Arbeidsnotater

STATISTISK SENTRALBYRÅ

IO 65|2 Oslo, 16. november 1965

Varehandelsstatistikken

Ny estimeringsmetode 1963 - alternativ metode og noen generelle kommentarer

av Hans Olav Egede Larssen

Innhold

Dette notat bør leses i direkte tilknytning til arbeidsnotat IO 64|4 av 11|6-64, hvor nærmere beskrivelse av situasjonen er gitt.

Notatet faller i 2 deler. Under 1. undersøkes en brøkestimat-variant av "korrigerte gjennomsnitts metode". Under 2. kommer en viktig kommentar vedrørende skjevheter som kan oppstå ved summering over alle omsetningsgrupper, og særlig i forbindelse med "ukorrigerte gjennomsnitts metode".

1. En brøkestimat-variant av "korrigerte gjennomsnitts metode"

1.1. Begrunnelse

Hensikten er som før å finne best mulige estimater for hvert aktuelt kjennetegn på undersøkelsestidspunkt.

Ved "korrigerte gjennomsnitts metode" ble brukt estimatoren

$$\hat{\bar{B}}_h = \frac{\hat{\xi}_h}{\hat{\xi}} \cdot \bar{y} = K_h \cdot \bar{y}$$

dvs.: Utvalgsgjennomsnitt innen omsetningsgruppe for hovedgruppe ble multiplisert med en "korreksjonsfaktor" for hver undergruppe (innen omsetningsgruppe). Denne faktor var forholdet mellom anslått forventning av totalomsetning innen omsetningsgruppe for undergruppe og tilsvarende størrelse for hovedgruppe - hvor forventning er i relasjon til "bakenforliggende" fordeling for totalomsetning.

Men estimatoren kan også skrives

$$\hat{\bar{B}}_h = \hat{\xi}_h \cdot \frac{\bar{y}}{\hat{\xi}}$$

og kan da oppfattes på en litt annen måte: Undergruppe-forventning for totalomsetning multipliseres med (et overslag over) forholdet mellom gjennomsnitt for det annet kjennetegn og gjennomsnittlig totalomsetning innen hovedgruppen, altså et anslag over $\frac{\eta}{\xi}$. Men denne størrelse kan også estimeres ved forhold mellom bare utvalgsgjennomsnitt, $\frac{\bar{y}}{\bar{x}}$. Og er x-er og y-er tilstrekkelig sterkt positivt korrelert, vil det - etter den vanlige teori for brøkformede estimatorer - være rimelig å vente at dette er et bedre estimat for $\frac{\eta}{\xi}$ enn $\frac{\bar{y}}{\hat{\xi}}$. Som estimator for $\bar{B}_h$ foreslås derfor:

$$\underline{\hat{\bar{B}}_h^{*} = \hat{\xi}_h \cdot \frac{\bar{y}}{\bar{x}}}$$

I praksis vil det oftest være slik at registeropplysninger og tall fra tellingen refererer til forskjellige tidspunkter. Men dette innebærer at $\bar{x}$ ikke kan finnes (hvis det da ikke direkte er spurt etter omsetning på siste revisjonstidspunkt for registeret).

Derimot kjennes utvalgsgjennomsnitt for omsetning på tellingstidspunkt. Dette betegnes med $\bar{z}$.

Følgende betegnelser brukes:

	Gjennomsnitt for totalomsetning på revisjons-tidspunkt	tellings-tidspunkt	kjennetegn som skal **undersøkes på tellingstidspunkt**
"Bakenforliggende" populasjon	ξ	θ	η
Faktisk, endelig, populasjon	$\bar{A}$		$\bar{B}$
Utvalg	$\bar{x}$	$\bar{z}$	$\bar{y}$
Varianser	σ	ω^2	τ^2

Tilsvarende betegnelser blir brukt innen hvert stratum, altså ξ_h, θ_h, η_h osv.

I estimatoren $\bar{B}_h^{*}$ blir nå $\bar{x}$ erstattet med $\bar{z}$. Derved fås en estimator som skal undersøkes nærmere:

$$\underline{\bar{B}^{*} = \xi_h \cdot \frac{\bar{y}}{\bar{z}}}$$

(Her sees det bort fra at ξ_h bestemmes ut fra registeret ved anslaget $\hat{\xi}_h$).

1.2. Bruttovariansen

Som mål for avvikelse fra $\bar{B}_h$ skal bruttovariansen finnes.

Først bestemmes forventet kvadrert avvik fra $\bar{B}_h$ for et gitt stratum, h (undergruppe nr. h, innen omsetningsgruppe nr. g, hvor g-en er utelatt).

$$E_h\left(\xi_h \cdot \frac{\bar{y}}{\bar{z}} - \bar{B}_h\right)^2 = E_h\left[\xi_h\left(\frac{\bar{y}}{\bar{z}} - \frac{\eta}{\theta}\right) - (\bar{B}_h - \eta_h) - \left(\eta_h - \xi_h \cdot \frac{\eta}{\theta}\right)\right]^2$$

$$\approx \xi_h^2\, E_h\left(\frac{\bar{y}}{\bar{z}} - \frac{\eta}{\theta}\right)^2 + E_h(\bar{B}_h - \eta_h)^2 - 2\,\xi_h E_h\left(\frac{\bar{y}}{\bar{z}} - \frac{\eta}{\theta}\right)(\bar{B}_h - \eta_h) + \left(\eta_h - \xi_h \cdot \frac{\eta}{\theta}\right)^2$$

(De øvrige ledd forsvinner helt eller tilnærmet under forventningstegnet fordi $E\,\bar{B}_h = \eta_h$ og $E\left(\frac{\bar{y}}{\bar{z}}\right) \quad \frac{\eta}{\theta}$).

Nå gjøres tilnærmelsen

$$\theta \cdot \frac{\bar{y}}{\bar{z}} \approx \bar{y} + \frac{\eta}{\theta}(\theta - \bar{z})$$

og da blir videre:

$$E_h(\xi_h \cdot \frac{\bar{y}}{\bar{z}} - \bar{B}_h)^2 \approx \frac{\xi_h^2}{\theta^2} E_h(\bar{y} - \frac{\eta}{\theta} \cdot \bar{z})^2 + E_h(\bar{B}_h - \eta_h)^2$$

$$- 2 \frac{\xi_h}{\theta} E_h\left[(\bar{y} - \eta) - \frac{\eta}{\theta}(\bar{z} - \theta)\right](\bar{B}_h - \eta_h) + (\eta_h - \xi_h \cdot \frac{\eta}{\theta})^2$$

$$= \frac{\xi_h^2}{\theta^2}\left[\text{var } \bar{y} - 2 \frac{\eta}{\theta} \text{ cov } (\bar{y}, \bar{z}) + (\frac{\eta}{\theta})^2 \text{ var } \bar{z}\right] + \text{var } \bar{B}_h$$

$$- 2 \frac{\xi_h}{\theta}\left[\text{cov } (\bar{y}, \bar{B}_h) - \frac{\eta}{\theta} \text{ cov } (\bar{z}, \bar{B}_h)\right]$$

$$+ \left[(\eta_h - \eta) - \frac{\eta}{\theta}(\xi_h - \xi) + \frac{\eta}{\theta}(\xi - \theta)\right]^2$$

$$= \frac{1}{n} \cdot \frac{\xi_h^2}{\theta^2}\left[\tau^2 - 2 \frac{\eta}{\theta} \cdot \rho' \tau \omega + (\frac{\eta}{\theta})^2 \cdot \omega^2\right] + \frac{\tau_h^2}{N_1}$$

$$- 2 \cdot \frac{1}{N} \cdot \frac{\xi_h}{\theta} (\tau_h^2 - \frac{\eta}{\theta} \rho'_h \tau_h \omega_h)$$

$$+ \left[(\eta_h - \eta) - \frac{\eta}{\theta}(\xi_h - \xi) + \frac{\eta}{\theta}(\xi - \theta)\right]^2$$

ρ' og ρ'_h betegner korrelasjon mellom undersøkt kjennetegn og totalomsetning på tellingstidspunktet, henholdsvis totalt (innen omsetningsgruppe nr. g) og for stratum nr. h. τ_h^2, ω_h^2, er varianser innen stratum nr. h.

Som i tidligere notat forutsettes her at alle strata er like store. Det totale antall enheter, N_1 er fordelt på L strata a N_1 enheter. Tilsvarende gjelder for antall utvalgsenheter, n og n_1, som tenkes trukket under proporsjonal allokering.

Forventet avvikelse tatt over alle strata skal så finnes:

$$E_2 = E\, E_h\, (\xi_h \cdot \frac{\bar{y}}{\bar{z}} - \bar{B}_h)^2$$

Som i det tidligere notat innføres

$$\tau_w^2 = E\, \tau_h^2 \qquad \tau_b^2 = E\, (\eta_h - \eta)^2$$

$$(\tau_w^2 + \tau_b^2 = \tau^2)$$

og andre tilsvarende betegnelser. Da fåes, med lignende tilnærmelser som før i ledd av orden $(\frac{1}{N})$:

$$\underline{E_2 = \frac{1}{n} \cdot \frac{\xi^2 + \sigma_b^2}{\theta^2}\left[\tau^2 - 2 \cdot \frac{\eta}{\theta} \rho' \tau \omega + (\frac{\eta}{\theta})^2 \omega^2\right] + \frac{\tau_w^2}{N_1}}$$

$$\underline{- 2 \cdot \frac{1}{N} \cdot \frac{\xi}{\theta} (\tau_w^2 - \frac{\eta}{\theta} \rho'_w \tau_w \omega_w)}$$

$$\underline{+ \tau_b^2 - 2 \cdot \frac{\eta}{\theta} \cdot \rho'_b \sigma_b \tau_b + (\frac{\eta}{\theta})^2 \cdot \sigma_b^2 + (\frac{\eta}{\theta})^2 (\xi - \theta)^2}$$

1.3. Sammenligning mellom de to varianter av "korrigerte gjennomsnitts metode". Tabeller.

For $\hat{\bar{B}}_h = \frac{\xi_h}{\xi} \cdot \bar{y}$ var

$$E_1 = E\left(\frac{\xi_h}{\xi} \cdot \bar{y} - \bar{B}_h\right)^2 = \left(1 + \frac{\sigma_b^2}{\xi^2}\right) \cdot \frac{\tau^2}{n} + \frac{\tau_w^2}{N_1} - 2 \cdot \frac{\tau_w^2}{N}$$

$$+ \tau_b^2 - 2\,\frac{\eta}{\xi} \cdot \rho_b\, \sigma_b\, \tau_b + \left(\frac{\eta}{\xi}\right)^2 \cdot \sigma_b^2$$

Betingelsen for at brøkestimatet $\bar{B}_h^{*}$ skal være bedre enn $\hat{\bar{B}}_h$ kan skrives: $E_2 - E_1 < 0$, og dette gir:

$$- \frac{\tau^2}{n} + \frac{1}{n} \cdot \frac{\xi^2}{\theta^2}\left(\tau^2 - 2\,\frac{\eta}{\theta}\,\rho'\,\tau\,\omega + \frac{\eta^2}{\theta^2} \cdot \omega^2\right) + \frac{1}{n} \cdot \frac{\sigma_b^2}{\theta^2}\left(-2 \cdot \frac{\eta}{\theta}\,\rho'\,\tau\,\omega + \frac{\eta^2}{\theta^2} \cdot \omega^2\right)$$

$$- 2 \cdot \frac{1}{N}\left(\frac{1}{\theta} - \frac{1}{\xi}\right) \cdot \xi \cdot \tau_w^2 + 2 \cdot \frac{1}{N} \cdot \frac{\xi}{\theta} \cdot \frac{\eta}{\theta} \cdot \rho_w\, \tau_w\, \omega_w + \frac{\eta^2}{\theta^2}(\xi - \theta)^2 < 0$$

Nå antas at forholdet mellom "gjennomsnittlig" varians innen stratum og total varians er av samme størrelsesorden for undersøkt kjennetegn som for omsetning på undersøkelsestidspunkt. Dvs.:

$$\frac{\tau_w^2}{\tau^2} = \frac{\omega_w^2}{\omega^2} = \psi^2$$

Etter innføring av

$$C' = \frac{\frac{\omega}{\theta}}{\frac{\tau}{\eta}}$$

fåes da, idet $\frac{n}{N} = \alpha$ = utvalgsbrøken, felles for alle strata:

$$\frac{\tau^2}{n}\Bigg[-1 + \frac{\xi^2}{\theta^2}\left(1 - 2\,C'\rho' + C'^2\right) + \frac{\sigma_b^2}{\theta^2}\left(-2\,C'\rho' + C'^2\right)$$

$$- 2\,\alpha\left(\frac{1}{\theta} - \frac{1}{\xi}\right) \cdot \xi \cdot \psi^2 + 2\,\alpha \cdot \frac{\xi}{\theta} \cdot C' \cdot \psi^2 \cdot \rho'_w + n\,C'^2 \cdot \frac{(\xi - \theta)^2}{\omega^2}\Bigg] \qquad 0$$

Her kan bestemmes hvilke restriksjoner ulikheten legger på verdisett av ρ'_w og ρ'.

I det følgende betraktes det tilfelle at forholdet mellom covarians innen stratum og total covarians er lik det tilsvarende forhold for variansene, altså:

$$\frac{\rho'_w\, \omega_w\, \tau_w}{\rho'\, \omega\, \tau} = \psi^2$$

Dette gir

$$\frac{\rho'_w}{\rho'} = \psi^2 \cdot \frac{\omega\,\tau}{\omega_w \tau_w} = \psi^2 \cdot \frac{1}{\psi} \cdot \frac{1}{\psi} = 1$$

dvs. $\rho'_w = \rho'$

altså at "gjennomsnittlig" korrelasjon innen stratum er lik total korrelasjon. Det er ikke urimelig at dette kan gjelde med tilnærmelse for strata som alle har felles grense oppad og nedad etter størrelsen av en variabel, totalomsetning.

Ulikheten vil da gi at

$$\rho' > \frac{\frac{\xi^2}{\theta^2} - 1 - 2\alpha\left(\frac{\xi}{\theta} - 1\right)\cdot\psi^2 + c'^2\left(\frac{\xi^2}{\theta^2} + \frac{\sigma_b^2}{\theta^2} + n\,\frac{(\xi - \theta)^2}{\omega^2}\right)}{2\,c'\left(\frac{\xi^2}{\theta^2} + \frac{\sigma_b^2}{\theta^2} - \alpha\cdot\frac{\xi}{\theta}\cdot\psi^2\right)}$$

Anta nå at endringene mellom registerrevisjon og tellingstidspunkt er <u>relativt</u> små, altså at $\frac{\xi}{\theta} \approx 1$. Da blir leddene som inneholder $\frac{\xi^2}{\theta^2} - 1$ og $\frac{\xi}{\theta} - 1$ små i forhold til andre ledd og kan derfor sløyfes. Derimot blir $\frac{(\xi - \theta)^2}{\omega^2} = \left(\frac{\xi}{\theta} - 1\right)^2 \cdot \frac{\theta^2}{\omega^2}$ multiplisert med n og vil derfor lett spille inn all den stund n må forutsettes å være relativt stor. I det undersøkte, konkrete tilfelle er også $\frac{\theta^2}{\omega^2}$ av størrelsesorden 10 og oppover. Leddet $n \cdot \frac{(\xi - \theta)^2}{\omega^2}$ må derfor beholdes, og man får:

$$\rho' > \frac{c'}{2} \cdot \frac{1 + \frac{\sigma_b^2}{\theta^2} + n \cdot \frac{(\xi - \theta)^2}{\omega^2}}{1 + \frac{\sigma_b^2}{\theta^2} - \alpha \cdot \frac{\xi}{\theta} \cdot \psi^2}$$

Man kan skrive: $\frac{\sigma_b^2}{\theta^2} = \frac{\sigma_h^2}{\sigma^2} \cdot \frac{\sigma^2}{\theta^2}$

$\frac{\sigma^2}{\theta^2}$ er i det konkrete tilfelle av størrelsesorden ca. 0,06, og $\frac{\sigma_b^2}{\sigma^2}$ tilnærmet av orden 0,01. Dvs.: $\frac{\sigma_b^2}{\theta^2}$ er av størrelsesorden 0,0006. Det er derfor forsvarlig å se bort fra $\frac{\sigma^2}{\theta^2}$, selv om den skulle kunne øke betydelig. Derfor kan i det foreliggende tilfelle settes

$$\rho' > \frac{c'}{2} \cdot \frac{1 + n \cdot \frac{(\xi - \theta)^2}{\omega^2}}{1 - \alpha \cdot \psi^2}$$

Er nå også $\frac{\tau_w^2}{\tau^2} = \psi^2$ av orden $(1 - 0{,}01) = 0{,}99$, blir $1 - \alpha\,\psi^2$ svært nær lik $1-\alpha$, og man får

$$\rho' > \frac{C'}{2} \cdot \frac{1 + n \cdot \frac{(\xi - \theta)^2}{\omega^2}}{1 - \alpha}$$

Tilfellet $C' = \frac{\frac{\omega}{\theta}}{\frac{\tau}{\eta}} = 1$

betraktes nå spesielt. Det skulle ikke være noe dårlig grunnlag for en vurdering, og resultater for $C \neq 1$ vil kunne fåes ved enkel multiplikasjon. De følgende tabeller er derfor beregnet under forutsetning av at $C = 1$.

Tabell 1 A gir - med utgangspunkt i en $\frac{\omega}{\theta}$ - verdi tilsvarende $\frac{\sigma_w}{\xi}$ i engros, omsetningsgruppe 2 - tallene for

$$\frac{1}{2}\left(1 + n\,\frac{(\xi - \theta)^2}{\omega^2}\right)$$

som funksjon av total utvalgsstørrelse n og som funksjon av $100 \cdot \frac{|\xi - \theta|}{\theta}$ - absoluttverdi av endring fra revisjonstidspunkt til tellingstidspunkt i prosent av verdi på tellingstidspunkt.

I tabell 1 B og 1 C er gitt

$$\frac{1}{2} \cdot \frac{1 + n \cdot \frac{(\xi - \theta)^2}{\omega^2}}{1 - \alpha}$$

for de samme n og $100 \cdot \frac{|\xi - \theta|}{\theta}$ og for to utvalgsbrøker, α, henholdsvis 0,05 og 0,20. Tallene angir størrelser som ρ', - korrelasjon mellom totalomsetning på tellingstidspunkt og undersøkt kjennetegn - må overstige for at

$$\bar{B}^* = \xi_h \cdot \frac{\bar{y}}{\bar{x}} \quad \text{(brøkestimat)}$$

skal være bedre enn

$$\hat{\bar{B}}_h = \frac{\xi_h}{\xi} \cdot \bar{y} \quad \text{(korrigerte gjennomsnitt, opprinnelige versjon).}$$

Tabell 1 A kan oppfattes tilsvarende for tilfellet $\alpha \longrightarrow 0$.

For praktiske formål er det nyttig å ha "ρ' - minimum" som funksjon av antall strata, L, og totalt antall enheter pr. stratum, som etter forutsetningene er konstant $= N_1$. I tabell 1 B og 1 C er det derfor innført verdisett av L og N_1 som sammen med $\alpha = 0{,}05$ eller $\alpha = 0{,}20$ gir den i hver horisontalrad oppgitte verdi av n.

Tabell 1 A-C

Størrelse som korrelasjonen mellom undersøkt kjennetegn og omsetning på tellingstidspunkt må overstige for at brøkestimatet $\overline{B}_h^*$ skal være bedre enn den tidligere versjon av „korrigerte gjennomsnitt" - estimat, $\overline{B}_h$.

Antall strata = L

Totalt antall enheter pr. stratum = N_l

Tabell 1 A. Utvalgsbrøk α = 0

n \ $100 \cdot \frac{\lvert \xi - \theta \rvert}{\theta}$	0,2	0,5	1,0	2,0	3,5	5,0
10	0,500	0,502	0,509	0,533	0,601	0,707
20	0,501	0,504	0,517	0,566	0,703	0,914
50	0,502	0,510	0,541	0,666	1,006	1,533
100	0,503	0,521	0,583	0,831	1,513	2,567
200	0,507	0,542	0,666	1,161	2,525	4,633
500	0,517	0,603	0,914	2,153	5,565	10,833
1000	0,533	0,706	1,328	3,805	10,625	21,166

Tabell 1 B. Utvalgsbrøk α = 0,05, dvs. 5 prosent utvalg

L=10 N_l	L=30 N_l	n \ $100 \cdot \frac{\lvert \xi - \theta \rvert}{\theta}$	0,2	0,5	1,0	2,0	3,5	5,0
20	7	10	0,527	0,529	0,536	0,561	0,633	0,744
40	13	20	0,528	0,531	0,544	0,596	0,740	0,962
100	33	50	0,529	0,537	0,570	0,701		
200	67	100	0,530	0,549	0,614	0,875		
400	133	200	0,534	0,571	0,701			
1000	333	500	0,544	0,635	0,962			
2000	667	1000	0,561	0,743				

Tabell 1 C. Utvalgsbrøk $\alpha = 0{,}20$, dvs. 20 prosent utvalg

L=10 N_1	L=30 N_1	n \ $100 \cdot \frac{\vert \xi - \theta \vert}{\theta}$	0,2	0,5	1,0	2,0	3,5	5,0
5	2	10	0,625	0,628	0,636	0,666	0,751	0,884
10	3	20	0,626	0,630	0,646	0,708	0,879	1,143
25	8	50	0,628	0,638	0,676	0,833		
50	17	100	0,629	0,651	0,729	1,039		
100	33	200	0,634	0,678	0,833			
250	83	500	0,646	0,754	1,143			
500	167	1000	0,666	0,883				

Resultatene gjelder under de presiserte forutsetninger. Men de burde også kunne gi en pekepinn under de forhold som er i praksis. Det ligger derfor nær å trekke omtrent disse konklusjoner:

Hvis endringene (i totalomsetning) fra siste registerrevisjon til tellingstidspunkt er svært små (f.eks. 0,2 prosent), er brøkestimatet å foretrekke, såsant ρ ikke er mindre enn 0,55 og 0,65 for henholdsvis $\alpha = 0{,}05$ og 0,20. Med $\alpha = 0{,}05$ er brøkestimatet også konkurransedyktig for noe større verdier av endringsprosenten (inntil ca. 1 prosent). Lite antall strata øker også brukbarheten av metoden.

Hvis endringene er noe større, og f.eks. oppe i 5 - 10 prosent, bør foretrekkes den tidligere versjon av „korrigerte gjennomsnitts metode", og med klassifisering av enhetene etter registeropplysninger.

2. Skjevheter som adderes opp ved summering over alle omsetningsgrupper

Utgangspunktet for vurdering av de forskjellige estimeringsmetoder har hittil vært egenskapene ved estimater for gjennomsnitt innen omsetningsgruppe for hver næringsgruppe (undergruppe). Vurderingen har bygget på hvordan metodene ville virke stort sett når alle undergrupper har vært betraktet under ett. Den metode har vært ansett som best som ga minste gjennomsnittlige (kvadrerte) avvikelser fra „sann" verdi i næringsgruppe innen omsetningsgruppe.

Nå er man imidlertid interessert i totaler for hele næringsgrupper (undergrupper). Man multipliserer da gjennomsnitt innen omsetningsgruppe for vedkommende næring med antall bedrifter - hentet fra registeret - og summerer over alle omsetningsgrupper.

Vil da de resultater som er utledet for en omsetningsgruppe fremdeles være gyldige, eller kan det tenkes at summeringen bringer endringer i forholdet mellom estimeringsmetodene ?

Hvis alle estimatene var forventningsrette, ville ingen problemer oppstå. Men i virkeligheten er både „ukorrigerte" og begge versjoner av „korrigerte gjennomsnitt" estimater vanligvis er belastet med skjevhet.

Bare metoden med rent - nå stratumveiet - gjennomsnitt gir forventningsrette estimater (forutsatt at antall bedrifter hentet fra registeret er identisk med det antall de beregnede gjennomsnitt er basert på).

Ukorrigerte gjennomsnitt er belastet med skjevhet. Forventning av estimat i undergruppe er populasjonsgjennomsnitt i hovedgruppe for vedkommende omsetningsgruppe.

Nå er det forutsatt at totalomsetning har logaritmisk-normal fordeling. Anta lignende fordeling for kjennetegn som skal undersøkes. Forutsett spesielt at den logaritmisk-normale fordeling innen undergruppe og den innen hovedgruppe har samme spredningsparameter ε. Hvis da medianen i fordeling innen hovedgruppe er f.eks. større enn i den aktuelle undergruppe, vil man i alle omsetningsgrupper få at forventning i hovedgruppe er litt større enn forventning i undergruppe. Brukes så utvalgsgjennomsnitt innen hovedgruppe som estimat for populasjonsgjennomsnitt i undergruppe, vil man innen alle omsetningsgrupper løpe stor risiko for (i dette tilfelle) overestimering. Ved addisjon over alle omsetningsgrupper, vil skjevheter som sterkt tenderer i samme retning lett føre til betydelig skjevhet på summen.

Problemet synes sterkt redusert ved bruk av korrigerte gjennomsnitt. „korreksjonen" av hovedgruppegjennomsnittet ($\bar{y}$) med totalomsetning innen undergruppe dividert med totalomsetning innen hovedgruppe, som anslås ved $\frac{\hat{\xi}_h}{\hat{\xi}}$ eller $\frac{\hat{\xi}_h}{Z}$, bør rimeligvis motvirke systematiske skjevheter. Skjevheten innen omsetningsgruppe er 0 med bruk av „korreksjonsfaktoren" $\frac{\xi_h}{\xi}$ hvis

$$\frac{\bar{B}_h}{\bar{B}} = \frac{\xi_h}{\xi} \approx \frac{\bar{A}_h}{\bar{A}}\text{, som også kan skrives } \frac{\bar{B}_h}{\bar{A}_h} \approx \frac{\bar{B}}{\bar{A}}.$$

Nærmere analyse må utstå. Men den konklusjon må i hvert fall kunne trekkes at når man ønsker estimat for gjennomsnitt eller totaler oppstått ved summering over alle omsetningsgrupper, tegner tabell 1A - 1C i notat av 11-6-64 et for lyst bilde av hva som kan ventes oppnådd ved ukorrigerte gjennomsnitts metode.