

Översikt av estimatorer för stora och små redovisningsgrupper

Sixten Lundström



R&D Report
Statistics Sweden
Research - Methods - Development
1991:2

INLEDNING

TILL

R & D report : research, methods, development / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1988-2004. – Nr. 1988:1-2004:2.

Häri ingår Abstracts : sammanfattningar av metodrapporter från SCB med egen numrering.

Föregångare:

Metodinformation : preliminär rapport från Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån. – 1984-1986. – Nr 1984:1-1986:8.

U/ADB / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1986-1987. – Nr E24-E26

R & D report : research, methods, development, U/STM / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1987. – Nr 29-41.

Efterföljare:

Research and development : methodology reports from Statistics Sweden. – Stockholm : Statistiska centralbyrån. – 2006-. – Nr 2006:1-.

R & D Report 1991:2. Översikt av estimatorer för stora och små redovisningsgrupper / Sixten Lundström.

Digitaliserad av Statistiska centralbyrån (SCB) 2016.

Översikt av estimatorer för stora och små redovisningsgrupper

Sixten Lundström



R&D Report
Statistics Sweden
Research - Methods - Development
1991:2

Från trycket
Producent
Ansvarig utgivare
Förfrågningar

Januari 1991
Statistiska centralbyrån, utvecklingsavdelningen
Åke Lönnqvist
Sexten Lundström, 019-17 64 96

© 1991, Statistiska centralbyrån
ISSN 0283-8680
Printed in Sweden

Översikt av estimatorer för stora och små redovisningsgrupper.

1. Inledning

Statistikanvändare nöjer sig vanligtvis inte med statistik för hela populationen, utan vill i allmänhet ha information även för redovisningsgrupper (delpopulationer). Det kan gälla en uppdelning efter kön och ålder vid en individundersökning eller efter näringsgren i en företagsundersökning. När statistiken hämtas från ett kvalitetsmässigt gott register eller totalundersökning är det normalt inga problem med framställningen. Det är inte heller några större problem om en urvalsundersökning är väl designad för att skatta parametrar för stora redovisningsgrupper. Om däremot skattningar önskas för små redovisningsgrupper kan man få problem eftersom antalet observationer kan bli litet i vissa redovisningsgrupper eller i värsta fall finns det inga observationer alls.

Sedan länge har det funnits metoder för att skatta parametrar för "stora" redovisningsgrupper. En gemensam egenskap hos dessa "konventionella" estimatorer är att de är väntevärdesriktiga eller åtminstone för stora urval är approximativt väntevärdesriktiga. Detta gör det möjligt att beräkna konfidensintervall, vilket är en god egenskap. Nackdelen är att de ger alltför stora urvalsfel för små redovisningsgrupper.

Dessa begränsningar är besvärande och därför har statistiker världen över börjat intressera sig för modellberoende estimatorer eller som de populärt kallas för syntetiska estimatorer. Ett markerat drag hos dessa estimatorer är att de "lånar" information från andra redovisningsgrupper, vanligtvis tillsammans med en modell. Därmed kan urvalsfelet reduceras kraftigt jämfört med konventionella estimatorer. Detta sker dock ofta till pris av införande av skevhet. Denna skevhet, som inte kan skattas utifrån aktuella data, leder till att konfidensintervallet blir ett irrelevant mått på kvaliteten. Mean Square Error (se (29)) kan dock bli mycket mindre för dessa estimatorer än för konventionella estimatorer vid små redovisningsgrupper.

Vad som är "stora" resp "små" redovisningsgrupper kan inte preciseras entydigt, eftersom det beror på flera omständigheter, t ex kan det vara så att undersökningen är designad för att med hjälp av en konventionell estimator ge separata skattningar för vissa små redovisningsgrupper. Det beror även på sambandet mellan undersökningsvariabeln och hjälpinformationen.

I skalan från stora till små redovisningsgrupper finner man ett ökat beroende av modeller. För att enkelt beskriva hur en modell kan se ut kan följande exempel ges. Antag att man vill skatta andelen arbetslösa i varje kommun. Arbetskraftsundersökningen (AKU) ger en hygglig skattning av andelen arbetslösa i varje landsdel. En mycket enkel modellberoende estimator är att för varje kommun inom en landsdel ta estimatet för landsdelen. Modellen består alltså i ett antagande om att varje kommun inom en landsdel har samma andel arbetslösa.

Konventionella estimatorer för stora redovisningsgrupper är beskrivna i de flesta böckerna inom survey sampling, som Cochran [1977], Kish [1965], och Des Raj [1972] och därför finns inte samma behov av översikt för dessa som för modellberoende estimatorer, som huvudsakligen finns beskrivna i artiklar. Trots detta beskrivs även de först nämnda i denna rapport. Framställningen baserar sig till stor del på manuskriptet till den kommande boken "Model Assisted Sampling Strategies" av Carl-Erik Särndal

(Université de Montréal, Kanada) Bengt Swensson (Högskolan i Örebro) och Jan Wretman (Stockholms universitet). (I fortsättningen kallas detta material för SSW). Deras framställning ger en sammanhållande ram för hela gruppen och utgör även ett bra utgångsläge för beskrivning av modellberoende estimatorer.

Vi vill skilja mellan modellbaserad och modellberoende estimator i denna rapport. Utvecklingen av en estimator för stora redovisningsgrupper kan vara modellbaserad (som i SSW) där modellansatsen syftar till att ge litet urvalsfel - väntevärdesriktig är estimatoren även om modellen avviker från verkligheten. Modellberoende estimatorer är däremot skeva, där skevheten är större ju sämre modellen stämmer med verkligheten.

Rapporten redovisar ett stort antal estimatorer utan att närmare gå in på deras egenskaper. Det finns tyvärr inte tillräckligt med kunskap och erfarenhet kring många av metoderna och framförallt är det svårt att dra generella slutsatser av genomförda studier.

I slutet av rapporten ges en lång lista av artiklar som behandlar problemet med skattningar för små redovisningsgrupper.

2. Definition av statistiska begrepp

Vi betraktar en ändlig population $U = \{1, \dots, k, \dots, N\}$. Populationen är uppdelad i de disjunkta redovisningsgrupperna $U_1, \dots, U_q, \dots, U_Q$, för vilka vi vill ha skattningar av totaler. Dessutom är populationen uppdelad i H disjunkta grupper, $U_1, \dots, U_h, \dots, U_H$ bestämda av hjälpvariabler. Dessa grupper låter vi utgöra design strata eller poststrata (för generalitetens skull borde designstrata och poststrata bestå av olika indelningar, men för enkelhets skull görs detta). Denna korsklassificering ger $H \times Q$ populationsceller U_{hq} ; $h=1, \dots, H$; $q=1, \dots, Q$. Låt N_{hq} vara storleken på U_{hq} , vilken genomgående antas vara känd.

Undersökningsvariabeln antar värdena y_k för k :te elementet; $k = 1, \dots, N$. Vi önskar skatta totalen för redovisningsgrupp q

$$t_q = \sum_{U_q} y_k \quad (1)$$

I detta syfte dras stickprovet s av storleken n från U enligt designen $p(s)$ med inklusionssannolikheterna $P(k \in s) = \pi_k (>0)$ för alla k och $P(k \text{ och } l \in s) = \pi_{kl} (>0)$ för alla $k \neq l$. (Vi sätter $\pi_{kk} = \pi_k$). En del av stickprovet s kommer att bestå av element från U_{hq} och betecknas s_{hq} . Antalet element i s_{hq} beror i allmänhet av slumpen. Däremot kan antalet i s_h (den del av s som består av element i U_h) vara en konstant, nämligen då U_h ($h = 1, \dots, H$) utgör designstrata och s_h dras med s k "fixed size design". Vi låter antalet enheter i s_{hq} vara n_{hq} . I denna rapport exemplifierar vi med designerna OSU och STOSU, där ett obundet slumpmässigt urval av storleken n dras från U resp ett OSU av den fixa storleken n_h dras från stratum U_h ($h = 1, \dots, H$).

3. Estimatorer för stora redovisningsgrupper.

I detta avsnitt redovisas en översikt av estimatorer för stora redovisningsgrupper, som alla är väntevärdesriktiga eller åtminstone asymptotiskt väntevärdesriktiga estimatorer.

En "byggsten" i den följande framställningen är den $s_k \pi$ -estimatorn (eller Horvitz-Thompson-estimatorn):

$$\hat{t}_{q\pi} = \sum_{s_q} y_k / \pi_k = \sum_{s_q} \check{y}_k \quad (\check{y}_k \text{ kallas för "}\pi\text{-expanderat" värde.}) \quad (2)$$

Det generella uttrycket för variansen är (se härledning i bilaga 1)

$$V(\hat{t}_{q\pi}) = \sum_{k \in U_q} \sum_{l \in U_q} \Delta_{kl} \check{y}_k \check{y}_l, \text{ där } \Delta_{kl} = \pi_{kl} - \pi_k \pi_l \quad (3)$$

En unbiased varians estimator ges av

$$\hat{V}(\hat{t}_{q\pi}) = \sum_{k \in s_q} \sum_{l \in s_q} \check{\Delta}_{kl} \check{y}_k \check{y}_l, \text{ där } \check{\Delta}_{kl} = \Delta_{kl} / \pi_{kl} \quad (4)$$

Vid fixed size design är oftast följande varians estimator att föredra

$$\hat{V}(\hat{t}_{q\pi}) = -\frac{1}{2} \sum_{k \in s_q} \sum_{l \in s_q} \check{\Delta}_{kl} (\check{y}_k - \check{y}_l)^2 \quad (5)$$

Vid OSU och STOSU ger båda alternativen samma resultat.

För att visa att variansen och variansskattningen är "gamla bekanta" utvecklar vi den för designen OSU. Då är $\pi_k = n / N$ och $\pi_{kl} = (n / N) (n-1) / (N-1)$ för alla $k \neq l$ [$\pi_{kk} = n/N$]. Detta ger

$$\hat{t}_{q\pi} = \frac{N}{n} \sum_{s_q} y_k \quad (6)$$

$$V(\hat{t}_{q\pi}) = N^2 \frac{1-n/N}{n} \frac{1}{N-1} \left\{ \sum_{U_q} y_k^2 - \frac{1}{N} \left(\sum_{U_q} y_k \right)^2 \right\} \quad (7)$$

och en skattning av variansen

$$\hat{V}(\hat{t}_{q\pi}) = N^2 \frac{1-n/N}{n} \frac{1}{n-1} \left\{ \sum_{s_q} y_k^2 - \frac{1}{n} \left(\sum_{s_q} y_k \right)^2 \right\} \quad (8)$$

Om vi har designen STOSU erhålles estimatorn

$$\hat{t}_{q\pi} = \sum_h \sum_{s_{hq}} y_k / \pi_k = \sum_h \frac{N_{h.}}{n_{h.}} \sum_{s_{hq}} y_k = \sum_h \hat{N}_{hq} \bar{y}_{s_{hq}}, \text{ där } \hat{N}_{hq} = \frac{N_{h.}}{n_{h.}} n_{hq} \quad (9)$$

Varians och variansestimator kan lätt härledas från (3) och (5).

SSW baserar en stor del av framställningen på en generaliserad regressionsansats. Den visar på ett pedagogiskt sätt hur man kan använda modelltänkade vid utvecklande av estimator och den ger också en naturlig brygga över till modellberoende estimatorer.

På ungefär följande sätt beskriver de ansatsen.

Om vi har tillgång till hjälpvariabler $x_1, \dots, x_j, \dots, x_J$ - dvs för varje $k \in U$ har tillgång till en vektor $\mathbf{x}_k = (x_{1k}, \dots, x_{jk}, \dots, x_{Jk})'$ - sådana att y och $\mathbf{x}$ uppvisar hyfsade samband kan vi utnyttja detta för att konstruera estimatorer som är effektivare än den enkla π -estimatoren. Utgångspunkten är ett antagande att utseendet av spridningsdiagrammet (se nedan) för de N punkterna $(y_k, \mathbf{x}_k)$ i den ändliga populationen ser ut som om det hade kunnat genereras av en regressionsmodell ξ sådan att $y_1, \dots, y_N$ utgör realiserade värden av oberoende slumpvariabler $Y_1, \dots, Y_N$, för vilka

$$E_{\xi}(Y_k) = \mathbf{x}_k' \boldsymbol{\beta} = \sum_j \beta_j x_{jk} \quad (k=1, \dots, N) \quad (10)$$

och

$$V_{\xi}(Y_k) = \sigma_k^2$$

Om populationen hade kunnat totalundersökas, skulle en vägd minsta kvadrat-skattning (under modellen ξ) av $\boldsymbol{\beta} = (\beta_1, \dots, \beta_J)'$ ges av

$$\mathbf{B} = (B_1, \dots, B_J)' = \left(\sum_U \frac{\mathbf{x}_k \mathbf{x}_k'}{\sigma_k^2} \right)^{-1} \sum_U \frac{\mathbf{x}_k y_k}{\sigma_k^2} = \mathbf{T}^{-1} \mathbf{t} \quad (11)$$

$\mathbf{T}$ och $\mathbf{t}$ skattas väntevärdesriktigt av

$$\hat{\mathbf{T}} = \sum_s \frac{\mathbf{x}_k \mathbf{x}_k'}{\sigma_k^2 \pi_k} \text{ resp. } \hat{\mathbf{t}} = \sum_s \frac{\mathbf{x}_k y_k}{\sigma_k^2 \pi_k} \quad (12)$$

vilka utnyttjas för att som en estimator (med viss men vanligen försumbar skevhet) för $\mathbf{B}$ bilda

$$\hat{\mathbf{B}} = (\hat{B}_1, \dots, \hat{B}_J)' = \left(\sum_s \frac{\mathbf{x}_k \mathbf{x}_k'}{\sigma_k^2 \pi_k} \right)^{-1} \sum_s \frac{\mathbf{x}_k y_k}{\sigma_k^2 \pi_k} \quad (13)$$

Den resulterande regressionsestimatorn för populationstotalen ges nu av

$$\hat{t}_{yr} = \hat{t}_{y\pi} + \sum_{j=1}^J \hat{B}_j (t_{x_j} - \hat{t}_{x_j\pi}), \text{ där } \hat{t}_{x_j\pi} \text{ är } \pi\text{-estimatorn för totalen för } x_j. \quad (14)$$

I SSW visas att regressionsestimatorn kan skrivas på olika sätt. Ett sätt är

$$\hat{t}_{yr} = \sum_U \hat{y}_k + \sum_s \hat{e}_{ks}, \text{ där } \hat{y}_k = \sum_j \hat{B}_j x_{jk} \text{ och } \hat{e}_{ks} = y_k - \hat{y}_k \quad (15)$$

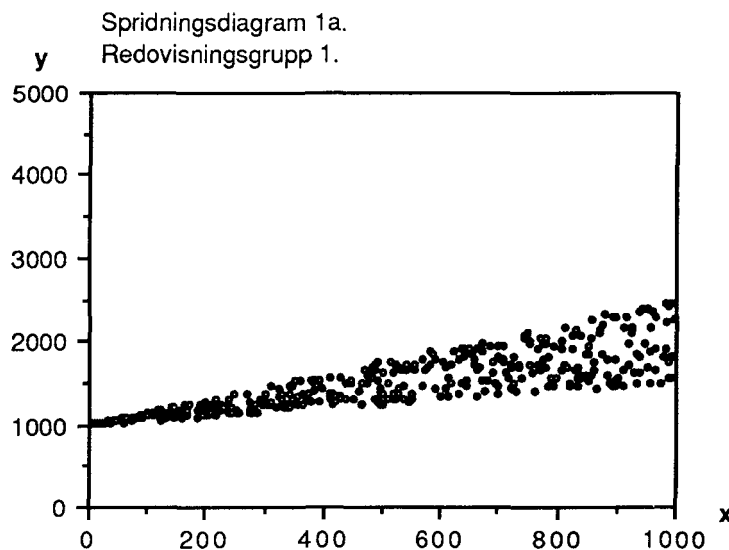
Motsvarande estimator för totalen i redovisningsgrupp q är

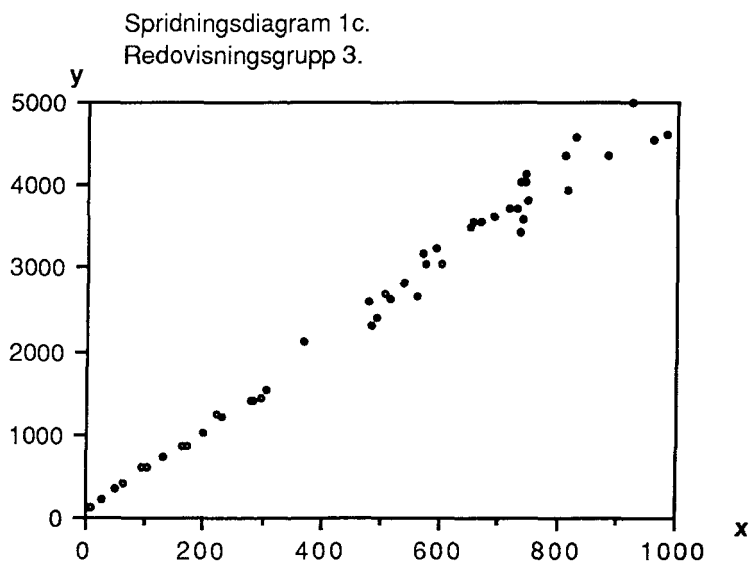
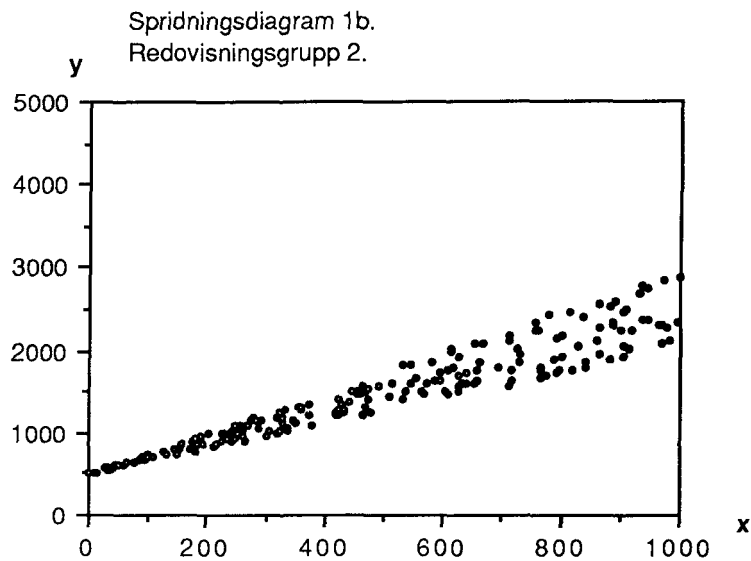
$$\hat{t}_{qr} = \sum_{U_q} \hat{y}_k + \sum_{s_q} \hat{e}_{ks} \quad (16)$$

I fortsättningen ska vi utnyttja denna generella estimator och för olika modeller ξ visa hur såväl kända som mindre kända estimatorer "faller ut". För att välja "rätt" modell krävs att man har kunskap om sambandet mellan undersökningsvariabeln och hjälpvariabeln. Helst vill man ha möjlighet att rita ett spridningsdiagram.

I (27) visas ett annat sätt att skriva estimatorn.

Antag att man har tre redovisningsgrupper och man har data som beter sig ungefär enligt följande diagram:





Då är följande modelltyp lämplig:

Modelltyp 1

$$E_{\xi}(Y_k) = \alpha_q + \beta_q x_k; \quad V_{\xi}(Y_k) = \sigma_q^2 x_k^v \quad \text{för } k \in U_q \text{ och där } v > 0. \quad (17)$$

Spridningsdiagrammen 1a - 1c är framställda utifrån en modell där β_1 , β_2 och β_3 har värdena 1, 2 resp 5 och $\alpha_1 = 1000$, $\alpha_2 = 500$ och $\alpha_3 = 100$. Dessutom är $v = 2$.

För att konkretisera estimatorn studerar vi fallet där linjen går genom origo ($\alpha_q = 0$ för alla q) och där $v=1$. Då kan uttrycket (12) skrivas

$$\hat{T} = \sum_{s_q} \frac{x_k}{\sigma_q^2 \pi_k} \text{ och } \hat{t} = \sum_{s_q} \frac{y_k}{\sigma_q^2 \pi_k} \quad (18)$$

$$\text{vilket ger } \hat{B} = \left(\sum_{s_q} \check{y}_k / \sum_{s_q} \check{x}_k \right). \text{ Dessutom erhålles } \sum_{s_q} \check{e}_{ks} = 0.$$

Det resulterande uttrycket för estimatorn är

$$\hat{t}_{qr}(1) = X_q \left(\sum_{s_q} y_k / \pi_k \right) / \sum_{s_q} x_k / \pi_k \quad (19)$$

$$\text{där } X_q = \sum_{U_q} x_k$$

Dessa modellantagande leder alltså till en konventionell kvotestimator.

Om designen är OSU och vi antar att $x_k = 1$ för $k \in U_q$ erhålles följande välkända estimator:

$$\hat{t}_{qr}(1) = N_q \sum_{s_q} y_k / n_q \quad (20)$$

Skulle däremot sambandet mellan y och x vara tydligare genom en uppdelning på hjälpgrupper bör följande modell väljas:

Modelltyp 2

$$E_{\xi}(Y_k) = \alpha_h + \beta_h x_k ; V_{\xi}(Y_k) = \sigma_h^2 x_k^v \text{ för } k \in U_h. \quad (21)$$

Ett intressant specialfall uppkommer när $\alpha_h = 0$ (för $h=1, \dots, H$) och $v = 1$. För OSU och STOSU erhålles:

$$\hat{t}_{qr}(2) = \sum_h \left[X_{hq} \frac{\bar{y}_{s_h}}{\bar{x}_{s_h}} + k_h n_{hq} \left\{ \bar{y}_{s_{hq}} - \frac{\bar{y}_{s_h}}{\bar{x}_{s_h}} \bar{x}_{s_{hq}} \right\} \right], \text{ där } k_h = \frac{N}{n} \text{ vid} \quad (22)$$

$$\text{OSU och } k_h = \frac{N_h}{n_h} \text{ vid STOSU.}$$

Den estimatorn användes senare i rapporten för att demonstrera övergången till modellberoende estimatorer. I fortsättningen kallar vi den första termen för "syntetisk term" och den andra för "korrektionsterm".

Om vi har $x_k = 1$ för alla k erhålles:

$$\hat{t}_{qr}(2) = \sum_h \left\{ N_{hq} \bar{y}_{s_h} + \frac{n_{hq}}{\pi_k} (\bar{y}_{s_{hq}} - \bar{y}_{s_h}) \right\} \quad (23)$$

Estimatorn (23) studeras i Särndal (1981).

Om vi även tror att "redovisningsgruppsdimensionen" har betydelse kan följande modelltyp vara lämplig.

Modelltyp 3

$$E_{\xi}(Y_k) = \alpha_{hq} + \beta_{hq}x_k; \quad V_{\xi}(Y_k) = \sigma_{hq}^2 x_k^v \quad \text{för } k \in U_{hq} \quad (24)$$

Om $x_k = 1$ (för alla k) och $\alpha_{hq} = 0$ ($h=1, \dots, H$ och $q=1, \dots, Q$) erhålles för både OSU och STOSU

$$\begin{aligned} \hat{B} &= \bar{y}_{s_{hq}} \text{ och } \sum_{s_q}^v e_{ks} = 0, \text{ vilket ger den poststratifierade estimatoren} \\ \hat{t}_{qr}(3) &= \sum_h N_{hq} \bar{y}_{s_{hq}} \end{aligned} \quad (25)$$

Särndal och Hidioglou (1987) fann att en i många fall bättre estimator än (16) (har mindre varians och de betingade egenskaperna är bättre) är

$$\hat{t}_{qrev} = \sum_{U_q} \hat{y}_k + N_q \frac{\sum_{s_q}^v e_{ks}}{\hat{N}_q} \quad (26)$$

I vissa fall är $\sum_{s_q}^v e_{ks} = 0$, vilket gör att $\hat{t}_{qr} = \hat{t}_{qrev}$. Det inträffar för modelltyp 1 och 3.

Estimatorn för modelltyp 2 erhålles lätt genom att summera den syntetiska delen i (22) med produkten av $N_q / \hat{N}_q$ och korrektionstermen.

I detta avsnitt har vi endast behandlat tre huvudmodeller, men det finns även andra alternativ.

Estimatorerna $\hat{t}_{qr}$ och $\hat{t}_{qrev}$ kan uttryckas i en generell estimator $\hat{t}_q$:

$$\hat{t}_q = \sum_s g_{qks} \check{y}_k \quad (27)$$

där

$$g_{qks} = \delta_q d_{qk} + \left(\sum_{U_q} x_k - \delta_q \sum_{s_q} \check{x}_k \right) \hat{T}^{-1} x_k / \sigma_k^2,$$

där $d_{qk} = 1$ om $k \in s_q$; $= 0$ f.ö. och $\delta_q = 1$ för $\hat{t}_{qr}$; $= \frac{N_q}{\hat{N}_q}$ för $\hat{t}_{qrev}$ då $k \in U_q$.

Denna estimator liknar i formen π -estimatorn och det går att visa att en lämplig varians estimator är (jmf med (4))

$$\hat{V}(\hat{t}_q) = \sum_s \sum_{kl} \check{\Delta}_{kl} g_{qks} \check{e}_{ks} g_{qls} \check{e}_{ls} \quad (28)$$

Denna varians estimator har alltså samma enkla form som varians estimatorn för π -estimatorn. Därför är det enkelt att utveckla de explicita formerna för olika modeller och designar.

4. Estimatorer för små redovisningsgrupper.

De estimatorer, som presenteras i föregående avsnitt är alla approximativt väntevärdesriktiga för stora stickprov inom varje redovisningsgrupp. Vid små redovisningsgrupper (=små stickprov) är vissa av estimatorerna behäftade med en skevhet och någon estimator är inte ens definierad då $n_{hq} = 0$ (se (25)). Den avgörande bristen är dock det stora urvalsfelet.

Kraven på att erhålla skattningar även för små redovisningsgrupper leder till att man får lämna gruppen av väntevärdesriktiga estimatorer och söka estimatorer som har små "totala" fel. Med totalt fel kan man t ex mena Mean Square Error (MSE):

$$MSE(\hat{t}_q) = E[(\hat{t}_q - t_q)^2] = V(\hat{t}_q) + \{E(\hat{t}_q) - t_q\}^2 \quad (29)$$

(Den senare termen är skevheten i kvadrat.)

Det innebär att även om estimatorn är behäftad med skevhet så kan det totala felet vara litet om variansen är liten.

Estimatorerna är genomgående modellberoende, vilket innebär att om modellen överensstämmer dåligt med verkligheten uppkommer (design) skevhet. Det är följaktligen mycket viktigt att noggrant undersöka om så är fallet.

Den, i litteraturen vanligast förekommande modellberoende estimator återfinns som första term i estimatorerna

$\hat{t}_{qr}$ och $\hat{t}_{qrev}$ för modelltyp 2. Den andra termen har vi kallat för "korrigerings-term"

och skälet till detta är att den är en skattning av skevheten för den modellberoende estimator. Denna modellberoende estimator kallas vanligtvis för "syntetisk" estimator.

Om vi har få urvalsobjekt i studerade redovisningsgrupp kommer skevhetsskattningen att vara behäftad med stort urvalsfel och därmed ge ett stort värde på MSE. Om man däremot litar på att modelltyp 2 bra beskriver verkligheten kan endast den syntetiska delen utnyttjas och därmed reduceras MSE. Den syntetiska estimatorn har ofta ett litet urvalsfel eftersom den baserar sig på en skattning av B utifrån urvalsobjekten i de stora populationsgrupperna U_h . Modellen i denna estimator innebär alltså ett antagande om att skevheten är liten, vilket i sin tur innebär att

$$\frac{\bar{Y}_{U_{hq}}}{\bar{X}_{U_{hq}}} - \frac{\bar{Y}_{U_h}}{\bar{X}_{U_h}} \approx 0 \quad (30)$$

Om vi antar att $x_k = 1$ för alla k förenklas uttryck (30) till

$$\bar{Y}_{U_{hq}} - \bar{Y}_{U_h} \approx 0 \quad (31)$$

I litteraturen ser man vanligtvis följande benämningar på de två syntetiska estimatorerna ($x_k = 1$ för alla k resp x_k kan anta andra värden) :

$$\text{Synthetic-count: } \sum_h N_{hq} \bar{y}_{s_h} \quad (32)$$

och

$$\text{Synthetic-ratio: } \sum_h X_{hq} \frac{\bar{y}_{s_h}}{\bar{x}_{s_h}} \quad (33)$$

Purcell and Linacre[1976] ger en syntetisk estimator som är något annorlunda -

$\bar{x}_{s_h}$ i (33) ersätts av den kända storheten $\bar{X}_{U_h}$. Oftast är dock synthetic-ratio

estimatorn bättre än denna estimator.

Holt, Smith, and Thomberlin [1979] visar i en s k "Prediction approach" att en bättre (mindre modellvarians under antagandet om modellväntevärdesriktighet) estimator, då $x_k = 1$ för alla k , är följande:

$$\hat{t}_{qPred} = \sum_h \{N_{hq} \bar{y}_{s_h} + n_{hq} (\bar{y}_{s_{hq}} - \bar{y}_{s_h})\} \quad (34)$$

Oftast är $N_{hq} \gg n_{hq}$ och då är denna estimator och synthetic-count-estimatorn i stort sett lika. Om man utnyttjar klustersampling kan dock skillnaden vara stor för vissa redovisningsgrupper. Skillnaden i sättet att konstruera de två estimatorerna är att den förra utnyttjar det faktum att urvalsobservationerna är kända exakt och därför behöver man endast prediktera de $(N_{hq} - n_{hq})$ ej observerade värdena.

Choudhry and Rao [1988] samt Prasad and Rao [1990] studerar mer komplicerade modeller (bl a stokastiska intercept) och baserat på dessa utvecklar de "Empirical best linear unbiased predictors" (EBLUPs). Man kan förvänta sig att dessa estimatorer, liksom

$\hat{t}_{qPred}$, fungerar bättre än vanliga syntetiska estimatorer när urvalsfraktionen är stor. I våra tillämpningar vid SCB är dock normalt urvalsfraktionen ytterst liten och därför torde de enklare modellberoende estimatorerna vara att rekommendera. För att ändå översikten ska bli någorlunda fullständig redovisas ett antal EBLUPs i bilaga 2.

Ovanstående estimatorer skattar en total, t_q , i redovisningsgrupp q . I fortsättningen av detta avsnitt ska vi presentera en grupp estimatorer som skattar hela fördelningen för en kategorisk variabel. Teorin baserar sig huvudsakligen på log-linjära modeller.

Antag att undersökningsvariabeln kan anta I kategorier och att vi dessutom har en hjälpvariabel som kan anta H kategorier. Denna grupp av s k SPREE-estimatorer (Structure Preserving Estimator - se Purcell [1979]) skattar N_{hiq} , d v s antalet objekt i redovisningsgrupp q som tillhör hjälpgrupp h och har egenskapen i . Därefter summeras över hjälpvariabelns kategorier

$$\hat{N}_{\cdot iq} = \sum_{h=1}^H \hat{N}_{hiq} \cdot$$

$\hat{N}_{hiq}$ beräknas på olika sätt beroende på vilken modell som utnyttjas.

Ovan talas om en undersökningsvariabel resp en hjälpvariabel, men av flera kategoriserade variabler kan man alltid bilda en variabel och därför ligger det inga begränsningar i en 3-dimensionell beskrivning av estimationsmetoden.

En 3-dimensionell tabell med antalen N_{hiq} kan med hjälp av en log-linjär ansats entydigt beskrivas med

$$\log N_{hiq} = u + u_{1(h)} + u_{2(i)} + u_{3(q)} + u_{12(hi)} + u_{13(hq)} + u_{23(iq)} + u_{123(hiq)} \quad (35)$$

För att i detalj förstå denna uppdelning måste någon bok i ämnet läsas t ex Fienberg [1978]. Här vill vi utnyttja (35) för att endast förklara principen i skattningsmetoden.

Log-linjära modeller används vanligtvis för att analysera komplicerade korstabeller. Om man har en tre-dimensionell korstabell startar man oftast med att testa om tre-faktor effekten är försumbar, d v s om $u_{123} \approx 0$. Om så är fallet fortsätter man med att "plocka

bort" insignifikanta termer för att slutligen erhålla en enkel och bra beskrivning av strukturen i tabellen.

I föreliggande fall utnyttjar vi tekniken med log-linjära modeller för att erhålla en bra skattning av den okända storheten N_{hiq} . I princip går det till så att vissa termer i (35) skattas med hjälp av ett urval, andra termer skattas från aktuell registerinformation och slutligen kan man ha inaktuell information som ger skattningar av några termer. Återgår vi till den 3-dimensionella tabellen kan det vara så att vi erhåller

- a) $\hat{N}_{hi}$, dvs antalet objekt i hjälpgrupp h som har egenskapen i , från en urvalsundersökning,
- b) $N_{h,q}$ från ett register
- och slutligen
- c) inaktuell information, som vi betecknar N'_{hiq} , som kan erhållas från t ex en folk- och bostadsräkning (FoB).

Om vi låter 1 stå för hjälpvariabeln, 2 för undersökningsvariabeln och 3 för redovisningsgrupper innehåller en SPREE-estimator som baserar sig på informationen a)-c) följande antaganden och osäkerheter:

- u_{123} och u_{23} skattas med motsvarande värde erhållet i FoBen, dvs man antar att tre-faktoreffekten och två-faktor effekten mellan undersökningsvariabeln och redovisningsgrupperna inte har förändrats sedan FoBen genomfördes.
- u_{12} är behäftad av urvalsfel som följd av att denna två-faktor effekt skattas utifrån urvalsinformation.
- u_{13} baserar sig på registerinformation.

Om man inte har all information a) -c) tvingas man skatta termerna på annat sätt. Ibland kan även skattningarna vara så osäkra att det är bättre att ta bort termen. I fortsättningen ska vi visa hur olika estimatorer faller ut från olika modeller. Men först beskriver vi tekniken för att skatta N_{hiq} i det fall informationen a)-c) finns. Detta ger en estimator, som baserar sig på antaganden vi kallar modell 1.

Det är inte möjligt att erhålla ett explicit uttryck för denna SPREE-skattning. Lösningen erhålles däremot från en iterativ process, den s k Iterative Proportional Fitting (IPF)-proceduren. Startvärdena i iterationen är

$$N_{hiq}^{(0)} = N'_{hiq} \quad (36)$$

Därefter startar själva iterationen

$$N_{hiq}^{(k)} = \frac{N_{hiq}^{(k-1)}}{N_{hi.}^{(k-1)}} \hat{N}_{hi.} \quad \text{där } N_{hi.}^{(k-1)} = \sum_{q=1}^Q N_{hiq}^{(k-1)} \quad (37)$$

och

$$N_{hiq}^{(k)} = \frac{1}{N_{h,q}^{(k)}} N_{hiq}^{(k)} \quad \text{där} \quad N_{h,q}^{(k)} = \sum_{i=1}^I N_{hiq}^{(k)} \quad (38)$$

Iterationen konvergerar (under tämligen allmänna villkor) relativt snabbt - oftast kan den avbrytas efter fem steg. Man kan också ha en gräns för hur mycket en förändring mellan två steg bör ha för att iterationen ska fortsätta. Antag att sista steget är $k = k_0$, då är SPREE-skattningen

$$\hat{N}_{hiq} = N_{hiq}^{(k_0)} \quad (39)$$

Om vi inte har tillgång till $N_{h,q}$ kan skattningen tas fram utifrån en IPF-procedur där (38) inte ingår (detta kallar vi modell 2). Det visar sig att en explicit form på estimatoren erhålles efter första halv-steget

$$\hat{N}_{hiq}(2) = \frac{N'_{hiq}}{N'_{hi.}} \hat{N}_{hi.} \quad (40)$$

I denna estimator antar vi att u_{13} inte förändras mellan FoBen och undersökningstillfället, men i övrigt är antagandena desamma som i modell 1.

Om man inte har tillgång till N'_{hiq} , men i övrigt tillgång till de aktuella data (modell 3) beskrivna i a)-b) låter man iterationen starta med

$$N_{hiq}^{(0)} = 1 \text{ för alla } h, i \text{ och } q \quad (41)$$

Det är samma sak som att sätta $u_{123} = u_{23} = 0$. Ett explicit uttryck för estimatoren erhålles

$$\hat{N}_{hiq}(3) = N_{h,q} \frac{\hat{N}_{hi.}}{N_{h..}} \quad (42)$$

Om vi har designen STOSU vid skattning av N_{hi} , kan vi se att denna estimator är samma som (32) - "synthetic / count"-estimatorn.

Det finns många alternativa uppsättningar av hjälpinformation, som leder till olika modeller och olika estimatorer. Alla kan lösas med versioner av IPF-proceduren.

5. Estimatorer som samutnyttjar väntevärdesriktiga och modellberoende estimatorer.

Estimatorerna i avsnitt 3 är alla (approximativt) väntevärdesriktiga, medan däremot estimatorerna i avsnitt 4 är skeva beroende på att modellen avviker från verkligheten. Den först nämnda typen har ofta stora urvalsfel och den senare små urvalsfel. I avsnitt 5.1 presenteras sammansatta estimatorer, d v s estimatorer som består i en sammanvägning av en estimator från vardera gruppen. I avsnitt 5.2 presenteras de estimatorer av typen generaliserad regressionsansats (se (26)) där man försöker dämpa det stora urvalsfelet i korrektionstermen. Tanken med dessa estimatorer är att de ska klara både stora och små

redovisningsgrupper. Detta ska ske genom rätt val av vikter, vilket dock inte är lätt att utföra i verkligheten.

5.1 Sammansatta estimatorer

Om vi låter en väntevärdesriktig estimator av t_q benämnas $\hat{t}_q$ och en modellberoende estimator för $\hat{\mu}_q$, då har den sammansatta estimatorn följande generella utseende:

$$\hat{t}_{q\text{Sam}} = C_q \hat{t}_q + (1 - C_q) \hat{\mu}_q \quad (43)$$

där C_q är en vikt som vanligtvis ligger i intervallet $[0,1]$. I den "konventionella" sammansatta estimatorn bestäms C_q apriori. I litteraturen är det dock vanligt att se C_q -värden bestämda utifrån utfallet av urvalet och då benämns de "estimatorer beroende av urvalsstorleken" (eng. Sample size dependent estimators). Den senare typen anses i vissa artiklar inte ens tillhöra gruppen av sammansatta estimatorer, vilket antyder att grupperingen av estimatorer inte är entydig.

Det stora problemet med sammansatta estimatorer är att bestämma C_q -värdena så att estimatorn ger små fel. Nedan presenteras ett antal alternativa lösningsförslag.

C_q bestämt apriori.

Mean square error för den sammansatta estimatorn kan skrivas

$$\text{MSE}(\hat{t}_{q\text{Sam}}) = C_q \text{MSE}(\hat{t}_q) + (1-C_q) \text{MSE}(\hat{\mu}_q) - C_q(1-C_q) E(\hat{t}_q - \hat{\mu}_q)^2 \quad (44)$$

MSE minimeras för

$$C_q = C_q^*, \text{ där}$$

$$C_q^* = \frac{\text{MSE}(\hat{\mu}_q) - E[(\hat{t}_q - t_q)(\hat{\mu}_q - t_q)]}{\text{MSE}(\hat{t}_q) + \text{MSE}(\hat{\mu}_q) - 2E[(\hat{t}_q - t_q)(\hat{\mu}_q - t_q)]} \quad (45)$$

Schaible [1978] visar att om C_q väljs i intervallet $[2C_q^* - 1, 2C_q^*]$ så ger den sammansatta estimatorn mindre MSE än båda komponentestimatorerna. T ex om $C_q^* = 1/2$ ger varje val av C_q i $[0,1]$ en "bättre" estimator än de ingående estimatorerna.

Den optimala vikten innehåller flera okända storheter och därför är det svårt att i praktiken bestämma det. Man får därför söka ett närmevärde. Choudry and Rao [1988] menar att

$E[(\hat{t}_q - t_q)(\hat{\mu}_q - t_q)]$ är relativt liten i de fall $\hat{t}_q$ baserar sig enbart på urvalsobjekten

i redovisningsgrupp q och $\hat{\mu}_q$ på samtliga urvalsobjekt. (Bailar[1978] tror inte på detta antagande.)

Med detta antagande förenklas formen på den optimala vikten, men för att få ett värde på vikten måste vi också veta storleken på $MSE(\hat{\mu}_q)$. Choudry and Rao [1988] föreslår följande väntevärdesriktiga skattning av den storheten:

$$mse(\hat{\mu}_q) = (\hat{\mu}_q - \hat{t}_q)^2 - \text{var}(\hat{t}_q), \text{ där } \text{var}(\hat{t}_q) \text{ är en skattning av variansen för } \hat{t}_q.$$

Denna ansats ger stora urvalsfel i skattningen av den optimala vikten.

I Lundström [1987] testas en sammansatt estimator att användas mellan två FoBar. Där visar sig de optimala vikterna baserade på föregående FoB-information fungera bra även vid undersökningstillfället.

Estimatorer beroende av urvalsstorleken.

Schaible [1978] förenklar uttrycket för optimal vikt på olika sätt. I sin enklaste form antar han att

$$MSE(\hat{t}_q) = k_1 / n_q \text{ och } MSE(\hat{\mu}_q) = k_2 \text{ där } k_1 \text{ och } k_2 \text{ är konstanter.}$$

(Även Lundström [1978] prövar denna variant).

Drew, Singh and Choudhry [1982] väger samman en design unbiased estimator av typ (19) med en syntetisk estimator av Purcell and Linacre [1976]-typ med en vikt som är beroende av sampelutfallet. (De utnyttjar också efterstratifiering, men för att enkelt kunna beskriva principen bakom estimatormen tas den "dimensionen" bort). Om

$$R_q = \left(\sum_{U_q} x_k \right) / \left(\sum_{s_q} x_k / \pi_k \right) > K \text{ sätts } C_q = K / R_q \text{ och i annat fall sätts } C_q = 1.$$

Parametern K väljs för att kontrollera påverkan från den syntetiska estimatormen, ju större K desto mer liknar den kvotestimatormen (19). En nackdel med denna typ av estimator är att R_q kan vara större än K trots att n_q är liten och då kommer estimatormen att vara behäftad med stort urvalsfel för denna redovisningsgrupp.

En annan typ av estimatorer baserar sig på Baysianska idéer, där man även utnyttjar den eventuella kunskapen om hur parametervärdena t_q fördelar sig. Den finns många varianter av estimatorer som faller ut från denna ansats, men för att konkretisera metoden presenteras här den estimator Fay and Herriot [1977] använder för att skatta per capita inkomst för små geografiska områden i USA. Den parameter man skattar är alltså av formen

$$\bar{y}_q = \frac{t_q}{N_q}$$

De har konventionella skattningar $\hat{y}_q$, som baserar sig på urvalsobservationerna.

Problemet är att de är behäftade med stora urvalsfel, dvs de (skattade) varianserna D_q är stora. Fay and Herriot [1977] antar att $\hat{y}_q$ ($q=1, \dots, Q$) är oberoende normalfördelade,

d v s $\hat{y}_q \sim \text{indN}(\bar{y}_q, D_q)$. Dessutom har man tillgång till hjälpvariabler $x_1, x_2, \dots, x_J$ -
 d v s för varje redovisningsgrupp q har man tillgång till en vektor $x_q = (x_{1q}, \dots, x_{Jq})'$, som
 man menar har ett linjärt samband med $\bar{y}_q$ av formen $E[\bar{y}_q] = x_q' \beta$. För enkelhets
 skull antar de att variansen för $\bar{y}_q$ är konstant ($=A$) och att $\bar{y}_q \sim \text{indN}(x_q' \beta, A)$.

Dessutom antar de att β har en apriorifördelning som är likformig på ett obestämt
 intervall (diffus fördelning). De okända parametrarna A resp β skattas sedan utifrån
 data, vilket är skälet till att metoden fått namnet Empirical Bayes. Utnyttjande av
 apriorikunskapen samt urvals- och hjälpinformationen ger en estimator, som har formen
 av en sammansatt estimator:

$$\hat{y}_{qEB} = [A^* / (A^* + D_q)] \hat{y}_q + (D_q / (A^* + D_q)) \bar{y}_q^*, \text{ där } A^* \text{ är en skattning av } A \text{ som}$$

$$\text{erhålles ur följande ekvation } \sum_q \frac{(\hat{y}_q - \bar{y}_q^*)^2}{A^* + D_q} = Q - J \text{ och } \bar{y}_q^* \text{ är vägda regres-}$$

$$\text{sionsestimater (vikten} = A + D_q).$$

Uträkningen av A^* kräver ett iterativt förfarande, vilket beskrivs i Fay and Herriot [1977].

5.2 Generaliserade regressionsestimatorer med dämpad korrektionsterm.

Särndal and Hidioglou [1989] föreslår följande "Dampened regression estimator":

$$\hat{t}_{qDre} = \sum_{U_q} \hat{y}_k + K_q \sum_{s_q}^v e_{ks}, \text{ där } K_q = (\hat{N}_q / N_q)^{\lambda-1} \text{ med} \quad (46)$$

$$\lambda = 0 \text{ om } \hat{N}_q \geq N_q$$

och

$$\lambda = \lambda_0 \text{ om } \hat{N}_q < N_q$$

λ_0 är en konstant som väljs på lämpligt sätt - Särndal and Hidioglou föreslår $\lambda_0 = 2$ för allmänt bruk.

Cassel [1984] föreslår att K_q får utgöra en skattning av den "optimala" faktorn (minimerar mean square error för estimatoren under förutsättning att den syntetiska termen och korrektionstermen är okorrelerade). K_q har då formen

$$K_q = 1 - \text{var}(\hat{B}_q) / (\hat{B}_q)^2, \text{ där } \hat{B}_q = \sum_{s_q}^{\vee} \hat{e}_{ks} \text{ och } \text{var}(\hat{B}_q) \text{ är en skattning}$$

av variansen för $\hat{B}_q$

6. Avslutning

Denna rapport är skriven i avsikt att ge en översikt av i första hand estimatorer för små redovisningsgrupper. För att hålla nere omfattningen av rapporten har den skrivits relativt komprimerat och av detta skäl har även en del mer udda estimatorer utelämnats. Naturligtvis finns det också estimatorer inom området som inte vi känner till. Därför är inte rapporten heltäckande, men förhoppningsvis ger den läsaren en hygglig inblick i området och hjälp att gå vidare i litteraturen.

Utvecklingen av estimatorer för små redovisningsgrupper har varit intensiv under det senaste decenniet och många tror att det blir ett viktigt område även i framtiden (se t ex Rao[1990]). Sverige, liksom övriga nordiska länder, har mycket hjälpinformation i register och därför kan vi bättre än andra utnyttja de tekniker som beskrivs i denna rapport.

Härledning av generellt variansuttryck.

Till varje design $p(s)$ kan en uppsättning stokastiska variabler införas genom $I_k = 1$ om objekt k kommer med i urvalet; $= 0$ annars.

Då är det lätt att visa följande

$$\text{Väntevärde: } E[I_k] = \pi_k$$

$$\text{Varians: } V(I_k) = \pi_k (1 - \pi_k)$$

$$\text{Kovarians: } C(I_k) = \pi_{kl} - \pi_k \pi_l = \Delta_{kl}$$

Med utnyttjande av denna indikatorvariabel kan π -estimatorn skrivas

$$\hat{t}_{q\pi} = \sum_{s_q} \check{y}_k = \sum_{U_q} I_k \check{y}_k$$

Variansen för π -estimatorn är då

$$\begin{aligned} V(\hat{t}_{q\pi}) &= \sum_{U_q} \check{y}_k^2 V(I_k) + 2 \sum_{k < l} \check{y}_k \check{y}_l C(I_k) = \\ &= \sum_{U_q} \pi_k (1 - \pi_k) \check{y}_k^2 + 2 \sum_{k < l} (\pi_{kl} - \pi_k \pi_l) \check{y}_k \check{y}_l \end{aligned}$$

Detta uttryck kan visas vara en utveckling av

$$\sum_{U_q} \sum_{k < l} \Delta_{kl} \check{y}_k \check{y}_l$$

Empirical best linear unbiased predictors.

På senare tid har ett antal komplicerade modeller med tillhörande estimatorer studerats, t ex Fuller and Harter (1987), Choudhry and Rao (1988). De består i regel av antagande om stokastiska intercept eller/och stokastiska lutningar på regressionslinjen. I denna bilaga presenteras ett antal av dessa ansatser, vilka huvudsakligen bygger på framställningen i Choudhry and Rao (1988). Genomgående antar de att urvalet är ett OSU och därmed finns inte π_k och π_{kl} i formlerna. Särndal and Hidiroglou (1989) har en mer generell ansats vad beträffar designen för en av nedanstående estimatortyper (modell 5).

Den estimator som erhålles från kriteriet "best linear unbiased predictor" är beroende av kvoten mellan två varianser. I den ansats som presenteras nedan har skattningar för den kvoten satts in. Därför har estimatorerna tilldelats prefixet "Empirical".

MODELL 1 och 2.

$Y_k = \beta_q x_k + \epsilon_k x_k^v$ för $k \in U_q$, där $v = 1/2$ för modell 1 och $v = 1$ för modell 2.

β_q är en stokastisk variabel med komponenterna $\beta_q = \beta + \alpha_q$, där β är en konstant.

Variablerna α_q är oberoende med $E[\alpha_q] = 0$ och $V[\alpha_q] = \sigma_a^2$.

Dessutom antas att $E[\epsilon_k] = 0$ och $V[\epsilon_k] = \sigma^2$.

Estimatorer

$$\hat{t}_{qEBLUP}(1, 2) = \sum_{s_q} y_k + \left(\sum_{U_q} x_k - \sum_{s_q} x_k \right) [\varphi_q b_q + (1 - \varphi_q) b], \text{ där}$$

Sätt $z_k = y_k / x_k^v$ och $w_k = x_k^{1-v}$

$$b_q = \sum_{s_q} z_k w_k / \sum_{s_q} w_k^2;$$

$$b = \left[\sum_s w_k z_k - \sum_q \left\{ (\varphi_q / \eta_q) \left(\sum_{s_q} z_k \right) \left(\sum_{s_q} w_k \right) \right\} \right] x$$

$$x \left[\sum_s w_k^2 - \sum_q \left\{ (\varphi_q / \eta_q) \left(\sum_{s_q} w_k \right)^2 \right\} \right]$$

där

$$\varphi_q = \hat{\sigma}_a^2 / [\hat{\sigma}_a^2 + \hat{\sigma}^2 / \eta_q]; \quad \eta_q = \sum_{s_q} w_k^2$$

och vidare

$$\hat{\sigma}^2 = (n - Q)^{-1} \left[\sum_s z_k^2 - \sum_q b_q^2 \eta_q \right];$$

$$\hat{\sigma}_a^2 = \left[\sum_s u_k^2 - (n - 1) \hat{\sigma}^2 \right] / \left[\sum_s w_k^2 - \sum_q \eta_q^2 / \sum_s w_k^2 \right]$$

där

$$\sum_s u_k^2 = \sum_q b_q^2 \eta_q - \left(\sum_s y_k \right)^2 \left(\sum_s w_k^2 \right) / \left(\sum_s x_k \right)^2$$

MODELL 3 och 4

$Y_k = \beta x_k + \alpha_q + \varepsilon_k x_k^v$ för $k \in U_q$, där $v = 1/2$ för modell 3 och $v = 1$ för modell 4.

Variablerna α_q är oberoende med $E[\alpha_q] = 0$ och $V[\alpha_q] = \sigma_a^2$.

Dessutom antas att $E[\varepsilon_k] = 0$ och $V[\varepsilon_k] = \sigma^2$.

Estimatorer

$$\hat{t}_{q\text{EBLUP}}(3, 4) = \sum_{s_q} y_k + b \left(\sum_{U_q} x_k - \sum_{s_q} x_k \right) + (N_q - n_q) a_q, \text{ där}$$

$$b = \left[\sum_s w_k z_k - \sum_q (\varphi_q / \eta_q) \left\{ \left(\sum_{s_q} w_k x_k^{-v} \right) \left(\sum_{s_q} z_k x_k^{-v} \right) \right\} \right] / \left[\sum_s w_k^2 - \sum_q (\varphi_q / \eta_q) \left(\sum_{s_q} w_k x_k^{-v} \right)^2 \right]$$

och

$$a_q = (\varphi_q / \eta_q) \sum_{s_q} (z_k x_k^{-v} - b w_k x_k^{-v})$$

där

$$\varphi_q = \hat{\sigma}_a^2 / [\hat{\sigma}_a^2 + \hat{\sigma}^2 / \eta_q] \text{ och } \eta_q = \sum_{s_q} x_k^{-2v}$$

och vidare

$$\hat{\sigma}^2 = \left(\sum_s e_k^2 \right) / (n - Q - 1) \text{ och}$$

$$\hat{\sigma}_a^2 = \left[\sum_s u_k^2 - (n - 1) \hat{\sigma}^2 \right] / \left[\sum_q \eta_q - \sum_q d_q^2 / \sum_s w_k^2 \right]$$

$$\sum_s u_k^2 = \sum_s z_k^2 - \left(\sum_s z_k w_k \right)^2 / \sum_s w_k^2$$

$$\sum_s e_k^2 = \sum_s z_k^2 - \left[\sum_s w_k^2 - \sum_q d_q^2 / \eta_q \right]^{-1} \left[\sum_s w_k z_k - \sum_q d_q c_q / \eta_q \right]^2 -$$

$$- \sum_q c_q^2 / \eta_q$$

där

$$d_q = \sum_{s_q} w_k x_k^{-v} \text{ och } c_q = \sum_{s_q} z_k x_k^{-v}$$

MODELL 5

$$Y_k = \beta x_k + \alpha_q x_k^{1/2} + \varepsilon_k x_k^{1/2} \text{ för } k \in U_q.$$

Variablerna α_q är oberoende med $E[\alpha_q] = 0$ och $V[\alpha_q] = \sigma_a^2$.

Dessutom antas att $E[\varepsilon_k] = 0$ och $V[\varepsilon_k] = \sigma^2$.

Estimator.

$$\hat{t}_{q\text{EBLUP}}(5) = \sum_{s_q} y_k + b \sum_{s_q} x_k + a_q \left(\sum_{U_q} x_k^{1/2} - \sum_{s_q} x_k^{1/2} \right)$$

där

$$a_q = (\varphi_q / n_q) \left(\sum_{s_q} z_k - b \sum_{s_q} w_k \right)$$

och

$$b = \left[\sum_s w_k z_k - \sum_q (\varphi_q / n_q) \left\{ \left(\sum_{s_q} w_k \right) \left(\sum_{s_q} z_k \right) \right\} \right] /$$

$$/ \left[\sum_s w_k^2 - \sum_q (\varphi_q / n_q) \left(\sum_{s_q} w_k \right)^2 \right]$$

där

$$\varphi_q = \hat{\sigma}_a^2 / \{ \hat{\sigma}_a^2 + \hat{\sigma}^2 / n_q \}$$

och vidare

$$\hat{\sigma}_a^2 = \left[\sum_s u_k^2 - (n-1) \hat{\sigma}^2 \right] / \left[n - \sum_q \left(\sum_{s_q} w_k \right)^2 / \sum_s w_k^2 \right]$$

$$\sum_s u_k^2 = \sum_s z_k^2 - \left(\sum_s z_k w_k \right)^2 / \sum_s w_k^2$$

$$\sum_s e_k^2 = \sum_s z_k^2 - \left[\sum_s w_k^2 - \sum_q d_q^2 / n_q \right]^{-1} \left[\sum_s w_k z_k - \sum_q d_q c_q / n_q \right]^2 -$$

$$- \sum_q c_q^2 / n_q$$

där

$$d_q = \sum_{s_q} w_k \text{ och } c_q = \sum_{s_q} z_k$$

Sammanställning av artiklar om skattningar för små redovisningsgrupper.

1. Inledning

Skattningar för "stora" redovisningsgrupper är sedan lång tid tillbaka behandlad i den statistiska litteraturen och de finns också i de flesta böcker om "Survey sampling". Det finns därför inget större skäl att göra en sammanställning av referenser inom det området (de viktigaste böckerna anges dock). Däremot är metoder för skattning av parametrar för små redovisningsgrupper av betydligt färskare datum och det är svårare att via böcker få en överblick av området.

Basen i följande sammanställning av artiklar över skattningar för små redovisningsgrupper utgör "A Bibliography for Small Area Estimation" presenterad i *Survey Methodology*, 1983, vol. 9, No. 2. Kompletteringen och uppdateringen baserar sig på rapportskrivarens deltagande i konferenser i ämnet, löpande uppföljning av tidskrifter, etc. Naturligtvis har det inte varit möjligt att göra en fullständig förteckning, men den bör ändå utgöra en god vägledning för att tränga in i ämnet.

En del referenser är inte fullständiga vad gäller tidsangivelse, var artikeln presenteras, etc, men i de fallen finns en kopia av artiklarna i U/STM-Ö:s samling.

2. Sammanställning

Bailar, B.A. (1979). Discussion. Synthetic Estimates for Small Areas, NIDA Research Monograph 24, 54-59.

Battese, G.E., and Fuller, W.A. (1981). Prediction of crop areas using survey a satellite data. 1981 Proc. Surv. Meth. Res. Sec., American Statistical Association, 500-505.

Brackstone, G.J. (1986). Small area data: Policy Issues and technical challenges, Small Area Statistics: An International Symposium, R.Platak, J.N.K. Rao, C.E. Särndal and M.P. Singh eds., Wiley, New York, 1-22.

Baxter, R., and Williams, I. (1978). Population forecasting and uncertainty at the national and local scale. Progress in Planning 4, Pergamon Press, Great Britain, 1978.

Ben-Akiva, M., and Lerman, S.R. (1975). Use of forecasting models in transportation planning. Conference on Population Forecasting for Small Areas, Oak Ridge, In., Oak Ridge Associated Universities, Inc., In.

Bender, R.K., and Verma, R.B.P. (1983). Translation for demographic data between overlapping subprovincial areas. Presented at the American Statistical Association meeting, Toronto.

Bogue, D.J. (1950). A technique for making extensive population estimates. JASA45, 149-163.

Bogue, D.J., and Duncan, B.D. (1959). A Composite Method for Estimating Postcensal Population of Small Areas by Age, Sex and Colour. National Office of Vital Statistics. Vital Statistics -- Special Reports 47, No. 6.

Bousfield, M.V. (1977). Intercensal estimation using a current sample and census data. Review of Public Data Use 5, 6-15.

Box, E.P., and Tiao, G.C. (1973). Bayesian Inference in Statistical Analysis. Addison Wesley, Cambridge, Mass.

Bresee, J.C. (1975). Forecasting population for energy plant sites. Conference on Population Forecasting for Small Areas, Oak Ridge, In., Oak Ridge Associated Universities, Inc., In.

Cardenas, M., Craig, M.E., and Blanchard, M. (1979). Small area estimators: county crop acreage estimates using LANDSAT data. Review of Public Data Use 7, 23-28.

Carrol, S.J., Caggiano, M.N., McCarty, K.F., Morrison, P.A., and Quint, B. (1980). City Data: a Catalogue of Data Sources for Small Cities. RAND Corp., Santa Monica, Ca.

Cassel, C.M. Kristiansson, K.E., Råbäck, G., and Wahlström, S.(1987). Using model-based estimation to improve the estimate of unemployment on a regional level in the Swedish Labor Force Survey, Small Area Statistics: An International Symposium, R.Platak, J.N.K. Rao, C.E. Särndal and M.P. Singh eds., Wiley, New York, 141-174.

Chambers, R.L., and Feeney, G.A. (1977). Log-linear models for small area estimation. Presented at the Joint Conference of the CSIRO Division of Mathematics & Statistics & the Australian Region of the Biometrics Society. Newcastle, Australia, Biometrics Abstract No. 2655.

Chambers, R.L. (1988). Limited estimation of domain means. Presented at a symposium on Small Area Statistics in New Orleans.

Choudhry, G.H. (1979). Synthetic estimation for special areas the economic region level. Unpublished memorandum, Census and Household Survey Methods Division, Statistic Canada.

Choudhry, G.H., and Hidioglou, M.A. (1987). Small area estimation: Some investigations at Statistics Canada. Bullentin of the International Statistical Institute, 52, Book 4, 451-468.

Choudhry, G.H., and Rao, J.N.K. (1988). Evaluation of small area estimators: An empirical study. Presented at a symposium on Small Area Statistics in New Orleans.

Chu, S.F. (1974). On the use of the regression method in estimating regional population. International Statistical Review 42, 17-28.

Cochran, W.G.(1977). Samplig techniques. John Wiley & Sons, New York.

Cohen, S.B. (1978). A modified approach to small area estimation. Synthetic Estimates for Small Areas, NIDA Research Monograph 24, 98-134.

Cohen, S.B., and Kalsbeek, W.D. (1977). An alternative strategy for estimating the parameters of local areas. 1977 Proc. Soc. Statist. Sec., American Statistical Association, 781-785.

Cohen, Reuben (1978). Drug Abuse Applications: Some regression explorations with National Survey Data. Synthetic Estimates for Small Areas, NIDA Research Monograph 24, 194-213.

Conopask, J.V. (1978). A Data-Pooling Approach to Estimate Unemployment Multipliers for Small Regional Economies. Economics, Statistics and Co-operatives Service, Washington, D.C.

Cronkhite, F.R. (1987). Use of regression techniques for developing state and unemployment estimates. Small Area Statistics: An International Symposium, R.Platak, J.N.K. Rao, C.E. Särndal and M.P. Singh eds., Wiley, New York, 160-174.

Crosetti, A.H., and Schmitt, R.C. (1956). A method of estimating the intercensal population of counties. JASA 51, 587-590.

Dagum, E.B., Hidioglou, and Morry, M. (1987). Sensitivity of small area estimators to misclassification and conceptual differences of variables. Small Area Statistics: An International Symposium, R.Platak, J.N.K. Rao, C.E. Särndal and M.P. Singh eds., Wiley, New York, 175-197.

Deer, G.E. (1975). Forecasting population of small areas to facilitate state health-care planning and development. Conference on Population Forecasting for Small Areas, Oak Ridge, In., Oak Ridge Associated Universities, Inc., In.

Delbert, J.E. (1981). A Technique for Estimating Age Specific Net Migration for Short-term Projections of Country Population. Population Research Center, University of Texas.

Deming, W.E., and Stephan, F.F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Annals of Mathematical Statistics 11, 427-444.

Dempster, A.P., and Raghunathan, T.E. (1987). Using a covariate for small area estimation: A common sense Bayesian approach, Small Area Statistics: An International Symposium, R.Platak, J.N.K. Rao, C.E. Särndal and M.P. Singh eds., Wiley, New York, 1-22.

Des Raj (1972). The design of sample surveys. McGraw-Hill.

Dominion Bureau of Statistics (1967). Population Estimates for Counties and Census Divisions, Cat. No. 91-206.

Drew, J.D. and Choudhry, G.H. (1979). Small area estimation. Unpublished memorandum, Census and Household Survey Methods Division, Statistics Canada.

Drew, J.D., Singh, M.P., and Choudhry, G.H. (1982). Evaluation of small area estimation techniques for the Canadian Labour Force Survey. Survey Methodology 8, 17-47.

Earwaker, S., Brien, N., and Gosselin, J.F. (?). Model-based small area estimation of employment: A Canadian experience.

Efron, B., and Morris, C. (1973). Stein's estimation rule and its competitors - and empirical Bayes approach. *JASA* 68, 117-130

Efron, B., and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *JASA* 70, 311-319.

Elliot, P. (1988). Small area statistics in the evaluation of environmental. Presented at a symposium on Small Area Statistics in New Orleans.

Elvers, E., Särndal, C.E., Wretman, J.H., and Örnberg, G. (1985). Regression analysis and ratio analysis for domains, a randomization theory approach. *Canadian Journal of Statistics*, 13, 185-199.

Ericksen, E.P. (1971). A method of combining sample survey data and symptomatic indicators to obtain estimates for local areas. Unpublished Ph.D. thesis, University of Michigan, Ann Arbor, Michigan.

Ericksen, E.P. (1973). A method of combining sample survey data and symptomatic indicators to obtain population estimates for local areas. *Demography* 10, 137-160.

Ericksen, E.P. (1973). Recent developments in estimation for local areas. *Proc. Soc. Statist. Sec., American Statistical Association*, 37-41.

Ericksen, E.P. (1974). A regression method for estimating population changes of local areas. *JASA* 69, 867-975.

Ericksen, E.P. (1974). Developments in statistical estimation for local areas. *Census Tract Papers Series GE-40 No. 10*, U.S. Government Printing Office, Washington, D.C.

Ericksen, E.P. (1975). Outliers in regression analysis when measurement error is large. 1975 *Proc. Soc. Statist. Sec., American Statistical Association*, 412-417.

Ericksen, E.P. and Kadane, J.B. (1987). Sensitivity analysis of local estimates of undercount in the 1980 U.S. Census. *Small Area Statistics: An International Symposium*, R. Platek, J.N.K. Rao, C.E. Särndal and M.P. Singh eds., Wiley, New York, 23-45.

Ervin, D.J. (1982). County population estimation with early release data: A new approach to data quality determination. Presented at the Population Association of America meeting, San Diego.

Espenshade, T., Hobbs, F.B. and Pol, L.G. (1981). An experiment in estimating postcensal age distribution of state population from death registration data. *Review of Public Data Use* 9, 97-114.

Fay, R.E. (1974). Statistical considerations in estimating the current population of the U.S. Unpublished PH.D. thesis, Department of Statistics, University of Chicago.

Fay, R.E., and Herriot, R. (1977). Estimates of income for small places: An application of James-Stein procedures to census data. *JASA* 74, 269-277.

Fay, R.E. (1978). Some recent Census Bureau applications regression techniques to estimation. *Synthetic Estimates for Small Areas*, NIDA Research Monograph 24, 155-184.

Fay, R.E. (1987). Small domain estimation through components of variance models, *Bulletin of the International Statistical Institute*, 52, Book 4, 451-468.

Fay, R.E. (1987). Application of multivariate regression to small domain estimation. *Small Area Statistics: An International Symposium*, R. Platek, J.N.K. Rao, C.E. Särndal and M.P. Singh eds., Wiley, New York, 91-102.

Fay, R.E. (1988). Empirical Bayes estimation for multiple characteristics. Presented at a symposium on Small Area Statistics in New Orleans.

Federal Statistical Office of Switzerland (1986). Small area estimation - An empirical study. Presented at the 1986 Conference on Statistical Methodology, United Nations.

Federal Statistical Office of the Federal Republic of Germany (1986). Problems of small area estimation, including the use of censuses and registers for such purposes. Presented at the 1986 Conference on Statistical Methodology, United Nations.

Feeney, G. (1983). A research proposal for the estimation of unemployment at the regional level. Unpublished, Australian Bureau of Statistics.

Feeney, G.A. (1987). The estimation of the number of unemployed at the small area level. *Small Area Statistics: An International Symposium*, R. Platek, J.N.K. Rao, C.E. Särndal and M.P. Singh eds., Wiley, New York, 198-238.

Ferguson, A.D. (1975). The necessity of forecasting for education planners. Conference on Forecasting for Small Areas, Oak Ridge, In., Oak Ridge Associated Universities, Inc., In.

Fienberg, S.E. (1978). The analysis of cross-classified categorical data. The MIT Press.

Fisher, A., Fisher, W., and Starler, N. (1975). Small area economic analysis: Are Dunn and Bradstreet the answer? *Northeast Regional Science Review* 5, 270-276.

Ford, B.L. (1981). The Development of County Estimates in North Carolina. USDA Staff Report AGES 811119, Washington, D.C.

Ford, B.L., Bond, D., and Carter, N. (1983). Combining Historical and Current Data to Make District and County Estimates for North Carolina. USDA Staff Report AGES 8301, Washington, D.C.

Ford, B.L., Bond, D., and Carter, N.J. (1983). Research into small area estimation at the U.S. Department of Agriculture. 1983 Proc. Sec. on Survey Research Methods, American Statistical Association, 232-237.

Freeman, D.H., and Koch, G.G. (1976). An asymptotic covariance structure for testing hypotheses on raked contingency tables for complex sample surveys. 1976 Proc. Soc. Statist. Sec., American Statistical Association, 330-335.

Froland, C. (1978). *Synthetic Estimates as an Approach to Needs Assessment: Issues and Experience*. Synthetic Estimates for Small Areas, NIDA Research Monograph 24, 246-258.

Fuller, W.A., and Battese, G. (1981). Regression estimation for small areas. Appendix G (pp. 572-586) in *Rural America in Passage: Statistics for Policy*, National Academy Press, Washington, D.C.

Fuller, W.A. and Harter, R. (1987). The multivariate components of variance model for small area estimation, *Small Area Statistics: An International Symposium*, R. Platek, J.N.K. Rao, C.E. Särndal and M.P. Singh eds., Wiley, New York, 103-123.

Fuller, W.A. (1986). Small area estimation as a measurement error problem. Presented at a symposium on small domain estimation in Oslo.

Ghangurde, P.D., and Singh, M.P. (1976). Synthetic estimation in the LFS. Technical Report, Household Surveys Development Division, Statistics Canada.

Ghangurde, P.D., and Singh, M.P. (1977). Synthetic estimation in periodic household surveys. *Survey Methodology* 3, 152-181.

Ghangurde, P.D., and Singh, M.P. (1977). Evaluation of synthetic estimation in the LFS. Technical Report, Household Surveys Development Division, Statistics Canada.

Ghangurde, P.D., and Singh, M.P. (1978). Evaluation of efficiency of synthetic estimates. 1978 Proc. Soc. Statist. Sec., American Statistical Association, 52-61.

Ghangurde, P.D. (1978). Evaluation of synthetic estimation in self-representing areas of LFS. Technical Memorandum, Household Surveys Development Division, Statistics Canada.

Ghangurde, P.D., and Gray, G.B. (1981). Estimation for small areas in household surveys. *Communications in Statistics A10*(22), 2327-2338.

Goldberg, D., and Balakrishnan, T.R. (1961). A partial evaluation of four estimating techniques. Department of Sociology, University of Michigan, Ann Arbor.

Goldberg, D., Rao, V.R., and Namboodiri, N.R. (1964). A test of the accuracy of the ratio-correlation population estimates. *Land Economics* 40, 100-102.

Gonzalez, M.E. (1973). Use and evaluation of synthetic estimates. 1973 Proc. Soc. Statist. Sec., American Statistical Association, 33-36.

Gonzalez, M.E., and Waksberg, J.E. (1973). Estimation of the error of synthetic estimates. Presented at the first meeting of the International Association of Survey Statisticians, Vienna, Austria.

Gonzalez, M.E., and Hoza, C. (1975). Small area estimation of unemployment. 1974 Proc. Statist., American Statistical Association, 437-443.

Gonzalez, M.E., and Hoza, C. (1978). Small area estimation with application to unemployment and housing estimates. *JASA* 73, 7-15.

Gonzalez, M.E. (1978). Case studies on the use and accuracy of synthetic estimates: unemployment and housing applications. *Synthetic Estimates for Small Areas*, NIDA Research Monograph 24, 142-154.

Government of British Columbia, Central Statistics Bureau (1980). *British Columbia School District Population Estimates*. Unpublished, Ministry of Industry and Small Business Development.

Government of British Columbia, Central Statistics Bureau (1980). *British Columbia Municipal Population Estimates*. Unpublished, Ministry of Industry and Small Business Development.

Gray, G.B., and Ghangurde, P.D. (1975). On a ratio estimate with post-stratified weighting. *Survey Methodology* 1, 134-144.

Greenberg, M.R. (1975). Use of the mathematical extrapolation methods to forecast population of small areas. Conference on Population Forecasting for Small Areas, Oak Ridge, In., Oak Ridge Associated Universities, Inc., In.

Greenberg, M.R., Krueckeber, D.A., Michaelson, C.O., Mautner, R., and Newman, N. (1975). *Local Population and Employment Projection Techniques*. The Center for Urban Policy Research.

Harms, L.T., James, R., and Springer, R.C. (1966). *Projective Models of Employment by Industry and by Occupation for Small Areas: A Case Study*. Temple University, Philadelphia, Pa.

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953). *Sample Survey Methods and Theory*. John Wiley & Sons, Inc., New York.

Heldal, J. (1983). Kombinasjon av registre og utvalg. Problemstillinger og metodearbeid i Norge. Inl  gg i Komet-m  te i Helsingfors.

Heldal, J. (1986). Combined estimation for small domains using regression estimators. Presented at a symposium on small domain estimation in Oslo.

Henry, L. (1976). *Population Analysis and Models*. Edward Arnold, London.

Heeringa, S.G. (1981). Small area estimation: Prospects for the survey of income and participation, A review of the literature and interim report on the methodology. Survey Research Center, Institute for Social Research, University of Michigan.

Heeringa, S.G. (1981). Synthetic estimation of totals for state subpopulations. Survey Research Center, Institute for Social Research, University of Michigan.

Heeringa, S.G. (1982). Error analysis in synthetic estimation of totals for state subpopulations. Survey Research Center, Institute for Social Research, University of Michigan.

Herring, D.E. (1974). *Procedures for preparing annual population estimates for counties and census divisions*. Unpublished, Demography Division, Statistics Canada.

Hidioglou, M.A., Morry, M., Dagum, E.B., Rao, J.N.K., and Särndal, C.E. (1984). Evaluation of alternative small area estimators using administrative records. *Proc. Soc. Statist. Sec., American Statistical Association*, 307-313.

Hidioglou, M.A., and Särndal, C.E. (1985). An empirical study of some regression estimators for small domains, *Survey Methodology*, 11, 15-31.

Holt, D., Smith, T.M.F., and Tomberlin, T.J. (1977). Synthetic estimation for small sub-groups of a population. Unpublished Technical Report, University of Southampton, England.

Holt, D., Smith, T.M.F., and Tomberlin, T.J. (1979). A model based approach to estimation for small sub-groups of a population. *JASA* 74, 405-410.

Hughes, P.J., and Choy, C.Y. (1982). Regression Techniques for Local Government Area Population Estimation, Australian Bureau of Statistics Occasional Paper 1982/1.

Isaki, T.C., Schultz, L.K., Smith, P.J., and Diffendal, G.J. (?). Small area estimation research for census undercount - progress report. Statistical Research Division, Bureau of the Census.

Irwin, R. (1978). Aggregate medicare enrollment by age, sex and race as a resource in analysing demographic change for local areas. Prepared for presentation at the NBER Workshop on Population Analysis with Social Security Research Files.

Irwin, R. (1975). Methods and data sources for small areas. Conference on Population Forecasting for Small Areas, Oak Ridge, In., Oak Ridge Associated Universities, Inc., Tn.

Irwin, R. (1975). Use of the cohort-component method in population projections for small areas. Conference on Population Forecasting for Small Areas, Oak Ridge, In., Oak Ridge Associated Universities, Inc., In.

Heeringa, S.G. (1982). Error analysis in synthetic estimation of totals for state subpopulations. Survey Research Center, Institute for Social Research, University of Michigan.

James, W, and Stein, C. (1961). Estimation with quadratic loss. *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 361-379, University of California Press, Berkeley.

Jones, H.L. (1974). Jackknife estimation of functions of stratum means. *Biometrika* 61, 343-348.

Jones, D.H., and Coopersmith, L.A. (1976). A ratio estimator of the total of a sub-population. *Communications in Statistics A5*(3), 251-260.

Kalsbeek, W.D. (1973). A method for obtaining local postcensal estimates for several types of variables. Unpublished PH.D. theses, University of Michigan, Ann Arbor, Michigan.

Kendal, M.C. (1975). Labour market models. Conference on Population Forecasting for Small Areas, Oak Ridge, In., Oak Ridge Associated Universities, Inc., In.

- Kish, L. (1965). *Survey Sampling*. John Wiley & Sons, Inc., New York.
- Kish, L. (1980). Design and estimation for domains. *The Statistician* 29, 209-222.
- Koch, G.G. (1973). An alternative approach to multivariate response error models for sample survey data with applications to estimators involving sub-class means. *JASA* 68, 906-913.
- Kott, P.S. (?). Robust small domain estimation using random effects modeling.
- Kovacevic, M.S. (1983). Estimation of totals on the commune's level in the livestock survey. Report to the 1983 Meeting on Statistical Methodology, Statistical Commission for Europe, U.N. Economic Council.
- Krebs, H.C., Feeney, D.T., Palit, C., Voss, P., and Kale, B. (1982). On measurement of error in estimating population change: a partial critique of the report of the panel on small area estimates of population and change. Presented at the Population Association of America meeting, San Diego.
- Kumar, S. and Rao, J.N.K. (1984). Logistic regression analysis of Labour Force Survey data. *Survey Methodology*, 10, 62-81.
- Laake, P. og Langva, H.K. (1976). Estimering av sysselsettning i geografiske regioner: Om estimatorers skjevhet, varians og bruttovarians. *Artikler fra Statistisk Centralbyrå* nr 88.
- Laake, P. (1978). An evaluation of synthetic estimates of employment. *Scandinavian Journal of Statistics* 5, 57-60.
- Laake, P. (1979). A prediction approach to subdomain estimation in infinite populations. *JASA* 74, 355-358.
- Lalu, N.M. (1977). *Monitoring Population Change in Postcensal Years*, Edmonton Population Research Lab., University of Alberta.
- Leblond, Y. (1989). *Contributions à la theorie d'estimation pour souspopulations*. Oppublicerad doktorsavhandling. Université de Montréal.
- Lee, E.S., and Goldsmith, H.F. (1978). Population estimates for small area analysis. Presented at the conference of the National Institute for Mental Health, Annapolis.
- Levy, P.S. (1971). The use of mortality data in evaluating synthetic estimates. 1971 *Proc. Soc. Statist. Sec.*, American Statistical Association, 328-331.
- Levy, P.S., and French, D.K. (1977). Synthetic estimation of state health characteristics based on the Health Interview Survey. *Vital and Health Statistics Series 2*, No. 75, U.S. Department of Health, Education & Welfare.
- Levy, P.S. (1978). Small area estimation - Synthetic and other procedures. *Synthetic Estimates for Small Areas*, NIDA Research Monograph 24, 4-19.
- Lui, K.U. and Cumberland, W.G. (?). A model-based approach: Composite estimators for small area.

Lundstrom, S. (1983). Estimation for small domains: two studies using combined data from censuses, surveys and registers. Report to the 1983 Meeting on Statistical Methodology, Statistical Commission for Europe, U.N. Economic Council.

Lundström, S. (1986). Estimating household characteristics in Swedish municipalities using survey and register data. Presented at the 1986 Conference on Statistical Methodology, United Nations.

Lundström, S. (1987). An evaluation of small area estimation methods: The case of estimating the number of non-married Cohabiting persons in Swedish municipalities. Small Area Statistics: An International Symposium, R.Platak, J.N.K. Rao, C.E. Särndal and M.P. Singh eds., Wiley, New York, 141-174.

Lundström, S. (1988). Experiences on small domain estimation methods at Statistics Sweden. Presented at a symposium on Small Area Statistics in New Orleans.

Lundström, S. (1990). Estimatorer för små redovisningsgrupper. URVAL nr 20.

Mandell, M., and Tayman, J. (1982). Measuring temporal stability in regression models of population estimation, *Demography* 19, 135-146.

Mandell, M., and Tayman, J. (1979). A test of the ratio-correlation methods of population estimation in a high growth state. Presented at the Southern Regional Demographic Group meeting, Myrtle Beach.

Marker, D.A. (1983). Organization of small area estimators. 1983 Proc. Soc. Statist. Sec., American Statistical Association.

Martin, J.H., and Serow, W.J. (1978). Estimating demographic characteristics using the ratio-correlation method. *Demography* 15, 223-233.

Martin J.H., and Spar, M.A. (1981). Increasing the of postcensal country population estimates: the case for area specific methods. Presented at the Population Association of America meeting, Washington, D.C.

Marton, A. (1986). Synthetic estimates for small areas: Problems and results of simulation experiments, *Statistical Journal of the United Nations*, 71-80.

McCleary, A. (1982). Small area population estimation: A note on the state of the art. *Scottish Geographical Magazine* 98 (3).

McCullagh, P. and Zidek, J.V. (1987). Regression methods and performance criteria for small area population estimation. Small Area Statistics: An International Symposium, R.Platak, J.N.K. Rao, C.E. Särndal and M.P. Singh eds., Wiley, New York, 62-74.

Morris, C.E. (1980). Evaluation design and the use of low-precision benchmarks, *Estimating Population and Income of Small Areas*, National Academy Press, Washington, D.C., 229-231.

Morrison, P.A. (1975). Forecasting population of small areas (an overview). Conference on Population Forecasting for Small Areas, Oak Ridge, In., Oak Ridge Associated Universities, Inc., In.

Morrison, P.A. (1982). *Different Approaches to Monitoring Local Demographic Change*, Rand Corporation Paper No. P.6743, Santa Monica.

Morrison, P.A., and Relles, D.A. (1975). A method for monitoring small area population changes in cities. *Review of Public Data Use* 3(2), 10-15.

Namboodiri, N.K. (1972). On the ratio-correlation and related methods of sub-national population estimation. *Demography* 9, 443-453.

Namboodiri, N.K., and Lalu, N.M. (1971). The average of several simple regression estimates as an alternative to multiple regression estimates in postcensal and intercensal population estimation: a case study. *Rural Sociology* 36, 187-194.

Nameketa, T. (1974). *Synthetic state estimates of work disability*. Unpublished Ph.D. thesis, University of Illinois, Champagne, Illinois.

National Center for Health Statistics (1968). *Synthetic State Estimates of Disability*. P.H.S. Publication No. 1759, U.S. Government Printing Office, Washington, D.C.

National Center for Health Statistic (1977). *State Estimator of Disability and Utilization of Medical Services, 1969-1971*. D.H.E.W. Publication No. (HRA) 77-1241, U.S. Government Printing Office, Washington, D.C.

National Center for Health Statistics (1977). *Synthetic Estimation of State Health Characteristics Based on the Health Interview Survey*. D.H.E.W. Publication No. (PHS) 78-1349, U.S. Government Printing Office, Washington, D.C.

National Center for Health Statistic (1978). *State Estimates of Disability and Utilization of Medical Services, 1974-1976*. D.H.E.W. Publication No. (PHS) 78-1241, U.S. Government Printing Office, Washington, D.C.

National Center for Health Statistic (1980). *Small Area Estimation: An Empirical Comparison of Conventional and Synthetic Estimators for States*. D.H.E.W. PUB. No. 80-1356, U.S. Government Printing Office, Washington, D.C.

National Research Council (1980). *Estimating Population and Income of Small Areas*, National Academy Press, Washington, D.C.

Nicholls, A. (1977). *A regression approach to small area estimation*. Unpublished, Australian Bureau of Statistics, Canberra, Australia.

Norris, D.A., Britton, M., and Verma, R. (1982). *The Use of Administrative Records for Estimating Migration and Population*, Internal Revenue Service, Washington.

Office of Population, Censuses and Surveys (1980). *Estimating the population of local authorities*. *Population Trends* 20, 12-17.

Office of Population, Censuses and Surveys (1980). *Local Authority Population Estimation Methodology*, Occasional Paper 18, Population Statistics Division, OPCS.

Office of Population, Censuses and Surveys (1981). *The Revised Mid-1971 Population Estimates for Local Authorities Compared with the Original Estimates*. Occasional Paper 22, Population Statistics Division, OPCS.

O'Hare, W. (1976). Report on a multiple regression method for making population estimates. *Demography* 13, 369-379.

O'Hare, W.P. (1980). A note on the use of regression methods in population estimation. *Demography* 17, 341-343.

Pannekoek, J. (1989). Empirical Bayes estimation for small domains using logit models with application to unemployment rates. Unpublished, Department of Statistical Methods, Netherlands Central Bureau of Statistics.

Pham, D.Q. and Thomsen, I. (1988). Small area estimation - possibilities and limitations. Presented at a symposium on Small Area Statistics in New Orleans.

Pittenger, D., et al (1977). Making the housing unit population method work: a progress report. Presented at the Population Association of America meeting St. Louis.

Platek, R. (1980). Small area estimation. Unpublished, Census and Household Survey Methods Division, Statistics Canada.

Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean square error of small-area estimators, *JASA* 85, 163-171.

Promisel, D.M. (1978). Applications of synthetic estimates to alcoholism and problem drinking. *Synthetic Estimates for Small Areas*, NIDA Research Monograph 24, 63-87.

Purcell, D.E. (1970). Improving population estimates with the use of dummy variables. *Demography* 7, 87-91.

Purcell, N.J., and Linacre, (1976). Techniques for the estimation of small area characteristics. Presented at the 3rd Statistical Conference, Melbourne, Australia.

Purcell, N.J. (1979). Efficient small domain estimation. A categorical data analysis approach. Unpublished Ph.D. thesis, University of Michigan, Ann Arbor, Michigan.

Purcell, N.J., and Kish, L. (1979). Estimation for small domains. *Biometrics* 35, 365-384.

Purcell, N.J., and Kish, L. (1980). Postcensal estimates for local areas (or domains). *International Statistical Review* 48, 3-18.

Raby, R. (1980). Intercensal estimates of the population of Canada and the provinces, census metropolitan areas and census divisions. Unpublished, Demography Division, Statistics Canada.

Rao, C.R., and Shirazaki, M. (1978). Precision of individual estimates in simultaneous estimation of parameters. *Biometrika* 65, 23-30.

Rao, J.N.K. (1984). Some thoughts on small area estimation. Unpublished.

Rao, J.N.K. (1986). Synthetic estimators, SPREE, and Bestmodel-Based Predictors of small area means. Presented at a symposium on small domain estimation in Oslo.

Rao, J.N.K.(1989). The history and development of theoretical foundations of survey based estimation and statistical analysis. Unpublished.

Rasmussen, N. (1974). The use of driver license address change records for estimating interstate and intercounty migration. Presented at the Small Area Statistic Conference, St. Louis.

de Ree, S.J.M. (1989). Statistiek werkzame personen, een nieuwe schattningsmethode. Unpublished, Department of Statistical Methods, Netherlands Central Bureau of Statistics.

Rives, N.W. (1976). A modified housing unit method for small area population estimation. 1976 Proc. Soc. Statist. Sec., American Statistical Association, 717-720.

Rosenberg, H. (1968). Improving current population estimates through stratification. Land Economics 44, 331-338.

Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. Biometrika 57, 377-387.

Royall, R.M. (1973). Discussion of two papers on recent developments in estimation of local areas. 1973 Proc. Soc. Sec., American Statistical Association, 43-44.

Royall, R.M. (1974). Discussion of papers by Gonzalez and Ericksen. Census Tract Papers Series GE-40, No. 10, U.S.B.C., U.S. Government Printing Office, Washington, D.C.

Royall, R.M. (1977). Statistical theory of small area estimates - use of predictor models. Unpublished technical report presented to National Center for Healt Statistics.

Royall, R.M. (1978). Prediction models in small area estimation. Synthetic Estimates for Small Areas, NIDA Research Monograph 24, 63-87.

Schaible, W.L. (1975). A comparison of the mean square errors of the postratified, synthetic and modified synthetic estimators. Unpublished report, Office of Statistical Research, National Center for Health Statistics.

Schaible, W.L., Brock, and Schnack, G.A. (1977). An empiriacal comparison of the simple inflation, synthetic and composite estimators for small area statistics. 1977 Proc. Soc. Statist. Sec., American Statistical Association, 1017-1021.

Schaible, Wesley L., Brock, D.B., and Schnack, G.A. (1977). An empirical comparison of two estimators for small areas. Presented at the Second Annual Data Use Conference of the National Center for Health Statistics, Dallas.

Schaible, W.L. (1978). Choosing weights for composite estimates for small area statistics. 1978 Proc. Soc. Statist. Sec., American Statistical Association, 741-746.

Schaible, W. L. (1979). A composite estimator for small area statistics. Synthetic Estimates for Small Areas, NIDA Research Monograph 24, 36-53.

Schaible, W.L., Brock, D.B., Casady, R.J., and Schack, G.A. (1979). Small area estimation: An empirical comparison of conventional and synthetic estimators for states. Public Health Service Series 2-82 (N0. 80-1356), NCHS, D.H.E.W., U.S. Government Printing Office, Washington, D.C.

Schmitt, R.C., and Crosetti, A.H. (1954). Accuracy of the ratio-correlation method of estimation postcensal population. *Land Economics* 30, 279-280.

Schmitt, R.C., and Grier, J.M. (1966). A method of estimating the population of minor civil divisions. *Rural Sociology* 31, 355-361.

Serow, W.J., and Martin, J.H. (1977). Estimating demographic characteristics using the ratio-correlation method. *Demography* 15, 223-233.

Shryock, H.S., Siegel, J.S., and Assoc. The Methods and Materials of Demography. U.S. Government Printing Office, Washington, D.C.

Simmons, W.P. (1973). Adjustment of data-synthetic estimates. Presented at the first meeting of the International Association of Survey Statisticians, Vienna, Austria.

Singh, M.P., and Tessier, R. (1975). Som estimators for domain totals. *Survey Methodology* 1, 77-86.

Smith, T.M.F. and Brunstond, T.M. (1986). Time series methods for small areas. Presented at a symposium on small domain estimation in Oslo.

Snow, E.C. (1911). The application of the method of multiple correlation to the estimator of postcensal population. *Journal of the Royal Statistical Society* 74, 575-620.

Spar, M.A., and Martin J.H. (1979). Refinements to regression based estimates of postcensal population characteristics. *Review of Public Data Use* 7, 16-22.

Spjotvoll, E. and Thomsen, I. (1987). Application of some empirical Bayes methods to small area statistics. *Bulletin of the International Statistical Institute*, 52, Book 4, 435-450.

Starsinic, D.E., and Zitter, M. (1968). Accuracy of the housing unit method in preparing population estimates for cities. *Demography* 5, 475-484.

Starsinic, D.E. (1974). Development of population estimates for revenue sharing area. Census Tract Series GE-40, No. 10, U.S.B.C., U.S. Government Printing Office, Washington, D.C.

Statistics Canada (1979). Revised Annual Estimates of Population for Census Divisions. Catalogue 91-521.

Steahr, T.E., and Heston, J.F. (1983). Adjusted estimates of population by age and sex for towns in Connecticut, 1940-1980. Demographic Technical Paper 75-5, University of Connecticut.

Steel, D. and Poulton, J. (1988). Geografiphic estimates of underenumeration. Presented at a symposium on Small Area Statistics in New Orleans.

- Steel, D. (1988). Approaches to small area estimation at the Australian Bureau of Statistics. Presented at a symposium on Small Area Statistics in New Orleans.
- Stroud, T.W.F. (1987). Bayes and Empirical Bayes approaches to small area estimation. *Small Area Statistics: An International Symposium*, R. Platek, J.N.K. Rao, C.E. Särndal and M.P. Singh eds., Wiley, New York, 124-137.
- Strohmenger, C. (1980). Estimates, C. (1989). Estimates of population for census metropolitan areas. Unpublished, Demography Division, Statistics Canada.
- Sturdevant, T.R. (1978). Using available resources to generate small area data. Presented at the American Statistical Association meeting, San Diego.
- Swanson, D.A. (1977). Preliminary results of an evaluation of the utility of ridge regression for making county population estimates. Staff Document 40, Office of Financial Management, State of Washington.
- Swanson, D.A. (1978). An evaluation of "ratio" and "difference" regression methods for estimating small, highly concentrated populations: the case of ethnic groups. *Review of Public Data Use* 6, 18-27.
- Swanson, D.A. (1980). Allocation accuracy in population estimates: an overlooked criteria with fiscal implications. Presented at the American Statistical Association meeting, Houston.
- Swanson, D.A. (1980). Improving accuracy in multiple regression estimates of population using principles from casual modelling. *Demography* 17, 413-427.
- Swanson, D.A., and Tedrow, L.M. (1983). Improving the measurement of temporal change in regression models used for county population estimates. Presented at the American Statistical Association meeting, Toronto.
- Swensen, A.R. (1986). Combining surveys and registers in the 1980 population and household census in Norway: Synthetic estimation. Presented at a symposium on small domain estimation in Oslo.
- Särndal, C.E. (1984). Design consistent versus model dependent estimation for small domains. *JASA* (79), 624-631.
- Särndal, C.E. and Hidiroglou, M.A. (1989). Small domain estimation: A conditional analysis, *JASA* (84), 266-275.
- Tam, S.M. (1982). Postcensal estimates for local areas using current sample with census as the source of sampling frame. *Int. Statist. Review* 50, 125-134.
- Tayman, J. (1980). Total and occupied housing units as useful symptomatic indicators in population estimation models. Presented at the Federal Population Estimates Cooperative Meeting, Denver, Colorado.
- Terasvitra, T. (1981). Superiority comparisons of homogeneous linear estimators. Research Report No. 30, Department of Statistics, University of Helsinki.
- Thibault, N. (1980). Population estimates for census divisions. Unpublished, Statistics Canada.

Thérour, G. (1979). Estimation of under-enumeration of the population in small areas. Unpublished, Census and Household Survey Methods Division, Statistics Canada.

Thomsen, I. (1986). On the combination of surveys and registers. Presented at the 1986 Conference on Statistical Methodology, United Nations.

Thompson, J.R. (1978). Some shrinkage techniques for estimating the mean. *JASA* 63, 113-122.

Treasury, Economic and Intergovernmental Affairs (Ontario) (1977). Demographic information from administrative files: recent developments in Ontario. *Demographic Bulletin*.

United Nations (1950). The Preparation of Sampling Survey Reports. *Statistical Papers, Series C*, No. 1.

U.S.B.C. (1981). Small area population estimates methods and their accuracy and new metropolitan area definitions and their impact on the private sector (GE-41, No.6). U.S. Government Printing Office, Washington, D.C.

U.S.B.C. (1980). Population and per capita money income estimates for local areas: detailed methodology and evaluation. *Current Population Reports, Series P-25*, No. 699, U.S. Government Printing Office, Washington, D.C.

U.S.B.C. (1980). Small area statistics papers: An evaluation of small area data forecasting models and 1980 census small area statistics program (GE-41, No.6). U.S. Government Printing Office, Washington, D.C.

U.S.B.C. (1977). Interrelationship among estimates, surveys and forecasts produced by federal agencies. Presented at the American Statistical Association Conference on Small Area Statistics.

U.S.B.C. (1975a). Population estimates and projection. *Current Population Reports, Series P-25*, No. 580, U.S. Government Printing Office, Washington, D.C.

U.S.B.C. (1974). Intercensal estimates for small areas and public data files for research. Presented at the American Statistical Association Conference on Small Area Statistics.

U.S.B.C. (1973). Federal State co-operative program for local population estimates: Test results April 7, *Current Population Reports, Series P-26*, No. 21, U.S. Government Printing Office, Washington, D.C.

U.S.B.C. (1970). Inventory of state and local agencies preparing population estimates, survey of 1969. *Current Population Reports, Series P-25*, No. 454, U.S. Government Printing Office, Washington, D.C.

U.S.B.C. (1969). Estimates of the population of counties and metropolitan areas, July 1, 1966: A summary report. *Current Population Reports, Series P-25*, No. 427, U.S. Government Printing Office, Washington, D.C.

U.S.B.C. (1966). Methods of population estimation: Part I, Illustrative procedure of the Census Bureau's component method II. *Current Population Reports, Series P-25*, No. 339, U.S. Government Printing Office, Washington, D.C.

U.S.B.C. (1949). Illustrative examples of two methods of estimating the current population of small areas. Current Population Reports, Series P-25, No. 20, U.S. Government Printing Office, Washington, D.C.

U.S.B.C. (1947). Population special reports. Series P-47, No. 4, U.S. Government Printing Office, Washington, D.C.

Verma, R.B.P., and Basavarajappa, K.G., and Bender, R. (1982). New approaches to methods of estimating the population of census metropolitan areas. Presented to the Ad hoc Committee on Demography, Ottawa.

Verma, R.B.P., and Basavarajappa, K.G. (1982). Sub-provincial estimation of population in Canada: a review of estimation methods and prospects for development. Presented at the American Public Health Association meeting, Montreal.

Verma, R.B.P., Basavarajappa, K.G., and Bender, R. (1982). New approaches to methods of estimating the population of census divisions. Unpublished, Statistics Canada.

Verma, R.B.P., Basavarajappa, K.G., and Bender, R. (1983). The regression estimates of population for subprovincial areas in Canada. Presented at the American Statistical Association meeting, Toronto.

Verma, R.B.P., Basavarajappa, K.G., Bender, R.K., and Stepien, B. (1983). Generalized System for evaluation and production of total population estimates for sub-provincial areas. Unpublished, Statistics Canada.

Verma, R.B.P., and Basavarajappa, K.G. (1987). Recent developments in the regression method for estimation of population for small areas in Canada. Small Area Statistics: An International Symposium, R. Platek, J.N.K. Rao, C.E. Särndal and M.P. Singh eds., Wiley, New York, 141-174.

Verma, B.P. (1988). An overview of methods of population estimation. Presented at a symposium on Small Area Statistics in New Orleans.

Woodruff, R.S. (1966). Use of a regression technique to produce area breakdowns of the monthly national estimates of retail trade. JASA 62, 496-504.

Woodruff, R.S. (1976). A simple method of approximating the variance of a complicated estimate. JASA 66, 411-414.

Young, M.E. (1975). Population Projections for Small Areas (A Bibliography with Abstracts). National Technical Information Service, Springfield, Va.

Zitter, M., and Shryock, H.S. (1964). Accuracy of methods of preparing postcensal estimates for states and local areas. Demography 1, 227-241.

R & D Reports är en för U/ADB och U/STM gemensam publikationsserie, som fr o m 1988-01-01 ersätter de tidigare "gula" och "gröna" serierna. I serien ingår även **Abstracts** (sammanfattning av metodrapporter från SCB).

R & D Reports Statistics Sweden are published by the Department of Research & Development within Statistics Sweden. Reports dealing with statistical methods have green (grön) covers. Reports dealing with EDP methods have yellow (gul) covers. In addition, abstracts are published three times a year (light brown /beige/ covers).

Reports published earlier during 1991:

1991:1 Computing elementary aggregates in the Swedish consumer price
(grön) index (**Jörgen Dalén**)

Kvarvarande **beige** och **gröna** exemplar av ovanstående promemorior kan rekvireras från Inga-Lill Pettersson, U/LEDN, SCB, 115 81 STOCKHOLM, eller per telefon 08-783 49 56.

Kvarvarande **gula** exemplar kan rekvireras från Ingvar Andersson, U/ADB, SCB, 115 81 STOCKHOLM, eller per telefon 08-783 41 47.