Arbeidsnotater

S T A T I S T I S K   S E N T R A L B Y R Å

Dronningensgt. 16, Dep, Oslo 1. Tlf.*(02) 41 38 20

IO 77/21 25. mai 1977

BRUK AV TIDSREKKEANALYSE TIL ESTIMERING OG PREDIKSJON VED LØPENDE UTVALGSUNDERSØKELSER. ANNEN DEL: NUMERISKE RESULTATER FOR BYRAETS ARBEIDSKRAFTUNDERSØKELSER.

av

John Dagsvik*)

INNHOLD

*) Kjetil Sørlie er ansvarlig for programmeringsarbeidet.

SAMMENDRAG

Ved estimering på grunnlag av løpende utvalgsundersøkelser kan en oppnå presisjonsgevinst ved å utnytte informasjon fra tidligere undersøkelser. Sammensatt estimering er en slik metode. Ved sammensatt estimering benytter en imidlertid bare informasjon om korrelasjonen mellom utvalgsfeilen ved ulike tidspunkter og en gjør ingen forutsetninger om tidsutviklingen for de variablene en ønsker å estimere. I dette notatet viser vi at ved forholdsvis svake forutsetninger om disse variablene kan en oppnå betydelig presisjonsgevinst. Metoden er basert på ideene fra teorien om filtrering av prosesser som er belagt med støy og er også behandlet i et tidligere arbeid, Dagsvik 1976.

1. Innledning

Vi redegjør her for annen del av et prosjekt som består i å utrede om teknikkene fra tidsrekkeanalysen kan benyttes til å forbedre estimeringsmetoden i Arbeidskraftundersøkelsene (AKU). Vårt utgangspunkt er at det er en vesentlig korrelasjon mellom de aktuelle variabelverdiene på ulike tidspunkter og at vi ønsker å utnytte denne til å redusere utvalgsfeilen i AKU.

Imidlertid ønsker vi ikke å gjøre for "sterke" forutsetninger om korrelasjonsstrukturen da det er ønskelig at den metoden vi kommer fram til er forholdsvis robust overfor feilspesifikasjoner og usikre parameterestimater.

I første del av prosjektet (I) (Dagsvik 1976) gjorde vi et forsøk på å identifisere autoregresjonsmodeller for antall sysselsatte fordelt på noen hovednæringer. Vi behandlet videre problemet med å finne den beste lineære estimator (filter) for en variabel x_t på grunnlag av en uendelig tidsrekke av observasjoner $y_t, y_{t-1}, \ldots$ der y_t er et estimat for variabelen x_t ved tidspunkt t basert på et utvalg. Vi har med andre ord

$$y_t = x_t + v_t$$

der v_t er utvalgsfeilen. I det foreliggende notatet behandles estimering og testing av modellene samt utledning av den beste lineære estimator (filter) for x_t basert på et endelig antall observasjoner $y_t, y_{t-1}, \ldots, y_{t-m}$. Vi vil kalle denne metoden for filtermetoden.
De variablene vi studerer kan representeres ved modeller som er spesialtilfeller av modelltypen

$$\Delta_{12}x_t = r\Delta_{12}x_{t-1} + \lambda\Delta_{12}x_{t-12} - \lambda r\Delta_{12}x_{t-13} + e_t$$

hvor $\Delta_{12}x_t = x_t - x_{t-12}$, $\{e_t\}$ er stasjonær og autokorrelert, $|r|\leq 1$, $|\lambda|<1$ og måned er tidsenhet. Vi har videre funnet en autokorrelasjonsstruktur for $\{e_t\}$. Siden tidsrekkene i AKU er svært korte er det ønskelig å utlede optimale filtre basert på et endelig sett av observerte y_t-verdier. I motsetning til (I) hvor vi benyttet kompleks funksjonsteori under utledningen, bruker vi her den tradisjonelle teorien for minste kvadraters metode (Se for eks. Whittle 1963). Det som er hovedproblemet er å finne en hensiktsmessig metode til å beregne de respektive kovariansmatriser som inngår i formlene. Vi har løst dette problemet i kapittel 8 ved å innføre spesielle matrisefunksjoner som gjør oss i stand til å uttrykke kovariansmatrisene ved parametrene i modellene for $\{x_t\}$ og $\{v_t\}$ ved formler som er spesielt velegnet for programmering.

De data som er brukt er delvis hentet fra Sysselsettingstatistikken 1955-1970 og delvis fra AKU 1975-1976. Dataene fra Sysselsettingstatistikken i perioden 1955-1970 er benyttet til å estimere og teste modeller for sysselsatte fordelt på næring. Vi har brukt data på kommunenivå fra denne statistikken fra perioden 1967-1970 til å estimere autokorrelasjonen for utvalgsfeilen mellom utvalgsområdene. Dataene fra AKU på individnivå er brukt til å estimere autokorrelasjonen for utvalgsfeilen innen utvalgsområdene.

Parameterestimatene på grunnlag av Sysselsettingstatistikken er usikre siden variabeldefinisjonene i dagens AKU er forskjellige fra definisjonene som ble brukt før 1971. Estimatene på grunnlag av AKU er også usikre fordi datamaterialet er lite. Det er derfor av stor interesse å studere hvor robuste estimatorene er med hensyn til feilspesifikasjon av modeller og gale parameterestimater. Vi har derfor utledet formler for presisjonstapet når de anvendte parameterverdiene er forskjellige fra antatt "sanne verdier" for nivå- og endringstallestimatorene. Resultatene av disse beregningene er gitt i kapittel 6 og viser at selv med store feil i parameterestimatene får vi lite tap i presisjon ved estimeringen.

I mange andre land har en brukt metoder basert på å kombinere estimater ved flere tidspunkter, såkalt sammensatt estimering (Se Dagsvik 1975). I sammensatt estimering utnytter en bare at utvalgsfeilen er autokorrelert og en gjør ingen forutsetninger om den variabelen en ønsker å estimere. Vi sammenlikner prinsippene sammensatt estimering bygger på med de som ligger til grunn for filtermetoden og vi studerer også effekten av å kombinere de to metodene ved først å lage sammensatte estimater og deretter benytte filtermetoden på tidsrekken av sammensatte estimater.

I framstillingen er de matematiske utledningene skilt ut som eget kapittel og appendiks. Konklusjonen i kapittel 2 kan i hovedsak leses uavhengig av de andre kapitlene.

Jeg takker for nyttige kommentarer fra Fridstrøm og Thomsen under bearbeidelsen av manuskriptet.

Jeg vil takke Børsum spesielt for programmeringshjelp under utprøving og testing av metoden. Sørlie har en stor del av æren for gjennomføringen av beregningene. Uten hans hjelp ville prosjektet i beste fall blitt sterkt forsinket.

2. Konklusjon

Datamaterialet som publiseres på grunnlag av AKU er "rå-data" i den forstand at det skal kunne anvendes til ulike typer analyser. Det er derfor viktig at produksjonen av AKU-estimatene ikke er basert på forutsetninger som kan være vanskelige å kontrollere og som erfaringsmessig har lett for å bli underslått av brukere.

På den andre siden ligger det vesentlig informasjon i det innsamlede datamaterialet om tidsutviklingen av de aktuelle variable. Problemet er hvordan en skal kunne utnytte denne informasjonen til å redusere utvalgsfeilen ved estimeringen uten å gjøre forutsetninger som, hvis de er gale vil medføre at datakvaliteten forringes. Det gjelder altså å gi en forholdsvis generell beskrivelse av de prosessene som studeres samtidig som forutsetningene må være sterke nok til å gi presisjonsgevinst av betydning. Hvorvidt disse motstridende hensyn med rimelighet lar seg forene er naturligvis avhengig av stabiliteten i de prosessene vi ønsker å estimere og av størrelsen på utvalgsvariansen i de oppblåste tall.

De studiene som er foretatt her viser imidlertid at en kan oppnå en betydelig presisjonsgevinst ved å ta utgangspunkt i forbausende svake forutsetninger. De vesentlige antagelser som gjøres er følgende: La x_t være den variabelen vi ønsker å estimere ved kvartal t. Prosessen $\{e_t\}$ definert ved

$$e_t = x_t - x_{t-4} - (x_{t-1} - x_{t-5})$$

er stasjonær med forventning null. Med andre ord antas det at den kvartalsvise endring av den årlige endringen svinger tilfeldig omkring nivået null. Videre antar vi at utvalgsfeilprosessen $\{v_t\}$ (uobserverbar) er stasjonær. For å beregne det optimale filter må vi ha et grovt estimat for var $\{e_t\}$ / var $\{v_t\}$. Det er også av betydning å kjenne autokorrelasjonsfunksjonen for prosessene $\{e_t\}$ og $\{v_t\}$ men konsekvensene av å bruke gale estimater her er små. For å beregne presisjonsgevinsten ved filtermetoden må vi ha estimater (grove) for autokorrelasjonsfunksjonen til $\{v_t\}$.

Vi kan upresist si at stasjonæritetsbetingelsen for $\{e_t\}$ betyr at trenden for $\{x_t\}$ er lineær i perioder og skifter nå og da retning.

På grunnlag av Sysselsettingstatistikken 1955-1970 har vi funnet at $\{e_t\}$ er stasjonær for antall sysselsatte lønnstakere fordelt på kjønn og næring. For alle disse variablene avtar autokorrelasjonen for $\{e_t\}$ mot null svært raskt, hvilket antyder at $\{e_t\}$ er langt fra ikke-stasjonæritet. Dersom autokorrelasjonsfunksjonen avtar langsomt mot null vil presisjonsgevinsten ved metoden reduseres. Presisjonsgevinsten ved metoden er avhengig av forholdet var $\{e_t\}$ / var $\{v_t\}$ og av autokorrelasjonen til $\{v_t\}$. I de aktuelle anvendelser er reduksjonen av utvalgsvariansen innen utvalgsområdene ca. 40-50% for nivåtall når en har tidsrekker på minst 20 kvartalers lengde. Dette tilsvarer en reduksjon av standardavviket på 20-30%. For endringstall vil utvalgsfeilen mellom utvalgsområdene elimineres siden denne er tilnærmet konstant i tid. Den totale variansen ved estimering av endringstall er derfor lik variansen innen utvalgsområdene og denne vil reduseres med 60-70% som tilsvarer en reduksjon i standardavviket på ca. 37-45%.

For korte tidsrekker (opp til 3 år) vil det som regel lønne seg å foreta sammensatt estimering først og dernest bruke filtermetoden på de sammensatte estimater. For lange tidsrekker får vi derimot ingen ekstra gevinst ved å kombinere disse metodene.

Vår metode kan også benyttes til kortsiktige prognoser på inntil fire kvartaler. For prediksjon lengre fram i tiden er prediksjonsintervallet så stort at metoden har liten verdi.

I Byrået ble en ny utvalgsplan tatt i bruk fra og med 1. kvartal 1975. Dette medfører brudd i autokorrelasjonsfunksjonen for utvalgsfeilen og kan utnyttes til å redusere utvalgsfeilen ytterligere for estimering av nivåtall mens vi ikke får noen ekstra gevinst ved estimering av endringer.

3. Estimering av modeller for antall sysselsatte fordelt på næring

Vi betrakter her en stokastisk prosess $\{x_t\}$ som er antall sysselsatte i en eller annen sysselsettingsgruppe (eventuelt relativ andel i sysselsettingsgruppen). Tidsenheten er én måned. Karakteristisk for denne prosessen er at den har sesongsvingninger og mer langsiktige bølgebevegelser. Ved å ta differensoperasjonen $\Delta_{12}x_t = x_t - x_{t-12}$ vil en som regel fjerne en stor del av sesongeffekten. Vi vil derfor forsøke å etablere en modell for de årlige endringer $\{\Delta_{12}\, x_t\}$ i stedet for å arbeide med den opprinnelige prosessen. Vi forutsetter ikke nødvendigvis at $\{\Delta_{12}\, x_t\}$ er stasjonær. La oss først betrakte en autoregresjonsmodell av typen

$$(3.1)\qquad \Delta_{12}\, x_t = r\Delta_{12}\, x_{t-1} + e_t$$

der e_t er et stokastisk restledd og r er en parameter, $0<r\leq 1$. Dersom $r = 1$ betyr det at prosessen $\{\Delta_{12}\, x_t\}$ er ikke-stasjonær mens $|r|<1$ medfører at $\{\Delta_{12}\, x_t\}$ svinger tilfeldig omkring null, dvs. vi har stasjonæritet. Dersom restleddprosessen $\{e_t\}$ har autokorrelasjonsfunksjon lik null definerer (3.1) en lineær Markov-prosess. Nå vil som regel operasjonen $\Delta_{12}\, x_t$ ikke fjerne hele sesongeffekten. Dette betyr at vi må vente å finne at e_t er spesielt høyt korrelert med e_{t-1}, e_{t-11}, e_{t-12} og e_{t-13}. Vi postulerer derfor følgende modell for e_t : $\{e_t\}$ er stasjonær med korrelasjonsstruktur

$$\text{Korr}\{e_t,\, e_{t-11}\} = \text{Korr}\{e_t,\, e_{t-13}\}$$

$$\text{Korr}\{e_t,\, e_{t-k}\} \neq 0 \text{ for } k = 1,\ 11,\ 12 \text{ og } 13$$

og

$$\text{Korr}\{e_t,\, e_{t-k}\} = 0 \quad \text{ellers.}$$

En slik antakelse er ekvivalent med at e_t kan uttrykkes på formen

$$(3.2)\qquad e_t = a_t - \alpha a_{t-1} - \beta(a_{t-12} - \alpha a_{t-13})$$

hvor a_t, a_{t-1},... er ukorrelerte med $E\{a_t\} = 0$, $\text{Var}\{a_t\} = \sigma^2$, α og β er parametre. Prosessen $\{a_t\}$ kalles gjerne "hvit støy" eller prosessen av tilfeldige sjokk. Modellen (3.1) og (3.2) er en generalisering av (4.1) i (I) idet vi får denne modellen som et spesialtilfelle ved å sette $r = 1$ og $\alpha = 0$. Fra (3.1) og (3.2) følger at den autoregressive representasjon av modellen er

$$(3.3)\qquad \Delta_{12}\, x_t = \sum_{j=1}^{\infty}\gamma_j\, \Delta_{12}\, x_{t-j} + a_t$$

der $\quad \gamma_j = \alpha^{j-1}(r-\alpha)$ for $j = 1, 2, ..11$

$$\gamma_{12} = \alpha^{11}\,(r-\alpha) - \beta,$$

$$\gamma_{13} = (\alpha^{12} + \beta)\,(r - \alpha)$$

og

$$\gamma_j = \alpha\gamma_{j-1} + \beta\gamma_{j-12} - \alpha\beta\gamma_{j-13} \text{ for } j\geq 14$$

Ønsker vi en regresjonsmodell for den opprinnelige prosessen får vi umiddelbart

$$(3.4)\qquad x_t = x_{t-12} + \sum_{j=1}^{\infty}(\gamma_j - \gamma_{j-12})\, x_{t-j} + a_t$$

der γ_j defineres lik null for $j\leq 0$. Det kan nå vises at restleddet a_t i (3.3) og (3.4) er ukorrelert med "strukturdelen". Den autoregressive representasjonen (3.3) viser at dersom modellen er riktig er $\{\Delta_{12} x_t\}$ ikke en lineær Markov-prosess.

Testing av modellen foregår på den måten at de ukjente parametrene estimeres og en kan derved finne estimater $\hat{a}_t$ for restleddet a_t. Under forutsetning om normalitet kan det vises at

$$(3.5) \quad Q_q = N \sum_{k=1}^{q} \hat{\rho}_k^2(\hat{a}) \qquad q<N$$

er tilnærmet χ^2 -fordelt der N er antall observasjoner og $\hat{\rho}_k(\hat{a})$ er den empiriske autokorrelasjonsfunksjonen.

Resultatene fra parameterestimeringen og testingen er gitt i tabell 1 med 95% konfidensintervall for parametrene. Det datamaterialet vi har brukt er hentet fra Sysselsettingstatistikken i perioden 1955-1970 (192 observasjoner). Dette data-materialet gir estimater for r som er tilnærmet lik 1. Testobservatoren i tabell 1 er tilnærmet χ^2 -fordelt med 22 frihetsgrader. 25%, 10% og 5% fraktilene er henholdsvis 26, 30,8 og 33,9. Vi ser at bortsett fra Bygge- og anleggsvirksomhet er alle testobservatorene mindre enn 25% fraktilen.

For ên næring, nemlig sysselsatte i industri gir (3.1) og (3.2) dårlig tilpassing og vi har følgelig generalisert (3.1) til modellen

$$(3.6) \quad \Delta_1\Delta_{12} x_t = \lambda\Delta_1\Delta_{12} x_{t-12} + a_t - \beta a_{t-12}, \qquad |\lambda|<1,$$

som viser seg å gi god tilpasning. Sysselsatte i industri er estimert på grunnlag av kvartalsdata og parameterestimatene er gitt i tabell 2.

Tabell 1. Parameterestimater og testobservatorer for modellene for noen hovednæringer

	$\hat{r}$	Konfidens-intervall	$\hat{\alpha}$	Konf. intervall	$\hat{\beta}$	Konf. intervall	Var a_t	Q_{24}
Kraft og vannforsyning ...	1		0.39	0,25 0,53	0,29	0,15 0,44	$1,96\cdot10^4$	17,5
Bygge- og anleggs-virksomhet	0,86	0,78 0,93	0		0,62	0,49 0,75	$3,55\cdot10^6$	31,8
Varehandel, hotell og restaurant	1		0,36	0,21 0,50	0,47	0,32 0,60	$0,84\cdot10^6$	20.0
Transport lagring, post og telekommunikasjoner	1		0,16	0 0,31	0,70	0,60 0,80	$0,59\cdot10^6$	21,2
Bank og finansierings-virksomhet m.v.	1		0,14	0 0,29	0,43	0,28 0,58	$1,37\cdot10^4$	12,3
Off., sos. og privat tjenesteyting	1		0,18	0,03 0,33	0,54	0,40 0,68	$0,61\cdot10^6$	23,3
Sum	1		0,25	0,01 0,39	0,69	0,57 0,80	$11,8\cdot10^6$	19,0

Hittil har vi studert de aktuelle prosessene når måned er tidsenhet. I AKU foretas imidlertid undersøkelsene kvartalsvis slik at vi trenger en representasjon av modellene når kvartal er tidsenhet. I appendiks B er det vist hvordan "månedsmodellen" som er behandlet ovenfor kan overføres til en kvartalsmodell med samme form og hvor parametrene i kvartalsmodellen enkelt kan uttrykkes ved parametrene i månedsmodellen. Siden sammenhengen mellom parametrene i den gamle og den nye modellen er en-entydig vil sannsynlighetsmaksimeringsestimatorene for den gamle modellen automatisk gi sannsynlighetsmaksimeringsestimatorer for den nye modellen (asymptotisk) ved å benytte (B.8) og (B.9). Modellen for sysselsatte i industri er som nevnt estimert direkte på grunnlag av kvartalsdata mens parametrene i kvartalsmodellen for de andre næringene er estimert ved bruk av resultatene i tabell 1, (B.8) og (B.9). De kvartalsparametrene som derved framkommer er gitt i tabell 2.

Tabell 2. Parameterestimater for noen hovednæringer for kvartalsmodellene

	$\hat{r}$	$\hat{\alpha},\hat{\lambda}$	Konfidensinterv.	$\hat{\beta}$	Konfidensinterv.	Var a'_t
Industri	1	0,32 α=0	0,07 0,58	0,77	0,59 0,95	$5{,}6 \cdot 10^6$
Kraft og vannforsyning	1	0,22 λ=0	0,12 0,34	0,29	0,15 0,44	$2{,}6 \cdot 10^4$
Bygge- og anleggsvirksomhet	0,7	0 λ=0		0,62	0,49 0,75	$5{,}2 \cdot 10^6$
Varehandel hotell og restaurant	1	0,19 λ=0	0,09 0,31	0,47	0,32 0,60	$1{,}6 \cdot 10^6$
Transport, lagring, post og telekommunikasjon	1	0,07 λ=0	0 0,16	0,70	0,60 0,80	$1{,}4 \cdot 10^6$
Bank og finansieringsvirksomhet m.v.	1	0,07 λ=0	0 0,16	0,43	0,28 0,58	$2{,}7 \cdot 10^4$
Off.,sos. og privat tjenesteyting	1	0,08 λ=0	0 0,17	0,54	0,40 0,68	$1{,}4 \cdot 10^6$
Sum		0,21 λ=0	0,04 0,22	0,69	0,57 0,80	$22 \cdot 10^6$

4. Utvalgsfeilen

Som beskrevet i (I) er Byråets' utvalgsplan totrinns hvor ett utvalgsområde trekkes i hvert stratum og holdt fast et visst antall år.

Den totale utvalgsfeilen v_t er altså summen av utvalgsfeilen innen og mellom utvalgsområdene, betegnet med henholdsvis v_{2t} og v_{1t}. Vi antar at $\left\{v_{2t}\right\}$ og $\left\{v_{1t}\right\}$ er stasjonære. La $x_t(j)$ være den aktuelle variabel knyttet til j-te person og la $\underset{\sim}{J} = (J_1, J_2, \ldots)$ være numrene på de respektive uttrukne utvalgsområdene. Vi antar videre at y_t er en forventningsrett estimator for x_t dvs.

$$E\left\{y_t \mid \underline{x}_t\right\} = x_t$$

der

$$\underline{x}_t = \left\{x_t(j) : j = 1, 2, \ldots\right\} .$$

I (I) side 10 viste vi at

$$\text{(4.1)} \quad \operatorname{cov}\left\{v_t, v_{t-r}\right\} = E \operatorname{cov}\left\{y_t, y_{t-r} \mid \underline{x}_t, \underline{x}_{t-r}\right\} .$$

Uttrykket

$$\text{(4.2)} \quad \operatorname{cov}\left\{y_t, y_{t-r} \mid \underline{x}_t, \underline{x}_{t-r}\right\}$$

er kovariansen i populasjonen mellom estimatorene y_t og y_{t-k} slik den er definert i utvalgsteorien. Autokovariansen mellom og innen utvalgsområdene i populasjonen er definert ved henholdsvis

$$\text{(4.3)} \quad \operatorname{cov}\left\{E\,(y_t \mid \underset{\sim}{J}, \underline{x}_t),\; E\,(y_{t-r} \mid \underset{\sim}{J}, \underline{x}_{t-r})\right\}$$

og

$$\text{(4.4)} \quad E \operatorname{cov}\left\{y_t, y_{t-r} \mid \underset{\sim}{J}, \underline{x}_t, \underline{x}_{t-r}\right\}$$

Hvor $\underline{x}_t$ og $\underline{x}_{t-r}$ er holdt faste og forventning og kovarians er tatt over alle mulige utvalg $\underset{\sim}{J}$.

Herav følger at

$$\text{(4.5)} \quad \operatorname{cov}\left\{v_{1t}, v_{2,t-r}\right\} = E \operatorname{cov}\left\{E\,(y_t \mid \underset{\sim}{J}, \underline{x}_t),\; E\,(y_{t-r} \mid \underset{\sim}{J}, \underline{x}_{t-r})\right\}$$

og

$$\text{(4.6)} \quad \operatorname{cov}\left\{v_{2t}, v_{2,t-r}\right\} = E\, E \operatorname{cov}\left\{y_t, y_{t-r} \mid \underset{\sim}{J}, \underline{x}_t, \underline{x}_{t-r}\right\}$$

er de respektive autokovariansfunksjoner innen og mellom utvalgsområdene. Fra (4.1), (4.5) og (4.6) får vi at

$$\operatorname{cov}\left\{v_t, v_{t-r}\right\} = \operatorname{cov}\left\{v_{1t}, v_{1,t-r}\right\} + \operatorname{cov}\left\{v_{2t}, v_{2,t-r}\right\}$$

hvilket betyr at $\left\{v_{1t}\right\}$ og $\left\{v_{2t}\right\}$ er ukorrelerte prosesser. Uttrykkene (4.3) og (4.4) kan estimeres eller beregnes eksakt ved standard metoder fra utvalgsteorien. Følgelig kan (4.5) og (4.6) estimeres ved å ta gjennomsnittet av de beregnede populasjonskovarianser over tid.

Beregninger av (4.3) er foretatt ved hjelp av Sysselsettingstatistikken i tidsrommet januar 1967 - desember 1970, og er framstilt i tabell 3 med kvartal som tidsenhet.

Tabell 3. Empiriske autokorrelasjoner mellom utvalgsområdene

	Industri		Bygge- og anleggsvirk.		Kraft og vannforsyning		Varehandel, hotell og rest.	
	M	K	M	K	M	K	M	K
r	0,99	0,99	0,97	0,97	0,98	0,99	0,99	0,99
2	0,99	0,99	0,94	0,95	0,94	0,96	0,99	0,99
3	0,99	0,99	0,82	0,84	0,94	0,92	0,99	0,99
4	0,99	0,99	0,82	0,81	0,94	0,89	0,99	0,99
5	0,99	0,99	0,80	0,77	0,93	0,90	0,99	0,99
6	0,99	0,99	0,75	0,70	0,96	0,88	0,99	0,99
7	0,99	0,99	0,62	0,60	0,97	0,92	0,99	0,99
8	0,98	0,99	0,63	0,50	0,96	0,92	0,99	0,99
9	0,98	0,98	0,60	0,53	0,95	0,92	0,98	0,98
10	0,98	0,98	0,61	0,52	0,95	0,91	0,98	0,98
11	0,98	0,98	0,58	0,51	0,96	0,92	0,98	0,98
12	0,98	0,98	0,60	0,46	0,93	0,91	0,98	0,98
13	0,97	0,98	0,60	0,48	0,92	0,90	0,97	0,98
14	0,97	0,98	0,56	0,50	0,91	0,89	0,97	0,98
15	0,97	0,98	0,50	0,36	0,90	0,88	0,97	0,98

	Bank og finansieringsvirksomhet		Transport, lagring, post og telekom.		Off., sos. og privat tjenesteyting		Sysselsatte lønnstakere Total	
	M	K	M	K	M	K	M	K
r	0,99	0,99	0,99	0,98	1	0,98	0,99	0,98
2	0,99	0,99	0,99	0,98	1	0,99	0,99	0,99
3	0,99	0,98	0,99	0,96	1	0,98	0,98	0,98
4	0,98	0,97	0,99	0,96	1	0,99	0,98	0,99
5	0,98	0,97	0,98	0,96	1	0,97	0,98	0,97
6	0,98	0,95	0,98	0,94	1	0,98	0,97	0,98
7	0,98	0,96	0,98	0,95	1	0,98	0,97	0,98
8	0,98	0,96	0,81	0,94	1	0,98	0,97	0,98
9	0,98	0,96	0,78	0,92	0,997	0,97	0,95	0,97
10	0,98	0,95	0,77	0,90	0,997	0,97	0,95	0,96
11	0,97	0,95	0,77	0,91	1	0,97	0,96	0,97
12	0,97	0,96	0,77	0,90	0,997	0,97	0,96	0,97
13	0,98	0,96	0,79	0,91	0,996	0,96	0,95	0,96
14	0,97	0,95	0,78	0,89	0,996	0,96	0,95	0,95
15	0,97	0,94	0,77	0,90	0,996	0,96	0,95	0,96

Vi ser at bortsett fra næringen Bygg og anlegg avtar disse autokorrelasjonene svært langsomt. For sysselsatte i Off., sosial og privat tjenesteyting er autokorrelasjonen ekstremt høy.

Når autokorrelasjonsfunksjonen har verdier som ligger nær 1 er utvalgsfeilen mellom utvalgsområdene tilnærmet konstant over tid såfremt utvalgsområdene holdes faste. Dette medfører at det er liten hjelp i å utnytte informasjon fra tidligere undersøkelser til å redusere utvalgsfeilen mellom utvalgsområdene såfremt utvalgsområdene holdes faste. Dette resultatet som intuitivt er opplagt er bevist i appendiks C.

Byråets utvalgsplan er imidlertid endret på et tidspunkt slik at utvalgsfeilen før og etter skifte av utvalgsplan er ukorrelert. Dette kan utnyttes til å redusere utvalgsfeilen også mellom utvalgsområdene. For endringstall vil utvalgsfeilen mellom utvalgsområdene praktisk talt elimineres (såfremt endringene ikke er beregnet for tidspunkter før og etter skifte av utvalgsplan).

Variansen mellom utvalgsområdene, er beregnet av Sæbø 1976 ved hjelp av folketellingsdata.

I Dagsvik 1974 er det gitt variansestimater for den gamle utvalgsplanen.

Innen hvert av de uttrukne utvalgsområdene (som er faste) trekkes et utvalg av husholdninger. Dette annet trinns utvalg roterer etter en spesiell rotasjonsplan, det vil si utvalget er overlappende for tidspunkter som legges ett, tre, fire og fem kvartaler fra hverandre. Rotasjonsplanen er beskrevet blant annet i appendiks A.2 og i (I). Følgen av en slik rotasjonsplan er at v_{2t} blir korrelert med $v_{2,t-1}$, $v_{2,t-3}$, $v_{2,t-4}$ og $v_{2,t-5}$.

I Dagsvik 1975 er det gjort autokorrelasjonsberegninger for noen næringer. Disse beregningene er basert på den gamle utvalgsplanen (før 1975) og en har ikke tatt hensyn til stratifiseringen. Vi må derfor regne med at disse estimatene er noe usikre.

I tabell 4 er det gitt estimater for den gjennomsnittlige autokorrelasjonen ρ_{2r} = korr $\left\{v_{2t}, v_{2,t-r}\right\}$, $r = 1$ og 4, for noen hovednæringer. For $r = 3$ og 5 har vi anslått verdier for ρ_{2r}. Legg merke til at den definisjonen av ρ_{2r} som er gitt her <u>ikke</u> samsvarer med definisjon (3) i Dagsvik 1975. Lar vi ρ'_r være autokorrelasjonen i populasjonen slik den er definert ved (3) i Dagsvik 1975 har vi i vår notasjon at

$$\rho'_r = \text{Korr}\left\{y_t,\ y_{t-r} \mid J, \underline{x}_t,\ \underline{x}_{t-r}\right\}$$

og vi har at

$$\rho_{2r} = 0{,}5\ E\left\{\rho'_r\right\} \quad \text{for} \quad r = 1 \text{ og } 4,$$

og

$$\rho_{2r} = 0{,}25\ E\left\{\rho'_r\right\} \quad \text{for} \quad r = 3 \text{ og } 5,$$

fordi halvparten av utvalget er felles ved tidspunktent t og t-1, t og t-4, og en fjerdedel av utvalget er felles ved tidspunktene t og t-3, t og t-5.

Tabell 4. Empiriske autokorrelasjoner innen utvalgsområdene

r	1	2	3	4	5
ρ_{2r}	0,43	0	0,2	0,4	0,18

Anta nå at vi først benytter den sammensatte estimator $\tilde{\tilde{x}}_t$ definert ved A.1 (Se appendiks A eller Dagsvik 1975) for å forbedre "råestimatet" y_t.

Da sammensatt estimering bare har effekt på utvalgsfeilen innen utvalgsområdene blir altså v_{1t} uendret mens v_{2t} reduseres til $\tilde{\tilde{v}}_{2t}$. Autokorrelasjonsfunksjonen blir vesentlig mer komplisert for $\left\{\tilde{\tilde{v}}_{2t}\right\}$ enn for $\left\{v_{2t}\right\}$. I appendiks A har vi utledet formler som kan benyttes til å beregne estimater for den "nye" autokorrelasjonsstrukturen når den "gamle" er estimert.

Med utgangspunkt i tabell 4 og med p = q = 0,3 har vi beregnet

$$\mathrm{Var}\left\{\tilde{\tilde{v}}_{2t}\right\} / \mathrm{Var}\left\{v_{2t}\right\} \text{ og } \mathrm{Korr}\left\{\tilde{\tilde{v}}_{2t}, \tilde{\tilde{v}}_{2,t-r}\right\}.$$

Beregningene er gjengitt i tabell 5.

Tabell 5. Empiriske autokorrelasjoner. Sammensatt estimering.

r	1	2	3	4	5	6	7	8	9	10
c_r/c_0	0,51	0,18	0,23	0,47	0,23	0,06	0,08	0,1	0,03	0

$\mathrm{Var}\left\{\tilde{\tilde{v}}_{2t}\right\} / \mathrm{Var}\left\{v_{2t}\right\} = 0{,}73$

Når dette skrives har vi ikke estimater for variansen til utvalgsfeilen innen utvalgsområdene for den nye utvalgsplanen.

5. Lineær filtrering

Vi beskriver her kort prinsippene som ligger til grunn for hva vi har kalt filtermetoden. Vi begrenser oss til filtrering idet løsning av prediksjonsproblemet er fullstendig analogt.

Den observerte prosessen $\{y_t\}$ er en sum av den latente prosessen $\{x_t\}$ som vi ønsker å estimere verdiene for, og utvalgsfeilen $\{v_t\}$. Altså

(5.1) $\quad y_t = x_t + v_t .$

Prosessen $\{y_t\}$ er "råestimator" for $\{x_t\}$ som er konstruert slik at den er forventningsrett i "utvalgsteoriforstand", dvs. over tenkte repeterte utvalg. Altså er

$$E\left\{y_t \mid x_t\right\} = x_t .$$

I praksis er denne forutsetningen av flere grunner bare tilnærmet riktig. Vi antar imidlertid at tilnærmelsen er god. Det følger nå umiddelbart at

$$E\left\{x_t\, v_t\right\} = E\left\{x_t\, E\left\{y_t - x_t | x_t\right\}\right\} = 0 = E\left\{x_t\right\} E\left\{v_t\right\},$$

dvs. x_t og v_t er ukorrelerte. Som nevnt i forrige avsnitt forutsetter vi at $\{v_t\}$ er en stasjonær prosess med varians s^2. Dette er en tilnærmelse til den virkelige situasjonen fordi variansen til utvalgsfeilen vil endre seg dersom x_t endrer seg mye. Vi regner likevel med at de relative endringene for x_t er så små at vi med god tilnærmelse kan gjøre antagelsen ovenfor.

Vårt primære mål er å bestemme vektene ζ_j i det lineære filter

(5.2) $\quad \hat{x}_t = \zeta_0\, y_t + \zeta_1\, y_{t-1} + .. + \zeta_m\, y_{t-m}$

som gir det "beste" estimat for x_t. Det beste filtret betyr i vår sammenheng det som minimerer forventet kvadratavvik

(5.3) $\quad E\left\{\hat{x}_t - x_t\right\}^2 .$

I (I) løste vi dette minimeringsproblemet for stor m ($m=\infty$) ved å bruke kompleks funksjonsteori. I det foreliggende notat har vi tatt utgangspunkt i den tradisjonelle teorien for minste kvadraters metode. Ved å innføre hensiktsmessige matriseoperatorer kan de formlene som utledes uttrykkes ved matriser som gjør dem velegnet til programmering. For enkelte modeller blir disse matrisene uendeligdimensjonale, men en kan i praksis oppnå god tilnærmelse ved endelig-dimensjonale matriser med lav dimensjon.

Innfører vi $\hat{v}_t = y_t - \hat{x}_t$ får vi at

$$E\left\{\hat{x}_t - x_t\right\}^2 = E\left\{\hat{v}_t - v_t\right\}^2,$$

dvs. $\hat{v}_t$ er det beste lineære filter for v_t.

Vi har valgt å estimere endringer ved å ta differansen mellom de respektive nivåtallsestimater. Dette gir ikke optimale estimater men en oppnår ved dette konsistens mellom nivå og endringstallestimatene.

La

(5.4a) $M_0 = 1 - E\left\{\hat{v}_t - v_t\right\}^2 / s^2$

og

(5.4b) $M_j = 1 - E\left\{\hat{v}_t - \hat{v}_{t-j} - v_t + v_{t-j}\right\}^2 / \mathrm{Var}\left\{v_t - v_{t-j}\right\}$

for j>0. M_0 og M_j er et mål for presisjonsgevinsten for henholdsvis nivå og endringstall. I tabellen har vi brukt disse størrelser som mål for presisjonsgevinsten.

Dersom utvalgsområdene holdes faste er det ikke mulig å redusere utvalgsfeilen mellom utvalgsområdene nevneverdig ved bruk av filtre siden denne utvalgsfeilen da har tilnærmet konstant verdi. Det beste en kan gjøre i denne situasjonen er derfor å konstruere det filtret som minimerer

$$E\left\{\hat{v}_t - v_{2t}\right\}^2.$$

Dette intuitive opplagte resultatet er formelt bevist i appendiks C. I vår situasjon derimot kan vi oppnå reduksjon av utvalgsfeilen mellom utvalgsområdene siden ny utvalgsplan er tatt i bruk pr. 1/1-75.

6. Robusthetsegenskaper

De modellene vi har identifisert og estimert på grunnlag av data er spesialtilfeller av den generelle modellen (kvartal er tidsenhet)

(6.1) $\Delta_4 x_t + \lambda_1 \Delta_4 x_{t-1} + \ldots + \lambda_5 \Delta_4 x_{t-5}$

$= a_t + \phi_1 a_{t-1} + \ldots + \phi_5 a_{t-5}$

hvor λ_j og ϕ_j er parametre. Vi tenker oss nå at prosessen $\left\{x_t\right\}$ genereres av denne modellen hvor parametrene λ_j, ϕ_j og $\mathrm{Var}\left\{a_t\right\}$ er ukjente.

La $\tilde{x}_t$ være et vilkårlig lineært filter av typen (5.2). Da er

$$E\left\{\tilde{x}_t - x_t\right\}^2 - E\left\{\hat{x}_t - x_t\right\}^2 \geq 0$$

og

(6.2a) $R_0 = \left[E\left\{\tilde{x}_t - x_t\right\}^2 - E\left\{\hat{x}_t - x_t\right\}^2\right]/s^2$

er et mål for hvor mye vi taper i presisjon ved å bruke $\tilde{x}_t$ istedet for $\hat{x}_t$. Tilsvarende er

(6.2b) $R_j = \left[E\left\{\tilde{x}_t - \tilde{x}_{t-j}\right\}^2 - E\left\{\hat{x}_t - \hat{x}_{t-j}\right\}^2\right]/ \mathrm{Var}\left\{v_t - v_{t-j}\right\}$

et mål for hvor mye vi taper i presisjon ved å bruke $\tilde{x}_t - \tilde{x}_{t-j}$ istedet for $\hat{x}_t - \hat{x}_{t-j}$ ved estimering av endringen $x_t - x_{t-j}$.

Anta nå at vi har estimert parametrene i modellen (6.1) men at disse estimatene avviker vesentlig fra de ukjente "sanne" verdier. Et viktig spørsmål er hvordan resultatene vi kommer fram til avhenger av avviket fra de sanne verdier i parameterestimatene. Vi er spesielt interessert i hvor robust filtrene $\hat{x}_t$ og $\hat{x}_t - \hat{x}_{t-j}$ som estimatorer for nivået x_t og endringen $x_t - x_{t-j}$, er med hensyn på feil i parametrene. La nå $\tilde{x}_t$ betegne filteret som er beregnet på grunnlag av de "gale" parameterestimatene. Som mål for robusthet bruker vi R_j definert ved (6.2). Formler for R_j er gitt i kapittel 8.

Dette formelapparatet kan benyttes for generelle modeller av typen (6.1) men vi har innskrenket oss til spesialtilfellet

(6.3) $\Delta_4 x_t = r\Delta_4 x_{t-1} + \lambda(\Delta_4 x_{t-4} - r\Delta_4 x_{t-5}) + e_t$

hvor

$$e_t = a_t - \alpha a_{t-1} + \beta(a_{t-4} - \alpha a_{t-5}).$$

I denne modellen inngår de fem parametrene r, λ, α, β og $\sigma^2 = \mathrm{Var}\{a_t\}$. De resterende parametrene som inngår i bestemmelsen av filtret er autokovariansfunksjonen for utvalgsfeilen. La s_{11}^2 og s_{12}^2 betegne utvalgsvariansene mellom og innen utvalgsområdene for den nye utvalgsplanen, og la s_{21}^2 og s_{22}^2 være de tilsvarende størrelser for den gamle utvalgsplanen. s_j^2 betegner var$\{v_{jt}\}$ når vi bare har én utvalgsplan. Vi har ikke beregnet R_j for alle rimelige variasjoner av parametrene men bare sett på noen, til dels ekstreme avvik, fra de sanne verdier for å få et grovt bilde av hvor store og hvilke typer feil som er av betydning. I tabell 8, 9, 10 og 11 har vi bare tatt hensyn til utvalgsfeilen innen utvalgsområdene dvs. vi har behandlet situasjonen hvor utvalgsområdene holdes faste og hvor autokorrelasjonen for utvalgsfeilen mellom utvalgsområdene er lik 1. La $\underset{\sim}{\rho}_2 = (1, \rho_{12}, \rho_{22}, \ldots)$ være vektoren som definerer autokorrelasjonsfunksjonen for utvalgsfeilen innen utvalgsområdene. I tabellene lar vi $\underset{\sim}{\rho}_2$ være lik henholdsvis $\underset{\sim}{\rho}_2'$, $\underset{\sim}{\rho}_2''$ og $\underset{\sim}{\rho}_2'''$ som er definerte i tabell 6.

I denne tabellen har vi for en gitt modell for x_t sett på hvordan feil i korrelasjonsstrukturen for utvalgsfeilen påvirker "utsagnskraften" for filtrene. "Korrelasjonsvektoren" $\underset{\sim}{\rho}_2'$ ligger nær den virkelige "korrelasjonsvektoren" for mange av variablene som estimeres i AKU. Korrelasjonsvektoren $\underset{\sim}{\rho}_2''$ representerer korrelasjonsstrukturen for utvalgsfeilen ved sammensatt estimering når $\underset{\sim}{\rho}_2'$ er korrelasjonsvektoren for den "opprinnelige" utvalgsfeilen. For enkelte variable i AKU er utvalgsfeilen tilnærmet ukorrelert og korrelasjonsvektoren i denne situasjonen har vi betegnet $\underset{\sim}{\rho}_2'''$.

Anta nå at $\underset{\sim}{\rho}_2'''$ er den sanne parametervektor men at $\underset{\sim}{\rho}_2'$ brukes i stedet. Da viser tabellen at vi får til dels betydelig presisjonstap, spesielt gjelder dette for estimering av årlige endringer. Dersom $\underset{\sim}{\rho}_2'$ er sann verdi og $\underset{\sim}{\rho}_2'''$ anvendt verdi får vi noe presisjonstap ved estimering av nivåer, mens vi derimot får øket presisjon ved estimering av endringstall.

For $\underset{\sim}{\rho}_2'''$ og $\underset{\sim}{\rho}_2''$ som henholdsvis sann og anvendt verdi er det til dels betydelig presisjonstap men dette presisjonstapet reduseres vesentlig når $\underset{\sim}{\rho}_2'''$ byttes ut med $\underset{\sim}{\rho}_2'$.

Fra tabell 7 og 8 konkluderer vi at feil i parametrene for restleddet e_t har liten betydning. Tabellene viser for øvrig at for estimering av endringer vil en få presisjonsgevinst dersom σ^2/s_2^2 underestimeres mens det motsatte er tilfelle når σ^2/s_2^2 overestimeres.

Tabell 9 viser at feil i den autoregressive parameteren r har stor betydning mens feil i parameteren λ har uvesentlig betydning dersom λ ikke er stor. Er λ nær 1 betyr det at $\{\Delta_1 \Delta_4 x_t\}$ er "nesten" ikkestasjonær. Vi ser av tabellen at vi i denne situasjonen får presisjonstap for estimering av nivå mens vi får presisjonsgevinst ved estimering av endringer.

Vi har også gjort noen få robusthetsstudier (Tabell 10) når vi tar hensyn til utvalgsfeilen mellom utvalgsområdene og til endringen av utvalgsplanen. Tabellen antyder at filteret er robust selv når parametrene avviker betydelig fra de sanne verdiene.

Tabell 6. Robusthet overfor feil i parametrene for $\{v_{2t}\}$ når r = 1, λ = 0 og β = 0,5, $\underset{\sim}{\rho}_2' = (1, 0.43, 0, 0.2, 0.4, 0.19)$, $\underset{\sim}{\rho}_2'' = (1, 0.51, 0.18, 0.23, 0.47, 0.23, 0.06, 0.08, 0.1, 0.03)$, $\underset{\sim}{\rho}_2''' = (1, 0, \ldots)$

Tidsrekkenes lengde, år	Sanne verdier σ^2/s_2^2	Sanne verdier $\underset{\sim}{\rho}_2$	Anvendte verdier σ^2/s_2^2	Anvendte verdier $\underset{\sim}{\rho}_2$	R_0	R_1	R_2	R_3	R_4
3	0,02	$\underset{\sim}{\rho}_2'''$	0,02	$\underset{\sim}{\rho}_2'$	0,11	0,15	0,11	0,11	0,09
4	"	"	"	"	0,08	0,10	0,09	0,14	0,18
5	"	"	"	"	0,07	0,09	0,08	0,11	0,13
6	"	"	"	"	0,06	0,06	0,07	0,10	0,14
3	"	$\underset{\sim}{\rho}_2'$	"	$\underset{\sim}{\rho}_2'''$	0,05	-0,02	-0,01	-0,02	0,03
4	"	"	"	"	0,05	-0,02	-0,01	-0,03	-0,01
5	"	"	"	"	0,04	-0,02	-0,01	-0,04	-0,03
6	"	"	"	"	0,04	-0,02	-0,01	-0,03	-0,04
3	0,03	ρ_2''	0,02	ρ_2'	0,01	0,01	-0,01	-0,01	-0,04
4	"	"	"	"	0,01	0	-0,02	0,01	0,02
5	"	"	"	"	0,01	0	-0,01	-0,01	-0,02
6	"	"	"	"	0,01	0	-0,02	-0,01	-0,02
3	"	ρ_2''	"	ρ_2'''	0,07	0	-0,02	-0,04	0
4	"	"	"	"	0,08	-0,01	-0,02	-0,03	0
5	"	"	"	"	0,07	-0,02	-0,02	-0,04	-0,04
6	"	"	"	"	0,06	-0,01	-0,02	-0,04	-0,06
3	0,05	$\underset{\sim}{\rho}_2'''$	0,05	$\underset{\sim}{\rho}_2''$	0,25	0,40	0,29	0,11	0,30
4	"	"	"	"	0,16	0,24	0,15	0,12	0,32
5	"	"	"	"	0,18	0,28	0,21	0,11	0,34
6	"	"	"	"	0,17	0,27	0,22	0,10	0,33
3	"	$\underset{\sim}{\rho}_2'$	"	"	0,02	0,08	0,04	0,01	0,09
4	"	"	"	"	0,02	0,07	0,04	0,01	0,04
5	"	"	"	"	0,02	0,06	0,04	0,01	0,08
6	"	"	"	"	0,03	0,08	0,05	0,01	0,08

Tabell 7. Robusthet overfor feil i parameteren σ^2/s_2^2 når r = 1, $\alpha = \lambda = 0$, β = 0.5, $\underset{\sim}{\rho}_2 = \underset{\sim}{\rho}_2'$

Tidsrekkens lengde, år	σ^2/s_2^2 Sann verdi	σ^2/s_2^2 Anvendt verdi	R_0	R_1	R_2	R_3	R_4
3	0,001	0,1	0,004	0,02	0,02	0,03	0,05
4	"	"	0,01	0,03	0,02	0,03	0,05
5	"	"	0,02	0,04	0,04	0,05	0,08
6	"	"	0,04	0,04	0,05	0,06	0,09
3	0,1	0,001	0,005	-0,02	-0,02	-0,02	-0,03
4	"	"	0,02	-0,03	-0,02	-0,02	-0,03
5	"	"	0,04	-0,03	-0,03	-0,03	-0,03
6	"	"	0,08	-0,04	-0,03	-0,04	-0,04
3	0,01	0,1	0,003	0,02	0,03	0,03	0,04
4	"	"	0,01	0,03	0,02	0,03	0,04
5	"	"	0,02	0,04	0,03	0,04	0,07
6	"	"	0,03	0,04	0,04	0,05	0,08
3	0,1	0,01	0,004	-0,02	-0,02	-0,02	-0,03
4	"	"	0,01	-0,02	-0,02	-0,02	-0,03
5	"	"	0,03	-0,03	-0,02	-0,03	-0,03
6	"	"	0,06	-0,04	-0,03	-0,04	-0,05

Tabell 8. Robusthet overfor feil i parametrene for $\{e_t\}$ når $r = 1$, $\lambda = 0$, $\underset{\sim}{\rho}_2 = \underset{\sim}{\rho}_2'$

Tidsrekkens lengde, år	Sanne verdier σ^2/s_2^2	α	β	Anvendte verdier σ^2/s_2^2	α	β	R_0	R_1	R_2	R_3	R_4
3	0,08	0	0,9	0,08	0	0	0	0	0	0	0
4	"	"	"	"	"	"	0	0	0,01	0,02	0,03
5	"	"	"	"	"	"	0,01	0,02	0,02	0,03	0,04
6	"	"	"	"	"	"	0,03	0,04	0,03	0,05	0,08
3	0,01	"	0	0,05	0	0,9	0	0,01	0,01	0,01	0,03
4	"	"	"	"	"	"	0	0,01	0,01	0,01	0,01
5	"	"	"	"	"	"	0	0,01	0,01	0,01	0,02
6	"	"	"	"	"	"	0	0,01	0,01	0,01	0,02
3	0,05	0,5	0,8	0,05	0	0,5	0	0	0,01	0,01	0,01
4	"	"	"	"	"	"	0	0,01	0,01	0,01	0,02
5	"	"	"	"	"	"	0	0,01	0,01	0,02	0,02
6	"	"	"	"	"	"	0	0,01	0,01	0,02	0,03

Tabell 9. Robusthet overfor feil i parametrene λ, r og σ^2/s_2^2 når $\beta = 0.5$, $\alpha = 0$, og $\underset{\sim}{\rho}_2 = \underset{\sim}{\rho}_2'$

Tidsrekkens lengde, år	Sanne verdier σ^2/s_2^2	λ	r	Anvendte verdier σ^2/s_2^2	λ	r	R_0	R_1	R_2	R_3	R_4
3	0,02	0	0,7	0,1	0	1	0,26	0,06	0,06	0,16	0,36
4	"	"	"	"	"	"	0,25	0,06	0,06	0,15	0,33
5	"	"	"	"	"	"	0,25	0,06	0,07	0,13	0,30
6	"	"	"	"	"	"	0,25	0,06	0,07	0,13	0,32
3	"	"	"	0,02	"	"	0,25	0,05	0,07	0,13	0,32
4	"	"	"	"	"	"	0,24	0,04	0,05	0,12	0,26
5	"	"	"	"	"	"	0,23	0,04	0,04	0,09	0,21
6	"	"	"	"	"	"	0,20	0,03	0,04	0,08	0,18
3	"	"	0,8	"	"	"	0,24	0,04	0,07	0,13	0,32
4	"	"	"	"	"	"	0,23	0,04	0,04	0,11	0,26
5	"	"	"	"	"	"	0,22	0,04	0,04	0,09	0,21
6	"	"	"	"	"	"	0,18	0,03	0,03	0,08	0,15
3	0,05	0,5	1	0,05	"	1	0	0	0	0,01	0,01
4	"	"	"	"	"	"	0	-0,01	-0,01	-0,02	-0,02
5	"	"	"	"	"	"	0,01	-0,01	-0,01	-0,02	-0,03
6	"	"	"	"	"	"	0,02	-0,02	-0,02	-0,03	-0,05
3	0,05	0,9	1	0,05	0	1	0	-0,01	-0,01	-0,01	-0,03
4	"	"	"	"	"	"	0,02	-0,02	-0,02	-0,04	-0,06
5	"	"	"	"	"	"	0,07	-0,03	-0,03	-0,05	-0,09
6	"	"	"	"	"	"	0,15	-0,03	-0,04	-0,06	-0,11

Tabell 10. Robusthet overfor feil i parametrene s_{11}^2, s_{12}^2, s_{21}^2 og s_{22}^2 når σ^2/s_{12}^2 = 0,05, β=0,5, α = 0, $\underset{\sim}{\rho}_2 = \underset{\sim}{\rho}_2'$ og autokorrelasjonen mellom utvalgsområdene er lik 1.

Tidsrekkens lengde, år	Sanne verdier s_{11}^2	s_{12}^2	s_{21}^2	s_{22}^2	Anvendte verdier s_{11}^2	s_{12}^2	s_{21}^2	s_{22}^2	R_0	R_1	R_2	R_3	R_4
5	1	1	1	1	0,3	1	0,3	1	0,06	-0,01	-0,01	-0,02	-0,04
6	"	"	"	"	"	"	"	"	0,05	-0,01	-0,01	-0,03	-0,06
5	0,3	1	0,3	1	1	1	1	1	0,05	0,01	0,01	0,02	0,04
6	"	"	"	"	"	"	"	"	0,05	0,01	0,02	0,03	0,06
5	0,1	1	0,3	1	0,3	1	1	1	0,01	0,01	0,01	0,01	0,02
6	"	"	"	"	"	"	"	"	0,01	0,01	0,01	0,01	0,02
5	0,3	1	1	1	0,1	1	0,3	1	0,01	0	0	-0,01	-0,01
6	"	"	"	"	"	"	"	"	0,02	-0,01	-0,01	-0,01	-0,02

7. Filtrering og prediksjon for Arbeidskraftvariable

For beregning av de optimale filtre trenger vi estimater for parametrene i modellen for $\{x_t\}$. Sysselsettingstatistikken (fra 1970 og bakover) gir imidlertid ikke opplysninger om selvstendig næringsdrivende, sysselsatte fordelt etter alder og andre variable som vi er interessert i å estimere i AKU. Tidsrekken av AKU-estimater er foreløpig for kort til å brukes til modellestimering. Fra robusthetstudiene ovenfor kan vi imidlertid slutte at vi kan bruke svært upresise anslag for modellparametrene og fremdeles få en vesentlig presisjonsgevinst ved metoden.

Videre er det rimelig å tro at det normalt er en viss "treghet" i bevegelsene for de fleste arbeidskraftvariable slik at σ^2 ikke er svært stor.

Tabell 11. Filterkoeffisienter 1975-1978

Koeffisient-indeks	1975 1.kv.	1975 2.kv.	1975 3.kv.	1975 4.kv.	1976 1.kv.	1976 2.kv.	1976 3.kv.	1976 4.kv.
0	0,368	0,395	0,438	0,492	0,446	0,470	0,503	0,506
1	0,163	0,291	0,347	0,400	0,441	0,357	0,390	0,425
2	0,151	0,175	0,242	0,304	0,296	0,232	0,220	0,219
3	0,051	0,113	0,136	0,212	0,255	0,194	0,162	0,187
4	0,198	0,230	0,346	0,406	0,567	0,486	0,533	0,569
5	0,099	0,130	0,188	0,309	0,377	0,453	0,406	0,483
6	0,032	0,018	0,103	0,164	0,279	0,265	0,332	0,365
7		-0,014	0,012	0,101	0,160	0,194	0,199	0,283
8			0,163	0,213	0,289	0,264	0,377	0,423
9				0,161	0,182	0,204	0,210	0,336
10					0,061	0,052	0,144	0,187
11						0	0,038	0,130
12							0,161	0,207
13								0,157

Tabell 11 (forts.). Filterkoeffisienter 1975-1978

Koeffisient-indeks	1977 1.kv.	2.kv.	3.kv.	4.kv.	1978 1.kv.	2.kv.	3.kv.	4.kv.
0	0,460	0,469	0,494	0,500	0,471	0,497	0,518	0,524
1	0,397	0,367	0,380	0,408	0,404	0,376	0,403	0,423
2	0,191	0,195	0,207	0,217	0,209	0,205	0,208	0,221
3	0,137	0,115	0,142	0,153	0,151	0,129	0,137	0,145
4	0,504	0,527	0,578	0,597	0,544	0,562	0,591	0,603
5	0,502	0,414	0,466	0,514	0,500	0,455	0,484	0,517
6	0,347	0,293	0,296	0,321	0,295	0,304	0,318	0,335
7	0,310	0,252	0,237	0,251	0,239	0,209	0,245	0,260
8	0,526	0,458	0,517	0,529	0,440	0,489	0,544	0,568
9	0,379	0,422	0,389	0,457	0,432	0,328	0,403	0,455
10	0,275	0,262	0,322	0,323	0,283	0,204	0,206	0,240
11	0,175	0,200	0,211	0,278	0,248	0,173	0,148	0,164
12	0,255	0,241	0,344	0,371	0,432	0,348	0,394	0,405
13	0,172	0,188	0,196	0,303	0,313	0,324	0,284	0,345
14	0,062	0,059	0,139	0,162	0,218	0,189	0,228	0,228
15		0,013	0,050	0,128	0,139	0,139	0,139	0,190
16			0,142	0,179	0,210	0,183	0,261	0,282
17				0,137	0,144	0,144	0,142	0,227
18					0,045	0,037	0,098	0,114
19						0	0,027	0,088
20							0,111	0,141
21								0,108

Tabell 12. Sysselsatte etter hovednæring. 1000 Råestimater y_t og filterestimater $\hat{x}_t$

	Jordbruk, skogbruk, fiske og fangst	Industri og bergverk	Kraft og vann-fors.	Bygge- og anleggsv.	Vare-handel, hotell og rest.	Transport, lagring, post og telekom.	Bank og finan-sierings-virk.	Off., sos. og privat tjeneste-yting	Total
1.kv. 1975 y_t	166	415	18	141	255	155	69	444	1663
1. " " $\hat{x}_t$	164	407	18	142	267	157	72	436	1663
2. " " y_t	176	423	17	149	257	161	72	456	1711
2. " " $\hat{x}_t$	171	420	17	150	267	161	72	437	1695
3. " " y_t	184	422	19	146	264	158	76	430	1699
3. " " $\hat{x}_t$	178	421	18	152	265	163	75	421	1693
4. " " y_t	163	407	16	153	269	150	75	465	1698
4. " " $\hat{x}_t$	161	409	18	152	266	157	76	458	1698
1. " 1976 y_t	171	404	20	146	290	153	77	469	1730
1. " " $\hat{x}_t$	164	405	19	144	284	155	78	462	1711
2. " " y_t	176	421	20	148	295	160	73	479	1772
2. " " $\hat{x}_t$	170	419	19	149	290	159	75	464	1745
3. " " y_t	180	432	17	145	271	166	83	475	1769
3. " " $\hat{x}_t$	176	424	18	151	278	164	80	457	1746
4. " " y_t	169	424	18	150	303	158	83	500	1805
4. " " $\hat{x}_t$	162	418	19	151	293	159	81	490	1774

Tabell 13. Personer i arbeidsstyrken etter kjønn og alder. 1000. Råestimater y_t og filterestimater $\hat{x}_t$

	16-29 år		30-49 år		50-74 år		I alt	
	M	K	M	K	M	K	M	K
1. kv. 1975 y_t	276	190	429	241	361	193	1066	624
1. " " $\hat{x}_t$	279	192	430	238	356	193	1066	623
2. " " y_t	301	214	429	252	359	197	1089	663
2. " " $\hat{x}_t$	297	207	428	244	359	191	1084	642
3. " " y_t	305	214	427	248	354	198	1086	660
3. " " $\hat{x}_t$	309	208	428	242	355	191	1092	641
4. " " y_t	304	211	428	257	345	200	1077	668
4. " " $\hat{x}_t$	303	210	427	254	344	196	1072	660
1. " 1976 y_t	278	201	442	265	361	216	1081	682
1. " " $\hat{x}_t$	279	200	437	256	356	208	1072	664
2. " " y_t	303	223	443	266	364	208	1110	697
2. " " $\hat{x}_t$	300	217	439	259	361	202	1100	678
3. " " y_t	305	217	445	271	355	212	1105	700
3. " " $\hat{x}_t$	310	215	441	262	355	205	1104	682
4. " " y_t	303	226	444	286	352	222	1099	734
4. " " $\hat{x}_t$	302	222	401	276	348	212	1091	710

Tabell 14. Relativ reduksjon i forventet kvadratavvik

	σ^2/s_2^2	α	β	$\underset{\sim}{\rho}_2$	r	λ	M_0	M_1	M_2	M_3	M_4
3	0,02	0	0,5	$\underset{\sim}{\rho}_2'$	1	0	0,27	0,50	0,49	0,48	0,48
4	"	"	"	"	"	"	0,36	0,60	0,59	0,59	0,60
5	"	"	"	"	"	"	0,43	0,66	0,65	0,65	0,65
6	"	"	"	"			0,48	0,70	0,70	0,70	0,70
3	0,03	"	"	ρ_2''	"	"	0,20	0,36	0,43	0,43	0,39
4	"	"	"	"	"	"	0,27	0,45	0,53	0,54	0,48
5	"	"	"	"	"	"	0,33	0,50	0,60	0,60	0,56
6	"	"	"	"	"	"	0,38	0,55	0,65	0,65	0,61
3	0,02	"	"	$\underset{\sim}{\rho}_2'''$	"	"	0,53	0,67	0,65	0,58	0,68
4	"	"	"	"	"	"	0,61	0,75	0,73	0,68	0,77
5	"	"	"	"	"	"	0,67	0,79	0,78	0,74	0,84
6	"	"	"	"	"	"	0,70	0,82	0,81	0,78	0,87
3	0,001	0	0,5	$\underset{\sim}{\rho}_2'$	"	"	0,28	0,50	0,46	0,45	0,50
4	"	"	"	"	"	"	0,38	0,61	0,60	0,60	0,59
5	"	"	"	"	"	"	0,45	0,68	0,67	0,68	0,68
6	"	"	"	"	"	"	0,51	0,72	0,72	0,73	0,73
3	0,05	"	"	"	"	"	0,26	0,48	0,49	0,48	0,45
4	"	"	"	"	"	"	0,35	0,58	0,58	0,57	0,54
5	"	"	"	"	"	"	0,40	0,64	0,64	0,63	0,62
6	"	"	"	"	"	"	0,44	0,68	0,68	0,67	0,66
3	0,1	0	0,5	$\underset{\sim}{\rho}_2'$	1	0	0,25	0,45	0,46	0,44	0,43
4	"	"	"	"	"	"	0,33	0,56	0,55	0,53	0,50
5	"	"	"	"	"	"	0,37	0,61	0,61	0,59	0,56
6	"	"	"	"	"	"	0,40	0,65	0,64	0,63	0,59
3	0,02	0	0,5	"	0,7	0	0,59	0,54	0,54	0,61	0,88
4	"	"	"	"	"	"	0,61	0,63	0,63	0,71	0,89
5	"	"	"	"	"	"	0,67	0,69	0,69	0,75	0,92
6	"	"	"	"	"	"	0,69	0,73	0,73	0,79	0,93
3	0,05	0	0,5	"	1	0,9	0,25	0,47	0,46	0,44	0,43
4	"	"	"	"	"	"	0,30	0,54	0,53	0,51	0,48
5	"	"	"	"	"	"	0,32	0,57	0,56	0,54	0,52
6	"	"	"	"	"	"	0,33	0,59	0,58	0,55	0,53

Tabell 15. Total relativ reduksjon i forventet kvadratavvik når $\alpha = 0$, $\beta = 0{,}5$, $r = 1$, $\lambda = 0$ $\sigma^2/s_{12}^2 = 0{,}05$ og $\underset{\sim}{\rho}_2 = \underset{\sim}{\rho}_2'$

Tidsrekkens lengde	s_{11}^2	s_{12}^2	s_{21}^2	s_{22}^2	M_0	M_1	M_2	M_3	M_4
5	1	1	1	1	0,31	0,65	0,64	0,61	0,57
6	"	"	"	"	0,33	0,68	0,67	0,63	0,58
5	0,3	1	0,3	1	0,32	0,66	0,65	0,63	0,62
6	"	"	"	"	0,34	0,69	0,68	0,66	0,63
5	0,3	1	1	1	0,31	0,67	0,65	0,63	0,63
6	"	"	"	"	0,31	0,69	0,68	0,66	0,64
5	0,1	1	0,3	1	0,37	0,67	0,66	0,64	0,64
6	"	"	"	"	0,39	0,70	0,69	0,67	0,67

Tabell 16. Relativ reduksjon i forventet kvadratavvik når en foretar sammensatt estimering først. $\alpha = 0$, $\beta = 0{,}5$, $r = 1$, $\lambda = 0$ og $\sigma^2/s_2^2 = 0{,}02$

Tidsrekkens lengde	ρ	Sammensatt estimering			Total variansreduksjon		
		Nivåtall	Kvartalsendring	Årlig endring	M_0	M_1	M_4
3	$\underset{\sim}{\rho}$	0,75	0,87	0,82	0,27	0,41	0,48
4	"	"	"	"	0,36	0,51	0,58
5	"	"	"	"	0,44	0,58	0,68
6	"	"	"	"	0,51	0,63	0,74

Vi har valgt å benytte følgende parametre til beregning av filterkoeffisientene: $\alpha = 0$, $\beta = 0{,}5, r = 1$ $\lambda = 0$, $\sigma^2/s_{12}^2 = 0{,}05$, $s_{11}^2 = 0{,}2s_{12}^2$ $s_{21}^2 = 0{,}5s_{22}^2$, $1{,}2s_{12}^2 = s_{22}^2$, $\underset{\sim}{\rho}_2 = (1,\ 0.43, 0, 0.2,\ 0.4,\ 0.19)$ og autokorrelasjonsfunksjonen mellom utvalgsområdene konstant lik 1 for den gamle og nye utvalgsplanen. De beregnede filterkoeffisienter er gjengitt i tabell 11. Tabell 12 og 13 gir estimater for henholdsvis sysselsatte eller næring og personer i arbeidstyrken etter kjønn og alder beregnet ved hjelp av tabell 11. Denne estimeringen foregår ved å beregne

$$\hat{v}_t\ (0) = \sum_{j=0}^{n} \nu_j\ (y_{t-j} - y_{t-j-1} - y_{t-j-4} + y_{t-j-5})$$

for hvert kvartal t hvor ν_j er filterkoeffisientene.

$$\hat{x}_t\ (0) = y_t - \hat{v}_t\ (0)$$

er nå det nye "filterestimatet" for x_t. I beregningene har vi brukt AKU-estimatene fra og med 2. kvartal 1972 til og med 4. kvartal 1976.

Disse filterestimatene avviker en del fra de offisielle estimatene for næringene Varehandel, hotell og restaurant, Off.,sos. og privat tjenesteyting, Transport, lagring m.v. og samlet antall sysselsatte. For de andre næringene er det liten forskjell i de to estimatene. Ser vi på samlet antall sysselsatte ser vi at filterestimatene ligger systematisk lavere enn AKU-estimatene. Dette er en følge av at filtermetoden har en utjevningseffekt. Den samme tendensen finner vi også for kvinner i arbeidsstyrken etter alder.

Tabell 14 gir presisjonsgevinsten ved metoden for ulike verdier av parametrene. Her har vi ikke tatt hensyn til utvalgsfeilen mellom utvalgsområdene slik at tabellen bare gir relativ reduksjon av variansen innen utvalgsområdene. Vi ser at for en tidsrekke av observasjoner på 6 års lengde ligger variansreduksjonen på mellom 40 og 70% for nivåtall når $\lambda = 0$. Når $\lambda = 0.9$, dvs. at prosessen $\left\{\Delta_1\ \Delta_4\ x_t\right\}$ er nær ikke-stasjonæritet, har vi fremdeles en variansreduksjon på ca. 30%. Størst reduksjon får vi når $\underset{\sim}{\rho}_2 = \underset{\sim}{\rho}_2'''$, dvs. når utvalgsfeilen ikke er autokorrelert og når $r < 1$ som betyr at $\{\Delta_4 x_t\}$ er stasjonær. For endringstall er variansreduksjonen betydelig, mellom 60 og 90% (når $\lambda = 0$) og størst når $r < 1$ og $\underset{\sim}{\rho}_2 = \underset{\sim}{\rho}_2'''$. Når det gjelder presisjonsgevinst ved endringstallestimering kan tabellen gi et noe misvisende inntrykk. Forholdet er jo at $\mathrm{Var}\{v_t - v_{t-j}\}$, $j = 1, 2, ..$ er betydelig større når autokorrelasjonen for utvalgsfeilen (innen utvalgsområdene) er null enn når denne er positiv. De høye verdiene av M_j når $\underset{\sim}{\rho}_2 = \underset{\sim}{\rho}_2'''$ skyldes derfor tildels at utvalgsvariansen til "råestimatet" $y_t - y_{t-j}$ har økt.

I beregningene som ligger til grunn for tabell 15 har vi tatt hensyn til Byråets totrinns utvalgsplan og overgangen til ny utvalgsplan i 1975. Den totale variansreduksjon for nivåtallsestimering ligger på mellom 30 og 40% og for estimering av endringer er reduksjonen stort sett av samme størrelsesorden som de tilsvarende verdier i tabell 14.

Tabell 16 gir presisjonsgevinsten når vi foretar sammensatt estimering først og deretter bruker filtermetoden på de sammensatte estimater. (Her har vi ikke tatt hensyn til utvalgsfeilen mellom utvalgsområdene). Dersom tidsrekken har en lengde på minst tre år tyder tabellen på at vi får ingen ekstragevinst ved å kombinere disse metodene. For kortere tidsrekker derimot får en liten presisjonsgevinst ved å bruke filtermetoden. Da vil det være aktuelt å bruke sammensatt estimering.

Vi har også beregnet prediktorer og prediksjonsintervall for 1, 2, 3 og 4 kvartaler fremover. I disse beregningene er det ikke tatt hensyn til endringen av utvalgsplanen og de er derfor ikke gjengitt her. Programmet som beregner koeffisientene i prediktorene vil imidlertid bli modifisert slik at det tar hensyn til denne endringen.

Beregningene tyder på at prognosen opptil to kvartaler framover har omtrent samme konfidensintervall som dagens råestimater. Prognosen for mer enn fire kvartaler framover har svært store konfidensintervall og har følgelig liten verdi.

8. Den matematiske teori for prediksjon av en tidsrekke ved hjelp av et endelig antall observasjoner av en annen

A. Problemstilling og definisjoner

I (I) utledet vi estimatorer (filtre), prediktorer og deres forventede kvadratavvik basert på et uendelig antall observasjoner $y_t, y_{t-1}, \ldots$. Ved å anvende kompleks funksjonsteori oppnådde vi formler på en forholdsvis kompakt form som delvis kan beregnes numerisk for hånd enskjønt dette er en tidkrevende operasjon. I dette kapitlet skal vi utlede tilsvarende resultater når vi bare har et endelig sett observasjoner $y_t, y_{t-1}, .., y_{t-m}$. I motsetning til (I) skal vi nå vise hvordan estimatorer, prediktorer og deres forventede kvadratavvik kan utledes ved hjelp av den tradisjonelle teorien for minste kvadraters metode. De formlene som derved oppnås er uttrykt ved matriser som er velegnede for programmering. Noen av disse matrisene er egentlig uendelig-dimensjonale men de kan tilnærmes ved endeligdimensjonale matriser med dimensjoner som ikke går utover datamaskinens kapasitet. Vårt problem er altså å predikere $x_{t+\ell}$, $\ell \geq 0$, på grunnlag av $y_t, y_{t-1}, \ldots, y_{t-m}$, der

(8.1) $\quad y_t = x_t + v_t$

(I dette kapitlet vil vi for enkelthets skyld benytte betegnelsen prediktor også når $\ell = 0$). Utvalgsfeilen v_t har forventning null og er ukorrelert med x_t. Vi antar at $\{v_t\}$ er stasjonær og vi vil tilpasse teorien til det tilfellet at det er brudd i autokorrelasjonsstrukturen til $\{v_t\}$. Dette er situasjonen i vår anvendelse fordi Byråets utvalgsplan ble endret i 1975.

I samsvar med definisjonene i (I) kapittel 6 innfører vi operatorene

$$\Delta_k = \Delta_k(B) = 1 - B^k$$

$$\Psi = \Psi(B) = (1 + \Psi_1 B + .. + \Psi_5 B^5)(1 + r(1-\delta_{1r})B) \;, \quad \delta_{1r} = \begin{cases} 1 & \text{når } r=-1 \\ 0 & \text{ellers} \end{cases}$$

$$\phi = \phi(B) = 1 + \phi_1 B + .. + \phi_5 B^5$$

$$\mu = \mu(B) = \Delta_4 (B) (1 + r \delta_{1r} B) , r \varepsilon [-1, 0]$$

hvor B er den baklengse skiftoperator, B $x_t = x_{t-1}$. Dersom B erstattes med en kompleks variabel z blir $\psi(z)$, $\phi(z)$ og $\mu(z)$ de karakteristiske polynomene til de respektive operatorene. $\phi(z)$ og $\psi(z)$ har røtter som ligger utenfor enhetsirkelen hvilket medfører at de respektive operatorene er invertible. Vi definerer autokovariansfunksjonen

$$g_{jw} = g_{jw}(t) = \text{cov}\{w_t, w_{t-j}\}, \; j = .. -1, 0, 1, ..$$

til en vilkårlig prosess $\{w_t\}$ og vi lar

(8.2) $\quad \underset{\sim}{G}_w = (g_{0w}, g_{1w}, g_{2w}, \ldots)^T$

og

(8.3) $\quad G_w (z) = \sum_{j=-\infty}^{\infty} g_{jw} z^j$

være den generende funksjonen til autokovariansfunksjonen for $\{w_t\}$. Her betyr T transponert. Vi antar i dette kapitlet at $\{x_t\}$ genereres av modellen

(8.4) $\quad \psi\mu x_t = \Delta_4 x_t + (r+\psi_1) \Delta_4 x_{t-1} + .. + (r\psi_4 + \psi_5) \Delta_4 x_{t-5}$

$$= \phi a_t = a_t + \phi_1 a_{t-1} + .. + \phi_5 a_{t-5}$$

hvor prosessen $\{a_t\}$ er en "hvit-støy" prosess, dvs. en ikke autokorrelert stasjonær prosess. a_t har forventning null og varians σ^2.

La videre

$$x'_t = \mu x_t, \; y'_t = \mu y_t \text{ og } v'_t = \mu v_t$$

Med denne notasjon får vi fra (8.4) at

$$\psi x'_t = \phi a_t$$

Siden ψ var antatt invertibel blir $\{x'_t\}$ stasjonær. Lar vi eksempelvis

$$\psi = 1 - \lambda B^4$$

og

$$\phi = (1 - \alpha B) (1 - \beta B^4) = 1 - \alpha B - \beta B^4 + \alpha\beta B^5$$

ser vi at modellene som ble studert i kapittel 3 er spesialtilfeller av (8.4). I følge sammenhengen mellom modelloperatorer og genererende funksjoner (se (I) side 13 og 14) har vi

(8.5) $\quad G_{x'} (z) = \dfrac{\phi (z) \; \phi (z^{-1})}{\psi (z) \; \psi (z^{-1})} \sigma^2$

B. Matrisefunksjoner

Vi skal her innføre en type matriser som viser seg å gi hensiktsmessige formler for prediktorene og deres forventede kvadratavvik.

Til en vilkårlig Laurentrekke (dvs. en potensrekke i z og i z^{-1})

$$h(z) = \sum_{j=-\infty}^{\infty} h_j z^j$$

definerer vi operatorene

$$[h(z)]_r^{(s)} = \sum_{j=r}^{s} h_j z^j \quad \text{og} \quad [h(z)]_+ = [h(z)]_0^{(\infty)}$$

Videre definerer vi vektoren $\underset{\sim}{h}$ tilhørende h(z) ved

$\underset{\sim}{h} = (h_0, h_1, \ldots)^T$, (uendeligdimensjonal),

operatorene P_k og Q_k ved

$P_k \underset{\sim}{h} = (h_0, h_1, \ldots h_k)^T$,

$Q_k \underset{\sim}{h} = (0, 0, \ldots, h_k, h_{k+1}, \ldots)^T$

og matrisene $M_{rk}(\underset{\sim}{h})$ og $S_{kr}(\underset{\sim}{h})$ ved

$$M_{rk} = \underbrace{\left.\begin{bmatrix} h_0 & & & & \\ h_1 & h_0 & & & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ h_r & & h_0 & 0 & \ldots \end{bmatrix}\right\} r+1}_{k+1}, \quad S_{kr}(\underset{\sim}{h}) = M_{rk}(\underset{\sim}{h})^T$$

Vi ser at disse matrisene er invertible dersom r = k og $h_0 \neq 0$. Dersom $\alpha(z)$, $\beta(z)$ og $\gamma(z)$ er potensrekker slik at

$$\alpha(z) = \beta(z)\,\gamma(z)$$

er det lett å verdifisere at

$$(8.6) \quad P_k \underset{\sim}{\alpha} = M_{kk}(\beta)\, P_k \underset{\sim}{\gamma} = M_{kk}(P_k \underset{\sim}{\beta})\, P_k \underset{\sim}{\gamma}$$

$$= M_{kk}(\underset{\sim}{\gamma})\, P_k \underset{\sim}{\beta} = M_{kk}(P_k \underset{\sim}{\gamma})\, P_k \underset{\sim}{\beta}.$$

Noen ganger vil vi operere med vektorer hvor bare en del av komponentene er definerte. Det er da underforstått at de udefinerte komponentene er lik null. For eksempel kan $P_k \underset{\sim}{h}$ bety

$(h_0, h_1, \ldots h_k, 0, 0, \ldots)$

enkelte steder. Setter vi

$$\delta_k(z) = \left[\, [\beta(z)]_0^{(k)} \, \gamma(z^{-1}) \,\right]_+$$

finner vi analogt til (8.6)

(8.7) $\quad \underset{\sim}{\delta}_k = S_{kk}(\underset{\sim}{\gamma}) \, P_k \underset{\sim}{\beta} = S_k(P_k \underset{\sim}{\gamma}) \, P_k \underset{\sim}{\beta}$.

La oss nå se hvordan vi kan bruke denne formalismen til å uttrykke $\underset{\sim}{G}_x{}'$. La

$$\gamma(z) = \phi(z)\, \psi^{-1}(z)$$

Fra (8.6) får vi

$$M_{kk}(\underset{\sim}{\psi}) \, P_k \underset{\sim}{\gamma} = P_k \underset{\sim}{\phi}$$

hvilket medfører at koeffisientene i $[\phi(z)\, \psi^{-1}(z)]_0^{(k)}$ er gitt ved vektoren

$$M_{kk}^{-1}(\underset{\sim}{\psi}) \, P_k \underset{\sim}{\phi} \, .$$

Bruker vi (8.7) ser vi at koeffisientene i rekkeutviklingen for

$$\left[\left[\frac{\phi(z)}{\psi(z)} \right]_0^{(k)} \frac{\phi(z^{-1})}{\psi(z^{-1})} \right]_+$$

kan uttrykkes ved vektoren

$$S_{kk}\left\{ M_{kk}^{-1}(\underset{\sim}{\psi}) \, P_k \underset{\sim}{\phi} \right\} M_{kk}^{-1}(\psi) \, P_k \underset{\sim}{\phi} \, .$$

Siden vi ifølge (8.5) har

$$[G_x{}'(z)]_+ = \sigma^2 \lim_{k\to\infty} \left[\left[\frac{\phi(z)}{\psi(z)} \right]_0^{(k)} \frac{\phi(z^{-1})}{\psi(z^{-1})} \right]_+$$

får vi

(8.8) $\quad \underset{\sim}{G}_x{}' = \sigma^2 \lim_{k\to\infty} S_{kk}\left\{ M_{kk}^{-1}(\underset{\sim}{\psi}) \, P_k \underset{\sim}{\phi} \right\} M_{kk}^{-1}(\underset{\sim}{\psi}) \, P_k \underset{\sim}{\phi}$.

C. Lineær prediksjon

Vi skal i dette avsnittet utlede det generelle formelapparat for den beste lineære prediktor

(8.9) $\quad \hat{x}_t(\ell) = \sum_{j\geq 0}^{m} \zeta_j(\ell) \, y_{t-j} \, , \quad \ell \geq 0,$

for $x_{t+\ell}$ ved tidspunkt t. Den beste prediktor betyr i vår forstand den prediktoren som minimerer forventet kvadratavvik

$$E\left\{ \hat{x}_t(\ell) - x_{t+\ell} \right\}^2 .$$

Vi skal se at vi med visse tillempninger kan benytte den tradisjonelle teorien for minste kvadraters metode. Siden $\{x_t\}$ ikke er stasjonær må vi modifisere minimeringsproblemet.

Dersom $\ell = 0$ er dette enkelt. Vi kan da innføre

$$\hat{v}_t(0) = y_t - \hat{x}_t(0)$$

som gir

$$E\left\{\hat{x}_t(0) - x_t\right\}^2 = E\left\{\hat{v}_t(0) - v_t\right\}^2.$$

Siden $\{v_t\}$ er stasjonær må $\hat{v}_t(0)$ følgelig være en funksjon av den stasjonære prosessen $\{y_t'\}$.*) La oss se på det generelle tilfellet. Vi har

$$x_{t+\ell} = (1-\mu)\, x_{t+\ell} + x_{t+\ell}' .$$

Nå vil $(1-\mu)\, x_{t+\ell}$ være en funksjon av prosessverdiene $x_{t+\ell-1}, x_{t+\ell-2}, \ldots,$ slik at $(1-\mu)^{\ell}\, x_{t+\ell}$ er en funksjon av prosessverdiene $x_t, x_{t-1}, \ldots,$. Fra likningen ovenfor får vi rekursivt

$$x_{t+\ell} = (1-\mu)^{\ell} x_{t+\ell} + \sum_{j=0}^{\ell-1} (1-\mu)^j x_{t+\ell}' .$$

Definerer vi $\hat{v}_t(\ell)$ ved

$$\hat{x}_t(\ell) = (1-\mu)^{\ell}\, y_{t+\ell} - \hat{v}_t(\ell) \tag{8.10}$$

der

$$\hat{v}_t(\ell) = \sum_{j=0}^{n} \nu_j(\ell)\, y_{t-j}' , \quad n = m - 4 - \delta_{1r}$$

får vi

$$E\left\{x_{t+\ell} - \hat{x}_t(\ell)\right\}^2 = E\left\{(1-\mu)^{\ell}\, x_{t+\ell} - (1-\mu)^{\ell}\, x_{t+\ell} - u_{t+\ell}(\ell) + \hat{v}_t(\ell)\right\}^2$$

$$= E\left\{\hat{v}_t(\ell) - u_{t+\ell}(\ell)\right\}^2$$

der

$$u_t(\ell) = (1-\mu)^{\ell} v_t - \sum_{j=0}^{\ell-1} (1-\mu)^j\, x_t' , \tag{8.11}$$

for $\ell \geq 1$. Både $\{v_t\}$ og $\{u_t\}$ er stasjonære prosesser slik at $\hat{v}_t(\ell)$ blir en funksjon av $\{y_t'\}$. Innfører vi vektornotasjonene

$$\underset{\sim}{\nu}(\ell) = (\nu_0(\ell), \nu_1(\ell), \ldots.)^T,$$

$$\underset{\sim}{y}_t = (y_t, y_{t-1}, \ldots)^T, \quad \underset{\sim}{x}_t = (x_t, x_{t-1}, \ldots)^T$$

og

$$\underset{\sim}{v}_t = (v_t, v_{t-1}, \ldots)^T$$

får vi

$$\hat{v}_t(\ell) = (P_n \underset{\sim}{\nu}(\ell))^T P_n \underset{\sim}{y}_t' = \underset{\sim}{\nu}(\ell)^T P_n \underset{\sim}{y}_t'$$

---

*) Dette trenger egentlig et bevis.

Med denne notasjonen får vi

$$E\left\{u_{t+\ell}(\ell) - \underset{\sim}{\nu}(\ell)^T P_n \underset{\sim}{y}_t'\right\}^2 = \mathrm{Var}\left\{u_t(\ell)\right\}$$

$$- 2\,\underset{\sim}{\nu}(\ell)^T E\left\{P_n \underset{\sim}{y}_t' u_{t+\ell}(\ell)\right\} + \underset{\sim}{\nu}(\ell)^T E\left\{P_n \underset{\sim}{y}_t' (P_n \underset{\sim}{y}_t')^T\right\} \underset{\sim}{\nu}(\ell).$$

Nå er

$$\sum_{j=0}^{n} \nu_j(\ell)\, v_{t-j}' = \sum_{i=0}^{m} \sum_{j=0}^{i} \mu_{i-j}\, \nu_j(\ell)\, v_{t-i}$$

hvilket medfører at

$$(8.12) \quad \underset{\sim}{\nu}(\ell)^T P_n \underset{\sim}{v}_t' = [\, M_{mn}(\underset{\sim}{\mu})\, \underset{\sim}{\nu}(\ell)\,]^T P_m \underset{\sim}{v}_t$$

$$= \underset{\sim}{\nu}(\ell)^T S_{nm}(\underset{\sim}{\mu})\, P_m \underset{\sim}{v}_t$$

Herav følger det at

$$\underset{\sim}{\nu}(\ell)^T E\left\{P_n \underset{\sim}{v}_t' (P_n \underset{\sim}{v}_t')^T\right\} \underset{\sim}{\nu}(\ell) = \underset{\sim}{\nu}(\ell)^T S_{nm}(\underset{\sim}{\mu})\; E\left\{P_m \underset{\sim}{v}_t (P_m \underset{\sim}{v}_t)^T\right\} M_{mn}(\underset{\sim}{\mu})\, \underset{\sim}{\nu}(\ell)$$

og

$$\underset{\sim}{\nu}(\ell)^T E\left\{P_n \underset{\sim}{y}_t' u_{t+\ell}(\ell)\right\} = \underset{\sim}{\nu}(\ell)^T E\left\{P_n \underset{\sim}{v}_t' u_{t+\ell}(\ell)\right\}$$

$$= \underset{\sim}{\nu}(\ell)^T S_{nm}(\underset{\sim}{\mu})\; E\left\{P_m \underset{\sim}{v}_t u_{t+\ell}(\ell)\right\}.$$

Lar vi

$$(8.13) \quad W_{22} = S_{nm}(\underset{\sim}{\mu})\, E\left\{P_m \underset{\sim}{v}_t P_m \underset{\sim}{v}_t^T\right\} M_{mn}(\underset{\sim}{\mu}) + E\left\{P_n \underset{\sim}{x}_t' P_n (\underset{\sim}{x}_t')^T\right\},$$

$$(8.14) \quad W_{12}(\ell) = S_{nm}(\underset{\sim}{\mu})\, E\left\{P_m \underset{\sim}{v}_t u_{t+\ell}(\ell)\right\} + E\left\{P_n \underset{\sim}{x}_t' u_{t+\ell}(\ell)\right\} \text{og } W_{21}(\ell) = W_{12}(\ell)^T$$

får vi etter en enkel omforming

$$E\left\{u_{t+\ell}(\ell) - \underset{\sim}{\nu}(\ell)^T P_n \underset{\sim}{y}_t'\right\} = \mathrm{Var}\left\{u_t(\ell)\right\} - W_{21}(\ell)\, W_{22}^{-1}\, W_{12}(\ell)$$

$$+ \left(\underset{\sim}{\nu}(\ell)^T - W_{21}(\ell)\, W_{22}^{-1}\right) W_{22} \left(\underset{\sim}{\nu}(\ell)^T - W_{21}(\ell)\, W_{22}^{-1}\right)^T.$$

Under antagelsen at W_{22} er positiv definit oppnår uttrykket ovenfor sitt entydige minimum når

$$(8.15) \quad \underset{\sim}{\nu}(\ell) = W_{22}^{-1}\, W_{12}(\ell)$$

med forventet kvadratavvik

$$(8.16) \quad \min\; E\left\{u_{t+\ell}(\ell) - \underset{\sim}{\nu}(\ell)^T \underset{\sim}{y}_t'\right\}^2 = \mathrm{Var}\left\{u_t(\ell)\right\} - W_{21}(\ell)\, W_{22}^{-1}\, W_{12}(\ell)$$

$$= \mathrm{var}\left\{u_t(\ell)\right\} - \underset{\sim}{\nu}(\ell)^T W_{12}(\ell).$$

Det gjenstår nå å finne $W_{12}(\ell)$ og W_{22}. Vi skal anta at ved tidspunkt t-k er det et brudd i korrelasjonsstrukturen for v_t dvs. at v_{t-i} og v_{t-j} er ukorrelerte når $i \leq k$ og $j > k$. La $G_v''(z)$ og $G_v'(z)$ være de respektive genererende funksjonene som identifiserer korrelasjonsstrukturen for $\{v_t\}$ før og etter tidspunkt t-k. Siden

$$\left[\sum_{i=-\infty}^{\infty} E\left\{v_{t-i}\,(1-\mu)^{\ell}\, v_{t+\ell}\right\} z^{i}\right]_{-\infty}^{(k)} = \left[z^{-\ell} \sum_{i=-\infty}^{\infty} E\left\{v_{t-i}\,(1-\mu)^{\ell} v_t\right\} z^{i}\right]_{-\infty}^{(k)}$$

$$= \left[z^{-\ell}\,(1-\mu\,(z)^{\ell})\; G_v'\,(z)\right]_{-\infty}^{(k)}$$

og koeffisientene i $(1-\mu\,(z))^{\ell}$ er gitt ved vektoren

$$-\,M_{rr}^{\ell-1}\,(-\,Q_1\,\underset{\sim}{\mu})\,Q_1\,P_r\,\underset{\sim}{\mu}$$

når r er stor nok, får vi fra (8.6) og (8.7) at koeffisientene i

$$(8.17)\quad \left[z^{-\ell}\,(1-\mu\,(z))^{\ell}\,G_v'\,(z)\right]_{0}^{(k)} = \left[z^{-\ell}\,(1-\mu\,(z))^{\ell}\,\left([G_v'\,(z)]_{+} + [G_v'\,(z^{-1})]_{1}^{(\ell)}\right)\right]_{0}^{(k)}$$

er gitt ved vektoren

$$-\,P_r\left[(S_{rr}\,(\underset{\sim}{L})\,M_{rr}\,(\underset{\sim}{G}_v') + S_{rr}\left\{M_{rr}\,(Q_1\,\underset{\sim}{G}_v')\,P_r\,\underset{\sim}{L}\right\})\,M_{rr}^{\ell-1}\,(-Q_1\,\underset{\sim}{\mu})\,Q_1\,P_r\,\underset{\sim}{\mu}\right]$$

når r er stor nok hvor $\underset{\sim}{L}$ er vektoren tilknyttet $L\,(z) = z^{\ell}$.

Tilsvarende finner vi at

$$\sum_{i=-\infty}^{\infty} E\left\{x_{t-i}'\,(1-\mu)^{j}\,x_{t+\ell}'\right\} z^{i} = z^{-\ell}\,(1-\mu(z))^{j}\,G_x'\,(z)$$

slik at koeffisientene i

$$(8.18)\quad \sum_{i=0}^{n} E\left\{x_{t-i}'\,(1-\mu)^{j}\,x_{t+\ell}'\right\} z^{i}$$

kan uttrykkes ved vektoren

$$-\lim_{r\to\infty} P_n\left[(S_{rr}\,(\underset{\sim}{L})\,M_{rr}\,(\underset{\sim}{G}_x') + S_{rr}\left\{M_{rr}\,(Q_1\,\underset{\sim}{G}_x')\,P_r\,\underset{\sim}{L}\right\})\,M_{rr}^{j-1}\,(-Q_1\,\underset{\sim}{\mu})\,Q_1\,P_r\,\underset{\sim}{\mu}\right].$$

Siden elementene i $E\left\{\underset{\sim}{v}_t\,u_{t+\ell}\,(\ell)\right\}$ og $E\left\{\underset{\sim}{x}_t'\,u_{t+\ell}\,(\ell)\right\}$ er gitt ved de respektive koeffisientene i (8.17) og (8.18) medfører utledningene ovenfor at for $\ell \geq 1$ er

$$(8.19a)\quad W_{12}\,(\ell) = -\,S_{nm}\,(\underset{\sim}{\mu})\,P_k\left[(S_{rr}\,(\underset{\sim}{L})\,M_{rr}\,(\underset{\sim}{G}_v') + S_{rr}\left\{M_{rr}\,(Q_1\,\underset{\sim}{G}_v')\,P_r\,\underset{\sim}{L}\right\})\,M_{rr}^{\ell-1}\,(-Q_1\,\underset{\sim}{\mu})\,Q_1\,P_r\,\underset{\sim}{\mu}\right]$$

$$-\lim_{j\to\infty} P_n\left[(S_{jj}\,(\underset{\sim}{L})\,M_{jj}\,(\underset{\sim}{G}_x') + S_{jj}\left\{M_{jj}\,(Q_1\,\underset{\sim}{G}_x')\,P_j\,\underset{\sim}{L}\right\}\underset{\sim}{A}_j\,(\ell)\right]$$

der $r \geq 5\ell + m$,

$$\underset{\sim}{A}_j\,(1) = P_j\,(1,\,0,\,0\,\ldots\;)^{T}$$

og

$$\underset{\sim}{A}_j(\ell) = \underset{\sim}{A}_j(1) - \sum_{i=0}^{\ell-2} M_{jj}^{i} (-Q_1 \underset{\sim}{\mu}) Q_1 P_j \underset{\sim}{\mu}$$

$$= M_{jj}^{-1}(\underset{\sim}{\mu}) \left[\underset{\sim}{A}_j(1) + M_{jj}^{\ell-1} (-Q_1 \underset{\sim}{\mu}) Q_1 P_j \underset{\sim}{\mu} \right], \ \ell > 1.$$

For $\ell = 0$ får vi

$$(8.19) \quad W_{12}(0) = S_{nm}(\underset{\sim}{\mu}) P_k \underset{\sim}{G}_v' .$$

La oss dernest utlede et uttrykk for W_{22}. Siden

$$E\left\{ \underset{\sim}{x}_t' (\underset{\sim}{x}_t')^T \right\}_{ij} = g_{x', |i-j|}$$

får vi umiddelbart at

$$(8.20) \quad E\left\{ P_n \underset{\sim}{x}_t' P_n (\underset{\sim}{x}_t')^T \right\} = M_{nn}(\underset{\sim}{G}_{x'}) + S_{nn}(Q_1 \underset{\sim}{G}_{x'}) .$$

Tilsvarende får vi at

$$E\left\{ P_m \underset{\sim}{v}_t P_m \underset{\sim}{v}_t^T \right\} = \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix} = V$$

der

$$(8.21a) \quad V_1 = M_{kk}(\underset{\sim}{G}_v') + S_{kk}(Q_1 \underset{\sim}{G}_v')$$

og

$$(8.21b) \quad V_2 = M_{pp}(\underset{\sim}{G}_v'') + S_{pp}(\underset{\sim}{Q}_1 \underset{\sim}{G}_v''), \ p = m - k .$$

Lar vi

$$\underset{\sim}{G}_v''' = \begin{pmatrix} 0 \\ \underset{\sim}{G}_v'' \end{pmatrix} \begin{matrix} \} \ k+1 \\ \end{matrix}$$

ser vi at V alternativt kan skrives som

$$(8.22) \quad V = M_{km}(\underset{\sim}{G}_v') + M_{mm}(\underset{\sim}{G}_v''') + S_{mk}(Q_1 \underset{\sim}{G}_v') + S_{mm}(Q_1 \underset{\sim}{G}_v''')$$

hvilket er en hensiktsmessig formel for programmering. Det gjenstår nå bare å beregne var $\left\{ u_t(\ell) \right\}$. Når $\ell = 0$ er $u_t = v_t$ som gir var $\left\{ u_t(0) \right\} = s^2$. For $\ell > 0$ kan $\underset{\sim}{G}_u$ uttrykkes ved matriseformler analoge til de vi har utledet ovenfor.

Dersom modellen for $\left\{ x_t \right\}$ har formen

$$\Delta_1 \Delta_4 x_t = a_t - \alpha a_{t-1} - \beta(a_{t-4} - \alpha a_{t-5})$$

og vi antar at $\rho_j = \text{korr}\left\{ v_t, v_{t-j} \right\} = 0$ for $j > 10$ finner vi ved å bruke (8.11) etter noe regning

$$(8.23a) \quad \text{Var}\left\{ u(1) \right\} = s^2 (3 - 2\rho_1 + 2\rho_3 - 2\rho_4) + \sigma^2 (1+\alpha^2)(1+\beta^2),$$

$$(8.23b) \quad \text{Var}\left\{ u(2) \right\} = s^2 (15 - 16\rho_1 - 2\rho_2 + 16\rho_3 - 16\rho_4 + 4\rho_5 + 2\rho_6 - 4\rho_7 + 2\rho_8)$$

$$+ 4\sigma^2 (1+\alpha^2)(1+\beta^2),$$

$$(8.23c)\quad Var\left\{u\ (3)\right\} = s^2\ (93 - 114\rho_1 - 18\rho_2 + 118\rho_3 - 120\rho_4 + 42\rho_5$$

$$+ 24\rho_6 - 48\rho_7 + 30\rho_8 - 4\rho_9\ ) + \sigma^2\ [\ 8\alpha\beta + 6\beta\ (\ 1+\alpha^2)$$

$$+ 6\alpha\ (\ 1+\beta^2\ ) + 15\ (\ 1+\alpha^2\ )\ (\ 1+\beta^2\ )\ ]\ ,$$

$$(8.23d)\quad Var\left\{u\ (4)\right\} = s^2\ (\ 631 - 808\rho_1 - 136\rho_2 + 864\rho_4$$

$$+ 352\rho_5 + 216\rho_6 - 448\rho_7 + 312\rho_8 - 56\rho_9)$$

$$+ \sigma^2\ [\ 84\ (\ 1+\alpha^2)\ (\ 1+\beta^2) + 92\alpha\ (\ 1+\beta^2) + 84\beta\ (\ 1+\alpha^2) + 129\alpha\beta\ ]\ .$$

Den følgende setningen oppsummerer resultatene ovenfor.

Setning 1: La $W_{12}\ (\ell)$, W_{22} og V være definert ved (8.13), (8.14) og (8.22). Koeffisientene i den optimale lineære prediktor

$$\hat{x}_t\ (\ell) = (1 - \mu)^\ell\ y_{t+\ell} - \hat{v}_t\ (\ell) = (1 - \mu)^\ell\ y_{t+\ell} - \underset{\sim}{\nu}\ (\ell)^T\ P_n\ \underset{\sim}{y}'_t$$

er gitt ved

$$\underset{\sim}{\nu}\ (\ell) = W_{22}^{-1}\ W_{12}\ (\ell)$$

og har forventet kvadratavvik

$$E\left\{\hat{x}_t(\ell) - x_{t+\ell}\right\}^2 = Var\left\{u_t\ (\ell)\right\} - \underset{\sim}{\nu}\ (\ell)^T\ W_{12}\ (\ell)$$

Korollar 2: For $\ell = 0$ er

$$\hat{x}_t\ (0) = y_t - \underset{\sim}{\nu}\ (0)^T\ P_n\ \underset{\sim}{v}'_t$$

der

$$\underset{\sim}{\nu}\ (0) = W_{22}^{-1}\ S_{nm}\ (\underset{\sim}{\mu})\ P_k\ \underset{\sim}{G}'_v\ .$$

Forventet kvadratavvik er gitt ved

$$E\left\{\hat{x}_t\ (0) - x_t\right\}^2 = s^2 - \underset{\sim}{\nu}\ (0)^T\ S_{nm}\ (\underset{\sim}{\mu})\ P_k\ \underset{\sim}{G}'_v$$

Dersom a_t og v_t er normalfordelte er

$$(8.24)\quad \hat{v}_t\ (\ell) = E\left\{u_{t+\ell}\ (\ell)\ |\ y_t, y_{t-\ell}, \ldots, y_{t-m}\right\}$$

og

$$(8.25)\quad E\left\{\hat{v}_t\ (\ell) - u_{t+\ell}\ (\ell)\right\}^2 = Var\left\{u_{t+\ell}\ (\ell)\ |\ y_t, y_{t-1}, \ldots, y_{t-m}\right\}.$$

Dette er velkjente resultater fra teorien for minste kvadraters metode.

D. Estimering av endringer

Estimering av endringen $x_t - x_{t-j}$ kan gjøres på flere måter men vi skal bare betrakte den naturlige estimatoren

$$\Delta_j \hat{x}_t (0) = \hat{x}_t (0) - \hat{x}_{t-j} (0) = \Delta_j y_t - \Delta_j \hat{v}_t (0)$$

og beregne forventet kvadratavvik for denne. Vi trenger da et uttrykk for kovariansmatrisen

$$E\left\{ P_m \underset{\sim}{v}_t \Delta_j v_t \right\} \qquad , \quad j > t-k .$$

Siden

$$\sum_{i=0}^{k} E\left\{ v_{t-i} \Delta_j v_t \right\} z^i = \left[\Delta_j (z) G_v' (z) \right]_0^{(k)}$$

får vi at

$$(8.26) \quad E\left\{ P_k \underset{\sim}{v}_t \Delta_j v_t \right\} = \left[M_{kk} (\underset{\sim}{G}_v') + S_{kk} (Q_1 \underset{\sim}{G}_v') \right] P_k \underset{\sim}{\Delta}_j .$$

La oss nå sette $\underset{\sim}{\nu} (0) = \underset{\sim}{\nu} (0, t)$ for å betegne at prediktoren er beregnet for tidspunkt t. Vi har da

$$\Delta_j \hat{v}_t (0) = \underset{\sim}{\nu} (0, t)^T P_n \underset{\sim}{y}_t' - \underset{\sim}{\nu} (0, t-j)^T P_n y_{t-j}'$$

$$= \left[\underset{\sim}{\nu} (0, t) + M_{nn} (Q_1 \underset{\sim}{\Delta}_j) \underset{\sim}{\nu} (0, t-j) \right]^T P_n \underset{\sim}{y}_t' .$$

Innfører vi

$$(8.27) \quad \underset{\sim}{\tau} = \underset{\sim}{\tau} (n, j) = \underset{\sim}{\nu} (0, t) + M_{nn} (Q_1 \underset{\sim}{\Delta}_j) \underset{\sim}{\nu} (0, t-j)$$

kan vi uttrykke forventet kvadratavvik for $\Delta_j \hat{v}_t (0)$ ved $\underset{\sim}{\tau}$.

$$E\left\{ \Delta_j v_t - \Delta_j \hat{v} (0) \right\}^2 = \mathrm{var}\left\{ \Delta_j v_t \right\} - 2 E\left\{ \Delta_j \hat{v}_t (0) \Delta_j v_t \right\} + E\left\{ \Delta_j \hat{v}_t (0) \right\}^2$$

$$= \mathrm{var}\left\{ \Delta_j v_t \right\} - 2 \left(M_{mn} (\underset{\sim}{\mu}) \underset{\sim}{\tau} \right)^T E\left\{ P_m \underset{\sim}{v}_t \Delta_j v_t \right\} + \underset{\sim}{\tau}^T E\left\{ P_n \underset{\sim}{x}_t' P_n (\underset{\sim}{x}_t')^T \right\} \underset{\sim}{\tau}$$

$$+ \left(M_{mn} (\underset{\sim}{\mu}) \underset{\sim}{\tau} \right)^T E\left\{ P_m \underset{\sim}{v}_t P_m \underset{\sim}{v}_t^T \right\} M_{mn} (\underset{\sim}{\mu}) \underset{\sim}{\tau} .$$

Herav følger setning 3.

Setning 3:

$$E\left\{ \Delta_j \hat{x}_t (0) - \Delta_j x_t \right\}^2 = \mathrm{Var}\left\{ \Delta_j v_t \right\}$$

$$- 2 \underset{\sim}{\tau} (n,j)^T S_{nk} (\underset{\sim}{\mu}) (M_{kk} (\underset{\sim}{G}_v') + S_{kk} (Q_1 \underset{\sim}{G}_v')) P_k \underset{\sim}{\Delta}_j + \underset{\sim}{\tau}(n,j)^T W_{22} \underset{\sim}{\tau}(n,j).$$

E. Robusthet

Parameterestimatene i modellene for $\{x_t\}$ og $\{v_t\}$ er ofte usikre slik at det er av vesentlig interesse å studere hvor robust prediktorene er med hensyn på feil i parametrene. La symbolet $\sim$ betegne at parameterestimater er innsatt istedet for de respektive sanne verdier. Det "totale" forventede kvadratavvik for $\Delta_j \tilde{v}_t (0)$ kan skrives

$$E\{\Delta_j \tilde{v}_t (0) - \Delta_j v_t\}^2 = \mathrm{Var}\{\Delta_j v_t\} + E\{\Delta_j \tilde{v}_t (0)\}^2 - 2\, E\{\Delta_j \tilde{v}_t (0)\, \Delta_j v_t\}$$

$$= \mathrm{Var}\{\Delta_j v_t\} - 2\, (m_{m\tilde{n}} (\underset{\sim}{\tilde{\mu}})\, \underset{\sim}{\tilde{\tau}})^T\, E\{P_m \underset{\sim}{v}_t\, \Delta_j v_t\}$$

$$+ [M_{nn}^{-1} (\underset{\sim}{\mu})\, M_{n\tilde{n}} (\underset{\sim}{\tilde{\mu}})\, \underset{\sim}{\tilde{\tau}}]^T\, E\{P_n x_t'\, P_n (\underset{\sim}{x}_t')^T\}\, M_{nn}^{-1} (\underset{\sim}{\mu})\, M_{n\tilde{n}} (\underset{\sim}{\tilde{\mu}})\, \underset{\sim}{\tilde{\tau}}$$

$$+ [M_{m\tilde{n}} (\underset{\sim}{\tilde{\mu}})\, \underset{\sim}{\tilde{\tau}}]^T\, E\{P_m \underset{\sim}{v}_t\, P_m \underset{\sim}{v}_t^T\}\, M_{m\tilde{n}} (\underset{\sim}{\tilde{\mu}})\, \underset{\sim}{\tilde{\tau}}.$$

Fra (8.20) og (8.26) får vi dermed

<u>Setning 4</u>:

$$E\{\Delta_j \tilde{v}_t (0) - \Delta_j v_t\}^2 = \mathrm{Var}\{\Delta_j v_t\} - 2\, \underset{\sim}{\tilde{\tau}} (\tilde{n},j)^T\, S_{\tilde{n}k} (\underset{\sim}{\tilde{\mu}})\, (M_{kk} (\underset{\sim}{G}_v')$$

$$+ S_{kk} (Q_1\, \underset{\sim}{G}_v'))\, P_k\, \underset{\sim}{\Delta}_j + \underset{\sim}{\tilde{\tau}} (\tilde{n},j)^T\, S_{\tilde{n}n} (\underset{\sim}{\tilde{\mu}})\, S_{nn}^{-1} (\underset{\sim}{\mu})\, W_{22}\, M_{nn}^{-1} (\underset{\sim}{\mu})\, M_{n\tilde{n}} (\underset{\sim}{\tilde{\mu}})\, \underset{\sim}{\tilde{\tau}} (\tilde{n},j)$$

for $j \geq 0$ og $j < k$, hvor μ og $\tilde{\mu}$ er slik at $\tilde{\mu}(z)/\mu(z)$ enten er lik 1 eller lik $\Delta_1(z)$.

9. Sammensatt estimering

Et sammensatt estimat er en veiet sum av estimater ved flere tidspunkter som dels er basert på hele utvalget ved de aktuelle tidspunkter, dels på den delen av utvalget som er felles ved flere av tidspunktene. Metoden er behandlet blant annet i Dagsvik 1975.

I sammensatt estimering betraktes $\underline{x}_t, \underline{x}_{t-1}, ..$ som gitte. (Se kap.4). Det forutsetter altså ikke noe om sammenhengen mellom påfølgende verdier av prosessen $\{x_t\}$. La x_t være en sammensatt estimator for x_t. Estimatoren x_t^* er slik at

$$(9.1) \quad E\, \mathrm{Var}\{x_t^* \mid \underline{x}_t, \underline{x}_{t-1}, \ldots\}$$

minimeres i klassen av estimatorer $\tilde{x}_t$ som tilfredstiller

$$(9.2) \quad E\{\tilde{x}_t \mid \underline{x}_t, \underline{x}_{t-1}, ..\} = x_t$$

Det er nettopp denne bibetingelsen som nødvendiggjør en oppsplitting av utvalget i flere deler. Fra (9.1) og (9.2) får vi umiddelbart

$$(9.3) \quad E\{x_t^* \mid x_t\} = E\, E\{x_t^* \mid \underline{x}_t, \underline{x}_{t-1}, ..\} = x_t$$

og

$$\mathrm{Var}\left\{ x_t^* \mid x_t \right\} = E\left[\mathrm{Var}\left\{ x_t^* \mid \underline{x}_t, \underline{x}_{t-1}, \ldots \right\} \mid x_t \right]$$

$$+ \mathrm{Var}\left[E\left\{ x_t^* \mid \underline{x}_t, \underline{x}_{t-1}, \ldots \right\} \mid x_t \right]$$

Det siste uttrykket er lik null slik at

$$\mathrm{Var}\left\{ x_t^* \mid x_t \right\} = E\left[\mathrm{Var}\left\{ x_t^* \mid \underline{x}_t, \underline{x}_{t-1}, \ldots \right\} \mid x_t \right] \tag{9.4}$$

Videre er

$$E\left\{ \tilde{x}_t - x_t \right\}^2 = E\, E\left\{ (\tilde{x}_t - E\left\{ \tilde{x}_t \mid \underline{x}_t, \underline{x}_{t-1}, \ldots \right\}) \mid \underline{x}_t, \underline{x}_{t-1}, \ldots \right\}$$

$$= E\, \mathrm{Var}\left\{ \tilde{x}_t \mid \underline{x}_t, \underline{x}_{t-1}, \ldots \right\} = \mathrm{Var}\left\{ \tilde{x}_t \mid x_t \right\}$$

for en vilkårlig lineær estimator som tilfredstiller (9.2). Altså er minimering av

$$E\, \mathrm{Var}\left\{ \tilde{x}_t \mid \underline{x}_t, \underline{x}_{t-1}, \ldots \right\}$$

det samme som å minimere forventet kvadratavvik under bibetingelsen (9.2).

I praksis innsnevres klassen av estimatorer som tilfredstiller (9.2) enda mer i det en betrakter spesielt enkle typer lineære estimatorer. Det er gjort studier som viser at den variansøkningen en dermed får er liten.

REFERANSER

Box, G.E.P. and Jenkins, G.M. 1970: Time Series Analysis, Forecasting and Control. Holden - Day, Inc. San Francisco.

Dagsvik, J. 1974: Variansestimering for nivåtallestimater og endringstallestimater ved Byråets arbeidskraftundersøkelser. Statistisk Sentralbyrå, Arbeidsnotat (IO 74/50).

Dagsvik, J. 1974: Presisjonsgevinst ved bruk av sammensatt estimering i Byråets arbeidskraft-undersøkelser. Statistisk Sentralbyrå, Arbeidsnotat (IO 75/24).

Dagsvik, J. 1976: Estimering og prediksjon ved bruk av metoder fra tidsserieanalysen i Byråets arbeidskraftundersøkelser. Statistisk Sentralbyrå, Arbeidsnotat (IO 76/4).

Scott, A.J. and Smith, T.M.F. 1974: Analysis of Repeated Surveys Using Time Series Methods. JASA **69**, 674-678.

Sæbø, H.V. 1976: Varianser og designeffekter for sysselsettingstall estim ved bruk av Byråets nye utvalgsplan. Statistisk Sentralbyrå, Arbeidsnotat (IO 76/1).

Whittle, P. 1963: Prediction and Regulation by Linear Least-Square Methods. English Universities Press Ltd, London.

Appendiks A. Autokorrelasjonsstrukturen for en sammensatt estimator tilpasset Byråets rotasjonsplan.

I Bjerve 1974 og Dagsvik 1975 er det gjort beregninger som viser hvor stor presisjonsgevinst en kan oppnå ved sammensatt estimering. Siden rotasjonsplanen i AKU er noe forskjellig fra rotasjonsplanene i andre land vil aktuelle sammensatte estimatorer i vår situasjon være mer kompliserte enn de som er i bruk for eksempel i Sverige og i USA. Beregninger av autokorrelasjonsformler for de aktuelle sammensatte estimatorer vil følgelig bli mer komplisert. Med en hensiktsmessig notasjon viser det seg at utledningene reduserer seg til et "regnskapsystem" hvor en setter opp antall symboler som har felles "indeks". Begrunnelsen for å dokumentere disse formlene er at noen av dem er brukt i Dagsvik 1975 uten bevis og i tillegg at vi trenger disse formlene for å bruke filtermetoden på tidsrekken av sammensatte estimater.

Utvalget i AKU trekkes som kjent i to trinn der første trinns enheter er faste fra undersøkelse til undersøkelse. Sammensatt estimering har derfor bare effekt på utvalgsfeilen innen utvalgsområdene. Utvalget innen annen trinns enheter roterer periodisk slik at halvparten av utvalget er felles ett kvartal og det påfølgende kvartal, og ved ett kvartal året etter. En fjerdedel av utvalget er felles ett kvartal og tre kvartaler senere, og ett kvartal og fem kvartaler senere mens utvalgene ett kvartal, to kvartaler og seks kvartaler senere er ikke-overlappende. En hensiktsmessig måte å illustrere rotasjonsplanen er følgende figur

figur 1

pulje \ kvartal	1	2	3	4	5	6	7	8	9	10	11
1 ...	x										
2 ...	x	x									
3 ...		x	x								
4 ...			x	x							
5 ...	x			x	x						
6 ...	x	x			x	x					
7 ...		x	x			x					
8 ...			x	x							
9 ...				x	x						
10 ...					x	x					

Dette betyr at puljenumrene for utvalget ved tidspunkt t er henholdsvis t, t+1, t+4 og t+5. La $y_t\,(t+j)$ være et estimat for det aktuelle populasjonsgjennomsnitt x_t ved tidspunkt t basert på pulje t-j. Den sammensatte estimator (9) i Dagsvik 1975 er lik

$$\text{(A.1)}\quad \tilde{\hat{x}}_t = p\,(y_{t-4} + \hat{d}_{t-4,t}) + q\,(y_{t-1} + \hat{d}_{t-1,t}) + (1-p-q)\,y_t \quad ,$$

hvor p+q, p og q ε $|0,1|$, y_t er "rå" estimatet for x_t og $\hat{d}_{t-j,t}$ er en estimator for endringen $x_t - x_{t-j}$ basert på de puljene som er felles ved tidspunktene t-j og t.

Spesielt får vi

$$\text{(A.2)}\quad \hat{d}_{t-1,t} = 1/2\left[\,y_t\,(t) + y_t\,(t+4) - y_{t-1}\,(t) - y_{t-1}\,(t+4)\,\right]$$

$$\text{(A.3)}\quad \hat{d}_{t-4,t} = 1/2\left[\,y_t\,(t) + y_t\,(t+1) - y_{t-4}\,(t) - y_{t-4}\,(t+1)\,\right]$$

og

$$\text{(A.4)}\quad y_t = 1/4\,\left[\,y_t\,(t) + y_t\,(t+1) + y_t\,(t+4) + y_t\,(t+5)\,\right]$$

hvilket innsatt i (A.1) gir

(A.5) $\tilde{\tilde{x}}_t = 0{,}25\ [\ p\ y_{t-4}\ (t-4) + p\ y_{t-4}\ (t-3) + q\ y_{t-1}\ (t-1)$

$+ (1+p+q)\ y_t\ (t) - p\ y_{t-4}\ (t) - q\ y_{t-1}\ (t) - p\ y_{t-4}\ (t+1)$

$+ (1+p-q)\ y_t\ (t+1) + q\ y_{t-1}\ (t+3) + (1+q-p)\ y_t\ (t+4)$

$- q\ y_{t-1}\ (t+4) + (1-p-q)\ y_t\ (t+5)\ ]$

La $\tilde{\tilde{v}}_{2t}$ være utvalgsfeilen innen utvalgsområdene når x_t estimeres ved $\tilde{\tilde{x}}_t$ og la

$$s_2^2 = \mathrm{Var}\left\{ v_{2t} \right\},$$

$$k_j = E\ \mathrm{cov}\left\{ y_t\ (t),\ y_{t-j}\ (t)\ \middle|\ \underset{\sim}{x}_t,\ \underset{\sim}{x}_{t-j} \right\} \Big/ E\ \mathrm{Var}\left\{ y_t\ (t)\ \middle|\ \underset{\sim}{x}_t \right\}$$

og

$$c_j = \mathrm{cov}\left\{ \tilde{\tilde{v}}_{2t},\ \tilde{\tilde{v}}_{2,t-j} \right\} = E\ \mathrm{cov}\left\{ \tilde{\tilde{x}}_t,\ \tilde{\tilde{x}}_{t-j}\ \middle|\ \underset{\sim}{x}_t,\ \underset{\sim}{x}_{t-1},\ \ldots,\ \underset{\sim}{J} \right\}$$

Vi antar at puljer med ulike puljenummer er stokastisk uavhengige. Fra (A.5) får vi nå umiddelbart

$$c_0 = 0{,}25\ s_2^2\ [\ 2p^2 + q^2 + (1+p+q)^2 + p^2 + q^2$$

$$- 2p\ (1+p+q)\ k_4 - 2q\ (1+p+q)\ k_1 + 2pq\ k_3 + p^2$$

$$+ (1+p-q)^2 - 2p\ (1+p-q)\ k_4 + q^2 + (1+q-p)^2 - q^2$$

$$- 2q\ (1+q-p)\ k_1 + (1-p-q)^2\ ]$$

som gir

(A.6) $c_0 = s_2^2\ (1+2p^2 + 2q^2 - q\ (1+q)\ k_1$

$+ \dfrac{pq\ k_3}{2} - p\ (1+p)\ k_4\ ).$

Sammenlikner vi nå $\tilde{\tilde{x}}_t$ ledd for ledd med

$\tilde{\tilde{x}}_{t-1} = 0{,}25\ [\ p\ y_{t-5}\ (t-5) + p\ y_{t-5}\ (t-4) + q\ y_{t-2}\ (t-2)$

$+ (1+p+q)\ y_{t-1}\ (t-1) - p\ y_{t-5}\ (t-1) - q\ y_{t-2}\ (t-1)$

$- p\ y_{t-5}\ (t) + (1+p-q)\ y_{t-1}\ (t) + q\ y_{t-2}\ (t+2)$

$+ (1+q-p)\ y_{t-1}\ (t+3) - q\ y_{t-2}\ (t+3) + (1-p-q)\ y_{t-1}\ (t+4)\ ]$

får vi

(A.7) $c_1 = 0{,}25\ s_2^2\ [\ 4\ q^2 + 2\ (1+2p^2 - 2q^2)\ k_1 - p\ (1+p-q)\ k_3$

$- p\ (1+p+q)\ k_5\ ].$

Utledningen av de resterende korrelasjonene er helt analog og vi tar ikke med detaljregningen her.

(A.8) $c_2 = 0{,}25\ s_2^2\ [\ 2q\ (1-q)\ k_1 + q\ (1+p-q)\ k_3 - pq\ k_5\ ]\ ,$

(A.9) $c_3 = 0{,}25\ s_2^2\ [\ 2p^2\ k_1 + (1-2pq - 2p^2 - 2q^2)\ k_3 + 2q^2\ k_4\ ]$,

(A.10) $c_4 = 0{,}25\ s_2^2\ [\ 4p^2 - q\ (1+q-p)\ k_3 + (2+4q^2 - 4p^2)\ k_4 - q\ (1+p+q)\ k_5\ ]$,

(A.11) $c_5 = 0{,}25\ s_2^2\ [\ 2p^2\ k_1 + 2q^2\ k_4 + (1-2p^2 - 2q^2 - 2pq)\ k_5\ ]$,

(A.12) $c_6 = 0{,}25\ s_2^2\ [\ pq\ k_3 + q\ (1-q-p)\ k_5\ ]$,

(A.13) $c_7 = 0{,}25\ s_2^2\ [\ 1+q-p\ ]\ p\ k_3$,

(A.14) $c_8 = 0{,}25\ s_2^2\ [\ 2p\ (1-p)\ k_4 - pq\ k_5\ ]$,

(A.15) $c_9 = 0{,}25\ s_2^2\ (1-p-q)\ p\ k_5$

og

$$c_j = 0 \text{ for } j \geq 10 .$$

Legg merke til at behandlingen av sammensatt estimering her skiller seg noe fra den som beskrives i Dagsvik 1975. Der betraktet vi $\underline{x}_t, \underline{x}_{t-1}, ..$ som <u>gitte</u> mens $\{\underline{x}_t\}$ her er en stokastisk prosess. For estimering av de respektive autokovarianser har dette imidlertid ingen betydning.

Appendiks B. Overgang fra månedsenhet til kvartalsenhet.

Vi skal her vise hvordan vi kan overføre "månedsmodellen" som ble behandlet i avsnittet foran til en "kvartalsmodell". Fra (3.1) og (3.2) får vi

$$\Delta_{12} x_t = e_t + r\Delta_{12} x_{t-1} = e_t + r\,(r\Delta_{12} x_{t-2} + e_{t-1})$$

$$= e_t + re_{t-1} + r^2\,(\,r\Delta_{12} x_{t-3} + e_{t-2}\,)$$

$$= r^3\,\Delta_{12} x_{t-3} + e_t'$$

der

(B.1) $$e_t' = e_t + re_{t-1} + r^2 e_{t-2} = a_t + (r-\alpha)\,a_{t-1} +$$

$$+ r\,(r-\alpha)\,a_{t-2} - \alpha r^2\,a_{t-3} - \beta[\,a_{t-12} + (r-\alpha)\,a_{t-13}$$

$$+ r\,(r-\alpha)\,a_{t-14} - \alpha\,r^2\,a_{t-15}\,].$$

Herav får vi

(B.2) $$\begin{cases} E\left\{e_t'\,e_{t-3}'\right\} = -\,\sigma^2\,\alpha\,r^2\,(1+\beta^2) \\ E\left\{e_t'\,e_{t-9}'\right\} = \sigma^2\,\alpha\,\beta\,r^2 \\ E\left\{e_t'\,e_{t-15}'\right\} = \sigma^2\,\alpha\,\beta\,r^2 \end{cases}$$

og

$$\mathrm{Var}\left\{e_t'\right\} = \sigma^2\,[\,1+(r-\alpha)^2\,(1+r^2) + \alpha^2 r^4\,]\,(1+\beta^2)$$

mens

$$E\left\{e_t'\,e_{t-3j}'\right\} = 0 \quad \text{for } j = 2 \text{ og } j \geq 6.$$

Går vi over til kvartalsenheten T ved å innføre tiden målt i kvartaler

for t =	t	t-3	t-6	t-9	t-12	t-15	t-18
er T	T	T-1	T-2	T-3	T-4	T-5	T-6

er det umiddelbart innlysende at vi kan skrive

$$e_T' = a_T' - \alpha_1\,a_{T-1}' - \alpha_4\,a_{T-4}' + \alpha_5\,a_{T-5}'$$

hvor a_T', a_{T-1}', , ... er ukorrelerte med $E\left\{a_T'\right\} = 0$ og $\mathrm{Var}\left\{a_T'\right\} = \sigma_1^2$.
Sammen med utledningene ovenfor gir dette

(B.3) $$\mathrm{Var}\left\{e_T'\right\} = \sigma_1^2\,(1+\alpha_1^2 + \alpha_4^2 + \alpha_5^2) = \sigma^2\,[\,1+(r-\alpha)^2\,(1+r^2)+\alpha^2 r^4\,]\,(1+\beta^2),$$

(B.4) $$E\left\{e_T'\,e_{T-1}'\right\} = -\,\sigma_1^2\,\alpha_1\,(1+\alpha_4^2) = -\,\sigma^2\,\alpha\,r^2\,(1+\beta^2)\,,$$

(B.5) $$E\left\{e_T'\,e_{T-3}'\right\} = \sigma_1^2\,\alpha_4\,\alpha_1 = \sigma^2\,\alpha\beta r^2 = \sigma_1^2\,\alpha_5 = E\left\{e_T'\,e_{T-5}'\right\},$$

(B.6) $$E\left\{e_T'\,e_{T-4}'\right\} = -\sigma_1^2\,\alpha_4\,(1+\alpha_1^2) = -\,\sigma^2\,\beta\,[\,1 + (r-\alpha)^2\,(1+r^2) + \alpha^2 r^4\,].$$

Kombinerer vi (B.2) med (B.3) og (B.4) med (B.3) får vi

(B.7) $$(1+\alpha_4^2)\,/\,\alpha_4 = (1+\beta^2)\,/\,\beta\,,$$

(B.8) $$(1+\alpha_1^2)\,/\,\alpha_1 = [\,1+(r-\alpha)^2\,(1+r^2) + \alpha^2\,r^4\,]\,/\,\alpha r^2.$$

Hver av disse likningene har to røtter med produkt lik 1. Vårt krav er at koeffisientene skal ha tallverdi mindre enn 1 hvilket medfører

(B.9) $\alpha_4 = \beta$,

(B.10) $$\alpha_1 = \begin{cases} 0 & \text{når } \alpha = 0 \\ \dfrac{1+\alpha^2 r^4 + (r-\alpha)^2 (1+r^2)}{2\alpha r^2} - \left[\left(\dfrac{1+\alpha^2 r^4 + (r-\alpha)^2 (1+r^2)}{2\alpha r^2} \right)^2 - 1 \right]^{\frac{1}{2}} & \text{når } \alpha \neq 0 \end{cases} ,$$

(B.11) $$\sigma_1^2 = \begin{cases} \sigma^2 (1 + r^2 + r^4) & \text{når } \alpha = 0 \\ \sigma^2 r^2 \alpha / \alpha_1 & \text{når } \alpha \neq 0 \end{cases}$$

og

(B.12) $\alpha_5 = \alpha_1 \alpha_4$.

Vi har altså kommet fram til modellen

(B.13) $$\Delta_4 x_T = r^3 \Delta_4 x_{T-1} + a_T' - \alpha_1 a_{T-1}' - \beta a_{T-4}' + \alpha_1 \beta a_{T-5}'$$

som har samme form "månedsmodellen".

Modellen (3.5) er tilsvarende invariant ovenfor endring av tidsenhet til kvartalsenhet. Fra (3.5) får vi

$$\Delta_{12} u_t = \Delta_{12} u_{t-1} + a_t - \beta a_{t-12}$$

$$= \Delta_{12} u_{t-3} + a_t + a_{t-1} + a_{t-2} - \beta (a_{t-12} + a_{t-13} + a_{t-14})$$

hvor

$$u_t = x_t - \lambda x_{t-12} .$$

Herav følger at med kvartal som tidsenhet er

$$\Delta_1 \Delta_4 x_T - \lambda \Delta_1 \Delta_4 x_t = a_T'' - \beta a_{T-4}''$$

hvor

$$a_T'' = a_t + a_{t-1} + a_{t-2} .$$

Appendiks C. Høy autokorrelasjon for utvalgsfeilen mellom utvalgsområdene.

Når autokorrelasjonen for utvalgsfeilen mellom utvalgsområdene er nær 1 vil vi ikke kunne redusere denne utvalgsfeilen for nivåtall noe særlig ved å benytte en lineær prediktor av typen foran. Dette er lett å innse men vi skal likevel gjennomføre et formelt resonement. Vi har

$$E\left\{\hat{v}_t(0) - v_t\right\}^2 = E\left\{\hat{v}_t(0) - v_{2t}\right\}^2 + \mathrm{Var}\left\{v_{1t}\right\} - 2\,E\left\{v_t(0)\,v_t\right\},$$

$$E\left\{y'_{t-j}\;v_{1t}\right\} = E\left\{v'_{1,t-j}\;v_{1t}\right\} = E\left\{\mu\, v_{1,t-j}\,v_{1t}\right\}$$

som gir

$$E\left\{\hat{v}_t(0) - v_t\right\}^2 = E\left\{\hat{v}_t(0) - v_{2t}\right\}^2 + \mathrm{Var}\left\{v_{1t}\right\}$$

$$- 2\sum_j \nu_j(0)\;E\left\{v'_{1,t-j}\;v_{1t}\right\}.$$

Dersom $\mathrm{korr}\left\{v_{1,t-j},\,v_{1t}\right\} \simeq 1$ og $\mathrm{Var}\left\{v_{1t}\right\}$ er tilnærmet konstant over tid vil

$$E\left\{\mu v_{1,t-j}\;v_{1t}\right\} \simeq 0 \quad \text{for } j = 0, 1, 2 \,..$$

slik at

$$E\left\{\hat{v}_t(0) - v_t\right\}^2 \simeq E\left\{\hat{v}_t(0) - v_{2t}\right\}^2 + \mathrm{Var}\left\{v_{1t}\right\}$$

Minimeringsproblemet er dermed redusert til å finne den lineære prediktor som reduserer utvalgsfeilen innen utvalgsområdene mest mulig. For estimering av endringer får vi

$$E\left\{\Delta_j\,\hat{v}_t(0) - \Delta_j\,v_t\right\}^2 \simeq E\left\{\Delta_j\,\hat{v}_t(0) - \Delta_j\,v_{2t}\right\}^2$$

$$+ \mathrm{Var}\left\{\Delta_j\,v_{1t}\right\} \simeq E\left\{\Delta_j\,\hat{v}_t(0) - \Delta_j\,v_{2t}\right\}^2$$

Utvalgsfeilen mellom utvalgsområdene er altså ubetydelig ved estimering av endringstall.