Interne notater

STATISTISK SENTRALBYRÅ

82/29 17. september 1982

AKU-NIVÅ-PROGRAM - ET PROGRAM TIL BEREGNING AV ENDRINGSTALL OG SAMMENSATTE NIVÅTALL FOR AKU. PROGRAMBESKRIVELSE OG BRUKERVEILEDNING

AV

ROY ØSTENSEN*)

INNHOLD

*) Jeg vil takke Tom Kirkerud for nyttige tips under programmeringsarbeidet.

1. INNLEDNING

Siden 1979 har Byrået brukt et program til beregninger av nivå- og endringstall i Byråets arbeidskraftundersøkelser (AKU). Disse beregningene forsøker å utnytte dataene fra tidligere kvartaler på en slik måte at usikkerheten blir redusert mest mulig. Resultatene av disse beregningene er ikke publisert, men er brukt som hjelp i analyser av de tall som blir publisert fra AKU.

Programmet tar utgangspunkt i teorien for sammensatt estimatorer. En sammensatt estimator er en veiet sum av estimatorer fra flere tidspunkter. Vektene er valgt slik at estimatoren blir forventningsrett for det vi vil måle. Hvis den gruppen vi er interessert i er relativt stabil, kan sammensatte estimatorer gi tildels ganske betydelig reduksjon i usikkerheten i forhold til et enkelt gjennomsnitt. Er derimot gruppen svært utstabil, slik tilfellet er med gruppen arbeidssøkende, kan usikkerheten isteden bli større hvis man bruker sammensatte estimatorer istedenfor det enkle gjennomsnitt. Det forannevnte programmet kan brukes for denne gruppen, men vi vil likevel advare mot at man gjør det. Derimot kan programmet med hell brukes for sysselsetningen i enkeltnæringer, noe som også Byrået har gjort hittil.

Etter en tids bruk har det vist seg at programmet burde revideres. Blant annet har det vært et ønske at programmet lagrer nivåtall og rådata fra tidligere kvartaler. Vi har derfor foretatt en fullstendig revisjon av programmet, hvilket bl.a. har bestått i å legge størstedelen av programmet over på en diskette, som for tiden har betegnelsen (SSB:I-JOHNSEN). På samme diskette blir også rådataene lagt, etter hvert som programmet leser dem, samt de nye nivåtallene som blir beregnet. Om man må foreta omkjøringer av tidligere kvartaler, er følgelig ikke det noen tidkrevende operasjon (så sant belastningen på maskinen ikke er for stor).

Dette notatet er delvis ment som en brukerveiledning av programmet og delvis som en dokumentasjon. Da notatet IN 80/17, "Sammensatt estimering i forbindelse med Byråets arbeidskraftundersøkelser" (Østensen, 1980) er svært teknisk, har vi også ønsket å gi en mindre teknisk innføring i teorien for sammensatt estimering. Vi har derfor forsøkt å begrense matematikken til et minium. Denne teoretiske bakgrunnen er gitt i kap. 2. Den som ønsker en mere matematisk innføring bør studere forannevnte notat, samt de referansene som gis der.

Under revisjonen av programmet fant vi noen trykkfeil og regnefeil i notatet IN 80/17. Disse feilene gjør vi rede for i vedlegg 1.

I kapittel 3 vil man finne en kort oversikt over hvordan programmet virker. Selve brukerveiledningen vil man finne i kapittel 4. En utlisting av programmet er tatt med i vedlegg 3.

2. ESTIMATORENE - EN TEORETISK BAKGRUNN

I dette kapitlet vil vi forsøke å gi en minst mulig teknisk innføring i teorien bak estimatorene som er brukt for brukere og andre ikke har behov for å kjenne så mye til det rent matematisk - statistiske grunnlaget for dem. Vi antar likevel at slike begreper som forventning, varians og korrelasjon ikke er helt ukjent for leseren. Skulle leseren ønske en dypere matematisk innføring, viser vi til IN 80/17 med de referansene som er gitt der. Vi henviser også til vedlegg 1 i dette notatet, hvor vi har gitt en oversikt over de trykkfeilene vi har funnet.

I dette kapitlet tar vi direkte utgangspunkt i det rotasjonsmønsteret som brukes i AKU. Fig. 1 viser skjematisk hvordan rulleringsplanen er lagt opp. Utvalget er delt opp i fire puljer, hver på ca. 1/4 av hele utvalget. Hver pulje er med ialt fire ganger, først to kvartaler, deretter et opphold på to kvartaler, og så to kvartaler til. I tillegg kommer det hvert kvartal med en helt ny pulje. På denne måten vil en halvpart av utvalget være med et kvartal senere og en halvpart være med et år senere. Hvis vi f.eks. ser på 1. kvartal (pulje 1, 2, 5, 6 på fig. 1), vil pulje 2 og 6 være med i 2. kvartal, og pulje 5 og 6 være med 1. kvartal et år senere.

Figur 1.

Kvartal	1	2	3	4	1	2
Pulje 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
11						x

Fordelen med et slikt rotasjonsmønster framfor å bytte ut hele utvalget for hvert kvartal, er at man dermed kan få mere presise anslag (estimater) for endringen i sysselsetningen fra forrige kvartal og samme kvartal et år tidligere. Dermed er det samtidig mulig å få mer presise anslag for sysselsetningen i inneværende kvartal ved å ta hensyn til estimatene for tidligere kvartaler. Dermed ledes vi til å bruke sammensatte estimatorer, som tar med estimatene for tidligere kvartaler når de nye estimatene skal beregnes. Som vi skal se reduserer dette usikkerheten til anslagene ganske betydelig.

2.1. Sammensatt estimering av nivåtall

Vi vil i dette avsnittet forsøke å illustrere idéen bak sammensatt estimering ved noen regneeksempler. Vi vil ta utgangspunkt i AKU-tallene for næring 1, jordbruk, skogbruk, fiske og fangst. Tallene er gitt i tabell 1 og 2.

Vi skylder å gjøre oppmerksom på at hvis man setter inn tallene i de enkelte uttrykk som følger, vil resultatene avvike noe fra de resultatene vi har fått, jfr. f.eks. uttrykket (2,1,2) hvor man vil få resultatet 155 389, som avrundet blir 155 400, mens vi gir resultatet lik 155 200 like nedenfor. Årsaken er at tallene som puttes inn er avrundet, mens maskinen så langt mulig regner med uavrundete tall. "Så langt mulig", fordi den mates med puljetall som er avrundet til nærmeste hundre. Det behøver følgelig ikke være feil hvis man kommer fram til svar som avviker litt fra det teksten gir.

Tabell 1. Observerte AKU-tall fordelt på puljer for jordbruk, skogbruk, fiske og fangst. 1 000 personer

	Sysselsatte i næringen			Antall i den totale arbeidsstyrke		
	1. kvartal 1982	4. kvartal 1981	1. kvartal 1981	1. kvartal 1982	4. kvartal 1981	1. kvartal 1981
I alt	154,8	156,2	160,2	2 948,9	2 925,8	2 925,8
Pulje 1	39,0	37,4	35,2	745,2	721,8	721,6
Pulje 2	36,9	39,1	47,2	741,2	741,6	771,7
Pulje 3	37,7	40,3	38,2	710,8	738,1	718,7
Pulje 4	41,2	39,4	39,6	751,7	724,3	713,8

Vi vil gjøre oppmerksom på at vi har gitt puljene nummer etter hvor mange kvartaler de har vært med til og med det kvartalet vi ser på. For pulje nr. 1 er det altså 1. gang puljen er med vedkommende kvartal, pulje 2 er med for 2. gang osv. Dette betyr at den puljen vi betegner som nr. 1 i 1. kvartal 1981, blir pulje 2 i 2. kvartal 1981, pulje 3 i 1. kvartal 1982 og pulje 4 i 2. kvartal 1982.

Av hensyn til de etterfølgende regneeksemplene er det hensiktsmessig å slå sammen pulje 1 og 2, pulje 1 og 3, pulje 2 og 4 og pulje 3 og 4 og beregne totalt antall sysselsatte i næring, ifølge disse sammenslåtte puljene. Det vil snart framgå hvorfor.

I tabell 2 har vi blåst opp puljetallene til å gjelde det totalt antall sysselsatte. Nøyaktig har vi blåst opp puljetallet med forholdet mellom antall i den totale arbeidsstyrken i alt og det antall puljen representerer i arbeidsstyrken. For pulje nr. 1 i 1. kvartal 1982 vil det si at de 39 000 som er sysselsatt i næringen er blåst opp med faktoren 2 948,9/745,2 eller 3,957. Ifølge pulje 1 skulle altså i alt 3,957 . 39 000 = 154 330 være sysselsatt i jordbruk, skogbruk etc. 1. kvartal 1982.

Tabell 2. Antall sysselsatte i alt i jordbruk, skogbruk, fiske og fangst ifølge de enkelte puljer i AKU. 1 000 personer

	1. kvartal 1982	4. kvartal 1981	1. kvartal 1981
Alle puljer	154,80	156,20	160,20
Pulje 1	154,33	151,60	142,72
Pulje 2	146,81	154,26	178,95
Pulje 3	156,41	159,75	155,51
Pulje 4	161,63	159,16	162,32
Pulje 1 + 2	150,58	152,95	161,45
Pulje 1 + 3	155,34	155,72	149,10
Pulje 2 + 4	154,27	156,68	170,96
Pulje 3 + 4	159,09	159,45	158,90

I følge tabell 1 er råestimatet for antall sysselsatte, Y, lik 154 000. Lar vi $\bar{Y}$ stå for den prosentvise andel av den totale arbeidsstyrken, kan vi anslå standardavviket til Y ved

$$SD(Y) = N\ SD(\bar{Y}) \stackrel{\wedge}{=} N\sqrt{1{,}5\bar{Y}(1-\bar{Y})/n}$$

($\stackrel{\wedge}{=}$ kan leses "anslås lik"), hvor N er totalt antall i arbeidsstyrken (i vårt eksempel 2 948 900 i 1. kvartal 1982) og n er antall observasjoner (innkomne svar). I 1. kvartal 1982 kom det inn i alt 10 683 brukbare svar i AKU, slik at n = 10 683 i vårt tilfelle. Faktoren 1,5 er en korreksjonsfaktor som skyldes at Byråets utvalgsplan ikke baserer seg på enkle tilfeldige utvalg. Av tabell 1 finner vi $\bar{Y}$ lik 0,0525, slik at

$$SD(Y) = 2\ 948\ 900\sqrt{1{,}5 x 0{,}0525\ (1-0{,}0525)/10\ 683} = 7\ 800$$

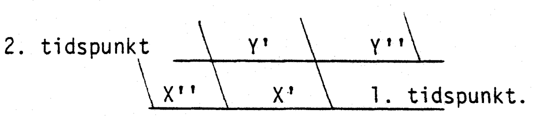
Et tilnærmet 95 prosent konfidensintervall for antall sysselsatt i næring 1 blir derfor

$$Y \pm 2\ SD(Y) = 154\ 800 \pm 15\ 600$$

eller (139 200, 170400). (Et 95 prosents konfidensintervall vil i 95 av 100 ganger vi gjør samme undersøkelse, dekke det faktisk antall sysselsatte.) Det framkommer ved at vi tar to ganger standardavviket SD(Y) og legger det til, henholdsvis trekker det fra, den anslåtte verdien Y, slik vi har gjort ovenfor.

Estimatoren Y ovenfor tar ikke hensyn til at vi har tall fra tidligere kvartaler. Nå vil folk stort sett ha samme type arbeid fra kvartal til kvartal. Hvis vi derfor tar hensyn til dataene fra tidligere kvartaler, kan vi følgelig redusere usikkerheten på anslaget vårt for inneværende kvartal. Dette leder oss til bruk av sammensatte estimatorer, det vil si estimatorer som er satt sammen av observasjonene fra inneværende og tidligere kvartaler.

La Y' og X' betegne antall sysselsatte i henholdsvis inneværende og forrige tidspunkt ifølge de puljene som er felles for disse to tidspunktene, og Y'' og X'' være antallet ifølge de andre puljene, på følgende måte:



En enkelt apostrof (') indikerer altså enhetene som er felles for de to tidspunktene, mens en dobbel apostrof ('') indikerer enhetene som er valgt uavhengig på de to tidspunktene.

I første omgang vil vi anta at undersøkelsen bare er foretatt på to tidspunkter, 4. kvartal 1981 og 1. kvartal 1982. I IN 80/17 ble det vist at under visse forutsetninger som er gjort rede for i avsnitt 2 i notatet, vil den forventningsrette estimator med minst varians kunne skrives som

$$\hat{P}_2 = [\lambda\mu\rho(X''-X') + \lambda Y' + \mu(1 - \rho^2\mu)Y'']/(1-\rho^2\mu^2)$$

hvor λ er den andel av utvalget som er felles på begge tidspunkter, $\mu = 1-\lambda$ er den andel som er byttet ut, og ρ er korrelasjonen mellom observasjonene for samme person fra forrige og inneværende kvartal. Vi har anslått korrelasjonen ρ til å bli 0,86, se avsnitt 5. Autokorrelasjon i IN 80/17. Antar vi at $\lambda = \mu = \frac{1}{2}$, altså at halvparten av utvalget ble byttet ut mellom de to kvartalene, finner vi

$$\hat{P}_2 = 0{,}39\ Y'' + 0{,}61Y' + 0{,}26\ (X''-X')$$

Vær oppmerksom på at $\hat{P}_2$ ovenfor avviker fra $\hat{p}_2$ i IN 80/17, på den måten at vi her bruker $Y' = N_2\bar{Y}'$, $Y'' = N_2\bar{Y}''$, $X' = N_1\bar{X}'$ og $X'' = N_1\bar{X}''$, istedenfor gjennomsnittene $\bar{Y}'$, $\bar{Y}''$, $\bar{X}'$ og $\bar{X}''$. N_1 og N_2 er totalt antall i arbeidsstyrken ved henholdsvis forrige og inneværende kvartal. Nå er $N_2 \approx N_1$, slik at $\hat{P}_2 \approx N_2\hat{p}_2$, slik at det har liten betydning om vi baserer estimeringen av nivåtall på $\hat{P}_2$ eller $N_2\hat{p}_2$. Derimot vil en liten forskjell i N_1 og N_2 kunne gi et meget stort utslag i endringsestimatene. Mens vi opprinnelig baserte estimeringen på gjennomsnittene, har vi derfor nå gått over til å bruke totaltallene direkte i estimeringen.

Slik rulleringsmønsteret i AKU er (se fig. 1) finner vi ifølge tabell 2 at

$$Y' = Y_{2+4} = 154\ 270$$

$$Y'' = Y_{1+3} = 155\ 340$$

$$X' = X_{1+3} = 155\ 720$$

$$X'' = X_{2+4} = 156\ 680$$

Vårt sammensatte estimat for antall sysselsatte i næring 1 blir dermed

$$\hat{P}_2 = (0{,}39\text{x}155{,}34+0{,}61\text{x}154{,}27+0{,}26(156{,}68-155{,}72))\text{x}10^3 = 154\ 940$$

Standardavviket vil P_2 er gitt som (ligning (2,6) i IN 80/17)

$$SD(\hat{P}_2) = \sqrt{0{,}77}\ SD(Y)$$
$$= 6\ 860$$

Et 95 prosents konfidensintervall basert på $\hat{P}_2$ blir derfor

$$\hat{P}_2 \pm 2\ SD\ (\hat{P}_2) = 154\ 940 \pm 13\ 720$$

eller (141 200, 168 600). Til sammenligning var konfidensavviket 15 600 da vi baserte beregningene på Y. Vi har dermed redusert avviket med ca. 12 prosent.

Vi må gjøre oppmerksom på at beregningene ovenfor hviler på visse forutsetninger vi ikke har nevnt, bl.a. at puljene er like store og at varians og korrelasjon er uavhengig av tiden. Vi kommer her stilltiende til å anta at de nødvendige forutsetningene er oppfylt, og vil bare vise til IN 80/17, hvor forutsetningene er gjort rede for.

Estimatoren $\hat{P}_2$ foran har som en forutsetning av undersøkelsen blir gjort på to og bare to tidspunkter. Når en undersøkelse foretas regelmessig, slik som i AKU, vil $\hat{P}_2$ ikke lenger være best; vi må ta hensyn til estimatene fra tidligere kvartaler. Nå gjør rotasjonsmønsteret i AKU at matematikken blir noe uoversiktlig når vi skal finne estimatoren med laveste varians i dette tilfellet. La $\hat{P}$ (t-s) være nivåtallsestimatoren for tidspunkt t-s (s<t) og X' (t-s) være antall sysselsatte ved samme tidspunkt ifølge de enhetene som er med ved inneværende tidspunkt t. Da ser det ut til at den "beste" estimatoren må skrives på formen

$$\hat{P}(t) = \sum_{i=1}^{4} a_i(t)Y_i(t) + \sum_{s=1}^{5} b_{t-s}(\hat{P}_{t-s}-X'_{t-s})$$

Med $Y_i(t)$ mener vi antall sysselsatte ifølge den puljen som ved inneværende tidspunkt er med for i'te gang, i = 1, 2, 3, 4.

Vi kan nevne at

$$X'_{t-1} = (X_1(t-1) + X_3(t-1))/2$$

$$X'_{t-2} = 0 \quad \text{(se fig. 1)}$$

$$X'_{t-3} = X_2(t-3)$$

$$X'_{t-4} = (X_1(t-4) + X_2(t-4))/2$$

$$X'_{t-5} = X_1(t-5)$$

Vi har forsøkt å finne koeffisientene $a_i(t)$ og b_{t-s}, men har ikke kommet fram. Vi har derfor isteden valgt å ta utgangspunkt i en estimator på formen

$$P^*_t = a(P^*_{t-1} + \hat{d}(t-1,t)) + b(P^*_{t-4} + \hat{d}(t-4,t)) + (1-a-b)Y_t$$

og minimere variansen til P^*_t m.h.p. a og b rent nummerisk.

De to estimatorene $\hat{d}(t-1,t)$ og $\hat{d}(t-4,t)$ er estimatorer for endringen i sysselsettingen fra henholdsvis tidspunkt t-1 og t-4 til tidspunkt t. Uttrykt ved puljetallene ser de slik ut:

$$\hat{d}(t-1,t) = [\{Y_2(t) - X_1(t-1)\} + \{Y_4(t)-X_3(t-1)\}]/2$$

$$\hat{d}(t-4,t) = [\{Y_3(t) - X_1(t-4)\} + \{Y_4(t)-X_2(t-4)\}]/2$$

Minimeringen gav som resultat (IN 80/17 avsn. 3.2.) at estimatoren får formen

$$P^*_2 = 0{,}2348(P^*_{t-4}-X'_{t-4}) + 0{,}3905(P^*_{t-1}-X'_{t-1}) + \frac{1}{4}(0{,}3746Y_1 + 1{,}1557Y_2 + 0{,}8443Y_3 + 1{,}6254Y_4) \quad (2,1,2)$$

$$(Y_t = (Y_1+Y_2+Y_3+Y_4)/4)$$

Standardavviket viste seg å bli

$$SD(P^*_t) = 0{,}7512\ SD(Y_t)$$

I vårt tilfelle er P^*_{t-4} = 165 900, P^*_{t-1} = 152 600. Vi finner dermed P^*_{t-4} = 155 200 med et standardavvik lik $SD(P^*_t)$ = 5 850.

Konfidensintervallet for antall sysselsatte blir dermed 155 200 ± 11 700, eller (143 500, 166 900).

I forhold til Y har vi altså redusert konfidensavviket med ca. en fjerdedel.

Som nevnt foran kan vi finne nivåtallsestimatorer med enda lavere varians (usikkerhet) enn dette. Vi tviler likevel på at den videre gevinst det gir å bruke estimatoren med minst varians står i noe rimelig forhold til det arbeid det krever å finne denne estimatoren. Programmet baserer seg derfor på P_t^* i (2,1,2).

2.2. Endringstall

Av publiseringsmessige grunner vil man vel normalt angi endringstall som differensen mellom de estimerte nivåtall for inneværende og forrige tidspunkt. Vi finner at endringsestimatoren

$$\tilde{d}(t-1,t) = Y_t - X_{t-1}$$

vil ha et standardavvik omtrent lik

$$SD(\tilde{d}(t-1,t)) = 1{,}07 SD(Y_t)$$

Av tabell 1 finner vi at endringen i næring 1 fra 4. kvartal 1981 til 1. kvartal 1982 blir

$$\tilde{d}(t-1,t) = 154\ 800 - 156\ 200 = -1\ 400,$$

med et standardavvik lik

$$SD(\tilde{d}(t-1,t)) = 8\ 300$$

Et konfidensintervall for endringen blir derfor

$$-1\ 400 \pm 16\ 600$$

eller (-18 000, 15 200). Vi har følgelig ingen grunn til å tro det har vært noen reel nedgang, så lenge vi bare betrakter disse to tidspunktene.

For endringen fra 1. kvartal 1981 til 1. kvartal 1982 har vi tilsvarende at

$$\begin{aligned} \tilde{d}(t-4,t) &= Y_t - X_{t-4} \\ &= 154\ 800 - 160\ 200 \\ &= -5\ 400 \end{aligned}$$

I dette tilfellet er standardavviket

$$\begin{aligned} SD(\tilde{d}(t-4,t)) &= 1{,}11\ SD(Y_t) \\ &= 8\ 650 \end{aligned}$$

Konfidensintervallet blir dermed

$$-5\ 400 \pm 17\ 300$$

eller (-22 700, 11 900). Vi kan nok ha større grunn til å tro at det har vært en nedgang her, men uten å trekke inn trender, kan vi likevel ikke påstå at sysselsetting har gått ned.

I AKU-NIVÅ-PROGRAM har vi gitt avkall på kravet om at estimatene for endringstallene skal stemme overens med differansen mellom de estimerte nivåtall. Dermed kan vi få redusert usikkerheten enda noe mer. Våre beregninger antyder at de estimatorene som har minst usikkerhet, er

$$\hat{d}(t-1,t) = [0{,}14(Y_t'' - X_{t-1}'') + (Y_t' - X_{t-1}')]/1{,}14$$

$$\hat{d}(t-4,t) = [0{,}23(Y_t'' - X_{t-4}'') + (Y_t' - X_{t-4}')]/1{,}23 \qquad (2,2,1)$$

med standardavvik lik henholdsvis

$$SD(\hat{d}(t-1,t)) = 0{,}70\ SD(Y_t)$$

og $$SD(\hat{d}(t-4,t)) = 0{,}87\ SD(Y_t)$$

For næring 1 finner vi

$$\hat{d}(t-1,t) = [0{,}14(155\ 340 - 156\ 680) + (154\ 270 - 155\ 720)]/1{,}14 = -1\ 440$$

$$SD(\hat{d}(t-1,t)) = 0{,}70\ SD(Y) = 5\ 460$$

$$\hat{d}(t-4,t) = [0{,}23(150\ 580 - 158\ 900) + (159\ 090 - 161\ 450)]/1{,}23 = -3\ 470$$

$$SD(\hat{d}(t-4,t)) = 6\ 750$$

Dette gir konfidensintervallene for endringene lik:
Fra 4. kvartal 1981 til 1. kvartal 1982: -1 440 ± 10 920 eller (-12 400, 9 500).
Fra 1. kvartal 1981 til 1. kvartal 1982: -3 470 ± 13 500 eller (-17 000, 10 000).
Når det gjelder $\tilde{d}(t-1,t)$ og $\hat{d}(t-1,t)$ ser vi at begge gir samme estimat for nedgangen fra forrige kvartal, nemlig en nedgang på 1 400 personer (for $\hat{d}(s,t)$ er tierne tatt med av hensyn til avrundingsfeil). Dette vil nok heller være unntaket enn regelen, siden de beregners på forskjellig måte. Derimot har $\hat{d}(t-1,t)$ redusert konfidensavviket fra 16 600 til 10 900. Vi har altså fått redusert usikkerheten med ca. 1/3, hvilket må sies å være betydelig.

Når det gjelder endringen fra 1. kvartal 1981 til 1. kvartal 1982, er estimatet endret fra en nedgang på 5 400 til en nedgang på 3 500. Tallene i seg selv betyr egentlig lite for oss. Det som betyr noe er at konfidensavviket er redusert fra 17 300 til 13 500 eller med 22 prosent. Gevinsten er mindre enn for $\hat{d}(t-1,t)$'s vedkommende, men fortsatt er den god. Nivåprogrammet baserer seg derfor på $\hat{d}(t-1,t)$ og $\hat{d}(t-4,t)$ ved beregning av endringstall.

Merk at i våre regneeksempler kan vi ikke påstå noen reell nedgang i antall sysselsatte i næring 1, hverken fra et kvartal tilbake eller fra et år tilbake.

Det er et spørsmål som er interessant, nemlig hvor god endringsestimatoren

$$d^*(s,t) = P_t^* - P_s^* \quad s<t$$

er i forhold til $\hat{d}\,(s,t)$.

Desverre er det ikke lett å finne hvor stor korrelasjonen mellom P_t^* og henholdsvis P_{t-1}^* og P_{t-4}^* er, så vi kjenner ikke standardavviket til $d^*(t-1,t)$ eller $d^*(t-4,t)$. I IN 80/17 har vi derfor isteden sett på

$$\hat{d}'(s,t) = \hat{P}(t) - \hat{P}(s)\ ,\ s<t$$

hvor

$$\hat{P}(t) = 0{,}34Y''(t) + 0{,}66\ [Y'(t) + 0{,}86\{\hat{P}(t-1) - X'(t-1)\}]$$

Det viser seg at $\hat{d}'(t-1,t)$ har et standardavvik (IN 80/17, tab. 3)

$$SD(\hat{d}'(t-1,t)) = 0{,}763\ SD(Y),$$

mens

$$SD(\hat{d}'(t-4-t)) = 1{,}029\ SD(Y).$$

Til sammenligning har altså estimatoren $\hat{d}(s,t)$ et standardavvik lik:

$$SD(\hat{d}(t-1,t)) = 0,70\ SD(Y)$$

og $$SD(\hat{d}(t-y,t)) = 0,87\ SD(Y)$$

Nå har $\hat{P}(t)$ et standardavvik (tab. 1, IN 80/17)

$$SD(\hat{P}(t)) = 0,822\ SD(Y)$$

mens

$$SD(P^*(t)) = 0,751\ SD(Y),$$

(s. 13 i IN 80/17) slik at $P^*(t)$ reduserer usikkerheten med ca. 9 prosent i forhold til $\hat{P}(t)$. Et tilnærmet anslag på standardavviket til $d^*(s,t)$ blir følgelig

$$SD(d^*(t-1,t)) = 0,70\ SD(Y),\ s=t-1$$

og $$SD(d^*(t-4,t)) = 0,94\ SD(Y),\ s=t-4$$

Vi skulle kunne regne at dette er øvre grenser for standardavviket. Vi har nemlig grunn til å tro at korrelasjonen mellom $P^*(t)$ og $P^*(s)$ $(s<t)$ (dvs. samvariasjonen mellom $P^*(t)$ og $P^*(s)$) er større enn mellom $\hat{P}(t)$ og $\hat{P}(s)$, siden uttrykket for $P^*(t)$ trekker inn ledd fra flere tidligere tidspunkter enn $\hat{P}(t)$. Spesielt gjelder dette antagelig korrelasjonen mellom $P^*(t)$ og $P^*(t-4)$ i forhold til korrelasjonen mellom $\hat{P}(t)$ og $\hat{P}(t-4)$. Dette gir en antydning om at $d^*(s,t) = P_t^* - P_s^*$ kan gi omtrent samme gevinst i usikkerheten som $\hat{d}(s,t)$. Programmet AKU-NIVA-PROGRAM baserer seg likevel på $\hat{d}(s,t)$ for $s=t-1$ og $s=t-4$.

3. AKU-NIVA-PROGRAM - BESKRIVELSE

3.1. Estimatorer

Først litt notasjon. Når det i fortsettelsen snakkes om et puljenummer, f.eks. pulje 4, mener vi den puljen som er med på vedkommende tidspunkt for 4. gang. Når vi refererer til pulje 2 for 4 kvartaler siden, er det altså samme pulje som pulje 4 ved inneværende kvartal.

Nivå: Med nivået i en næring vil vi mene det totale antall som er sysselsatt i den næringen vi er interessert i.

La da

Y_i = nivået i næringen i inneværende kvartal ifølge pulje i i utvalget, i = 1, 2, 3, 4.

X_i = Tilsvarende nivå i næringen ved forrige kvartal, i = 1, 2, 3, 4.

Z_i = Nivået i næringen for ett år siden i pulje i, i = 1, 2, 3, 4.

Definer videre

Y_1'' = Nivået i pulje 1+3 ved inneværende kvartal

Y_2' = Nivået i pulje 2+4

Y_3'' = Nivået i pulje 1+2

Y_4' = Nivået i pulje 3+4

X_1' = Nivået i pulje 1+3 ved forrige kvartal

X_2'' = Det tilsvarende nivå i pulje 2+4

Z_1' = Nivået i pulje 1+2 for ett år siden

Z_2'' = Det tilsvarende nivå i pulje 3+4

Årsaken til at vi slår sammen puljene som foran, er at Y_2' og X_1' refererer seg til de samme enhetene, og at Y_4' og Z_1' gjør det samme. Y_1'' og X_2'' er derimot bare med henholdsvis ved inneværende og forrige kvartal. Det samme gjelder for Y_3'' og Z_2''.

De estimatorene vi har basert oss på, har dermed fått følgende form:

Nivåtall:

$$P_{1t} = 0,2348(P_{t-4}-Z_1')+0,3905(P_{t-1}-X_1') + (0,3746Y_1+1,1557Y_2+0,8443Y_3+1,6254Y_4)/4$$

$$P_{2t} = \bar{Y} = (Y_1+Y_2+Y_3+Y_4)/4$$

Endringstall:

$$D(t-1,t) = [Y_2'-X_1'+0,14(Y_1''-X_2'')]/1,14$$

$$D(t-4,t) = [Y_4'-Z_1'+0,23(Y_3''-Z_2'')]/1,23$$

P_{2t} er identisk med råestimatet. Vi har tatt P_{2t} med for å lette sammenligningen mellom råestimatene og de sammensatte nivåtallsestimatorene. Ved å kontorllere P_{2t} vil man også lett oppdage om puljetallene $Y_1,\dots,Y_4$ er blitt tastet inn korrekt. Man bør derfor straks kontrollere dem når man får tabellene fra maskinen for tidligst mulig å oppdage eventuelle feil.

3.2. Programforløp

Programmet er kalt AKU-NIVAA-PROG, og når dette notatet skrives, ligger det på området til bruker I-JOHNSEN i Nord-maskinen i Falbesgate.

Programmet har valg mellom to operasjoner. Man kan velge bare å liste ut tabellen over nivå- og endringstall. Dette er aktuelt om man allerede har fått beregnet nivå- og endringstallene og bare skal ha en kopi av tabellen.

Eller man kan få beregnet nye nivå- og endringstall. Skal programmet beregne nye estimater, får man spørsmål om det er første gang man kjører programmet for vedkommende kvartal. Da må maskinen først klargjøre filene med data fra tidligere kvartaler. Programmet leser nemlig dataene for forrige kvartal og forrige år fra datafilene AKU-KVARTAL-1 og AKU-KVARTAL-4. Hvis man ikke passer på å få klargjort datafilene vil dataene fra forrige kvartal og forrige år ligge på henholdsvis AKU-KVARTAL-0 og AKU-KVARTAL-3. Før kjøring må altså maskinene foreta følgende overføringer mellom filer:

AKU-KVARTAL-3 ⟶ AKU-KVARTAL-4
AKU-KVARTAL-2 ⟶ AKU-KVARTAL-3
AKU-KVARTAL-1 ⟶ AKU-KVARTAL-2
AKU-KVARYAL-0 ⟶ AKU-KVARTAL-1

Når så programmet kjøres, skriver det dataene for inneværende kvartal på filen AKU-KVARTAL-0. Disse filene ligger for tiden på området til bruker I-JOHNSEN.

En tilsvarende klargjøring av estimerte nivåtall for tidligere kvartaler blir også utført.

Hvis man etter kjøringen oppdager at det er tastet inn feil, er det mulig å foreta en omkjøring etter at dataene er rettet. Da skal programmet ikke foreta forannevnte klargjøring av filene, siden de allerede er på rett sted. Når programmet spør om det er første gangs kjøring på kvartalet, må man altså svare "nei".

Rent skjematisk er programmet bygget opp som på figur 2 neste side

4. BRUKERVEILEDNING

(Vi tror det vil lette forståelsen å studere figur 2 neste side i forbindelse med veiledningen nedenfor).

i) Før programmet AKU-NIVÅ-PROG (A-N-P) kan kjøres, må dataene for siste kvartal ligge tilgjengelig på en file med navn AKU-DATA. Det format som må brukes ser slik ut:

```
PULJEDATA FOR 2. KVARTAL 1981
TOTALT ANTALL I ARBEIDSSTYRKEN I PULJENE
 7517 7317 7416 7010
PULJEVIST ANT. SYSSELSATTE I HVER NÆRING
  372  367  363  309
   22   26   36   19
   46   35   58   32
    :    :    :    :
    :    :    :    :
    :    :    :    :
```

Det har ingen betydning hvilken tekst man bruker, bare man passer på følgende:

a) Kvartalsnummeret (i dette tilfellet 2) må komme i posisjon 15 i 1. linje på filen.

b) Årstallet (1981 i eksemplet) må komme i posisjonene 26-29 i 1. linje

c) Totalt antall i arbeidsstyrken i hver av puljene må komme på linje 3, og må være oppgitt i hele hundre. Programmet forutsetter 5 posisjoner til hver av puljene, og at hvert puljetall er høyrejustert. I eksemplet er puljetallene 4-sifret, slik at det må en blank foran. At tallene er i hele hundre vil si at det f.eks. i 1. pulje var i alt 751 500 (avrundet til nærmeste hundre) i arbeidsstyrken i vedkommende pulje. Dette tastes følgelig inn som 7515.

d) Antall sysselsatte i hver næring i hver pulje kommer så fra og med linje 5, en næring pr. linje. Igjen forutsetter programmet 5 posisjoner, høyrejustert, til hvert puljetall. Videre er programmet laget for i alt 60 næringer. Skulle antall næringer bli forandret, f.eks. ved at noen 2-sifrede næringer splittes opp på 3-sifffernivå, må programmet endres. Dette er en meget liten operasjon, da det kun er parametrene i linje 24 i programmet AKU-NIV-PROG (se vedlegg 3) som må endres. For å endre dette, skulle det være nok å søke bistand hos en som kjenner programmeringsspróget FORTRAN.

ii) Kall opp styreprogrammet med kommandoen

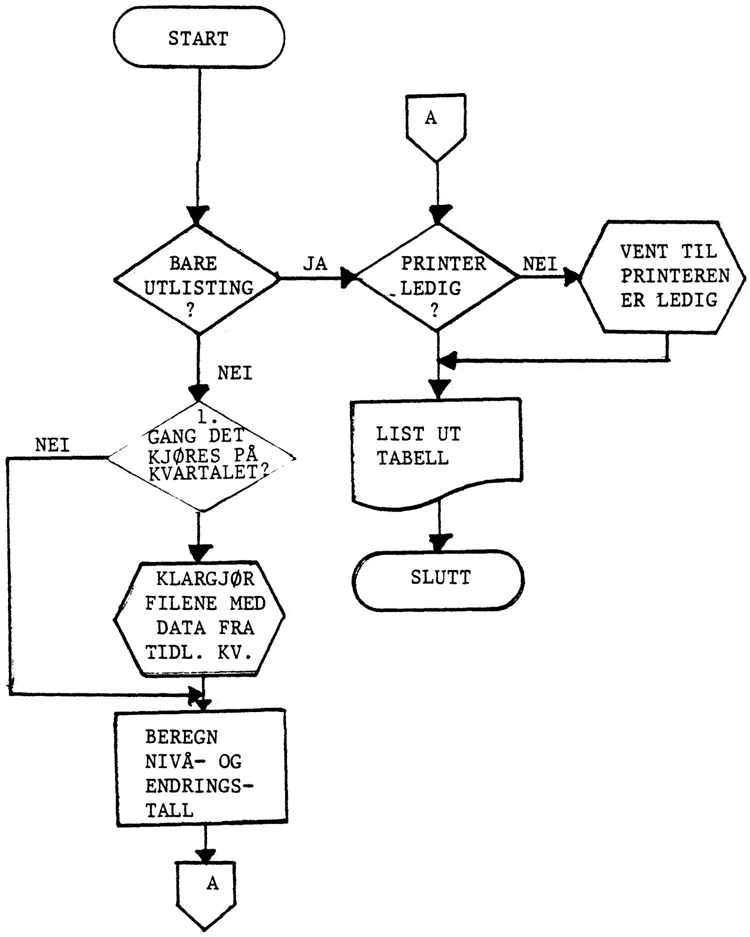
αAKU-NIV-PROG

(α er et sintrantegn som ikke skal tastes inn. Det forteller bare at maskinen venter på kommando). (Det er en underforstått forutsetning at man først har logget seg på. Vi antar brukeren har kjennskap til hvordan dette gjøres, og vil ikke komme inn på det her).

Da kommer følgende melding fram på skjermen:

Velkommen til et program for beregning av sammensatte nivå- og endringstall i AKU.

Figur 2. Flytdiagram for programmet AKU-NIVÅ-PROG



Deretter kommer spørsmålet:

Skal du bare liste ut tabellene?

Skal du bare ha en ny kopi av tabellene for siste kvartals nivå- og endringstall, svarer du JA ("J" er nok). Da blir tabellene listet ut straks. Hvis printeren er opptatt venter programmet på tur. Om printeren er opptatt med en meget lang utlisting, kan det følgelig lønne seg å avbryte programmet ved å bruke BREAK-knappen og isteden forsøke å få listet tabellen ut noe senere.

Skal det beregnes nye nivå- og endringstall, svarer du isteden med NEI eller ved å trykke på CR-knappen. Da får du spørsmålet

Er det første gang du kjører på kvartalet?

Som nevnt i avsnitt 3.2., må maskinen klargjøre datafilene hvis dette er første gangs kjøring på dataene for kvartalet. Hvis dette derimot er en omkjøring fordi dataene er blitt korrigert, skal maskinen ikke foreta denne klargjøringen, siden dette ble gjort ved forrige gangs kjøring.

Svarer man JA på spørsmålet, vil dataene fra tidligere kvartaler flyttes mellom filer. Deretter vil selve beregningene starte, og endelig vil tabellene bli utlistet. Hele operasjonen, fra man starter programmet til man får tabellene, vil normalt ikke ta mer enn 3-4 minutter, så sant belastningen på maskinen ikke er usedvanlig stor. Mesteparten av denne tiden går med til å klargjøre disketten og flytte dataene mellom de nødvendige filene.

Svarer man med å tyrkke pr CR-knappen, starter beregningen direkte. Da blir tiden redusert til under det halve.

For at brukeren til enhver tid skal vite hvilken aktivitet som maskinen holder på med, gir maskinen en melding på skjermen hver gang den starter en ny aktivitet.

NB! Hvis belastningen på maskinen er meget stor, kan man risikere at tidene som ble nevnt foran, nærmere 10-dobles. Da vi vi anbefale at man venter med å kjøre programmet til belastningen er blitt mindre.

iii) Kontroller tabellene. Dette bør du alltid huske på, siden det vil kunne medføre en del unødvendig merarbeide å kjøre tidligere kvartaler på nytt (bl.a. må man på forhånd sørge å få riktige data fra tidligere kvartaler på rett plass). Denne kontrollen gjøres meget enkelt ved at man sjekker at råestimatet for hver næring er lik summen av puljetallene for vedkommende næring. Ta også en ekstra titt på de tallene som er tastet inn for totalt antall i arbeidsstyrken i hver næring. Hvis de sistnevnte tallene er gale, vil dette kunne føre til at de beregnede nivå- og endringstall for alle næringene i vedkommende kvartal blir gale. Oppdager du feil, retter du dette på filen AKU-DATA og kjører så programmet på nytt (omkjøring).

Vedlegg 1

Tilnærmede 95 prosents konfidensavvik på nivå- og endringstall i AKU.

Til hjelp for vurderingen av estimerte nivå- og endringstall i AKU følger to tabeller for tilnærmede 95 prosents konfidensavvik i AKU. Som basis for konfidensavvikene har vi antatt at det er i alt 3 millioner personer i arbeidsstyrken og at man har 10 000 observasjoner i hvert kvartal. (For 1. kvartal 1982 var de korrekte tallene: 2 948 900 i arbeidsstyrken og i alt 10 683 observasjoner i utvalget.)

La så

Y_t = estimert antall sysselsatte i næringen ved tidspunkt t.

$\bar{Y}_t$ = Y/3 000 000 = Andelen av arbeidsstyrken som er sysselsatt i næringen.

Et tilnærmet 95 prosents konfidensintervall for antall sysselsatte har grensene

$$Y_t \pm ZN\sqrt{1{,}5\bar{Y}_t(1-\bar{Y}_t)/n} = Y_t \pm 1{,}96 \cdot 3\ 000\ 000 \sqrt{1{,}5\bar{Y}(1-\bar{Y})/10\ 000} = Y_t \pm 72\ 000 \sqrt{\bar{Y}_t(1-\bar{Y}_t)} \qquad (1)$$

hvor Z er 0,975 - fraktilen i normalfordelingen, dvs. Z = 1,96, N = 3 000 000 er antall i arbeidsstyrken, mens n = 10 000 er antall observasjoner i alt.

Konfidensintervallet i (1) bygger på en empirisk formel som etter erfaringer i andre land har vist seg å gi rimelig god tilnærmelse for mange kjennemerker. Uttrykket i (1) gir derfor antagelig ingen god tilnærmelse til den faktiske samplingsvariasjon, men den antyder størrelsesordenen av samplingsfeilen i AKU.

Tabell A nedenfor gir anslag for konfidensavviket til Y_t og P_t^* for forskjellige verdier av Y_t. P_t^* er den sammensatte nivåestimator og er på formen

$$P_t^* = 0{,}2\ 348\ (P_{t-4}^* - X_{t-4}^!) + 0{,}3\ 905\ (P_{t-1}^* - X_{t-1}^!) + (0{,}3\ 746Y_1 + 1{,}\ 1\ 557Y_2 + 0{,}8443Y_3 + 1{,}6\ 254Y_4)/4 \qquad (2)$$

hvor $Y_i = N\bar{Y}_i$ er antall sysselsatte i næringen ifølge pulje i, i = 1, 2, 3, 4, og $X_{t-1}^!$ og $X_{t-4}^!$ er antall sysselsatte i næringen ved tidspunkt t-1 og t-4 ifølge puljene som er felles for t og t-1, henholdsvis t og t-4. Standardavviket til P_t^*, SD(P_t^*), viser seg å bli

$$SD(P_t^*) = 0{,}7\ 512\ SD(Y_t),$$

slik at P_t^* får et tilnærmet 95 prosents konfidensintervall lik

$$P_t^* \pm 0{,}\ 7\ 512 \cdot 72\ 000 \sqrt{\bar{Y}(1-\bar{Y})} = P_t^* \pm 54\ 000 \sqrt{\bar{Y}(1-\bar{Y})}$$

Tabell A. Tilnærmede 95 prosents konfidensavvik for estimert antall sysselsatte i AKU. 1 000 personer

Estimert antall sysselsatte, 1 000 personer	Konfidensavvik	
	Sammensatt estimat (P^*_t)	AKU (Y_t)
0,5	0,7	0,9
1,0	1,0	1,3
1,5	1,2	1,6
2,0	1,4	1,9
3,0	1,7	2,3
5,0	2,2	2,9
7,5	2,7	3,6
10,0	3,1	4,1
15,0	3,8	5,1
20,0	4,4	5,9
30,0	5,4	7,2
50,0	6,9	9,2
70,0	8,2	10,9
100,0	9,7	12,9
150,0	11,8	15,7
200,0	13,5	18,0
300,0	16,2	21,6
500,0	20,1	26,8
700,0	22,9	30,5
1 000,0	25,5	33,9

Tabell B gir anslag for konfidensavviket til følgende endringsestimatorer:

$$\hat{d}(t-1,t) = [0,14(Y''_t - X''_{t-1}) + (Y'_t - X'_{t-1})]/1,14$$

$$\hat{d}(t-4,t) = [0,23(Y''_t - X''_{t-4}) + (Y'_t - X'_{t-4})]/1,23$$

$$\tilde{d}(t-1,t) = Y_t - X_{t-1}$$

$$\tilde{d}(t-4,t) = Y_t - X_{t-4}$$

Y''_t og X''_s (s=t-1 eller s = t-4) er det estimerte antall sysselsatte i næringen ved henholdsvis tidspunkt t og tidspunkt s ifølge de delene av utvalget som ikke er felles for de to tidspunktene, mens Y'_t og X'_s som før er antall sysselsatte ifølge den del av utvalget som går igjen ved begge tidspunkter. Y_t og X_s er det tilsvarende antall ifølge hele utvalget ved begge tidspunkter. (Se forøvrig kap. 2.)

Konfidensavvikene til estimatorene blir

$$KA(\hat{d}(t-1,t)) = 0,70 \cdot KA(Y_t) = 50\ 000 \sqrt{\bar{Y}_t(1-\bar{Y}_t)}$$

$$KA(\hat{d}(t-4,t)) = 0,87 \cdot KA(Y_t) = 62\ 000 \sqrt{\bar{Y}_t(1-\bar{Y}_t)}$$

$$KA(\tilde{d}(t-1,t)) = 1,07 \cdot KA(\bar{Y}_t) = 77\ 000 \sqrt{\bar{Y}_t(1-\bar{Y}_t)}$$

$$KA(\tilde{d}(t-4,t)) = 1,11 \cdot KA(\bar{Y}_t) = 80\ 000 \sqrt{\bar{Y}_t((1-\bar{Y}_t)}$$

Tabell B. Tilnærmede 95 prosents konfidensavvik for estimerte endringstall i AKU. 1 000 personer

Estimert antall sysselsatte i næringen, 1 000 personer	Sammensatte endringstall		Endringstall, AKU	
	Fra ett kvartal siden $\hat{d}(t-1,t)$	Fra fire kvartaler siden $\tilde{d}(t-4,t)$	Fra ett kvartal siden $\hat{d}(t-1,t)$	Fra fire kvartaler siden $\tilde{d}(t-4,t)$
0,5	0,7	0,8	1,0	1,0
1,0	0,9	1,1	1,4	1,5
1,5	1,1	1,4	1,7	1,8
2,0	1,3	1,6	2,0	2,1
3,0	1,6	2,0	2,4	2,5
5,0	2,1	2,5	3,1	3,3
7,5	2,5	3,1	3,8	4,0
10,0	2,9	3,6	4,4	4,6
15,0	3,6	4,4	5,4	5,6
20,0	4,1	5,1	6,3	6,5
30,0	5,0	6,2	7,6	7,9
50,0	6,5	8,0	9,8	10,2
70,0	7,6	9,4	11,6	12,1
100,0	9,1	11,2	13,8	14,3
150,0	11,0	13,6	16,8	17,4
200,0	12,6	15,6	19,2	19,9
300,0	15,1	18,7	23,1	24,0
500,0	18,8	23,2	28,6	29,8
700,0	21,3	26,4	32,5	33,8
1 000,0	23,8	29,4	36,2	37,7

Vedlegg 2

Rettelser til IN 80/17

Ved revisjonen av programmet som beregner sammensatte nivåtall har vi oppdaget enkelte feil i IN 80/17, noe som endrer enkelte av konklusjonene i notatet. På s. 14, midt på siden står det

For s = t-1 får vi spesielt

$$\begin{aligned}\hat{d}(t-1,t) &= \hat{p}(t) - \hat{p}(t-1)\\ &= \phi(t)\bar{Y}''(t) + [1-\phi(t)]\bar{Y}'(t)-\rho_1\bar{X}'(t-1)]\\ &= \{1-\rho_1[1-\phi(t)]\}\hat{p}(t-1)\end{aligned}$$

Det skal stå

For s = t-1 får vi spesielt

$$\begin{aligned}\hat{d}(t-1,t) &= \hat{p}(t) - \hat{p}(t-1)\\ &= \phi(t)\bar{Y}''(t) + [1-\phi(t)][\bar{Y}'(t)-\rho_1\bar{X}'(t-1)]-\{1-\rho_1[1-\phi(t)]\}\hat{p}(t-1)\end{aligned}$$

På neste side (s. 15) står det

$$\hat{p}(t) = \rho(1-\phi)\hat{p}(t-1) + R(t) \tag{4.6}$$

hvor

$$\begin{aligned}R(t) = {} & \phi\bar{Y}(1,t) + (1-\phi)\bar{Y}(2,t) + \phi\bar{Y}(3,t) + (1-\phi)\bar{Y}(4,t)\\ & - \rho(1-\phi)[\bar{X}(1,t-1) + \bar{X}(3,t-1)]\end{aligned} \tag{4.7}$$

Det skal være

$$\hat{p}(t) = \rho(1-\phi)\hat{p}(t-1) + R(t) \tag{4.6}$$

hvor

$$\begin{aligned}R(t) = {} & \{\phi\bar{Y}(1,t) + (1-\phi)\bar{Y}(2,t) + \phi\bar{Y}(3,t) + (1-\phi)\bar{Y}(4,t)\\ & - \rho(1-\phi)[\bar{X}(1,t-1) + \bar{X}(3,t-1)]\}/2\end{aligned} \tag{4.7}$$

Denne siste uteglemmelsen av faktoren 1/2 har ledet til feil også i senere uttrykk. De korrekte uttrykkene skal være

$$\kappa(1) = 0$$

$$\kappa(2) = 0$$

$$\kappa(3) = (\rho_3-\rho_1\rho_4)\phi(1-\phi)/4 \tag{4.15}$$

$$\kappa(4) = \rho_4\phi^2 + [\rho_4(1+\rho_1^2) - \rho_1(\rho_3+\rho_5)](1-\phi)^2/4$$

$$\kappa(5) = (\rho_5-\rho_1\rho_4)\phi(1-\phi)/4$$

$$\begin{aligned}\gamma(4) &= \rho_1^4(1-\phi)^4G + \{[1-\rho_1^7(1-\phi)^7][\rho_1(1-\phi)\kappa(5) + \kappa(4) + [1-\rho_1^5(1-\phi)^5]\kappa(3)\}/\{4[1-\rho_1(1-\phi)]\}\\ &= \rho_1^4(1-\phi)^4G + \{[1-\rho_1^7(1-\phi)^7](\rho_4\phi^2 + (1-\phi)^2[(\rho_4-\rho_1\rho_3) - (\rho_5-\rho_1\rho_4)\rho_1(1-\phi)])\\ &\quad + [1-\rho_1^5(1-\phi)^5](\rho_3-\rho_1\rho_4)\phi(1-\phi)\}/\{4[1-\rho_1(1-\phi)]\}\end{aligned} \tag{4.16}$$

Tabell 4 s. 18 og tabell 6 s. 21 blir derfor også gale. Vi gjengir derfor tabell 6 med de korrekte tallene.

Tabell 6. De asymptotiske verdiene for de normerte varianser til $\hat{d}(t-4,t)$ og $\tilde{d}$ $(t-4,t)$ for noen verdier av ρ_4

ρ	$\rho 4$	H'(4)	H(4)	H'(4)/H(4)
0,25	0,09	1,91	1,93	0,99
0,50	0,30	1,65	1,75	0,94
0,70	0,54	1,27	1,48	0,86
0,80	0,68	0,97	1,25	0,78
0,86	0,77	0,75	1,06	0,71
0,90	0,83	0,57	0,88	0,65
0,95	0,91	0,31	0,57	0,55

De konklusjonene som er trukket rett under tabell 6 er følgelig også gale. Spesielt står det at vi av tabell 6 kan slutte at

$$\tilde{d}'(t-4,t) = \tilde{p}(t) - \tilde{p}(t-4)$$

er den beste estimatoren for endringstallene fra tidspunkt t-4 til t blant de estimatorene notatet har sett på. Denne konklusjonen er følgelig feilaktig.

Vedlegg 3

Programlisting til programmet AKU-NIVA-PROG

På de neste sidene finner man en programlisting av programmet AKU-NIVÅ-PROG. Den eneste kommentar vi tror er nødvendig, er at parametrene i linje 24 gir antall næringer på henholdsvis 1-siffers (N1), 2-siffers (N2) og 3-siffersnivå (N3). Hvis man gjør endringer i antall næringer, må man følgelig korrigere verdien til disse parametrene. Husk også å gjøre de nødvendige korreksjoner på filen AKU-NAERINGER. AKU-NAERINGER inneholder næringskoden til de enkelte næringene på 1-, 2- og 3-siffersnivå.

```
      PROGRAM AKUNIV
* PROGRAMMET STARTER BEREGNING AV SAMMENSATTE NIVÅTALL FOR AKU,
* PLUSS TILHØRENDE RUTINER.
      CHARACTER SVAR*1
      WRITE (1,10)
      READ (1,'(A1)') SVAR
      IF (SVAR.EQ.'J') GO TO 2
      WRITE (1,20)
      READ (1,'(A1)') SVAR
      IF (SVAR.EQ.'J') GO TO 1
      CALL COPY
1     CALL NIVAA
2     CALL LIST
10    FORMAT (' Velkommen til et program for beregning av'/
     &' sammensatte nivå- og endringstall for AKU.'//
     &' Skal du bare liste ut tabellene?'/)
20    FORMAT (' Annet spørsmål gjelder hvorvidt dataene ligger på',
     &' rett plass.'/' Har du gjort opprettinger i dataene rett før',
     &' denne kjøringen?'/' (Hvis dataene på rett plass,'
     &' svar ''JA''.)'/)
      END
```

```
      SUBROUTINE COPY
* SUBRUTINEN KOPIERER DATA MELLOM FILER FOR AKU
      CALL COMND("COPY KVARTAL-4 KVARTAL-3'")
      WRITE (1,*) 'KOPIERING KVARTAL-3 => KVARTAL-4 FERDIG.'
      CALL COMND("COPY KVARTAL-3 KVARTAL-2'")
      WRITE (1,*) '    "     KVARTAL-2 => KVARTAL-3    "'
      CALL COMND("COPY KVARTAL-2 KVARTAL-1'")
      WRITE (1,*) '    "     KVARTAL-1 => KVARTAL-2    "'
      CALL COMND("COPY KVARTAL-1 KVARTAL-0'")
      WRITE (1,*) '    "     KVARTAL-0 => KVARTAL-1    "'
      CALL COMND("COPY NIVAA-4 NIVAA-3'")
      WRITE (1,*) '    "     NIVAA-3   => NIVAA-4      "'
      CALL COMND("COPY NIVAA-3 NIVAA-2'")
      WRITE (1,*) '    "     NIVAA-2   => NIVAA-3      "'
      CALL COMND("COPY.NIVAA-2 NIVAA-1'")
      WRITE (1,*) '    "     NIVAA-1   => NIVAA-2      "'
      CALL COMND("COPY NIVAA-1 NIVAA-0'")
      WRITE (1,*) '    "     NIVAA-0   => NIVAA-1      "'
      WRITE (1,*) 'Nå er datafilene klargjort.'
      RETURN
      END
```

```
      SUBROUTINE NIVAA
* VARIABELLISTE:
*   NANT2(I):ANTALL 2-SIFREDE NÆRINGER MED I SOM 1. SIFFER, I=1-9
*   NANT3(I):  "    3-   "        "      SOM HØRER INN UNDER I-TE 2-SIFFERNÆRING
*   NERNRJ(I): NÆRINGSKODE PÅ J SIFRE
*   TOTAL(J) : TOTALT ANTALL I ARB.STYRKEN VED TIDSPKT T,T-1,T-4
*   ATOT(I,J): PULJETOTAL I PULJE 1-4 FOR TIDSPKT T,T-1,T-4
*   PN(I,J)   : TOTALT ANTALL SYSSELSATTE I NÆRING I VED TIDSPUNKT:
*              J=1: INNEV. KVART J=2: FORRIGE KVART; J=3 FORRIGE ÅR
* ANTALL OG ANDEL SYSSELSATTE I PULJE I TIL NÆRING J PÅ K SIFRE, K=0-3
*   YK(I,J)  : INNEV. KVARTAL
*   XK(I,J)  : 1 KVARTAL SIDEN
*   ZK(I,J)  : 1 ÅR SIDEN
*   PK(I)    : SAMMENSATTE NIVÅTALL FOR 1 KVARTAL SIDEN
*   QK(I)    :      "          "     "  1 ÅR SIDEN
* TOTALT ANTALL PERSONER I NÆRING NR I PÅ K SIFRE IFØLGE PULJER:
*   YKA(J,I) : INNEV. KVARTAL. J=1 PULJE 1+3; J=2 PULJE 2+4
*                              J=3   "   1+2; J=4   "   3+4
*   XK(J,I)  : 1 KVART. SIDEN. J=1   "   1+3; J=2   "   2+4
*   ZK(J,I)  : 1   ÅR    "
* NIVÅ- OG ENDRINGSTALL FOR NÆRING I, J SIFRE, INNEV. KVARTAL
*  PT(1),PJ(I) : SAMMENSATT NIVÅTALLESTIMATOR
*  PT(2)       : RÅESTIMAT
*  PT(3)       : SAMMENSATT ENDRINGSTALL FRA FORRIGE KVARTAL
*  PT(4)       :     "          "      "     "     ÅR
*  PT(5)       : ENDRINGSTALL, AKU FRA FORRIGE KVARTAL
*  PT(6)       :     "       ,  "   "     "     ÅR
      PARAMETER N1=9,N2=33,N3=38
      PARAMETER M1=2+N1,M2=2+N1+N2,M3=N2+N3,M4=1+N1+N2+N3
      INTEGER NERNR2(N2),NERNR3(N3),ITID(7),NANT2(N1),NANT3(N2),
     &KVART(2,3),NY(4,M3),NTOT(4)
      REAL P(M4),P1(N1),P2(N2),P3(N3),Q(M4),Q1(N1),Q2(N2),Q3(N3),
     &X0(2),X1(2,N1),X2(4,N2),X3(4,N3),Z0(2),Z1(2,N1),Z2(4,N2),Z3(4,N3),
     &Y0(4),Y1(4,N1),Y2(4,N2),Y3(4,N3),Y0A(4),Y1A(4,N1),Y2A(4,N2),
     &Y3A(4,N3),P0T(3),P1T(N1,3),P2T(N2,3),P3T(N3,3),ATOT(4,3),TOTAL(3)
      COMMON PT(6)
      EQUIVALENCE (P(1),P0),(P(2),P1),(P(M1),P2),(P(M2),P3),
     &(Q(1),Q0),(Q(2),Q1),(Q(M1),Q2),(Q(M2),Q3)
      DATA IN1,IN2,IN3,IN4,IN5,IN6/11,12,13,14,15,16/,
     &IUT0,IUT1,IUT2/20,21,22/,X0,X1,Y0,Y1,Y2,Z0,Z1,Y0A,Y1A/252*0./,
     &P0T,P1T,P2T,P3T/243*0./,TOTAL,PT/9*0./
      WRITE (1,*) 'Beregning av nye nivåtall starter.'
      OPEN (IN1,FILE='NAERING',ACCESS='R')
      OPEN (IN2,FILE='AKU-DATA',ACCESS='R')
      OPEN (IN3,FILE='KVARTAL-1',ACCESS='R')
      OPEN (IN4,FILE='KVARTAL-4',ACCESS='R')
      OPEN (IN5,FILE='NIVAA-1',ACCESS='R')
      OPEN (IN6,FILE='NIVAA-4',ACCESS='R')
      OPEN (IUT0,FILE='KVARTAL-0',ACCESS='W')
      OPEN (IUT1,FILE='NIVAA-0',ACCESS='W')
      OPEN (IUT2,FILE='AKU-TABELL:SYMB',STATUS='UNKNOWN',ACCESS='W')
* INNLES FORDELING AV NÆRINGSKODER
40    READ (IN1,100) (NANT2(I),I=1,N1),(NANT3(I),I=1,N2),
     &(NERNR2(I),I=1,N2),(NERNR3(I),I=1,N3)
* INNLES SAMMENSATTE ESTIMATER FRA TIDL. KVARTALER
      READ (IN5) (P(I),I=1,M4)
      READ (IN6) (Q(I),I=1,M4)
* INNLES PULJETALL FOR TIDLIGERE KVARTALER
      READ (IN3,1000) (KVART(I,2),I=1,2),(ATOT(I,2),I=1,4),
     &((X2(I,J),I=1,4),J=1,N2),((X3(I,J),I=1,4),J=1,N3)
      READ (IN4,1000) (KVART(I,3),I=1,2),(ATOT(I,3),I=1,4),
     &((Z2(I,J),I=1,4),J=1,N2),((Z3(I,J),I=1,4),J=1,N3)
* INNLES PULJETALL FOR INNEVÆRENDE KVARTAL
*     READ (IN2,1000) (KVART(I,1),I=1,2),(ATOT(I,1),I=1,4),
*    &((Y2(I,J),I=1,4),J=1,N2),((Y3(I,J),I=1,4),J=1,N3)
      READ (IN2,1100) (KVART(I,1),I=1,2),(ATOT(I,1),I=1,4)
      DO 7 I=1,4
      DO 71 J=1,3
71    TOTAL(J)=TOTAL(J)+ATOT(I,J)
7     NTOT(I)=NINT(ATOT(I,1)*10.)
      K2=0
      L2=0
      L3=33
      DO 4 I1=1,9
      K1=K2+1
      K2=K2+NANT2(I1)
      DO 4 I2=K1,K2
      IF (NANT3(I2).GT.0) GO TO 1
      READ (IN2,1200) (Y2(J,I2),J=1,4)
      GO TO 3
```

```
1     L1=L2+1
      L2=L2+NANT3(I2)
      DO 2 I3=L1,L2
      READ (IN2,1200) (Y3(J,I3),J=1,4)
      L3=L3+1
      DO 2 J=1,4
      Y2(J,I2)=Y2(J,I2)+Y3(J,I3)
2     NY(J,L3)=NINT(Y3(J,I3)*10.)
3     DO 4 J=1,4
4     NY(J,I2)=NINT(Y2(J,I2)*10.)
      WRITE (IUTO,1300) (KVART(I,1),I=1,2),(NTOT(I),I=1,4),
     &((NY(I,J),I=1,4),J=1,M3)
* OPPBLÅS PULJETALLENE TIL TOTALTALL
      DO 8 I=1,N3
      Y3A(1,I)=TOTAL(1)*(Y3(1,I)+Y3(3,I))/(ATOT(1,1)+ATOT(3,1))
      Y3A(2,I)=TOTAL(1)*(Y3(2,I)+Y3(4,I))/(ATOT(2,1)+ATOT(4,1))
      Y3A(3,I)=TOTAL(1)*(Y3(1,I)+Y3(2,I))/(ATOT(1,1)+ATOT(2,1))
      Y3A(4,I)=TOTAL(1)*(Y3(3,I)+Y3(4,I))/(ATOT(3,1)+ATOT(4,1))
      DO 81 J=1,4
      P3T(I,1)=P3T(I,1)+Y3(J,I)
      P3T(I,2)=P3T(I,2)+X3(J,I)
      P3T(I,3)=P3T(I,3)+Z3(J,I)
81    Y3(J,I)=TOTAL(1)*Y3(J,I)/ATOT(J,1)
      X3(1,I)=TOTAL(2)*(X3(1,I)+X3(3,I))/(ATOT(1,2)+ATOT(3,2))
      X3(2,I)=TOTAL(2)*(X3(2,I)+X3(4,I))/(ATOT(2,2)+ATOT(4,2))
      Z3(1,I)=TOTAL(3)*(Z3(1,I)+Z3(2,I))/(ATOT(1,3)+ATOT(2,3))
8     Z3(2,I)=TOTAL(3)*(Z3(3,I)+Z3(4,I))/(ATOT(3,3)+ATOT(4,3))
      DO 9 I=1,N2
      Y2A(1,I)=TOTAL(1)*(Y2(1,I)+Y2(3,I))/(ATOT(1,1)+ATOT(3,1))
      Y2A(2,I)=TOTAL(1)*(Y2(2,I)+Y2(4,I))/(ATOT(2,1)+ATOT(4,1))
      Y2A(3,I)=TOTAL(1)*(Y2(1,I)+Y2(2,I))/(ATOT(1,1)+ATOT(2,1))
      Y2A(4,I)=TOTAL(1)*(Y2(3,I)+Y2(4,I))/(ATOT(3,1)+ATOT(4,1))
      DO 91 J=1,4
      P2T(I,1)=P2T(I,1)+Y2(J,I)
      P2T(I,2)=P2T(I,2)+X2(J,I)
      P2T(I,3)=P2T(I,3)+Z2(J,I)
91    Y2(J,I)=TOTAL(1)*Y2(J,I)/ATOT(J,1)
      X2(1,I)=TOTAL(2)*(X2(1,I)+X2(3,I))/(ATOT(1,2)+ATOT(3,2))
      X2(2,I)=TOTAL(2)*(X2(2,I)+X2(4,I))/(ATOT(2,2)+ATOT(4,2))
      Z2(1,I)=TOTAL(3)*(Z2(1,I)+Z2(2,I))/(ATOT(1,3)+ATOT(2,3))
9     Z2(2,I)=TOTAL(3)*(Z2(3,I)+Z2(4,I))/(ATOT(3,3)+ATOT(4,3))
      L2=0
      DO 13 I1=1,N1
      L1=L2+1
      L2=L2+NANT2(I1)
      DO 11 I2=L1,L2
      DO 10 J=1,2
      X1(J,I1)=X1(J,I1)+X2(J,I2)
10    Z1(J,I1)=Z1(J,I1)+Z2(J,I2)
      DO 17 J=1,3
17    P1T(I1,J)=P1T(I1,J)+P2T(I2,J)
      DO 11 J=1,4
      Y1A(J,I1)=Y1A(J,I1)+Y2A(J,I2)
11    Y1(J,I1)=Y1(J,I1)+Y2(J,I2)
      DO 12 J=1,2
      X0(J)=X0(J)+X1(J,I1)
12    Z0(J)=Z0(J)+Z1(J,I1)
      DO 14 J=1,3
14    P0T(J)=P0T(J)+P1T(I1,J)
      DO 13 J=1,4
      Y0A(J)=Y0A(J)+Y1A(J,I1)
13    Y0(J)=Y0(J)+Y1(J,I1)
* HENT DATO
      CALL CLOCK(ITID)
* BEREGN NIVÅ- OG ENDRINGSTALL FOR HVER NÆRING OG SKRIV UT RESULTATENE
      WRITE (IUT2,200) (KVART(I,1),I=1,2),(ITID(I),I=5,7),
     &((KVART(I,J),J=2,3),(KVART(I,J),J=2,3),I=1,2)
      K2=0
      L2=0
      DO 20 I1=1,N1
      WRITE (IUT2,300)
      K1=K2+1
      K2=K2+NANT2(I1)
      DO 18 I2=K1,K2
      IF (NANT3(I2).EQ.0) GO TO 16
      WRITE (IUT2,300)
      L1=L2+1
      L2=L2+NANT3(I2)
      DO 15 I3=L1,L2
```

```
      M=I3
      N=N3
      CALL NIV(X3,Y3,Z3,Y3A,P3T,P3,Q3,M,4,N)
15    WRITE (IUT2,500) NERNR3(M),(PT(I),I=1,6)
      WRITE (IUT2,300)
16    M=I2
      N=N2
      CALL NIV(X2,Y2,Z2,Y2A,P2T,P2,Q2,M,4,N)
      WRITE (IUT2,600) NERNR2(M),(PT(I),I=1,6)
      IF (NERNR2(M).EQ.33.OR.NERNR2(M).EQ.96) WRITE (IUT2,400)
18    CONTINUE
      WRITE (IUT2,300)
      M=I1
      N=N1
      CALL NIV(X1,Y1,Z1,Y1A,P1T,P1,Q1,M,2,N)
      WRITE (IUT2,700) M,(PT(I),I=1,6)
      IF (M.EQ.6) WRITE (IUT2,400)
20    CONTINUE
      WRITE (IUT2,300)
      CALL NIV(X0,Y0,Z0,Y0A,P0T,P0,Q0,1,2,1)
      WRITE (IUT2,800) (PT(I),I=1,6)
      WRITE (IUT2,400)
      WRITE (IUT1) (P(I),I=1,M4)
      CLOSE (-1)
      WRITE (1,*) 'Tabellene er ferdige.'
      RETURN
100   FORMAT (/14X,9I1/14X,33I1,2/20I3/13I3/4(/10I4))
200   FORMAT ('1NIVÅ- OG ENDRINGSTALL FOR A.K.U.',I2,'. KVARTAL',I5,
     &'. 1000 PERSONER.'/' KJØREDATO:',I3,'/',I2,'-',I4/1X,66('-')/
     &12X,'NIVÅTALL',20X,'ENDRINGSTALL'/9X,15('-'),3X,40('-')/
     &31X,'SAMMENSATT',15X,'AKU'/' NÆRING  SAMMEN-',9X,2(2X,19('-'))/
     &10X,'SATT',6X,'AKU',2(2(7X,'FRA')1X)/
     &9X,'ESTIMAT',9X,2(1X,2(I2,'. KVART.'))/23X,2(1X,2I10)/1X,66('-'))
300   FORMAT (' ')
400   FORMAT ('1')
500   FORMAT (I6,1X,2F8.1,1X,2F10.1,2X,2F10.1)
600   FORMAT (I5,2X,2F8.1,1X,2F10.1,2X,2F10.1)
700   FORMAT (I4,3X,2F8.1,1X,2F10.1,2X,2F10.1)
800   FORMAT (' TOTAL',1X,2F8.1,1X,2F10.1,2X,2F10.1)
1000  FORMAT (14X,I1,10X,I4,2/4F5.1/8(/16F4.1)/4F4.1/9(16F4.1/),8F4.1)
1100  FORMAT (14X,I1,10X,I4,2/4F5.1/1X)
1200  FORMAT (4F5.1)
1300  FORMAT ('PULJEDATA FOR ',I1,'. KVARTAL ',I4/'TOTALT ANTALL I ',
     &'ARBEIDSSTYRKEN I HVER PULJE'/4I5/'PULJEVIST ANT. SYSSELSATTE I',
     &' HVER NÆRING',8(/16I4)/4I4/9(16I4/),8I4)
      END



      SUBROUTINE NIV(X,Y,Z,YA,PN,P,Q,M,N1,N2)
* SUBRUTINEN BEREGNER NIVÅTALL
      REAL X(N1,N2),Y(4,N2),Z(N1,N2),YA(4,N2),PN(N2,3),P(N2),Q(N2)
      COMMON PT(6)
*     PT(1)=PN(M,1)
*     PT(1)=.34*YA(1,M)+.66*(YA(2,M)+.86*(P(M)-X(1,M)))
      PT(1)=.2348*(Q(M)-Z(1,M))+.3905*(P(M)-X(1,M))+(.3746*Y(1,M)+
     &1.1557*Y(2,M)+.8443*Y(3,M)+1.6254*Y(4,M))/4.
      IF (PT(1).LT.0.) PT(1)=0.
      P(M)=PT(1)
      PT(2)=PN(M,1)
      PT(3)=(YA(2,M)-X(1,M)+.14*(YA(1,M)-X(2,M)))/1.14
      PT(4)=(YA(4,M)-Z(1,M)+.23*(YA(3,M)-Z(2,M)))/1.23
      PT(5)=PN(M,1)-PN(M,2)
      PT(6)=PN(M,1)-PN(M,3)
      RETURN
      END
```

```
      SUBROUTINE LIST
* SUBRUTINEN LESER EN TABELLFIL OG SKRIVER DEN UT PÅ LINJESKRIVER
      CHARACTER REC*80
      WRITE (1,*)'Hvor mange lister ønsker du? (0 - 9)'
      READ (1,'(I1)') N
      IF (N .EQ. 0) GO TO 4
      WRITE (1,*) 'Utlistingen starter.'
      OPEN(5,FILE='LINE-PRINT',ACCESS='W',ERR=3)
      OPEN(2,FILE='AKU-TABELL',ACCESS='R')
      DO 2 I=1,N
1     READ(2,'(A80)',END=2) REC
      WRITE(5,'(A80)') REC
      GOTO 1
2     REWIND 2
      CLOSE (-1)
      WRITE (1,*) 'Utlistingen er ferdig.'
      GO TO 4
3     WRITE (1,*) 'Linjeskriveren er ikke ledig.'
      WRITE (1,*) 'Prøv å liste ut tabellene senere.'
4     RETURN
      END
```

R E F E R A N S E R

Des Raj (1968): "Sampling Theory". Mc. Graw-Hill, New York.

Østensen, Roy: "Sammensatt estimering i forbindelse med Byråets arbeidskraftundersøkelser". Statistisk Sentralbyrå, Arbeidsnotat (IN 80/17).

Nord-10 Fortran System: Reference Manual. Norsk Data A.S., Oslo, ND-60.074.01.