



Kalibrering av vikter

**- beskrivning av tekniken och
de SCB-fall den prövats i**

Sixten Lundström

Med bidrag från

Stina Andersson

Pär Brundell

Jan Hörngren

Charlotte Jansson

Klas Lindström

Peter Lundqvist

Hassan Mirza

Britt och Anders Wallgren

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 2000:1. Kalibrering av vikter – beskrivning av tekniken och de SCB-fall den prövats i / Sixten Lundström.

Digitalt skapad fil, anpassad efter de digitaliserade delarna i serien. Statistiska centralbyrån (SCB) 2016.

Kalibrering av vikter

- beskrivning av tekniken och
de SCB-fall den prövats i

Sixten Lundström

Med bidrag från

Stina Andersson

Pär Brundell

Jan Hörngren

Charlotte Jansson

Klas Lindström

Peter Lundqvist

Hassan Mirza

Britt och Anders Wallgren

R&D Report 2000:1

Research - Methods - Development

Calibration of Weights

- description of the method and a review of calibration cases

Från trycket

Mars 2000

Producent

Statistiska centralbyrån, *Statistics Sweden*, metodenheten

Box 24300, SE-104 51 STOCKHOLM

Utgivare

Lars Lyberg

Förfrågningar

Sixten Lundström

sixten.lundstrom@scb.se

telefon 019- 17 64 96

Innehåll

	Sida
FÖRORD	1
1. Inledning	2
2. Huvudprinciperna i kalibreringstekniken	4
3. Teknisk beskrivning av kalibrering av vikter	6
4. CLAN97	9
5. Produkter där kalibrering har prövats	10
6. Sammanfattning och kommentarer	27
7. Kalibreringstekniken i framtiden	29
8. Referenser	31

FÖRORD

I föreliggande rapport beskrivs tekniken med kalibrering av vikter samt fall där tekniken har prövats. Vi kan inte säkert säga att alla fall har kommit med, men kartläggningen har skett genom en ”bred” förfrågan, vilket gör att vi tror att det uteblivna antalet är litet. De olika fallen har beskrivits av de personer som genomfört kalibreringen. Rapporten i övrigt har skrivits av Sixten Lundström, METOD¹. För att få ytterligare information om ett specifikt fall tas lämpligen kontakt med namngiven person. I de fall det finns bakomliggande rapporter har dessa nämnts i skrivningen.

Många personer har alltså varit inblandade i skrivningen och därför kan beskrivningarna ge ett heterogent intryck. Dessutom har fallen genomförts vid olika tidpunkter och därmed baserar sig på olika kunskapsläge.

Statistiska centralbyrån i mars 2000

Sixten Lundström

¹ Eva Elvers och Lennart Nordberg, båda ES/SES, har gett värdefulla synpunkter på tidigare versioner av denna rapport.

Kalibrering av vikter -

- beskrivning av tekniken och de SCB-fall den prövats i

1. Inledning

I en urvalsundersökning erhålles observationer för endast en delmängd av populationen, vilket innebär att varje observation i estimationen ”representerar” flera objekt i populationen (ibland har man dock totalundersökta strata). Vid beräkningen av skattningar multipliceras därför varje värde på målvariablerna med en vikt. Delmängden bestäms i första hand av urvalsförfarandet, men det sker också en reduktion genom att inte alla utvalda objekt svarar. I regel uppträder också täckningsbrister i en undersökning och för att reducera åtföljande fel i skattningarna krävs att vissa korrigeringar görs; i många fall görs korrigeringen med en delvikt i estimation. Inom ”traditionell” metodik består vikten för ett givet objekt av produkten av flera delvikter, vanligtvis av de första tre delvikterna i följande uppräkningsordning. Ibland används också den fjärde delvikten.

(i) *Designvikten*. Vid t.ex. obundet slumpmässigt urval (OSU) av 10 objekt ur en population bestående av 1000 objekt är designvikten ($1000/10=$) 100. Vid stratifierat OSU (STOSU) får varje objekt i ett givet stratum samma designvikt, men vikterna kan vara olika stora för skilda strata. Urvalsdesignen bestäms med målsättningen att få så bra skattningar som möjligt av vissa parametrar, ofta avseende viktiga redovisningsgrupper.

(ii) *Bortfallsvikten*. Vanligtvis har man skäl att misstänka att bortfallet är snedvridande och därför söks delvikter som reducerar den effekten. Detta görs ofta genom att dela upp de svarande i grupper och inom varje grupp bilda en korrigeringsvikt. Målsättningen med indelningen är att föra samman objekt som har samma sannolikhet att svara.

(iii) *Täckningsvikten*. I en statistisk undersökning uppträder också vanligtvis under- och övertäckningsproblem, som kan leda till skevhet i skattningarna. Undertäckningen är det allvarigare problemet av de två, men även övertäckningen kan vara besvärlig eftersom den i regel inte kan identifieras i bortfallet. För att i möjligaste mån undvika skevhet orsakad av täckningsproblem bildar man korrigeringar delvikter. (Det finns dock andra alternativ, som t.ex. att efter

estimationen lägga till en separat modellberoende uppskattning av bidraget från undertäckningen.)

(iv) *Postsamplingsvikten*. Med syfte att reducera både urvals- och bortfallsfelet används ibland exempelvis en poststratifierad estimator eller en kvotestimator. Även dessa estimatorer, samt många andra estimatorer, kan skrivas med hjälp av postsamplingsvikten. Denna delvikt ger också *konsistenta* skattningar i den meningen att skattningarna överensstämmer exakt med vissa registertotaler för populationen.

Vikter används inte bara i urvalsundersökningar utan också ibland i totalundersökningar; t.ex. kan bortfall eller täckningsproblem kräva att värdena viktas. I denna rapport visar vi också hur vikter används i prognosmodeller och vid länkning av tidsserier.

De flesta läroböcker i statistik förutsätter att sannolikhetsurval dras från populationen och att varken bortfall eller täckningsbrister förekommer. Därmed kan väntevärdesriktiga (unbiased) estimatorer konstrueras och relevanta konfidensintervall beräknas. I verkligheten förekommer dock både bortfall och täckningsproblem som vanligtvis kan misstänkas ge skeva (biased) estimatorer om inte effektiva korrigeringar görs. Tyvärr har vi aldrig tillräcklig information för att vara säkra på att inte skevhet kvarstår.

Samtliga ovan beskrivna delvikter bestäms vanligtvis med hjälp av registervariabler, exempelvis för stratifiering, i bortfallskompensationen för att bilda grupper, etc. Den information som utnyttjas kallas i fortsättningen för *hjälpinformation*. Nödvändig hjälpinformation består av två former, nämligen (i) värden för registervariabler (*hjälpvariabler*) för varje svarsobjekt samt (ii) populationstotaler (*hjälptotaler*) för dessa registervariabler.

Ett mindre krävande alternativ till att bilda var och en av dessa vikter är att utnyttja tekniken med kalibrering av vikter. Metoden är ytterst generell, vilket betyder att de varianter som beskrivits ovan, som produkten av ett antal delvikter, oftast kan erhållas i ett enda steg inom kalibreringstekniken. Den rymmer ett stort antal sätt att kompensera för bortfallet samt ett stort antal estimator typer.

Det bör dock nämnas att alla varianter av metoder inte utgör specialfall till kalibreringstekniken. Eftersom hanteringen av bortfall och täckningsbrister i estimatorn grundar sig på modeller (antaganden) kan naturligtvis vilka varianter som helst vara tänkbara. Har man en unik kunskap kring dessa brister som inte kan beskrivas inom ramen för kalibreringstekniken ska man naturligtvis använda annan teknik. Ett exempel på när inte kalibreringsestimatorn (se (3.8) och (3.12)) kan

användas, men väl regressionsestimatorn (se (3.6) och (3.10)), är följande: I en löpande undersökning baserad på samma urval bestäms närmevärden för svarssannolikheter utifrån gammalt material.

SCB har tillgång till omfattande hjälpinformation och goda programvaror för beräkning av skattningarna. För närvarande varierar utnyttjandet av dessa möjligheter kraftigt mellan olika undersökningar vid SCB. Vi vill med denna rapport visa på vad som pågår (pågått) inom området. Rapporten inspirerar förhoppningsvis till ytterligare utnyttjande av hjälpinformation med användning av kalibreringstekniken.

I denna rapport används begreppet ”kalibrering av vikter” på ett vidsträckt sätt. Exempelvis kommer vi i begreppet att också inkludera den generella regressionsestimatorn (Särndal, Swensson and Wretman, 1992). Både kalibreringsestimatorn (Deville and Särndal, 1992) och den generella regressionsestimatorn är utvecklade för att ge (approximativt) väntevärdesriktiga estimatorer i en urvalsundersökning. Här kommer vi också att använda begreppet kalibreringsteknik i mer modellberoende situationer, där dess förmåga till att ge konsistens i skattningarna används.

I avsnitt 2 beskriver vi kalibreringstekniken rent principiellt och i avsnitt 3 i tekniska termer. I avsnitt 4 beskriver vi programvaran CLAN97 med vars hjälp punkt- och variansskattningar kan beräknas. Därefter redogör vi i avsnitt 5 för olika fall där tekniken har prövats. Avsnitt 6 innehåller sammanfattning och kommentarer och avsnitt 7 pekar på de framtida möjligheterna. Vi har försökt skriva rapporten så att den i ”formler” mindre bevandrade personen kan hoppa över avsnitt 3 och ändå få en hygglig insikt i innehållet.

2. Huvudprinciperna i kalibreringstekniken

Låt oss börja med ett exempel, som beskriver principen för kalibrering av vikter.

Exempel 2.1. Antag att vi vill göra en urvalsundersökning om hälsa och skatta antalet personer som har varit sjuka under året i olika sjukdomar. Förutom skattningar för hela riket vill vi också ha skattningar för länen. Vi genomför datainsamlingen och trots stora ansträngningar, tvingas vi konstatera att ett visst bortfall föreligger.

Vi vet att ohälsa samvarierar med ålder och att män och kvinnor i viss mån drabbas av olika sjukdomar. Variablerna ålder och kön bör därför ingå som hjälpinformation eftersom de kommer att reducera både variansen och bortfallsskevheten.

Som vanligt får vi större bortfall i storstadsområden än i övriga riket. Vi ser också att såväl låginkomsttagare som höginkomsttagare svarar i mindre grad än medelinkomsttagarna. De variabler som bestämmer dessa uppdelningar bör vara med eftersom de kan förväntas reducera bortfallsskevheten..

Vi vet också från teorin att de viktigaste redovisningsgrupperna bör i möjligaste mån identifieras i hjälpinformation och därför bör länsuppdelningen tas med. Det leder till att variansen för skattningar för redovisningsgrupper minskar.

Vi har för hela populationen kunskap om kön, ålder, inkomst, storstadsområde/övriga riket, län. Därmed har vi möjlighet att på olika sätt samutnyttja registerinformationen för att både reducera urvalsfelet och bortfallsskevheten. Man kan dock räkna med att en korstabulering av alla dessa variabler skulle resultera i vissa mycket små grupper, som kanske helt saknar observationer. Kalibreringstekniken ger dock många möjligheter att utnyttja hjälpvariablerna. Det enklaste förfarandet är att utnyttja alla en-dimensionella marginaler till denna korstabell. Mellan dessa alternativ finns ett stort antal varianter. Man kan också ta med den ursprungliga (kontinuerliga) inkomstvariabeln som hjälpinformation. När man bestämt sig för en uppsättning beräknas kalibrerade vikter, punkt- och variansestimater med hjälp av CLAN97 (se avsnitt 4).

?

Kalibreringsvikterna är bestämda så att de ”avviker så lite som möjligt” från designvikterna och så att de återskapar också de kända populationstotalerna. Om t.ex. variabeln könstillhörighet i exempel 2.1 används som hjälpinformation kommer skattningen av antal män och kvinnor att helt överensstämma med antal män och kvinnor i populationen. Som vi tidigare nämnt kallar vi denna egenskap för konsistens.

Vi vill med exempel 2.1 visa att flexibiliteten vad gäller utnyttjande av hjälpinformation är stor. Vilken uppsättning av hjälpinformation som är ”optimal” är inte självklart. I arbetet att söka en bra uppsättning ingår ofta att beräkna kalibreringsvikter utifrån olika uppsättningar hjälpinformation. Även om det är tekniskt och beräkningsmässigt möjligt att utnyttja stora mängder hjälpinformation är det inte rekommendabelt att utan eftertanke låta mängden bli för stor. Det finns risk att vissa vikter blir mycket stora eller negativa, vilket kan ge orimligt stora variansskattningar för någon redovisningsgrupp. Därför bör följande principer för val av hjälpinformation noga följas. Hjälpvariablerna ska i möjligaste mån (Lundström, 1997, och Lundström and Särndal, 1999)

$$\left. \begin{array}{l} (i) \text{ förklara variationen i de viktigaste målvariablerna} \\ (ii) \text{ förklara variationen i svarssannolikheterna} \\ (iii) \text{ identifiera de viktigaste redovisningsgrupperna} \end{array} \right\} \quad (2.1)$$

Det är också viktigt att analysera vikterna och i det fall det finns avvikande vikter utvärdera huruvida det finns onödig (redundant) hjälpinformation, någon hjälpgrupp är för liten, e.d.

I fortsättningen säger vi att hjälpinformationen är ”stark” när (2.1) är väl uppfyllt.

3. Teknisk beskrivning av kalibrering av vikter

Betrakta den ändliga populationen bestående av N element, nämligen

$U = \{u_1, \dots, u_k, \dots, u_N\}$. Vi vill skatta en total $Y = \sum_U y_k$, där y_k är värdet på

målvariabeln y för k :te elementet. Ibland är vi intresserade av kvoten mellan två

totaler, säg $R = \sum_U y_k / \sum_U z_k$, där z är en annan målvariabel. När $z_k = 1$, för alla k ,

erhålles medelvärdet $\bar{Y} = \frac{1}{N} \sum_U y_k$. Vanligtvis är vi också intresserade av parametrar

för redovisningsgrupper. Låt oss definiera redovisningsgrupperna med

$U_1, \dots, U_d, \dots, U_D$, där $U = \bigcup_{d=1}^D U_d$. Totalen för redovisningsgrupp d kan skrivas

$Y_d = \sum_{U_d} y_k = \sum_U y_{dk}$, där

$$y_{dk} = \begin{cases} y_k & \text{för } k \in U_d \\ 0 & \text{f.ö.} \end{cases} \quad (3.1)$$

Låt s vara ett stickprov av storleken n draget från U med sannolikheten $p(s)$.

Inklusionssannolikheterna är då $\pi_k = \sum_{s \ni k} p(s)$ och $\pi_{kl} = \sum_{s \ni \{k,l\}} p(s)$. Låt $d_k = 1 / \pi_k$

beteckna *designvikten* för element k och låt $d_{kl} = 1 / \pi_{kl}$. Nu uppstår bortfall och

svarsmängden r av storleken m erhålles, där $r \subseteq s$. Låt oss tills vidare anta att vi

känner sannolikheten för varje objekt i i s att svara och att de betecknas:

$$\Pr(k \in r | s) = \theta_k \quad \text{och} \quad \Pr(k \& l \in r | s) = \theta_{kl} \quad (3.2)$$

Svarsmängden r kan alltså ses som resultatet av ett två-fas-urval och därmed kan följande väntevärdesriktiga estimator bildas

$$\hat{Y} = \sum_r d_k y_k / \theta_k \quad (3.3)$$

Eftersom svarssannolikheten θ_k aldrig är känd måste den ersättas med ett närmevärde $\hat{\theta}_k$ och följande estimator erhålles

$$\hat{Y}_\pi = \sum_r d_k y_k / \hat{\theta}_k \quad (3.4)$$

Exempel 3.1. Ett vanligt förfarande att bestämma $\hat{\theta}_k$ består i att dela upp urvalet s i s.k. svarshomogenitetsgrupper (SHG). Gruppindelningen görs så att man inom en given grupp har skäl att tro att svarssannolikheterna är ungefär lika stora. Låt oss definiera dessa grupper $s_1, \dots, s_p, \dots, s_p$ och motsvarande svarsmängder $r_1, \dots, r_p, \dots, r_p$. Storleken på dessa mängder är $n_1, \dots, n_p, \dots, n_p$ respektive $m_1, \dots, m_p, \dots, m_p$. Det närmevärde av θ_k som används i samband med SHGs är $\hat{\theta}_k = m_p / n_p$ för $k \in r_p$.

?

En motsvarande estimator för totalen för redovisningsgrupp d är

$$\hat{Y}_{d\pi} = \sum_r d_k y_{dk} / \hat{\theta}_k \quad (3.5)$$

Antag vidare att vi har en hjälpvektor \mathbf{x}_k för $k \in s$ och att vi känner motsvarande populationsstorheter $t_{\mathbf{x}}$. Vanligtvis känner man \mathbf{x}_k för alla objekt i populationen och kan därmed bilda $t_{\mathbf{x}} = \sum_U \mathbf{x}_k$, men det kan också vara så att $t_{\mathbf{x}}$ hämtas från en annan källa och därför inte är exakt en summa av \mathbf{x}_k -vektorerna. Inom vissa områden, som exempelvis ekonomisk statistik, förändras populationen relativt snabbt och därmed kan en hjälptotal baserad på urvalsramen avvika en del från en total avseende målpopulationen.

En generell regressionsestimator som utnyttjar hjälpinformationen kan skrivas:

$$\hat{Y}_{reg,U} = \sum_r d_k g_{Uk} y_k / \hat{\theta}_k \quad (3.6)$$

där

$$g_{Uk} = 1 + c_k (t_{\mathbf{x}} - \sum_r d_k \mathbf{x}_k / \hat{\theta}_k)' (\sum_r d_k c_k \mathbf{x}_k \mathbf{x}_k' / \hat{\theta}_k)^{-1} \mathbf{x}_k \quad (3.7)$$

Anm.: c_k är en positiv faktor som gör estimatorn mer flexibel. Vi återkommer till den senare.

?

I estimatorn (3.6) används hjälpinformation både i hjälpvektorn \mathbf{x}_k och i bestämningen av $\hat{\theta}_k$. Ett alternativt sätt är att i hjälpvektorn också ta med hjälpinformationen som förklarar variationen i svarssannolikheterna och med hjälp av kalibrering bilda estimatorn

$$\hat{Y}_{kal,U} = \sum_r d_k v_{Uk} y_k \quad (3.8)$$

där

$$v_{Uk} = 1 + c_k (t_{\mathbf{x}} - \sum_r d_k \mathbf{x}_k)' (\sum_r d_k c_k \mathbf{x}_k \mathbf{x}_k')^{-1} \mathbf{x}_k \quad (3.9)$$

I vissa mer ovanliga fall känner vi inte $t_{\mathbf{x}}$ utan endast \mathbf{x}_k för $k \in s$. Med den informationen kan vi bilda en väntevärdesriktig estimator av dessa populationstotaler, nämligen $\hat{t}_{\mathbf{x}} = \sum_s d_k \mathbf{x}_k$. Motsvarande regressionsestimatorer och kalibreringsestimator till (3.6) resp. (3.8) är

$$\hat{Y}_{reg,s} = \sum_r d_k g_{sk} y_k / \hat{\theta}_k \quad (3.10)$$

där

$$g_{sk} = 1 + c_k (\hat{t}_{\mathbf{x}} - \sum_r d_k \mathbf{x}_k / \hat{\theta}_k)' (\sum_r d_k c_k \mathbf{x}_k \mathbf{x}_k' / \hat{\theta}_k)^{-1} \mathbf{x}_k \quad (3.11)$$

respektive

$$\hat{Y}_{kal,s} = \sum_r d_k v_{sk} y_k \quad (3.12)$$

där

$$v_{sk} = 1 + c_k (\hat{t}_{\mathbf{x}} - \sum_r d_k \mathbf{x}_k)' (\sum_r d_k c_k \mathbf{x}_k \mathbf{x}_k')^{-1} \mathbf{x}_k \quad (3.13)$$

Motsvarande skattningar av Y_d erhålls genom att ersätta y_k med y_{dk} i (3.6), (3.8), (3.10) och (3.12).

I det fall hjälpinformationen är stark har både (3.6) och (3.8) flera goda egenskaper. Jämfört med estimatoren (3.4) gäller att $\hat{Y}_{reg,U}$ och $\hat{Y}_{kal,U}$

(i) har mindre urvalsfel

(ii) har mindre bortfallsbias

(iii) är konsistenta, d.v.s. $\sum_r d_k g_{Uk} \mathbf{x}_k / \hat{\theta}_k = \sum_r d_k v_{Uk} \mathbf{x}_k = t_{\mathbf{x}}$.

Vad beträffar estimatorerna (3.10) och (3.12) så leder stark hjälpinformation till att (ii) uppfylls i lika grad som för estimatorerna (3.6) och (3.8). Urvalsfelet minskar också något och de är konsistenta i så motto att $\sum_r d_k g_{sk} \mathbf{x}_k / \hat{\theta}_k = \sum_r d_k v_{sk} \mathbf{x}_k = \sum_s d_k \mathbf{x}_k$.

Förutom egenskaperna (i)-(iii) torde också utnyttjande av hjälpinformation i estimatorerna $\hat{Y}_{reg,U}$ och $\hat{Y}_{kal,U}$ reducera täckningsfelen i vissa fall. Vid löpande uppdatering av urvalsramen kan \mathbf{x}_k -totalen ($t_{\mathbf{x}}$) vara närmare populationstotalen $\sum_U \mathbf{x}_k$ än \mathbf{x}_k -totalen i urvalsramen, tack vare att estimationen görs sent i produktionsprocessen. Även estimatorerna $\hat{Y}_{reg,s}$ och $\hat{Y}_{kal,s}$ kan uppvisa samma effekt. Fortfarande finns dock inte tillräckligt teoretiskt och empiriskt stöd för att säkert säga att även täckningsfelen kan reduceras med hjälp av kalibreringstekniken.

4. CLAN97

När SHGs (se exempel 3.1) används vid bestämning av närmevärden av svarssannolikheter kan såväl punkt- som variansskattningar för estimatoren av typ (3.4) och för kalibreringsestimatoren av typ (3.6) beräknas med hjälp av CLAN97 (Andersson and Nordberg, 1998). I CLAN97 finns två varianter på SHGs (se exempel 3.1), nämligen (i) där p -grupperna är identiska med strata och (ii) där p -grupperna utgör subgrupper inom strata.

CLAN97 klarar i själva verket alla parametrar som är (rationella) funktioner av totaler, där totalerna skattas med någon av ovan nämnda alternativ.

Variansskattningarna baserar sig på tekniken med Taylor-linearisering.

CLAN97 är också mycket flexibel vad gäller utnyttjande av hjälpinformation. Såväl kontinuerliga som kategoriska variabler kan ingå i hjälpinformationen. Programmet ger också stöd för dikotomiseringen av kategoriska variabler, som är nödvändig för bildandet av hjälpvektorn.

Med hjälp av CLAN97 kan också $\hat{Y}_{kal,U}$, se (3.8), beräknas. Inom kort kommer CLAN97 också att inkludera en procedur för variansberäkning av denna estimator. Likaså kan $\hat{Y}_{reg,s}$ och $\hat{Y}_{kal,s}$ beräknas, men däremot inte variansestimaten för dessa estimatorer.

Skattning av parametrar för redovisningsgrupper samt beräkningen av motsvarande variansestimata görs i CLAN97 genom utnyttjande av variabel (3.1).

Ibland kan vikterna g_{Uk} , g_{sk} , v_{Uk} och v_{sk} bli mycket stora och i vissa fall också negativa och då ge upphov till en instabil estimator. För att undvika detta kan man i CLAN97 sätta en övre och en undre gräns för dessa vikter. Trots denna begränsning kommer konsistensvillkoret i (iii), avsnitt 3, att vara uppfyllt.

5. Produkter där kalibrering av vikter har prövats

I detta avsnitt beskrivs kortfattat olika fall där kalibreringstekniken används eller har prövats. I vissa fall består arbetet endast av en metodstudie och i andra fall är metoden helt implementerad. Mellan dessa ytterlighetsfall finns flera varianter.

Vi delar upp fallen efter avdelningstillhörighet. För varje fall anges vem/vilka som har genomfört kalibreringen och även gjort beskrivningen.

AM-avdelningen

Estimationsförfarandet i Arbetskraftsundersökningarna (AKU)

Jan Hörngren, AM/LEDN och Hassan Mirza, AM/AKU

Från och med 1993 används hjälpinformation från SCB:s sysselsättningsregister (SREG) och AMS register (PRESO) i estimationsförfarandet. Från och med fjärde kvartalet 1999 modifierades estimationssystemet ytterligare. Nuvarande estimationssystem består av två GREG-estimatorer:

- GREG1 för skattningar av antalet arbetslösa. I estimationen används hjälpinformation från PRESO (arbetssökande utan arbete som kan ta arbete direkt) och RTB (kön, ålder och region)
- GREG2 för skattningar av sysselsatta och undergrupper av sysselsatta. I estimationen används hjälpinformation från SREG (sysselsatt i olika näringsgrensgrupper/ej sysselsatt) och RTB (kön, ålder och region).

Vissa typer av parametrar skattas som en funktion av GREG1 och GREG2. Målet är att från och med år 2001 skapa en enhetlig GREG-estimator för alla önskvärda skattningar.

Införandet av den nytillkomna hjälpinformationen medförde betydelsefulla precisionsvinster för skattningar av antalet arbetslösa och för skattningar av antalet sysselsatta uppdelade efter näringsgrenar. Vidare är det påvisat att den nya hjälpinformationen reducerar bortfallsbias för skattningar av antalet arbetslösa och antalet sysselsatta jämfört med tidigare estimationsförfarande som baserades på enbart RTB-uppgifter. Modifieringen av estimationen 1999 innebar dessutom bättre precision för regionala skattningar. Den nya estimationen (1999) innebär också en övergång till ett PC-baserat system där CLAN spelar huvudrollen.

För ytterligare information rörande kalibrering av vikter i AKU hänvisas till Hörngren och Mirza (1998) och Mirza (1999).

Att länka ett system av tidsserier -

Omräkning av Arbetskraftsundersökningarna 1987- 92

Britt och Anders Wallgren, METOD

Bakgrund och syfte

Surveyundersökningar måste ibland revideras. Mellan december 1992 och januari 1993 har AKU genomgått stora förändringar med avseende på *definitioner*, *skattningsmetoder* och *mätperioder*. Avsikten var att öka undersökningarnas kvalitet. Som resultat av förändringarna uppstod brott i tidsserier, både som nivåförändringar men också som ändrade säsongmönster. Alla dessa tidsseriebrott innebär en mycket stor komplikation för användarna. Vetskapen om sådana komplikationer bromsar införandet av nödvändiga metodförändringar.

I AKU-systemet rapporteras ca 30 000 tidsserier varje månad. Alla dessa serier har räknats om med bibehållen inbördes konsistens och för detta ändamål har vi utvecklat en ny metodik. Syftet med omräkningarna är dels att öka jämförbarhet och tillförlitlighet hos gamla skattningar och dels att underlätta säsongrensning av de nya serierna. För att kunna säsongrensa en tidsserie behövs ca 5 år med månadsdata - utan omräkning av gamla värden skulle man behöva vänta till 1998 med att säsongrensa de nya serierna.

Metodöversikt

Den nya metoden för länkning av tidsseriesystem består av två delar:

1. På *tidsserienivå* länkas 106 nyckelserier och länkningarna görs så att nyckelserierna blir konsistenta. För var och en av dessa serier får vi då 72 omräknade månadsvärden för perioden 1987-1992 som är jämförbara med motsvarande nya serier från och med 1993.
2. På *individnivå* görs omräkning av varje intervjuad persons uppräkningsvikt med de omräknade nyckelserierna som styrvillkor. Detta innebär att för var och en av de 72 undersökningarna från 1987-1992 kalibreras uppräkningsvikterna.

Därigenom blir alla 30 000 serier länkade med bibehållen inbördes konsistens. Att de länkade serierna är konsistenta innebär att för varje månad är exempelvis $\text{antalet sysselsatta} + \text{antalet arbetslösa} + \text{antalet ej i arbetskraften} = \text{antalet i befolkningen}$ och detta gäller för alla uppdelningar efter ålder och kön.

Genom att kalibrera på individnivå kommer alla tidsserier som tas fram med de nya vikterna att bli konsistenta. De nyckelserier som vi använt kommer att förbli som de beräknats under punkt 1. Alla övriga tidsserier som kan bildas kommer att länkas i överensstämmelse med nyckelserierna. De naturliga sambanden mellan variablerna på individnivå styr dessa länkningar.

Steg 1: Länkning av nyckelserier

Länkingsarbete har gjorts i flera steg. Dels har vi ändrat för nya definitioner, och dels har vi korrigerat för bortfallsfel och periodiceringsfel och slutligen har vi gjort de korrigeringar som beror på att hela månaden numera undersöks. Som resultat har vi nu fått korrigerade nyckelserier som är jämförbara över tiden.

Steg 2: Kalibrering av uppräkningsvikter

Kalibreringsteknikens förmåga att ge konsistenta skattningar används för att länka resterande tidsserier. De kända hjälptotalerna (t_x ; se avsnitt 3) motsvaras här av de färdiglänkade värdena på nyckelserierna för en viss månad. T.ex. har länkningen i steg 1 resulterat i att antal fast anställda kvinnor 16-19 år skall vara 38 957 för en viss månad. Ytterligare 105 hjälptotaler utnyttjas i kalibreringen av vikter för denna månads svarsfil. (För ytterligare detaljer, se Wallgren och Wallgren, 1998)

Kortperiodisk sysselsättningsstatistik för privat sektor (KSP)

Sixten Lundström, METOD

KSP är en kvartalsundersökning inom den privata sektorn. Huvuduppgiften består i att skatta antalet anställda inom olika branscher. Undersökningen presenterar också resultat för redovisningsgrupper som bildas av bransch*kön, bransch*anställningsförhållande*kön och arbetsställets storleksklass*kön. Dessutom efterfrågas statistik för län och även län*bransch. Den utnyttjade estimatorn av formen (3.4), där SHGs är lika med strata, ger dock länsestimater med alltför stora urvalsfel och därför framställde man inte sådan statistik.

Urvalet av arbetsställen dras med STOSU och är av storleken 11500. Strata bestäms av bransch*arbetsställets storleksklass. Ramuppgiften om antal anställda ("CFAR-anställda") används som allokeringssvariabel vid försöken att bestämma en optimal allokering vid skattning av antal anställda inom varje bransch. Korrelationskoefficienten för svarsobjekten (ingen viktning görs) mellan antal anställda och CFAR-anställda är så hög som 0.96, vilket gör att förutsättningarna är gynnsamma för att erhålla små urvalsfel även med estimatorn (3.4). Svarsandelen är ca 92%.

I syfte att reducera urvalsfelen och då speciellt att uppnå kvalitetskraven på läns-skattningarna studerades effekten av kalibrering av vikter. Även en reduktion av bortfallsbiasen förväntades. Den estimator vi använde var (3.6) med SHGs identiska med strata. Troligen skulle också bortfallsbiasen reduceras, men detta gjordes inget försök att studera.

Den hjälpinformation som fanns tillgänglig var CFAR-anställda, länsstillhörighet och bransch. En mängd olika uppsättningar av hjälpinformation studerades. Beräkningen av kalibrerade vikter gjordes med det SAS-baserade programmet CALMAR och beräkningen av medelfel och andra mått utfördes med egenkonstruerade APL-program. (CLAN innehöll vid den tiden inte kalibreringsestimatorn.)

Det visade sig att de minsta medelfelen erhöles för de "specialsydda" estimatorerna, d.v.s. branschindelningen av CFAR-sysselsatta användes i hjälpvektorn när antal anställda för olika branscher skulle skattas, när länsindelningen användes vid skattning av antal anställda för olika län, etc. Det kräver dock olika viktsystem för olika variabler vilket är opraktiskt. Det visade sig dock att alternativet med att utnyttja de tre (en-dimensionella) marginalerna över antal CFAR-anställda i korstabellen storleksklass*bransch*län var en god

kompromiss. Den genomsnittliga ökningen av medelfelen när man lämnar de specialsydda estimatorerna var 3.2% för länsredovisningen och 4.9% för branschredovisningen. Trots detta gav ”kompromisslösningen” medelfel som var mindre än för den ordinarie estimatorn, nämligen i genomsnitt 54% mindre för länsredovisningen och 7% för branschredovisningen.

Vi provade också att ta med antal arbetsställen i olika kategorier i uppsättningen av hjälptotaler (intercept i regressionsmodellen) men det gav ingen förbättring. Vi jämförde också skattningar där c_k var lika med en konstant och där c_k var omvänt proportionell till CFAR-anställda på arbetsstället k , men vi fann inga större skillnader. Vi fick också några få negativa vikter. Eftersom CALMAR tillåter att man sätter gränser för hur små (och stora) vikterna får vara kunde vi jämföra viktsystemet som innehöll negativa vikter med det som inte innehöll sådana. Skillnaden i estimaten var ytterst liten.

Undersökning bland deltagare i studiecirklar för arbetslösa år 94/95

Sixten Lundström, METOD

Syftet med undersökningen var att få information om vad deltagarna tyckte om genomgången studiecirkel. I huvudsak bestod parametrarna av totaler.

Det fanns ingen gemensam förteckning över deltagare i studiecirklar för arbetslösa, men däremot fanns det inom varje kommun och studieförbund en förteckning över sådana studiecirklar. För varje cirkel fanns en deltagarförteckning, men av sekretesskäl kunde inte SCB ta del av dessa. Därför fick vi ta hjälp av personer ansvariga för dessa studiecirklar. För att minska på administrationen ville vi undvika en inblandning av alltför många personer och därför drogs ett urval av kommuner (steg 1) inom vart och ett av de 11 studieförbunden. Inom varje utvald kommun drogs ett urval av cirklar (steg 2) och inom varje utvald cirkel drogs en deltagare. Populationen av kommuner (för varje förbund) stratifierades med hjälp av uppgiften om antal studiecirklar föregående år. För bestämning av stratumgränser användes cum- \sqrt{f} metoden. I vissa strata gjordes en totalundersökning och i andra allokerades urvalet med hjälp av Neyman-regeln. Urvalet av cirklar och deltagare gjordes av de studiecirkelansvariga personerna utifrån slumpstalstabeller framtagna av SCB.

Bortfallet av deltagare och därmed cirklar var stort - nästan 50% i vissa studieförbund! För att i någon mån reducera bortfallsbiasen utnyttjade vi hjälpinformation till att kalibrera vikter. För hela urvalet hade vi uppgift om kön,

cirkeltyp och huvudämnesgrupp. Vi kunde inte utnyttja den korstabulerade informationen eftersom vissa celler blev för små. Därför utnyttjade vi enbart marginalinformationen för att bilda \mathbf{x}_k . Denna hjälpvektor definierades inte på samma sätt för varje förbund beroende på att ibland var vissa kategorier tomma eller innehöll ett ytterst litet antal. I sådana fall har vissa kategorier slagits samman. Hjälpinformationen förklarade relativt bra svarsbeteendet.

Den punkttestimator vi använde var (3.12), där designvikten d_k bestämdes av produkten av inklusionssannolikheterna i de tre urvalsstegen. Vi lät genomgående $c_k = 1$.

Alternativet till estimator (3.12) hade varit (3.10), men eftersom i några fall ingen av de utvalda studiecirkelarna inom vissa strata (och förbund) svarade hade vi varit tvungna att bortfallskorrigera även första-steps-vikten. Vi tror att det skulle ha gett sämre resultat än utnyttjade metod.

BV-avdelningen

FASIT-modellen

Klas Lindström, BV/IF

FASIT (Fördelnings - Analytiskt Statistiskt System för Inkomster och Transfereringar) är en så kallad mikrosimuleringsmodell. Med detta menas att befintliga eller simulerade skatter, bidrag och ersättningar beräknas för varje individ och varje hushåll som ingår i databasen varefter beräkningarna summeras upp till riksnivå.

Modellen används för att kunna beräkna effekterna av förändringar i skatte- och transfereringssystemen. Dessa beräkningar önskar man göra för nästa budgetår. De data som finns tillgängligt för beräkningarna är HINK-data som är tre eller fyra år gamla.

I modellen vill man därför skriva fram HINK-data från år t till år $t+1$, $t+2$, $t+3$ och $t+4$. Framskrivningen görs genom att ekonomiska variabler skrivs fram med koefficienter. För att behandla befolkningsförändringar används däremot prognoser för befolkningsutvecklingen tillsammans med prognoser över förändringar på arbetsmarknaden (antal sysselsatta och antal arbetslösa) dessa prognoser används som hjälptotaler i modellen. Med hjälp av dessa hjälptotaler kalibreras designvikterna för HINK-datat. Som resultat skapas nya viktfiler som avser år t , $t+1$, $t+2$, $t+3$ och $t+4$.

I modellen används befolkningens fördelning på ålder och kön som hjälptotaler, dessutom anges antalet sysselsatta i statlig och kommunal sektor, för privat sektor anges antalet sysselsatta inom industrin, byggnadsverksamhet och privat tjänstsektor. För personer över 64 år används även uppgift om man antalet med ATP och utan ATP som hjälptotaler.

Med hjälp av dessa nya vikter och den ekonomiska framskrivningen med koefficienter får man en framskrivning av databasen för år t+1, t+2, t+3 och t+4 som sedan används för att beräkna effekter av förändringar i skatte- och transfererings-systemen.

Anpassning av estimat till kända storheter i ULF

Peter Lundqvist, METOD

En intressant fråga som ofta vill studeras i panelundersökningar är ifall det har skett en förändring av befolkningens beteende mellan två mättillfällen. I ULFs fall vill man studera olika faktorer som påverkar hälsan t.ex. kan man titta på andelen rökare mellan två mättillfällen.

Nedan ges ett exempel på en typ av kalibrering som utnyttjar två undersökningar med överlappande urval. Ca 40 procent av individerna har intervjuats vid båda tillfällena. I ULF 80/81 skattades andelen rökare/icke-rökare med $N_{1.}^*$ alternativt $N_{2.}^*$ och i ULF 88/89 skattades motsvarande andelar med $N_{.1}^*$ och $N_{.2}^*$. Dessa skattningar är baserade på ca 12000 individer. I en korstabell kommer då cellernas skattningar, \hat{N}_{11} , \hat{N}_{12} , \hat{N}_{21} , \hat{N}_{22} , att vara skattade på färre individer (ca 40 procent av 12000) än för de separata undersökningarna. I tabellen nedan beskrivs situationen.

	ULF 88/89		
ULF 80/81	Rökare	Icke-rökare	Summa
Rökare	\hat{N}_{11}	\hat{N}_{12}	$N_{1.}^*$
Icke-rökare	\hat{N}_{21}	\hat{N}_{22}	$N_{2.}^*$
Summa	$N_{.1}^*$	$N_{.2}^*$	

Denna situation har tidigare studerats med hjälp av Raking-Ratio-skattningar (jmf Lundström, 1993). Genom att använda $N_{1.}^*$, $N_{2.}^*$, $N_{.1}^*$, $N_{.2}^*$ som hjälpinformation har vi den typ av estimator som beskrivs i (3.12). Variansestimaten kan tyvärr inte räknas på ett korrekt sätt. En jämförelse av variansestimaten, ett simulerat (1000 replikat) baserat på ett Raking-Ratio estimat och ett baserat på ett kalibrerat estimat, visade på små variationer i det här fallet.

Kalibrering av vikter i Hushållens utgifter 1992 inför Undersökningen av hushållens utgifter 1995

Peter Lundqvist, METOD

Från och med 1995 planerades en årlig Undersökningen av hushållens utgifter (UTB). Undersökningar med liknande innehåll och uppläggning (tidigare kallad HUT) har genomförts 1985, 1988 och 1992. Man beslöt att behålla samma urvalsdesign i UTB men med ett reducerat urval, ca 2000 hushåll d.v.s. ungefär en tredjedel av de tidigare urvalen i HUT. Inför UTB genomfördes en studie med syftet att se om en ny modellberoende estimator kunde ersätta den tidigare använda estimatorn, som var baserad på efterstratifieringen.

Den tidigare använda efterstratifieringen i HUT bildades m.h.a. logistisk regression från Registret över totalbefolkningen (RTB-S) samt en justering över bokföringsomgångarna för att kompensera för variationen av antalet deltagande hushåll över tiden och få representativa årsskattningar. Syftet med efterstratifieringen var först och främst att reducera bortfallsbiasen.

HUT 92 är baserat på ett systematiskt urval från RTB-S av individer, i åldern 0-74 år. Urvalet kan betraktas som ett approximativt OSU. Vid identifieringen av hushållets sammansättning kan sedan hushållets urvalssannolikhet beräknas (betecknas ibland som nätverksurval).

Hjälpinformationen som studerades var definierad både på individ och hushållsnivå, vilket måste tas hänsyn till i estimationen (jmf t.ex. Andersson, C. (1997)). Ett önskemål var att gå igenom den tillgängliga NR-informationen och se om det fanns någon möjlighet att utnyttja denna. Det visade sig att mindre än 10 procent av den totala konsumtionen kunde beskrivas med hjälp av de register som NR använde i sin estimation. En huvudgrupp som skulle kunna förklaras bättre med hjälp av informationen från NR var utgifter för 'Sprit och tobak'. Denna huvudgrupp utgjorde mindre än 5 procent av den totala konsumtionen. En annan variabel som identifierades var antalet småhusenheter enligt fastighetstaxeringsregistret.

I den preliminära rapporten av UTB 95 användes estimatorn (3.8), $c_k = 1$ för alla objekt, som baserades på hjälpinformation från NR för sprit och tobak, fastighetstaxeringsregistret, RTB och variationen av antalet hushåll över tiden. En del kalibreringsvikter blev ganska stora och negativa vikter uppstod. Eftersom man inte ville ha negativa vikter användes den trunkerade linjära metoden i CALMAR (studien genomfördes innan CLAN97 innehöll kalibreringsestimatorn). I den slutliga rapporten av UTB 95 bedömde man att den valda kalibreringsestimatorn gav en del

svårtolkade resultat och valde därför att återgå till en traditionell efterstratifiering baserad på RTB-S samt en justering av antalet hushåll över tiden. Man har dock behov av att testa vad inkomster från IoF kan ge samt att utnyttja en större undersökning för att skatta antalet hushåll (eventuellt HEK).

Kom SOM-undersökningen

Sixten Lundström, METOD

SOM-institutionen vid Göteborgs universitet genomför varje år tre undersökningar om allmänhetens inställning i olika samhällsfrågor. Undersökningarna skiljer sig åt bl.a. beträffande avgränsningen av geografiskt område. SCB genomförde den del som kallas KomSOM-undersökningen och avsåg befolkningen i tre väst-svenska kommuner. Inom varje kommun drogs ett OSU av ca 1200 personer. Vid bortfallsanalysen fann man att antalet män och kvinnor skilde sig åt kraftigt i den ena kommunen, detta beroende på ”otur” i urvalsdragningen, vilket sedan ytterligare förstärktes av bortfallet. Vi genomförde en kalibrering efter kön och en annan kalibrering efter kön*åldersklass och överlämnade vikterna till uppdragsgivaren. Vilket viktsystem uppdragsgivaren använder vet vi inte.

Undersökning om nya hälsobesvär

Sixten Lundström, METOD

Sociologiska institutionen vid Umeå universitet var ansvarig för undersökningen. Undersökningen baserade sig på ett OSU av ca 3100 personer, där 69.5% svarade. För att reducera bortfallsbiasen och kanske urvalsfelet önskade uppdragsgivaren att kalibrerade vikter framställdes.

I undersökningen hade vi tre olika register att hämta hjälpinformation ifrån, nämligen Registret över totalbefolkningen (RTB), Utbildningsregistret (UTB) och SCB:s sysselsättningsregister (ÅRSYS). Från RTB hämtade vi variablerna kön, ålder, civilstånd, nationalitet, inkomst samt geografisk tillhörighet och från UTB hämtade vi två variabler, nämligen utbildning - nivå samt utbildning - inriktning. Variablerna sysselsättningsstatus, sektorstillhörighet och näringsgren (för förvärvsarbetande) hämtades från ÅRSYS.

I analysen av effekten på bortfallsbiasen studerade vi sambandet mellan den dikotoma variabeln svarande/bortfall och hjälpvariablerna. Andelen svarande i olika grupper bestämda av respektive hjälpvariabel beräknades och vid stora skillnader mellan svarsandelarna (och när antalet observationer var tillräckligt stort) bestämdes att variabeln skulle utnyttjas i kalibreringen. I vissa fall kunde man se att antalet kategorier i en variabel var för stort och därför gjordes sammanslagningar.

I analysen av sambandet mellan målvariablerna och olika hjälpvariabler studerades fem viktiga dikotoma målvariabler, nämligen (i) ofta huvudvärk, (ii) ofta klåda, sveda, irritation i ögonen, (iii) ofta klåda, stickningar, stramningar, hettande känsla i ansikte, (iv) ofta munsveda, (v) ofta besvär i nacke, skuldror/axlar. Vi studerade frekvensen av dessa åkommor i olika hjälpgrupper.

I tabell 1 anges de hjälpvariabler som ovanstående två analyser visat bör vara med i kalibreringen. I kommentarerna anges ett U i det fall en reduktion av urvalsfelet kan förväntas och B i det fall bortfallsbiasen kan förväntas reduceras.

Tabell 1. Utnyttjad hjälpinformation

Hjälpvariabler	Kommentarer
Kön	U och B
Civilstånd	B [omkodning (1, övriga)]
Nationalitet	B [omkodning (svenskar, övriga)]
Utbildning -nivå	B [omkodning (4-7, övriga)]
Sysstatus	B [omkodning (1+5, övriga)]
Åldersklasser	U och B
Inkomstklasser	U och B
Storstadsområde	B

Utbildning - inriktning, sektorstillhörighet och näringsgren (en-siffer-nivån) visade sig inte innehålla någon värdefull information och togs därför inte med. Vi studerade även variablerna H-region och riksområden och fann att den geografiska informationen speglas bättre av storstadsområde.

I kalibreringen använde vi estimator (3.8) med $c_k = 1$ för alla objekt. Ingen kalibreringsvikt blev särskilt stor, men däremot uppstod några negativa vikter. Eftersom det kan vara störande utnyttjade vi en teknik att erhålla enbart positiva vikter (Singh and Mohl, 1996, method 5).

SPEL-undersökningen

Sixten Lundström, METOD

Undersökningen syftade till att öka kunskapen om spelbeteendet i landet och framför allt var man intresserad av antalet ”spelberoende” personer. Totaler och medelvärden var de parametertyper som förekom i undersökningen. Urvalet bestod av tre delurval, där det första var ett proportionellt stratifierat urval från hela populationen, det andra bestod av ett OSU av unga personer och det tredje av ett OSU av födda i utlandet. Stratifieringen i första delurvalet gjordes med variablerna kön, åldersklasser, och utbildning. Totalt ingick i urvalet ca 9900 personer, varav ca 72 % svarade.

Förutom stratifieringsinformationen fanns det en mängd registerinformation att hämta från RTB och det s.k. Louise-registret. Vid analys av samband mellan svarssannolikheter och hjälpvariabler samt samband mellan viktiga målvariabler och hjälpvariabler fann vi att följande hjälpvariabler var viktiga: kön, åldersklasser, utbildning samt de dikotoma variablerna födda utomlands, boende i storstad, mottar socialbidrag och studerande. Analysen visade att denna hjälpinformation är stark. I syfte att reducera urvalsfelet och bortfallsbiasen genomförde vi en kalibrering av vikter. Hjälptotalerna från Louise var något äldre än de från RTB och dessutom saknades den yngsta åldersgruppen och därför fick vi göra vissa (modellberoende) justeringar av de förra för att erhålla hjälptotalerna t_x .

De två sista delurvalen var inte stratifierade (som det första är) och därför var det komplicerat att beräkna inklusionssannolikheterna för t.ex. unga personer födda utomlands. För att hitta en approximation av dessa delade vi upp de två sista delurvalen (svarande) i efterstrata enligt samma principer som stratifieringen i delurval 1 och antog att alla svarande i ett givet stratum var draget med OSU. Därmed hade vi närmevärden till $d_k / \hat{\theta}_k$ i estimator (3.4).

Det var inte möjligt att använda den fullständiga cellinformationen i korstabellen med alla hjälpvariablerna, eftersom antalet observationer blev ytterst litet i vissa celler. Därför utnyttjade vi enbart de en-dimensionella marginalerna till den korstabellen. De kalibrerade vikterna g_{uk} (se (3.7)) blev alla positiva och ingen var större än 1.9.

Försök att utveckla estimator för skattning av hushålls- och lägenhetsparametrar för kommuner med hjälp av modellberoende estimatorer

Sixten Lundström, METOD

Den enda undersökning som ger statistik över hushåll och lägenheter för små redovisningsgrupper, som exempelvis kommuner, är Folk- och Bostadsräkningen (FoB). Tidigare har dessa totalundersökningar genomförts vart femte år, men nu har denna regelbundenhet brutits och den senaste genomfördes år 1990. I syfte att erhålla viss statistik år 1995 planerades år 1996 ett test av modellberoende estimatorer. En av dessa estimatorer hade formen (3.8), där $d_k = 1$ för alla k , och mängden r motsvarar en särskild mängd i FoB90. I detta fall hämtas också hjälptotalerna t_x från en annan tidpunkt än hjälpvektorn använd på r -mängden och därför kan vi förvänta ganska stor skillnad mellan t_x och $\sum_U \mathbf{x}_k$.

En stor del av befolkningen och hushållen kan "skrivas fram" till undersökningstidpunkten med hjälp av bl.a. flyttningsaviseringar. För resterande del skulle kalibreringstekniken användas. Mängden r består av den del av FoB90 som inte skulle kunna skrivas fram från FoB85. I den delen finns alltså målvariablerna, som skulle räknas upp med hjälp av kalibrerade vikter till att motsvara situationen år 1995. Dessa vikter v_{Uk} skulle alltså uppfylla konsistensvillkoret $\sum_r v_{Uk} \mathbf{x}_k = t_x$, d.v.s. att skattningarna med hjälpvektorn insatt i estimatorm överensstämmer exakt med de aktuella hjälptotalerna.

Hjälptotalerna skulle hämtas huvudsakligen från RTB år 1995 och avse totaler för "region", kön, åldersklasser, civilstånd, antal barn, etc. Försöket fick dock avbrytas eftersom det visade sig att ingen av SCBs PCar/Servrar klarade av de krävande beräkningarna. (Även superdatoren vid Linköpings Universitet skulle få problem, enligt uppgift från säker källa.)

Undersökningen "Miljö Hälsa Allergi"

Sixten Lundström, METOD

Institutet för miljömedicin (IMM) vid Karolinska institutet bedriver forskning och utbildning inom det miljömedicinska området. I syfte att öka kunskapen om miljöfaktorer som luftföroreningar, metaller och klorerade organiska föreningar genomfördes denna undersökning. Till ett obundet slumpmässigt urval av ca 15700 personer skickades en enkät. Av dessa svarade ca 11250 personer, d.v.s. ca 71% av det totala urvalet.

I undersökningen hade vi två olika register att hämta hjälpinformation ifrån, nämligen Registret över totalbefolkningen (RTB) och Utbildningsregistret (UTB). De variabler som vi i en första gallring fann intressanta var RTB-variablerna kön, ålder, civilstånd, födelseland, inkomst, storstadsområde och län samt UTB-variabeln utbildningsnivå.

Vid val av hjälpvariabler beaktade vi tre principer (se (2.1)), nämligen (i) att variabeln samvarierar med svarssannolikheterna, eller/och (ii) att den samvarierar med (viktiga) målvariabler, eller/och (iii) att den avgränsar (viktiga) redovisningsgrupper.

För att uppfylla princip (i) studerade vi sambandet mellan den dikotoma variabeln svarande/bortfall och hjälpvariablerna. Andelen svarande i olika grupper bestämda av respektive hjälpvariabel beräknades och vid stora skillnader mellan svarsandelarna (givet att antalet observationer var tillräckligt stort) utnyttjades variabeln i kalibreringen.

För att uppfylla princip (ii) analyserades sambandet mellan några målvariabler och olika hjälpvariabler. De målvariabler som studerades var fyra dikotoma målvariabler, nämligen (i) svårt att somna p.g.a. trafikbuller, (ii) astma, (iii) känslig för pollen samt (iv) känslig för katt. Variablerna (i) och (iii)-(iv) pekade ut en egenskap så snart värdena 1-3 var angivet på motsvarande frågor. Vi studerade frekvensen av dessa åkommor i olika hjälpgrupper.

Län bestämmer viktiga redovisningsgrupper och därför prövades den variabeln (princip (iii)).

I nedanstående tabell anges de hjälpvariabler som visade sig bör vara med i kalibreringen. Vi utgick från fler kategorier i de olika variablerna, men efter analys gjordes vissa sammanslagningar. I kommentarerna anges ett U i det fall en reduktion av urvalsfelet kan förväntas och B i det fall bortfallsbiasen kan förväntas reduceras.

Tabell Utnyttjade hjälpvariabler

Hjälpvariabler	Kommentarer
Kön *Åldersklasser	U och B [åldersklasser: 18-25, 26-45, 46-65, 66-75, 76-80]
Civilstånd	B [gifta, övriga]
Utbildning -nivå	B [förgymnasial+gymnasial högst 2 år; eftergymnasial 3 år och mer; övriga;]
Inkomstklasser	U och B [0-140, 141- tkr]
Län	U

Storstadsområde visade sig vara en svag hjälpvektor. Födelseland (Sverige; övriga Norden; övriga världen) förklarade både variationen i svarsbenägenhet och variationen i vissa av studerade målvariabler, men antalet observationer i ”övriga världen” var så litet att vikterna förväntades bli instabila (stora), vilket i sin tur kunde ge orealistiska skattningar för någon (liten) redovisningsgrupp. Därför användes inte födelseland.

ES-avdelningen

Användningen av administrativt datamaterial som hjälpinformation vid estimation - en studie av användningen av SRU-materialet som hjälpinformation till finansstatistiken

Stina Andersson, AM/AKU (f.d. ES)

Sammandrag av studien

Ett Standardiserat Räkenskaps-Utdrag (SRU) utgörs av ett antal företagsekonomiska uppgifter - balansräkning, resultaträkning etc. - som samtliga företag måste lämna till RSV i samband med skattedeklaration.

SRU-materialet har varit tillgängligt för SCB sedan 1993. Finansstatistiken (och sedermera dess efterföljare Företagsstatistiken) vid ES-avdelningen samlar in uppgifter som till stora (men inte till alla) delar finns i SRU-materialet. Det har därför legat nära till hands att undersöka om man eventuellt kunde ersätta hela eller delar av den direkta datainsamlingen med uppgifter direkt ur SRU. Den studie som refereras här är en studie av kvaliteten i SRU-93 (Andersson, 1996). Studien gjordes i två etapper.

Etapp 1 studerade möjligheterna att helt slopa datainsamlingen för den stora mängden mindre företag (< 50 anställda) och ersätta den med SRU-uppgifter. En avstämning mot Finansstatistiken (FS) gjordes på så sätt att för samtliga företag med <50 anställda i urvalet till FS-93 byttes - för de viktigaste delposterna på resultat- och balansräkning - den SCB-insamlade uppgiften ut mot motsvarande SRU-uppgift. Därefter skedde uppräknings på samma sätt (H-T-skattning) för det modifierade materialet som för den ordinarie FS-93. Skillnaderna mellan de båda skattningarna jämfördes i ett antal redovisningsgrupper (delbranscher).

Analysen visade att SRU-materialet höll en hygglig till god kvalitet (åtminstone med kvaliteten i FS som måttstock). Eftersom SRUs variabelinnehåll hade vissa luckor var det dock inte aktuellt att på detta stadium helt övergå till SRU. Däremot borde SRU-materialet kunna vara en utmärkt hjälpinformation för GREG-estimation (motsvarar estimatorn (3.8)) i FS.

Etapp2 är en studie av GREG-estimation i FS-93 med SRU-93 som hjälpinformation. Slutsatsen av etapp2 blev att SRU skulle vara ett utomordentligt bra material som hjälpinformation för FS. Variansreduktionen blev i många fall 90% eller mer. En studie gav också empiriskt stöd för att GREG-tekniken kan ge ett skydd mot bortfallsbias.

Hur gick det sedan ?

Förhandlingar mellan SCB och RSV ledde så småningom fram till en situation där SRU fr.o.m. årgång 96 kom att få ett variabelinnehåll som var bättre anpassat än tidigare till SCBs behov. En bidragande orsak till detta var sannolikt också att den nya årsredovisningslagen påverkade RSV och SCB så att man från varsitt håll gick i riktning mot bättre samstämmighet. Fortsatta kvalitetstest visade att också kvaliteten i SRU-materialet förbättrats ytterligare sedan årgång 1993.

För årgång 96 av FS (och fr.o.m. årgång 97 Företagsstatistiken) sker datainsamling direkt ur SRU för samtliga företag med <50 anställda. Blankettinsamling görs nu enbart för företag med minst 50 anställda.

GREG-estimatorn används visserligen inte f.n. i Företagsstatistiken men studien av GREG-estimatorns möjligheter var en viktig etapp i utvecklingsarbetet av Finansstatistiken och dess efterföljare Företagsstatistiken.

GREG-tekniken finns också med som en framtida garanti för SCB om RSV skulle besluta minska SRUs variabelinnehåll, av besparingsskäl eller för att man anser att man kan sköta sina egna uppgifter (nämligen att inkräva skatter) utan omfattande SRU-insamling.

METOD-enheten

Metodarbete syftande till kalibrering av vikter i turistdatabasens undersökning (TDB).

Sixten Lundström, METOD

Syftet med arbetet bestod i att förbättra skattningstekniken inom TDB. I TDB skattas omfattningen av resor utförda av invånare i Sverige, övernattningsform, kostnad, etc. Urvalet i undersökningen genomförs som ett två-stegsurval av individer. I första steget tillämpas Random Digit Dialing (RDD) tekniken för urval av hushåll och inom hushåll slumpas en person ut och intervjuas. Skattningar görs för både hushållsparametrar och individparametrar. Undersökningen genomförs av TIMI, Åre. Uppdragsgivare är Turistdelegationen.

TDB använder hjälpinformation både vid skattning av hushållsparametrar och individparametrar i en metod som man kallar RIM-weighting. Metoden är dåligt beskriven, men man anar raking ratio teknik bakom detta begrepp.

Den hjälpinformation som används är antal individer (i) boende i hushåll av olika storlekar, (ii) boende i olika län, (iii) i olika celler av kön*åldersklasser. Vid skattning av hushållsparametrar används uppgifter om antal hushåll (a) av olika storlekar samt (b) i olika län.

Baserat på tidigare undersökningar genomfördes en analys av hur ”stark” hjälpinformationen är. Detta gjordes genom att beräkna multipla determinationskoefficienten mellan ”antal resor” (av olika slag) och hjälpvariablerna inom svarsmängden. Det visade sig vara ytterligt liten i samtliga fall och därför kunde konstateras att hjälpinformationen i stort sett var oanvändbar. Dessutom var det inte möjligt att direkt identifiera bortfallet och därför kunde vi inte studera i vilken grad hjälpinformationen förklarade variationen i svarssannolikheterna. Vi kunde inte heller finna ytterligare hjälpinformation, som skulle vara värdefull.

Uppdragsgivaren hade antagit att svarsmängden kan betraktas som ett OSU från populationen. Om så vore fallet skulle en väntevärdesriktig skattning, trots svag hjälpinformation, ändå kunna erhållas. Analysen av svarsmängden visade att antagandet om OSU är felaktigt. Vi observerade t.ex. att små hushåll var mycket kraftigt underrepresenterade.

Vi såg därför ingen möjlighet att förbättra estimationen och avbröt uppdraget.

MR-avdelningen

Användning av hjälpinformation i Jordbruksekonomiska undersökningen (JEU) i syfte att öka precisionen

Pär Brundell, MR/LEDN

I Jordbruksekonomiska undersökningen (JEU) insamlas uppgifter från jordbruksföretag ur ett företagsekonomiskt perspektiv och man är intresserad av medelvärden för ett tiotal redovisningsgrupper. Medverkan i undersökningen är frivillig och varar över flera år.

En studie gjordes hösten -97 (Brundell, 1997) för att se om hjälpinformation (i form av ett par storleksmått som finns för samtliga företag i urvalsramen) kunde minska urvalsfelen för några utvalda målvariabler. Slutsatsen blev att hjälpinformationen inte var tillräckligt stark för att påverka urvalsfelen annat än i enstaka fall.

Någon studie av att använda hjälpinformationen i bortfallsfelskorrigerande syfte gjordes ej vid detta tillfälle, dock funderar man på att också titta på detta.

Skattningsmetoder i skördeuppskattningar av spannmålsgrödor

Charlotte Jansson, MR/LEDN

Statistik över totalskördar och hektarskördar har under många år byggts enbart på objektiva skördeuppskattningar, vilket innebär att avkastningen på utvalda ytor (med storleken 1 kvadratmeter) skördas, tröskas och vägs. Under åren 1995-1997 gjordes objektiva mätningar i ett permanent urval av 32 skördeområden (SKO) samt en intervjuundersökning, riktad till jordbrukare, som täckte hela landets produktion. (Sverige är uppdelat i 106 stycken SKO.) Omfattningen av de objektiva skördeuppskattningarna i spannmål har alltså successivt minskat under senare år för att helt försvinna 1998.

I och med omläggningen av skördestatistiken 1995 lades arbete ned på att finna ett optimalt skattningsförfarande för beräkning av skördestatistiken. Urvalsarbetet var genomfört då estimationsmetod skulle bestämmas och styrde valet av skattningsmetod. Målet var att kombinera de båda undersökningarna - objektiva mätningar och intervjuundersökningen - för att på bästa sätt utnyttja den information som fanns tillgänglig (Jansson, 1997). De objektiva mätningarna gav tidig estimation

av skörden på riksnivå (i september). Resultat från intervjuundersökningen finns ej tillgängligt förrän i november.

Vid en jämförelse mellan metoderna är det den objektiva undersökningen som anses som den mest korrekta. I undersökningen finns en relativt god uppfattning om de olika osäkerhetskällorna. Risken för mätfel är liten i jämförelse med intervjuundersökningen där brukaren ska ange skördad kvantitet samt vattenhalt för respektive gröda. För den kvantitet som levereras till spannmålsuppköpare har brukaren ofta mycket tillförlitliga uppgifter. Däremot för den lagrade spannmålen på gården (ofta till foderändamål) görs uppskattningar av brukaren, ibland med hjälp av intervjuaren som har enklare mätredskap med sig vid besök.

Analys i form av beräkningar av korrelationskoefficienten mellan resultat från intervjuundersökningen och den objektiva undersökningen (uppdelat per gröda) visade genomgående höga korrelationer (0.7 - 0.9).

I beräkningarna till definitiv statistik på riksnivå användes de objektiva mätresultaten som hjälpinformation vid skattning av mätvariabeln y (hektarskörd från intervjuundersökningen) för att få ökad säkerhet/precision i skattningen. (På regional nivå räckte inte uppgifterna till för en regressions-skattning utan en enklare metod fick användas.)

Från och med 1998 utförs endast intervjuundersökningen, (som en besöksintervju eller en telefonintervju), för att ge statistik över totalskördar och hektarskördar i spannmål.

6. Sammanfattning och kommentarer

Kalibreringsestimaten är generell i så motto att alla estimatorer från den enklaste Horvitz-Thompson-estimatoren till den mest komplicerade generella regressions-estimatoren utgör specialfall. Olika hjälpinformation ger olika specialfall. Styrkan i hjälpinformationen (och förmågan att använda den) avgör hur man lyckas reducera felen i estimaten. Har man ingen hjälpinformation ger inte kalibreringstekniken någon förbättring och har man endast svag hjälpinformation kan endast marginella förbättringar göras. Tillgången till stark hjälpinformation är helt avgörande för att nå framgång.

Hjälpvariablerna hämtas vanligtvis från register, men ibland kan de hämtas från annan källa.

Hittar vi hjälpvariabler med egenskap (i) i (2.1) i reduceras bortfallsbiasen för alla parameterskattningar. Eftersom vi vanligtvis har många målvariabler i en undersökning kommer egenskap (ii) i (2.1) att variera för olika målvariabler; för vissa målvariabler ger kanske inte hjälpvariablerna någon information alls och för andra mycket bra information. Det innebär att reduktionen av felen varierar för olika målvariabler. Även egenskap (iii) i (2.1) varierar för olika redovisningsgrupper. Vid val av hjälpinformation bör man försöka beakta de viktigaste målvariablerna och redovisningsgrupperna.

Utnyttjande av hjälpinformation i estimationen i syfte att reducera urvalsfel och bortfallsbias har troligen den största effekten vid skattning av totaler och mindre effekt vid skattning av kvoter.

I föregående avsnitt har vi gett en kortfattad beskrivning av kalibreringstekniken och vi har också försökt finna och beskriva alla SCB-fall där kalibrering av vikter använts eller endast har prövats. Naturligtvis kan vi ha missat något fall, men vi tror att dessa fall är få. Beskrivningen av fallen visar att kalibreringstekniken är vida känd och att den prövats i många olika situationer. I vissa fall har det stannat med en metodstudie men i andra fall har metoden implementerats.

Beskrivningen av kalibreringsfallen kan ge ett något heterogent intryck. Ett skäl till detta är att fallen har utförts vid olika tidpunkter; det tidigaste fallet avser år 1994 och det senaste avser slutet av år 1999 och att kunskapen kring tekniken har ökat över tiden. Ett annat skäl är att många personer har bidragit till skrivningen.

Tyvärr är det inte möjligt att kvantifiera minskningen av bortfallsbiasen men däremot är urvalsfelen mätbart och det får därför ofta en framträdande roll i studierna. Litteratur inom området visar dock att den mest betydande effekten av utnyttjande av hjälpinformation i estimationen kan vara reduktion av bortfallsbiasen. De fall som beskrivs i avsnitt 5 lider också i hög grad av att inte bortfallsbiasen kan kvantifieras.

En del av fallen utnyttjar endast konsistensegenskapen hos kalibreringsestimatorn. Det görs t.ex. i de fall som är modellberoende (man skulle dock kunna analysera bakomliggande modell och diskutera gjorda antaganden) Kalibreringsestimatorn är mycket flexibel vad gäller utnyttjande av hjälpinformation och är därför intressant även i sådana fall.

För vissa undersökningar är tekniken enkel och billig att använda. De flesta individundersökningarna uppvisar samma bortfallsmönster, svarsandelen är låg i storstäder, bland äldre personer, bland frånskilda, etc. Det är också relativt lätt att utifrån ett saklogiskt resonemang eller baserad på erfarenhet från tidigare undersökningar välja ut hjälpvariabler som samvarierar med de viktiga

målvariablerna. Dessutom vet vi i regel vilka de viktiga redovisningsgrupperna är. När hjälpinformation hämtas från urvalsramen blir kostnaden för kalibreringen och tillhörande dokumentation relativt låg.

I andra fall kräver kalibreringen mer arbete och större kostnad. Detta kan vara fallet när hjälpinformation hämtas från flera olika källor. Önskar vi korstabulera variabler från olika källor krävs att källorna sambearbetas, vilket kan vara mycket besvärligt. Däremot om denna korstabulering inte är nödvändig kan hjälpinformation från olika källor relativt lätt utnyttjas. Då matchas endast svarsfilen med respektive källa och hjälptotalerna hämtas separat från varje källa.

Inom ekonomisk statistik är populationerna relativt föränderliga genom att objekten (företag, arbetsställen, etc.) uppstår, försvinner och delas, vilket skapar många problem i undersökningarna. Därför kräver användning av hjälpinformation i ekonomiska undersökningar en betydligt större arbetsinsats än för exempelvis individundersökningar. Eftersom populationerna förändras snabbt inom ekonomisk statistik är det också vanligare än i andra typer av undersökningar att hjälptotalerna t_x är dåliga approximationer av $\sum_U \mathbf{x}_k$. Om avvikelsen mellan dessa storheter är alltför stor bör man undvika att använda kalibreringstekniken.

Valet av hjälpinformation är inte självklar och därför visar sig en omsorgsfull analys vara nödvändig i många fall. Genomgången av fallen visar att ibland blir vikterna negativa eller mycket stora. I de fallen måste i första hand hjälpvektorn förändras, vilket kan bestå av att grupper slås samman eller variabler plockas bort. Ibland kan det vara riktigt att använda CLAN97-proceduren som begränsar vikterna till ett specificerat intervall.

7. Kalibreringstekniken i framtiden

I stort sett allt arbete med att förbättra skattningar består i att mer effektivt utnyttja hjälpinformation. I urvalsdesignen används hjälpinformation för stratifiering, för allokering av urvalet, som storleksmått i pps-urval, etc. och efter datainsamlingen används hjälpinformation i imputering, i viktning, etc. Tillgången till registerinformation är stor vid SCB och därför är den presumtiva mängden hjälpinformation stor. Detta borde kunna bli en konkurrensfördel för SCB i förhållande till andra institut.

Det arbete som utförs inom SCB för att underlätta utnyttjandet av registerinformation (t.ex. Registerprojektet) kommer förhoppningsvis att göra hjälpinformationen mer lättillgänglig.

Kalibrering av vikter visar sig i många fall ha en god effekt på kvaliteten. I individundersökningar är också utnyttjandet av tekniken billig och därför borde den där bli ett stående inslag. Även i andra undersökningstyper vore det önskvärt om utnyttjandet av hjälpinformation ökade. Det vore också bra om erfarenheter från dessa arbeten dokumenterades väl för att underlätta framtida arbeten.

I vilken mån kalibreringstekniken också reducerar täckningsfelen måste studeras ytterligare.

8. Referenser

Andersson, C. (1997). Continuous labour force surveys: Results and conclusions of the analysis of the effects of the single weight - per household - estimation. Statistics Sweden 1997-06-05).

Andersson, C. och Nordberg, L. (1998). CLAN97 - a SAS-program for computation of point- and standard error estimates in sample surveys. Statistics Sweden.

Andersson, S. (1996). Användning av administrativa datamaterial som hjälpinformation vid estimation - en studie av användningen av SRU-materialet som hjälpmedel till finansstatistiken. *ES-metod nr 39*.

Brundell, P. (1997). Användning av hjälpinformation i Jordbruksekonomiska undersökningen (JEU) i syfte att öka precisionen (utkast). SCB.

Deville, J.C. and Särndal, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association* **87**, 376-382.

Hörngren, Jan och Mirza, Hassan (1998). Sikte mot ett modifierat estimationsförfarande i AKU. SCB.

Jansson, C. (1997). Skattningsmetoder i skördeuppskattningar av spannmålsgrödor (utkast). SCB.

Lundström, S. (1993). Metoder att få konsistens mellan olika skattningar och samtidigt öka precisionen. *R&D Report 1993:6*. SCB.

Lundström, S. (1997), Calibration as a Standard Method for Treatment of Nonresponse, doctoral dissertation. Stockholms universitet.

Lundström, S. and Särndal, C.E. (1999). Calibration as a Standard Method for Treatment of Nonresponse. *Journal of Official Statistics* **15**, 305-327.

Mirza, Hassan (1999). Instruktioner inför det nya estimationsförfarandet i AKU från oktoberundersökningen 1999. SCB.

SCB (1980). *Räkna med bortfall*. En handbok om statistiska metoder i samband med bortfall.

Singh, A.C. and Mohl, C.A. (1996). Understanding calibration estimators in survey sampling. *Survey Methodology* **22**, 107-115.

Särndal, C.E., Swensson, B. and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Wallgren, B. och Wallgren, A. (1998). Att länka ett system av tidsserier - omräkning av Arbetskraftsundersökningarna 1987-92. *Bakgrundsfakta till Arbetsmarknads- och utbildningsstatistiken 1998:2*. SCB.

Förteckning över utkomna R&D Reports

R&D Reports är en för IT-enheten och Metodenheten gemensam publikationsserie, som 1988-01-01 ersatte de tidigare ”gula” och ”gröna” serierna. I serien ingick fram till årsskiftet 1992-93 även **Abstracts** (sammanfattning av metodrapporter från SCB).

Reports published during 1997 and onwards:

- 1997:1 (grön) Bortfallsbarometern nr 12 (*Anti Ahtiainen, Stefan Berg, Mats Bergdahl, Fredrik Granström, Dan Hedlin, Lena Otterskog och Monica Rennermalm*)
- 1997:2 (grön) Quality Concept for Official Statistics - Entry in the forthcoming update of the Encyclopedia of Statistical Sciences, Wiley & Sons (*Eva Elvers and Bengt Rosén*)
- 1998:1 (grön) Preliminär statistik: Nybyggnadskostnader - en simuleringstudie (*Catarina Elffors*)
- 1998:2 (grön) On Inclusion Probabilities for Order Sampling (*Bengt Rosén*)
- 1998:3 (grön) On the Stratification of Highly Skewed Populations (*Dan Hedlin*)
- 1998:4 (grön) Bortfallsbarometer nr 13 (*Per Nilsson, Antti Ahtiainen, Stefan Berg, Mats Bergdahl, Monica Rennermalm och Marcus Vingren*)
- 1998:5 (grön) Estimation from Order π ps Samples with Non - Response (*Bengt Rosén and Pär Lundqvist*)
- 1998:6 (grön) On variance estimation for measures of change when samples are coordinated by a permanent random numbers technique (*Lennart Nordberg*)
- 1999:1 (grön) Täckningsproblem i Registret över totalbefolkning RTB. Skattning av övertäckning med en indirekt metod (*Jan Qvist*)
- 1999:2 (grön) Bortfallsbarometer nr 14 (*Per Nilsson, Antti Ahtiainen, Mats Bergdahl, Tomas Garås, Jan Qvist och Charlotte Strömstedt*)
- 1999:3 (grön) Att mäta statistikens kvalitet (*Claes Andersson, Håkan L. Lindström och Thomas Polfeldt*)
- 2000:1 (grön) Kalibrering av vikter – beskrivning av tekniken och de SCB-fall den prövats i (*Sixten Lundström et al*)

ISSN 0283-8680

Tidigare utgivna **R&D Reports** kan beställas genom Katarina Klingberg, SCB, MET, Box 24 300, 104 51 STOCKHOLM (telefon 08-506 942 82, fax 08-506 945 99, e-post katarina.klingberg@scb.se). **R&D Reports** from 1988-1996 can - in case they are still in stock - be ordered from Statistics Sweden, attn. Katarina Klingberg, MET, Box 24 300, SE-104 51 STOCKHOLM (telephone +46 8 506 942 82, fax +46 8 506 945 99, e-mail katarina.klingberg@scb.se).