

Interne notater

STATISTISK SENTRALBYRÅ

87/46

17. november 1987

KALMANFILTERET SOM TEKNIKK FOR ØKONOMETRISK ANALYSE¹⁾

AV

LARS LOE, HILDE OLSEN OG TORE SCHWEDER

INNHold

	Side
1. Innledning	1
2. Tilstandsmodeller, noen eksempler	2
3. Kalman-filter	7
4. Framskrivning - glatting	11
5. Estimering av ukjente parametre	12
6. Estimering av de relative variansene	14
7. Filteret brukt på en feil-korreksjonsmodell	17
8. Prediksjon av regresjonsresidualer	19
9. Linearisering av feilkorreksjonsmodellen	22
10. Fler-likningsmodeller	25
11. Empiriske eksempler	25
Referanser	38

1) Dette notatet inneholder en dokumentasjon av et arbeid som er gjort i Statistisk Sentralbyrå med å bygge opp kompetanse på Kalmanfilterteknikk og i bruk av slike modeller på utvalgte problemstillinger. Erik Biørn og Anders Rygh Swensen har også deltatt i prosjektet.

1. Innledning

Kalmanfilteret ble opprinnelig utviklet og tatt i bruk av kontrollingeniører. Både når det gjelder tradisjonelle anvendelsesområder og terminologi, har det derfor liten tilknytning til økonomifaget. Dette har også i en viss grad preget holdningen til å bruke det for å belyse økonomiske problemstillinger.

Rudolf Kalman har hevdet at Kalmanfilteret er en regneteknikk og at det ikke har noen sammenheng med økonomiske modeller. Andrew C. Harvey har hevdet noe av det samme: Kalmanfiltermodeller er ikke spesielt interessante fra en økonomisk-teoretisk synsvinkel, men som statistisk modell har de flere gode egenskaper.

Likevel har Kalmanfilterteknikk i de seinere år fått en viss utbredelse i økonomiske anvendelser. Dette arbeidet har som formål å bygge opp kompetanse på dette området, og vurdere hvilke muligheter og eventuelle fordeler bruk av denne teknikken kan ha for anvendelser i Byrået.

I hovedsak baserer denne studien seg på Harveys bok "Time Series Models" (Harvey (1981a)), og på forskjellige andre arbeider han har skrevet. Utgangspunktet for gjennomgangen er at Kalmanfilteret:

1. Gir en felles ramme for å spesifisere mange forskjellige økonomiske modeller.
2. Ved å skrive ut den aktuelle modellen i tilstandsroms form vil man ofte bli hjulpet til å brette opp modellen i deler som har en intuitiv forståelse. Man spesifiserer både observerbare og latente variable i modellen. Man kan sette navn på modellens enkelte deler og man kan forstå hvordan disse samvirker.
3. Filteret gir anledning til effektiv estimering av underliggende parametre ved hjelp av Maximum-Likelihood-(ML-)metoden.
4. Filteret vil gi estimater (prediksjoner) for underliggende, latente størrelser. Seriene av disse estimatene kan være instruktive for å forstå de økonomiske mekanismene som modellen skal beskrive.
5. Filteret gir anledning til rekursiv oppdatering og estimering. Man trenger ikke full reestimering for å inkorporere siste års data!

2. Tilstandsmodeller, noen eksempler

For at Kalmanfilter-teknikk skal kunne brukes, må modellen kunne skrives på tilstandsroms form. Den sentrale komponenten i tilstandsmodellen er et sett av tilstandsvariable, α_t , som endres over tid. Disse er ofte parametre eller latente variable. For å kunne tallfeste disse effektene må de knyttes til målbare størrelser. En slik spesifisering av sammenhenger mellom observerbare og ikke observerbare variable kalles en tilstandsmodell.

Tilstandsmodellen består av to likninger:

$$(M) \quad y_t = Z_t' \alpha_t + \zeta_t \quad \text{Målelikning}$$

$$(T) \quad \alpha_t = T_t \alpha_{t-1} + \eta_t \quad \text{Transisjonslikning}$$

$$t = 1, \dots, T$$

der

y_t = responsvariabel (en observerbar, endogen variabel)

Z_t' = vektor av koeffisienter eller eksogene variable

α_t = vektor av tilstandsvariable, noen ganger parametre (latente variable uten variasjon)

T_t = matrise som spesifiserer dynamikken i den underliggende prosessen

ζ og η -ene, uavhengige restledd, $\xi_t \sim N(0, \sigma^2 h_t)$
 $\eta_t \sim N(0, \sigma^2 Q_t)$

h_t er en fast skalar, Q_t en kovariansmatrise og σ^2 en (ukjent) varians.

Målelikningen knytter sammenhengen mellom de målbare og de latente variable (tilstandsvariable), mens transisjonslikningen styrer dynamikken i de tilstandsvariable (de ukjente størrelsene). Tilstandsmodeller inne-

holder således ikke noen "ny teori", men er en modellformulering som ofte vil være hensiktsmessig for estimering, prediksjon og tolkning av den teori som skal modelleres. Vi skal nå vise eksempler på hvordan noen kjente modeller kan formuleres i tilstandsform.

Eksempel 1 Standard regresjon i Kalmanfilterform:

Her er

$Z_t = x_t$; vektoren av forklaringsvariable - observasjoner på tidspunkt t .

$\alpha_t = \beta$; vektoren av regresjonsparametre (tilstandsvariable).

$T_t = I$ (identitetsmatrisen); dvs. ingen dynamikk i β .

$Q_t = 0$; dvs. ingen fluktusjon i β .

Tilstandsmodellen kan dermed skrives:

$$(M) \quad y_t = x_t' \beta + \zeta_t$$

$$[(T) \quad \beta = \beta]$$

Eksempel 2 ARMA(p,q) modell:

For

$$p = 2, \quad q = 2$$

er

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t$$

som kan skrives:

$$(T) \begin{bmatrix} y_t \\ y_{t-1} \\ \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix} = \begin{bmatrix} \varphi_1 & \varphi_2 & \Theta_1 & \Theta_2 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \varepsilon_{t-1} \\ \varepsilon_{t-2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \varepsilon_t$$

eller på matriseform:

$$(T) \quad \alpha_t = T\alpha_{t-1} + \eta_t$$

$$(M) \quad y_t = [1 \ 0 \ 0 \ 0]\alpha_t$$

På tilsvarende vis kan den generelle ARMA(p,q) modellen

$$y_t = \sum_{j=1}^p \varphi_j y_{t-j} + \sum_{j=1}^q \Theta_j \varepsilon_{t-j} + \varepsilon_t$$

skrives på tilstandsform.

Kalmanfilteret kan også brukes for å forene regresjonsmetodikk og Box-Jenkins ARMA-prosessmetodikk:

Eksempel 3 Regresjonsmodell med ARMA(p,q) restledd:

$$y_t = x_t \beta + \zeta_t$$

$$\zeta_t = \varphi_1 \zeta_{t-1} + \dots + \varphi_p \zeta_{t-p} + \varepsilon_t + \Theta_1 \varepsilon_{t-1}$$

For å forenkle forutsettes det å inngå bare én eksogen variabel.

Her kan vi skrive (p=2, q=2):

$$\alpha_t = \begin{bmatrix} \beta \\ \zeta_t \\ \zeta_{t-1} \\ \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \varphi_1 & \varphi_2 & \Theta_1 & \Theta_2 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta \\ \zeta_{t-1} \\ \zeta_{t-2} \\ \varepsilon_{t-1} \\ \varepsilon_{t-2} \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \varepsilon_t$$

eller modellen på matriseform:

$$(T) \quad \alpha_t = T\alpha_{t-1} + \eta_t$$

$$(M) \quad y_t = [x_t \ 1 \ 0 \dots 0]\alpha_t = x_t\beta + \zeta_t$$

Filtermodellene trenger ikke bare ha form av slike direkte kryssninger mellom ARMA-prosesser og lineære regresjonsstrukturer. Som eksempel på en annen type modell presenterer vi:

Eksempel 4 Harvey's "strukturmodell" for sesongjustering og trendframskrivning (se f.eks. Harvey (1981a):

Etter gammel "oppskrift" som f.eks. i X11-metoden (se Shiskin et al. (1967)) tenker en seg at serien (muligens etter logaritmisk transformasjon) kan skrives som summen av trend, sesong og tilfeldig komponent:

$$y_t = \mu_t + \gamma_t + \varepsilon_t$$

- trendnivået μ_t har en veksthastighet på β :

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t^\mu$$

- veksten er gjennomgående stabil: $\beta_t = \beta_{t-1} + \eta_t^\beta$

- og (de kvartalsvise) sesongkomponentene, γ , nuller hverandre tilnærmet ut over året:

$$\gamma_t = -\gamma_{t-1} - \gamma_{t-2} - \gamma_{t-3} + \eta_t^\gamma$$

$\eta_t^H \sim N(0, \sigma_H^2)$ og uavhengige; $H = \mu, \beta, \gamma$

På matriseform kan modellen skrives:

$$(T) \alpha_t = \begin{bmatrix} \mu_t \\ \beta_t \\ \gamma_t \\ \gamma_{t-1} \\ \gamma_{t-2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \gamma_{t-1} \\ \gamma_{t-2} \\ \gamma_{t-3} \end{bmatrix} + \begin{bmatrix} \eta_t^\mu \\ \eta_t^\beta \\ \eta_t^\gamma \\ 0 \\ 0 \end{bmatrix}$$

$$(M) y_t = [1 \ 0 \ 1 \ 0 \ 0] \alpha_t + \varepsilon_t$$

Denne modellen inneholder stokastiske trend- og sesongkomponenter som gjør den mer fleksibel enn modeller med deterministiske trend- og sesongkomponenter. Eksempler på bruk av denne modellen til sesongjustering, prediksjon etc. er gitt i avsn. 11. Se også Loe (1987).

I en slik modell kan sesongmønsteret endres over tid samtidig som forventningsverdien av sesongkomponentene, summert over et år, er tilnærmet lik null. Hvis σ_Y^2 er svært liten, er sesongmønsteret stabilt. Stor verdi på σ_Y^2 indikerer derimot at sesongmønsteret er varierende. Hvordan variasjoner i observasjonsmaterialet kommer til syne i estimatene av de ulike koeffisientene avhenger av de relative variansene til de 3 komponentene ettersom vi tidligere har uttrykt kovariansmatrisen ved:

$$\Sigma_{\alpha_t} = \begin{bmatrix} \sigma_\mu^2 & \dots & \dots & \dots & 0 \\ & \sigma_\beta^2 & & & \\ & & \sigma_\gamma^2 & & \\ & & & 0 & \\ 0 & \dots & \dots & \dots & 0 \end{bmatrix} = \sigma^2 Q$$

I standardmodellen for Kalmanfilteret antas det at kovariansmatrisen til ε_t er kjent med unntak av den felles skalafaktoren σ^2 . Dette innebærer at de relative variansene σ_μ^2/σ^2 og σ_β^2/σ^2 antas kjent. Noen ganger kan denne antakelsen være urimelig og det er umulig å angi matrisen Q . Da må de relative variansene estimeres, noe som kompliserer tilpassingen av modellen. Dette kommer vi tilbake til i avsnitt 6.

Denne modellen finnes i ulike versjoner. En modifikasjon består i å la trenden bestå av en stokastisk komponent og en globalt stasjonær

autoregressivt bestemt komponent (se Gersch og Kitagawa (1983)). Utgangspunktet for denne formuleringen er at trenden sett i et langsiktig perspektiv er svært glatt, mens en lokal trend kan vise betydelige svingninger. Hos Gersch og Kitagawa er det således eksempler som viser at det kan være optimalt å velge ulike modellformuleringer, avhengig av tidshorisonen på den analysen som skal gjennomføres. Dette blir også kommentert av de Vos (1986). En tilsvarende modifikasjon kan gjøres i formuleringen av sesongkomponenter.

Eksempel 5

Målevariabelen y kan også være en vektor. Modellen:

$$(M) \quad y_t = x_t B_t + \zeta_t$$

$$(T) \quad B_t = B_{t-1} + \eta_t$$

der y_t er vektoren av endogene variable og x_t vektoren av eksogene variable, vil for eksempel representere et simultant likningssystem på redusert form, men hvor koeffisientmatrisen B tillates å endres over tid.

Mange forskjellige typer modeller kan altså skrives på tilstandsmodellform. Kravet er:

1. Modellens dynamikk er lineær.
2. Ved å ta med mange nok lag i tilstandsvektoren (la den ha høy nok dimensjon) er avhengigheten Markovsk (ett-trinns).
3. Målemodellen er lineær.

3. Kalmanfilter

Selve Kalmanfilteret er en regnerutine for å estimere slike tilstandsmodeller som ble beskrevet i forrige avsnitt og evt. bruke dem til f.eks. prediksjonsformål. Filteret består av et sett av rekursjonslikninger som brukes for å oppdatere estimatene av de tilstandsvariable etter hvert som nye observasjoner blir tilgjengelige. For å illustrere denne algo-

ritmen tar vi utgangspunkt i modellen med univariat y (når y er multivariat blir likningene tilsvarende, se Harvey (1981a)):

$$(M) \quad y_t = Z_t' \alpha_t + \zeta_t; \quad \zeta_t \sim N(0, \sigma^2 h_t)$$

$$(T) \quad \alpha_t = T_t \alpha_{t-1} + \eta_t; \quad \eta_t \sim N(0, \sigma^2 Q_t)$$

For en gitt lineær tilstandsmodell kan tilstandsvektoren fram- og tilbakeskrives ved hjelp av Kalmanfilteret. Disse framskrivingene er optimale i minste kvadraters forstand.

Vi definerer:

a_{t-1} = minste kvadraters prediktoren for α_{t-1} , gitt all informasjon opp til og med tidspunkt $t-1$.

H_{t-1} = observasjoner av den målevariable opp til og med tidspunkt $t-1$, (y_{t-1}, y_{t-2}, \dots).

$$a_{t-1} = E(\alpha_{t-1} | H_{t-1}).$$

$$P_{t-1} = \frac{1}{\sigma^2} \cdot \text{kovariansmatrisen til } a_{t-1} - \alpha_{t-1}.$$

Filteret består nå av 1. trinns prediksjoner og oppdatering:

$$a_{t|t-1} = E(\alpha_t | H_{t-1}) \text{ er minste kvadraters 1. skrittsprediksjonen}$$

På tidspunkt $t-1$ er all tilgjengelig informasjon innkorporert i a_{t-1} , og kovariansmatrisen er $\sigma^2 P_{t-1}$. Ved å bruke transisjonslikningen kan vi dermed gjøre følgende betingede prediksjoner av tilstandsvektoren og kovariansmatrisen:

$$\text{Prediksjonslikning} \quad \begin{cases} a_{t|t-1} = T_t a_{t-1} \\ P_{t|t-1} = T_t P_{t-1} T_t' + Q_t \end{cases}$$

$$\sigma^2 P_{t|t-1} \text{ er kovariansmatrisen til } \alpha_t - a_{t|t-1}$$

Når ny informasjon er tilgjengelig, oppdateres estimatene for tilstandsvektoren og kovariansmatrisen:

$$\text{Oppdateringslikning} \quad \begin{cases} a_t &= a_{t|t-1} + P_{t|t-1} Z_t' (y_t - Z_t' a_{t|t-1}) / f_t \\ P_t &= P_{t|t-1} - P_{t|t-1} Z_t Z_t' P_{t|t-1} / f_t \\ f_t &= Z_t' P_{t|t-1} Z_t + h_t \end{cases}$$

f_t har tolkning som varians til ett-trinnsprediksjonen.

De rekursive residualene eller innovasjonene som kan uttrykkes ved

$$v_t = y_t - \hat{y}_{t|t-1} = y_t - Z_t' a_{t|t-1}$$

er ukorrelerte og har varians $\sigma^2 f_t$. Disse inneholder all den informasjon som den nye observasjonen, y_t har bidratt med og som ikke allerede ligger i tidligere observasjoner. Innovasjonen utnyttes til å oppdatere estimatene.

Vi skal nå bruke den enkle regresjonsmodellen fra eksempel 1 ovenfor for å vise Kalmanfilterformuleringen av en enkel tilstandsmodell:

Vi har da modellen:

$$(M) \quad y_t = x_t \beta + \zeta_t; \quad h_t=1, \quad Z_t=x_t$$

$$(T) \quad \beta = \beta; \quad T_t=I, \quad Q_t=0$$

La b_{t-1} være MK-estimatoren basert på H_{t-1} (historien opp til og med tidspunkt $t-1$), med kovariansmatrise $\sigma^2 P_{t-1}$.

Selve Kalmanfilteret kan da skrives:

$$\left. \begin{aligned} b_{t|t-1} &= T_t b_{t-1} = b_{t-1} \\ P_{t|t-1} &= P_{t-1} \end{aligned} \right\} \text{Prediksjonslikninger}$$

$$\left. \begin{aligned} b_t &= b_{t-1} + P_{t-1} x_t (y_t - x_t' b_{t-1}) / f_t \\ P_t &= P_{t-1} - P_{t-1} (x_t x_t') P_{t-1} / f_t \end{aligned} \right\} \text{Oppdateringslikninger}$$

$$f_t = x_t' P_{t-1} x_t + 1$$

Dersom en initierer filteret med startverdiene

$$b_p = (X_p' X_p)^{-1} X_p' Y_p$$

$$P_p = (X_p' X_p)^{-1}$$

hvor

$$Y_p = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \\ y_{p+1} \\ \vdots \\ y_T \end{bmatrix} ; \begin{bmatrix} x_1' \\ \vdots \\ x_p' \\ x_{p+1}' \\ \vdots \\ x_T' \end{bmatrix} = X_p$$

vil nettopp $\sigma^2 P_p$ være kovariansmatrisen til b_p .

Størrelsen b_T som framkommer ved å rulle filteret helt fram, er nettopp MK-estimatoren for β . Den har kovariansmatrise

$$\sigma^2 P_T = \sigma^2 (X_T' X_T)^{-1}$$

De rekursive residualene eller innovasjonene

$$v_t = y_t - x_t' b_{t-1} \quad \text{er uavhengige og } N(0, \sigma^2 f_t)$$

Dersom vi hadde startet med valgte verdier

b_0 og P_0

kunne vi oppfatte b_T som Bayesestimatorens for β når apriorifordelingen til $\beta \sim N(b_0, \sigma^2 P_0)$.

En annen tolkning er å oppfatte b_T som ridgeestimatoren til β , der β er trukket mot b_0 med relativ styrke P_0^{-1} .

4. Framskriving og glatting

Når en har data fram til $t = T$, kan det være aktuelt å predikere α_{t+s} hvor s er et bestemt antall perioder inn i framtida. Dette gjøres ved å utelate oppdateringslikningene fra filteret og bare bruke prediksjonslikningene.

$$\begin{aligned} a_{T+s}|T &= E(\alpha_{T+s}|H_t) \\ &= T_{T+s} \cdot a_{T+s-1}|T; \quad (a_T|T = a_T) \end{aligned}$$

$$\hat{y}_{T+s}|T = Z_{T+s}' a_{T+s}|T$$

$$s = 1, \dots, n$$

Når data foreligger fram til $t = T$, kan det også være av interesse å tilbakeskrive serien. Dvs. at også informasjon om serien etter beregningstidspunktet benyttes i estimeringen ved at det tas utgangspunkt i a_T og P_T , og beregner:

$$a_t|T = E(\alpha_t|H_t) \quad t < T$$

Dette gjøres rekursivt ved lineære operasjoner.

En slik rekursjon ved å bruke Kalmanfilteret kan være nyttig hvis det skal foretas en glatting av serien, f.eks. sesongjustering.

5. Estimering av ukjente parametre

Bortsett fra den enkle regresjonen har vi hittil bare sett på modeller med spesifiserte parametre i Z_t , T_t , h_t og Q_t . For gitte parametre fram- og tilbakeskriver filteret tilstandsvektoren som da består av latente variable. Når ukjente parametre inngår i tilstandsmodellen, må de estimeres. Dette kan illustreres med utgangspunkt i regresjonsmodellen med AR-feilledd fra eksempel 3 ovenfor:

$$y_t = x_t \beta + \zeta_t ; \quad \zeta_t = \varphi(L)\zeta_{t-1} + \varepsilon_t$$

Tilstandsmodellen kan skrives:

$$\alpha_t = \begin{bmatrix} \beta \\ \zeta_t \\ \zeta_{t-1} \\ \vdots \\ \zeta_{t-p-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \varphi_1 & \varphi_2 & \dots & 0 & \dots & \varphi_{t-p} & \\ 0 & 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta \\ \zeta_{t-1} \\ \vdots \\ \zeta_{t-p} \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \varepsilon_t$$

Her kan vi indeksere filteret med parametervektoren φ . For gitt verdi av φ kan β lett estimeres, for eksempel ved hjelp av filteret.

For hver gitt φ kan vi kjøre filteret og få fram de rekursive residualene:

$$V_t(\varphi) = y_t - \hat{y}_{t|t-1}(\varphi)$$

med

$$\text{var}V_t(\varphi) = \sigma^2 f_t(\varphi).$$

V_t er uavhengig og $N(0, \sigma^2 f_t)$ så sant φ er den sanne parameter. Det viser seg at dersom de k første observasjonene oppfattes som gitt (initialverdien for filteret er α_k og P_k som er dannet ut fra disse observasjonene), er likelihood-funksjonen:

$$\begin{aligned} \ln L(\varphi | y_{k+1}, \dots, y_T) &= \ln \prod_{t=k+1}^T \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2 f_t}} e^{-V_t^2 / 2\sigma^2 f_t} \\ &= -(T-k) \frac{1}{2} \ln \sigma^2 - \frac{1}{2} \sum \ln f_t(\varphi) - \frac{1}{2} \frac{1}{\sigma^2} \sum_{t=k+1}^T \frac{V_t(\varphi)^2}{f_t} \end{aligned}$$

hvor $\sigma^2 f_t$ er variansen til V_t . f_t er gitt ved $f_t = Z_t' P_t |_{t-1} Z_t + h_t$

Følgelig er; se f.eks. Harvey (1984a):

$$\hat{\sigma}^2 = \frac{1}{T-k} \sum_{k+1}^T \frac{V_t^2(\hat{\varphi})}{f_t(\hat{\varphi})^2}$$

Den konsentrerte likelihoodfunksjonen blir dermed:

$$\ln L_c(\varphi) = -\frac{1}{2} [\sum \ln f_t(\varphi) + (T-k) \ln \hat{\sigma}^2(\varphi)]$$

ML-estimatoren for φ finnes ved å maksimere $\ln L_c$ m.h.p. φ . Likelihoodfunksjonen regnes deretter ut ved hjelp av Kalmanfilteret. Denne prosedyren gjentas inntil verdien av likelihoodfunksjonen ikke lenger endres særlig fra én iterasjon til den neste.

Sekvensen

filter (med gitt φ) X maksimering av $L_c(\varphi)$

gir altså ML-estimatorer for parametrene i modellen. Dette er, ifølge Harvey og Phillips (1979), en effektiv måte å estimere ukjente parametre som ikke lar seg konsentrere ut til tilstandsvektoren. En utvidelse til å se på en ARMA-modell fører ikke inn noe nytt bortsett fra at det må foretas en simultan estimering av AR- og MA-parametrene. Denne estimeringen innebærer en ikke-lineær maksimering m.h.p. Q_t . Dette krever tilgang til gode optimeringsalgoritmer (jfr. avsn. 6).

I følge Harvey (1984b) er en av de bekvemme egenskapene ved de rekursive residualene fra Kalmanfiltreringen at de kan brukes for å uttrykke likelihoodfunksjonen. Dette kan også utnyttes ved testing av modellspesifikasjonen der f.eks. Lagrangemultiplikatortester kan brukes.

Denne måten å tilpasse tilstandsmodeller med ukjente parametre er

generell, og Harvey anbefaler den for tilpassing av hans strukturmodell med spesifisert trend og sesong. Dette skal vi se nærmere på i det følgende.

6. Estimering av de relative variansene

For å se hvordan maksimum likelihood estimering av de relative variansene i strukturmodellen (Eksempel 4) virker, har vi gjennomført et meget begrenset Monte Carlo eksperiment. En serie på i alt 1 200 observasjoner av y ble generert ut fra strukturmodellen ved å trekke uavhengige, normalfordelte restledd ved hjelp av rutinen SNORM i TROLL (se TROLL 1981). De variansparametrene vi benyttet er gitt i første kolonne i tabell 1. I tabellen er videre maksimum likelihood estimatet for σ^2 og de relative variansene angitt etter at $T=30, 150, 400$ og 1200 observasjoner er brukt i estimeringen. Den numeriske optimeringsalgoritmen E04 JAF i NAG-biblioteket ble benyttet (se NAG (1981)). Vi har ingen garanti for at maksimeringen av likelihoodfunksjonen førte fram til et globalt maksimum (se Loe (1987)).

Tabell 1

Estimat:	"Sann" verdi	Antall observasjoner:			
		30	150	400	1200
σ_{μ}^2/σ^2	1	0.000	0.0000	0.000	0.094
$\sigma_{\beta}^2/\sigma^2$	0.25	0.015	0.055	0.102	0.127
$\sigma_{\gamma}^2/\sigma^2$	3	1.888	1.273	1.339	1.578
σ^2	100	1449.04	590.52	331.28	223.35
$\text{LnL}_c(\tilde{\varphi})$	-	- 138.56	- 620.68	- 1566.44	- 4556.90
$\text{LnL}_c(\varphi)$	-	- 143.06	- 628.08	- 1573.94	- 4565.73

Her er $\text{LnL}_c(\tilde{\varphi})$ funksjonsverdien til den konsentrerte likelihood-funksjonen for de estimerte parametrene, mens $\text{LnL}_c(\varphi)$ er funksjonsverdien

for de "sanne" parameterverdiene.

Som en ser, virker metoden meget dårlig i den forstand at σ^2 blir grovt overestimert og de relative variansene grovt underestimert. En viss tendens til konvergens mot mer akseptable verdier når T øker, er heldigvis tilstede, men fra et praktisk synspunkt er denne tendensen meget svak. At den observerte log likelihood-verdien er litt større enn den teoretiske er normalt, faktisk bør forskjellen være av størrelsesorden 1/2 ganger antall estimerte parametre - og det stemmer grovt sett. Det er følgelig liten grunn til å tvile på at estimatene er maksimum likelihood estimater (i alle fall tilnærmet).

For å se om maksimum likelihood metoden virker like dårlig i et enklere problem, sløyfet vi vekstkomponenten β og sesongkomponenten γ og genererte på liknende måte som over en serie på 1 200 observasjoner. De sanne verdier og estimatene etter 30, 150, 400 og 1200 observasjoner er angitt i tabell 2.

Tabell 2

Estimat:	"Sann" verdi	Antall observasjoner:			
		30	150	400	1200
σ_Y^2 / σ^2	1	(ingen konvergens)	3.68	1.81	1.19
σ^2	100	-----	163.34	139.58	116.19
$\text{Ln}L_c(\tilde{\varphi})$	-	-----	- 510.03	- 1239.83	- 3479.40
$\text{Ln}L_c(\varphi)$	-	-----	- 511.23	- 1241.12	- 3479.80

Også i den enkle strukturmodellen virker maksimum likelihood metoden forbausende dårlig.

En skal naturligvis vokte seg for å trekke vidtgående konklusjoner på grunnlag av et så avgrenset eksperiment som det vi har rapportert her. Men på bakgrunn av litteraturen må det kunne hevdes at estimatene fra maksimum likelihood tilpassing av strukturmodellen til økonomiske (kvar-tals)serier av vanlig lengde, er meget upålitelige. Dette kommer av at likelihood-funksjonen er flat i nærheten av optimum og at den i mange tilfeller er meget stor på randen av parameterrommet ($\sigma=0$, $\sigma=\infty$). Slike parametre er rett og slett vanskelige å estimere.

Den modellen som blir tilpasset ved hjelp av maksimum likelihood

metoden, vil imidlertid ikke nødvendigvis virke dårlig som prediksjons-instrument. Loe (1987) fant faktisk at prediksjonene fra den tilpassede strukturmodellen konkurrerte godt med de som er basert på framskrivninger av serier som er sesongjustert med X11-metoden for et utvalg vanskelig predikerbare handelsstatistikk-serier.

Når formålet er sesongjustering og trendestimering, er det imidlertid ikke opplagt at maksimum likelihood metoden er et godt kriterium for modelltilpassing. En vesentlig egenskap ved en trend er at den varierer sakte og at den er glatt. Når variansene σ_{μ}^2 og σ_{β}^2 er store, vil trendestimateret fra strukturmodellen bli alt annet enn glatt. For dekomponeringsformål bør altså tilpassingen foretas under den bibetingelse at disse variansene er små. Denne problemstilling likner ikke så lite på den en har ved maksimum likelihood estimering av en ikke-parametrisk sannsynlighetstetthet. Der har en det paradoksale at ingen tetthet maksimerer likelihoodfunksjonen, men at den blir maksimert for en diskret likefordeling over observasjonene. For å tvinge estimatet til å bli en glatt tetthet opererer man derfor med såkalt straffet likelihood maksimering. Det vil si, en maksimerer:

$$\ln L - G,$$

der $\ln L$ er log likelihood og G er en indeks for glatthet: jo glattere desto mindre G , se Silverman (1986).

I vårt problem kunne en tenke seg å måle glattheten av trenden μ_t relativt til utslagene i den eratiske komponenten ε_t . Et mulig mål på absolutt glatthet av serien μ_t , $t=1, \dots, T$ er:

$$\frac{1}{T-2} \sum_{t=3}^T (\Delta^2 \mu_t)^2 \approx E[\Delta^2 \mu_t]^2,$$

der;

$$\Delta^2 \mu_t = \mu_t - 2\mu_{t-1} + \mu_{t-2}$$

er annenordens differensen. Når den eratiske seriens glatthet måles på samme måte blir:

$$\delta = \frac{1}{6\sigma^2} (2\sigma_{\mu}^2 + \sigma_{\beta}^2)$$

et mål på trendens relative glatthet. Et straffet likelihood kriterium vil følgelig være

$$\text{maksimer } \left[\frac{1}{T} \ln L - g \cdot \frac{1}{6\sigma^2} (2\sigma_\mu^2 + \sigma_\beta^2) \right],$$

der $g \geq 0$ er et tall som bestemmes av brukeren. Dette tallet angir hvor mange enheter tilpasning, $\frac{1}{T} \ln L$, en enhet relativ glatthet er verdt. Merk at glatthetsmålet δ bare avhenger av de relative variansene, og σ kan følgelig også konsentreres ut av den straffede likelihooden.

En annen metode er på forhånd å bestemme den relative glattheten til trendestimatet. Vi får dermed en parameter mindre å bestemme i det for eksempel σ_μ^2/σ^2 regnes som fri parameter,

$$0 \leq \frac{\sigma_\mu^2}{\sigma^2} \leq 3\delta,$$

og

$$\frac{\sigma_\beta^2}{\sigma^2} = 6\delta - 2\frac{\sigma_\mu^2}{\sigma^2}.$$

Å fastsette δ kunne være en oppgave for Modellutvalget. Hvor glatt kurven skal være for å kalles en trend, er et spørsmål som godt kan gjøres til gjenstand for byråkratisk eller politisk beslutning.

7. Filteret brukt på en feil-korreksjonsmodell

Et annet eksempel på hvordan Kalmanfilterteknikk kan brukes for å gi ny informasjon om en økonomisk sammenheng får vi ved å skrive en økonomisk relasjon som en tilstandsmoell, og vi kan dermed undersøke den dynamiske strukturen i interessante komponenter i modellen. Vi har her valgt å se på en feilkorreksjonsmodell slik den er utformet hos Biørn og Olsen (1986).

Modellen er for skalare y_t og x_t (kvartalsserier):

$$y_t = (1-\gamma)\rho(L)y_t + \alpha x_t + (\gamma\beta - \alpha)\rho(L)x_t + u_t$$

der

$$\varrho(L) = 1 - (1 - \varrho_1 L)(1 - \varrho_4 L^4) = \varrho_1 L + \varrho_4 L^4 - \varrho_1 \varrho_4 L^5$$

L er lag-operator; $LX_t = X_{t-1}$.

Når $\varrho(1) = \varrho_1 + \varrho_4 - \varrho_1 \varrho_4 \neq 1$, vil den langsiktige likevektsløsningen være:

$$y = Ax + B$$

$$A = \beta + \frac{(\alpha - \beta)(1 - \varrho_1(1))}{1 - (1 - \gamma)\varrho_1(1)}$$

$$= \alpha \quad \text{hvis } \varrho_1(1) = 0 \\ \text{eller } \gamma = 0$$

β kan gis tolkning som en langsiktig og α som en kortsiktig virkningskoeffisient.

Ved å skrive modellen

$$y_t = \alpha x_t + \varrho(L)[(y_t - \alpha x_t) - \gamma(y_t - \beta x_t)] + u_t$$

ser en hvordan banen til y korrigeres i henhold til "kortsiktige" feil: $y_t - \alpha x_t$ og "langsiktige" feil: $y_t - \beta x_t$.

For å få skrevet denne modellen på tilstandsform innfører vi:

$$v_t = y_t - \beta x_t$$

$$w_t = y_t - \alpha x_t$$

Transisjonsformelen kan dermed skrives:

$$(T) \quad v_t = (\alpha - \beta)x_t + w_t$$

$$w_t = \varrho(L)(w_t - \gamma v_t) + u_t$$

Tilsvarende kan målelikningen skrives:

$$(M) \quad y_t = \frac{\beta}{\beta-\alpha} w_t - \frac{\alpha}{\beta-\alpha} v_t \quad (y_t = y_t)$$

Transisjonslikningene (T) kan nå skrives på matriseform med en tilstandsvariabel som er 11-dimensjonal: $(w_t, w_{t-1}, \dots, w_{t-4}, v_t, \dots, v_{t-4}, \alpha-\beta)$, og med en transisjonsmatrise som inneholder X_t samt parametrene ρ_1, ρ_4 og produkter av disse.

For estimerte parametre $\rho_1, \rho_4, \alpha, \beta$ og γ kan filteret kjøres, og vi kan se på de "kortsiktige" og de "langsiktigs" feilene v_t og w_t .

Vi kunne også legge inn en "myk" sesongkomponent (se f.eks. avsn. 2):

$$y_t = \frac{\beta}{\beta-\alpha} w_t - \frac{\alpha}{\beta-\alpha} v_t + \gamma_t$$

$$\gamma_t = -\gamma_{t-1} - \gamma_{t-2} - \gamma_{t-3} + \eta_t ; \quad (\text{kvartalseffekt})$$

Fordelen med denne modellspesifikasjonen er at for estimerte (gitte) koeffisienter kan filteret brukes til prediksjoner etc. samtidig som kjennskap til utviklingen i og størrelsen på de latente komponentene kan være nyttig.

8. Prediksjon av regresjonsresidualer

Hvis det er noen form for samvariasjon i restleddene, vil vi forbedre prediksjonsresultatene ved å ta hensyn til dette. Problemet er hvordan vi skal kunne modellere denne restleddsprosessen.

La oss anta at vi har en modell:

$$y_t = x_t' \beta + \varepsilon_t ; \quad t=1, \dots, T$$

hvor

$$E\varepsilon_t = 0.$$

I utgangspunktet forutsetter vi at restleddene ε_t er ukorrelerte med samme varians, slik at parametrene i modellen kan estimeres ved å bruke OLS.

Når Y er en vektor bestående av y_1, \dots, y_T og X er designmatrisen, kan vi skrive:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$E = Y - X\hat{\beta} = (I-H)Y$$

E er en vektor med residualer, og:

$$H = X(X'X)^{-1}X'$$

er projeksjonsmatrisen på rommet spent ut av kolonnene i X :

Vi antar at H har full rang, $T-k$, der k er dimensjonen på X og β .

Projeksjonen $I-H$ er ortogonal på kolonnene i X , dvs. at:

$$E = (I-H)(X\beta + \varepsilon) = (I-H)\varepsilon$$

Vi antar at ε har en kovariansmatrise Σ_ε som kan modelleres ved en enkel ARMA-modell. Kovariansmatrisen til E vil være:

$$\Sigma_E = (I-H)\Sigma_\varepsilon(I-H)$$

Dersom vi bruker E som substitutt for ε , og tilpasser en ARMA-modell, vil vi få estimert en annen ARMA-prosess enn den som beskriver den egentlige restleddsprosessen ε . Når k/T ikke er nær null, vil $I-H$ kunne være ganske forskjellig fra I , og dermed også Σ_E ganske forskjellig fra Σ_ε . Vi vil følgelig kunne få en betydelig feil i ARMA-modelleringen. Hvis f.eks. $\Sigma_\varepsilon = \sigma^2 I$, slik at residualene egentlig ikke inneholder noen prediksjonsinformasjon, vil $\Sigma_E = \sigma^2(I-H)$ likevel kunne gi opphav til en ARMA-modell. Spørsmålet er om vi kan utnytte informasjon om regresjonsresidualenes korrelasjonsstruktur i framskrivingen av de egentlige feilleddene.

Hvis vi f.eks. estimerer parametrene i modellen ved å bruke Kalman-filteret, kan vi få tilgang til de rekursive residualene fra estimeringen,

$$s_t = y_t - y_{t|t-1} = y_t - x_t' \hat{\beta}_{t|t-1}; \quad t > k$$

De sekvensielle residualene er ikke definert for $t \leq k$. Vi har valgt å

definere dem lik null. Dersom vi også definerer:

$$x_t'(X_{t-1}' X_{t-1})^{-1} X_{t-1}' = [0 \dots 0 \underset{\substack{\downarrow \\ \text{element nr. } t}}{1} 0 \dots 0], \quad t \leq k,$$

$$X_t = [x_1 \dots x_t \ 0 \dots 0]'$$

vil vektoren av rekursive residualer kunne skrives

$$S = \begin{bmatrix} I - \begin{bmatrix} x_1'(X_0' X_0)^{-1} X_0' \\ \vdots \\ x_T'(X_{T-1}' X_{T-1})^{-1} X_{T-1}' \end{bmatrix} \end{bmatrix} Y \\ = VY$$

Det kan vises at

$$VX = 0$$

og

$$VV' = \begin{bmatrix} 0 & \vdots & 0 \\ \dots & \vdots & \dots \\ 0 & \vdots & D^2 \end{bmatrix}$$

hvor

$$D = \text{diag} (1 + x_t'(X_{t-1}' X_{t-1})^{-1} x_t')$$

Linjene i V er altså ortogonale, men ikke av en enhets lengde. De standardiserte rekursive residualene, R , er gitt ved:

$$R = D^{-1}S = D^{-1}V\varepsilon \quad (\text{siden } VX=0)$$

Kovariansmatrisen til disse standardiserte residualene blir:

$$\Sigma_R = D^{-1}V\Sigma_\varepsilon V' D^{-1}.$$

Hvis $\Sigma_\varepsilon = \sigma^2 I$, er også $\Sigma_R = \sigma^2 I$.

Dersom vi nå tilpasser en ARMA (p,q)-modell til de standardiserte rekursive residualene r_{k+1}, \dots, r_T , vil vi kunne framskrive den f.eks.

v.h.j.a. prediksjonsreglene til Box-Jenkins. Vi kan dermed få en forlenget serie:

$$r_{k+1}, \dots, r_T, \hat{r}_{T+1}, \dots, \hat{r}_{T+\tau}$$

La oss kalle den vektoren som inneholder denne serien, \hat{R} . Problemet er nå å transformere de standardiserte prediksjonene til prediksjoner på de egentlige residualene, $\hat{\varepsilon}_{T+1}, \dots, \hat{\varepsilon}_{T+\tau}$.

Vi har nå at:

$$\hat{R} = \hat{D}^{-1} \hat{V} \hat{\varepsilon}$$

der ε -serien er forlenget med prediksjonene av de egentlige residualene. \hat{V} er en ortogonal $(t+\tau) \times (T+\tau)$ -matrise som kan lages ved å bruke en Cholesky dekomposisjon. Følgelig har vi:

$$\hat{V}' \hat{D} \hat{R} = \hat{V}' \hat{V} \hat{\varepsilon} = \hat{D} \hat{\varepsilon}$$

siden \hat{V} er ortogonal. $\hat{D} = \hat{V}' \hat{V}$ er diagonal. Den ønskede transformeringen kan foretas ved

$$\hat{\varepsilon} = \hat{D}^{-1} \hat{V}' \hat{D} \hat{R}$$

(Ettersom \hat{V} er øvre triangulær, vil de k første komponentene i \hat{R} ikke komme inn ved beregningen av $\hat{\varepsilon}_t$, $t > T$.) Beregningen av \hat{V} , \hat{R} og \hat{D} vil kunne forenkles ved å benytte rekursjonsformlene for OLS-estimering, uttrykt som et spesialtilfelle av et Kalmanfilter (se Harvey (1981a s. 192)).

9. Linearisering av feilkorreksjonsmodellen

Som det går fram av avsnitt 7, er feilkorreksjonsmodellen ikke-lineær i parametrene. Den helspesifiserte Kalman-filtertechnikken kan imidlertid bare brukes for estimering av regresjonsparametre som inngår lineært i modellen. Ved å linearisere modellen rundt et punkt, kan vi imidlertid få en modelltilnærming der også parametre eller latente variable som inngår ikke-lineært i den opprinnelige modellen, kan estimeres v.h.j.a. Kalman-filtertechnik.

I dette tilfellet vil imidlertid de optimalitetsegenskaper som estimatorene i den lineære modellen har, bare være tilnæringer. Tilsvarende vil en likelihood-funksjon som er beregnet via prediksjonsfeilene, bare være en tilnærming (se f.eks. Harvey og Pierse (1984)).

$$Y_t = (1-\gamma)(L^3 + e_1(1-L^4))Y_{t-1} + \alpha X_t \\ + (\gamma\beta - \alpha)(L^3 + e_1(1-L^4))X_{t-1} + \sum_{j=1}^3 \delta_j D_j + \text{restledd}$$

der

$$D_j = DKV_j - DKV_4$$

er kvartalseffekten relativt til 4. kvartal.

Denne modellen skal lineariseres i parametrene. Ved direkte derivasjon får vi:

$$\frac{dY_t}{de_1} = (1-\gamma)(1-L^4)Y_{t-1} + (\gamma\beta - \alpha)(1-L^4)X_{t-1}$$

$$\frac{dY_t}{d\alpha} = X_t - (L^3 + e_1(1-L^4))X_{t-1}$$

$$\frac{dY_t}{d\beta} = \gamma(L^3 + e_1(1-L^4))X_{t-1}$$

$$\frac{dY_t}{d\gamma} = (L^3 + e_1(1-L^4))(\beta X_{t-1} - Y_{t-1})$$

$$\frac{dY_t}{d\delta_j} = D_j$$

For en gitt parametervektor $\theta^0 = (e_1^0, \alpha^0, \beta^0, \gamma^0, \delta_1^0, \delta_2^0, \delta_3^0)$, definerer vi:

$$Y_t^0 = (1-\gamma^0)(L^3 + e_1^0(1-L^4))Y_{t-1} + \alpha^0 X_t \\ + (\gamma^0\beta^0 - \alpha^0)(L^3 + e_1^0(1-L^4))X_{t-1} + \sum_1^3 \delta_j^0 D_j,$$

og "forklaringsvariable":

$$\frac{dY_t^0}{de_1} = (1-\gamma_0 (1-L^4)Y_{t-1} + (\gamma_0^0 \beta^0 - \alpha^0)(1-L^4)X_{t-1})$$

$$\frac{dY_t^0}{d\alpha} = X_t - (L^3 + e_1^0 (1-L^4))X_{t-1}$$

$$\frac{dY_t^0}{d\beta} = \gamma^0 (L^3 + e_1^0 (1-L^4))X_{t-1}$$

osv.

Den lineariserte modellen kan nå skrives:

$$Y_t - Y_t^0 = de_1 \cdot \frac{dY_t^0}{de_1} + d\alpha \cdot \frac{dY_t^0}{d\alpha} + d\beta \cdot \frac{dY_t^0}{d\beta} + d\gamma \cdot \frac{dY_t^0}{d\gamma} \\ + \sum_{j=1}^3 d\delta_j \frac{dY_t^0}{d\delta_j} + \text{restledd}$$

Koeffisientene de_1 , $d\alpha$, $d\beta$ osv. kan nå estimeres ved lineær regresjon, for eks. ved å bruke Kalman-filteret.

Estimering av feilkorreksjonsmodellen kan nå gjøres ved å iterere følgende sykel:

For startparameteren θ^0 , beregn variablene $\frac{dY_t^0}{de_1}$ osv.

Estimer $de_1, \dots, d\delta_3$.

Sett $\theta^1 = \theta^0 + d\theta$.

Denne sykkelen kan itereres enten et bestemt antall ganger eller til $d\theta$ får en verdi som er lavere enn et valgt konvergensnivå. Denne teknikken likner på det som gjøres i vanlig ikke-lineær regresjon.

10. Flerlikningsmodeller

Til nå har vi bare sett på én-likningsmodeller. En generalisering av Kalmanfilteret til å omfatte flerlikningsmodeller innebærer ikke noe prinsipielt nytt.

Målelikningen kan være en vektorlikning

$$y_t = Z_t' \alpha_t + \zeta_t \quad \zeta_t \sim N(0, \sigma R_t^2)$$

Her kan R_t være ikke-diagonal. Z_t er nå en matrise og y_t en vektor. Kalmanfiltreringen gir således en vektor med sekvensielle residualer av samme dimensjon som y_t . Bruken av flerlikningssystemer begrenses imidlertid av at det raskt blir vanskelig å skrive ut modellen i tilstandsform.

11. Empiriske eksempler

11.1. Feilkorreksjonsmodeller

Som et eksempel på bruk av Kalman-filteret har vi formulert den feilkorreksjonsmodellen som ble gjennomgått ovenfor, innenfor rammen av et Kalman-filter, og brukt denne for estimering og prediksjon. Vi har her sett på sammenhengen mellom totalt privat konsum og realdisponibel inntekt for husholdningene. Det er brukt tall fra det kvartalsvise nasjonalregnskapet, kjedet til 1984-priser, for perioden 1969.1-1983.4. Som det framgår av de faktiske tallseriene for konsum og inntekt som er vist i fig. 1, er det klare sesongsvingninger. Dette er det tatt hensyn til på en enkel måte ved å inkludere en additiv sesongkomponent. Det kan nevnes at i Byråets makroøkonometriske modell, KVARTS, er det spesifisert multiplikative sesongkomponenter i makrokonsumfunksjonen.

Feilkorreksjonsmodellen får dermed følgende utforming:

$$(1-\rho(L)) y_t = \alpha(1-\rho(L)) x_t + \gamma\rho(L)(\beta x_t - y_t) + \mu_t + u_t$$

$$\mu_t = \sum_{i=1}^3 \mu_{t-i} + \varepsilon_t$$

der

$$\varrho(L) = \varrho_1 L(1-L^4) + \varrho_4 L^4$$

y_t = privat konsum, volum

x_t = realdisponibel inntekt

μ_t = sesongkomponent

$\alpha, \beta, \gamma, \varrho_1, \varrho_4$ - konstante parametre.

u_t, ε_t - restledd

Denne modellen er ikke-lineær i parametrene. Som vist i kap. 7 kan imidlertid modellen lineariseres, slik at hvis ønskelig kan Kalman-filtertechnik brukes for å estimere koeffisientene også i denne modellen. I dette eksemplet har vi imidlertid valgt å oppfatte koeffisientene som faste parametre, og formulere modellen slik at den kan brukes til å belyse andre sammenhenger mellom konsum og inntekt.

Følgende nye variable er definert:

$$v_t = y_t - \beta x_t$$

$$w_t = y_t - \alpha x_t$$

I den opprinnelige modellen har α og β tolkning som h.h.v. kort-siktig og langsiktig konsumtilbøyelighet. v_t vil derfor kunne tolkes som avviket mellom faktisk konsum og det konsumet som svarer til langtidstilpasningen, mens w_t kan tolkes som avviket mellom faktisk konsum og korttidstilpasningen. v_t og w_t er latente variable som det kan være interessant å tallfeste, kanskje særlig for å kunne undersøke om det har vært noen systematiske endringer i disse komponentene over tid. Modellen er derfor reformulert innenfor rammen av et Kalman-filter:

$$\text{Målelikning: } y_t = \frac{\alpha}{\beta - \alpha} v_t + \frac{\beta}{\beta - \alpha} w_t + u_t$$

$$\text{Transisjonslikning: (i) } v_t = (\alpha - \beta)x_t + w_t$$

$$\text{(ii) } w_t = \varrho(L)(w_t + \gamma v_t)$$

$$\text{(iii) } \mu_t^* = - \sum_{i=1}^3 \mu_{t-i}^* + \varepsilon_t^*$$

x_t er behandlet som en rent eksogen variabel. Estimerer for koeffisientene er hentet fra tidligere estimeringer av en tilsvarende modell som er

foretatt ved en kombinert grid-search-NLS-prosedyre i TROLL (se Biørn og Olsen (1986)). Denne estimeringen ga følgende resultater:

$$e_1 = 0.4$$

$$e_4 = 1$$

$$\gamma = 0.6$$

$$\alpha = 0.62$$

$$\beta = 0.96$$

Den modellen som ble brukt hos Biørn og Olsen (1986) hadde en litt annen formulering av sesongbevegelsene enn i denne Kalman-filter-versjonen slik at det er mulig at dette kan påvirke resultatene noe. I dette eksemplet har vi benyttet enkle additive sesongdymmies som definert i (iii). Innføring av et restledd impliserer at sesongkomponentene ikke nødvendigvis summerer seg opp eksakt til null over året, og kan gis en tolkning av å være stokastiske sesongvariasjoner.

Alle beregninger er foretatt i TROLL. Et problem som må løses for å kunne gjøre en praktisk bruk av Kalman-filteret er tallfesting av elementene i Q-matrisen. Som nevnt i kap. 6 vil det beste estimatet for elementene i Q framkomme ved en maksimering av likelihood-funksjonen til y_1, \dots, y_t . Det krever imidlertid tilgang til optimeringsalgoritmer som ikke finnes i TROLL. På det tidspunktet disse beregningene ble gjort, måtte vi derfor velge forsøksvise verdier og prøve om disse ga rimelige resulater. Dette er tidkrevende, og det har delvis vært vanskelig å nå fram. Det trengs også startverdier for elementene i tilstandsvektoren α_t og kovariansmatrisen P. Elementene i tilstandsvektoren er gitt anslagsvise verdier, mens elementene i kovariansmatrisen er gitt høye verdier, slik at startverdiene "glemmes" forholdsvis raskt.

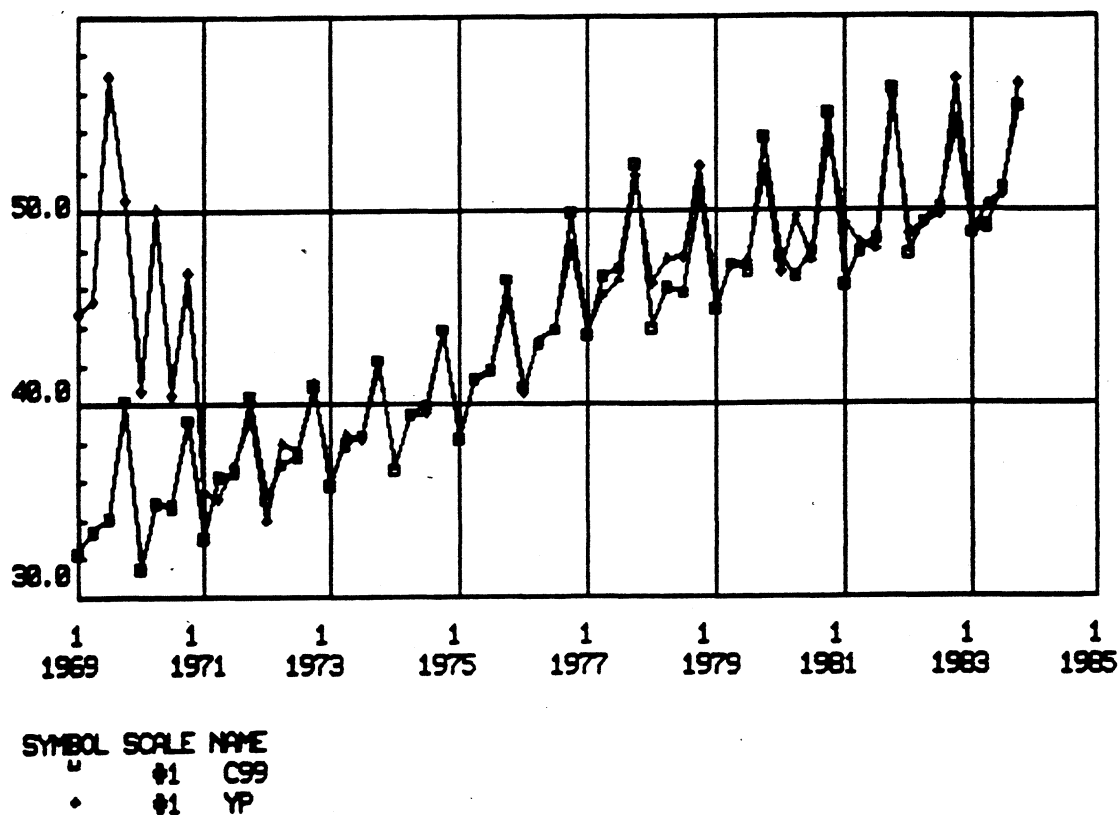
Resultater fra beregningene er gitt i tabell 3 og figur 1. Det er beregnet betingede prediksjoner en periode framover for v_t og w_t for perioden 1969.1 til 1983.4: dvs. predikert verdi på tidspunkt t, gitt forhistorien. Det er også gitt tall for den betingede prediksjonsfeilen på hvert tidspunkt i prosent av faktisk nivå for konsumet i perioden. Resultatet kan kanskje synes noe skuffende når en tar hensyn til at dette er ex post-prediksjoner av en serie med en forholdsvis regelmessig bevegelse. Det er imidlertid grunn til å understreke betydningen av de tekniske problemene som er nevnt ovenfor og den enkle framgangsmåten som er brukt. Erfaring tilsier at betydelig forbedring av resultatene kan

forventes når tallfesting av de relative variansene er mer tilfredsstillende (se Loe (1987)). Det er også verdt å merke seg at filteret trenger bare et par år med observasjoner før det er rimelig godt på "sporet".

Tabell 3. Betinget prediksjonsfeil for privat konsum, prosent.

1969.1	-37,7	-35,3	-66,8	-26,0
1970.1	-29,2	-43,2	-16,2	-19,4
1971.1	-7,2	3,3	-0,8	2,2
1972.1	3,1	-3,0	-0,5	-0,2
1973.1	0,4	-1,2	0,5	0,6
1974.1	-0,7	0,3	0,8	0,1
1975.1	0,3	0,2	0,3	2,0
1976.1	0,8	-0,4	-0,1	3,4
1977.1	-0,2	2,1	1,2	1,3
1978.1	-5,0	-3,2	-3,9	-2,7
1979.1	-0,5	0,2	-1,0	3,1
1980.1	1,8	-6,5	1,0	2,3
1981.1	-6,5	-0,9	1,3	1,0
1982.1	-2,0	0,3	0,8	-3,9
1983.1	0,2	-2,1	0,7	-1,9

Figur 1: Totalt privat konsum, faktisk (C99) og predikert (YP)



11.2. Sesongjustering

I avsnitt 4 er det vist hvordan Kalman-filteeret kan brukes for å foreta en tilbakeregning av en serie. Den ordinære framskrivningen av Kalman-filteeret gir på hvert tidspunkt $t = 1, \dots, T$ minimum minste kvadraters estimatorer for elementene i tilstandsvektoren, α_t , gitt all informasjon opp til det aktuelle tidspunktet, t . Den eneste estimatoren som benytter all tilgjengelig informasjon, er estimatene fra den aller siste sampeleperioden T . Ved å foreta en rekursjon av filteeret med T som startpunkt, oppnås på hvert tidspunkt bakover minimum minste kvadraters estimatorer, gitt all tilgjengelig informasjon.

En slik bruk av Kalman-filteertechnik vil særlig kunne gi gevinst når en serie skal glattes i én eller annen forstand; dvs. at vi vil trekke ut enkelte komponenter i serien, f.eks. de som skyldes irregulære bevegelser og/eller systematiske sesongsvingninger. Det finnes mange metoder for å foreta en slik filtrering av serien. Felles for alle er imidlertid at resultatene forbedres ved å kunne utnytte informasjon om

hvordan utviklingen i serien fortsetter etter det tidspunktet hvor vi foretar glattingen. Et slikt to-sidig filter inngår f.eks. i X11-metoden¹⁾ som brukes for sesongjustering bl.a. i Byrået.

Som et eksempel på hvordan Kalman-filtertechnik kan brukes for sesongjusteringsformål, har vi benyttet Harvey's enkle strukturmodell. I denne inngår eksplisitt spesifiserte sesongkomponenter. Som utgangspunkt foretas en ordinær framskriving av Kalman-filteret slik det er beskrevet i avsnitt 4. Fra denne beregningen oppbevares estimatene på de oppdaterte elementene i tilstandsvektoren, a_T , kovariansmatrisen, P_T , og den betingete kovariansmatrisen, $P_{t|T}$ på hvert tidspunkt $t=1, \dots, T$. Når filteret er ført så langt fram som det finnes data, foretas det en ny estimering av komponentene på hvert tidspunkt, gitt historien helt fram til T . Denne beregningen starter i T med utgangspunkt i a_T og P_T , og tilbakeføres suksessivt til starttidspunktet. Dermed tas det i sesongjusteringen ikke bare hensyn til hvordan sesongbevegelsene har vært opp til tidspunkt t , men også hvordan de utvikler seg videre til tidspunkt T .

Hvis $a_{t|T}$ og $P_{t|T}$ betegner de tilbakeberegnete estimatorene for tilstandsvektoren og kovariansmatrisen, kan rekursjonsformlene skrives:

$$a_{t|T} = a_t + P_t^*(a_{t+1|T} - T_{t+1}a_t)$$

$$P_{t|T} = P_t + P_t^*(P_{t+1|T} - P_{t+1|t})P_t^{*'}$$

der

$$P_t^* = P_t T_{t+1}' P_{t+1|T}^{-1}$$

$$t = T-1, \dots, 1$$

og

$$a_{T|T} = a_T, \quad P_{T|T} = P_T$$

1) X11-metoden ble utarbeidet ved U.S. Bureau of the Census i slutten av 1960-årene. En omtale av modellen og litteraturhenvisninger finnes f.eks. hos Jensen et al. (1985).

Den tilbakeregnete estimatoren, $a_{t|T}$, er optimal, gitt all informasjon inklusive den siste observasjonen, y_T . Dvs. at for alle tidspunkt unntatt det siste kan vi bruke informasjon om den videre utvikling i serien for å beregne sesongjusterte tall;

$$S_{t|T} = Y_t - \gamma_{t|T} ; t = T-1, \dots, 1.$$

En "god" sesongjusteringsmetode bør oppfylle to krav:

- i) Den skal rense ut mest mulig av systematiske sesongutslag i en tidsserie.
- ii) Den skal gi minst mulig revisjoner av tidligere sesongjusterte tall etter hvert som det blir tilgjengelig informasjon om den videre utviklingen i serien.

Dette siste kravet er bekvemt ut fra praktiske formål, og bygger i virkeligheten på en forestilling om at trend- og sesongmønsteret i en serie endres lite over tid. Slik den såkalte strukturmodellen er formulert inneholder den lokale trend- og sesongfaktorer som avhengig av verdien på variansene til de ulike restleddene i prosessen, vil kunne endres mye eller lite. Det trenger derfor ikke være noen svakhet ved sesongjusteringsmetoden at det forekommer store revisjoner.

Vi har sesongjustert de 6 seriene som er listet opp i tabell 4 med den additive varianten av X11-metoden og Harvey's strukturmodell og sammenliknet resultatet av disse to prosedyrene.

For å studere nivået på revisjonene foretok vi først sesongjustering fra 1978.2 som er det tidligste tidspunktet Kalman-filtreringen kan startes, og til og med 1984.4. Deretter gjentok vi sesongjusteringen ved å forlenge perioden suksessivt med ett kvartal om gangen fram til og med 1985.3. For hvert kvartal ble revisjonen i den sesongjusterte serien, sammenliknet med foregående sesongjustering av den samme observasjonen, beregnet. Resultatet av disse beregningene er gjengitt i tabell 4.

Tabell 4.

Serie	Metode	Revisjon nr. ¹⁾		
		1	2	3
Eksport av metaller, verdi	Kalman	0,18	0,05	0,07
	X11	0,44	0,04	0,13
SAS' kvartalregnskap	Kalman	0,53	0,08	0,06
	X11	0,17	0,06	0
Omsetning av røyketobakk, volum	Kalman	1,09	0,29	0,09
	X11	0,41	0,32	0,03
Kjørelengde pr. dieseldrevet kjøretøy	Kalman	1,78	0,53	0,93
	X11	0,58	0,19	0,24
Eksport av elektrisk kraft, verdi	Kalman	8,79	0,84	0,35
	X11	2,27	1,30	9,06
Antall solgte symaskiner	Kalman	0,24	0,12	0,05
	X11	0,61	0,18	0,23
Totalt, ekskl. eksport av elektrisk kraft	Kalman	0,76	0,19	0,24
	X11	0,44	0,16	0,13

1) Revisjon nr. 1 er revisjonen fra 1. til 2. gang en observasjon er sesongjustert osv.

Antall observasjoner er lite, noe som gjør at resultatene må tolkes med varsomhet. Beregningene er gjort på grunnlag av tre 1.gangs-, to 2.gangs- og en 3.gangsrevisjoner for hver av de 6 seriene. Vi legger i første omgang merke til de ekstreme revisjonene i den sesongjusterte serien for eksport av elektrisk kraft. Dette skyldes de uvanlig store irregulære svingningene i denne serien som følge av variable nedbørsforhold, noe som også gir seg utslag i store prediksjonsfeil. En mekanisk sesongjusteringsmetode vil aldri kunne gi gode resultater for en slik serie. Den er derfor utelatt i den videre drøftingen.

Resultatene tyder på at, målt ved absoluttverdien av den gjennomsnittlige, prosentvise endringen i det sesongjusterte tallet fra første til annen gang observasjonen justeres, er revisjonen større ved bruk av Kalman-filter enn ved X11-metoden. Når vi har tilgang til flere datapunkter etter justeringstidspunktet, er endringene mindre og forskjellene ikke så markerte. Ved å se nærmere på de enkelte seriene framgår det imidlertid at når den relative variansen på sesongkomponenten, $\tilde{\sigma}_Y^2$, er svært liten i forhold til den relative variansen på trendkomponenten, $\tilde{\sigma}_\mu^2$, er revisjonene i den Kalman-filter-justerte serien mindre enn i den X11-justerte serien.

Dette tyder på at hvis det er lite stokastikk i sesongkomponentene, og de derfor påvirkes lite av dynamikken i Kalman-filteret, er det liten forskjell mellom å bruke Kalman-filteret eller X11-metoden. Hvis det derimot er mye stokastikk i sesongkomponentene, er endringene som følge av filterets dynamikk stor og komponentene blir ustabile i forhold til hva brukeren av sesongjusterte tall ofte finner akseptabelt. Dette er en utbredt kritikk mot Harvey's enkle strukturmodell spesifisert slik vi har brukt den (se f.eks. C.F. Ansley (1983)). Harvey har sjøl foreslått å legge struktur på restleddet til sesongkomponenten slik at bevegelsene blir langsommere (Harvey og Todd (1983)). Dette er en utvidelse som vi ennå ikke har fulgt opp, men som antakelig gir betydelige muligheter for forbedring av de Kalman-filter-baserte tallene. Vi må også være oppmerksomme på at X11-metoden kan tenkes å glatte for sterkt slik at revisjonene blir små, men at den sesongjusterte serien ikke fanger opp endringer som faktisk skjer.

Metoden må også være i stand til å rense ut mest mulig av de systematiske sesongutslagene i serien; dvs. den må kunne skille mellom hva som er irregulære og systematiske svingninger. Når serien er sesongjustert ved å bruke X11-metoden, faller det naturlig å undersøke om det er systematiske svingninger i den irregulære komponenten som er beregnet. I serier som blir sesongjustert ved Kalman-filtrering blir derimot den estimerte usikkerheten "spredt ut" på hver enkelt av de kildene til stokastikk som finnes i modellen. Fra tilstandslikningene i filteret har vi følgende sammenheng

$$\alpha_t = T_t \alpha_{t-1} + \eta_t$$

der η_t er restleddene i tilstandsvektoren. Hvis det er sesongsvingninger i ett eller flere av de estimerte restleddene, tyder dette på at sesongkomponentene i Kalman-filteret ikke er tilstrekkelig godt spesifisert. De restleddene som er analysert, er beregnet ved den rekursive filtreringen av serien, og framkommer fra uttrykket

$$a_t|T = T_t a_{t-1}|T + \hat{\eta}_t, \quad t = T-1, \dots, 1$$

Det vil her være nødvendig å foreta testing av restleddene til alle de stokastiske komponentene. I de seriene som her er undersøkt, er imidlertid nivået på trendendringskomponenten stort sett så lite i forhold til nivået i de andre at vi har valgt å utelate den med unntak av vareeksport av metaller. Vi har begrenset oss til å foreta en svært overfladisk test av

restleddsstrukturen. Vi har plottet residualene og sett at de i gjennomsnitt ligger nær 0. Dessuten har vi brukt en Box-Pierce-test for å undersøke om det er autokorrelasjon i restleddene. Særlig er vi da interessert i om det er 4.ordens autokorrelasjon. For en nærmere beskrivelse av denne testen henvises det til Pierce (1971). Ifølge denne testen kan vi ved et valgt signifikansnivå på 10 prosent forkaste en hypotese om at det ikke er 4.ordens autokorrelasjon i minst ett av restleddene for eksport av metaller, omsetning av røyketobakk og antall solgte symaskiner. Det kan imidlertid for alle disse varene se ut som det er 1.ordens autokorrelasjon i noen av restleddene. Dette kan skyldes at det er trend i noen av komponentene som ikke er fanget opp. Vi har også summert restleddene til hvert av elementene i tilstandsvektoren for hver vare og gjort tilsvarende test av denne variabelen. Resultatet av dette kan tyde på at systematikken i de ulike restleddene i en viss grad opphever hverandre, slik at testobservatorene beregnet på aggregert nivå, får lavere verdi og det er mindre grad av trend.

Ved bruk av Kalman-filteret kan vi i en viss utstrekning finne ut hvilken av tilstandslikningene som har et ikke-hvitt restledd og som evt. bør reformuleres. Hvis vi bruker X11-metoden, får vi bare estimert én irregulær komponent.

Materialet tyder på at de X11-beregnete restleddene inneholder betydelig 1.ordens autokorrelasjon. Denne tendensen er langt sterkere enn for de Kalman-filter-beregnete restleddene. Det vil generelt være et problem at en slik utstrakt mekanisk glatting av tallene vil kunne skape kunstig autokorrelasjon mellom de variable, se f.eks. Wallis (1974). Det synes imidlertid som de systematiske sesongeffektene er rensset ut.

Selv om dette materialet er for begrenset til å trekke noen vidtrekkende konklusjoner, synes det å være tendenser til at Kalman-filter-restleddene inneholder mindre systematikk enn X11-restleddene, noe som antakelig kan tilskrives selve spesifikasjonen av modellen.

Tabell 5. Testobservatorer for Box-Pierce-testing av autokorrelasjon i restleddene.

Serie	Grad	Kalmanfilter			X11
		Trendkomp.	Sesongkomp.	Sum	
Omsetning av røyketobakk, volum	1	9,07	3,12	6,16	8,84
	2	9,34	3,74	6,56	8,88
	3	19,87	3,78	7,21	9,02
	4	32,29	4,24	7,31	9,38
Antall solgte symaskiner	1	6,75			10,78
	2	6,85			10,83
	3	7,95			11,79
	4	13,49			13,95
Kjørelengde pr. diesel- drevet kjøretøy	1	0,07	9,61	0,02	17,38
	2	0,18	9,73	0,06	22,35
	3	0,20	9,87	1,49	22,90
	4	2,96	10,88	1,83	22,95
SAS' kvar- talsregnskap	1	2,61	8,56	0,00	11,10
	2	2,61	8,58	0,05	12,58
	3	4,87	8,62	0,16	13,07
	4	9,03	10,57	1,22	13,47
Eksport av metaller, verdi	1	0,01	21,66	0,01	11,69
	2	2,31	36,11	2,31	13,27
	3	2,59	42,24	2,60	13,30
	4	4,88	43,27	4,88	13,76

(trendendring)

(svært lavt nivå)

En av fordelene ved Kalman-filtrering som sesongjusteringsmetode, er at det eksplisitt kan beregnes et mål for usikkerheten i komponentene. Som det framgår i avsnitt 3 kan variansen til elementene i tilstandsvektoren estimeres ved

$$\text{var } \hat{\alpha}_{t|T} = \sigma^2 P_{t|T}$$

der $P_{t|T}$ er den estimerte kovariansmatrisen som framkommer ved filtreringen og σ^2 er estimert ved maksimering av likelihoodfunksjonen (se avsn. 6). På denne måten kan vi finne et uttrykk for hvor mye av stokastikken i serien som kan tilskrives sesongkomponentene og dermed hvor mye usikkerhet som er forbundet med de sesongjusterte tallene. Det er her av interesse å undersøke både hvor stor usikkerheten i den siste sesongjusterte observasjonen

er, og hvordan denne usikkerheten endres når vi suksessivt får flere observasjoner og dermed informasjon om den videre utvikling i serien. Vi har derfor beregnet standardavviket til sesongkomponentene for våre testserier. Vi har beregnet standardavviket for observasjoner fra 1985.1 første gang serien sesongjusteres dvs. i 1985.1, annen gang (i 1985.2) osv. Resultatene som er gjengitt i tabell 6 gir to hovedinntrykk:

- i) det beregnede standardavviket til sesongkomponenten er ofte stort i forhold til den suksessive endringen i den sesongjusterte serien.
- ii) verdien av standardavviket endres forholdsvis mye når vi får informasjon om utviklingen i serien i kvartalet som følger etter første sesongjusteringstidspunktet. Deretter reduseres imidlertid ikke standardavviket vesentlig når mer av den videre utviklingen blir kjent.

Dvs. at når vi bruker sesongjusterte tall for å analysere den økonomiske utviklingen, må vi være forsiktig med å legge for stor vekt på den aller siste observasjonen. Ofte er endringen målt fra foregående kvartal ikke så stor sammenliknet med usikkerheten i anslaget på sesongkomponenten. Det er kanskje ikke engang grunnlag for å si om det er oppgang eller nedgang i serien. I hvert fall må anslag på vekstprosenten brukes med stor varsomhet.

Tabell 6. Standardavvik til sesongkomponentene

Serie	Tidspkt.	Antall ganger sesongjustert		
		1	2	3
Omsetning av røyketobakk	1985.1	27,05	19,21	19,46
Antall solgte symaskiner		6,54	6,33	6,29
Kjørelengde pr. dieseldrevet kjøretøy		504,43	97,02	93,77
Eksport av metaller		42,77	42,49	42,47
SAS' kvartalsregnskap		9,26	7,11	7,00

Muligheten for å beregne anslag på usikkerheten i de ulike komponentene kan også være til nytte ved en nærmere analyse av de suksessive revisjonene i de sesongjusterte tallene. Etersom revisjonen er uavhengig av feilen i det endelige estimatet kan vi bruke (se Burridge og Wallis (1984)):

$$\text{var}(r_t^{j,h}) = \text{var}(S_{t,t+h} - S_{t,t+j}) = \text{var} \hat{y}_{t,t+h} - \text{var} \hat{y}_{t,t+j}, \quad j = 0, \dots, h$$

REFERANSER:

- Ansley, C.F. (1983): Comment til: A.C. Harvey and P.H.J. Todd (1983): Journal of Business and Economic Statistics, Vol. 1., No. 4.
- Biørn, E. og H. Olsen (1986): A Generalized Single Equation Error Correction Model and its Application on Quarterly Data. Discussion Paper no. 19/86 fra Statistisk Sentralbyrå.
- Burridge, P. og K.F. Wallis (1984): Calculating the Variance of Seasonally Adjusted Series. Paper presented at the Econometric Society Meeting in Madrid, September 1984.
- Chow, C.C. (1981): Econometric analysis by control methods, Wiley.
- Gersch, W. og G. Kitagawa (1983): The Prediction of Time Series With Trends and Seasonalities. Journal of Business & Economic Statistics, Vol. 1, No 3, 1983.
- Engle, R.F. (1978): Estimating Structural Models of Seasonality, Seasonal Analysis of Economic Time Series, ed. A. Zellner, Washington DC; US Dept of Commerce, Bureau of the Census, pp. 281-297.
- NAG (1981): NAG FORTRAN Library Manual.
- Harvey, A.C. (1981a): Time Series Models. Deddington.
- Harvey, A.C. (1981b): The Econometric Analysis of Time Series Phillip Allen Publ., Oxford.
- Harvey, A.C. (1984a): A unified view of Statistical Forecasting. Journal of Forecasting. 3. pp 245-75.
- Harvey, A.C. (1984b): Dynamic Models, the Prediction Error Decomposition and Quantitative Economics. D.F. Hendry and K.F. Wallis (Eds.).

- Harvey, A.C. og G.D.A. Phillips (1979): Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. Biometrika. 66. pp 49-58.
- Harvey, A.C. og R.G. Pierse (1984): Estimating missing observations in economic time series. Journal of the American Statistical Society, 79 pp. 125-131.
- Harvey, A.C. og P.H.J. Todd (1983): Forecasting Economic Time Series With Structural and Box-Jenkins Models: A Case Study Journal of Business & Economic Statistics, Vol 1. No 4 October 1983.
- Jensen, M. og L. Wahl (1985): Dokumentasjon av dataserier og dataarkiver knyttet til modellprosjektet KVARTS: 1975-versjonen. Interne notater nr. 85/2 fra Statistisk Sentralbyrå.
- Jensen, M., V. Knudsen, H. Olsen og T. Schweder (1985): Sesongjustering ved X11-metoden. Rapport fra en arbeidsgruppe. Interne notater nr. 85/19 fra Statistisk Sentralbyrå.
- Loe, L. (1987): Framskrivning av tidsseriedata i kvartalsvis nasjonalregnskap. Interne notater nr. 87/1 fra Statistisk Sentralbyrå.
- Olsen, H. og A. Skjæveland (1985): Teknisk dokumentasjon av beregningsopplegget for det kvartalsvise nasjonalregnskapet. Interne notater nr. 85/15 fra Statistisk Sentralbyrå.
- Pierce, D.A. (1971): Distribution of Residual Autocorrelation in the Regression Model with Autoregressive-Moving Average Errors. Journal of the Royal Statistics Society, Series B, 33 (1971).
- Shiskin, J., A.H. Young og J.C. Musgrave (1967): The X11 variant of the Census method II seasonal adjustment program. Technical paper no. 15. Bureau of Census, February 1967.
- Silverman, B.W. (1986): Density Estimation for Statistics and Data Analysis. London; Chapman and Hall.

TROLL (1981): Technical Report No. 32. Massachusetts Institute of Technology, Center for Computational Research in Economics and Management Science, October 1981.

de Vos, A.F., (1986): Forecasting the daily balance of the Dutch Giro. Paper presentert på The Sixth International Symposium On Forecasting, Paris.

Wallis, K.F., (1974): Seasonal Adjustment and Relations Between Variables. Journal of the American Statistical Association, 69 (1979).