

# Arbeidsnotater

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

Dronningensgt. 16, Oslo-Dep., Oslo 1. Tlf. 41 38 20

IO 75/24

25. juni 1975

## PRESISJONGEVINST VED BRUK AV SAMMENSATT ESTIMERING I BYRÅETS ARBEIDSKRAFTUNDERSØKELSER

Av

John Dagsvik

### INNHold

	Side
1. Innledning .....	2
2. Autokorrelasjonen .....	3
3. Sammensatt estimering av nivåer .....	5
4. Sammensatt estimering av endringer .....	9
5. Total varians og standardavvik .....	13
6. Konklusjon .....	14
7. Variansformler .....	15
Referanser .....	16

## SAMMENDRAG

Ved estimering av befolkningens fordeling på "arbeidsmarkedtilstander" har en i tillegg til observasjoner fra det aktuelle tidspunkt også observasjoner av endringer fra tidligere tidspunkter for den delen av utvalget som er felles ved disse tidspunktene.

Ved å ta utgangspunkt i den spesielle rotasjonsplanen som benyttes i Byråets arbeidskraftundersøkelser har en studert en metode til å utnytte denne informasjonen. Vi har funnet at presisjonsgevinsten ved bruk av denne metoden kan være betydelig.

## 1. Innledning

Sammensatte estimeringsmetoder er aktuelle ved løpende undersøkelser hvor utvalget er roterende slik som ved Arbeidskraftundersøkelsen (AKU).

En sammensatt estimator er en veiet sum av estimatorer 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. Vektene er valgt slikt at den sammensatte estimator blir forventningsrett.

Når vi estimerer endringer av nivå tall i AKU ved å sammenlikne resultatene fra utvalgene på flere tidspunkter, har vi både "indirekte" og "direkte" observasjoner av endringer.

Vi har "direkte" observasjoner for den delen av utvalget som er felles ved de aktuelle tidspunktene og "indirekte" observasjoner for den resterende delen av det samlede utvalget ved disse tidspunktene. Det er derfor nærliggende å undersøke om et sammensatt estimat som bygger på begge typer estimator kan forbedre det opprinnelige "rå"-estimatet både for estimering av nivå og endringstall.

Sammensatt estimering er i bruk i flere land, blant annet i USA og Sverige. Erfaringene fra disse landene kan imidlertid ikke uten videre overføres til norske forhold da tildels utvalgsplaner og rotasjonsplaner er ulike. I USA er utvalgsplan og rotasjonsplan av samme type som den norske, men med ulik rotasjonshastighet (rotasjonsrate). Som naturlig er, har rotasjonsplanen stor betydning for gevinsten ved sammensatt estimering.

Siden det norske utvalget blir trukket i to trinn og første trinns enheter er felles fra undersøkelse til undersøkelse vil sammensatt estimering bare ha virkning på varianskomponenten innen utvalgsområdene. Dersom varianskomponenten mellom utvalgsområdene er stor vil derfor effekten av sammensatt estimering på den totale variansen bli liten.

Dette notatet er skrevet med utgangspunkt i et arbeidsnotat laget sommeren 1971 av Steinar Bjerve (1974). Bjerve har imidlertid bare delvis behandlet estimering ved utvalgsplaner av den typen som Byrået benytter da denne ikke var fastlagt dengang hans notat ble skrevet. Vi har derfor mer spesielt behandlet hvordan Byråets utvalgsplan kan utnyttes til å forbedre estimatorene og hvordan eventuelle gevinster avhenger av autokorrelasjonen. Disse studiene er basert på variansformler

vi har utledet og som er gjengitt i siste avsnitt uten bevis. Resultatene som presenteres her er bare ment å kunne gi en kort oversikt over de aktuelle metodene.

## 2. Autokorrelasjonen

Vi tenker oss at det ved tidspunkt  $t$  er knyttet en verdi  $a_t(k,j)$  til enhet nr.  $k$  i utvalgsområde nr.  $j$ . Denne verdien kan f.eks. være 1 dersom enheten har et bestemt kjennetegn og null ellers. Størrelsen  $a_t(j)$  er summen av disse  $a$ -verdier for alle enhetene i det  $j$ -te utvalgsområde og  $a_t$  er summen av disse igjen, dvs. samlet  $a_t$ -verdi for populasjonen ved tidspunkt  $t$ . (Nivået ved tidspunkt  $t$ ). La  $N_j$  være antall enheter i det  $j$ -te utvalgsområde. Vi definerer

$$(1) \quad h_j(s,t) = \frac{\sum_k (a_t(k,j) - \bar{a}_t(j))(a_s(k,j) - \bar{a}_s(j))}{(N_j - 1)}$$

og

$$(2) \quad h(s,t) = \sum_j h_j(s,t)$$

hvor  $\bar{a}_t(j)$  som vanlig betyr  $a_t(j)/N_j$ .

Autokorrelasjonen innen utvalgsområdene mellom tidspunktene  $s$  og  $t$  kan nå tilnærmet uttrykkes ved

$$(3) \quad \rho_{t,s} \approx h(s,t) / \sqrt{h(s,s) h(t,t)}.$$

Som nevnt i innledningen gir sammensatt estimering bare reduksjon av variansen innen utvalgsområdene og vi er derfor interessert i å studere autokorrelasjonen innen utvalgsområdene. For å belyse hva autokorrelasjonsfunksjonen betyr vil vi omskrive uttrykket for  $h_j$ . La  $p_j(t)$  være sannsynligheten for at en person i  $j$ -te utvalgsområde tilhører en bestemt gruppe  $G$  ved tidspunkt  $t$ . La videre  $p_j(t|s)$  være sannsynligheten for at en person er i gruppe  $G$  ved tidspunkt  $t$  gitt at personen var i gruppe  $G$  ved tidspunkt  $s$  ( $s < t$ ). Benyttes tilnærmelsen  $N_j/(N_j - 1) \approx 1$  har vi dermed

$$(4) \quad h_j(s,t) \approx p_j(s) \{p_j(t|s) - p_j(t)\}.$$

Vi kan derfor slutte at dersom  $p_j(s|t)$  gjennomsnittlig er mye større enn  $p_j(t)$  så blir autokorrelasjonen stor. For eksempel dersom sannsynligheten for å være sysselsatt ved tidspunkt  $t$  totalt sett er liten sammenliknet med sannsynligheten for å være sysselsatt ved tidspunkt  $t$ , gitt at en var sysselsatt ved tidspunkt  $s$ , er altså autokorrelasjonen for sysselsatte stor. Autokorrelasjonen er derfor et mål for "treggheten" i arbeidsmarkedbevegelsene.

På grunnlag av Vannebo's bruttostrømtabeller (Vannebo 1975) har vi beregnet estimater for autokorrelasjonen for ulike variable. I disse estimatene er det ikke tatt hensyn til stratifiseringen i utvalgsplanen.

La  $\hat{p}(t)$  betegne det relative antall personer i utvalget som tilhører en gruppe  $G$  ved tidspunkt  $t$ . La tilsvarende  $\hat{p}(s,t)$  være det relative antall personer i utvalget som tilhører  $G$  både ved tidspunkt  $s$  og  $t$  ( $t > s$ ).  $\hat{p}(t)$  og  $\hat{p}(s,t)$  er altså estimatører for  $p(t)$  og  $p(s,t) = p(s) p(t|s)$ . Ved innsetting i (3) finner vi derfor en estimator for  $\rho_{t,s}$ :

$$(5) \quad \hat{\rho}_{t,s} = \frac{\hat{p}(s,t) - \hat{p}(s)\hat{p}(t)}{\sqrt{\hat{p}(s)(1-\hat{p}(s))(1-\hat{p}(t))\hat{p}(t)}}$$

Estimatene er gjengitt i tabell 1 og 2 nedenfor.

Tabell 1. Gjennomsnittlig årlig korrelasjon for tidspunktene 1972<sup>I</sup>-73<sup>I</sup>, 72<sup>II</sup>-73<sup>II</sup> og 72<sup>III</sup>-73<sup>III</sup>

Jordbruk, skogbruk, fiske og fangst .....	0,80
Bergverk, industri, vann og kraftforsyning .....	0,80
Bygg og anlegg .....	0,75
Varehandel, hotell og restaurant .....	0,76
Transport .....	0,79
Bank, forsikring, finans og eiendomsdrift .....	0,80
Off.adm., politi og rettsvesen .....	0,80

Tabell 2. Gjennomsnittlig kvartalsvis korrelasjon for perioden 1972<sup>I</sup>-  
1974<sup>II</sup>

Jordbruk, skogbruk, fiske og fangst .....	0,83
Bergverk, industri, vann og kraftforsyning .....	0,86
Bygg og anlegg .....	0,82
Varehandel, hotell og restaurant .....	0,82
Transport .....	0,86
Bank, forsikring, finans og eiendomsdrift .....	0,92
Off.adm. politi og rettsvesen .....	0,87

Disse tallene er svært like de tilsvarende svenske korrelasjonsberegningene. Vi har ikke beregnet autokorrelasjonen for variabelen "arbeidsøkere uten arbeidsinntekt", men de svenske beregningene tyder på at denne er svært liten (mellom 0,1 og 0,2).

Byrået har et variansberegningsprogram som er laget på grunnlag av en forventningsrett varians estimator for stratifisert to-trinns utvalg. Når dette programmet tilpasses Byråets nye utvalgsplan (1975) vil det være mulig å beregne estimater for autokorrelasjonen som tar hensyn til stratifiseringen.

### 3. Sammensatt estimering av nivåer

Vi innfører ytterligere følgende notasjoner.  $\hat{a}_t$  er en estimator for  $a_t$  basert på utvalget ved tidspunkt  $t$ ,  $\hat{a}_{t,s}$  er en estimator for  $a_t$  basert på den delen av utvalget som er felles ved tidspunktene  $t$  og  $s$ . Estimatoren

$$(6) \quad \hat{d}_{t,s} = \hat{a}_{s,t} - \hat{a}_{t,s}$$

er derfor en estimator for endringen  $a_s - a_t$ .

I det følgende forutsettes at variansen innen utvalgsområdene er tilnærmet konstant over tid samt at autokorrelasjonen  $\rho_{t,s}$  mellom tidspunktene  $t$  og  $s$  bare avhenger av differansen mellom disse tidspunktene, dvs.:  $\rho_{t,s} = \rho_{s-t}$ .

La  $k$  være en konstant,  $0 < k < 1$ . Betrakt estimatoren

$$(7) \quad a_t^* = k(\hat{a}_{t-1} + \hat{d}_{t-1,t}) + (1-k)\hat{a}_t$$

Siden både  $(\hat{a}_{t-1} + \hat{d}_{t-1,t})$  og  $\hat{a}_t$  er forventningsrette estimatører for  $a_t$  blir  $a_t^*$  forventningsrett for  $a_t$ . Denne estimatoren er den enkleste av de sammensatte estimatorene i det en her bare benytter observasjoner fra to tidspunkter. Denne estimatoren er forøvrig behandlet hos Bjerve (1974).

Forholdet mellom variansene til to estimatører  $u$  og  $v$  kalles effisiensen for  $u$  og  $v$ . La  $V_1$  betegne variansen innen utvalgsområdene. Av tabell 1 kan vi se sammenhengen mellom effisiensen for  $a_t^*$  og  $\hat{a}_t$  og parametrene  $\rho_1$  og  $k$ .

Tabell 3.

$\rho_1$	0,9	0,8	0,7	0,6	0,5	0,6	0,5
$k$	0,5	0,3	0,3	0,3	0,3	0,2	0,2
$V_1(a_t^*)/V_1(\hat{a}_t)$	0,65	0,87	0,91	0,95	0,99	0,94	0,96

Isteden for estimatoren  $a_t^*$  er det mer naturlig å bruke estimatoren  $a_t^{**}$  definert ved

$$(8) \quad a_t^{**} = k(a_{t-1}^{**} + \hat{d}_{t-1,t}) + (1-k)\hat{a}_t$$

Dvs. at en istedet for å benytte "rå-estimatet" ved tidspunkt  $t-1$  benytter det sammensatte estimatet ved tidspunkt  $t-1$ . Ved tiden  $t=0$  defineres  $a_0^{**} = \hat{a}_0$ . Denne estimatoren er i bruk i de amerikanske og svenske arbeidskraftundersøkelser. Siden en her benytter observasjoner fra hele "forhistorien" er det rimelig å tro at denne estimatoren er bedre enn  $a_t^*$ . Tabell 4 viser imidlertid at selv om  $a_t^{**}$  er bedre enn  $a_t^*$  er forskjellene små unntatt når korrelasjonene er store.

Tabell 4.

k	0,3	0,4	0,4	0,4	0,3	0,3	0,3	0,2	0,2	0,5	0,5	0,5	0,5
$\rho_1$	0,8	0,8	0,7	0,6	0,7	0,6	0,6	0,6	0,5	0,8	0,7	0,8	0,9
$\rho_3$	0,8	0,7	0,6	0,6	0,6	0,6	0,5	0,5	0,4	0,7	0,6	0,8	0,8
$\rho_4$	0,8	0,7	0,6	0,5	0,6	0,5	0,5	0,5	0,4	0,7	0,6	0,8	0,8
$\rho_5$	0,7	0,6	0,5	0,5	0,5	0,5	0,4	0,4	0,3	0,6	0,5	0,7	0,7
$V_1(a_t^{***})/$													
$V_1(\hat{a}_t)$	0,82	0,81	0,90	1	0,88	0,94	0,94	0,92	0,95	0,87	1	0,87	0,72

Som kjent roterer utvalget i AKU "periodisk" slik at halvparten av utvalget er felles ett kvartal og det påfølgende kvartal, og ved ett kvartal og samme kvartal året etter. Rotasjonsplanen kan beskrives ved følgende figur:

Kvartal	1	2	3	4	1
Pulje 1	x				
2	x	x			
3		x	x		
4			x	x	
5	x			x	x
6	x	x			x
7		x	x		
8			x	x	
9				x	x
10					x

Med en slik rotasjonsplan er det naturlig å generalisere estimatoren  $a_t^*$  ved å benytte observasjonene både fra forrige kvartal og fra samme kvartal året før. En slik estimator får formen

$$(9) \quad \tilde{a}_t = p(\hat{a}_{t-4} + \hat{d}_{t-4,t}) + q(\hat{a}_{t-1} + \hat{d}_{t-1,t}) + (1 - p - q)\hat{a}_t$$

hvor  $p$ ,  $q$  og  $p + q$  er konstanter som ligger i intervallet  $[0,1]$ . Når  $p = 0$  blir  $\tilde{a}_t$  identisk med  $a_t^*$ . Tabellene 5, 6, 7 og 8 nedenfor viser at  $\tilde{a}_t$  er bedre enn både  $a_t^*$  og  $a_t^{**}$ .



Tabell 5.

 $p = 0,2, q = 0,3$ 

$\rho_1$	0,2	0,5	0,6	0,7	0,7	0,8	0,8	0,8	0,9
$\rho_3$	0,1	0,4	0,5	0,6	0,6	0,7	0,7	0,8	0,8
$\rho_4$	0	0,3	0,5	0,5	0,6	0,6	0,7	0,8	0,8
$V_1(\tilde{a}_t)/V_1(\hat{a}_t)$	1,19	1	0,92	0,89	0,86	0,83	0,80	0,78	0,74

Tabell 6.

 $p = q = 0,2$ 

$\rho_1$	0,2	0,4	0,6	0,6	0,7	0,7	0,8	0,8	0,8	0,9
$\rho_3$	0,1	0,4	0,5	0,6	0,6	0,6	0,7	0,7	0,8	0,8
$\rho_4$	0	0,3	0,5	0,5	0,5	0,6	0,6	0,7	0,8	0,8
$V_1(\tilde{a}_t)/V_1(\hat{a}_t)$	1,09	1	0,91	0,91	0,88	0,86	0,84	0,81	0,79	0,77

Tabell 7.

 $p = 0,3, q = 0,4$ 

$\rho_1$	0,2	0,5	0,6	0,7	0,8	0,8	0,8	0,9	0,9	0,9
$\rho_3$	0,1	0,4	0,5	0,6	0,7	0,8	0,8	0,8	0,8	0,9
$\rho_4$	0	0,3	0,5	0,6	0,7	0,7	0,8	0,7	0,8	0,8
$V_1(\tilde{a}_t)/V_1(\hat{a}_t)$	1,45	1,09	1	0,91	0,82	0,83	0,79	0,77	0,73	0,74

Tabell 8.

 $p = q = 0,3$ 

$\rho_1$	0,2	0,6	0,7	0,8	0,8	0,8	0,9	0,9	0,9
$\rho_3$	0,1	0,5	0,6	0,7	0,8	0,8	0,8	0,8	0,9
$\rho_4$	0	0,5	0,6	0,7	0,7	0,8	0,7	0,8	0,8
$V_1(\tilde{a}_t)/V_1(\hat{a}_t)$	1,33	0,95	0,88	0,81	0,81	0,77	0,77	0,73	0,74

På grunnlag av tabell 1 og 2 kan en beregne variansreduksjonen innen utvalgsområdene for de enkelte næringer.

Tabell 9. Variansreduksjon innen utvalgsområder. Nivå tall

Jordbruk, skogbruk, fiske og fangst .....	24%
Bergverk, industri, vann og kraftforsyning .....	25%
Bygg og anlegg .....	22%
p=q=0,3 Varehandel, hotell og restaurant .....	22%
Transport .....	25%
Bank, forsikring, finans og eiendomsdrift .....	27%
Off.adm. politi og rettsvesen .....	25%

Tilsvarende til  $a_t^{xx}$  er estimatoren

$$\hat{a}_t = p(\hat{a}_{t-4} + \hat{d}_{t-4,t}) + q(\hat{a}_{t-1} + \hat{d}_{t-1,t}) + (1-p-q)\hat{a}_t$$

med

$$\hat{a}_0 = \hat{a}_0 \text{ og } \hat{a}_t = a_t^{xx} \text{ for } t=1,2,3,$$

en naturlig generalisering av  $\hat{a}_t$ . Vi har ikke studert denne estimatoren nærmere, men ved å sammenlikne  $\hat{a}_t^{xx}$  med  $a_t^{xx}$  syns det rimelig å anta at en ved bruk av  $\hat{a}_t$  muligens vil redusere variansen med 3-4 % i forhold til bruk av  $\hat{a}_t$ .

#### 4. Sammensatt estimering av endringer

Den enkleste måten å estimere endringer på er å ta differansen mellom de respektive nivå tallestimater. Vi skal følgelig sammenlikne variansene til estimatorene

$$(11) \quad \hat{a}_t - \hat{a}_{t-r}, \text{ for } r=1, 4$$

og

$$(12) \quad \hat{a}_t - \hat{a}_{t-r}, \text{ for } r=1,4.$$

Ved estimering av kvartalsvise endringer gir tabell 5 og 6 følgende effisienser:

Tabell 10.

$p = 0,2, q = 0,3$

$\rho_1$	0,9	0,8	0,7	0,7	0,6	0,6
$\rho_3$	0,8	0,7	0,6	0,6	0,6	0,5
$\rho_4$	0,8	0,7	0,6	0,5	0,5	0,5
$\rho_5$	0,7	0,6	0,5	0,5	0,5	0,4
$V_1(\tilde{a}_t - \tilde{a}_{t-1})/V_1(\hat{a}_t - \hat{a}_{t-1})$	0,76	0,83	0,89	0,93	0,96	0,95

Tabell 11.

$p = q = 0,2$

$\rho_1$	0,8	0,7	0,7	0,6	0,6	0,5
$\rho_3$	0,7	0,6	0,6	0,6	0,5	0,5
$\rho_4$	0,7	0,6	0,5	0,5	0,5	0,4
$\rho_5$	0,6	0,5	0,5	0,5	0,4	0,4
$V_1(\tilde{a}_t - \tilde{a}_{t-1})/V_1(\hat{a}_t - \hat{a}_{t-1})$	0,83	0,88	0,92	0,95	0,94	0,99

Tabell 12.

$p = q = 0,3$

$\rho_1$	0,9	0,9	0,8	0,7	0,7	0,6
$\rho_3$	0,8	0,8	0,7	0,6	0,6	0,6
$\rho_4$	0,8	0,7	0,7	0,6	0,5	0,5
$\rho_5$	0,7	0,7	0,6	0,5	0,5	0,5
$V_1(\tilde{a}_t - \tilde{a}_{t-1})/V_1(\hat{a}_t - \hat{a}_{t-1})$	0,73	0,81	0,84	0,93	0,99	1,02

Tabellene 13, 14 og 15 viser gevinster ved estimering av årlige endringer.

Tabell 13.

$p = 0,2, q = 0,3$

$\rho_1$	0,9	0,9	0,8	0,7	0,7	0,6
$\rho_3$	0,8	0,8	0,7	0,6	0,6	0,6
$\rho_4$	0,8	0,7	0,7	0,6	0,5	0,5
$\rho_5$	0,7	0,7	0,6	0,5	0,5	0,5
$V_1(\tilde{a}_t - \tilde{a}_{t-4})/V_1(\hat{a}_t - \hat{a}_{t-4})$	0,68	0,75	0,79	0,86	0,90	0,96

Tabell 14.

p = q = 0,2

$\rho_1$	0,9	0,9	0,8	0,7	0,7	0,6
$\rho_3$	0,8	0,8	0,7	0,6	0,6	0,6
$\rho_4$	0,8	0,7	0,7	0,6	0,5	0,5
$\rho_5$	0,7	0,7	0,6	0,5	0,5	0,5
$V_1(\tilde{a}_t - \tilde{a}_{t-4}) / V_1(\hat{a}_t - \hat{a}_{t-4})$	0,70	0,76	0,84	0,88	0,92	0,95

Tabell 15.

p = q = 0,3

$\rho_1$	0,9	0,8	0,9	0,8	0,7	0,7	0,6
$\rho_3$	0,8	0,8	0,8	0,7	0,6	0,6	0,6
$\rho_4$	0,8	0,8	0,7	0,7	0,6	0,5	0,5
$\rho_5$	0,7	0,7	0,7	0,6	0,5	0,5	0,5
$V_1(\tilde{a}_t - \tilde{a}_{t-4}) / V_1(\hat{a}_t - \hat{a}_{t-4})$	0,66	0,71	0,73	0,76	0,85	0,91	0,96

Med autokorrelasjoner av størrelsesordner som i tabell 1 og 2 viser tabellene overfor at p=q=0,3 er nær de optimale verdier av p og q både for estimering av nivå- og endringstall. På grunnlag av tabell 1 og 2 finner vi følgende variansreduksjonstabell.

Tabell 16. Variansreduksjon innen utvalgsområder. Årlige endringer

Jordbruk, skogbruk, fiske og fangst .....	30%
Bergverk, industri, vann og kraftforsyning .....	32%
Bygg og anlegg .....	25%
p=q=0,3 Varehandel, hotell og restaurant .....	25%
Transport .....	32%
Bank, forsikring, finans og eiendomsdrift .....	34%
Off.adm. politi og rettsvesen .....	33%

Dersom en gir avkall på kravet om at estimatet for endringer skal stemme overens med differensen mellom de respektive nivå-tall-estimatene vil estimatoren

$$(13) \quad \hat{\delta}_{t-r,t} = (1-k)(\hat{a}_t - \hat{a}_{t-r}) + k\hat{d}_{t-r,t}, \quad r = 1,4,$$

være bedre enn estimatorene  $\hat{a}_t - \hat{a}_{t-r}$  og  $\hat{a}_t - \hat{a}_{t-r}$  noe som best illustreres ved tabell 17 og 18 nedenfor.

Tabell 17.

k	0,4	0,5	0,5	0,6	0,6	0,6	0,6	0,7	0,7	0,7
$\rho_r$	0,5	0,6	0,7	0,6	0,7	0,8	0,9	0,7	0,8	0,9
$V_1(\hat{\delta}_{t-r,t})/V_1(\hat{a}_t - \hat{a}_{t-r})$	0,89	0,82	0,71	0,85	0,71	0,56	0,39	0,74	0,56	0,35

Tabell 18. Variansreduksjon innen utvalgsområdene ved bruk av  $\hat{\delta}$ 

	Kvartals- vise endringer	Årlige endringer
Jordbruk, skogbruk, fiske og fangst .....	50%	44%
Bergverk, industri, vann og kraftfor- syning .....	56%	44%
Bygg og anlegg .....	50%	37%
k=0,7 Varehandel, hotell og restaurant .....	50%	37%
Transport .....	56%	44%
Bank, forsikring, finans .....	70%	44%
Off.adm. politi og rettsvesen .....	59%	44%

Vi ser at allerede ved korrelasjoner av størrelsesorden 0,6 er gevinsten betydelig.

Estimatoren

$$(14) \quad a_t^o = a_{t-1}^o + \hat{\delta}_{t-1,t}$$

er derfor et nærliggende alternativ til de nivå-tallestimatorene vi har behandlet tidligere. Det viser seg imidlertid at denne estimatoren har

vesentlig større varians enn estimatoren  $\hat{a}_t$  slik at den ikke er noe brukbart alternativ.

Ett praktisk eksempel illustrerer at estimatene beregnet ved den siste metoden kan bli vesentlig forskjellig fra estimatene beregnet ved (11). I AKU var nedgangen i antall sysselsatte i jordbruk fra 1. kvartal 1973 til 1. kvartal 1974 estimert til 30 000, dvs.  $\hat{a}_t - \hat{a}_{t-4} = -30\ 000$ . Ved å bruke observasjonene fra den delen av utvalget som var felles ved de to tidspunktene fikk en et estimat på 18 000, dvs.  $\hat{d}_{t-4,t} = -18\ 000$ . Det sammensatte estimatet beregnet ved (13) med  $k=0,7$  blir følgelig lik 21 600. Ved å benytte resultatene foran kan vi konkludere at dette sammensatte estimatet er bedre enn de to enkle estimatene.

#### 5. Total varians og standardavvik

Som nevnt er det bare variansen innen utvalgsområdene som reduseres ved bruk av sammensatt estimering.

Dersom varianskomponenten mellom utvalgsområdene er stor vil derfor gevinsten ved bruk av denne metoden bli redusert.

Tabellene nedenfor viser reduksjonen av total varians og standardavvik når variansen innen utvalgsområdene reduseres med henholdsvis 25 prosent og 50 prosent.

Tabell 19. Variansreduksjon innen utvalgsområder

	Variansen innen utvalgsområdene	Variansen mellom utvalgsområdene	Variansreduksjon	Standardavvik reduksjon
25%	50%	50%	12%	6%
	60%	40%	15%	8%
	70%	30%	17%	9%
	80%	20%	20%	11%
50%	50%	50%	25%	13%
	60%	40%	30%	16%
	70%	30%	35%	19%
	80%	20%	40%	22%

Vi har foreløpig ikke estimater av varianskomponenten innen og mellom utvalgsområdene for den nye utvalgsplanen (1975) men det er rimelig å tro at varianskomponenten innen utvalgsområdene er mellom 20 prosent og 50 prosent av den totale varians.

## 6. Konklusjon

For estimering av nivåer peker estimatoren  $\hat{a}_t$  seg ut som den beste av de estimatorene vi har studert. Ved bruk av denne estimatoren vil en kunne redusere variansen innen utvalgsområdene med ca. 22-30 prosent ved estimering av sysselsatte etter næring. Dersom en skulle oppnå den samme presisjonsgevinst ved å øke utvalget måtte en altså øke utvalget (innen utvalgsområdene) fra 6 000 husholdninger til ca. 8 000 husholdninger. Ved estimering av antall arbeidssøkere uten arbeidsinntekt kan sammensatt estimering medføre betydelig variansøkning slik at det for denne variabelen er best å bruke de vanlige "råestimater". Dermed oppstår problemet med at en ikke nødvendigvis vil få samsvar i tabellene. Dette problemet kunne imidlertid tenkes løst ved en passende justering av estimatene.

Ved estimering av endringer er  $\hat{\delta}_{t-r,t}$  den beste estimatoren, men her har en det problemet at estimatene sjelden vil være lik differensen mellom nivåtallestimatene. Eksemplet i avsnittet 4 viser at forskjellen mellom estimatene ved de to estimeringsmetodene kan bli stor. Dersom en velger å bruke  $\hat{a}_t - \hat{a}_{t-r}$  til å estimere endringer viser tabell 16 at variansen mellom utvalgsområdene reduseres med 25-30 prosent. Tabellene viser videre at de optimale verdier av vektene p og q ligger nær  $p = q = 0,3$  både for estimering av nivå og endringer. (Unntatt "arbeidssøkere uten arbeidsinntekt").

## 7. Variansformler

De variansformlene vi har utledet og som ligger til grunn for tabellberegningene er gjengitt uten bevis i dette avsnittet.

$$(15) \quad V_1(a_t^{\times})/V_1(\hat{a}_t) \approx 1 + 2k^2 - k(1+k)\rho_1,$$

$$(16) \quad V_1(a_t^{\times\times})/V_1(\hat{a}_t) \approx \frac{1-k+2k^2}{1-k} - \frac{k(1+k)}{1-k} \rho_1 + \frac{k^3}{2} (1-k^2)\rho_3 \\ + \frac{k^5}{2} (1+k) \left\{ \frac{2-k}{1-k} \rho_4 - \rho_5 \right\},$$

$$(17) \quad V_1(\tilde{a}_t)/V_1(\hat{a}_t) \approx 1 + 2p^2 + 2q^2 - q(1+q)\rho_1 \\ + \frac{pq}{2} \rho_3 - p(1+p)\rho_4,$$

$$(18) \quad V_1(\hat{a}_t - \hat{a}_{t-r})/V_1(\hat{a}_t) \approx 2 - \rho_r \quad r=1,4$$

$$(19) \quad V_1(\tilde{a}_t - \tilde{a}_{t-1})/V_1(\hat{a}_t) \approx 2 + 4p^2 + 2q^2 - (1+2p^2+q)\rho_1 \\ + \frac{p}{2} (1+p+q)\rho_3 - 2p(1+p)\rho_4 + \frac{pq}{2} \rho_5,$$

$$(20) \quad V_1(\tilde{a}_t - \tilde{a}_{t-4})/V_1(\hat{a}_t) \approx 2 + 2p^2 + 4q^2 - 2q(1+q)\rho_1 \\ + \frac{q}{2} (1+p+q) (\rho_3 + \rho_5) - (1+2p+2q^2)\rho_4,$$

$$(21) \quad V_1(\hat{\delta}_{t-r,t})/V_1(\hat{a}_t) \approx 2(1+k^2) - (1+k)^2 \rho_r.$$



Referanser

- Bjerve, Steinar 1974: Sammensatt estimering ved roterende utvalg. Matematisk-statistiske problemer knyttet til arbeidskraftundersøkelsen. Statistisk Sentralbyrå, Arbeidsnotat (IO 74/36).
- Dagsvik, John 1974: Variansestimering for nivåtallestimater og endringstallestimater ved Byråets arbeidskraftundersøkelser. Statistisk Sentralbyrå, Arbeidsnotat (IO 74/50).
- Dahmstrøm, P. og Malmberg, S. 1971: Utredningsarbeidet rörande sammensatta skattningar i AKU. Statistiska Centralbyrån, Utredningsinstitutet (PM 25/5-71).
- Malmberg, S. 1970: Korrelasjonsberäkningar i arbetskraftundersökningarna (AKU) 1970-1971. Statistiska Centralbyrån, Utredningsinstitutet, Rapport 26/3-1971.
- U.S. Department of Commerce, Bureau of the Census 1963: The current population survey. A report on methodology. Technical paper no. 7.
- Vannebo, Olav 1975: Bruttostrømmer i arbeidsmarkedet 1. kvartal 1972 - 1. kvartal 1974. Statistisk Sentralbyrå, Arbeidsnotat (IO 75/2).