysen i denne oppgaven.

¹¹ Før denne tiden publiserte Statistisk sentralbyrå (SSB) en levekostnadsindeks som refererte seg til forbruket i arbeiderfamilier i byer og industristeder.

Ved en grafisk inspeksjon av tidsserien ser man at utviklingen i konsumprisindeksen (KPI) har endret seg over tid. I figur 4.1.1 er den naturlige logaritmen av konsumprisindeksen gitt for perioden januar 1960 til desember 1999. Man kan her se at veksten i tidsserien endrer seg rundt 1988. For å se nærmere på dette har jeg, i figur 4.1.2, differensiert denne tidsserien og kortet ned tidsperioden til perioden januar 1983 til desember 1993. Man ser her at gjennomsnittet av den differensierte tidsserien blir lavere i løpet av 1988. Noe som også kommer frem i tabell 4.1.

Tabell 4.1 Deskriptiv statistikk for DLKPI for perioden 1983:01 til 1993:12

Periode	Gjennomsnitt	Standardavvik
1983:01 – 1988:06	0,005765	0,004304
1988:07 – 1893:12	0,002495	0,003571
I hele perioden	0,004130	0,004268

Dette kan tyde på at den underliggende prosessen som bestemmer utviklingen i tidsserien har endret seg. Det kan være flere mulige årsaker til dette. Etter flere år med en ekspansiv finanspolitikk ble det, ved en tilleggs melding til Langtidsprogrammet i 1986, varslet en tilstramning. Myndighetene tok dessuten, mot slutten av 1980-tallet, i bruk en inntektspolitikk som hadde til hensikt å dempe veksten i pris og kostnadsutviklingen¹².

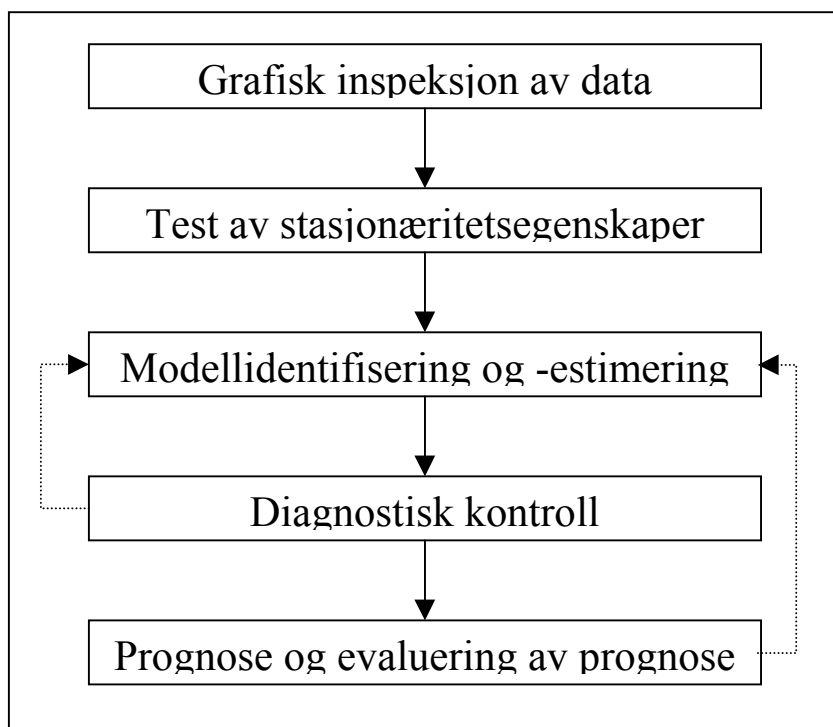
Dette problemet kan håndteres enten ved hjelp av intervensjonsanalyse eller ved å korte ned datamaterialet. Da tidsserien likevel vil være innenfor den vanlige minstekravet ARIMA-modeller setter til antall observasjoner, har jeg sett det som mest hensiktsmessig å velge den siste av disse. Jeg har derfor valgt å basere analysen i denne oppgaven på perioden januar 1988 til desember 1999.

¹² Se NOU 1989:1 *Penger og kredit i en omstillingstid* og SSB Økonomiske analyser nr.1 – 1989: *Økonomisk utsyn over året 1988*.

5. ARIMA - prognoser i praksis

I dette kapittelet vil jeg bruke den generelle ARIMA-modellen jeg har gjort rede for i kapittel 3 til å modellere inflasjonen i Norge. Denne prosessen kan deles i fem steg: *grafisk inspeksjon av data, test av stasjonaritetsegenskaper, modellidentifisering og -estimering, diagnostisk kontroll og prognose og evaluering av prognose*. Figur 1 illustrerer denne prosessen grafisk.

Figur 5.1 – ARIMA-prosedyre



Det første steget er innsamling og grafisk analyse av tidsserien som skal brukes i analysen. I det andre steget tester man om tidsserien er stasjonær eller om det er nødvendig å differensiere den. Når man har en stasjonær tidsserie, kan man forsøke å identifisere og estimere en ARMA-modell til denne. Deretter må de identifiserte modellene testes diagnostisk (vanligvis ved residualtester) og sensitivitetstestes. Om modellene består de diagnostiske testene, kan man bruke disse til å lage prognoser for tidsserien. Det er fordelaktig å spare en del av datamaterialet mot slutten av tidsserien, slik at man kan evaluere hvor godt disse modellene predikerer utviklingen i tidsserien. Legg imidlertid merke til at denne prosessen ikke er en enkel sekvensiell prosess. Det er vanlig å bruke steg fire og fem for å

evaluere de modellene som ble plukket ut i steg tre. De tre siste stegene får derfor en repeterende karakter, der også de to siste stegene inngår som en del av identifiseringsprosessen.

5.1 Innsamling og undersøkelse av data.

En grafisk inspeksjon av dataene vil fortelle oss mye om hvilken type modell man kan vente seg å finne. I den tradisjonelle *Box-Jenkins*-metoden sto dette steget også sentralt i identifiseringsprosessen. På denne tiden ble modelleringsprosessen delt inn i tre stadier: identifisering, estimering og diagnostisk kontroll. Identifisering refererer her til å finne graden p , d og q av $ARIMA(p,d,q)$ -modellen. Til dette brukte man grafisk inspeksjon av tidsserien, autokorrelasjonsfunksjonen (AKF) og den partielle autokorrelasjonsfunksjonen (PAKF).

I figur 5.2 til 5.5 har jeg vist grafen og korrelogrammet av AKF og PAKF for konsumprisindeksen på nivå og differensiert form, med og uten sesongjustering.

Når man ser på grafen i figur 5.2.1, ser man at tidsserien stiger jevnt med tiden. Det er da vanskelig å tenke seg at tidsserien kan ha et konstant gjennomsnitt, og dermed er stasjonær. Dette bli også mer tydelig når man ser på AKF og PAKF i figur 5.2.2 og 5.2.3. Man ser her at AKF er nesten lik én for den første tidsforsinkelsen, og at den så dør langsomt ut når antallet tidsforsinkelser øker, og at PAKF er nesten lik én for den første tidsforsinkelsen. Det betyr at verdien av en observasjon er sterkt avhengig av verdien i forrige periode, noe man også hadde forventet å se for en tidsserie som stiger jevnt med tiden. Jeg forventer derfor ikke at denne tidsserien er stasjonær på nivåform, når jeg skal teste dette i kapittel 5.2.

I figur 5.3 er tidsserien differensiert én gang. Det er her ikke lenger så lett å avgjøre om tidsserien er stasjonær eller ikke, og det må mer formelle tester til for å avgjøre dette. Når man ser på AKF og PAKF, ser man to ting: tidsserien inneholder en markant sesongkomponent med seks måneders mellomrom, og at det er ingen utpregede spor av hverken $AR(p)$ -prosesser eller $MA(q)$ -prosesser utenom sesongkomponentene.

Fig. 5.2.1: Logaritmen av kpi: lkpi.

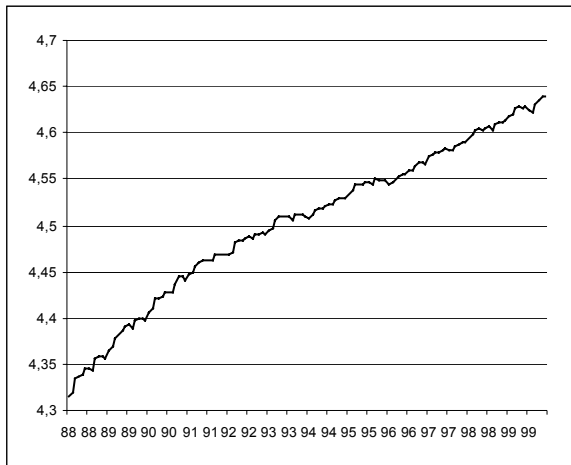


Fig. 5.3.1: 1.diff. av lkpi: dlkpi.

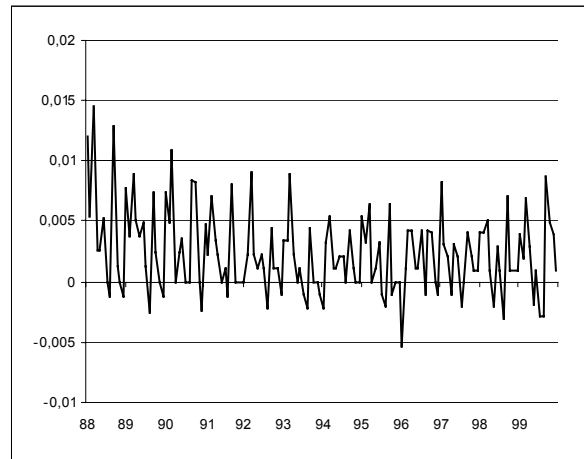


Fig. 5.2.2: AKF av lkpi.

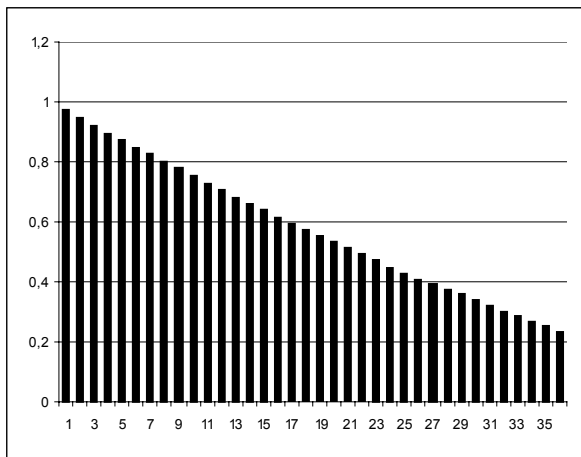


Fig. 5.3.2: AKF av dlkpi.

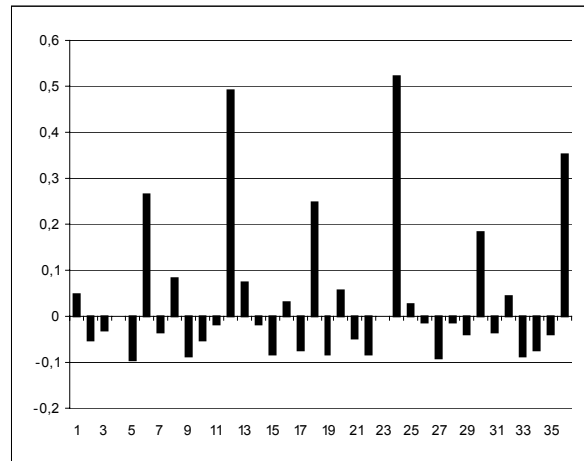


Fig. 5.2.3: PAKF av lkpi

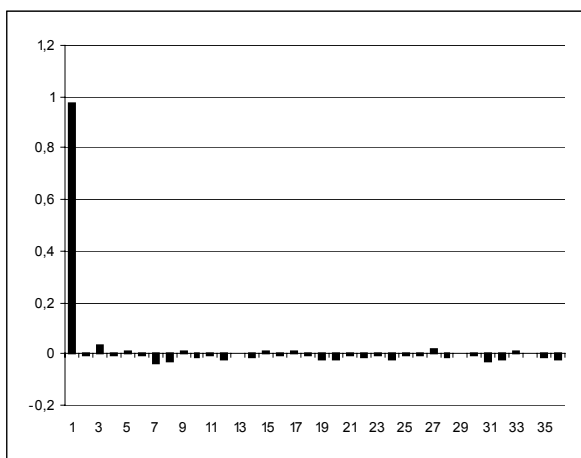
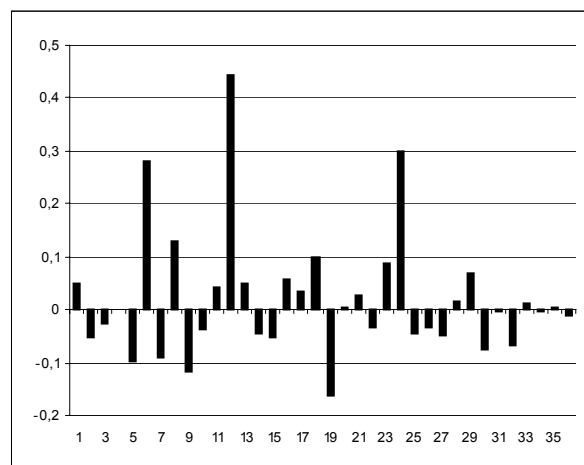
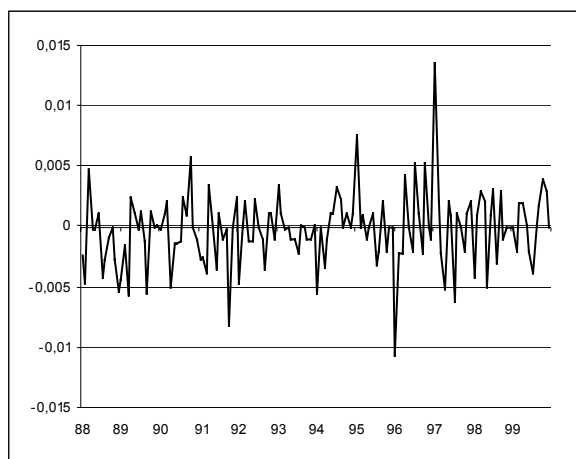


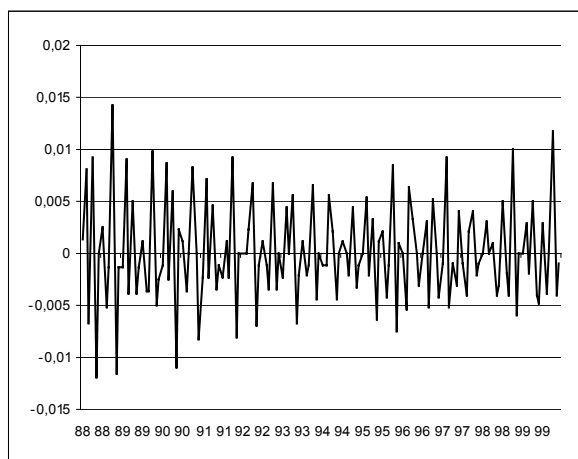
Fig. 5.3.3: PAKF av dlkpi.



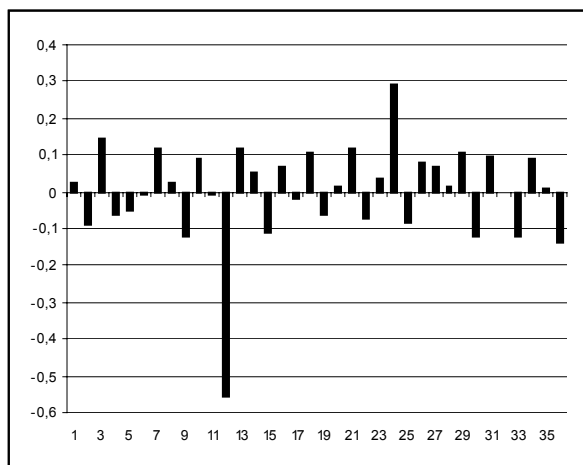
Figur 5.4.1: Sesongdiff. av dlkpi: dsdlkpi.



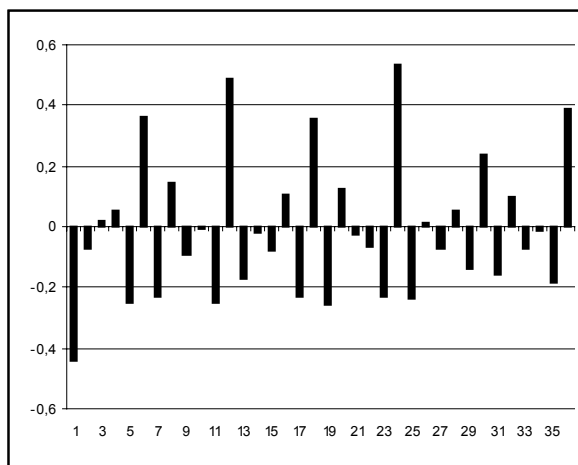
Figur 5.5.1: 2.diff. av lkpi: ddklpi.



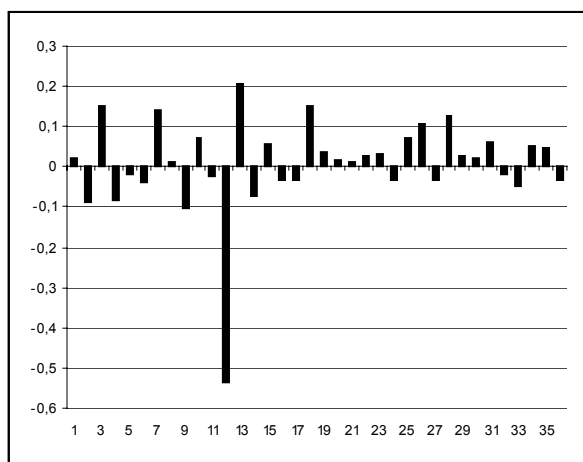
Figur 5.4.2: AKF av dsdlkpi.



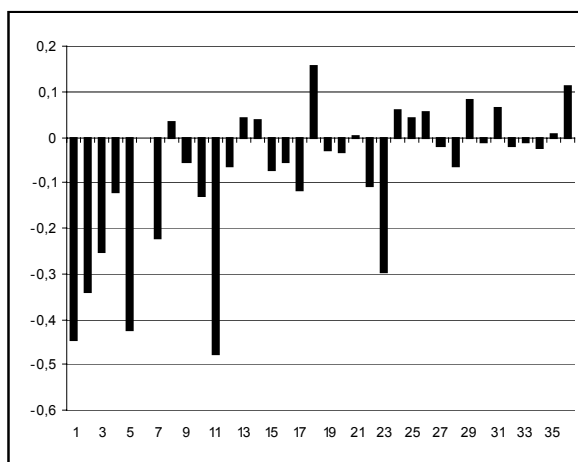
Figur 5.5.2: AKF av ddklpi.



Figur 5.4.3: PAKF av dsdlkpi



Figur 5.5.3: PAKF av ddklpi.



I figur 5.4 har jeg gjengitt grafen og korrellogrammene til den sesongdifferensierte av den førstedifferensierte tidsserien. Ved grafisk inspeksjon ser denne tidsserien ut til å være stasjonær, men man kan også se av korrellogrammene at det er et stort utslag ved den tolvte tidsforsinkelsen. Dette kan tyde på at det også vil bli behov for å inkludere flere sesongledd i modellen. I figur 5.4.1 ser man også at det er relativt store utslag for januar 1996 og januar 1997 i forhold til resten av tidsserien. Dette må trolig ses i sammenheng med en reduksjon i bilavgiftene med virkning fra januar 1996 og økte avgifter og elektrisitetspriser rundt årsskiftet 1996/1997¹³. Jeg vil derfor senere korrigerere for disse avvikene ved hjelp av intervasjonsanalyse.

Figur 5.5 viser andredifferansen av KPI. Man ser her at utslagene av AKF og PAKF er større enn i figur 5.3 der tidsserien bare var differensiert én gang. Dette kan tyde på at å andredifferensiere denne tidsserien vil resultere i en overdifferensiering.

Etter en grafisk inspeksjon av tidsserien kommer det fram at konsumprisindeksen sannsynligvis må differensieres og sesongdifferensieres for å bli stasjonær. Det er dessuten få tegn på av man vil finne AR(p)- eller MA(q)-prosesser i tidsserien. Tidsserien inneholder imidlertid en sterk sesongkomponent, det er derfor sannsynlig at modellen må inneholde flere sesongledd, i tillegg til sesongdifferensiering.

5.2 Test av stasjonaritetsegenskaper

For å kunne identifisere en passende ARIMA-modell, må man først identifisere stasjonaritetsegenskapene til tidsserien. Det vil si at man for en generell ARIMA(p,d,q)x(P,D,Q) må identifisere d og D .

Den mest anvendte testen for stasjonaritet er den såkalte Dickey-Fuller-testen (DF-testen). Denne testen antar som nullhypotese at tidsserien er ikke-stasjonær, mens alternativhypotesen er at den er stasjonær. Dickey-Fuller-testen tar utgangspunkt i prosessen

$$(5.1) \quad y_t = \rho y_{t-1} + \varepsilon_t$$

¹³ Se SSBs månedlige kommentar til konsumprisindeksen.

der restleddet ε_t er hvit støy. Nullhypotesen er at $\rho = 1$ mot alternativhypotesen $|\rho| < 1$. Under nullhypotesen er (5.1) identisk med *random walk*-prosessen beskrevet ved (3.1). Forkastes nullhypotesen, er prosessen som generer y_t lik

$$(5.2) \quad y_t = \sum_{i=0}^{t-1} \rho^i \varepsilon_{t-i}$$

Fordi den stokastiske innovasjonen i periode t har mindre vekt jo større i blir, er prosessen stasjonær. Siden prosessen har en enhetsrot når $\rho = 1$, kalles testen en enhetsrottest.

DF-testen baserer seg på estimering av en regresjonslikning som er ekvivalent med (5.1). Trekkes y_{t-1} fra begge sider i (5.1) får vi:

$$(5.3) \quad \Delta y_t = (\rho - 1)y_{t-1} + \varepsilon_t.$$

Nullhypotesen er her $(\rho - 1) = 0$ mot alternativhypotesen $(\rho - 1) < 0$.

For å teste om enhetsroten eksisterer, kan det virke naturlig å anvende en vanlig t -test. Dette blir imidlertid ikke riktig. En t -test vil under nullhypotesen gi en forventingskjev testobservator nedover, noe som fører til at en kan feilaktig slutte at $\rho < 1$ selv om $\rho = 1$. Man må derfor bruke en egen fordeling, beregnet av Dickey og Fuller (1976). Dersom alternativhypotesen ikke kan forkastes, har vi en AR(1)-prosess som fluktuerer rundt et gjennomsnitt lik null. DF-testen kan også ta hensyn til modeller som inkluderer deterministiske komponenter som konstant, trend og sesong.

Dickey-Fuller-testen slik den er presentert så langt, har en svakhet: Den tar ikke hensyn til eventuell autokorrelasjon i restleddet, ε_t . Dersom det er autokorrelasjon til stede, noe som er svært sannsynlig i en tidsserieanalyse, vil minste kvadrats metode (MKM) estimatene for ligning (5.3) være ineffisiente. Dickey og Fuller løser dette problemet ved å inkludere tidsforsinkede venstresidevariabler som ekstra forklaringsvariabler. Denne testen går under navnet utvidet Dickey-Fuller-test (ADF-test av det engelske "Augmented Dickey-Fuller test").

I en modell der både konstant og trend er inkludert, er ADF-testen gitt ved

$$(5.4) \quad \Delta y_t = \alpha_0 + \alpha_1 t + \rho * y_{t-1} + \sum_{i=1}^k \varphi_i \Delta y_{t-i} + \varepsilon_t$$

Et spørsmål som må avklares før testen kan utføres, er hvor mange perioder Δy_t skal tidsforsinkes med. En gylden regel er å velge k (antall perioder tidsforsinkelse) så lav som mulig for ikke å miste frihetsgrader, men samtidig stor nok til å fjerne eventuell autokorrelasjon i residualen. Campell og Perron (1991, s. 15) anbefaler at man *a priori* bestemmer en øvre grense for k , k -max, for så å velge den lengste tidsforsinkelsen der den siste ekstra forklaringsvariabelen opptrer signifikant. Siden jeg opererer med månedsdata er k -max satt til 13, slik at det er tatt hensyn til deterministiske sesongpåvirkninger. For å finne den siste ekstra forklaringsvariabelen som opptrer signifikant, har jeg testet restleddet for autokorrelasjon med en Lagrange Multiplikator (LM) test. Om det viser seg at den siste ekstra forklaringsvariabelen opptrer signifikant, er dette blitt fjernet og LM-testen repetert inntil den siste ekstra forklaringsvariabelen ikke opptrer signifikant lenger. Jeg har så satt k i ADF-testen slik at den siste ekstra forklaringsvariabelen opptrer signifikant.

Resultatet av ADF-testen er vist i tabell 5.1. Det er ikke gitt hvilke deterministiske komponenter som bør inkluderes i testen. Flere alternativer er derfor rapportert. Symbolene * og ** indikerer at nullhypotesen om enhetsrot er forkastet med en p -verdi på henholdsvis fem og en prosent. Tallene i parentes angir k .

Tabell 5.1: ADF- tester

Variabel- definisjon	ADF		ADF med konstant		ADF med konstant og trend	
LKPI	1,978	(12)	1,286	(12)	-3,053	(12)
DLKPI	-1,324	(11)	-2,298	(11)	-2,245	(11)
DDLKPI	-6,703**	(12)	-6,712**	(12)	-9,726**	(10)
DsDLKPI	-5,857**	(11)	-6,123**	(11)	-7,266**	(11)

ADF-testene bekrefter at tidsserien ikke er stasjonær på nivåform, slik det ble konkludert fra den grafiske inspeksjonen i kapittel 5.1. Testene viser også at det ikke er nok å differensiere tidsserien én gang. Blir tidsserien derimot differensiert to ganger eller både første- og sesongdifferensiert er tidsserien stasjonær, uavhengig av om de deterministiske

komponentene inkluderes i testen. Da tidsserien viste tydelige tegn til å inneholde en sesongkomponent i kapittel 5.1, vil jeg i resten av oppgaven benytte den første- og sesongdifferensiere tidsserien (DsDLKPI). Jeg velger deterministiske komponenter ut i fra hva som er hensiktsmessig.

5.3 Modellidentifisering og -estimering

Når man har oppnådd en stasjonær tidsserie w_t , kan man bruke denne til å identifisere en passende ARMA-modell. Det er hovedsakelig to måter dette kan gjøres på. Den tradisjonelle metoden for modellidentifisering fra Box-Jenkins prosedyren benytter stort sett grafisk inspeksjon av autokorrellogrammer og partielle autokorrellogrammer, slik det er vist i kapittel 3. Denne metoden er senere blitt oppfattet som både subjektiv og vanskelig å bruke. Det er derfor senere utviklet flere statistiske metoder for å gjøre identifiseringsprosedyren lettere og mer objektiv.

5.3.1 Box-Jenkins-prosedyren

Box-Jenkins-prosedyren involverer hovedsakelig grafisk analyse av tidsseriens autokorrellogram og partielle autokorrellogram til å identifisere den beste ARMA-modellen. I tillegg benyttes diagnostisk testing og evaluering av prognoseegenskaper som en aktiv del av identifikasjonsprosessen. Som jeg har vist i kapittel 3, er Box-Jenkins modellidentifisering uproblematisk hvis man har en ren AR(p)-prosess eller en ren MA(q)-prosess. Det blir imidlertid adskillig mer komplisert når man har en blandet ARMA-modell, spesielt av høyere orden, da det kan bli svært vanskelig å tolke AKF og PAKF.

5.3.2 En ren AR(p)-prosess

I kapittel 3.2.4 har vi sett at autokorrelasjonen av en ren AR(p)-prosess synker gradvis etter hvert som antall tidsforsinkelser øker. Det er derfor ikke mulig å bruke et autokorrellogram for å skille mellom en ren AR(3)-modell og en ren AR(4)-modell. Det partielle autokorrellogrammet, derimot, faller til null etter p tidsforsinkelser. Det vil si at PAKF av en ren AR(3)-prosess er lik null etter tre tidsforsinkelser, mens en ren AR(4)-prosess er lik null etter fire tidsforsinkelser.

Det vil si at for en ren AR(p)-prosess skal den teoretiske AKF og PAKF være:

$$\begin{aligned} \text{AKF}(i) &\neq 0 \forall i \\ \text{PAKF}(i) &\neq 0 \forall i=1, \dots, p \\ \text{PAKF}(i) &= 0 \forall i > p \end{aligned}$$

hvor i angir antall tidsforsinkelser.

5.3.3 En ren MA(q)-prosess

Autokorrellogrammet og det partielle autokorrellogrammet av en ren MA(q)-prosess vil oppføre seg akkurat motsatt av korrellogrammene til en ren AR(p)-prosess. Autokorrellogrammet til en ren MA(q)-prosess vil synke til null etter q tidsforsinkelser. Det partielle autokorrellogrammet av en ren MA(q)-prosess, derimot, vil synke langsomt over tid. Det vil si at det vil være umulig å skille PAKF av en MA(3)- og en MA(4)-prosess, mens AKF av en MA(3)-prosess vil falle til null etter tre tidsforsinkelser, og en MA(4)-prosess vil falle til null etter fire tidsforsinkelser.

Det vil si at for en ren MA(q)-prosess skal den teoretiske AKF og PAKF være:

$$\begin{aligned} \text{AKF}(i) &\neq 0 \forall i = 1, \dots, q \\ \text{AKF}(i) &= 0 \forall i > q \\ \text{PAKF}(i) &\neq 0 \forall i \end{aligned}$$

Det skulle derfor være relativt enkelt å identifisere rene AR(p)- eller MA(q)-prosesser.

5.3.4 Blandede ARMA-prosesser

Beklageligvis er modellidentifikasjon mye mer komplisert for en blandet ARMA-prosess. Mønsteret til AKF og PAKF av en høyere ordens ARMA-modell er ofte svært vanskelig å tolke. Derfor vil modellidentifikasjon ved bruk av *Box-Jenkins*-prosedyren ofte bli en gjentakende prosess, der man bruker diagnostisk kontroll og evaluering av prognoseegenskaper for å velge ut hvilke modeller som skal undersøkes videre. Dette medfører at modellidentifisering under den opprinnelige *Box-Jenkins*-prosedyren ofte får et subjektivt preg og er vanskelig å bruke. Dette problemet har tidsserieanalytikere senere forsøkt å løse ved å finne alternative, mer formelle statistiske metoder for å identifisere ARMA-modeller.

En måte å gjøre dette på er å sammenligne modellenes føyningsgrad, ved å sammenligne residualvariansen av de konkurrerende modellene. Om man betegner residualen av en estimert ARMA(p, q)-modell ved $\hat{\varepsilon}_i(p, q)$, er residualvariansen $\hat{\sigma}_{p,q}^2$ gitt ved

$$(5.5) \quad \hat{\sigma}_{p,q}^2 = \frac{\sum_{i=1}^N \hat{\varepsilon}_i^2(p, q)}{N}$$

Men en enkel sammenligning av residualvariansen av konkurrerende modeller vil favorisere modeller med mange forklaringsvariabler fremfor enkle, lavere-ordens modeller. Dette er et kjent fenomen fra regresjonsanalysen, hvor det å inkludere flere forklaringsvariabler typisk vil bedre modellens føyning. Man vil her kunne oppnå en perfekt modell ved å inkludere like mange forklaringsvariabler som vi har observasjoner. Disse modellene vil imidlertid være overtilpasset til datasettet ved at de også inkluderer ikke-systematiske trekk ved tidsserien. Om man ønsker å estimere modeller som kan forklare tidsserien på generell form, må modellene begrenses til bare å inneholde den systematiske informasjon i tidsserien.

Der er blitt introdusert flere utvelgelseskriterier hvor man prøver å overkomme problemet med *overspesifisering*, som beskrevet ovenfor. Et av de meste brukte kriteriene i denne sammenheng er Akaike Information Criterion (AIC), som er gitt ved

$$(5.6) \quad AIC = \log\left(\frac{\sum \hat{\varepsilon}_i^2}{N}\right) + \frac{2k}{N}$$

der $\sum \hat{\varepsilon}_i^2$ er kvadratsummen av residualene, k er antallet estimerte parametre og N er antall observasjoner. Her måler det første leddet modellens føyning, mens det andre leddet er et ”straffeled” som øker med antall parametre. Hensikten med straffeledet er å unngå overspesifiserte modellene. Man skal her velge den modellen som minimerer verdien av AIC.

Monte Carlo-simulering rapportert av Shibata (1976) indikerer at til tross for straffeledet har AIC en tendens til å favorisere overspesifiserte modeller. Det kan derfor være ønskelig å bruke et kriterium som inneholder et sterkere straffeled.

Schwarz-kriteriet (SC) er gitt ved

$$(5.7) \quad SC = \log\left(\frac{\sum \hat{\epsilon}_i^2}{N}\right) + \frac{k \log N}{N}$$

der $\sum \hat{\epsilon}_i^2$ er kvadratsummen av restleddene, k er antall estimerte parametre og N er antall observasjoner. Hvis $\log N > 2$ vil Schwarz-kriteriet gi en tyngre straff, for hver ekstra parameter man legger til, enn AIC.

Box, Jenkins og Reinsel (1994, s. 16) vektlegger at man bør være så sparsom som mulig med antall parametre i en ARIMA-modell, definert ved *"the smallest number of parameters for adequate representation"* av datasettet. De vektlegger dette prinsippet fordi hovedmålet med ARIMA-modellering er å predikere den fremtidige utviklingen i tidsseien. Det er da viktig å rendyrke systematiske trekk i tidsserien, fremfor å legge til ekstra parametre for å øke forklaringsgraden av de historiske verdiene av tidsserien. Et godt eksempel på denne tanken er den såkalte Airline-modellen¹⁴. Jeg vil på denne bakgrunn begrense antall AR(p)- og MA(q)-ledd til fire av hver, pluss sesongledd, når jeg skal identifisere hvilke modeller som passer tidsserien best. Jeg vil også vektlegge Schwarz-kriteriet i denne prosessen, da dette kriteriet favoriserer sparsomme modeller.

Jeg har estimert ARIMA-modeller for tidsserien DsDLKPI. Resultatet av disse testene er gjengitt i Appendiks 1. Det første man legger merke til er at AIC og SC rangerer modellene forskjellig. SC anbefaler modeller med færre parametre enn AIC. Det andre man legger merke til er at selv om det er én modell som peker seg ut med lavest verdi for hver av de to kriteriene, er det flere modeller som får tilnærmet likt utslag. Med utgangspunkt i dette problemet foreslo Poskitt og Tremayne (1987) å bruke en portefølje av modeller. Dette medfører at man ikke bare går videre med den modellen som fikk lavest verdi ved informasjonskriteriene, men en portefølje av modeller som gjorde det tilnærmet like bra. Man kan så skille mellom disse modellene etter hvilke prognoseegenskaper de har.

¹⁴ En ARIMA(0,1,1)x(0,1,1)-modell blir i litteraturen betegnet som en Airline-modell fordi den ble brukt av Box og Jenkins i et studie av passasjertrafikk på fly. Se Box, Jenkins og Reinsel (1994).

Tabell 5.2 – ARIMA(p,d,q)x(P,D,Q).

Rangering	Modell	AIC	Modell	SC
1	(3,1,0)x(1,1,1)	-9,493022	(0,1,0)x(1,1,1)	-9,423096
2	(4,1,2)x(1,1,1)	-9,491443	(0,1,0)x(0,1,1)	-9,411376
3	(3,1,4)x(1,1,1)	-9,487839	(1,1,0)x(1,1,1)	-9,404231

I tabell 5.2 har jeg gjengitt de tre modellene som oppnådde lavest verdi i AIC og i SC-testen, i rangert rekkefølge. Om man ser på de to første modellene som er valgt ut ved hjelp av Schwarz-kriteriet, ser man at disse utelukkende inneholder sesongledd. Dette samsvarer godt med de slutninger som ble trukket under den grafiske inspeksjonen.

Tabell 5.3 – ARIMA(p,d,q)x(P,D,Q) med et konstantledd.

Rangering	Modell	AIC	Modell	SC
1	(3,1,0)x(1,1,1)	-9,480268	(0,1,0)x(0,1,1)	-9,402357
2	(3,1,4)x(1,1,1)	-9,475380	(0,1,0)x(1,1,1)	-9,398441
3	(2,1,0)x(1,1,1)	-9,473637	(0,1,1)x(1,1,1)	-9,374037

Jeg har så gjennomført den samme prosessen en gang til, men denne gangen har jeg lagt til et konstantledd. Som beskrevet i kapittel 3.2.7, vil et konstantledd i en ARIMA-modell utgjøre en deterministisk trend. Av resultatene i tabell 5.3 ser vi at et slikt trendledd ikke forbedrer modellenes forklaringssevne.

Tabell 5.4 – ARIMA(p,d,q)x(P,D,Q) med du1 og du2.

Rangering	Modell	AIC	Modell	SC
1	(3,1,0)x(1,1,1)	-9,674932	(0,1,0)x(1,1,1)	-9,545787
2	(4,1,2)x(1,1,1)	-9,665304	(0,1,1)x(1,1,1)	-9,516536
3	(4,1,0)x(1,1,1)	-9,662748	(1,1,0)x(1,1,1)	-9,514963

I kapittel 5.1 og figur 5.4.1 har jeg påpekt at det er avvikende verdier for januar 1996 (du1) og januar 1997 (du2). I tabell 5.4 har jeg gjengitt hva resultatene av identifikasjonsprosessen blir, om jeg korrigerer for disse ved hjelp av intervensjonsanalyse. Man ser at dette forbedrer resultatene.

Tabell 5.5 – ARIMA(p,d,q)x(P,D,Q) med konstant, du1 og du2.

Rangering	Modell	AIC	Modell	SC
1	(3,1,0)x(1,1,1)	-9,665737	(0,1,0)x(1,1,1)	-9,528915
2	(4,1,2)x(1,1,1)	-9,655437	(0,1,0)x(0,1,1)	-9,527835
3	(4,1,0)x(1,1,1)	-9,653071	(0,1,1)x(1,1,1)	-9,494455

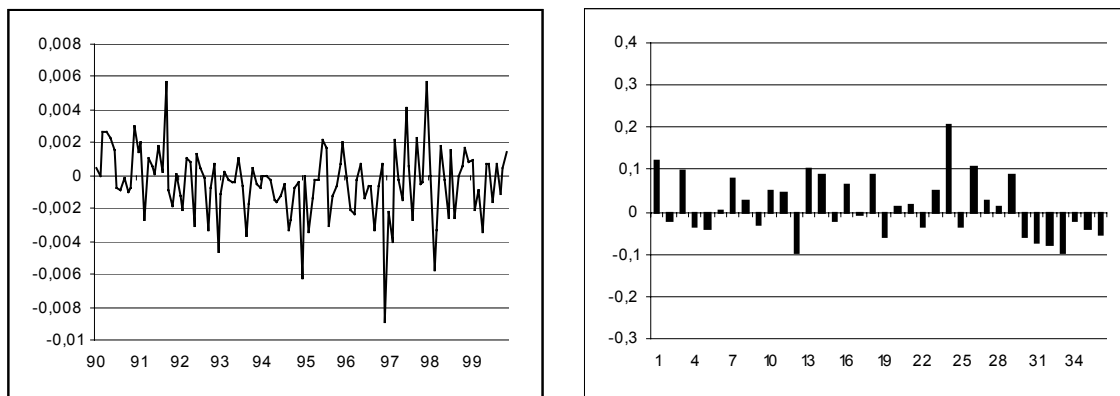
I tabell 5.5 har jeg gjengitt resultatet av analysen om man tar i bruk både et konstantledd og intervensjonsanalyse. Heller ikke her forbedrer bruken av et konstantledd resultatene av analysen.

5.4 Diagnostisk kontroll av modellen

Når man har identifisert modellene man ønsker å bruke, må man kontrollere at disse modellene oppfyller kravene for univariate tidsseriemodeller. Det vil si at man kontrollerer at modellene har klart å trekke ut all den systematiske informasjonen i tidsserien og at det bare er usystematisk informasjon igjen i restleddet. Dette gjør man ved å teste om restleddet oppfyller kravet til en *hvit-støy*-prosess, se kapittel 3.1. Det er et stort antall diagnostiske tester man kan bruke til dette formålet.

En grafisk inspeksjon av de estimerte modellenes residualer er en god begynnelse. Det vil kunne gi en god indikasjon på om det finnes avvikende observasjoner som kan påvirke de estimerte parameterne, og om det fremdeles er autokorrelasjon igjen i restleddet.

Figur 5.2 – Residualen og korrellogrammet av residualen av ARIMA(0,1,0)x(1,1,1)



I figur 5.2 har jeg vist grafen og korrellogrammet for modellen ARIMA(0,1,0)x(0,1,1). Figuren tyder på at residualen følger en *hvit-støy*-prosess, men for å avgjøre dette med sikkerhet må man bruke formelle statistiske tester. Her finnes det flere å velge mellom, men jeg har her valgt å bruke en Ljung-Box Q-test.

Ljung-Box Q-test er gitt ved:

$$(5.8) \quad Q = T(T+2) \sum_{k=1}^K \frac{\hat{r}_k^2}{T-k} \quad \text{hvor} \quad \hat{r}_k = \frac{\sum_{t=k+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{j=1}^T \hat{\varepsilon}_j^2}.$$

Q-testen tester om residualen følger en *hvit støy*-prosess eller ikke. Testresultatet følger en *kjikkvadratfordeling* med k frihetsgrader. Eviews oppgir denne verdien sammen med sannsynligheten Prob(Q-test) for at det ikke er autokorrelasjon igjen i feilleddet. Det er vanlig å bruke en kritisk verdi på 10 prosent for denne testen¹⁵, det vil si en Prob(Q - test) $\geq 0,1$.

Tabell 5.6 – ARIMA(p,d,q)x(P,D,Q).

Rangering	Modell (AIC)	Prob(Q-test)	Modell (SC)	Prob(Q-test)
1	(3,1,0)x(1,1,1)	0,951	(0,1,0)x(1,1,1)	0,817
2	(4,1,2)x(1,1,1)	0,804	(0,1,0)x(0,1,1)	0,779
3	(3,1,4)x(1,1,1)	0,748	(1,1,0)x(1,1,1)	0,835

Tabell 5.7 – ARIMA(p,d,q)x(P,D,Q) med et konstantledd.

Rangering	Modell (AIC)	Prob(Q-test)	Modell (SC)	Prob(Q-test)
1	(3,1,0)x(1,1,1)	0,953	(0,1,0)x(0,1,1)	0,752
2	(3,1,4)x(1,1,1)	0,735	(0,1,0)x(1,1,1)	0,843
3	(2,1,0)x(1,1,1)	0,903	(0,1,1)x(1,1,1)	0,883

Tabell 5.8 – ARIMA(p,d,q)x(P,D,Q) med du1 og du2.

Rangering	Modell (AIC)	Prob(Q-test)	Modell (SC)	Prob(Q-test)
1	(3,1,0)x(1,1,1)	0,491	(0,1,0)x(1,1,1)	0,134
2	(4,1,2)x(1,1,1)	0,319	(0,1,1)x(1,1,1)	0,152
3	(4,1,0)x(1,1,1)	0,496	(1,1,0)x(1,1,1)	0,103

Tabell 5.9 – ARIMA(p,d,q)x(P,D,Q) med konstant, du1 og du2.

Rangering	Modell (AIC)	Prob(Q-test)	Modell (SC)	Prob(Q-test)
1	(3,1,0)x(1,1,1)	0,498	(0,1,0)x(1,1,1)	0,216
2	(4,1,2)x(1,1,1)	0,356	(0,1,0)x(0,1,1)	0,438
3	(4,1,0)x(1,1,1)	0,548	(0,1,1)x(1,1,1)	0,212

¹⁵ Se Pindyck og Rubinfeld (1998).

I tabell 5.6 til 5.9 har jeg rapportert alle modellene fra tabell 5.2 til 5.5 med Prob(Q-test). Man ser i tabell 5.6 og 5.7 at modellene passerer Q - testen uten problemer og at det ikke spiller noe rolle hvorvidt det er et konstantledd med eller ikke.

I tabell 5.8 ser man at verdien av Prob(Q-test) synker betraktelig når man tar i bruk intervensjonsanalyse for å korrigere for avvikende observasjoner. Dette er spesielt tydelig for de modellene som er plukket ut ved hjelp av Schwarz-kriteriet, der verdien av Prob(Q-test) ligger akkurat over grensen for det som er godkjent.

I tabell 5.9 ser vi at resultatene for tabell 5.8 blir bedre om man legger til et konstantledd, men resultatene er fremdeles på langt nær så gode som de jeg har rapportert i tabell 5.6 og 5.7.

5.5 Prognose og evaluering av prognose

Det siste steget i prosessen er å lage en prognose $\hat{y}_T(l)$, l perioder frem fra periode T , for konsumprisindeksen. Målet er å predikere fremtidige verdier for tidsserien med så lite feil som mulig. For å gjøre dette, antar vi at den optimale prognosen vil være den prognosen som minimerer gjennomsnittlig kvadrert prognosefeil. Men siden prognosefeilen er en stokastisk variabel, minimerer vi den forventede verdien, hvilket vil si at vi ønsker å velge prognosen $\hat{y}_T(l)$ slik at forventet prognosefeil $E[e_T^2(l)] = E\{[y_{T+l} - \hat{y}_T(l)]^2\}$ minimeres. Denne prognose er gitt ved den betingete forventningen av y_{T+l} :

$$(5.9) \quad \hat{y}_T(l) = E(y_{T+l} | y_T, y_{T-1}, \dots, y_1)$$

For å bevise at minimum gjennomsnittlig kvadrert prognosefeil er gitt ved ligning (5.9), tar man utgangspunkt i ARIMA-modellen fra ligning (3.70). Denne kan skrive som

$$(5.10) \quad \phi(B)(1-B)^d y_t = \theta(B)\varepsilon_t$$

siden $\Delta = 1 - B$, som man så kan løse for y_t

$$(5.11) \quad y_t = \phi^{-1}(B)(1-B)^{-d} \theta(B)\varepsilon_t = \psi(B)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

Her er ARIMA-modellen uttrykt som en ren glidende gjennomsnittsprcess av uendelig grad¹⁶. Dette gir

$$(5.12) \quad y_{T+l} = \psi_0 \varepsilon_{T+l} + \psi_1 \varepsilon_{T+l-1} + \cdots + \psi_l \varepsilon_T + \psi_{l+1} \varepsilon_{T-1} + \cdots \\ = \psi_0 \varepsilon_{T+l} + \psi_1 \varepsilon_{T+l-1} + \cdots + \psi_{l-1} \varepsilon_{T+1} + \sum_{j=0}^{\infty} \psi_{l+j} \varepsilon_{T-j}$$

I likning (5.12) deler man den uendelige summen i to deler, der den andre delen begynner med leddet $\psi_l \varepsilon_T$ og beskriver dermed informasjonen opp til og med periode T . Prognosen $\hat{y}_T(l)$ kan selvfølgelig bare være basert på informasjon opp til tidspunkt T . Målet er her å sammenligne de predikerte verdiene med de faktiske verdiene y_{T+l} som er uttrykt i ligning (5.12). For å gjøre dette, setter vi opp prognosen som en veid sum av de restleddene vi kan estimere, $\varepsilon_T, \varepsilon_{T-1}, \dots$. Dette gir

$$(5.13) \quad \hat{y}_T(l) = \sum_{j=l}^{\infty} \psi_{l+j}^* \varepsilon_{T-j}$$

hvor restleddet er veid optimalt ved ψ_{l+j}^* for å minimere gjennomsnittlig kvadrert prognosefeil. Ved å kombinere ligning (5.11) og (5.12) kan vi nå finne et uttrykk for prognosefeilen, $e_T(l)$,

$$(5.14) \quad e_T(l) = y_{T+l} - \hat{y}_T(l) = \psi_0 \varepsilon_{T+l} + \psi_1 \varepsilon_{T+l-1} + \cdots + \psi_{l-1} \varepsilon_{T+1} + \sum_{j=0}^{\infty} (\psi_{l+j} - \psi_{l+j}^*) \varepsilon_{T-j}$$

Siden modellen forutsetter at $E(\varepsilon_i \varepsilon_j) = 0$ for $i \neq j$, er den gjennomsnittlige kvadrerte prognosefeilen gitt ved

$$(5.15) \quad E(e_T^2(l)) = (\psi_0^2 + \psi_1^2 + \cdots + \psi_{l-1}^2) \sigma_\varepsilon^2 + \sum_{j=0}^{\infty} (\psi_{l+j} - \psi_{l+j}^*)^2 \sigma_\varepsilon^2$$

¹⁶ Enhver ARIMA-modell kan uttrykkes som en ren glidende gjennomsnittsprcess eller en ren autoregressiv prcess av uendelig grad, men man kan ikke bruke dette i praksis fordi man da ville ha et uendelig antall parametre å estimere.

Dette uttrykket blir minimert ved å sette de optimale vektene ψ_{l+j}^* lik de faktiske vektene ψ_{l+j} for $j = 0, 1, \dots$. Men da er den optimale prognosen $\hat{y}_T(l)$ gitt ved den betingede forventningen av y_{T+l} . Dette kan bli vist ved å ta den betingete forventningen av y_{T+l} i ligning (5.12). De forventede verdiene av $\varepsilon_T, \dots, \varepsilon_{T+1}$ er her alle null, mens de forventede verdiene av $\varepsilon_T, \varepsilon_{T-1}, \dots$, er residualene fra den estimerte ligningen. Dette gir at

$$(5.16) \quad \hat{y}_T(l) = \sum_{j=0}^{\infty} \psi_{l+j} \hat{\varepsilon}_{T-j} = E(y_{T+l} | y_T, \dots, y_1)$$

Dette gir det underliggende prinsippet for hvordan man kalkulerer prognoser fra ARIMA-modellene.

5.5.1 Beregning av prognoser basert på ARIMA-modeller

Beregningen av prognosen $\hat{y}_T(l)$ kan i praksis bli gjort ved gjentagende bruk av den estimerte ARIMA-modellen. Det vil si at man først beregner en prognose én periode frem i tid, for så å legge denne prognosen til tidsserien. Man bruke så den nye tidsserien til å beregne en prognose to perioder frem i tid. Slik fortsetter man inntil man har oppnådd en prognose for l perioder. Om vi skriver en ARIMA(q,d,p)-modell uten sesong, der w_t er den transformerte tidsserien av y_t , som

$$(5.17) \quad w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} + \delta \quad \text{der} \quad y_t = \sum^d w_t.$$

For å beregne prognosen $\hat{y}_T(l)$, begynner man med å beregne en prognose én periode frem i tid for w_t , $\hat{w}_T(1)$. Dette gjør vi ved å skrive ligning (5.17) som

$$(5.18) \quad w_{T+1} = \phi_1 w_T + \dots + \phi_p w_{T-p+1} + \varepsilon_{T+1} - \theta_1 \varepsilon_T - \dots - \theta_q \varepsilon_{T-q+1} + \delta$$

Vi kan så bruke ligning (5.18) til å lage en prognose én periode frem i tid

$$(5.19) \quad \hat{w}_T(1) = E(w_{T+1} | w_T, \dots) = \phi_1 w_T + \dots + \phi_p w_{T-p+1} - \theta_1 \hat{\varepsilon}_T - \dots - \theta_q \hat{\varepsilon}_{T-q+1} + \delta$$

der $\hat{\varepsilon}_T, \hat{\varepsilon}_{T-1}, \dots$ er observerte residualer. Legg merke til at den forventede verdien av ε_{T+1} er null. Man kan nå bruke prognosen i (5.17) til å lage en prognose $\hat{w}_T(2)$ som går to perioder frem i tid:

$$(5.20) \quad \hat{w}_T(2) = E(w_{T+2} | w_T, \dots) = \phi_1 \hat{w}_T(1) + \phi_2 w_T + \dots + \phi_p w_{T-p+2} - \theta_2 \hat{\varepsilon}_T - \dots - \phi_q \hat{\varepsilon}_{T-q+2} + \delta$$

Denne prognosen kan så brukes til å produsere en prognose tre perioder frem i tid, og så videre inntil man har en prognose $\hat{w}_T(l)$ som går l perioder frem i tid:

$$(5.21) \quad \hat{w}_T(l) = \phi_1 \hat{w}_T(l-1) + \dots + \phi_l w_T + \dots + \phi_p w_{T-p+1} - \theta_l \hat{\varepsilon}_T - \dots - \theta_q \hat{\varepsilon}_{T-q+l} + \delta$$

Når man har laget en prognose for den differensierte serien w_t , kan man lage en prognose for den opprinnelige tidsserien y_t ved å bruke summeringsoperatoren til w_t , det vil si å summere w_t d ganger. Sett for eksempel at $d = 1$. Da vil prognosen for y_t over l perioder være gitt ved

$$(5.22) \quad \hat{y}_T(l) = y_T + \hat{w}_T(1) + \hat{w}_T(2) + \dots + \hat{w}_T(l).$$

5.5.2 Prognoseevaluering

For å vurdere modellens *ex post*-prognoseegenskaper er det fordelaktig å spare en del observasjoner mot slutten av tidsserien, som ikke blir brukt til å estimere modellen. Man lager her en prognose basert på den estimerte modellen for så å sammenligne de predikerte verdiene med de faktiske verdiene for denne perioden. Det finnes et utvalg av statistiske mål for å evaluere hvor godt de predikerte verdiene samsvarer med faktiske observasjonene. Et mål som ofte blir brukt, er rms (root-mean-square) prognosefeil. For variabelen y_t er rms gitt ved

$$(5.23) \quad \text{rms prognosefeil} = \sqrt{\frac{1}{F} \sum_{l=1}^F (\hat{y}_T(l) - y_{T+l})^2}$$

der $\hat{y}_T(l)$ er den predikerte verdien av y_t for periode $T+l$, mens y_{T+l} er den faktiske verdien, og F angir antall predikerte perioder. Dette gir et mål på avviket mellom de predikerte verdiene og den faktiske tidsserien. Dette målet er imidlertid avhengig av hvilken målestokk den avhengige variabelen er gitt i, og bør derfor bare brukes for å sammenligne egenskapene til forskjellige prognosemodeller for den samme tidsserien.

Et alternativ er *Theils ulikhetskoeffisient*, som er gitt ved

$$(5.24) \quad U = \frac{\sqrt{\frac{1}{F} \sum_{t=1}^F (\hat{y}_T(l) - y_{T+l})^2}}{\sqrt{\frac{1}{F} \sum_{t=1}^F (\hat{y}_T(l))^2 + \frac{1}{F} \sum_{t=1}^F (y_{T+l})^2}}.$$

Fordelen med *Theils ulikhetskoeffisient* er at den er uavhengig av måleenheten tidsserien er gitt i, slik at man kan sammenligne modellenes prognoseegenskaper over forskjellige tidsserier. *Theils ulikhetskoeffisient* vil alltid ligge mellom null og én, der null indikerer at den predikerte verdien og den faktiske verdien sammenfaller, $\hat{y}_t = y_t$, for alle t .

Theils ulikhetskoeffisient kan også dekomponeres, slik at man kan sammenligne den predikerte og den observerte tidsserien med hensyn til gjennomsnitt, varians og kovarians. Om man ser på ligning (5.24), ser man at *Theils ulikhetskoeffisient* er lik rms-prognosefeil, men at denne er skalert slik at den alltid vil være mellom null og én. Man kan vise at

$$(5.25) \quad \frac{1}{F} \sum (\hat{y}_T(l) - y_{T+l})^2 = (\bar{\hat{y}} - \bar{y})^2 + (\sigma_{\hat{y}} - \sigma_y)^2 + 2(1 - \rho)\sigma_{\hat{y}}\sigma_y$$

der $\bar{\hat{y}}, \bar{y}, \sigma_{\hat{y}}$ og σ_y er gjennomsnittet og variansen til den predikerte og den observerte tidsserien, og ρ er korrelasjonskoeffisienten, gitt ved $\rho = (1/\sigma_{\hat{y}}\sigma_y F) \sum (\hat{y}_T(l) - \bar{\hat{y}})(y_{T+l} - \bar{y})$.

De enkelte delene av *Theils ulikhetskoeffisient* kan da defineres som:

$$\text{Forventningsskjevhet} \quad U^M = \frac{(\bar{\hat{y}} - \bar{y})^2}{(1/F) \sum (\hat{y}_T(l) - y_{T+l})^2}$$

Varians
$$U^S = \frac{(\sigma_{\hat{y}} - \sigma_y)^2}{(1/F)\sum(\hat{y}_T(l) - y_T)^2}$$

Kovarianse
$$U^K = \frac{2(1-\rho)\sigma_{\hat{y}}\sigma_y}{(1/F)\sum(\hat{y}_T(l) - y_{T+l})^2}$$

Forventningsskjevhetsdelen forteller oss her hvor stort avvik det er mellom gjennomsnittet til den predikerte tidsserien og den observerte tidsserien, mens variandsdelen forteller oss hvor stort avvik det er mellom variansen til den predikerte tidsserien og den observerte tidsserien. Kovariansdelen måler den resterende usystematiske prognosefeilen. Fordelen ved disse er at man her kan dele opp feilleddet med hensyn til hvilken kilde de er et resultat av. Legg merke til at $U^M + U^S + U^K = 1$. Det er her ønskelig at verdien av U^M og U^S er så lav som mulig siden disse representerer systematiske forskjeller mellom prognosen og den faktiske tidsserien. U^K derimot, representerer de resterende usystematiske kildene til prognosefeil. Det er derfor ønskelig at mest mulig av feilleddet skal tilhøre denne kategorien. Det vil si at for en hver verdi $U > 0$ vil det være ideelt om $U^M = U^S = 0$ og $U^K = 1$.

Modellene som er identifisert i kapittel 5.3 er alle basert på perioden januar 1988 til desember 1999. For å evaluere prognoseegenskapene til disse modellene har jeg her laget *ex post*-prognoser for perioden januar til desember 2000, for så å sammenligne disse med de faktiske observasjonene i denne perioden. Resultatet av denne evalueringen er gjengitt i appendiks 2.

Her finner jeg fire sentrale resultater:

- 1) Modellene fra tabell 5.2 predikerer generelt sett godt for de første ti månedene av 2000.
- 2) Modeller identifisert ved hjelp av AIC har ikke bedre prognoseegenskaper enn modeller basert på SC, selv om disse inneholder flere ARMA-ledd.
- 3) Å legge en konstant til modellene gjør at prognoseegenskapene generelt sett blir dårligere. Et unntak her er ARIMA(3,1,4)x(1,1,1).
- 4) Å bruke intervensjonsanalyse for å korrigere for avvikende observasjoner forbedrer prognoseegenskapene, til de modellene som er identifisert ved hjelp av AIC, litt. Men gjør prognoseegenskapene til modellene som er identifisert ved hjelp av SC generelt sett dårligere.

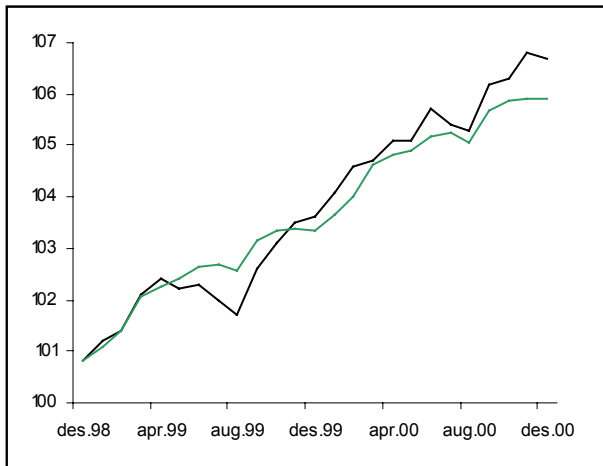
På bakgrunn av dette vil jeg benytte de modellene som er identifisert ved hjelp av Schwartz-kriteriet uten bruk av konstantledd eller intervensjonsanalyse. Etter å ha vurdert modellenes prognoseegenskaper ser man også at $ARIMA(0,1,0) \times (0,1,1)$ -modellen har gitt en bedre prognose for år 2000 selv om denne modellen ble rangert som nest best etter modellen $ARIMA(0,1,0) \times (1,1,1)$.

At modellene gir et dårlig bilde av situasjonen i november og desember, kan skyldes at modellene ikke gir noe god beskrivelse av utviklingen i tidsserien på lengre sikt enn ti måneder. Jeg har derfor også prøvd de fire modellene som gjorde det best med hensyn på prognoseegenskaper for en toårs periode. Jeg har her estimert modellene for perioden januar 1988 til desember 1998, og brukt disse til å lage *ex post*-prognoser for perioden januar 1999 til desember 2000. Resultatene av dette forsøket er gjengitt i figur 5.7. Man ser her at modellene, med unntak av $ARIMA(3,1,4) \times (1,1,1)$ med konstantledd, gjør det relativt bra. Man ser likevel at modellene overpredikere for september 1999 og underpredikerer for november og desember 2000. Som før, ser vi også her at $ARIMA(0,1,0) \times (0,1,1)$ har bedre prognoseegenskaper enn $ARIMA(0,1,0) \times (1,1,1)$, selv om de ble rangert motsatt av SC i tabell 5.2.

En annen forklaring på at modellene gir et dårlig bilde av inflasjonsutviklingen i november og desember 2000, er at man har hatt en høy prisvekst på elektrisitet¹⁷ og husleie i november. Dette kan ha medført en unormalt høy inflasjon i november og desember. Siden ARIMA-modeller krever en stabil utvikling i tidsserien for å kunne fungere optimalt, har ikke modellen klart å forutsi denne utviklingen.

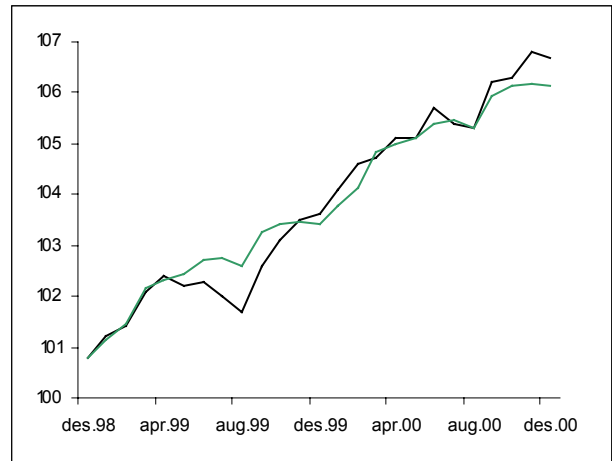
¹⁷ Endringen i elektrisitetprisene fra oktober til november 2000 var i følge SSB på 6,1 prosent.

Fig. 5.7.1 – ARIMA(0,1,0)x(1,1,1)



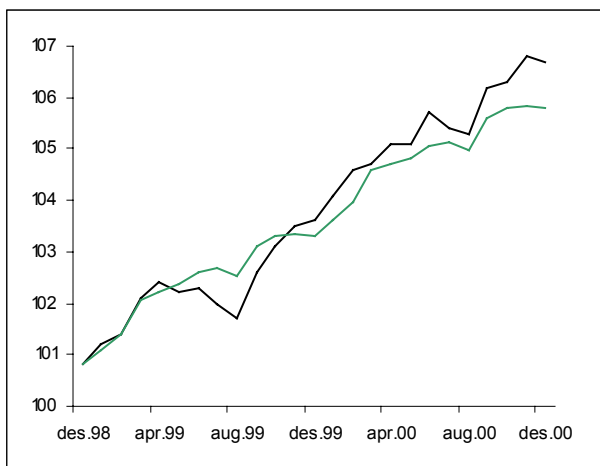
mrs	U	U^M	U^S	U^K
0,445484	0,002144	0,072158	0,461034	0,466808

Fig. 5.7.2 – ARIMA(0,1,0)x(0,1,1)



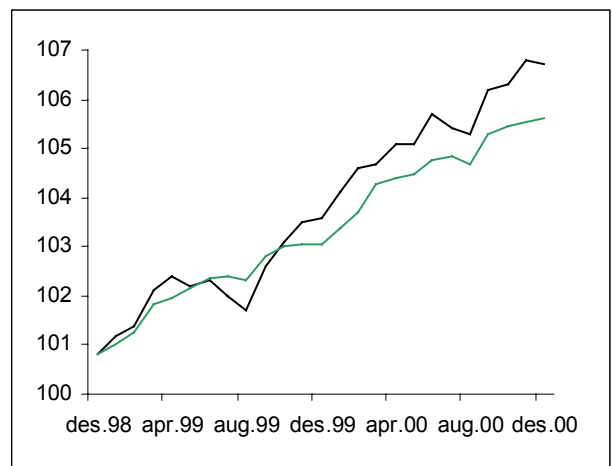
mrs	U	U^M	U^S	U^K
0,377937	0,001818	0,003451	0,362754	0,633796

Fig. 5.7.3 – ARIMA(3,1,0)x(1,1,1)



mrs	U	U^M	U^S	U^K
0,483530	0,002328	0,128609	0,482311	0,389380

Fig. 5.7.4 - ARIMA(3,1,4)x(1,1,1), konst.



mrs	U	U^M	U^S	U^K
0,634707	0,003060	0,458351	0,359745	0,181905

5.5.3 Prognoser for år 2001

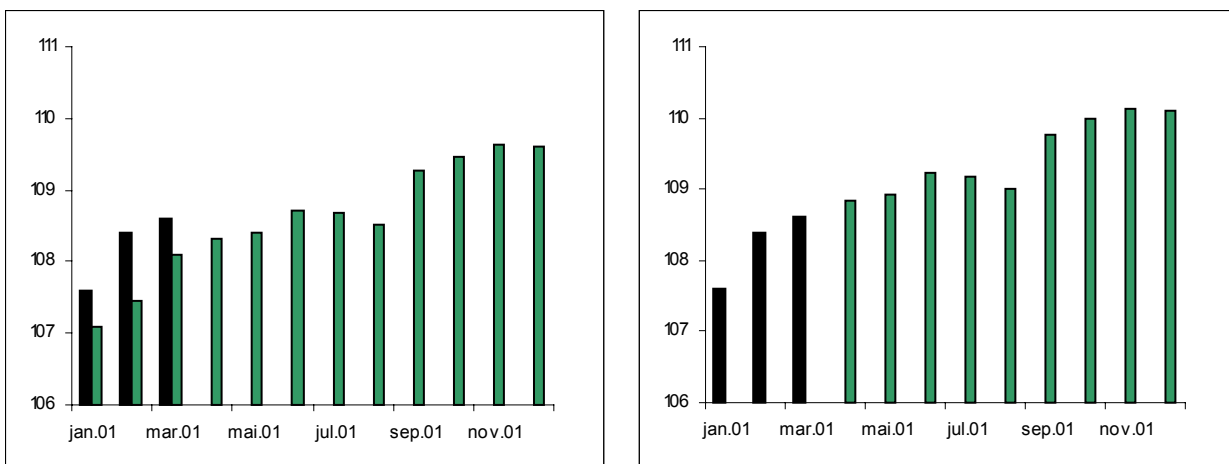
Til slutt vil jeg benytte denne metoden for å lage en prognose for 2001. Jeg har derfor estimert modellene på nytt for perioden januar 1988 til desember 2000¹⁸. Jeg har her på samme måte som tidligere rangert de tre beste modellene, ifølge AIC og SC, i tabell 5.10.

Tabell 5.10 – ARIMA(p,d,q)x(P,D,Q)

Ranking	Modell	AIC (Prob(Q-test))	Modell	SC (Prob(Q-test))
1	(3,1,0)x(1,1,1)	-9,406631 (0,913)	(0,1,0)x(0,1,1)	-9,377780 (0,794)
2	(0,1,0)x(1,1,1)	-9,402444 (0,790)	(0,1,0)x(1,1,1)	-9,358548 (0,790)
3	(4,1,0)x(1,1,1)	-9,402108 (0,792)	(0,1,1)x(0,1,1)	-9,346683 (0,730)

Her ser man at ARIMA(0,1,0)x(0,1,1) denne gangen er blitt rangert som den beste modellen av SC. Siden denne modellen også gjorde det best når jeg evaluerte modellenes prognoseegenskaper i kapittel 5.5.2, vil jeg bruke denne modellen til å predikere inflasjonen i 2001.

Figur 5.8.1 og 5.8.2 – Prognose for 2001 ved ARIMA(0,1,0)x(0,1,1)



I figur 5.8.1 er prognosen for 2001, basert på en ARIMA(0,1,0)x(0,1,1)-modell, gjengitt som grå stolper. Dersom denne prognosen gir et riktig bilde av inflasjonsutviklingen i år vil dette tilsi en inflasjon på 2,9 prosent i 2001.

¹⁸ Se Appendiks 3.

I figur 5.8.1 har jeg også gjengitt de observerte verdiene for konsumprisindeksen, for første kvartal, som svarte stolper. Når man sammenligner prognosen med de observerte verdiene, ser man at modellen underpredikerer inflasjonsutviklingen for første kvartal. Det er hovedsakelig to grunner det dette. Ved årsskiftet ble den generelle momsen satt opp fra 23 til 24 prosent. Dette ble forsterket av en uvanlig høy prisstigning på elektrisk kraft i vintermånedene 2000/01. Man får her demonstrert en av svakhetene med ARIMA-modeller. Siden modellen utelukkende baserer seg på den historiske utviklingen i tidsserien, har den ingen forutsetninger for å predikere et skift i tidsserien.

Men modellens svakhet, er også modellens styrke. Siden modellen utelukkende baserer seg på den historisk utviklingen i tidsserien, er det lett å inkludere nye observasjoner i modellen. I figur 5.8.2 er observasjonene for første kvartal tatt med i prognosegrunnlaget. Man ser her at anslaget for inflasjonsutviklingen blir høyere enn de var i det forrige anslaget og at man her får en gjennomsnittlig vekst i inflasjonen for 2001, på 3,4 prosent.

Det knytter seg imidlertid stor usikkerhet til inflasjonsutviklingen i år, på grunn av endringer i avgiftssystemet. I tillegg til en økning i det generelle momsnivået fra 23 til 24 prosent, vil moms på mat bli halvert etter 1. juli, og det vil samtidig bli introdusert moms på tjenester. Dette er endringer som ikke kan inkluderes i modellen før de inntreffer. Det er derfor sannsynlig at prognosen i figur 5.8.2 vil overpredikere inflasjonsutviklingen i tredje kvartal, på samme måte som prognosen i figur 5.2.1 underpredikerte inflasjonsutviklingen i første kvartal.

Dette er imidlertid i tråd med konklusjonene fra den økonometriske litteraturen, som ble diskutert i kapittel 2.2. Man mener her at ARIMA-modeller egner seg best til å modellere tidsserier som er relativt stabile, og at siden modellen ikke har noen basis i økonomisk teori, kan man ikke bruke disse modellene til å predikere større endringer i tidsserien.

Modellene inkluderer imidlertid raskt ny informasjon og justerer prognosen deretter. Det er derfor forventet at modellen vil kunne predikere inflasjonsutviklingen i andre og fjerde kvartal minst like bra som prognosene for 2000 i kapittel 5.5.2.

6. AVSLUTNING

29. mars 2001 innførte Norge inflasjonsmål som mandat for den norske pengepolitikken. Dette har gjort at muligheten for å lage gode prognoser for norsk inflasjon er viktigere nå enn tidligere. Det har derfor blitt viktigere å vurdere flere alternative metoder å predikere inflasjon på. Hovedmålet med denne oppgave var å undersøke muligheten for å predikere norsk inflasjon ved hjelp av ARIMA-modeller.

I første delen av oppgaven er det, i korte trekk, gjort rede for den utviklingen man har sett innen arbeidet med å lage økonomiske prognoser, og modellene disse baserer seg på. Man kan her se at man først tok i bruk modeller som i stor grad var basert på økonomisk teori, og som en reaksjon på disse ble det utviklet modeller som var basert på rene statistiske prinsipper, deriblant ARIMA-modellen. Denne diskusjonen endte i en syntese mellom disse to paradigmene og man begynte, i større grad enn tidligere, å inkludere statistiske prinsipper i de økonomisk modellene. Denne utviklingen er også lett å spore i Norge. Men det er, så lagt jeg kjenner til, ikke gjort noen forsøk på å predikere norsk inflasjon med rene statistiske modeller.

Utenlandske studier har vist at ARIMA-modeller, med gode resultater, kan brukes til å prediker inflasjon. Modellene fungerer imidlertid best på kort sikt og om inflasjonsutviklingen er relativt stabil. Siden modellen utelukkende baserer seg på den historiske utviklingen i tidsserien kan man ikke bruke denne modellen til å predikere større endringer i inflasjonsutviklingen, ved for eksempel endringer i den økonomiske politikken eller økonomiske sjokk.

Etter en gjennomgang av modellens virkemåte og datasettet, har jeg i kapittel 5, gjort en analyse for konsumprisindeksen i praksis. Jeg har her funnet at en $ARIMA(0,1,0) \times (0,1,1)$ -modell er best egnet til å modellere og predikere konsumprisindeksen i Norge. Man ser at modellen ikke inneholder $AR(p)$ - eller $MA(q)$ -ledd. Det vil si at modellen bare har vært i stand til å identifisere en tidstrend og en sesongvariasjon. Modellen demonstrerte imidlertid gode prognoseegenskaper, når jeg i kapittel 5.5.2, evaluerte modellenes *ex post*-prognoseegenskaper.

I kapittel 5.5.2 har jeg forsøkt å lage en prognose for 2001. En sammenlikning med de observerte verdiene for konsumprisindeksen de tre første månedene i år, viser imidlertid at modellen underpredikerer inflasjonen i første kvartal. Grunnen til dette er at den generelle momsen ble øket fra 1. januar i år, forsterket av en uvanlig høy prisstigning på elektrisitet i vinterhalvåret 2000/01. Man ser her at prognoseegenskapene til modellen reduseres i perioder der inflasjonsutviklingen er ustabil. Dette er imidlertid helt i tråd med den økonomiske litteraturen.

Jeg vil derfor konkludere med at ARIMA-modellen kan være et godt alternativ til de økonomiske modellene, om man ønsker å predikere inflasjon i Norge på kort sikt. Men man må ta et forbehold om at inflasjonsutviklingen må være relativt stabil i den perioden man ønsker å predikere.

REFERANSER

Batchelor, R. og P. Dua (1996), "Empirical measures of inflation uncertainty: a cautionary note," *Applied Economics*, 28, 333-341.

Box, G. E. P., G. M. Jenkins og G. C. Reinsel (1994), *Time Series Analysis: Forecasting and Control*, Prentice Hall, Englewood Cliffs.

Bråten, A. og K. Olsen (1997), "Ulike metoder for beregning av en indikator forunderliggende inflasjon," Rapport 97/7, Statistisk Sentralbyrå.

Bårdsen, G., E. S. Jansen og R. Nymoen (1999), "Econometric inflation targeting," Arbeidsnotat 1999/5, Norges Bank.

Campbell, J. Y. og P. Perron (1991), "Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots," Technical Working Paper No. 100, National Bureau of Economic Research.

Cecchetti, S. G. (1995), "Inflation Indicators and Inflation Policy," NBER Working Paper No. 5161. I NBER's Thenth Annual Conference on Macroeconomics.

Cecchetti, S. G., R. S. Chu og C. Steindel (2000), "The Unreliability of Inflation Indicators," *Current Issues in economics and finance*, 6, No. 4, Federal Reserve Bank of New York.

Dickey, D. A. og Fuller W. A. (1979), "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, 74, 427-431.

Diebold, F. X. (1998), "The Past, Present, and Future of Macroeconomic Forecasting," *Journal of Economic Perspectives*, 12, 175-192.

Freeman, D. G. (1998), "Do core inflation measures help forecast inflation," *Economics Letter*, 58, 143-147.

Granger, C. W. J. og P. Newbold (1986), *Forecasting Economic Time Series*, Academic Press Inc., San Diego.

Granger, C. W. J. (1990), *Modelling Economic Series*, Clarendon Press, Oxford.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton.

Hendry, D. F. (1993), *Econometrics. Alchemy or Science?*, Blackwell, Oxford.

Hendry, D. F. (1995), *Dynamic Econometrics*, Oxford University Press, Oxford.

Isachsen, A. J., S. O. Stoknes og G. H. Bjønnes (1999), "Den store gjettekonkurransen," Notat nr. 16/1999, Statistisk Sentralbyrå.

Jansen, E. S. (2000), "Statistical issues in macroeconomic modelling," Arbeidsnotat 2000/12, Norges Bank.

Kvilekval, E., K. Vaage og E. Vårdal (1998), "Sammenhengen mellom pengemengdeveksten og inflasjon i Norge," *Norsk Økonomisk Tidsskrift*, 112: 185-211.

Meyler, A., G. Kenny og T. Quinn (1999), "Forecasting Irish Inflation Using ARIMA Models," Technical Paper 3/RT/99, Central Bank of Ireland.

Norge Bank: Inflasjonsrapport, flere hefter.

Pindyck, R. S. og D. L. Rubinfeld (1998), *Econometric Models and Econometric Forecasts*, McGraw-Hill.

Poskitt, D. og A. Tremayne (1987), "Determining a Portfolio of Linear Time Series Models," *Biometrika*, 74:125-137.

Rodriguez, J. (1997), "Sesongjustering i praksis - en innføring," Notat nr. 29/1997, Statistisk Sentralbyrå.

Statistisk Sentralbyrå (1991): "Konsumprisindeksen," Rapport 91/1.

Statistisk Sentralbyrå: Økonomiske analyser, flere hefter.

Shibata, R. (1976), "Selection of the order of an autoregressive model by Akaike's information criterion," *Biometrika*, 63: 117-126.

Quinn, T., G. Kenny og A. Meyler (1999), "Inflation Analysis: An Overview," Technical Paper 1/RT/99, Central Bank of Ireland.

Appendiks 1: De estimerte modellene

Tabell 1: ARIMA(p,d,q)x(P,D,Q).

(p,d,q)x(P,D,Q)	(0,1,0)x(1,1,1)	(0,1,1)x(1,1,1)	(0,1,2)x(1,1,1)	(0,1,3)x(1,1,1)	(0,1,4)x(1,1,1)
AIC	-9,469803	-9,472412	-9,457765	-9,446682	-9,429888
SC	-9,423096	-9,402350	-9,364349	-9,329912	-9,289764
Prob(Q-test)	0,817	0,857	0,817	0,856	0,826
(p,d,q)x(P,D,Q)	(1,1,0)x(1,1,1)	(1,1,1)x(1,1,1)	(1,1,2)x(1,1,1)	(1,1,3)x(1,1,1)	(1,1,4)x(1,1,1)
AIC	-9,474672	-9,458623	-9,445934	-9,437179	-9,420245
SC	-9,404231	-9,364701	-9,328532	-9,296297	-9,255882
Prob(Q-test)	0,835	0,817	0,749	0,784	0,738
(p,d,q)x(P,D,Q)	(2,1,0)x(1,1,1)	(2,1,1)x(1,1,1)	(2,1,2)x(1,1,1)	(2,1,3)x(1,1,1)	(2,1,4)x(1,1,1)
AIC	-9,484191	-9,465969	-9,453070	-9,437208	-9,419673
SC	-9,389758	-9,347928	-9,311420	-9,271950	-9,230806
Prob(Q-test)	0,890	0,865	0,795	0,776	0,765
(p,d,q)x(P,D,Q)	(3,1,0)x(1,1,1)	(3,1,1)x(1,1,1)	(3,1,2)x(1,1,1)	(3,1,3)x(1,1,1)	(3,1,4)x(1,1,1)
AIC	-9,493022	-9,477957	-9,463730	-9,462177	-9,487839
SC	-9,374333	-9,335530	-9,297566	-9,272275	-9,274198
Prob(Q-test)	0,951	0,930	0,895	0,658	0,748
(p,d,q)x(P,D,Q)	(4,1,0)x(1,1,1)	(4,1,1)x(1,1,1)	(4,1,2)x(1,1,1)	(4,1,3)x(1,1,1)	(4,1,4)x(1,1,1)
AIC	-9,483425	-9,474221	-9,491443	-9,474329	-9,431210
SC	-9,340211	-9,307138	-9,300492	-9,259508	-9,192520
Prob(Q-test)	0,934	0,849	0,804	0,778	0,766

Tabell 2: ARIMA(p,d,q)x(0,D,Q).

(p,d,q)x(P,D,Q)	(0,1,0)x(0,1,1)	(0,1,1)x(0,1,1)	(0,1,2)x(0,1,1)	(0,1,3)x(0,1,1)	(0,1,4)x(0,1,1)
AIC	-9,433324	-9,427912	-9,420955	-9,420797	-9,405973
SC	-9,411376	-9,384016	-9,355111	-9,333005	-9,296232
Prob(Q-test)	0,779	0,742	0,588	0,597	0,545
(p,d,q)x(P,D,Q)	(1,1,0)x(0,1,1)	(1,1,1)x(0,1,1)	(1,1,2)x(0,1,1)	(1,1,3)x(0,1,1)	(1,1,4)x(0,1,1)
AIC	-9,433809	-9,429891	-9,430037	-9,429227	-9,416347
SC	-9,389693	-9,363718	-9,341805	-9,318937	-9,283999
Prob(Q-test)	0,727	0,617	0,417	0,645	0,529
(p,d,q)x(P,D,Q)	(2,1,0)x(0,1,1)	(2,1,1)x(0,1,1)	(2,1,2)x(0,1,1)	(2,1,3)x(0,1,1)	(2,1,4)x(0,1,1)
AIC	-9,433568	-9,422100	-9,406272	-9,416663	-9,401108
SC	-9,367061	-9,333423	-9,295426	-9,283649	-9,245924
Prob(Q-test)	0,630	0,560	0,533	0,616	0,566
(p,d,q)x(P,D,Q)	(3,1,0)x(0,1,1)	(3,1,1)x(0,1,1)	(3,1,2)x(0,1,1)	(3,1,3)x(0,1,1)	(3,1,4)x(0,1,1)
AIC	-9,437753	-9,428128	-9,414480	-9,400594	-9,423461
SC	-9,348628	-9,316720	-9,280791	-9,244623	-9,245209
Prob(Q-test)	0,818	0,709	0,585	0,614	0,513
(p,d,q)x(P,D,Q)	(4,1,0)x(0,1,1)	(4,1,1)x(0,1,1)	(4,1,2)x(0,1,1)	(4,1,3)x(0,1,1)	(4,1,4)x(0,1,1)
AIC	-9,424641	-9,411030	-9,391108	-9,382822	-9,384005
SC	-9,312665	-9,276659	-9,234341	-9,203660	-9,182449
Prob(Q-test)	0,819	0,691	0,590	0,589	0,609

Tabell 3: ARIMA(p,d,q)x(P,D,0).

(p,d,q)x(P,D,Q)	(0,1,0)x(1,1,0)	(0,1,1)x(1,1,0)	(0,1,2)x(1,1,0)	(0,1,3)x(1,1,0)	(0,1,4)x(1,1,0)
AIC	-9,294910	-9,284167	-9,282805	-9,280167	-9,268952
SC	-9,271556	-9,237459	-9,212743	-9,186751	-9,152182
Prob(Q-test)	0,919	0,875	0,928	0,947	0,947
(p,d,q)x(P,D,Q)	(1,1,0)x(1,1,0)	(1,1,1)x(1,1,0)	(1,1,2)x(1,1,0)	(1,1,3)x(1,1,0)	(1,1,4)x(1,1,0)
AIC	-9,274152	-9,303150	-9,286434	-9,282550	-9,296173
SC	-9,227191	-9,232709	-9,192512	-9,165148	-9,155291
Prob(Q-test)	0,880	0,941	0,930	0,907	0,942
(p,d,q)x(P,D,Q)	(2,1,0)x(1,1,0)	(2,1,1)x(1,1,0)	(2,1,2)x(1,1,0)	(2,1,3)x(1,1,0)	(2,1,4)x(1,1,0)
AIC	-9,262457	-9,278117	-9,259632	-9,243332	-9,228045
SC	-9,191632	-9,183684	-9,141591	-9,101682	-9,062787
Prob(Q-test)	0,900	0,941	0,938	0,917	0,915
(p,d,q)x(P,D,Q)	(3,1,0)x(1,1,0)	(3,1,1)x(1,1,0)	(3,1,2)x(1,1,0)	(3,1,3)x(1,1,0)	(3,1,4)x(1,1,0)
AIC	-9,269501	-9,254200	-9,310345	-9,316935	-9,302103
SC	-9,174550	-9,135511	-9,167918	-9,150770	-9,112200
Prob(Q-test)	0,975	0,961	0,947	0,881	0,872
(p,d,q)x(P,D,Q)	(4,1,0)x(1,1,0)	(4,1,1)x(1,1,0)	(4,1,2)x(1,1,0)	(4,1,3)x(1,1,0)	(4,1,4)x(1,1,0)
AIC	-9,248251	-9,232851	-9,265494	-9,262763	-9,279885
SC	-9,128906	-9,089637	-9,098412	-9,071811	-9,065064
Prob(Q-test)	0,976	0,971	0,957	0,863	0,779

Tabell 4: ARIMA(p,d,q)x(0,D,0).

(p,d,q)x(P,D,Q)	(0,1,0)x(0,1,0)	(0,1,1)x(0,1,0)	(0,1,2)x(0,1,0)	(0,1,3)x(0,1,0)	(0,1,4)x(0,1,0)
AIC		-8,804177	-8,809147	-8,815137	-8,825365
SC		-8,782228	-8,765251	-8,749292	-8,737573
Prob(Q-test)		0,000	0,000	0,000	0,000
(p,d,q)x(P,D,Q)	(1,1,0)x(0,1,0)	(1,1,1)x(0,1,0)	(1,1,2)x(0,1,0)	(1,1,3)x(0,1,0)	(1,1,4)x(0,1,0)
AIC	-8,798446	-8,930393	-8,868244	-8,853923	-8,827196
SC	-8,776388	-8,886277	-8,802070	-8,765692	-8,716906
Prob(Q-test)	0,000	0,000	0,000	0,000	0,0000
(p,d,q)x(P,D,Q)	(2,1,0)x(0,1,0)	(2,1,1)x(0,1,0)	(2,1,2)x(0,1,0)	(2,1,3)x(0,1,0)	(2,1,4)x(0,1,0)
AIC	-8,819386	-8,879064	-8,864199	-8,847537	-8,832034
SC	-8,775047	-8,812556	-8,775523	-8,736692	-8,699019
Prob(Q-test)	0,000	0,000	0,000	0,000	0,000
(p,d,q)x(P,D,Q)	(3,1,0)x(0,1,0)	(3,1,1)x(0,1,0)	(3,1,2)x(0,1,0)	(3,1,3)x(0,1,0)	(3,1,4)x(0,1,0)
AIC	-8,819963	-8,853600	-8,918065	-8,909802	-8,907231
SC	-8,753119	-8,764474	-8,806657	-8,776113	-8,751261
Prob(Q-test)	0,000	0,000	0,000	0,000	0,000
(p,d,q)x(P,D,Q)	(4,1,0)x(0,1,0)	(4,1,1)x(0,1,0)	(4,1,2)x(0,1,0)	(4,1,3)x(0,1,0)	(4,1,4)x(0,1,0)
AIC	-8,802739	-8,788353	-8,852616	-8,888794	-8,826881
SC	-8,713159	-8,676377	-8,718244	-8,732027	-8,647720
Prob(Q-test)	0,000	0,000	0,000	0,000	0,000

Tabell 5: ARIMA(p,d,q)x(P,D,Q) med konstant.

(p,d,q)x(P,D,Q)	(0,1,0)x(1,1,1)	(0,1,1)x(1,1,1)	(0,1,2)x(1,1,1)	(0,1,3)x(1,1,1)	(0,1,4)x(1,1,1)
AIC	-9,468503	-9,467453	-9,454188	-9,441537	-9,424761
SC	-9,398441	-9,374037	-9,337418	-9,301413	-9,261283
Prob(Q-test)	0,843	0,883	0,837	0,862	0,834
(p,d,q)x(P,D,Q)	(1,1,0)x(1,1,1)	(1,1,1)x(1,1,1)	(1,1,2)x(1,1,1)	(1,1,3)x(1,1,1)	(1,1,4)x(1,1,1)
AIC	-9,467503	-9,450579	-9,443182	-9,428391	-9,417518
SC	-9,373581	-9,333177	-9,302299	-9,264029	-9,229675
Prob(Q-test)	0,870	0,843	0,746	0,771	0,632
(p,d,q)x(P,D,Q)	(2,1,0)x(1,1,1)	(2,1,1)x(1,1,1)	(2,1,2)x(1,1,1)	(2,1,3)x(1,1,1)	(2,1,4)x(1,1,1)
AIC	-9,473637	-9,458867	-9,442756	-9,426532	-9,409521
SC	-9,355595	-9,317217	-9,277497	-9,237665	-9,197046
Prob(Q-test)	0,902	0,860	0,802	0,796	0,731
(p,d,q)x(P,D,Q)	(3,1,0)x(1,1,1)	(3,1,1)x(1,1,1)	(3,1,2)x(1,1,1)	(3,1,3)x(1,1,1)	(3,1,4)x(1,1,1)
AIC	-9,480268	-9,468676	-9,450605	-9,454596	-9,475380
SC	-9,337841	-9,302511	-9,260702	-9,240956	-9,238002
Prob(Q-test)	0,953	0,914	0,895	0,689	0,735
(p,d,q)x(P,D,Q)	(4,1,0)x(1,1,1)	(4,1,1)x(1,1,1)	(4,1,2)x(1,1,1)	(4,1,3)x(1,1,1)	(4,1,4)x(1,1,1)
AIC	-9,471479	-9,466298	-9,478670	-9,456155	-9,448404
SC	-9,304396	-9,275346	-9,263849	-9,217465	-9,185845
Prob(Q-test)	0,941	0,857	0,799	0,799	0,621

Tabell 6: ARIMA(p,d,q)x(0,D,Q) med konstant.

(p,d,q)x(P,D,Q)	(0,1,0)x(0,1,1)	(0,1,1)x(0,1,1)	(0,1,2)x(0,1,1)	(0,1,3)x(0,1,1)	(0,1,4)x(0,1,1)
AIC	-9,446253	-9,435459	-9,433285	-9,428540	-9,413303
SC	-9,402357	-9,369615	-9,345493	-9,318800	-9,281615
Prob(Q-test)	0,752	0,758	0,584	0,599	0,548
(p,d,q)x(P,D,Q)	(1,1,0)x(0,1,1)	(1,1,1)x(0,1,1)	(1,1,2)x(0,1,1)	(1,1,3)x(0,1,1)	(1,1,4)x(0,1,1)
AIC	-9,438697	-9,442173	-9,452194	-9,440776	-9,424920
SC	-9,372523	-9,353942	-9,341904	-9,308428	-9,270515
Prob(Q-test)	0,726	0,679	0,452	0,560	0,547
(p,d,q)x(P,D,Q)	(2,1,0)x(0,1,1)	(2,1,1)x(0,1,1)	(2,1,2)x(0,1,1)	(2,1,3)x(0,1,1)	(2,1,4)x(0,1,1)
AIC	-9,438249	-9,428367	-9,411242	-9,414708	-9,399716
SC	-9,349573	-9,317521	-9,278227	-9,259524	-9,222363
Prob(Q-test)	0,638	0,571	0,534	0,538	0,493
(p,d,q)x(P,D,Q)	(3,1,0)x(0,1,1)	(3,1,1)x(0,1,1)	(3,1,2)x(0,1,1)	(3,1,3)x(0,1,1)	(3,1,4)x(0,1,1)
AIC	-9,432892	-9,426708	-9,417329	-9,406200	-9,386568
SC	-9,321484	-9,293019	-9,261359	-9,227948	-9,186035
Prob(Q-test)	0,808	0,673	0,525	0,463	0,654
(p,d,q)x(P,D,Q)	(4,1,0)x(0,1,1)	(4,1,1)x(0,1,1)	(4,1,2)x(0,1,1)	(4,1,3)x(0,1,1)	(4,1,4)x(0,1,1)
AIC	-9,418893	-9,409229	-9,394020	-9,379724	-9,375273
SC	-9,284522	-9,252463	-9,214859	-9,178167	-9,151321
Prob(Q-test)	0,801	0,648	0,512	0,404	0,491

Tabell 7: ARIMA(p,d,q)x(P,D,0) med konstant.

(p,d,q)x(P,D,Q)	(0,1,0)x(1,1,0)	(0,1,1)x(1,1,0)	(0,1,2)x(1,1,0)	(0,1,3)x(1,1,0)	(0,1,4)x(1,1,0)
AIC	-9,290603	-9,278126	-9,280189	-9,273873	-9,265217
SC	-9,243895	-9,208064	-9,186773	-9,157103	-9,125093
Prob(Q-test)	0,917	0,879	0,929	0,948	0,944
(p,d,q)x(P,D,Q)	(1,1,0)x(1,1,0)	(1,1,1)x(1,1,0)	(1,1,2)x(1,1,0)	(1,1,3)x(1,1,0)	(1,1,4)x(1,1,0)
AIC	-9,268424	-9,297530	-9,281488	-9,278815	-9,245271
SC	-9,197983	-9,203609	-9,164086	-9,137932	-9,080909
Prob(Q-test)	0,884	0,940	0,934	0,902	0,920
(p,d,q)x(P,D,Q)	(2,1,0)x(1,1,0)	(2,1,1)x(1,1,0)	(2,1,2)x(1,1,0)	(2,1,3)x(1,1,0)	(2,1,4)x(1,1,0)
AIC	-9,258451	-9,272760	-9,253524	-9,236588	-9,263409
SC	-9,164018	-9,154718	-9,111874	-9,071330	-9,074542
Prob(Q-test)	0,905	0,942	0,937	0,918	0,768
(p,d,q)x(P,D,Q)	(3,1,0)x(1,1,0)	(3,1,1)x(1,1,0)	(3,1,2)x(1,1,0)	(3,1,3)x(1,1,0)	(3,1,4)x(1,1,0)
AIC	-9,259852	-9,245839	-9,298720	-9,311552	-9,298014
SC	-9,141163	-9,103412	-9,132555	-9,121649	-9,084373
Prob(Q-test)	0,975	0,960	0,918	0,895	0,877
(p,d,q)x(P,D,Q)	(4,1,0)x(1,1,0)	(4,1,1)x(1,1,0)	(4,1,2)x(1,1,0)	(4,1,3)x(1,1,0)	(4,1,4)x(1,1,0)
AIC	-9,240473	-9,225828	-9,214675	-9,326603	-9,266166
SC	-9,097259	-9,058745	-9,023723	-9,111782	-9,027476
Prob(Q-test)	0,976	0,969	0,954	0,838	0,766

Tabell 8: ARIMA(p,d,q)x(0,D,0) med konstant.

(p,d,q)x(P,D,Q)	(0,1,0)x(0,1,0)	(0,1,1)x(0,1,0)	(0,1,2)x(0,1,0)	(0,1,3)x(0,1,0)	(0,1,4)x(0,1,0)
AIC	-8,807485	-8,792242	-8,798084	-8,802680	-8,814460
SC	-8,785537	-8,748346	-8,732240	-8,714888	-8,704720
Prob(Q-test)	0,000	0,000	0,000	0,000	0,000
(p,d,q)x(P,D,Q)	(1,1,0)x(0,1,0)	(1,1,1)x(0,1,0)	(1,1,2)x(0,1,0)	(1,1,3)x(0,1,0)	(1,1,4)x(0,1,0)
AIC	-8,785976	-8,888123	-8,855756	-8,841061	-8,857637
SC	-8,741860	-8,821949	-8,767524	-8,730772	-8,725289
Prob(Q-test)	0,000	0,000	0,000	0,000	0,000
(p,d,q)x(P,D,Q)	(2,1,0)x(0,1,0)	(2,1,1)x(0,1,0)	(2,1,2)x(0,1,0)	(2,1,3)x(0,1,0)	(2,1,4)x(0,1,0)
AIC	-8,806264	-8,865780	-8,850821	-8,834039	-8,818539
SC	-8,739756	-8,777103	-8,739976	-8,701024	-8,663355
Prob(Q-test)	0,0000	0,000	0,000	0,000	0,000
(p,d,q)x(P,D,Q)	(3,1,0)x(0,1,0)	(3,1,1)x(0,1,0)	(3,1,2)x(0,1,0)	(3,1,3)x(0,1,0)	(3,1,4)x(0,1,0)
AIC	-8,806203	-8,843283	-8,827775	-8,818450	-8,895312
SC	-8,717077	-8,731876	-8,694086	-8,662480	-8,717060
Prob(Q-test)	0,000	0,000	0,000	0,000	0,000
(p,d,q)x(P,D,Q)	(4,1,0)x(0,1,0)	(4,1,1)x(0,1,0)	(4,1,2)x(0,1,0)	(4,1,3)x(0,1,0)	(4,1,4)x(0,1,0)
AIC	-8,789688	-8,775864	-8,837840	-8,875106	-8,850700
SC	-8,677712	-8,641493	-8,681074	-8,695944	-8,649144
Prob(Q-test)	0,000	0,000	0,000	0,000	0,000

Tabell 9: ARIMA(p,d,q)x(P,D,Q) med du1 og du2.

(p,d,q)x(P,D,Q)	(0,1,0)x(1,1,1)	(0,1,1)x(1,1,1)	(0,1,2)x(1,1,1)	(0,1,3)x(1,1,1)	(0,1,4)x(1,1,1)
AIC	-9,639203	-9,633306	-9,619281	-9,609357	-9,594623
SC	-9,545787	-9,516536	-9,479157	-9,445880	-9,407791
Prob(Q-test)	0,134	0,152	0,142	0,209	0,110
(p,d,q)x(P,D,Q)	(1,1,0)x(1,1,1)	(1,1,1)x(1,1,1)	(1,1,2)x(1,1,1)	(1,1,3)x(1,1,1)	(1,1,4)x(1,1,1)
AIC	-9,632365	-9,616963	-9,606320	-9,598843	-9,571109
SC	-9,514963	-9,476080	-9,441957	-9,411000	-9,359785
Prob(Q-test)	0,103	0,100	0,101	0,158	0,710
(p,d,q)x(P,D,Q)	(2,1,0)x(1,1,1)	(2,1,1)x(1,1,1)	(2,1,2)x(1,1,1)	(2,1,3)x(1,1,1)	(2,1,4)x(1,1,1)
AIC	-9,652362	-9,632926	-9,622283	-9,605953	-9,589471
SC	-9,510712	-9,467668	-9,433417	-9,393478	-9,353388
Prob(Q-test)	0,253	0,204	0,126	0,133	0,098
(p,d,q)x(P,D,Q)	(3,1,0)x(1,1,1)	(3,1,1)x(1,1,1)	(3,1,2)x(1,1,1)	(3,1,3)x(1,1,1)	(3,1,4)x(1,1,1)
AIC	-9,674932	-9,659143	-9,625751	-9,620823	-9,651259
SC	-9,508767	-9,469240	-9,412111	-9,383445	-9,390143
Prob(Q-test)	0,491	0,443	0,360	0,400	0,265
(p,d,q)x(P,D,Q)	(4,1,0)x(1,1,1)	(4,1,1)x(1,1,1)	(4,1,2)x(1,1,1)	(4,1,3)x(1,1,1)	(4,1,4)x(1,1,1)
AIC	-9,662748	-9,649571	-9,665304	-9,643264	-9,616297
SC	-9,471796	-9,434750	-9,426615	-9,380705	-9,329870
Prob(Q-test)	0,496	0,293	0,319	0,256	0,252

Tabell 10: ARIMA(p,d,q)x(0,D,Q) med du1 og du2.

(p,d,q)x(P,D,Q)	(0,1,0)x(0,1,1)	(0,1,1)x(0,1,1)	(0,1,2)x(0,1,1)	(0,1,3)x(0,1,1)	(0,1,4)x(0,1,1)
AIC	-9,586476	-9,580299	-9,574640	-9,577382	-9,565263
SC	-9,520632	-9,492506	-9,464899	-9,445693	-9,411626
Prob(Q-test)	0,392	0,409	0,323	0,294	0,205
(p,d,q)x(P,D,Q)	(1,1,0)x(0,1,1)	(1,1,1)x(0,1,1)	(1,1,2)x(0,1,1)	(1,1,3)x(0,1,1)	(1,1,4)x(0,1,1)
AIC	-9,583902	-9,580853	-9,581853	-9,582960	-9,573309
SC	-9,495670	-9,470563	-9,449505	-9,428554	-9,396845
Prob(Q-test)	0,360	0,231	0,121	0,266	0,142
(p,d,q)x(P,D,Q)	(2,1,0)x(0,1,1)	(2,1,1)x(0,1,1)	(2,1,2)x(0,1,1)	(2,1,3)x(0,1,1)	(2,1,4)x(0,1,1)
AIC	-9,589955	-9,577735	-9,560395	-9,575630	-9,561956
SC	-9,479109	-9,444721	-9,405211	-9,398277	-9,362434
Prob(Q-test)	0,340	0,259	0,251	0,271	0,183
(p,d,q)x(P,D,Q)	(3,1,0)x(0,1,1)	(3,1,1)x(0,1,1)	(3,1,2)x(0,1,1)	(3,1,3)x(0,1,1)	(3,1,4)x(0,1,1)
AIC	-9,604690	-9,596515	-9,581757	-9,569480	-9,609068
SC	-9,471001	-9,440544	-9,403505	-9,368946	-9,386253
Prob(Q-test)	0,602	0,501	0,373	0,443	0,386
(p,d,q)x(P,D,Q)	(4,1,0)x(0,1,1)	(4,1,1)x(0,1,1)	(4,1,2)x(0,1,1)	(4,1,3)x(0,1,1)	(4,1,4)x(0,1,1)
AIC	-9,592713	-9,578905	-9,574113	-9,563746	-9,545360
SC	-9,435946	-9,399744	-9,372557	-9,339798	-9,299013
Prob(Q-test)	0,537	0,380	0,338	0,445	0,347

Tabell 11: ARIMA(p,d,q)x(P,D,0) med du1 og du2.

(p,d,q)x(P,D,Q)	(0,1,0)x(1,1,0)	(0,1,1)x(1,1,0)	(0,1,2)x(1,1,0)	(0,1,3)x(1,1,0)	(0,1,4)x(1,1,0)
AIC	-9,433086	-9,417177	-9,418221	-9,415902	-9,399266
SC	-9,363024	-9,323762	-9,301451	-9,275778	-9,235788
Prob(Q-test)	0,437	0,340	0,565	0,787	0,768
(p,d,q)x(P,D,Q)	(1,1,0)x(1,1,0)	(1,1,1)x(1,1,0)	(1,1,2)x(1,1,0)	(1,1,3)x(1,1,0)	(1,1,4)x(1,1,0)
AIC	-9,408005	-9,463957	-9,405495	-9,389764	-9,448734
SC	-9,314084	-9,346556	-9,264613	-9,225401	-9,260891
Prob(Q-test)	0,348	0,312	0,534	0,748	0,697
(p,d,q)x(P,D,Q)	(2,1,0)x(1,1,0)	(2,1,1)x(1,1,0)	(2,1,2)x(1,1,0)	(2,1,3)x(1,1,0)	(2,1,4)x(1,1,0)
AIC	-9,399817	-9,398170	-9,403587	-9,396058	-9,379373
SC	-9,281776	-9,256520	-9,238329	-9,207191	-9,166898
Prob(Q-test)	0,517	0,693	0,698	0,829	0,778
(p,d,q)x(P,D,Q)	(3,1,0)x(1,1,0)	(3,1,1)x(1,1,0)	(3,1,2)x(1,1,0)	(3,1,3)x(1,1,0)	(3,1,4)x(1,1,0)
AIC	-9,420971	-9,404135	-9,387526	-9,532187	-9,514026
SC	-9,278544	-9,237970	-9,197623	-9,318547	-9,276647
Prob(Q-test)	0,854	0,816	0,786	0,836	0,765
(p,d,q)x(P,D,Q)	(4,1,0)x(1,1,0)	(4,1,1)x(1,1,0)	(4,1,2)x(1,1,0)	(4,1,3)x(1,1,0)	(4,1,4)x(1,1,0)
AIC	-9,395241	-9,381031	-9,374869	-9,430070	-9,497419
SC	-9,228159	-9,190079	-9,160048	-9,191380	-9,234861
Prob(Q-test)	0,847	0,849	0,795	0,291	0,159

Tabell 12: ARIMA(p,d,q)x(0,D,0) med du1 og du2.

(p,d,q)x(P,D,Q)	(0,1,0)x(0,1,0)	(0,1,1)x(0,1,0)	(0,1,2)x(0,1,0)	(0,1,3)x(0,1,0)	(0,1,4)x(0,1,0)
AIC	-9,094549	-9,082222	-9,096293	-9,115022	-9,123283
SC	-9,050653	-9,016378	-9,008500	-9,005282	-8,991594
Prob(Q-test)	0,000	0,000	0,000	0,000	0,002
(p,d,q)x(P,D,Q)	(1,1,0)x(0,1,0)	(1,1,1)x(0,1,0)	(1,1,2)x(0,1,0)	(1,1,3)x(0,1,0)	(1,1,4)x(0,1,0)
AIC	-9,076118	-9,077655	-9,138677	-9,122179	-9,142430
SC	-9,009944	-8,989424	-9,028388	-8,989831	-8,988025
Prob(Q-test)	0,000	0,000	0,000	0,000	0,006
(p,d,q)x(P,D,Q)	(2,1,0)x(0,1,0)	(2,1,1)x(0,1,0)	(2,1,2)x(0,1,0)	(2,1,3)x(0,1,0)	(2,1,4)x(0,1,0)
AIC	-9,120624	-9,157027	-9,230530	-9,219980	-9,159175
SC	-9,031948	-9,046181	-9,097516	-9,064797	-8,981822
Prob(Q-test)	0,000	0,000	0,000	0,000	0,000
(p,d,q)x(P,D,Q)	(3,1,0)x(0,1,0)	(3,1,1)x(0,1,0)	(3,1,2)x(0,1,0)	(3,1,3)x(0,1,0)	(3,1,4)x(0,1,0)
AIC	-9,145227	-9,131141	-9,186100	-9,190440	-9,155327
SC	-9,033820	-8,997452	-9,030129	-9,012188	-8,954793
Prob(Q-test)	0,000	0,000	0,000	0,000	0,000
(p,d,q)x(P,D,Q)	(4,1,0)x(0,1,0)	(4,1,1)x(0,1,0)	(4,1,2)x(0,1,0)	(4,1,3)x(0,1,0)	(4,1,4)x(0,1,0)
AIC	-9,094549	-9,082222	-9,096293	-9,115022	-9,123283
SC	-9,050653	-9,016378	-9,008500	-9,005282	-8,991594
Prob(Q-test)	0,000	0,000	0,000	0,000	0,002

Tabell 13: ARIMA(p,d,q)x(P,D,Q) med konstant, du1 og du2.

(p,d,q)x(P,D,Q)	(0,1,0)x(1,1,1)	(0,1,1)x(1,1,1)	(0,1,2)x(1,1,1)	(0,1,3)x(1,1,1)	(0,1,4)x(1,1,1)
AIC	-9,645685	-9,634579	-9,622820	-9,609140	-9,595149
SC	-9,528915	-9,494455	-9,459342	-9,422308	-9,384963
Prob(Q-test)	0,216	0,212	0,217	0,235	0,120
(p,d,q)x(P,D,Q)	(1,1,0)x(1,1,1)	(1,1,1)x(1,1,1)	(1,1,2)x(1,1,1)	(1,1,3)x(1,1,1)	(1,1,4)x(1,1,1)
AIC	-9,631742	-9,614973	-9,608272	-9,593308	-9,552025
SC	-9,490860	-9,450611	-9,420429	-9,381984	-9,317221
Prob(Q-test)	0,172	0,149	0,148	0,167	0,767
(p,d,q)x(P,D,Q)	(2,1,0)x(1,1,1)	(2,1,1)x(1,1,1)	(2,1,2)x(1,1,1)	(2,1,3)x(1,1,1)	(2,1,4)x(1,1,1)
AIC	-9,648957	-9,631417	-9,618503	-9,601881	-9,586852
SC	-9,483698	-9,442551	-9,406028	-9,365798	-9,327161
Prob(Q-test)	0,359	0,287	0,222	0,199	0,119
(p,d,q)x(P,D,Q)	(3,1,0)x(1,1,1)	(3,1,1)x(1,1,1)	(3,1,2)x(1,1,1)	(3,1,3)x(1,1,1)	(3,1,4)x(1,1,1)
AIC	-9,665737	-9,650651	-9,634238	-9,604834	-9,634687
SC	-9,475834	-9,437011	-9,396860	-9,343718	-9,349832
Prob(Q-test)	0,498	0,442	0,287	0,390	0,265
(p,d,q)x(P,D,Q)	(4,1,0)x(1,1,1)	(4,1,1)x(1,1,1)	(4,1,2)x(1,1,1)	(4,1,3)x(1,1,1)	(4,1,4)x(1,1,1)
AIC	-9,653071	-9,644224	-9,655437	-9,637336	-9,553576
SC	-9,438251	-9,405534	-9,392879	-9,350908	-9,243279
Prob(Q-test)	0,548	0,379	0,356	0,330	0,363

Tabell 14: ARIMA(p,d,q)x(0,D,Q) med konstant, du1 og du2.

(p,d,q)x(P,D,Q)	(0,1,0)x(0,1,1)	(0,1,1)x(0,1,1)	(0,1,2)x(0,1,1)	(0,1,3)x(0,1,1)	(0,1,4)x(0,1,1)
AIC	-9,615627	-9,601318	-9,602173	-9,596822	-9,582852
SC	-9,527835	-9,491578	-9,470485	-9,443186	-9,407267
Prob(Q-test)	0,438	0,414	0,336	0,296	0,221
(p,d,q)x(P,D,Q)	(1,1,0)x(0,1,1)	(1,1,1)x(0,1,1)	(1,1,2)x(0,1,1)	(1,1,3)x(0,1,1)	(1,1,4)x(0,1,1)
AIC	-9,600197	-9,605804	-9,617656	-9,604696	-9,589616
SC	-9,489907	-9,473457	-9,463251	-9,428232	-9,391094
Prob(Q-test)	0,387	0,317	0,201	0,234	0,181
(p,d,q)x(P,D,Q)	(2,1,0)x(0,1,1)	(2,1,1)x(0,1,1)	(2,1,2)x(0,1,1)	(2,1,3)x(0,1,1)	(2,1,4)x(0,1,1)
AIC	-9,609233	-9,602289	-9,592735	-9,584870	-9,569750
SC	-9,476219	-9,447105	-9,415382	-9,385348	-9,348059
Prob(Q-test)	0,404	0,304	0,207	0,282	0,219
(p,d,q)x(P,D,Q)	(3,1,0)x(0,1,1)	(3,1,1)x(0,1,1)	(3,1,2)x(0,1,1)	(3,1,3)x(0,1,1)	(3,1,4)x(0,1,1)
AIC	-9,606489	-9,602062	-9,591870	-9,579294	-9,556007
SC	-9,450518	-9,423811	-9,391337	-9,356479	-9,310910
Prob(Q-test)	0,593	0,503	0,363	0,293	0,379
(p,d,q)x(P,D,Q)	(4,1,0)x(0,1,1)	(4,1,1)x(0,1,1)	(4,1,2)x(0,1,1)	(4,1,3)x(0,1,1)	(4,1,4)x(0,1,1)
AIC	-9,593781	-9,583526	-9,562326	-9,550546	-9,548528
SC	-9,414620	-9,381969	-9,338374	-9,304199	-9,279786
Prob(Q-test)	0,594	0,449	0,291	0,247	0,265

Tabell 15: ARIMA(p,d,q)x(P,D,0) med konstant, du1 og du2.

(p,d,q)x(P,D,Q)	(0,1,0)x(1,1,0)	(0,1,1)x(1,1,0)	(0,1,2)x(1,1,0)	(0,1,3)x(1,1,0)	(0,1,4)x(1,1,0)
AIC	-9,436649	-9,419890	-9,426350	-9,417251	-9,401650
SC	-9,343233	-9,303120	-9,286227	-9,253773	-9,214819
Prob(Q-test)	0,470	0,411	0,657	0,789	0,795
(p,d,q)x(P,D,Q)	(1,1,0)x(1,1,0)	(1,1,1)x(1,1,0)	(1,1,2)x(1,1,0)	(1,1,3)x(1,1,0)	(1,1,4)x(1,1,0)
AIC	-9,411160	-9,446992	-9,406293	-9,391392	-9,387854
SC	-9,293758	-9,306110	-9,241930	-9,203549	-9,176531
Prob(Q-test)	0,412	0,313	0,746	0,754	0,772
(p,d,q)x(P,D,Q)	(2,1,0)x(1,1,0)	(2,1,1)x(1,1,0)	(2,1,2)x(1,1,0)	(2,1,3)x(1,1,0)	(2,1,4)x(1,1,0)
AIC	-9,407633	-9,403343	-9,477372	-9,394992	-9,378273
SC	-9,265983	-9,238085	-9,288506	-9,182517	-9,142189
Prob(Q-test)	0,610	0,748	0,516	0,851	0,819
(p,d,q)x(P,D,Q)	(3,1,0)x(1,1,0)	(3,1,1)x(1,1,0)	(3,1,2)x(1,1,0)	(3,1,3)x(1,1,0)	(3,1,4)x(1,1,0)
AIC	-9,417259	-9,400018	-9,383020	-9,515347	-9,385811
SC	-9,251094	-9,210115	-9,169379	-9,277968	-9,124695
Prob(Q-test)	0,860	0,828	0,789	0,832	0,733
(p,d,q)x(P,D,Q)	(4,1,0)x(1,1,0)	(4,1,1)x(1,1,0)	(4,1,2)x(1,1,0)	(4,1,3)x(1,1,0)	(4,1,4)x(1,1,0)
AIC	-9,391760	-9,379243	-9,377918	-9,457557	-9,454070
SC	-9,200808	-9,164422	-9,139229	-9,194998	-9,167642
Prob(Q-test)	0,865	0,867	0,824	0,404	0,267

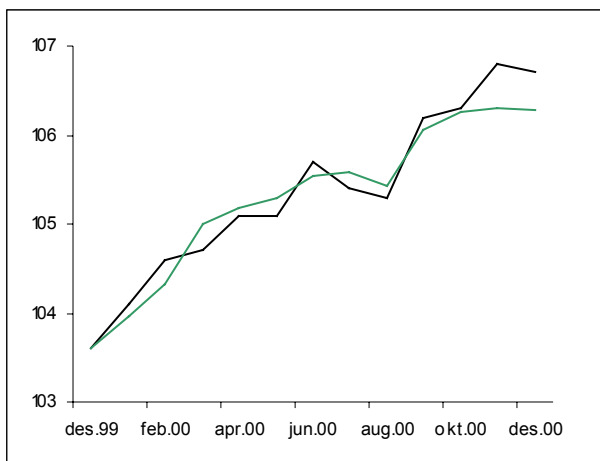
Tabell 16: ARIMA(p,d,q)x(0,D,0) med du1 og du2.

(p,d,q)x(P,D,Q)	(0,1,0)x(0,1,0)	(0,1,1)x(0,1,0)	(0,1,2)x(0,1,0)	(0,1,3)x(0,1,0)	(0,1,4)x(0,1,0)
AIC	-9,085123	-9,073607	-9,089308	-9,104344	-9,115219
SC	-9,019279	-8,985815	-8,979567	-8,972656	-8,961583
Prob(Q-test)	0,000	0,000	0,000	0,000	0,002
(p,d,q)x(P,D,Q)	(1,1,0)x(0,1,0)	(1,1,1)x(0,1,0)	(1,1,2)x(0,1,0)	(1,1,3)x(0,1,0)	(1,1,4)x(0,1,0)
AIC	-9,066482	-9,196247	-9,127567	-9,113322	-9,135540
SC	-8,978250	-9,085957	-8,995219	-8,958916	-8,959076
Prob(Q-test)	0,0000	0,0000	0,000	0,000	0,007
(p,d,q)x(P,D,Q)	(2,1,0)x(0,1,0)	(2,1,1)x(0,1,0)	(2,1,2)x(0,1,0)	(2,1,3)x(0,1,0)	(2,1,4)x(0,1,0)
AIC	-9,110559	-9,128506	-9,224635	-9,214484	-9,150797
SC	-8,999713	-8,995492	-9,069451	-9,037132	-8,951275
Prob(Q-test)	0,000	0,000	0,000	0,000	0,000
(p,d,q)x(P,D,Q)	(3,1,0)x(0,1,0)	(3,1,1)x(0,1,0)	(3,1,2)x(0,1,0)	(3,1,3)x(0,1,0)	(3,1,4)x(0,1,0)
AIC	-9,132719	-9,118829	-9,174408	-9,174408	-9,164501
SC	-8,999030	-8,962859	-8,996156	-8,996156	-8,941686
Prob(Q-test)	0,000	0,000	0,000	0,000	0,000
(p,d,q)x(P,D,Q)	(4,1,0)x(0,1,0)	(4,1,1)x(0,1,0)	(4,1,2)x(0,1,0)	(4,1,3)x(0,1,0)	(4,1,4)x(0,1,0)
AIC	-9,114368	-9,131892	-9,146131	-9,336474	-9,278759
SC	-8,957602	-8,952731	-8,944575	-9,112522	-9,032412
Prob(Q-test)	0,000	0,000	0,000	0,000	0,000

**Appendiks 2: Evaluering av
prognoseegenskaper.**

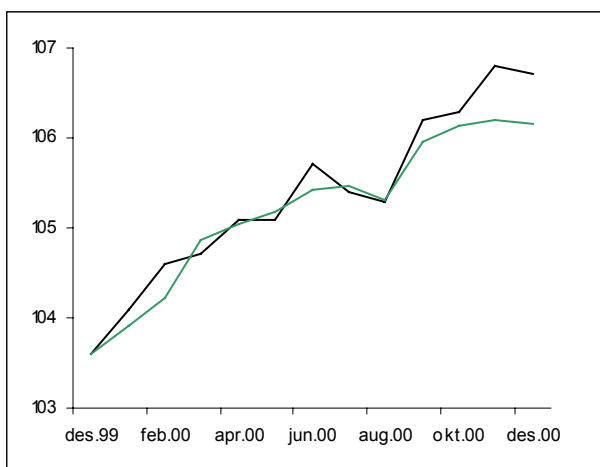
Modeller valgt ut på bakgrunn av AIC.

Figur 1 – ARIMA(3,1,0)x(1,1,1)



	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,252735	0,208221	0,192530	0,249145
Theil's U	0,001210	0,000993	0,000916	0,001181
U^M	0,023932	0,000154	0,009587	0,066365
U^S	0,466526	0,083335	0,022038	0,147280
U^K	0,509541	0,916510	0,968375	0,786355

Figur 2 – ARIMA(4,1,2)x(1,1,1)



	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,268595	0,269539	0,264356	0,401506
Theil's U	0,001287	0,001286	0,001258	0,001906
U^M	0,389506	0,511302	0,546898	0,566593
U^S	0,217491	0,001663	0,019718	0,169334
U^K	0,393002	0,487035	0,433384	0,264073

Figur 3 – ARIMA(3,1,4)x(1,1,1)

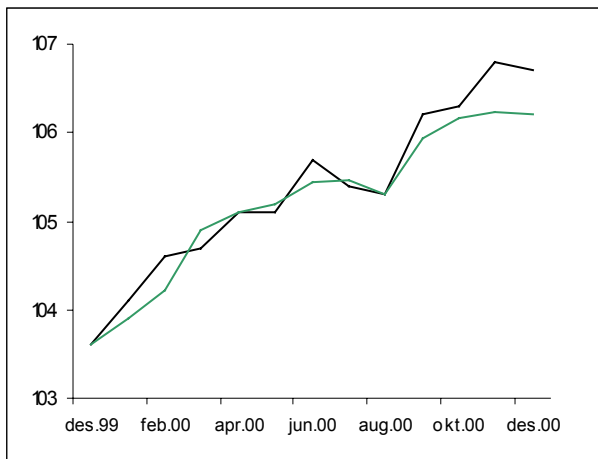


	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,274163	0,258424	0,255545	0,410302
Theil's U	0,001313	0,001233	0,001216	0,001947
U^M	0,418735	0,480084	0,491801	0,530801
U^S	0,204558	0,001069	0,010318	0,178379
U^K	0,376707	0,518847	0,497881	0,290819

Tidsserien simulert på bakgrunn av de estimerte modellene er gitt med grå grafer, mens de faktiske observasjonene er gitt med svarte grafer.

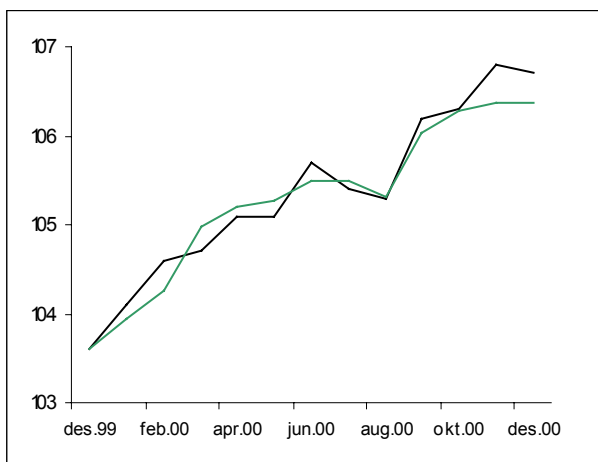
Modeller valgt ut på bakgrunn av SC.

Figur 4 – ARIMA(0,1,0)x(1,1,1)



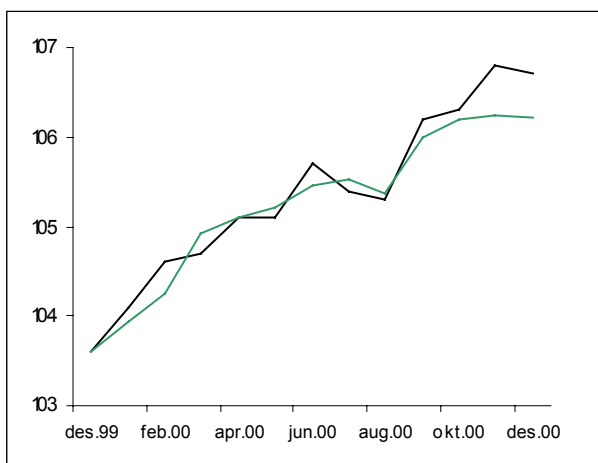
	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,266950	0,221529	0,201080	0,282402
Theil's U	0,001278	0,001057	0,000957	0,001339
U^M	0,187189	0,154617	0,146530	0,314176
U^S	0,325421	0,039992	0,002292	0,139544
U^K	0,487391	0,805391	0,851177	0,546280

Figur 5 – ARIMA(0,1,0)x(0,1,1)



	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,268633	0,220439	0,192005	0,228814
Theil's U	0,001286	0,001051	0,000913	0,001085
U^M	0,059083	0,010270	0,011088	0,125602
U^S	0,408317	0,080599	0,009378	0,107206
U^K	0,532600	0,909131	0,979534	0,767192

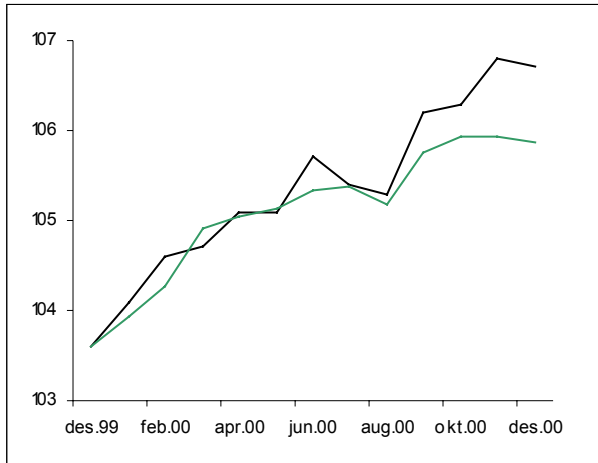
Figur 6 – ARIMA(1,1,0)x(1,1,1)



	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,258786	0,212538	0,192532	0,272736
Theil's U	0,001239	0,001014	0,000916	0,001293
U^M	0,138471	0,095333	0,059085	0,231066
U^S	0,343987	0,048828	0,009036	0,144384
U^K	0,517542	0,855840	0,931879	0,624550

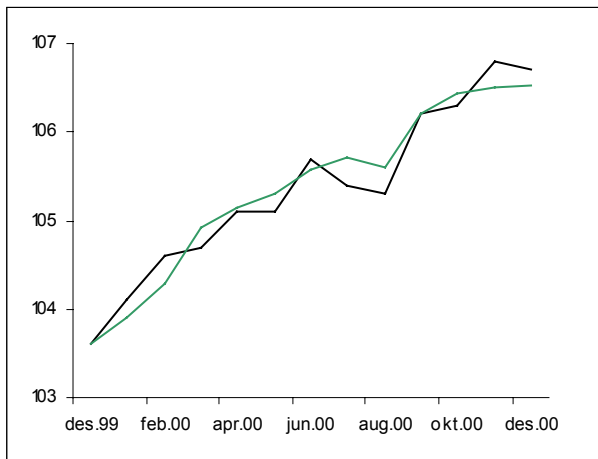
**Modeller med et konstantledd valgt ut
på bakgrunn av AIC.**

Figur 7 – ARIMA(3,1,0)x(0,1,1)



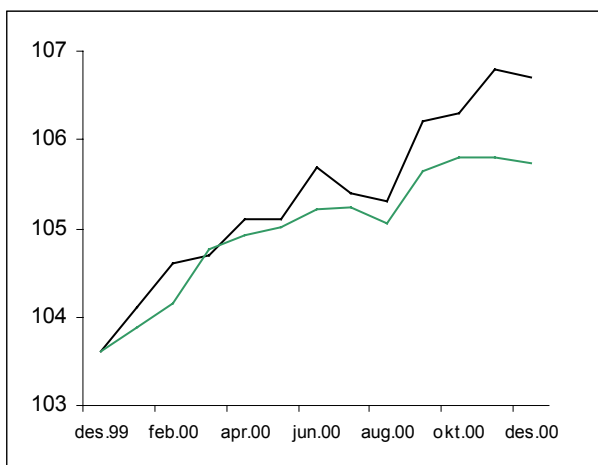
	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,253116	0,232832	0,243095	0,418979
Theil's U	0,001212	0,001111	0,001157	0,001988
U^M	0,169545	0,253367	0,351163	0,447784
U^S	0,332766	0,000737	0,040844	0,244421
U^K	0,497689	0,745897	0,607993	0,307795

Figur 8 – ARIMA(3,1,4)x(1,1,1)



	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,250838	0,204149	0,219504	0,217102
Theil's U	0,001201	0,000973	0,001044	0,001029
U^M	0,136937	0,013544	0,058875	0,003088
U^S	0,404139	0,177852	0,188774	0,000279
U^K	0,458924	0,808604	0,752351	0,996633

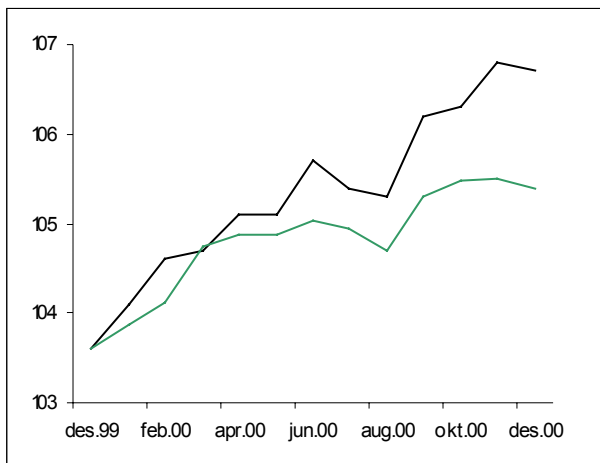
Figur 9 – ARIMA(2,1,0)x(1,1,1)



	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,289970	0,298164	0,322258	0,511498
Theil's U	0,001389	0,001423	0,001534	0,002429
U^M	0,463259	0,581036	0,643920	0,613573
U^S	0,147945	0,005011	0,047549	0,197780
U^K	0,388796	0,413953	0,308532	0,188647

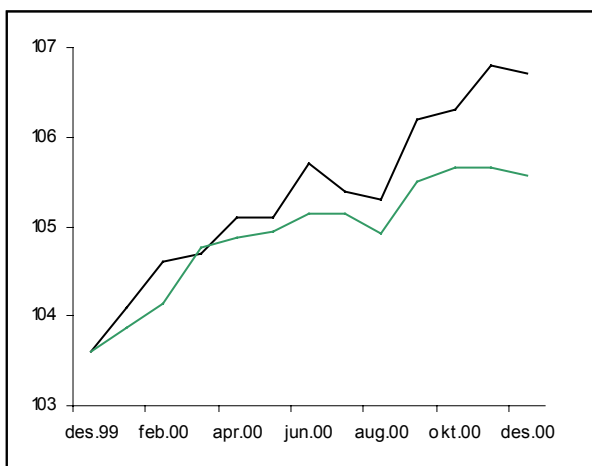
Modeller med et konstantledd valgt ut på bakgrunn av SC.

Figur 10 – ARIMA(0,1,0)x(0,1,1)



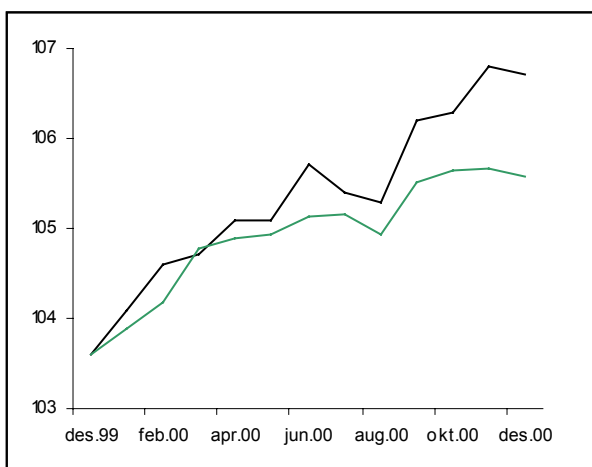
	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,310740	0,374407	0,495465	0,723623
Theil's U	0,001489	0,001787	0,002361	0,003439
U^M	0,495357	0,625281	0,699290	0,679782
U^S	0,121078	0,031357	0,106685	0,211527
U^K	0,383566	0,343362	0,194025	0,108691

Figur 11 – ARIMA(0,1,0)x(1,1,1)



	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,301986	0,331708	0,386797	0,600934
Theil's U	0,001447	0,001583	0,001842	0,002855
U^M	0,503849	0,625898	0,692708	0,649136
U^S	0,134560	0,013388	0,073814	0,206402
U^K	0,361591	0,360713	0,233477	0,144462

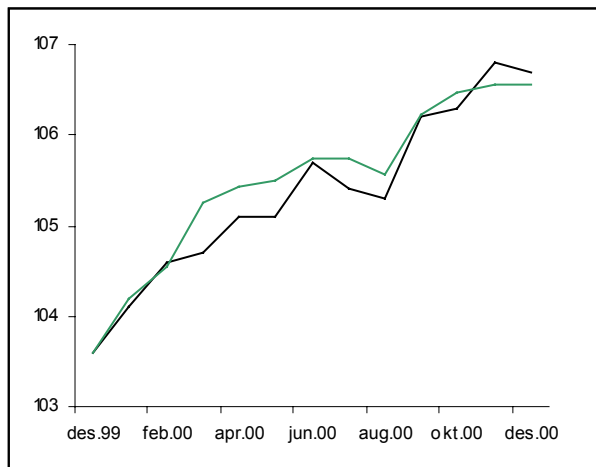
Figur 12 – ARIMA(0,1,1)x(1,1,1)



	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,281713	0,324214	0,380390	0,598230
Theil's U	0,001350	0,001547	0,001812	0,002842
U^M	0,460923	0,604414	0,677835	0,637889
U^S	0,145055	0,023977	0,089764	0,221437
U^K	0,394022	0,371609	0,232402	0,140675

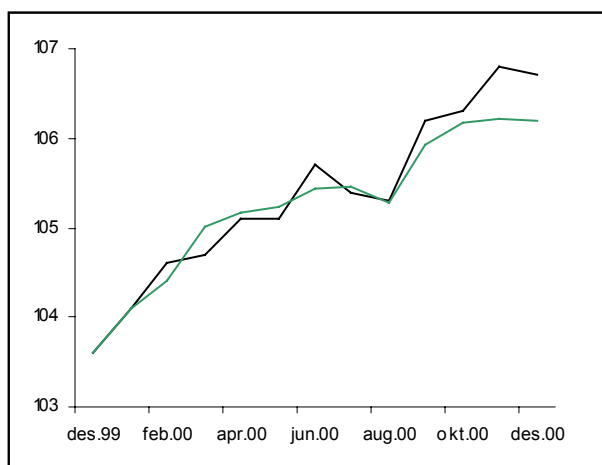
Modeller med du1 og du2 valgt ut på bakgrunn av AIC.

Figur 13 – ARIMA(3,1,0)x(1,1,1)



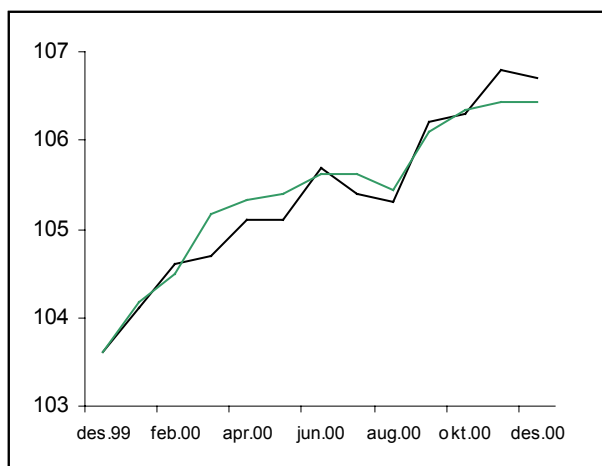
	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,321491	0,311270	0,290012	0,267671
Theil's U	0,001537	0,001482	0,001378	0,001268
U^M	0,386194	0,531269	0,568444	0,299343
U^S	0,284563	0,024899	0,000020	0,142295
U^K	0,329243	0,443832	0,431536	0,558362

Figur14 – ARIMA(4,1,2)x(1,1,1)



	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,217956	0,196560	0,186240	0,274840
Theil's U	0,001043	0,000937	0,000886	0,001303
U^M	0,032915	0,002863	0,009159	0,170987
U^S	0,320937	0,007453	0,106910	0,367282
U^K	0,646147	0,989684	0,883931	0,461731

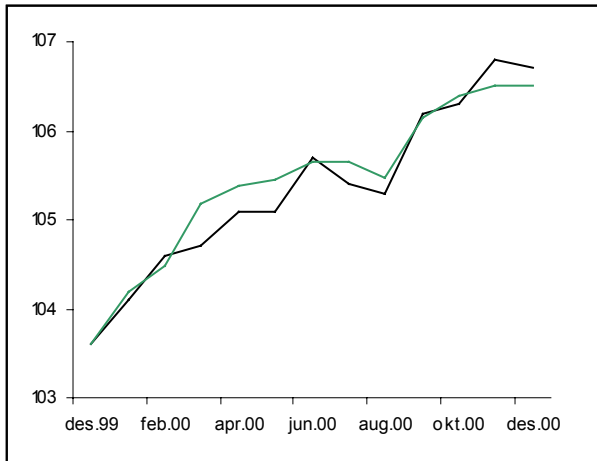
Figur 15 – ARIMA(4,1,0)x(1,1,1)



	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,279893	0,252820	0,226244	0,234951
Theil's U	0,001339	0,001204	0,001075	0,001113
U^M	0,264049	0,330495	0,309769	0,038612
U^S	0,303091	0,006480	0,016064	0,291883
U^K	0,432860	0,663025	0,674167	0,669504

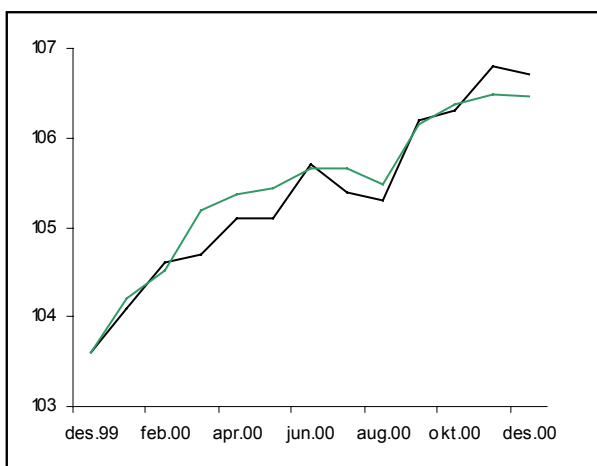
Modeller med du1 og du2 valgt ut på bakgrunn av SC.

Figur 16 – ARIMA(0,1,0)x(1,1,1)



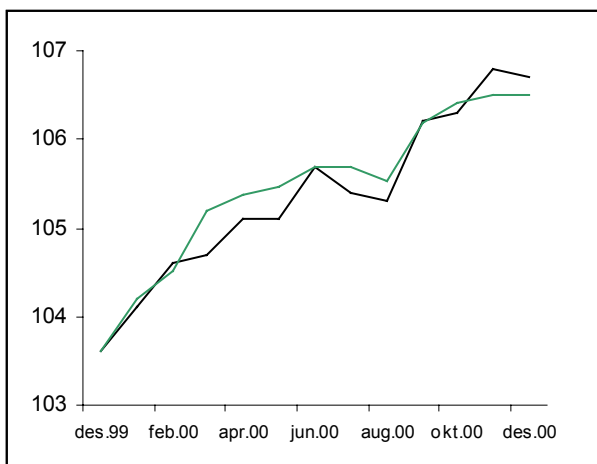
	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,284652	0,269755	0,245072	0,237834
Theil's U	0,001361	0,001285	0,001165	0,001127
U^M	0,284558	0,407310	0,417310	0,129479
U^S	0,273044	0,015845	0,003876	0,214135
U^K	0,442398	0,576844	0,578814	0,656386

Figur 17 – ARIMA(0,1,1)x(1,1,1)



	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,291911	0,272529	0,247143	0,245226
Theil's U	0,001396	0,001298	0,001174	0,001162
U^M	0,344423	0,432569	0,429094	0,108180
U^S	0,252649	0,006941	0,010835	0,268290
U^K	0,402929	0,560490	0,560071	0,623530

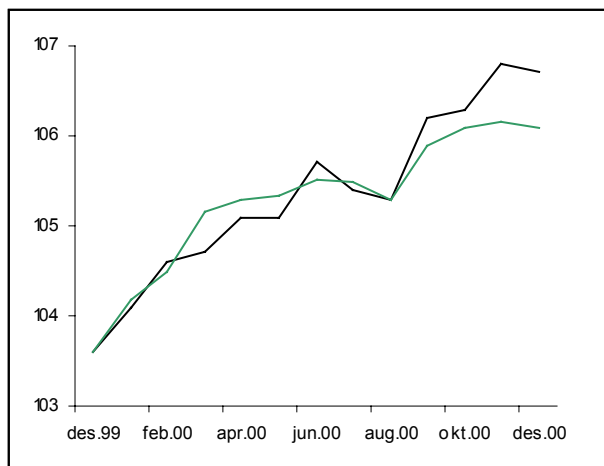
Figur 18 – ARIMA(1,1,0)x(1,1,1)



	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,289638	0,276015	0,257063	0,247249
Theil's U	0,001385	0,001315	0,001222	0,001171
U^M	0,312211	0,444426	0,479727	0,171815
U^S	0,260506	0,016914	0,000985	0,195399
U^K	0,427282	0,538660	0,519288	0,632786

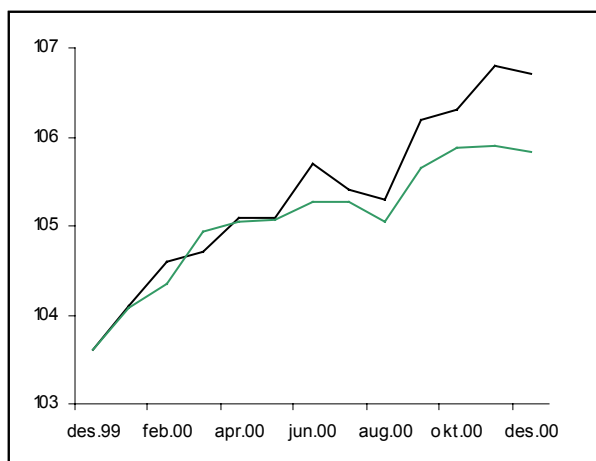
**Modeller med konstant, du_1 og du_2
valgt ut på bakgrunn av AIC.**

Figur 19 – ARIMA(3,1,0)x(1,1,1)



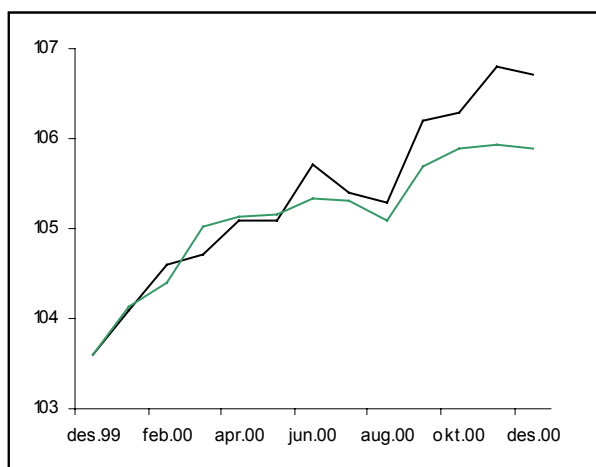
	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,269943	0,237736	0,221102	0,326041
Theil's U	0,001291	0,001133	0,001051	0,001546
U^M	0,269885	0,214423	0,046946	0,069606
U^S	0,268070	0,002914	0,156560	0,494158
U^K	0,462045	0,782663	0,796495	0,436236

Figur 20 – ARIMA(4,1,2)x(1,1,1)



	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,199851	0,229511	0,277238	0,452796
Theil's U	0,000957	0,001095	0,001320	0,002149
U^M	0,011760	0,180865	0,369552	0,470513
U^S	0,219820	0,089996	0,222652	0,345182
U^K	0,768420	0,729139	0,407796	0,184305

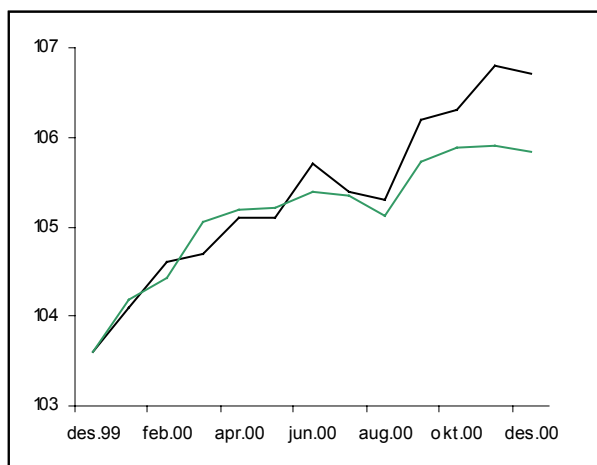
Figur 21 – ARIMA(4,1,0)x(1,1,1)



	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,218860	0,216775	0,255887	0,423660
Theil's U	0,001047	0,001034	0,001218	0,002010
U^M	0,043445	0,011640	0,171709	0,355877
U^S	0,254806	0,069469	0,264979	0,412812
U^K	0,701749	0,918891	0,563312	0,231312

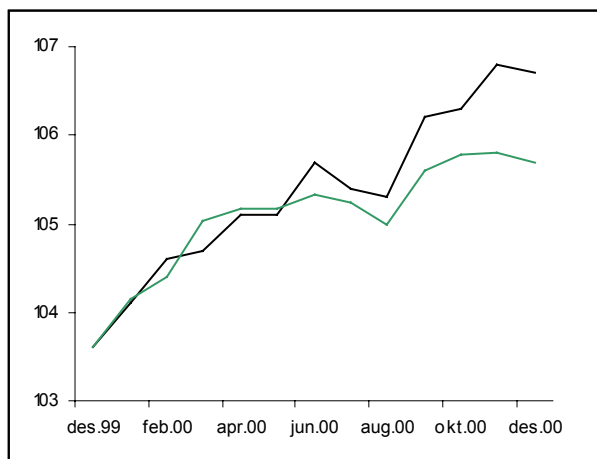
**Modeller med konstant, du_1 og du_2
valgt ut på bakgrunn av SC.**

Figur 22 – ARIMA(0,1,0)x(1,1,1)



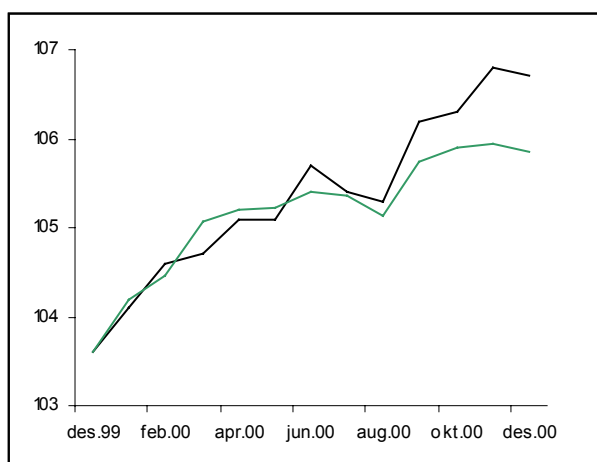
	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,230455	0,215442	0,246638	0,431772
Theil's U	0,001103	0,001027	0,001173	0,002048
U^M	0,152278	0,016327	0,060330	0,271587
U^S	0,205894	0,064068	0,303584	0,469019
U^K	0,641828	0,919604	0,633086	0,259394

Figur 23 – ARIMA(0,1,0)x(0,1,1)



	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,230697	0,228477	0,297057	0,507145
Theil's U	0,001104	0,001089	0,001414	0,002407
U^M	0,072951	0,002658	0,183732	0,365861
U^S	0,228833	0,075223	0,280724	0,414903
U^K	0,698216	0,922119	0,535544	0,219236

Figur 24 – ARIMA(0,1,0)x(1,1,1)



	3 mnd	6 mnd	9 mnd	12 mnd
mrs	0,234335	0,217013	0,241522	0,420401
Theil's U	0,001121	0,001034	0,001149	0,001994
U^M	0,191991	0,029566	0,039900	0,250206
U^S	0,204791	0,067251	0,312236	0,483476
U^K	0,603218	0,903183	0,647864	0,266318

Appendiks 3: Prognose for 2001¹⁹

Figur 1: ARIMA (p,d,q)x(P,D,Q).

(p,d,q)x(P,D,Q)	(0,1,0)x(1,1,1)	(0,1,1)x(1,1,1)	(0,1,2)x(1,1,1)	(0,1,3)x(1,1,1)	(0,1,4)x(1,1,1)
AIC	-9,402444	-9,394528	-9,379654	-9,369534	-9,355679
SC	-9,358548	-9,328683	-9,291861	-9,259793	-9,223991
Prob(Q-test)	0,790	0,702	0,660	0,742	0,726
(p,d,q)x(P,D,Q)	(1,1,0)x(1,1,1)	(1,1,1)x(1,1,1)	(1,1,2)x(1,1,1)	(1,1,3)x(1,1,1)	(1,1,4)x(1,1,1)
AIC	-9,385346	-9,381532	-9,367003	-9,354813	-9,341564
SC	-9,319172	-9,293300	-9,256713	-9,222465	-9,187158
Prob(Q-test)	0,712	0,735	0,699	0,760	0,738
(p,d,q)x(P,D,Q)	(2,1,0)x(1,1,1)	(2,1,1)x(1,1,1)	(2,1,2)x(1,1,1)	(2,1,3)x(1,1,1)	(2,1,4)x(1,1,1)
AIC	-9,402108	-9,394310	-9,386552	-9,372066	-9,346411
SC	-9,313432	-9,283464	-9,253537	-9,216882	-9,169058
Prob(Q-test)	0,792	0,769	0,738	0,666	0,626
(p,d,q)x(P,D,Q)	(3,1,0)x(1,1,1)	(3,1,1)x(1,1,1)	(3,1,2)x(1,1,1)	(3,1,3)x(1,1,1)	(3,1,4)x(1,1,1)
AIC	-9,406631	-9,399296	-9,353089	-9,380890	-9,383137
SC	-9,295223	-9,265607	-9,197119	-9,202638	-9,182604
Prob(Q-test)	0,913	0,890	0,948	0,880	0,843
(p,d,q)x(P,D,Q)	(4,1,0)x(1,1,1)	(4,1,1)x(1,1,1)	(4,1,2)x(1,1,1)	(4,1,3)x(1,1,1)	(4,1,4)x(1,1,1)
AIC	-9,389406	-9,373680	-9,367904	-9,309718	-9,364115
SC	-9,255035	-9,216914	-9,188742	-9,108162	-9,140163
Prob(Q-test)	0,879	0,849	0,834	0,429	0,732

Figur 2: ARIMA (p,d,q)x(0,D,Q).

(p,d,q)x(P,D,Q)	(0,1,0)x(0,1,1)	(0,1,1)x(0,1,1)	(0,1,2)x(0,1,1)	(0,1,3)x(0,1,1)	(0,1,4)x(0,1,1)
AIC	-9,398500	-9,388122	-9,374729	-9,381571	-9,367822
SC	-9,377780	-9,346683	-9,312572	-9,298694	-9,264226
Prob(Q-test)	0,794	0,730	0,682	0,765	0,715
(p,d,q)x(P,D,Q)	(1,1,0)x(0,1,1)	(1,1,1)x(0,1,1)	(1,1,2)x(0,1,1)	(1,1,3)x(0,1,1)	(1,1,4)x(0,1,1)
AIC	-9,384275	-9,371247	-9,357744	-9,362885	-9,349022
SC	-9,342644	-9,308800	-9,274481	-9,258806	-9,224128
Prob(Q-test)	0,716	0,620	0,558	0,744	0,687
(p,d,q)x(P,D,Q)	(2,1,0)x(0,1,1)	(2,1,1)x(0,1,1)	(2,1,2)x(0,1,1)	(2,1,3)x(0,1,1)	(2,1,4)x(0,1,1)
AIC	-9,372947	-9,360745	-9,351536	-9,344246	-9,330403
SC	-9,310208	-9,277092	-9,246970	-9,218767	-9,184010
Prob(Q-test)	0,608	0,573	0,501	0,660	0,571
(p,d,q)x(P,D,Q)	(3,1,0)x(0,1,1)	(3,1,1)x(0,1,1)	(3,1,2)x(0,1,1)	(3,1,3)x(0,1,1)	(3,1,4)x(0,1,1)
AIC	-9,372536	-9,358833	-9,346044	-9,335723	-9,323431
SC	-9,288489	-9,253775	-9,219974	-9,188641	-9,155337
Prob(Q-test)	0,779	0,751	0,680	0,757	0,775
(p,d,q)x(P,D,Q)	(4,1,0)x(0,1,1)	(4,1,1)x(0,1,1)	(4,1,2)x(0,1,1)	(4,1,3)x(0,1,1)	(4,1,4)x(0,1,1)
AIC	-9,359180	-9,344891	-9,330728	-9,318350	-9,325544
SC	-9,253624	-9,218223	-9,182949	-9,149460	-9,135542
Prob(Q-test)	0,771	0,732	0,676	0,615	0,675

¹⁹ Jeg har her bare prøvd dette datasettet for ARIMA(p,d,q)x(P,D,Q)- og ARIMA(p,d,q)x(0,D,Q)-modeller da erfaringer fra tidligere tilsier at ARIMA(p,d,q)x(P,D,0)- og ARIMA(p,d,q)x(0,D,0) ikke er relevante.